Taking a step backward after taking a step forward isn’t a disaster…it’s cha-cha

Robert Brault

Examination committee: Prof. dr. ir. Paul Van Der Meeren (chairman)

Prof. dr. ir. Wim Soetaert (promotor)

Dr. ir. Inge Van Bogaert (co-promotor)

Prof. dr. ir. Monica Höfte (secretary)

Prof. dr. Ibrahim Banat

Prof. dr. Bart Devreese

Dr. ir. Sofie De Maeseneire

Promotors: Prof. dr. ir. Wim Soetaert (promotor)

Dr. ir. Inge Van Bogaert (co-promotor)

Center of Expertise-Biotechnology and Biocatalysis

Department of Biochemical and Microbial Technology

Faculty of Bioscience Engineering

Ghent University

Dean: Prof. dr. ir. Guido Van Huylenbroeck

Rector: Prof. Paul Van Cauwenberge

Sophie Roelants was supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) 'Biosurf' grant number IWT80050 and the European FP7 project 'Biosurfing', grant number 289219

Ghent University

Faculty of Bioscience Engineering

Department of Biochemical and Microbial Technology

ir. Sophie Roelants

Starmerella bombicola as a platform organism for the production of biobased compounds

Thesis submitted for the fulfilment of the requirements for the degree of

Doctor (PhD) in Applied Biological Sciences

Academic year: 2012-2013

Titel van het doctoraatsproefschrift in het Nederlands:

Starmerella bombicola als platform organisme voor de productie van biogebaseerde moleculen.

Cover design and illustrations: Stijn Verweire and Ariane Mol

To refer to this thesis:

Roelants, S. (2013) *Starmerella bombicola* as a platform organism for the production of biobased compounds . PhD-thesis, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium.

ISBN-number: 978-90-5989-620-8

LIST OF ABBREVIATIONS

at	Acetyltransferase gene
BS	Biosurfactant
bp	Basepairs
CBL	Cellobioselipid
CFU	Colony forming units
CMC	Critical micelle concentration
CSL	Corn steep liquor
cyp	Cytochrome P450 mono oxygenase gene
DSP	Down stream processing
EDTA	Ethylene Diamine Tetraacetic Acid
ELSD	Evaporative Light Scattering Detector
FA-OH	Hydroxylated fatty acid
5'-FOA	5-fluoroorotic acid
gDNA	Genomic DNA
gt	Glycosyltransferase gene
HDAC	Histon deacetylase
HLB	Hydrophobic/hydrophilic balance
HPLC	High Pressure Liquid Chromatography
IP	Intelectual property
kb	Kilobases
KO	Knock Out
lip	Lipase gene (gene responsible for lactonisation of SLs)
LB	Luria Bertani (medium)
Mb	Megabases
MELs	Mannosylerythritollipids
min	minutes
MTA	Material transfer agreement
NaCl	Sodiumchloride
OD	Optical density
OE	overexpression
orf	Open reading frame
PCR	Polymerase chain reaction
P*gki*	Promotor of phosphoglycerate kinase
P*gapd*	Promotor of glyceraldehyde 3 phosphate dehydrogenase
PHA	Polyhydroxyalkanoates
RLs	rhamnolipids
rpm	Revolutions per minute
sec	seconds

SCE	Sorbitol Citrate EDTA (buffer)
SD	Synthetic dextrose medium
SM	Secondary metabolites
SLs	sophorolipids
TK	Tyrosine kinase
ura3	orotidine 5-phosphate decarboxylase gene
UA	Ustilagic acid (cellobioselipid variant)
UDP-glucose	Uridinediphosphate-glucose (activated sugar)
WT	Wild type
YPD	Yeast peptone dextrose medium

OUTLINE

The organism of interest in this manuscript is the yeast *Starmerella bombicola*, which is the teleomorph of *Candida bombicola* (Rosa and Lachance, 1998). At the 'International Botanical Congress in Melbourne (2011) it was decided that starting from the 1st of January 2013 one fungus can only have one name, being the name of the teleomorph. Whenever *Candida bombicola* is thus mentioned in older references, this refers to the same species and in this manuscript and future publications the name of the teleomorph: *Starmerella bombicola* will be used.

This yeast attracted interest for its ability to produce high amounts (above 400 g/L) of a mixture of glycolipid molecules (sophorolipids) with biosurfactant properties. In Chapter I of this manuscript other glycolipids and more generally, biosurfactant producing organisms (both pro- and eukaryotic), their genetic background, the regulatory mechanisms influencing biosynthesis and the biotechnological opportunities these microorganisms present, will be discussed. This will furthermore be discussed in the larger context of secondary metabolite production by eukaryotes as many analogies exist with glycolipid production by yeasts.

These sophorolipids produced by *Starmerella bombicola* are one of the only biosurfactants that are commercialized nowadays. Although numerous publications and patent applications deal with the optimization of fermentation parameters to steer sophorolipid productivity and - composition, very little is known about the physiological role and (molecular) regulation of the production of these secondary metabolites. Moreover, almost no genetic information (besides rDNA sequences) is publically available, let alone the existence of molecular techniques to engineer this industrially important though unconventional yeast. The challenge to do something about this was accepted by researchers at InBio.be and the first achievement was the development of a transformation and selection protocol for *S. bombicola* (Van Bogaert *et al.*, 2008b) a few years ago.

The work presented in this manuscript was performed in the framework of the national IWT-SBO project 'BIOSURF' and European FP7 project 'BIOSURFING', which shared a common goal: the transformation of *Starmerella bombicola* into a platform organism for the production of tailor-made biomolecules. To do so a thorough knowledge of the conditions promoting sophorolipid production in addition to genetic information of the yeast were required. The complete *S. bombicola* genome was thus sequenced and annotated in the framework of the IWT-SBO project, which led to the discovery of the sophorolipid biosynthetic gene cluster.

In the first part of this manuscript (Chapter II and Chapter III) the sophorolipid metabolism of *S. bombicola* is investigated. In Chapter II the genetic background and possible regulation of

sophorolipid production is discussed, whereas in Chapter III one of the possible physiological roles of sophorolipid production, i.e. the creation of an extracellular carbon source, is investigated.

The second part of this manuscript (Chapter IV and Chapter V) describes the development of a portfolio of molecular tools for the further genetic engineering of *S. bombicola*. In Chapter IV the development of a reporter system and a method for the reuse of the auxotrophic *ura3* marker is discussed, whereas in Chapter V de development of a qPCR platform is described. The latter will not only serve for the further investigation of sophorolipid metabolism in the wild type organism, but can also be used to determine possible bottlenecks in new strains producing (new-to-nature) biosurfactants.

In the third and last part of this manuscript proof of concept is delivered that this yeast can be transformed into a platform organism for the production of interesting biomolecules. In Chapter VI the possibility to fundamentally modify the sophorolipid mixture by applying genetic engineering illustrates the vast biotechnological possibilities this yeast offers. In the seventh and last chapter the conversion of this biosurfactant producer into a bioplastic producer by making use of its efficient SL machinery is described. This second example illustrates that this yeast can not only be used for the production of tailor-made-biosurfactants, but can also serve for the production of completely different biological molecules with interesting perspectives for the biobased economy.

TABLE OF CONTENTS

CHAPTER I:

LITERATURE REVIEW

Parts of this chapter are in preparation for publication as two mini-reviews in "Applied Microbiology and Biotechnology" (proposal accepted).

Roelants, S., De Maeseneire L. S., Ciesielska, K., Saerens, K., Van Bogaert, I. and Soetaert, W. Biosurfactant gene clusters in eukaryotes: regulation and biotechnological potential.

Roelants, S., De Maeseneire L. S., Gheys R., Van Bogaert, I. and Soetaert, W. Biosurfactant gene clusters in prokaryotes: regulation and biotechnological potential.

Chapter I.

LITERATURE REVIEW

I.1. Introduction

Secondary metabolites produced by microorganisms consist of a wide range of molecules, which have found applications in very diverse sectors. One of the best known examples is the production of penicillin by certain fungi, which formed the basis of the modern era of antibiotic discovery. Production of secondary metabolites is mostly not necessary for central metabolic pathways of living cells, but can play an important role in an organism's functioning and the industrial relevance of the producing strains. Secondary metabolites are often only produced during a specific life cycle phase or under certain environmental conditions and can help cells to cope with stress or to compete with other microorganisms. For endophytes alone -microorganisms residing in the tissue of living organisms- this accounts to 313 compounds, mainly alkaloids (Gutierrez *et al.*, 2012). Other secondary metabolites produced by microorganisms include benzopyrans, anthraquinones terpenoids and polyketides (Barros *et al.*, 2012). These molecules span a wide variety of structures and find applications in the medical field as antibiotics, antifungals, antiparasitics, cholesterol-lowering agents, antioxidants or as immuno suppressants. Another family of very diverse secondary metabolites produced by a wide variety of microorganisms are biosurfactants. These surface-active agents are capable of reducing surface and interfacial tension at the interfaces between immiscible liquids, solids and gases allowing them to disperse readily as emulsions in water or other liquids. Their biodegradability and the fact that they can be produced from renewable resources gives them an advantage over their chemical counterparts and may therefore make them suitable to partly replace chemicals (Banat *et al.*, 2000). Biosurfactants have been described to find (potential) applications in the medical world, personal care sector, mining processes, food industry, cosmetics, crop protection, pharmaceuticals, bio-remediation, household detergents, paper and pulp industry, textiles, paint industries, etc. (Banat *et al.*, 2000). Especially glycolipid biosurfactants, like sophorolipids (SLs), rhamnolipids (RLs), mannosylerythritol lipids (MELS) and cellobioselipids (CBLs) have been described to provide significant opportunities to replace chemical surfactants in several sectors (Marchant and Banat, 2012). Despite the obvious

opportunities represented by microbial production of biosurfactants, large chemical companies for a long time did not show much interest in such biobased technologies (Banat *et al.*, 2000). However, growing environmental awareness and rising oil prices combined with the fact that our dependency on this finite source of energy and chemical compounds could lead to a catastrophe once we run out of oil, has resulted in a new trend. Large companies like BASF, Evonik-Degussa, Unilever, Henkel, Cargill, etc. have initiated R&D projects focused on green alternatives of chemical surfactants. Especially glycolipid biosurfactants are described in patent applications filed by these companies, which emphasizes the industrial relevance of these compounds.

The major two factors currently limiting the penetration of bio-surfactants into the market are firstly the limited structural variety and secondly the rather high production price. However:

- researchers recently suggested that the natural diversity of molecules produced by micro-organisms is potentially much larger than appreciated from culture-based studies (Reddy *et al.*, 2012) and when it comes to structural diversity of secondary metabolites, we are probably only able to see the tip of the iceberg at this point. In practice this means that nature offers the possibility of producing a vast variety of complex structures, which cannot easily be synthesized by chemical processes (Banat *et al.*, 2000). As micro-organisms are already considered one of the world's richest sources of bioactive natural products, such predictions could have profound implications for the future.
- Furthermore biotechnological approaches and genetic engineering of natural producers could not only result in more efficient recombinant producers, but also in a diversification of the spectrum of available biosurfactants. One of the critical factors for the application of biosurfactants in industry is a detailed knowledge of their genetics, as this may hold the key to their future (enhanced) economical production using enhanced microbial strains.

This work focuses on the efficient natural biosurfactant producer *Starmerella bombicola,* and in particular on the biological opportunities this yeast offers. Therefore this literature review aims to provide the reader with a framework and will hence focus on the microbial production of biosurfactants and the regulation thereof (I.2.) Furthermore, both the natural (I.3) and industrial relevance (I.4) of prokaryotic and eukaryotic biosurfactants will be addressed.

Finally, the regulation of the some of the more extensively studied non-biosurfactant fungal secondary metabolites will be briefly discussed (I.5).

I.2. Biosynthesis of biosurfactants and regulation thereof

Biosurfactants (BS) are a class of secondary metabolites representing a wide variety of structures that can be produced from renewable feedstock by a wide variety of micro-organisms, both prokaryotes and eukaryotes. Many bacterial species secrete high-molecular-weight biosurfactants consisting mainly of polysaccharides, lipoproteins, or lipopolysaccharides. Low-molecular-weight biosurfactants are generally glycolipids or lipopeptides. The hydrophilic part of BS is usually composed of sugars, amino acids, or polar functional groups like carboxylic acid groups. The hydrophobic part is typically an aliphatic (hydroxy) fatty acid. Surfactants are generally classified according to the charge carried by their polar groups (head) into cationic, anionic, amphoteric, and nonionic surfactants. Biosurfactants are also classified according to their molecular structure into mainly glycolipids (e.g. rhamnolipids (RLs), sophorolipids (SLs), etc.), lipopeptides (e.g. surfactin), polymeric biosurfactants (e.g., emulsan and alasan), fatty acids ((e.g. 3-(3-hydroxyalkanoyloxy) alkanoic acids (HAAs)), and phospholipids (e.g., phosphatidylethanolamine) (Desai and Banat, 1997; Lang and Wullbrandt, 1999). An overview of some representative examples of biosurfactants and the corresponding organisms is given in Table I. 1. It is not the intention to provide the reader with a complete overview of all the discovered biosurfactants, but to give an idea about the structural and microbial variety of biosurfactants and their producers respectively.

Unraveling the genetics behind the biosynthesis of these interesting biological compounds is indispensable for the tinkering, fine tuning and rearrangement of these biological pathways with the aim of obtaining higher yields and a more extensive structural variety. In the next few pages the biosynthesis and regulation of some prominent and less prominent members of the family of the biosurfactants will be discussed.

Table I. 1 Examples of biosurfactants and their microbial origin.

	Biosurfactant	**Selection of producing micro-organisms**
GLYCOLIPIDS	Trehalose lipids	*Arthrobacter paraffineus, Corynebacterium spp., Mycobacterium spp., Rhodococus erythropolis, Nocardia sp.*
	Rhamnolipids	*Pseudomonas aeruginosa, Pseudomonas spp., Serratia rubidea*
	Sophorose lipids	*Candida apicola, Starmerella bombicola, Rhodutorola bogoriensis*
	Glycolipids	*Alcanivorax borkumensis, Arthrobacter sp., Corynebacterium sp., R. erythropolix, Serratia marcescens, Tsukamurella sp.*
	Cellobiose lipids	*Ustilago maydis, Pseudozyma flocculosa*
	Polyol lipids	*Rhodotorula glutinis, Rhodotorula graminus*
	Diglycosyl diglycerides	*Lactobacillus fermentii*
	Lipopolysaccharides	*Acinetobacter calcoaceticus (RAG1), Pseudomonas sp., Starmerella lipolytica*
	Mannosylerythritol lipids	*Pseudozyma antarctica, Ustilago maydis*
	Oligosaccharide lipids	*Tsukumurella sp.*
	Fructose-lipids	*Arthrobacter sp., Corynebacterium sp., Nocardia sp., Mycobacterium sp.*
LIPOPROTEINS	Arthrofactin	*Arthrobacter sp.*
	LichenysinA	*Bacillus licheniformis*
	Serrawettin	*Serratia marcescens*
	Gramicidins	*Bacillus brevis*
	Polymyxins	*Bacillus polymyxa*
	Lichenysin B, C	*Bacillus sp.*
	Surfactin	*Bacillus subtilis, Bacillus pumilus, Bacillus licheniformis*
	Surfactant BL86	*Bacillus licheniformis*
	Plipastatin	*Bacillus licheniformis*
	Massetolides	*Pseudomonas fluorescens*
	Iturin	*Bacillus amyloliquefaciens, Bacillus subtilis*
	Bamylomycin	*Bacillus amyloliquefaciens*
	Halobacillin	*Marine Bacillus sp.*
	Isohalobacillin	*Bacillus sp.*
	Bioemulsifier	*Bacillus stearothermophilus, Starmerella lipolytica*
PHOSPHOLIPIDS	Phospholipids	*Acinetobacter sp., Thiobacillus thiooxidans*
NEUTRAL LIPIDS	Sulfonylipids	*Thiobacillus thiooxidans, Corynebacterium alkanolyticum*
& FATTY ACIDS	Fatty acids	*Capnocytophage sp., Penicillium spiculisporum, Corynebacterium lepus, Arthrobacter paraffineus, Talaramyces trachyspermus, Nocardia erythropolis*
POLYMERIC SURFACTANTS	Emulsan	*Acinetobacter calcoaceticus, Acinetobacter venetianus*
	Liposan	*Candida lipolytica*
	Biodispersan	*A. Calcoaceticus*
	Carbohydrate-protein-lipid	*Pseudomonas fluorescens, Debaryomyces polymorphis*
	Protein PA	*Pseudomonas aeruginosa*
	Bioemulsan	*Gordonia sp.*
	Circulocin	*Bacillus circulans*
	AP-6	*Pseudomonas fluorescens*

I.2.1. *Biosurfactants produced by prokaryotes*

I.2.1.1. *Rhamnolipids*

Glycolipid biosurfactants which are comprised of L-rhamnose and β-hydroxyalkanoic acid residues are termed rhamnolipids (RL) (Figure I. 1). They were initially found as exoproducts of the opportunistic pathogen *Pseudomonas aeruginosa* (Jarvis and Johnson, 1949) and the structure was elucidated as a mixture of four congeners: alpha-L-rhamnopyranosyl-alpha-L-rhamnopyranosyl-beta-hydroxydecanoyl-beta-hydroxydecanoate (Rha-Rha-C(10)-C(10)), alpha-L-rhamnopyranosyl-alpha-L-rhamnopyranosyl-beta-hydroxydecanoate (Rha-Rha-C(10)), and their corresponding mono-rhamnolipid congeners Rha-C(10)-C(10) and Rha-C(10) (Edwards and Hayashi, 1965; Itoh *et al.*, 1971; Syldatk *et al.*, 1985). Nowadays it is known that about 60 variants, varying in the lengths of the incorporated fatty acids, are produced in mixtures at different concentrations by various *Pseudomonas* species and bacteria belonging to several families, classes or even phyla. For example, various *Burkholderia* species have been shown to produce rhamnolipids that have longer alkyl chains than those produced by *P. aeruginosa* (Dubeau *et al.*, 2009).

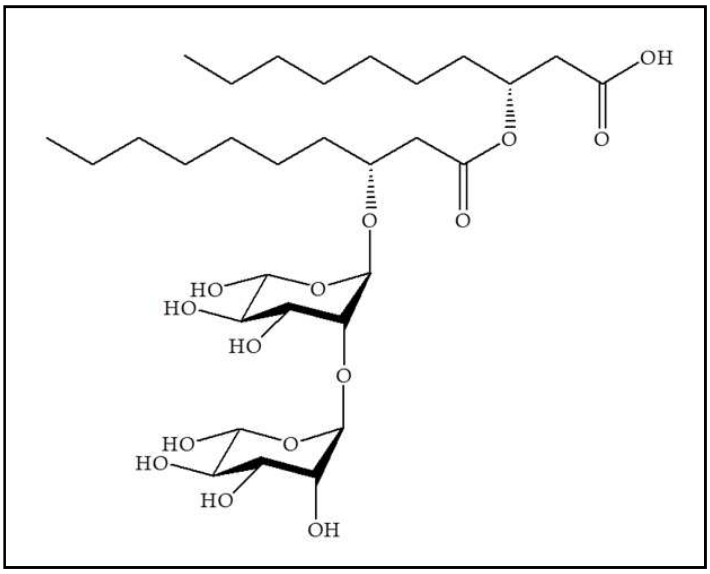


Figure I. 1 Di-rhamnolipid produced by *Pseudomonas aeruginosa*

Biosynthesis of RLs in *P. aeruginosa* occurs through three major synthesis steps (Burger *et al.*, 1963; Rahim *et al.*, 2001; Soberon-Chavez *et al.*, 2005). First RhlA (encoded by *rhl*A), an inner-membrane bound protein is involved in the biosynthesis of the fatty acid dimer (β-hydroxyalkanoyl-β-hydroxyalkanoyl-ACP (HAA-ACP or CoA)) from β-hydroxy fatty acid precursors (Lepine *et al.*, 2002; Deziel *et al.*, 2003; Zhu and Rock, 2008). Secondly, mono-RL

are synthetized by the action of the membrane-bound RhlB rhamnosyltransferase (encoded by *rhl*B) which uses dTDP-L-rhamnose and a HAA molecule as precursors (Ochsner *et al.*, 1994a). The *rhlA* and *rhlB* gene are part of the same operon and are hence transcribed together. A third and last step is catalysed by the action of a second rhamnosyltransferase (encoded by *rhl*C), located elsewhere in the genome, that uses the formed mono-RLs - together with dTDP-L-rhamnose- as substrates to produce di-RLs (Rahim *et al.*, 2001).

Until now no detailed information about the exact mechanisms that lead to the secretion of RLs to the extracellular space is known. RhlA was hypothesized to be involved in synchronized RL synthesis and transport (Muller and Hausmann, 2011). However, the *PA1131* gene, located in the same operon as the *rhlC* gene, encodes a probable transporter of the major facilitator superfamily (MFS) containing 11 transmembrane domains, thus the *PA1131* gene product is likely to be involved in RL secretion (Rahim *et al.*, 2001).

RL biosynthesis in *P. aeruginosa* is controlled by an N-acyl homoserine lactone (AHL) dependent quorum sensing (QS) system (cfr. Figure I. 2). The AHLs mediate the cellular responsiveness towards environmental signals and cell density in particular. Induction of the *rhl*AB operon directly depends on the quorum-sensing transcription activator RhlR which, complexed with the autoinducer N-butyryl homoserine lactone (C4-HSL) or *Pseudomonas* autoinducer PAI-2 synthesized by RhlI (Pearson *et al.*, 1995), positively regulates *rhl*A and B by binding a 20 bp palindromic sequence (lux-box) in the promotor region of *rhl*A (Pearson *et al.*, 1997). These regulatory genes, *rhl*R and *rhl*I, are located directly downstream of the *rhl*AB operon and are transcribed in the same direction but from a different promotor (Ochsner *et al.*, 1994b; Ochsner and Reiser, 1995). They are in turn influenced by another quorum sensing system encoded by *las*I and *las*R (Latifi *et al.*, 1995) (Figure I. 2). A third *P. aeruginosa* signal, 2-heptyl-3-hydroxy-quinolone or *Pseudomonas* quinolone signal (PQS) also plays a significant role in the transcription of multiple *P. aeruginosa* virulence genes (not shown in Figure I. 2). PQS is intertwined with the *P. aeruginosa* quorum sensing hierarchy as transcription of *pqsA* and subsequent production of PQS is regulated by the las and rhl quorum sensing system and PQS production is furthermore dependent of the ratio of PAI-1 to PAI-2 suggesting a regulatory balance between QS systems (McGrath *et al.*, 2004)

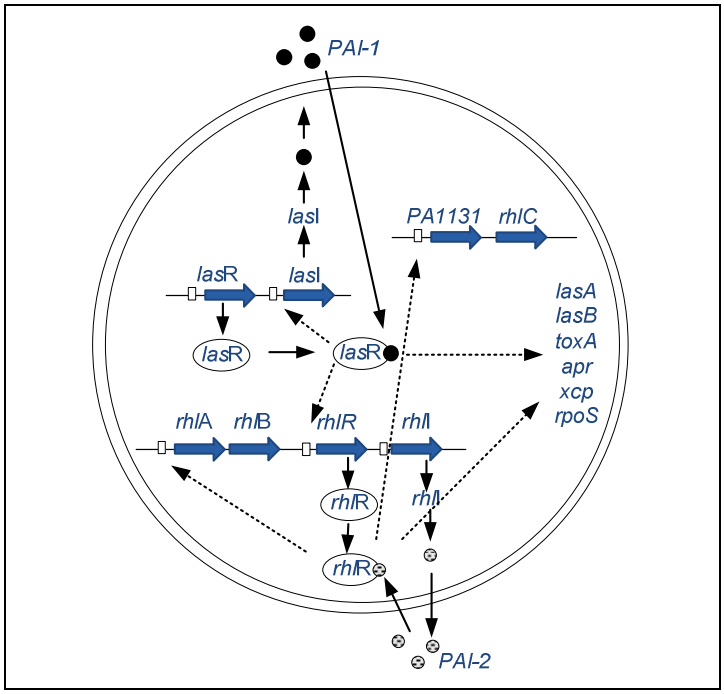


Figure I. 2 Simplified schematic representation of the regulation of rhamnolipid production in *P. aeruginosa* by the two hierarchical quorum sensing systems *rhl* and *las*. Blue arrows represent genes on the *Pseudomonas* chromosome. Small white boxes represent the 'lux-box' in the promoter regions. Continuous arrows show protein production from a gene and the dotted arrows show the places where the autoinducer bound to its cognate transcription activator positively controls gene expression; according to (Sullivan, 1998; Muller and Hausmann, 2011).

Regulation begins with the activation of *las*R transcription, which encodes a transcriptional activator, by autoinduction from its own system and by cyclic AMP (cAMP) levels, as indicated by the presence of both a lux-box and a binding consensus sequence for the cAMP receptor protein in the *las*R promotor region (Albus *et al.*, 1997). The LasI protein synthesizes *N*-oxododecanoyl homoserine lactone (OdDHL, 3OC12HSL or also called *Pseudomonas* autoinducer PAI-1) (Pearson *et al.*, 1994; Winson *et al.*, 1995). When PAI-1 autoinducer concentrations are high due to high cell density, LasR binds with the autoinducer (PAI-1) and regulates *rhl*R transcription by binding the 'lux-box' in its promotor. RhlR subsequently binds to its autoinducer PAI-2 which, as said, leads to activation of the *rhl*AB operon (Sullivan, 1998). The *las* system furthermore controls the *rhl* system post-translationally because PAI-1 blocks PAI-2 from binding RhlR until enough PAI-2 and/or RhlR is produced to overcome the blocking (Pesci *et al.*, 1997). The interaction between both systems is described as a

hierarchical quorum sensing cascade with LasR and I as the master regulators (Pesci and Iglewski, 1997). Aside from the above mentioned positive regulation by the *las* system, a basal level of *rhl*R expression is dependent upon mechanisms other than quorum sensing (Pearson *et al.*, 1997). Although located elsewhere in the genome, expression of the *PA1131-rhlC* operon is independently transcribed though coordinately regulated with the *rhl*AB operon by the same quorum sensing system (Rahim *et al.*, 2001).
In addition to the complex regulation by the hierarchical quorum sensing cascade, induction of *rhl*AB is also dependent on the culture's growth phase and on environmental conditions:

- Induction does not occur in the logarithmic phase of growth even in presence of RhlR and C4-HSL (Das *et al.*, 2008).
- Transcription of *rhl*AB genes involves σ^{54}, as the region upstream of the *rhl*A transcription start contains a putative σ^{70} and σ^{54}-type promotor, which is over-expressed under nitrogen limiting conditions (Ochsner *et al.*, 1994a; Pearson *et al.*, 1997). However N-limiting conditions do not favor RL production per se: while some reports show that RL production starts after nitrogen exhaust, other show that even though nitrogen was still present in excess, RL productivity was already high (Muller and Hausmann, 2011).
- Rhamnolipid production is promoted by enhanced C/N ratio (Winson *et al.*, 1995)
- The use of ammonium as a nitrogen source instead of nitrate (Deziel *et al.*, 2003) leads to a decrease in *rhl*A expression.
- Rhamnolipid production is inhibited by an excess of available iron (Guerra-Santos *et al.*, 1984),
- RLs are synthesized in oil contaminated environments, which is not that easy to explain: although cell densities are presumably low under these conditions, production of RL enhances degradation of hydrocarbons. It was thus suggested that other regulatory factors triggered by e.g. cAMP levels or environmental conditions such as nitrogen availability can override the quorum sensing mechanism (Sullivan, 1998).

The genes responsible for RL synthesis in *Burkholderia thailandensis* and *Burkholderia pseudomallei* were recently described by (Dubeau *et al.*, 2009) and while the molecular regulation of RL biosynthesis for *P. aeruginosa* is in large part known, these studies propose a different regulation for *Burkholderia* sp. These bacterial species harbor two paralogous *rhl* operons carrying homologues of *rhl*A, *rhl*B, and *rhl*C. However, unlike in *P. aeruginosa*,

these three genes are grouped together in the same operon. Interestingly the ratio of di-RLs to mono-RLs found in *B. thailandensis* is approximately 13:1, whereas that in *P. aeruginosa* is only 4:1 (Dubeau *et al.*, 2009). It is hypothesized that the fact that *P. aeruginosa* harbors *rhl*A and *rhl*B in one operon and *rhl*C in another could result in different levels and timing of expression of the first and second rhamnosyltransferase (RhlB and RhlC, respectively), thus decoupling both activities. *Burkholderia* species on the other hand co-express *rhl*B and *rhl*C, which might favor the simultaneous production of all the enzymes of the biosynthetic pathway in the same stoichiometric ratio (Das *et al.*, 2008). These gene clusters also harbor several genes with a predicted function as multidrug resistance proteins (Dubeau *et al.*, 2009), which are thus most likely involved in RL secretion, as such proteins have been reported to be involved in the secretion of other glycolipid biosurfactants (Teichmann *et al.*, 2007; Van Bogaert *et al.*, 2013).

I.2.1.2. *Lipopeptides*

Cyclic lipopeptides generally consist of a cyclic peptide or a peptide macrolactone of L and D amino acids linked to a variable hydrophobic alkyl chain. These lipopeptides are a large class of biosurfactants, which are surface-active, have antimicrobial activity, are involved in biofilm formation and swarming, function as pheromones in quorum-sensing and even act as a killing factor effectuating programmed cell death in sister cells (Stein, 2005). This class consists amongst others of surfactins (heptapeptides interlinked with a β-hydroxy fatty acid (C_{13}-C_{15}) to form a cyclic lactone-ring structure (Figure I. 3), fengycins (lipodecapeptides), iturins (lipoheptapeptides) and similar compounds. In general, each lipopeptide family consists of several different isoforms arising from variable lengths and branching of the alkyl chains and conservative changes in the amino acids or stereochemistry (Li *et al.*, 2010).

Lipopeptides are produced by a variety of bacterial genera, including *Pseudomonas* spp. and *Bacillus* strains. Gene clusters responsible for the biosynthesis of said biosurfactants were amongst others identified in *Bacillus subtilis* strain AS 43.3 and *Bacillus amyloliquefaciens* (Dunlap *et al.*, 2012). The genome of the latter contains five gene clusters encoding non-ribosomal peptide synthetase (NRPS) responsible for the synthesis of different lipopeptide variants. Three of these clusters were also present in the genome of the former strain and comprise of genes that encode three predicted lipopeptides: surfactin, iturin and fengycin (cfr. Figure I. 4). The remaining two identified gene clusters possess genes that were predicted to encode the antibacterial dipeptide bacilysin and the siderophore bacillibactin. All of these

compounds were found in the culture supernatants, confirming that the genes are functional and metabolically active. Surfactin, the best known member of this group of biomolecules is one of the most powerful biosurfactants as it can lower the surface tension of water (72.8) to 27.9 mN/m at concentrations of only 0.005% (Arima *et al.*, 1968) (cfr. Figure I. 3).

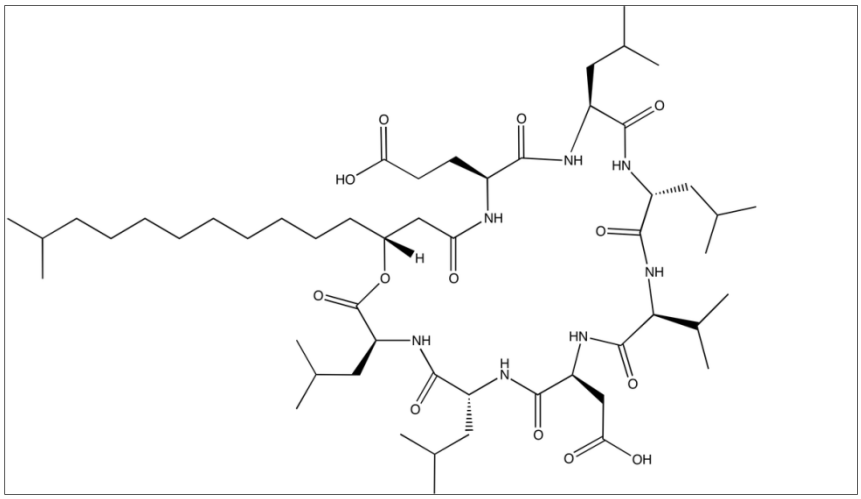


Figure I. 3 The lipopeptide surfactin produced by *Bacillus subtilis*

- Surfactin

Surfactin biosynthesis in *B. subtilis* is catalyzed non-ribosomally by the large multi-enzyme surfactin synthetase complex consisting of three protein subunits (srfA-C) which employs the mechanism of nonribosomal peptide synthetases (NRPSs) by incorporating amino acids into the lipopeptide biosurfactant by forming ester and amide bonds. There is an ATP dependant adenylation domain, which activates amino acids, a condensation domain catalyzing peptide bond formation, an epimerization domain epimerizing amino acids before their addition to the growing peptide chain and a thioesterase domain which cleaves the growing peptide chain. The folding of the peptide chain is subsequently stabilized by an intramolecular lactonization possibly involving a second thioesterase (sfrA-D) (Peypoux *et al.*, 1999). The proteins required for biosynthesis of the heptapeptide amino acid moiety of surfactin are encoded by four ORFs in the *srf*A operon (15 kb) (cfr. Figure I. 4), namely *srf*AA, *srf*AB, *srf*AC and s*rf*AD or s*rf*A-TE. SrfAA of the surfactin peptide synthetase contains the amino acid activation domains for the addition of Glu, Leu and D-Leu, SrfAB for addition of Val, Asp and D-Leu, and SrfAC for the addition of Leu (Galli *et al.*, 1994). *Srf*AC also encodes a thioesterase of a type I motif responsible for peptide termination (deFerra *et al.*, 1997). While

the thioesterase SrfAD is not essential for surfactin biosynthesis, the other three ORFs are absolutely essential for this process as was proven by deletion analysis (Sullivan, 1998). The *srfA* operon furthermore contains the *Sfp* gene which encodes a 4'-phosphopantetheinyl transferase required for activation of surfactin synthetase by posttranslational modification (Lambalot *et al.*, 1996). It posttranslationally converts inactive apoenzyme peptide synthetases into the active holoenzymes (Finking and Marahiel, 2004) and is absolutely essential for surfactin production. Finally, the operon contains *com*S, a gene lying within and out-of-frame with *srf*B, which is also dependent upon the *srf*A promotor for expression (Dsouza *et al.*, 1994; Hamoen *et al.*, 1995). ComS is required for competence (the ability of cells to bind and take up exogenous DNA) and could thus give a hint about the natural role of these biosurfactants (cfr. I. 3). Regulation of surfactin production by *B. subtilis* has been found to occur through two cell density-responsive mechanisms (quorum sensing), which are not based on homoserine lactone (cfr. RL biosynthesis), but on the peptide pheromone ComX (Menkhaus *et al.*, 1993) and the competence and sporulation stimulating factor (CSF) (Solomon *et al.*, 1995).

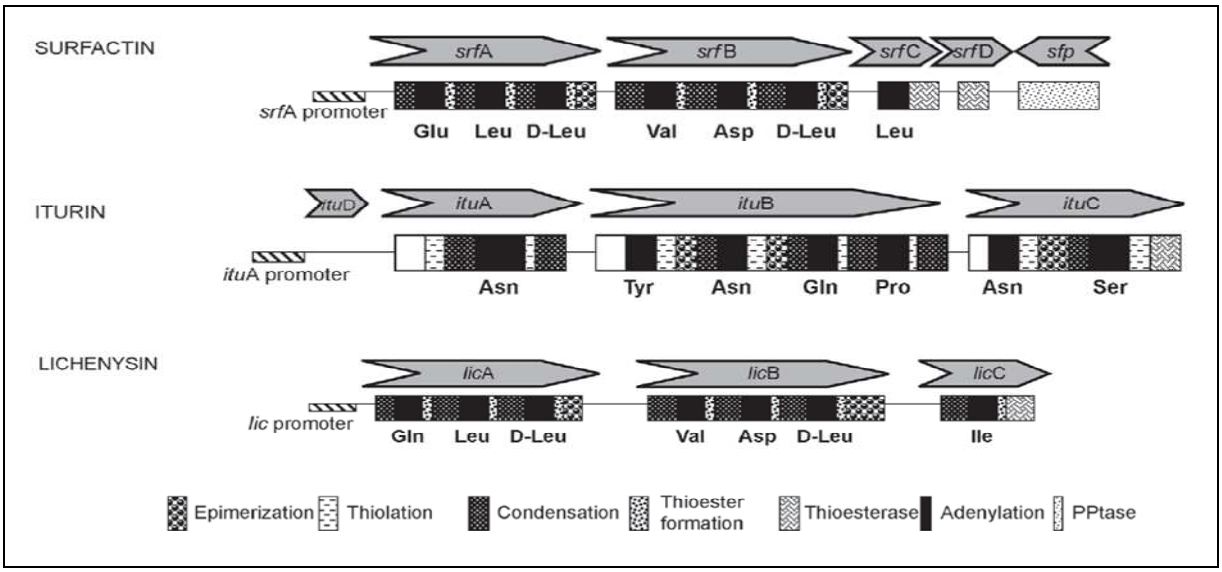


Figure. I. 4 Structural organization of the operons encoding various lipopeptide biosurfactant synthetases. These operons show a high degree of structural similarity. The ***SrfA* operon** (>15 kb) of *Bacillus subtilis* comprises four ORFs: *srf*A, *srf*B, *srf*C and *srf*D, which encode the surfactin synthetase enzyme. It also contains the *sfp* gene encoding phosphopantetheinyl transferase enzyme, required for posttranscriptional modification of surfactin (Das *et al.*, 2008). The genes consist of multiple regions coding for functional domains namely, epimerization, adenylation, condensation, thioester formation, etc. in the corresponding lipopeptide biosurfactant synthetase enzymes (Das *et al.*, 2008).

In the initiation reaction of surfactin biosynthesis, the β-hydroxy fatty acid substrate is transferred to SrfA from coenzyme A, leading to the formation of β-hydroxyacyl-glutamate (Steller *et al.*, 2004). (Kraas *et al.*, 2010) showed that the condensation domain of the initiation module, of which the formation was shown to be stimulated by SrfD (Steller *et al.*, 2004), catalyzes the transfer of the CoA-activated β-hydroxy fatty acid to the peptidyl carrier protein-bound glutamate. Furthermore, another gene located elsewhere in the genome was recently found to be involved in surfactin biosynthesis. The respective gene (*ybd*T), encoding a cytochrome P450 enzyme, was found to be responsible for β-hydroxylation of the incorporated fatty acids (Youssef *et al.*, 2011). These hydroxylated fatty acids are subsequently activated by the action of two acyl CoA ligases as such supplying the building blocks necessary for surfactin production (Kraas *et al.*, 2010).

- Lichenysin

Lichenysin produced by *Bacillus licheniformis* structurally resembles surfactin from *Bacillus subtilis*. The typical difference between the two lipopeptides is the presence of a Gln residue in position 1 or an Asn residue in position 5 of the peptide sequence instead of Glu or Asp in surfactin and the lipophilicic moiety which consist of a mixture of straight and branched beta-hydroxy fatty acid. These variations cause significant changes in the properties of lichenysin compared to surfactin (Li *et al.*, 2010). Two types of lichenysin, surfactant BL86 and lichenysin B, have recorded the lowest CMC ever (10 mg/L) for any surfactant under optimal conditions (Nerurkar, 2010). About eight types of lichenysin isoforms were reported to be produced by *Bacillus* strains (Li *et al.*, 2010). They are named lichenysin A, B, C, D, E, F, G and surfactant BL86 based on species specific variations.

Structural genes required for lichenysin synthesis are organized in the *lch*A (lic) operon and show high sequence homology with those of surfactin (Yakimov *et al.*, 1998). Furthermore, the *lch*A promotor region contains an imperfect inverted repeat with homology to the dyad symmetry of the *sfr*A promotor comA-box. Therefore, it can be expected that the biosynthesis of both substances follows similar pathways (Sullivan, 1998). The *lch*A operon consists of *lch*AA-AC (lic A-C) and *lch*A TE (licTE) genes encoding the proteins LchAA, LchAB, LchAC and thioesterase LchA-TE (Nerurkar, 2010) (cfr. Figure I. 4).

I.2.1.3. *Heterocyst envelope glycolipids (HGLs)*

Heterocyst envelope glycolipids (HGLs) consist of a glucose molecule coupled to a long chain (C26) triol or the 3-ketotautomer. An example is given in Figure I. 5 a :1-(O-α-D-glucopyranosyl)-3,25-hexacosanediol and 1-(O-α-D-glucopyranosyl)-3-keto-25-hexacosanolare secreted during heterocyst formation by filamentous cyanobacteria (Wolk *et al.*, 1994). Heterocysts are a differentiated cell type enabling bacteria to fix N_2 under oxic conditions. The development of a specific envelope outside the Gram-negative cell wall decreases the amount of O_2 entering the cell and as such protects nitrogenase from inactivation by O_2 (Haury and Wolk, 1978). The envelope consists of two distinct layers: the outer layer is composed of polysaccharides (heterocyst envelope polysaccharides) and protects a so-called laminated layer below. This laminated layer represents the actual barrier for O_2 diffusion rendering the heterocyst interior micro-oxic (Haury and Wolk, 1978; Walsby, 1985; Murry and Wolk, 1989) and is composed of specific glycolipids: heterocyst envelope glycolipids (HGLs). Heterocyst formation occurs upon nitrogen depletion by some of the vegetative cells of the *Anabaena* sp. filament, which results in patterning of the filament with heterocyst cells (Wolk *et al.*, 1994).

All structural genes necessary for the synthesis of HGLs in *Anabaena* sp. strain PCC7120 have been identified and are clustered in a ± 28 kb chromosomal region, spanning *hgl*E_A through *hgl*B, containing fatty acid synthases, polyketide synthases, ketoreductases, dehydrases, acyl transferases or thioesterases involved in the synthesis of the two closely related aglycones of heterocyst envelope glycolipids (Fan *et al.*, 2005) (Figure I. 5 b). Chain termination was attributed to a thioester reductase domain in *hgl*B in the apparent absence of any other mechanism. The resulting aldehyde would then either undergo a reduction or dismutation reaction producing a long-chain poly-ol (if dismutation is involved, it would also require a mechanism to reduce the resulting acid). The last step of HGL synthesis is the glycosylation of the hexacosane-1,3,25-triol to generate the mature HGL (cfr. Figure I. 5 a). The responsible glycosyltransferase, HglT, was also found in the cluster (Awai and Wolk, 2007). In other HGL producing families the glucose head of HGLs is rarely replaced by a galactose, mannose or β-glucose and sometimes the aglycon part of HGLs is dominated by tetrols and the corresponding ketones at the omega-3 position (Gambacorta *et al.*, 1996; Gambacorta *et al.*, 1998).

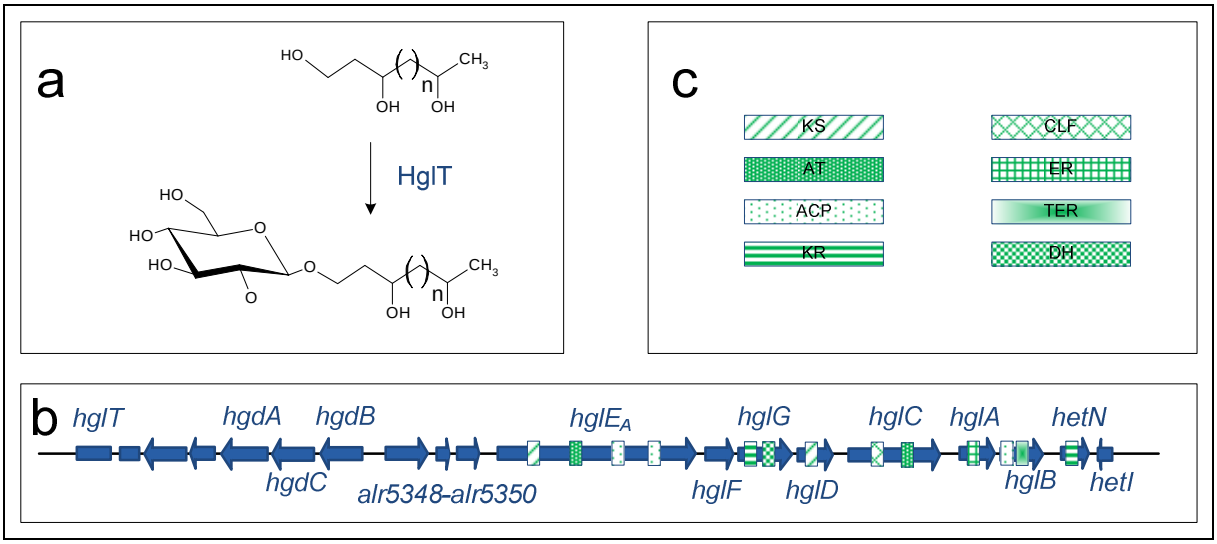


Figure I. 5 **(a)** Last step in the biosynthesis of HGLs consisting of the glycosylation of the formed long chain triol by the glycosyltransferase HglT **(b)** HGL gene cluster containing all the genes necessary for HGL formation **(c)** legend of loci with similarity to domains of polyketidesynthases that cumulatively may account for the synthesis of the HGL-aglycons of *Anabaena sp.* KS: R-ketoacyl syntase, AT: acyltransferase; ACP: acyl carrier protein; KR: ketoacyl reductase; DH: dehydrase; CLF: chain length factor; ER: enoyl reductase; TER: thioester reductase; adapted from (Fan *et al.*, 2005).

Transport of entire HGLs from the location of their synthesis, the cytoplasmic membrane, to beyond the outer membrane is mediated by DevB, DevC, DevA, and TolC (*alr*2887), which form an ATP-driven efflux pump (ABC-type exporter) (Moslavac *et al.*, 2007; Staron *et al.*, 2011). The exporter requires a distinct stoichiometry that allows the hexameric membrane fusion protein DevB to recognize and export its substrate (Staron *et al.*, 2011). The *dev*BCA genes are organised as an operon (*alr*3710-*alr*3712) elsewhere in the *Anabaena* genome than the biosynthetic gene cluster (Fiedler *et al.*, 1998). Mutations in *devBCA* or *tolC* result in the absence of the HGL layer, but do not impair HGL synthesis.

The rightmost border of the biosynthetic cluster contains the genes *het*N and *het*I. A certain regulation of heterocyst formation, more specifically the patterning of heterocysts, was found to depend upon HetN, which resembles short-chain dehydrogenases and contains a keto-reductase domain (Black and Wolk, 1994). HetN was suggested to metabolize the aglycone(s) and as such inhibit heterocyst formation in neighbouring cells of the filament resulting in patterning. However, the amino acid motif RGSGR of HetN (Higa *et al.*, 2012) and not the predicted ADH reductase activity (Liu and Chen, 2009) was found to be necessary for

patterning of heterocysts. HetI is related to phosphopantetheinyl transferases but no function has been attributed to this gene yet (Lambalot *et al.*, 1996; Fan *et al.*, 2005).

Transcription of the biosynthetic gene cluster as a whole (Ehira *et al.*, 2003) in addition to *tol*C (Moslavac *et al.*, 2007; Staron *et al.*, 2011) and the *dev*BCA operon (Staron *et al.*, 2011) is up-regulated in response to nitrogen deprivation. Upon perception of N-deficiency, the *ntcA* and *hetR* genes -encoding two regulators with key roles in heterocyst differentiation- increase their expression in a mutually dependent manner (Muro-Pastor *et al.*, 2002). Moreover, their expression is concentrated in certain spatially distributed cells that eventually become heterocysts (Black *et al.*, 1993; Olmedo-Verd *et al.*, 2006).

The global transcription factor NtcA, a member of the cyclic AMP receptor protein (CRP) family of bacterial regulators, exerts nitrogen control in cyanobacteria, and activates or in some cases represses genes in response to nitrogen limitation (Herrero *et al.*, 2001). NtcA binds to specific DNA sites, with the sequence signature $GTAN_8TAC$, present in the regulated promoters (Luque *et al.*, 1994; Vazquez-Bermudez *et al.*, 2002; Olmedo-Verd *et al.*, 2008). DNA binding of NtcA increases by interaction with 2-oxoglutarate (2-OG), a metabolic signal for the N status of the cell (Zhao *et al.*; Tanigawa *et al.*, 2002). These canonical NtcA-activated promoters include an NtcA-binding site centered about -41 nucleotides (nt) from the transcription start point of the regulated gene and a -10 box with the consensus sequence TAN_3T (Herrero *et al.*, 2001). In these promoters, NtcA and 2-OG have a positive influence on RNA polymerase (RNAP) recruitment, and are both required for obtaining transcriptionally active open promoter complexes with RNAP and for *in vitro* transcript production (Olmedo-Verd *et al.*, 2008). HetR has been reported to bind DNA upstream of some regulated genes *in vitro* (Risser and Callahan, 2007), however the precise mechanism by which HetR influences gene expression remains unknown. Expression of the HetR-dependent promoters likely takes place in differentiating cells, having a role in cell-specific gene activation during heterocyst differentiation.

A remarkable feature of many genes involved in heterocyst differentiation and function is that they bear multiple promoters (Flores and Herrero, 2010), and so does the *dev*BCA operon (Muro-Pastor *et al.*, 2002). These promoters include canonical NtcA-activated promoters (which do not require HetR), σ^{70} consensus-type promoters, and a special type of promoters that, depending *in vivo* on both NtcA and HetR, do not show any recognizable promoter

structure except, in some cases, a putative -10 promoter box (Camargo *et al.*, 2012). The mechanism of the dependence on NtcA of promoters that do not exhibit recognizable NtcA-binding sites remains unknown (Camargo *et al.*, 2012).

Expression of the *devBCA* operon of *Anabaena* sp. strain PCC 7120 from its dual promoter region was investigated by (Camargo *et al.*, 2012). Upon ammonium deprivation, the upstream canonical class II promoter is activated by NtcA at low levels in all cells of the filament, thereafter increasing in the differentiating cells in response to the localized increase of NtcA levels that result from localized increase of HetR levels and the mutual activation loop of the *hetR* and *ntcA* genes. The proximal *devBCA* promoter is also directly activated by NtcA, with a possible direct positive effect of HetR. Thus, through direct effects of NtcA on both promoters, an indirect effect of HetR in the distal one, and a direct effect of HetR in the proximal one, both promoters would contribute to expression of the *devBCA* operon in differentiating heterocysts.

I.2.2. *Biosurfactants produced by eukaryotes*

The best known biosurfactants produced by eukaryotes are glycolipids. Despite the fact that most of these molecules were already discovered more than half a century ago, the genetic background of their production and regulation remained largely unknown for a long time. However, some major scientific breakthroughs have been achieved the last five to ten years. Biosynthetic pathways linked to the responsible genes are now described for cellobioselipids like ustillagic acid (UA) and flocculosin, for mannosylerythritol lipids (MEL) and for sophorolipids (SL), but remain unknown for trehaloselipids, which will therefore not be included in this review.

I.2.2.1. *Mannosylerythritol lipids (MELs)*

Mannosylerythritol lipids (MELs) are produced by a variety of fungal species, *e.g. Schizonella melanogramma*, *Pseudozyma antarctica*, *Ustilago maydis*, *Kurtzmanomyces sp.*, *Geotrichum candidum* and *Pseudozyma aphidis* (Haskins, 1950; Kitamoto *et al.*, 1990; Kakugawa *et al.*, 2002b; Kurz *et al.*, 2003; Rau *et al.*, 2005). The dimorphic basidiomycete *U. maydis* is unique among fungal producers of biosurfactants because besides MELs it also produces a second structurally different class of glycolipids (cellobioselipids), which will be described in the next section.

MELs are built up of a mannosylerythritol disaccharide which is acylated with short-chain (C_2 to C_8) and medium-chain (C_{10} to C_{18}) fatty acids at the mannosyl moiety (position R-2 and R-3). Depending on the number of acetyl groups, mannosylerythritol lipids can be differentiated in four different varieties: MEL-A (fully acetylated), MEL-B and MEL-C (mono-acetylated at R-6 and R-4, respectively) and the fully deacetylated MEL-D (Kurz *et al.*, 2003) (cfr. Figure I. 6).

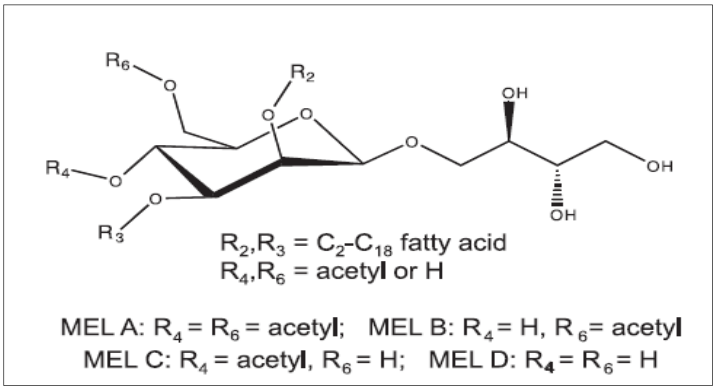


Figure I. 6 Molecular structure of MELs produced by *U. maydis.* MELs consist of the disaccharide mannosylerythritol, which is esterified with short-chain (C2 to C8) and medium-chain (C10 to C18) fatty acids at positions R-2 and R-3. Depending on the degree of acetylation at positions R-4 and R-6, MELs are differentiated into MEL A, MEL B, MEL C, and MEL D.

MELs, an extracellular oil heavier than water, were the first glycolipids produced by fungi or yeast for which a biosynthetic gene cluster was discovered (Hewald *et al.*, 2006). This cluster consists of five genes (cfr. Figure I. 7) coding for:

- a glucosyltransferase (*emt1*) responsible for stereospecific mannosylation (derived from GDP-mannose) of the meso-erythritol moiety at the C_4 position, yielding mannosyl-D-erythritol
- an acetyltransferase (*mat1*) responsible for acetylation both at the R-4 and R-6 positions of the mannose moiety
- two acyltransferases (*mac1* and *mac2*) responsible for specific acylation at position R-2 with a short chain fatty acid (C_2 to C_8) or R-3 with a medium or long chain length fatty acid (C_{10} to C_{18}). These last two enzymes appear to be highly specific, both in regioselectivity as in their preference for the length of the acyl-CoA cofactor, as they cannot replace each other. This specificity is in contrast with that of the

acetyltransferase Mat1, which is responsible for acetylation both at positions R-4 and R-6. For *Pseudozyma antarctica*, it has been shown that the medium-length fatty acids are derived from longer fatty acids by partial peroxisomal β-oxidation. Inspection of the genome sequence revealed that *U. maydis* contains both mitochondrial and peroxisomal proteins for fatty acid degradation. Thus, it was assumed that a similar peroxisomal chain-shortening pathway also exists in *U. maydis*.

- a protein of the family of the major facilitators (*mmf1*). Deletion of the *mmf1* gene leads to complete abolishment of extracellular MEL production, indicating that the major facilitator Mmf1 is essential for secretion and that no other secretion systems exist for MELs (Hewald *et al.*, 2005). This protein is able to secrete fully unacetylated MELs (*Δmat1*) and a large fraction of MELs acetylated at only one position (wild type), which indicates limited specificity for its substrates typically for members of the large family of major facilitators. However, the lack of MEL production for *Δmac1* or *Δmac2* mutants is probably attributed to the selectivity of the transporter protein. Acylation at both positions thus appears to be a prerequisite for glycolipid secretion (Hewald *et al.*, 2006). The intermediate mannosylerythritol has furthermore been isolated in significant amounts from MEL-producing cells (Boothroyd *et al.*, 1956), but not extracellularly, again suggesting that acylation is necessary for secretion. The secretion of MELs which are only acetylated at one position indicates that the second acetylation reaction is significantly slower than the first one. In this scenario, partially acetylated glycolipids are secreted before the second acetylation could occur.

Expression of all the members of this gene cluster is highly induced under conditions of nitrogen starvation (cfr. Figure I. 7). RNA derived from *emt1* for example was detectable 4h after shifting to nitrogen starvation medium. Although no obvious conserved sequence motifs, possibly involved in the co-regulation of the genes, were revealed when comparing the promotor regions of the cluster genes, several GATA sequences were found in all these promotor regions. This resulted in the suggestion that a GATA factor homologous to the general nitrogen regulator AreA from *A. nidulans* might be involved (cfr. I. 5). A candidate protein for the AreA function, XP_401867, was also reported to be present in the *U. maydis* genome. Some of these GATA elements are located as pairs within 30 bp of each other and thus are potential strong binding sites for GATA transcription factors. The hypothetical control by such GATA factors could be direct or indirect.

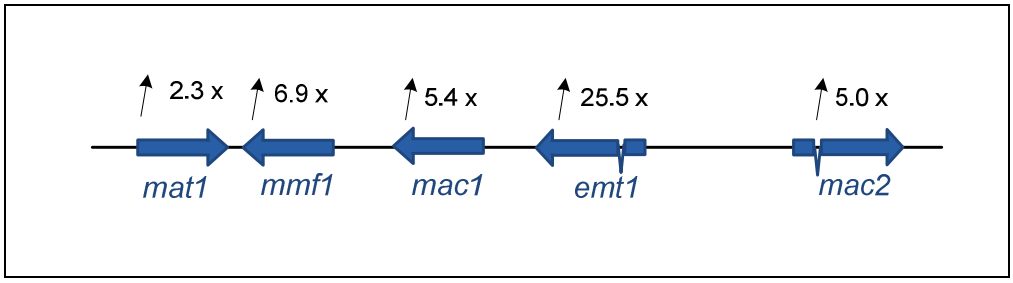


Figure I. 7 The MEL biosynthesis gene cluster comprises the *mat1* acetyltransferase gene, the *mmf1* gene, which specifies a member of the major facilitator family, *mac1* and *mac2*, encoding putative acyltransferases, and the previously identified glycosyltransferase gene *emt1*. Induction of transcription under conditions of nitrogen limitation is indicated above the open reading frames. *emt1* and *mac2* each contain a single intron; adapted from (Hewald *et al.*, 2006).

Remarkably, four of the five cluster genes have clustered homologs in the phylogenetically quite distant ascomycete *Aspergillus nidulans*, which are not present in *A. fumigatus* and *A. oryzae*, though the latter are closely related to *A. nidulans* (Hewald *et al.*, 2006). The genetic synteny shared between *U. maydis* and *A. nidulans* points to a common evolutionary origin of the MEL cluster genes. Horizontal gene transfer between these species or some of their recent progenitors is the most probable explanation for the presence of the cluster in these two quite distantly related organisms. It is thus possible that *A. nidulans* also synthesizes glycolipids similar or identical to MELs; however, it is not known whether or under which conditions the *A. nidulans* genes are expressed. Since the *A. nidulans* cluster lacks a potential export protein, it is feasible that the putative glycolipids might not be secreted.

The dimorphic basidiomycete *U. maydis* is unique among fungal producers of biosurfactants because besides MELs it also produces a second structurally different class of glycolipids (cellobioselipids) which will be described in the next section.

I.2.2.2. *Cellobioselipids*

A small number of basidiomycetous yeasts, mostly *Ustilaginales* secrete cellobioselipids. Ustilagic acid (UA) was the first one to be described by (Haskins, 1950) and consists of a cellobiose moiety O-glycosidically linked to the terminal hydroxyl group of a tri (2,15,16-) or di (15,16-) hydroxypalmitic acid. The sugar moiety is additionally esterified with an acetyl and a short chain β-hydroxy fatty acid (C_6 or C_8) (cfr. Figure I. 8 b).

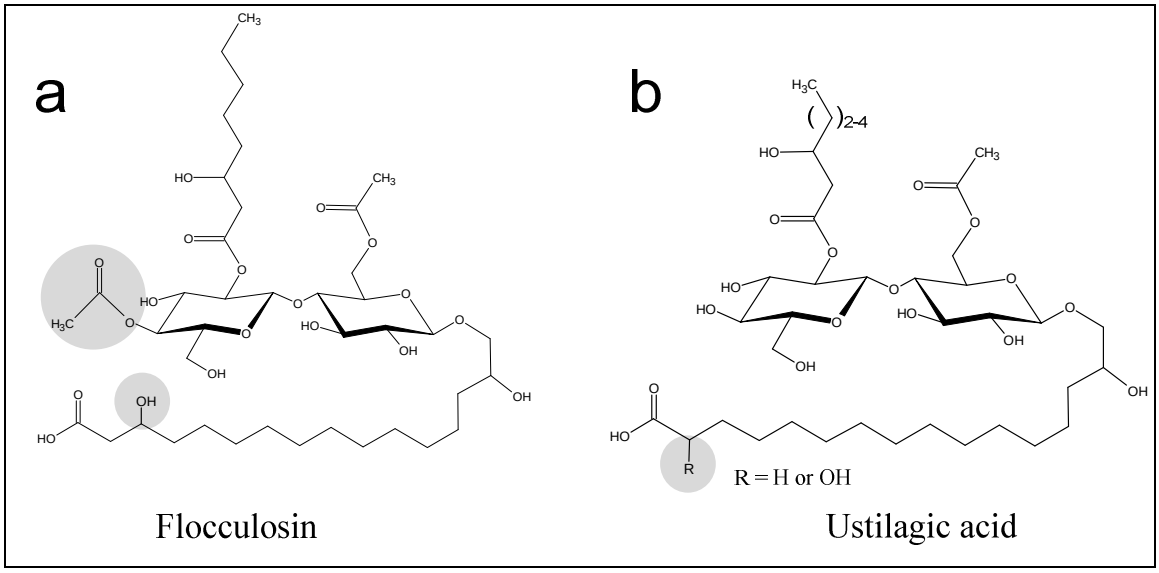


Figure I. 8 Structure of **(a)** flocculosin and **(b)** ustilagic acid (UA), which consist of a cellobiose moiety O-glycosidically linked to a long-chain fatty acid. The cellobiose of flocculosin is esterified with β-hydroxy-octanoic acid and acetylated at two positions. UA contains a cellobiose esterified with β-hydroxy-hexanoic or -octanoic acid and is acetylated at only one position.

A mixture of four major UA variants differing in the length of the short chain fatty acid molecule and the presence or absence of the α-hydroxy group on the fatty acid is secreted. Production of cellobiose glycolipids has since been described for other fungal species like *Pseudozyma graminicola, P. fusiformata, P. flocculosa, Trichosporon porosum, Sympodiomycopsis paphiopedili* and *Cryptotoccus humicola* (Puchkov *et al.*, 2002; Kulakovskaya *et al.*, 2004; Kulakovskaya *et al.*, 2005; Mimee *et al.*, 2005; Golubev *et al.*, 2008; Kulakovskaya *et al.*, 2010) and the products vary in the decorations of the sugar moiety and the hydroxylation pattern of the lipid tail. Amongst them is the fungal biocontrol agent *P. flocculosa* which produces flocculosin, a rare cellobiose lipid with antifungal activity (Cheng *et al.*, 2003; Mimee *et al.*, 2005). The lipid tail in this case consists of a (3,15,16-) trihydroxypalmitic acid. Flocculosin is acetylated at the 6' and 3'' position in contrast with UA, which is only acetylated at the 6' position. Another difference is the acylation at the 2'' position, which in the case of flocculosin only consists of acylation with a C_8 β-hydroxyfatty acid (Figure I.8 a).

The genetic background of the producing strains remained unraveled for a long time. However, Teichman *et al*. quite recently identified a biosynthetic gene cluster responsible for CBL production in *Ustilago maydis* (Teichmann *et al.*, 2007). This led scientists to suggest

the existence of a similar cluster in *Pseudozyma flocculosa* (Marchand *et al.*, 2009). Such a gene cluster was indeed recently identified and the genes are highly similar to those of *U. maydis* (Teichmann *et al.*, 2011a) (Figure I. 9). Moreover, all the genes in these 2 clusters are involved in CBL production, decoration and/or secretion. Biosynthesis is hence very similar and will be described here combined for both cases (the reader is referred to Figure I. 9 for respective gene names). For every discussed activity the gene names of *P. flocculosa* will be preceded by the respective ones of *U. maydis*.

- Hydroxylation of palmitic or β-hydroxy palmitic acid, for *U. maydis* and *P. flocculosa* respectively, at the terminal (ω) and subterminal (ω-1) position (*cyp1* and *cyp2*, respectively and for both organisms) is followed by

- glycosylation at the terminal hydroxylgroup by either the sequential addition of two glucose molecules derived from UDP-glucose or a cellobiose moiety (this remains to be determined) (Teichmann *et al.*, 2007; Teichmann *et al.*, 2011a) by a glucosyltransferase (*ugt1* vs. *fgt1*).

- This cellobioselipid is subsequently further decorated on the cellobiose moiety at position 2'' by the action of an acyltransferase (*uat1* vs. *fat1*) and at the 6' position by an acetyltransferase (*uat2* vs. *fat2*). The presence of an additional acetylgroup at the 3'' position of flocculosin was found to be caused by a second acetyltransferase encoded by *fat3* in the floccolusin gene cluster, which has no homologue in the UA gene cluster. The *fat2* and *fat3* acetyltransferases show high specificity in acetylating only their position of the cellobiose moiety, as *fat3* was not able to complement the *Δuat2* mutant, in contrast to *fat2*, which moreover failed to add a second acetylgroup on UA. Fat 3 also only seemed to be able to acetylate the second glucose molecule when the first one was already acetylated as introduction of *fat3* in a *Δuat2* mutant did not change the phenotype.

- The clusters additionally contain genes responsible for synthesis (*fas2* for both) and hydroxylation (*uhd1* vs. *fhd1*) (Teichmann *et al.*, 2011b) of the short-chain fatty acid at position 2''.

- Both also contain an ABC transporter (*atr1* for both) responsible for the export of the respective glycolipids.

- For UA, the last step in the biosynthetic pathway consists of hydroxylation at the α-position of the C_{16} dihydroxy fatty acid (*ahd1*), whereas for flocculosin the hydroxyl group at the β-position was already present on the substrate (β-hydroxy palmitic acid) before hydroxylation carried out by *cyp1* occurs. The gene responsible for this β-hydroxylation remains to be found. It was suggested that β-hydroxy fatty acids could be derived from *de novo* fatty acid synthesis (Teichmann *et al.*, 2011a), as is described for rhamnolipid biosynthesis trough the action of the PhaG enzyme (Rehm *et al.*, 1998; Deziel *et al.*, 2003).

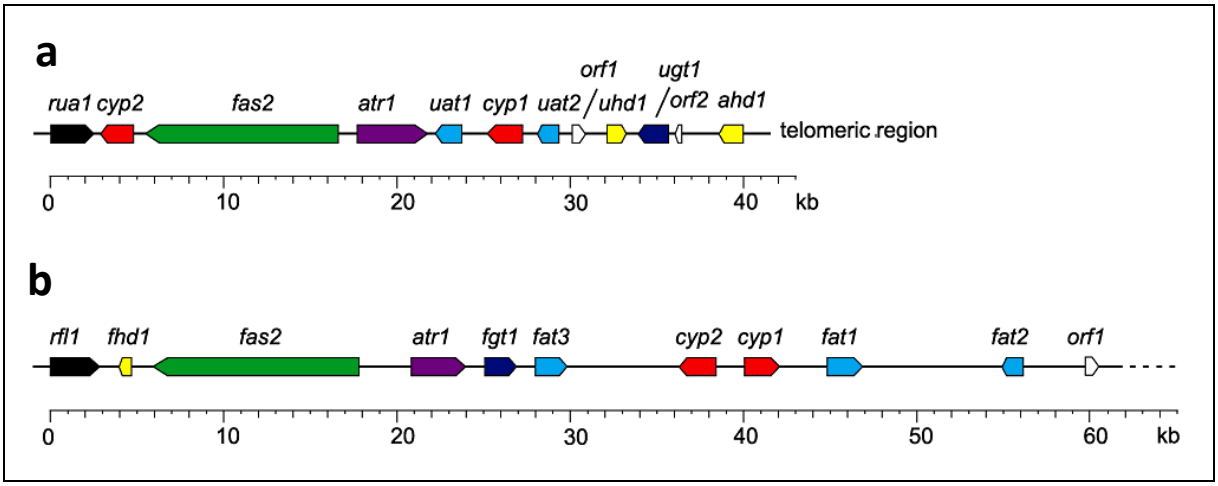


Figure I. 9 Molecular structure of the cellobioselipid biosynthetic gene cluster from **(a)** *Ustilago maydis* and **(b)** *Pzeudozyma flocculosa* (Teichmann *et al.*, 2011a).

The only two differences between the clusters, the unique presence of *ahd1* and *fat3* in the UA and flocculosin gene clusters, respectively, accounts for the structural differences between the two molecules. Both clusters are very similar in gene presence and function and likely evolved from a common origin. Although the gene sequence is relatively conserved, the order of the genes is more random. Furthermore, the UA cluster is more compact than the one for flocculosin (40 kb vs. 60 kb) and the single genes appear to contain more introns for the latter.

Clustering of genes has been proposed to promote selective transfer of certain genes via homologous gene transfer (Walton, 2000). It has also already been described that DNA alteration such as recombination processes, gene duplications and gene losses appear during fungal gene cluster evolution (Kroken *et al.*, 2003). This would explain the additional acetyltransferase *fat3* as well as the missing *ahd1* homologue in the flocculosin gene cluster. In spite of these differences, both fungi have maintained the ability to produce very similar

glycolipids, indicating a common origin for both gene clusters and a selection pressure to maintain this trait.

Since glycolipid production in *U. maydis* occurs under nitrogen starvation and expression of the first gene of the pathway (*cyp1*) was found to be strongly induced under such conditions (Hewald *et al.*, 2005), regulation of CBL synthesis by nitrogen starvation was suggested. The presence of a gene highly similar to the central regulator for nitrogen metabolism in *Aspergillus nidulans*, AreA, in the genome sequence of *U. maydis* strengthens this assumption (XP_401867, cfr. I. 5). Moreover, several GATA DNA motifs are present in the promotor sequences of genes of the biosynthetic pathway, which led the authors to suggest direct or indirect regulation by a GATA factor like AreA. However a very interesting finding was done by Teichman *et al.* in 2010. When these researchers first described the CBL gene cluster of *U. maydis* in 2007 and determined the borders of this cluster, they suggested that the leftmost gene of the cluster, with no clear role in biosynthesis, could possibly be a regulatory protein. This gene namely contains a Cys/His rich region, potentially constituting a zinc finger motif (Teichmann *et al.*, 2007). They therefore termed the corresponding gene '*rua1*'. The regulatory function of Rua1 was recently confirmed, as deletion of the *rua*1 gene leads to complete loss of UA production, whereas overexpression promotes increased UA synthesis, even in the presence of a good nitrogen source (Teichmann *et al.*, 2010). Moreover, a conserved sequence element found in all promotor sequences of the structural genes was confirmed to be involved in UA biosynthesis. This DNA element serves as an upstream activating sequence (UAS) mediating Rua1-dependent expression: Rua1 binds directly to the UAS of all genes of the cluster, except for *rua*1 itself, and this in a nitrogen dependent way as *rua1* expression is dependent on nitrogen limitation. MEL production is not affected, so Rua1 appears to act as a cluster-specific regulator. Rua1 is thus both necessary and sufficient to trigger UA biosynthesis which indicates that all environmental signals that affect UA biosynthesis are integrated at the *rua1* promotor.

The exact mechanisms regulating the expression of *rua*1 remain to be elucidated, but regulatory proteins that sense global and more specific nitrogen availability like the *U. maydis* AreA/Nit2 homolog and the homolog of the pathway specific activator Nit4 were suggested to be involved. Additional transcription factors, specifically involved in the nutrient control of secondary metabolism genes might exist, as was suggested by (Caddick *et al.*, 2006). A possible candidate transcription factor was identified in the genome of *Ustilago maydis* and

termed Nit1. Nit1 is suggested to be involved in nitrogen dependent transcription of Rua1. Interestingly, deletion of this gene affects both UA and MEL biosynthesis. Further investigation is needed to determine if Nit1 is directly involved in nitrogen sensing. On another level, it was suggested that *rua1* expression could be subject to posttranscriptional control as is described for the global nitrogen regulator AreA in *A. nidulans*: AreA mRNA is specifically degraded in response to intracellular glutamine (Caddick *et al.*, 2006). Finally, a positive feedback regulation for Rua1 was suggested, as a point mutation in Rua1 not only led to abolishment of UA production, but also to the abolishment of Rua1 expression.

Besides nitrogen limitation, also the effect of the carbon source on cellobioselipid production has been investigated. UA production by *U. maydis* is presumably dependent on the presence of glucose in the medium, while fatty acids (sunflower oil) will lead to the production of MELs (Spoeckner *et al.*, 1999). The yield and ratio of both classes of glycolipids depend on the available carbon source and can be shifted towards either of these biosurfactants (Haskins, 1950). In the case of flocculosin, studies on the carbon source revealed that glucose was preferred for both biomass and flocculosin production, and that the presence of carbohydrates influenced both the yield and the induction of flocculosin synthesis.

Nitrogen limitation is described as a common denominator for production of cellobioselipids by *U. maydis* and *P. flocculosa*. However, doubling the ammonium sulfate concentration in the medium, which was thus expected to delay or inhibit flocculosin production, did not have the expected effect (Hammami *et al.*, 2008). Instead, N consumption came to a halt after 48h of cultivation even with residual presence in the medium and flocculosin synthesis was switched on anyway. The authors first hypothesized organic N instead of inorganic N to be the factor that acts as a switch between flocculosin production and cell growth. However, as yeast extract nitrogen base gave the same results, they decided that the answer lies at another level: a physiological condition rather than by limitation for a certain compound, two events that possibly coincided in other research reports, could possibly be more determining for flocculosin production. Indeed, the cellular stage of the culture (i.e. resting cells versus growing cells) has been suggested to be of great influence for the biosynthesis of glycolipids (Kitamoto *et al.*, 1992a). For example, flocculosin production by *P. flocculosa* was proven to be dependent on the inoculum density (Hammami *et al.*, 2008): 0.4 g/L is the optimal start-up inoculum size for flocculosin production. Below that level production was reduced, while above that level it started do decrease rapidly. The authors suggested that *P. flocculosa* must

first reach a specific physiological state before flocculosin synthesis, a highly demanding energetic requirement, is initiated. Moreover, the activity of cells having reached the specific physiological state remains high enough for flocculosin production to resume when the carbon source is resupplied after exhaustion. The following was concluded: as a first step *P. flocculosa* needs carbohydrates and nitrogen to change its morphology either for flocculosin production or for cellular growth. As a second step, if *P. flocculosa* is provided with more sugars, regardless of the presence of nitrogen, flocculosin synthesis is switched on and will continue until sugar is depleted. However, if yeast extract is supplied together with sugars, cellular growth is favored over flocculosin production. The presence of a factor from yeast extract was thus suggested to activate the metabolic pathways supporting the production of biomass, leading to the rapid production of conidia rather than flocculosin. Finally, when growth is blocked by stress, *i.e.* the absence of that factor, flocculosin excretion might constitute an overflow metabolism for the yeast, which regulates the intracellular energy level. This hypothesis was also proposed for SL production by *S. bombicola*.

In conclusion the nitrogen and carbon source as well as the cellular stage of the culture influence cellobioselipid production.

I.2.2.3. *Sophorolipids*

Glycolipids consisting of a sophorose molecule linked O-glycosidically to a fatty acid, i.e. sophorolipids (SLs), were first identified as glycolipids produced by *Torulopsis* species (Gorin *et al.*, 1961). The hydrophobic part was identified as 17-hydroxystearate (C18:0) and 17-hydroxyoleate (C18:1), but also intermediates with 16 C-atoms were detected (Tulloch *et al.*, 1962). These SLs contain acetate groups at the 6’ and 6’’ positions (Tulloch *et al.*, 1967) and a macrocyclic lactone structure formed between the 4’’ hydroxyl group of the terminal glucose molecule and the hydroxyl acid carboxyl group can also be present (Tulloch *et al.*, 1967) (cfr. Figure I. 10 a). Other sophorose-containing glycolipids produced by *Rhodutorola bogoriensis* (formerly *Candida borogiensis*) were later identified (Tulloch *et al.*, 1968). These SLs differed in the fatty acid attached to the sophorose moiety, which in this case consisted of 13-docosanoic acid (C22:0) (cfr. Figure I. 10 b). Likewise as for the SLs produced by *Torulopsis*, they can be fully, partially or non-acetylated at the 6’ and 6’’ positions.

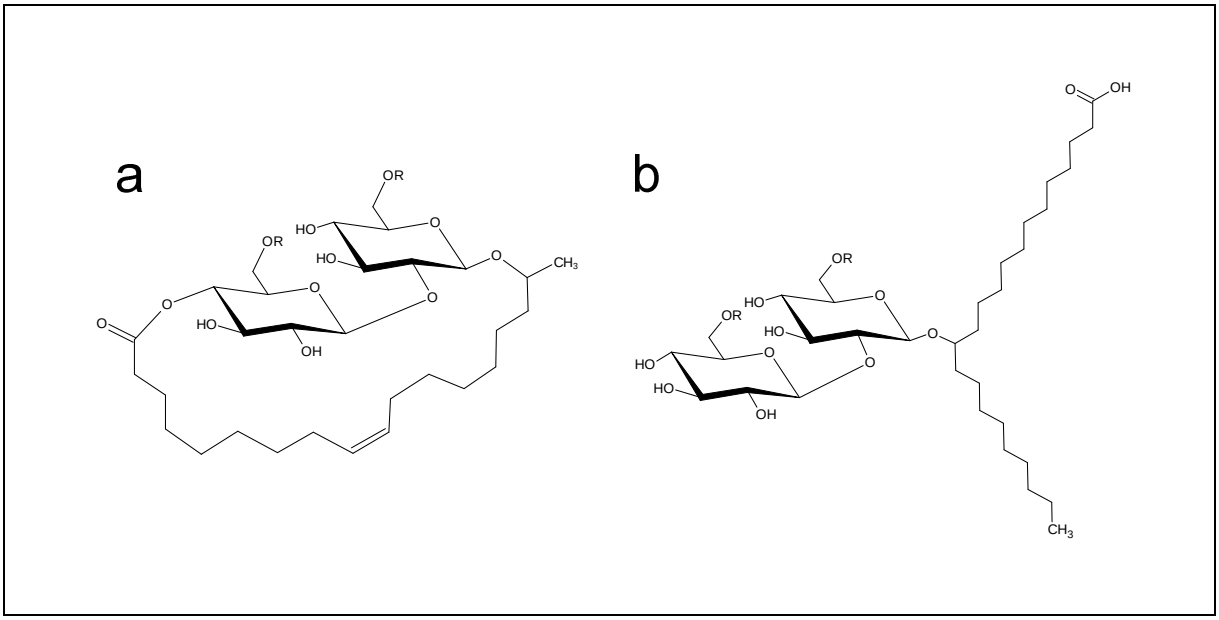


Figure I. 10 Structure of sophorolipids produced by **(a)** *S. bombicola* and **(b)** *R. bogoriensis* (R=H or $COCH_3$)

As mentioned above, a somewhat structurally different form of SLs, when compared to *R. bogoriensis*, are synthesized by a phylogenetically diverse group of yeasts. The earliest observations were done by Gorin *et al.* in 1961, who demonstrated SL production by the anamorphic ascomycetous yeast *Candida apicola* (former *Candida magnolia*) (Gorin *et al.*, 1961). Several closely related SL producing yeasts like *Starmerella bombicola* (formerly *Candida bombicola*) and *Candida batistae* were discovered. Besides these 3 species other species of the *Starmerella* clade produce significant amounts of SLs, namely *Candida riodocensis*, *Candida stellata* and *Candida sp.* NRRL Y-27208. SL production was only demonstrated for members of the *S. bombicola* subclade of the *Starmerella* clade (Kurtzman *et al.*, 2010). Finally, also less related organisms like *Wickerhamiella domercqiae* (Chen *et al.*, 2006) were shown to produce SLs.

Although these yeasts all produce SLs mainly consisting of a C18 hydroxyfatty acid (sub)terminally linked to a sophorose molecule, a clear structural diversity for the produced SLs was demonstrated by (Kurtzman *et al.*, 2010):

- *S. bombicola* produces a major di-acetylated lactone form plus a minor component of the free acid form.

- *C. stellata*, *Candida sp.* and *C. riodocensis* produce very little of the lactone form and the major product formed is the di-acetylated free acid form, besides minor amounts of mono- and non-acetylated SLs. The last two also produce mono-acetylated lactonic SLs.
- *C. apicola* produces the most heterogeneous SL mixture, mainly consisting of lactonic SLs, but in constrast to *S. bombicola* the mono- and non-acetylated forms are also observed in addition to minor amounts of the free acid forms of these three SLs.
- *C. batistae* primarily produces free acidic SLs (Konishi *et al.*, 2008)
- *Candida sp.* NRRL Y-27208 SLs contain an ω-hydroxy-linked acyl group (typically 18-hydroxy-Delta 9-octadecenoate) and occur predominantly in the free acidic form. In addition, 17 dimeric and trimeric sophoroses were identified from this strain (Price *et al.*, 2012).

These similarities and respective differences are probably a reflection of the genetic background of the producing organisms as was demonstrated to be the reason for the structural differences between UA and flocculosin, two very similar cellobioselipids produced by *U. maydis* and *P. flocculosa*, respectively (cfr. I.2.2.2).

Biosynthesis of SLs by *R. bogoriensis* was first investigated by (Esders and Light, 1972b) who demonstrated the involvement of glucosyl- and acetyltransferases. They were the first to propose a biosynthetic pathway consisting of the formation of a glucolipid from UDP-glucose and a hydroxylated fatty acid, followed by a second glycosylation (by the same or a different protein), yielding (unacetylated) SL. The disaccharide can subsequently be acetylated by the action of one or more acetyltransferases yielding a diacetylated SL. They suggested this di-acetylated SL to be the principal product excreted and suggested that the detected extracellular mono- and non-acetylated SLs are formed by deacetylation (for more information see I.3.2.).

For *R. bogoriensis* both the glucosyltransferase activities were found to appear in young cultures to both disappear at about the same time with a peak of activity after 1.5 days of cultivation when the cells approach the stationary growth phase. Synthesis of glycolipids thus occurs during the exponential growth phase and is not associated with a tropophase to idophase transition common to the production of many antibiotics and fungal secondary metabolites. The activity of the acetyltransferase(s) reaches a maximum in the early stationary phase (after 3 days of cultivation) after which the activity in crude cell extracts starts to diminish. This maximum activity coincides with the maximum amount of di-acetylated SLs

detected, which after three days also starts to diminish, whereas the amount of mono- and non-acetylated SLs continues to increase. Glucosyltransferase activities were found to be depressed on low glucose media as was the production of hydroxylated docosanoic acid and thus the activity of the enzyme responsible for the formation of the latter. This was in contradiction with the levels of acetyltransferase, which remained high on low glucose medium. Not just higher glucose concentrations, but more specifically the C/N ratio (or another compound of yeast extract) seems to regulate SL biosynthesis in *Rhodutorola bogoriensis*, as the ratio between glucose and yeast extract was proven to be important for regulation and an optimum for this ratio exists (Cutler and Light, 1979) (similar observations were made for cellobioselipids, as described in I.2.2.2). The fact that variations in the culture medium have a regulatory effect on the production of SLs does suggest some role in helping the organism to respond to changes in its environment (cfr. I.3).

Although the involvement of glucosyl- and acetyltranferases in SL biosynthesis of *R. bogoriensis* was already proven in the early seventies (see above), only in 2009 the discovery of a gene involved in SL biosynthesis by *S. bombicola* was reported (Van Bogaert *et al.*, 2009a). Today the SL biosynthetic pathway of *S. bombicola* is almost completely clarified and the responsible genes are identified (cfr. Chapter II, Figure II. 1). Very similarly to cellobioselipids, the genes are all found in one large subtelomeric gene cluster and also the steps of SL biosynthesis are highly similar to those of CBL biosynthesis:

- The first step in SL biosynthesis consists of (sub)terminal hydroxylation of a fatty acid by the action of a cytochrome P450 monooxygenase (*cyp52M1*) (Van Bogaert *et al.*, 2009a).

- Subsequent glycosylation of the hydroxy fatty acid, in contrast to CBL biosynthesis, involves two glucosylstransferases, instead of one. The first one (*ugta1*) (Saerens *et al.*, 2011a) is responsible for the transfer of a glucose molecule from UDP-glucose to the hydroxylated fatty acid yielding a glucolipid and UDP while the second one (*ugtb1*) (Saerens *et al.*, 2011c) specifically transfers a second glucose molecule from UDP-glucose to the formed glucolipid (and not to the hydroxylated fatty acid). The activity measured for the second glucosyltransferase (UgtB1) after 50 hours of cultivation was two times higher than that for the first glucosyltransferase (UgtA1), which could indicate that the first glucosylation step might be rate limiting (Saerens, 2012).

- the SLs are subsequently acetylated by the action of an acetyltransferase (*at*) (Saerens *et al.*, 2011b) and can be further

- Lactonised by the action of an only recently identified cell wall-bound lipase (*lip*) (unpublished results, cfr. chapter VI). The involvement of a lipase in lactonisation of SLs was already suggested for SLs produced by *Candida apicola* (Hommel *et al.*, 1994b). In contrast to all the other genes involved in SL biosynthesis by *S. bombicola*, the *lip* gene is not located in the biosynthetic gene cluster and seems to be differently regulated. As said, all the other genes of this biosynthetic pathway were found to be clustered, in analogy to the above described examples. This will be discussed in more detail in Chapter II.

Knocking out *ugta1* (or *cyp52M1*) results in a complete abolishment of SL production. The former indicates that the second glucosyltransferase is highly specific as it cannot rescue the *Δugta1* phenotype. This result might also give some indications for UA and flocculosin biosynthesis as here only one glucosyltransferase is found in the gene clusters and it is not known yet if this enzyme transfers one glucose molecule (derived of UDP-glucose) at a time to the hydroxylated fatty acid moiety or if it transfers a cellobiose molecule as such (cfr. I.2.2.2). The results for *S. bombicola* would point in the direction of the second assumption. Interestingly, UGTB1 activity was still detected in lysates of a Δ*ugta1* mutant (Saerens, 2012). The second glycosyltransferase was hence detected, but in contrast to assays of the wild type, no lactonic SLs were formed in these in vitro assays, which could indicate a regulatory effect of the formed SLs in the wild type on the transcription of the lipase gene, but not on expression of the glycosyltransferase. However, the lipase is secreted into the extracellular space, and these experiments were performed with lysates, which might also be an explanation for the abovementioned observations although lactonisation was detected for the wild type lysates in these experiments.

Similarly as described above for cellobioselipid biosynthesis and for SL production by *Rhodoturula*, the nitrogen and carbon source and the ratio thereof, as well as the cellular stage of the culture influence the production of SLs by *S. bombicola*. Although SL production by *S. bombicola* is often described as a two stage process for which growth and production are clearly separated (Cooper and Paddock, 1984; Davila *et al.*, 1992), activities of both the glucosyltransferases (UGTA1 and UGTB1) and the acetyltransferase (AT) are detected in cell lysates as soon as from the mid exponential phase (Saerens, 2012). The measured activities

increase towards the stationary phase, after which observations were unfortunately stopped, so no information is available about the further absolute course of the activity beyond 50 hours of cultivation is available. These results are in line with the conclusions of Albrecht et al. (Albrecht *et al.*, 1996) about the initial steps of SL biosynthesis in *S. bombicola*. They suggested that the enzymes involved in SL formation are, at least at a low basal level, constitutive and that nitrogen (and/or phosphate) depletion indirectly leads to enhanced SL synthesis trough an intracellular citrate accumulation (cfr. below), which is needed to supply acetyl-CoA for fatty acid biosynthesis. When the cells are growing exponentially, large amounts of these fatty acids are needed for the generation of membranes, but some of them can already be oxidized by the CYP52M1 enzyme, leading to the formation of SLs in exponentially growing cells. None of the activities are detected when the cells are grown in general yeast medium (YPD) (Saerens, 2012), which explains the absence of SL production on this general yeast medium and indicates the existence of regulatory effects exerted by the composition of the latter on the expression of the genes of the SL biosynthetic pathway. Furthermore, regulation of enzymes of the biosynthetic pathway at the transition from exponential to stationary growth was also demonstrated to be triggered by the carbohydrate (Hommel and Huse, 1993). Finally, the composition of the SL mixture varies with time: open SL structures are formed first and are then cycled into closed ones. The latter effect was faster for the resting cell method as compared to shake flasks experiments (Casas and Garcia-Ochoa, 1999).

This work focuses on *S. bombicola* and more specifically aims for the transformation of this yeast into a platform organism for the production of interesting biomolecules by using its very efficient SL machinery. The factors influencing the regulation of SL production are thus of paramount importance for this and further research. A more detailed description of several factors influencing SL production will be discussed below.

- Influence of the medium composition on SL production

A lot of research effort has been put in the investigation of the influence of the medium composition on SL biosynthesis for several SL producers, and predominantly for *S. bombicola* and *C. apicola* (Cooper and Paddock, 1984; Gobbert *et al.*, 1984; Zhou *et al.*, 1992; Hommel and Huse, 1993; Hommel *et al.*, 1994a; Hommel *et al.*, 1994b; Casas and Garcia-Ochoa, 1999; Cavalero and Cooper, 2003; Ashby *et al.*, 2005; Sudha *et al.*, 2010a; Sudha *et al.*, 2010b; Van Bogaert *et al.*, 2010; Van Bogaert *et al.*, 2011b; Zhang *et al.*, 2011). These

molecules have also been produced from renewable resources and even waste streams (Daniel *et al.*, 1998; Daniel *et al.*, 1999; Solaiman *et al.*, 2004; Ashby *et al.*, 2005; Kim *et al.*, 2005; Pekin *et al.*, 2005; Fleurackers, 2006; Felse *et al.*, 2007; Daverey and Pakshirajan, 2009a; Daverey and Pakshirajan, 2010; Daverey, 2011; Takahashi *et al.*, 2011).

Production of SLs is optimal when two carbon sources (a lipidic one and a glucidic one) are simultaneously provided during the growth phase and secretion phase. The distribution of SLs into structural classes may constitute gradual responses of the cells to imposed energetic conditions resulting from substrate supply.

S. bombicola and *C. apicola* are able to grow on medium with high osmotic pressure and the optimal glucose concentration for SL production was determined to be 100 g/L (Stüwer *et al.*, 1987; Zhou *et al.*, 1992; Casas and Garcia-Ochoa, 1999) or 120 g/L (Lang *et al.*, 2000). These high glucose concentrations were suggested to prevent fatty acids to be metabolized by the β-oxidation cycle. However, SL production also occurs on fructose, mannose (Gobbert *et al.*, 1984), sucrose (Klekner *et al.*, 1991), lactose (Zhou and Kosaric, 1993) and glycerol (Ashby *et al.*, 2006) or even soy (Solaiman *et al.*, 2007) and sugarcane molasses (Daverey and Pakshirajan, 2010).

Several oils can be used for SL production like olive oil, sunflower oil, corn oil, soybean oil, safflower oil, rapeseed oil, canola oil, coconut oil and even animal oils (Davila *et al.*, 1994; Casas and Garcia-Ochoa, 1999). SL production is similar for safflower oil, soybean oil and sunflower oil (Casas and Garcia-Ochoa, 1999). However, free fatty acids (or the methyl or ethyl esters thereof) like palmitic acid, oleic acid, hexadecenoic and octadecenoic acid alcohols and alkanes (Ito and Inoue, 1982; Cavalero and Cooper, 2003) are also suitable carbon sources for SL production. The nature of the lipophilic substrate influences the type of biosurfactant produced as well as the morphology and yield of the final product. Changing the co-substrate from sunflower to canola oil for example results in a large increase of the lactonic portion of SLs (Zhou and Kosaric, 1995) while using oleic acid (alone or with glucose) on the other hand increases the fraction of non-acetylated lactonic SL (Asmer *et al.*, 1988). Higher SL concentrations or yields have been reported for rapeseed oil ethyl esters or oleic acid, but these lipophilic carbon sources are more expensive than vegetable oils. Even though rapeseed esters are hardly consumed for cell growth, they appear to be potent inducers of SL biosynthesis. The derived fatty acids can be directly incorporated into the SLs, when they have a delimited length between 22.55 and 23.31 Å (Heinz *et al.*, 1969), but these can

also be synthesized *de novo* from any carbon source. In the last case C16 and C18 fatty acids are the main components of the produced SLs. The influence of rapeseed ester availability during the production phase on the SL lipid composition was also investigated. When residual lipid substrates accumulate in the medium as fatty acids resulting from hydrolysis of rapeseed methyl esters, acid forms and lactone forms of SLs are simultaneously produced. However, when the lipidic substrate is exhausted only acidic forms are produced.

Next to a lipidic and a glucidic carbon source, a source of vitamins, nitrogen and trace elements is also needed. Different compounds such as corn steep liquor, peptone, ammonium sulphate, yeast extract (YE) and mixtures of YE with urea or different salts have been employed. YE is the most frequently used nitrogen source, with an optimal concentration of 5 g/L (Cooper and Paddock, 1984), while Zhou et al. obtained higher SL concentrations using a lower YE concentration of between 2 and 3 g/L (Zhou *et al.*, 1992). Casas and Garcia-Ochoa on the other hand reported a concentration of 1 g/L to be optimal. Higher YE concentrations promote biomass formation, whereas SL production decreases probably because carbon sources have been depleted by yeast growth. When YE concentration is low, SL production is favored, while biomass concentration quickly reaches a stationary stage. Under these conditions the closed form of SLs is mainly synthesized while higher YE concentrations (> 5 g/L) favor the formation of opened structures (Casas and Garcia-Ochoa, 1999). Moreover, both growth and SL synthesis in *C. apicola* were reported to be strongly affected by increasing ammonium concentrations under constant YE concentrations (1 g/L) (Hommel *et al.*, 1994a). Increasing ammonium in the medium not only affects the formation of biomass, but also modulates enzyme activities of sophorose lipid synthesis as the altered ratio of the two hydroxyl fatty acid isomers of SL (ω and ω-1) indicates at least an altered activity of the P450 enzyme responsible for FA hydroxylation (Hommel *et al.*, 1994a). These authors suggested that not only the C/N ratio, but also the absolute quantity of N could possibly be important in SL formation by *C. apicola*.

- Influence of temperature, pH, pO_2 and culturing methods on SL production

The highest SL production with *S. bombicola* is obtained at 25°C, an incubation temperature close to the optimal growth temperature (22°C-25°C), and a pH value of 3.5 (Davila *et al.*, 1997). Varying these two parameters under excess of glucose and rapeseed esters decreases SL production, but does not result in a change of the composition of the SL mixture. When the pH is maintained constant at 5.0 instead of 3.5 for the whole fermentation period, the

performances are significantly reduced (Stüwer *et al.*, 1987). This was hypothesized to be a result of end-product inhibition, which might occur under these conditions since the solubility of SLs increases with increasing pH values.

Another important parameter is the aeration of the culture. Generally, higher levels of oxygenation result in increased SL formation, but a window of oxygenation that is optimal for SL biosynthesis was suggested to exist (Guilmanov *et al.*, 2002). The SL composition on the other hand is highly conserved over a wide range of aeration conditions. However, lower oxygenation in the initial periods of SL production can control the degree of FA unsaturation resulting in the increase of saturated FAs in the SLs.

The yield of SLs is also highly dependent on the imposed culture method: batch, fed-batch, fed-batch culture with pulse additions and resting cell methods have been considered of which the fed-batch method results in the highest SL production (Casas and Garcia-Ochoa, 1999).

- The effect of citrate on SL production

Addition of sodium citrate (2.5 to 10 g/L) to the culture medium at the beginning of the cultivation results in a slightly increased biomass, but also in significant higher biosurfactant yields. The optimal concentration was determined to be 5 g/L. A second effect of the addition of sodium citrate to the culture medium was a drastic alteration of the composition of the produced SL mixture. Cultivation of *S. bombicola* and *C. apicola* on glucose as sole carbon source without addition of citrate leads to the predominant production of acidic SLs. Upon addition of citrate, the amount of lactonic SLs increases and with prolonged cultivation time (over 5 days) water-insoluble SL microcrystals start to appear, which coincides with a drop of the concentration of water-soluble micelle forming lipids (Hommel *et al.*, 1994b). The *C. apicola* excreted microcrystals form crystalline platelets, which tend to aggregate. Electron microscope pictures of stationary glucose grown cells demonstrate so called hedgehog cells: the cell surface of most of the cells was covered by a crystal-layer consisting of the lactonic SLs. Using a mixed carbon source, comparable structured cell surfaces were extremely rare because the addition of hydrophobic substances liberates these cell-bound SLs into the culture medium. Cells furthermore possess the ability to convert water-soluble biosurfactant into crystalline lipids after a pH shift, but not vice versa and alteration of the pH in the cell free culture liquid does not influence secreted glycolipids. Lactonisation of the ionic forms of SLs furthermore proceeds only in the presence of whole cells grown in medium supplemented

with citrate (Stüwer *et al.*, 1987). (Baldé, 1989) detected a cell-wall bound lipase activity in *C. apicola* with synthetic capacities in apolar media. The involvement of a cell wall bound lipase in lactone formation was thus suggested. (Stüwer *et al.*, 1987) suggested that the effect of citrate merely is a pH effect: the SL composition was shown to be indirectly altered by altering the pH of the culture, resulting in either water-soluble acidic SLs (pH below 2) or crystalline lactonic SLs at higher pH values. The presence of citrate in the culture medium was furthermore shown to reduce the pH drop as compared to a medium in which no citrate is present (Stüwer *et al.*, 1987), which lead the authors to suggest that the effects observed for citrate are mere consequences of the altered pH. In Chapter VI of this manuscript these observations will be further investigated.

- Initiation of SL synthesis

The initiation of SL synthesis was investigated by Albrecht et al. (Albrecht *et al.*, 1996). They suggested, as stated above, that the enzymes involved in SL formation are, at least at a low basal level, constitutive and that nitrogen (and/or phosphate) depletion indirectly leads to enhanced SL synthesis. The latter was suggested to occur through a decline of the specific activities of NAD- and NADP dependent isocitrate dehydrogenases (NAD-ICDH and NADP-ICDH, respectively), which decrease in the exponential growth phase to 2 % (after 65 h of cultivation) and 0 % (after 45 h of cultivation) of the initial activities respectively. The unaltered constitutive high activity of citrate synthase subsequently causes an accumulation of isocitrate and citrate in the mitochondria and in the medium. Both acids are transported into the cytosol where citrate is cleaved by ATP:citrate lyase, giving rise to acetyl-CoA and oxaloacetate, the precursor for fatty acid synthesis by the cytosolic fatty acid synthase complex. Citrate furthermore allosterically (positively) modulates acetyl-CoA carboxylase, which is the regulating enzyme in the conversion of acetyl-CoA to malonyl-CoA (the commitment step in fatty acid synthesis). The formed fatty acids can then enter the SL biosynthetic pathway as was described above. NADPH for fatty acid synthesis might be generated by further metabolism of oxaloacetate, the other product of the citrate-cleaving reaction. The authors hypothesized that ATP:citrate lyase, NAD-ICDH and NADP-ICDH are involved in SL biosynthesis in a similar way as in the model for lipid accumulation by oleaginous yeasts formulated by (Evans and Ratledge, 1985). However, in contrast to the model of Ratledge, the enzyme activities are not regulated at the level of enzyme activity, but at the level of enzyme synthesis. The authors also hypothesized that SL yield could possibly

be even higher if all the citrate transferred into the cytosol would be metabolized by the ATP-citrate lyase, which they hypothetised to be the bottleneck leading to citrate secretion in the medium. It was suggested that this could be caused by the properties of the ATP:citrate lyase as there, as mentioned, was no activation of the enzyme, possibly in combination with the relatively high K_m value for citrate in comparison with that of the enzyme from the oleagineous yeast *Lipomyces starkeyi*. Another reason for citrate accumulation in the medium might be a rate-limiting step in fatty acid synthesis or SL formation.

I.3. Natural role of biosurfactants and other secondary metabolites

The role secondary metabolites play in the biology of cells is elusive. As mentioned above, they are mostly not essential and are often only produced during a specific phase of the organism's life cycle or under certain environmental conditions. Biosynthesis of secondary metabolites is usually subject to natural selection for the benefits that it confers to the cell producing them. Many secondary metabolites are produced by pathogenic fungi. The most likely advantage of secondary metabolites to a producing organism is that they may allow an organism to survive in its ecological niche, for example by competing successfully with other organisms in its natural habitat. Accordingly, many secondary metabolites tend to be compounds that have toxic or inhibitory effects on other organisms. A LaeA mutant of *Aspergillus nidudulans* (cfr. I. 5), which produces lower levels of secondary metabolites when compared to the wild type, was preferentially consumed over the wild type strain by the fungivorous arthropod, *Folsomia candida* (Rohlfs *et al.*, 2007). The secondary metabolites thus protect the fungus from predation. Virulence on a host may thus confer an advantage to the fungus, however in some cases the detrimental effect conferred by mycotoxins on the host only occurs after the fungus is dead, which does not confer a benefit to the fungus that produced the metabolite (Fox and Howlett, 2008).

Competition for nutrients, which is a factor of major importance in the microbial world, could also be a reason for secondary metabolite production as a way of denying carbon to competitors under conditions where the producing organism has more than enough for its energetic needs.

It is reasonable to assume that different groups of secondary metabolites and more specifically biosurfactants have different natural roles in the growth of the producing microorganisms. A general overview of some of the functions attributed to biosurfactants, is given below. It is not

the intention to give a complete overview, as this will take us too far, but to give the reader an idea about the possible natural roles of these fascinating molecules.

I.3.1. *Niche occupation and –protection and survival*

- Rhamnolipids

Although numerous studies have been performed on the biosynthesis and biochemistry of rhamnolipids, the exact physiological functions of these molecules for the producing bacteria are still not precisely defined. They seem to play multiple roles, as their presence promotes uptake of hydrophobic substrates (Beal and Betts, 2000; Noordman and Janssen, 2002) and alters cell surface polarity (Zhang and Miller, 1994; Al-Tahhan *et al.*, 2000). Rhamnolipids also have antimicrobial activities against other bacteria (Stanghellini, 1998; Abalos *et al.*, 2001; Haba *et al.*, 2003) and are able to disrupt *Bordetella bronchiseptica* biofilms (Irie *et al.*, 2005). In *Pseudomonas aeruginosa* biofilms, rhamnolipids are suspected to play a role in maintaining fluid channels and the detachment of cells from the biofilm community (Davey *et al.*, 2003; Boles *et al.*, 2005). Furthermore, rhamnolipids (and HAAs) seem to be required for swarming motility by acting as surface-modifying agents. The reduction of the surface tension also causes the surface conditioning needed for efficient colonization (Matsuyama and Nakagawa, 1996). Many of these functions are derived from the well known physicochemical properties of RLs such as surface activity, wetting ability, detergency, and other amphipathic related properties. Taken together, the abovementioned examples suggest that these molecules might play multifunctional roles for the producing organisms (Wilhelm *et al.*, 2007).

- Surfactin

Several lipopeptide surfactants including the cyclic lipopeptides (CPLs) of *B. subtilis* are potent antibiotics (Mukherjee and Das, 2005) and it has been suggested that some, if not all, peptide antibiotics produced by the genus *Bacillus* play a role in their sporulation (Grossman, 1995). Interestingly, the antibiotic potency and microbial specificity of the crude lipopeptides from *B. subtilis* DM-03 and DM-04 strains differed, which may be due to the presence of diverse CLP isoforms. It is reasonable to assume that this antibiotic specificity of lipopeptides may have a natural role in enhancing the growth of the producing bacteria by conferring some kind of competitive advantage over the other microbes, which are competitors of the producer in a given environment (Maier, 2003). (Mukherjee and Das, 2005) provided evidence that one

group of CLPs confer an advantage to the producing *B. subtilis* strain in a specific ecological niche, whereas another group of CLPs would be more appropriate for sustaining the growth of producing *B. subtilis* strain in a different ecological niche.

- Cellobioselipids

It has been hypothesized that the fungus *Pseudozyma flocculosa* produces the biosurfactant flocculosin in its antagonistic activity against powdery mildews, as it was shown to possess important antifungal activity in vitro (Cheng *et al.*, 2003); a natural function not applicable to *Ustilago maydis*, which produces structurally closely related glycolipids. Whether or not the glycolipids play a similar role in both fungi, their intricate synthesis through an elaborate gene cluster conserved between two related fungi with different lifestyles certainly supports the fact that they have an important function in their overall fitness. Toxic or inhibitory effects of cellobioselipids produced by various microorganisms on other organisms were extensively studied and reported (Kulakovskaya *et al.*, 2004; Mimee *et al.*, 2005; Kulakovskaya *et al.*, 2007; Golubev *et al.*, 2008; Kulakovskaya *et al.*, 2009; Mimee *et al.*, 2009b; Kulakovskaya *et al.*, 2010). However, these proven activities do not necessarily have this antagonistic function in the producing organism as they may be beneficial from a completely different point of view. The *U. maydis* Mmf1 protein (transporter of MEL glycolipids cfr. I.2.2.1) f. e. displays the highest level of sequence similarity to Mfs1-1 of *Coprinus cinereus*, which is located within the mating type-determining region of this basidiomycetous fungus (Halsall *et al.*, 2000). This could indicate a potential role of this exporter in the function of the mating type locus, i.e., for secretion of glycolipids, which enhance diffusion of the hydrophobic lipopeptide pheromones, as has been suggested for *U. maydis* (Hewald *et al.*, 2005).

- Sophorolipids

(Hommel *et al.*, 1994b) suggested SLs produced by *Candida apicola* to be helpful in the adaptation to high osmotic strengths in the natural environment of this yeast (honey) similarly as the production of polyol compounds by halotolerant yeasts in fermented foods regulating the osmotic pressure of the environment. (Davila *et al.*, 1997) suggested that the physiological role of SL excretion might constitute an overflow metabolism for the yeast which regulates the intracellular energy level (cfr. I.3.2.).

The regulation of the production of extracellular di-acetylated SLs by the nutrient composition of the growth medium may represent an important property in the adaptation of *Rhodotorula bogoriensis* to its natural environment, the phyllosphere (Cutler and Light, 1979). This hypothesis was based upon the observation that growth of *Rhodotorula glutinis* was particularly active and accompanied by the copious secretion of extracellular slime on damaged segments of seed coats, exposing the nutritionally rich endosperm (Cutler and Light, 1979). It was suggested that these wounded areas of plant surfaces release glucose, which in turn would induce the yeast to produce extracellular lipid that could help to provide a protective coating for the wounded area, which would not only protect against desiccation but also against attack by pathogenic fungi.

Inhibitory and/or toxic effects towards *Ustilago maydis* (Tran, 2012), yeasts (Ito *et al.*, 1980) and (gram positive) bacteria (Gross and Shah, 2005; Shah *et al.*, 2007) were reported for sophorolipids (SLs). Especially the lactonic SLs were reported to have measurable biocide activity (Hu and Ju, 2001).

I.3.2. *Competition for and storage of nutrients*

Although a lot of research effort has been dedicated to the elucidation of the natural function(s) of biosurfactants like sophorolipids, rhamnolipids, cellobioselipids, etc. the possibility that one of the functions of these compounds could be the formation of storage compounds under favorable conditions has long been underexposed. However, microorganisms are known to be capable of accumulating a wide variety of energy- and carbon storage compounds when certain nutrient supplies, other than carbon, are limiting. These intra- or extracellular carbon-storage compounds can be mobilized and used again under carbon starvation conditions temporarily occurring in the environment.

Well known examples include polyhydroxyalkanoates (PHA) (Chen, 2010), produced by a variety of taxonomically different microorganisms, and storage carbohydrates like glycogen (Wilson *et al.*, 2010), curdlan (Buller and Voepel, 1990) and trehalose (Panek, 1963; Arguelles, 2000). The ecological significance and regulation of these storage compounds has been widely described. In addition, these molecules often play a significant role in bacterial fitness. Accumulation of PHA for example is more than just an intracellular carbon storage strategy, but additionally enhances the survival of several bacteria under environmental stress conditions (Zhao *et al.*, 2007). Furthermore, yeast cells that can accumulate glycogen have a

growth advantage over cells that cannot, suggesting that glycogen makes a contribution toward overall fitness (Anderson and Tatchell, 2001).

The possibility that biosurfactants could also have a function as storage compounds was suggested for sophorolipids (Hommel *et al.*, 1994b), cellobioselipids (Mimee *et al.*, 2009a), MELS (Hommel *et al.*, 1994b) and surfactin (Abdel-Mawgoud *et al.*, 2008).

As early as in 1961 SL glycolipids were reported to disappear from the culture medium of old *R. bogoriensis* cultures (Deinema, 1961). Ten years later (Esders and Light, 1972a) reported the detection of mono- and non-acetylated SLs, in addition to the di-acetylated ones already characterized by Tulloch *et al*. in 1968 (Tulloch *et al.*, 1968), and suggested these to be part of a catabolic rather than of the biosynthetic pathway. They proposed possible degradation of the di-acetylated SL as an explanation for the gradual disappearance of these glycolipids after 3.5 days of cultivation of *R. bogoriensis* cultures and the simultaneous appearance of mono- and non-acetylated derivatives in these older cultures. This catabolic process was suggested to occur through a sequential development and (possible) release of hydrolytic enzymes responsible for hydrolysis of acetyl bonds and glycosidic linkages respectively. They also suggested that the mono-acetylated derivative would predominantly or completely be the one acetylated at the internal glucose moiety. (Bucholtz and Light, 1976) subsequently identified a catabolic acetylesterase responsible for hydrolysis of both acetyl groups of SLs in cell extracts of *R. bogoriensis*. This isolated enzyme was not specific for hydrolysis of the acetyl groups present in the glycolipid as p-nitrophenyl acetate even proved to be a better substrate, but did show certain specificity for the glycolipid as compared to other esterases from the same organism that did not show any activity towards the acetylated glycolipids. The acetylesterase activity was a lot lower than the acetyltransferase activity in the same extracts indicating that the biosynthetic pathway and catabolic pathway are overlapping.

Other SL producing species were identified in the years to follow (Gorin *et al.*, 1961; Jones, 1967; Tulloch *et al.*, 1967) but the function of these glycolipids was mostly described in connection with growth of the producing microorganisms on hydrophobic carbon sources like n-alkanes (Ito and Inoue, 1982; Hommel, 1990; Dekoster *et al.*, 1995). However, (Hommel *et al.*, 1994b) suggested SLs produced by *C. apicola* to be extracellular storage material. This hypothesis was investigated for *Starmerella bombicola*, another SL producing yeast isolated from honey, by prolonged incubation (Hu, 2000). At this point the authors could not conclude if SLs were degraded by *S. bombicola* at the starvation stage. They did mention that the cells

apparently switched to a different metabolism when nutrients were depleted from the medium and that this switch was influenced by pH. Researchers from the same lab later stated that degradation SLs produced by *S. bombicola* were not consumed by the yeast even after prolonged cultivations (Lo and Ju, 2009). (Garcia-Ochoa, 1996) on the other hand stated that SLs can be used as the sole carbon source by *S. bombicola* and even filed a patent on the production of sophorose from SLs in which they stated that cultivation of *S. bombicola* on SLs as the sole carbon source led to a release of sophorose in the culture medium. The possible catabolism of SLs by the producing organism *S. bombicola*, will be further investigated and discussed in Chapter III of this manuscript.

As for cellobioselipids, (Hammami *et al.*, 2008) reported the disappearance of flocculosin, from the culture medium of the glycolipid producing yeast-like fungus *Pseudozyma flocculosa* upon complete consumption of the carbon source. Flocculosin, a secreted cellobioselipid mostly described to be part of the fungus' biocontrol arsenal against other fungi, thus also appeared to be degraded by its producer when no other C-source was available (anymore). The same group recently indeed confirmed glycolipid degradation by the producing organism. This proves that flocculosin can be recycled by *P. flocculosa* as a nutrient in addition to its role in the protection of the ecological niche of the fungus (Mimee *et al.*, 2009a). Flocculosin was fully metabolized into 5 major products within 24 hours in a medium with the glycolipid as sole carbon source and flocculosin started to disappear from the culture medium after the first 12 hours. The pH of such cultures rapidly increased from 6 to 7.9 within the first 12 h and reached 8.6 after 96 h. When the medium was buffered at 6 no degradation was observed at all. The catabolic pathway consisted of removal of the decorations on the 'external' glucose molecule within the first 24 hours. Fully deac(et)ylated cellobioselipid and glucolipid started to appear after 24 hours of incubation to peak after 72 hours, after which these intermediates also gradually disappeared from the medium. Incubation of flocculosin with cell free solution of a 72 hours old culture results in complete digestion of fully ac(et)ylated flocculosin to glucolipid and the trihydroxylated C_{16} fatty acid after 24 hour of incubation. The formation of the hydroxylated fatty acid depends on initial formation of the glucolipid and no cellobiose was detected in any of the analyses suggesting a stepwise release of glucose molecules.

Similar catabolic pathways are likely to exist in other biosurfactant producers. The biosynthetic pathway of cellobioslipids produces by *Ustilago maydis* is highly similar to that of *P. flocculosa,* (Teichmann *et al.*, 2011a) so a similar catabolic pathway as the one

mentioned above could possibly also be present in *U. maydis*. Erythritol lipids of *Candida antarctica* were stored intracellularly up to 12 % of the dry weight after growth on glucose which was also suggested to be (partial) proof for a storage function of MELs (Hommel *et al.*, 1994b). Last but not least it was reported that surfactin decreased after depletion of a nutrient source, even though cell growth slightly increased. The authors suggested that the cells might assimilate surfactin as the carbon source for additional cell growth when (a certain) nutrient source was not available (Abdel-Mawgoud *et al.*, 2008). Rhamnolipids produced by *Pseudomonas aeruginosa* on the other hand remained unchanged in cultures for up to 33 days (Hauser and Karnovsky, 1954).

An important fact that has to be emphasized in this context is that the antagonistic activity of biosurfactants like flocculosin, UA, surfactin, etc. was proven to result from a membrane-destroying effect of these biosurfactants. Treatment of yeast cells with cellobiose glycolipids results in ATP leakage caused by increased cytoplasmatic membrane permeability (Kulakovskaya *et al.*, 2003). The question thus rises what causes some organisms and especially the producing organism itself to be insensitive to this membrane destroying effect. A different composition of the cell wall could be an explanation. Selfprotection (immunity) of *Bacillus* strains against produced lantibiotics was stated to be based on the presence of proteins (LanFEG) homologous to ATP-binding cassette (ABC) transporters that export the lantibiotic from the cytoplasmic membrane into the extracellular space (Stein *et al.*, 2003; Stein, 2005). YerP, a homolog to the resistance nodulation-cell division (RND) family efflux pumps, on the other hand was found to be implicated in surfactin resistance by secretion (Tsuge *et al.*, 2001). Furthermore, several lantibiotic producers possess membrane-bound lipoproteins LanI, which exhibit a sequestering-like function that prevents high local concentrations of the lantibiotic close to the cytoplasmic membrane and/or interferes with lantibiotic lipid II pore formation (Stein *et al.*, 2003; Koponen *et al.*, 2004; Stein, 2005). However, another plausible explanation is that insensitive organisms are capable of degrading the membrane compromising agents due to their ability to produce enzymes capable of hydrolyzing (part) of these molecules. (Eveleigh *et al.*, 1964) described the release of enzymes capable of degrading UA by fungi such as *Fusarium sp*. The authors surmised that those fungi avoided the deleterious effect of cellobioslipids in an effort to maintain their ecological niche. This could thus also explain why flocculosin has no adverse effect against *P. flocculosa* itself. (Belanger and Deacon, 1996) speculated that insensitivity to flocculosin by some fungal pathogens might indeed be explained by enzymatic degradation of the molecule.

Similar assumptions can be made for other biosurfactants. (Grangemard *et al.*, 1999) for example reported on the enzymatic hydrolysis of surfactin by bacterial endoproteases

Next to serving as a storage component, glycolipidic secondary metabolites can also be helpful in nutrient assimilation in others ways. Glycolipid production by cyanobacteria (HGLs) have a clear function in the elaboration of heterocysts, which facilitates the spatial separation of an oxygen labile metabolic process, nitrogen fixation, from one that evolves molecular oxygen, photosynthesis with photosystem II. Fixed nitrogen is supplied to vegetative cells from heterocysts, and in return, heterocysts receive a source of carbon and reductant to compensate for their lack of PS II and the Calvin cycle (Wolk *et al.*, 1994).

I.4. Biotechnological opportunities and applications for biosurfactants.

There is a wide range of applications for biosurfactants in several industrial fields (Desai and Banat, 1997). Biosurfactants represent ecological alternatives to their synthetic counterparts as they exhibit lower toxicity, potentially high activities, and stability at extremes of temperature, pH, and salinity. Most importantly, they are biodegradable, making them environmentally friendly “green” chemicals. All this in combination with the fact that they can be produced from renewable resources gives them an advantage over their chemical counterparts and may therefore make them suitable to partly replace chemicals (Banat *et al.*, 2000). Biosurfactants have been described to find (potential) applications in the medical world, personal care sector, mining processes, food industry, cosmetics, crop protection, pharmaceuticals, bio-remediation, household detergents, paper and pulp industry, textiles, paint industries, etc. (Banat *et al.*, 2000). Especially glycolipid biosurfactants, like sophorolipids (SLs), rhamnolipids (RLs), mannosylerythritol lipids (MELS) and cellobioselipids (CBLs) have been described to provide significant opportunities to replace chemical surfactants in several sectors (Marchant and Banat, 2012).

The suggested opportunities for some of the biosurfactants that have been described in the previous sections will only briefly be discussed here and the reader is referred to extensive reviews for more information:

- General reviews: (Kitamoto *et al.*, 2002; Schramm *et al.*, 2003; Muthusamy *et al.*, 2008; Kitamoto *et al.*, 2009; Banat *et al.*, 2010; Nguyen and Sabatini, 2011; Dreja *et al.*, 2012)

- Reviews on environmental applications: (Mulligan, 2005; Mulligan, 2009; Pacwa-Plociniczak *et al.*, 2011)
- Review on applications in the petroleum industry: (Perfumo *et al.*, 2010)
- Review on medical applications: (Cameotra and Makkar, 2004; Rodrigues and Teixeira, 2010)
- Reviews on more specific applications: (Xu *et al.*, 2011)

The major two factors currently limiting the penetration of bio-surfactants into the market are firstly the limited structural variety and secondly the rather high production price (due to low productivity). However, the expanding knowledge of the genetics of the producing organisms may be of significant importance as this knowledge represents the necessary base for the genetic engineering of biosurfactant producers. The latter is indispensable for the development of enhanced recombinant strains, which may well become industrial strains. Some of the efforts made in this respect for some of the abovementioned biosurfactants are discussed below.

I.4.1. *Biotechnological potential of engineered glycolipid producing strains*

I.4.1.1. *Rhamnolipids*

Rhamnolipids (RLs) have been described to find applications in foods, cosmetics, agricultural substances, pharmaceuticals as well as in bioremediation processes (Banat *et al.*, 2010). However, as described above, the complex regulation of the biosynthetic pathway in combination with a tight linkage with other biosynthetic pathways ànd the fact that the best described producer is an opportunistic human pathogen, hampers a real breakthrough of industrial production processes for RLs. Rhamnolipid overproduction was first obtained in the natural RL producer *P. aeruginosa* after induced random mutagenesis (Koch *et al.*, 1991; Iqbal *et al.*, 1995) and later e.g. by overexpression of *estA*, encoding an esterase involved in RL biosynthesis in *P. aeruginosa* (Wilhelm *et al.*, 2007). Heterologous RL production can be a strategy to avoid the complex regulation mechanisms and safety issues associated with the pathogenic *P. aeruginosa*. Knowledge of the involved genes thus resulted in recombinant *P. putida* and *P. fluorescens* strains producing rhamnolipids (Ochsner *et al.*, 1994a; Ochsner *et al.*, 1995; Cha *et al.*, 2008). (Ochsner *et al.*, 1995) reported final RL concentrations of 0.25 g/L with recombinant *P. fluorescens*, 0.6 g/L with *P. putida* KT2442 and no RLs with *E. coli.* Introduction of the *rmlBDAC* operon responsible for the synthesis of the dTDP-L-rhamnose

precursor in *E. coli* in addition to the *rhlAB* operon resulted in final yields of mono-rhamnolipid of 0.12 g/L, so it was concluded that the availability of L-rhamnose was the limiting factor for RL production in *E. coli* (Cabrera-Valladares *et al.*, 2006). The highest heterologous RL production was obtained with *P. putida* KCTC1067 expressing the *rhlAB* operon and *rhlI* of *P. aeruginosa* EMS1 (Cha *et al.*, 2008).

The fact that both precursors of RLs are derived from central metabolic pathways makes it difficult to over-produce RLs in heterologous hosts (Ochsner *et al.*, 1994a; Ochsner and Reiser, 1995). Additionally, byproducts (like PHA) which compete for the precursor pools, are also available in *P. putida* strains (Muller and Hausmann, 2011). Therefore, for now, only few bacterial species can be considered good RL producers. Finding new RL-producing bacteria is beneficial from a biotechnological point of view as it might result in the discovery of natural non-pathogenic producers which compared to the pathogenic *P. aeruginosa* strains could become more appropriate candidates for the industrially-safe production of RLs. Screening efforts will require the development of wide spectrum approaches, probably based on the detection of surface tension lowering substances or, theoretically and ideally, the presence of the *rhl*A, *rhl*B, and *rhl*C homologues in the genomes of bacteria. Finally, it is expected that studies on the genetic regulation of RL synthesis, especially in the context of specific environmental conditions affecting the production of these biosurfactants, will help in deciphering the exact genetic control of RL production. Consequently, this understanding should lead to a more successful control of regulatory and growth factors directing the optimized large-scale production with a safe host (Abdel-Mawgoud *et al.*, 2010).

I.4.1.2. *Surfactin*

Surfactin is a biological surfactant with numerous potential applications for various medical applications. Besides its antifungal and antibacterial effect, surfactin can inhibit fibrin clot formation, induce the formation of ion channels in lipid bilayer membranes, block the activity of cyclic adenosine monophosphate and furthermore exhibits antiviral and antitumor activities (Seydlova and Svobodova, 2008). A significant obstacle to the large-scale industrial application of surfactin is the high production cost coupled with a lower production rate compared with commercially available synthetic surfactin (Desai and Banat, 1997). A production cost competitive with the synthetic surfactant would lead to an increase in the commercial availability of surfactin, which then again would lead to a tremendous growth in the industrial use of surfactin in the coming decade (Whang *et al.*, 2009). To achieve this goal,

recent efforts have focused on the reduction of surfactin production costs through improving the yield and the use of either cost-free or low-cost feed stocks, such as cassava flour wastewater (Makkar et al., 2011). Most previous studies have focused on strain selection, random mutation, or manipulation of nutritional factors to enhance surfactin production, which only resulted in marginal increases (Abdel-Mawgoud et al., 2008; Moran et al., 2010). The direct application of genetic manipulation to improve surfactin production and quantitative analysis thereof in recombinant microbial hosts is only a recent trend (Coutte et al., 2010; Sun et al., 2009).

*Bacillus subtili*s was e.g. engineered to improve surfactin production by overexpressing two regulators, *com*X and *phr*C encoding ComX and CSF respectively, of the srfA operon (encoding the surfactin biosynthesis structural genes) with the aim of stimulating transcription of the *srf*A operon. Surfactin production for the transformants was 6.4-fold higher than that in the wild type strain, with approximately 135 mg/L surfactin produced after 48 h cultivation. Surfactin produced by the engineered strains showed functional groups similar to the commercially available surfactin. To reduce the production costs of surfactin, synthetic wastewater was also used, from which the engineered *B. subtilis* strain produced approximately 140 mg/L surfactin (Jung *et al.*, 2012).

Other studies focus on introducing changes in the amino acid and fatty acid composition of lipopeptides because they have a pronounced effect on the lipopeptide activity (Youssef *et al.*, 2011). For instance, surfactin with a pentadecanoic fatty acid side chain was found to be the most active with regards to hemolytic activity (Deleu *et al.*, 2003). Also, increasing the percentage of iso-even-numbered fatty acids compared to even-numbered fatty acids influences the lipopeptide activity as determined by the oil spreading assay (Youssef *et al.*, 2005). A study of the structure/function relationships associated with the three-dimensional structure has led to the recognition of the specific residues required for activity. Further studies will assist researchers in the selection of molecules with improved and/or refined properties useful in oil andbiomedical industries (Peypoux *et al.*, 1999). The required changes could e.g. be introduced by manipulating the specificity of YbdT, f.e. by increasing the specificity of this enzyme towards branched versus straight chain and even- versus odd-numbered fatty acids. By doing so, surfactin variants with a tailored mixture of 3-hydroxy fatty acyl moieties could be produced as such providing a new route to produce biosurfactants with activities tailored for specific functions (Youssef *et al.*, 2011). In terms of altering the peptide part of the molecule, knowledge about the modular arrangement of the peptide

synthetases is of utmost importance for combinatorial biosynthetic approaches. In this context, a systematic search for variants with altered peptide sequences was undertaken by hybrid-gene (Stachelhaus *et al.*, 1995; Schneider *et al.*, 1998) and directed-biosynthesis strategies (Peypoux *et al.*, 1994; Grangemard *et al.*, 1997). An enzymatically driven amino acid substitution was additionally suggested to be an interesting complement of molecular genetics or directed biosynthesis strategies used for variant engineering (Grangemard *et al.*, 1999). Due to its very interesting properties, organic chemists also tried to synthesize this compound, which was successfully performed by (Nagai *et al.*, 1996). However, the synthetic process requires many steps and therefore production costs are high, so the problem must be solved by a biotechnological approach.

One can imagine that a parallel with the abovementioned opportunities can be drawn for other lipopeptides like lichenysin. Studies on various aspects of other lipopeptides including an elucidation of the regulation of such biosurfactants are expected to swell in the coming years (Nerurkar, 2010).

I.4.1.3. *MELs*

In recent years, interest in MELs has increased due to their pharmaceutical applications. For example, Aventis Pharma Deutschland GmbH published a patent for a MEL, ustilipid, that is produced by *U. maydis* DSM 11494 and can be used in the treatment of schizophrenia or diseases caused by dopamine metabolic dysfunction (Vertesy *et al.*, 1998). Moreover, Japanese companies like Kanebo and Toyobo have commercialized these molecules in their products (cosmetics). Other applications of MEL, such as chemical tools for purification of proteins or as anti-agglomeration agents of ice-slurry are also known (Rau *et al.*, 2005). The first example of succeeded metabolic engineering of extracellular glycolipids produced by yeasts or fungi was presented by Hewald *et al*. (Hewald *et al.*, 2006) with the exclusive production of fully deacetylated MELs (MEL-D) by knocking out one gene (*mat1*) of the biosynthetic pathway. They further suggested production of fully acetylated MEL-A through overexpression of this Mat1 acetyltransferase as a large fraction of MELs acetylated at only one position is secreted by wild type *U. maydis* cells, indicating that the second acetylation is significantly slower than the first one. The importance of such achievements was stressed as MEL-A has some interesting properties not shared by the other less acetylated variants (e.g. a dramatic increase in gene transfection efficiency of liposomes, a property not shown for the partially acetylated MEL-B and MEL-C). They also attempted to stimulate MEL production

by *U. maydis* by overexpressing the glycosyltransferase gene *emt*1 with the arabinose-inducible *crg* promotor. Although this is a strong promotor only weak MEL production could be observed in arabinose medium which was most likely attributed to low glycolipid production in the presence of arabinose as carbon source (Hewald *et al.*, 2005). Mannosylerythritol compounds have been used for the production of diverse products and the Emt1 protein was therefore also suggested to have some interesting biotechnological applications (Hewald *et al.*, 2006).

I.4.1.4. *Cellobioselipids*

Cellobiose lipids are especially interesting with respect to development of novel biological antifungal preparations. (Kulakovskaya *et al.*, 2010). A biocontrol product, Sporodex ® based on the conidia of the basidiomycetous *P. flocculosa* active against powdery mildews is already commercialized and registered in some countries (Kiss, 2003). The most species producing these compounds belong to the order of the Ustilaginales: *Ustilago maydis, Pseudozyma fusiformata, Pseudozyma graminicola, Pseudozyma flocculosa and Sympodiomycopsis paphiopedili* (Kulakovskaya *et al.*, 2010). Only *U. maydis* (production of ustilagic acid (UA)) and *P. flocculosa* (production of flocculosin) have hitherto been subject to engineering approaches as the biosynthetic gene clusters were only quite recently discovered (Teichmann *et al.*, 2007; Teichmann *et al.*, 2011a). The reader is referred to I.2.2.2. for the respective gene names mentioned in the following overview of the results thereof.

A very interesting finding was the fact that constitutive expression of the regulator of UA biosynthesis in *U. maydis* (Rua1) gave rise to production of cellobioselipids even under non-inducing conditions (Teichmann *et al.*, 2010). Other literature data focus on modification of the cellobioselipid biosynthetic pathway with the aim of producing tailor-made CBL**s**. For example, expression of the *P. flocculosa fat3* gene in *U. maydis* results in the production of four additional UA derivatives corresponding to molecules carrying an extra acetyl group when compared to the wild type molecules. Flocculosin-like molecules were thus produced by *U. maydis*, which were never observed before in the culture supernatant of this fungus (Teichmann *et al.*, 2011a). In another study, the high selectivity of *cyp1* and *cyp2* for hydroxylation either at the ω or ω-1 position, respectively, was suggested to be attractive for biotechnological applications that require regio-specific introduction of hydroxyl groups (Teichmann *et al.*, 2007).

Furthermore, certain deletion strains of *U. maydis* represent interesting opportunities:

- deletion of *ahd1* responsible for α-hydroxylation leads to secretion of the naturally occurring UA derivatives lacking the α-hydroxyl group on the long chain fatty acid.
- deletion of the *cyp2* gene leads to the secretion of novel UA variants lacking the subterminal hydroxyl group.
- deletion of *uat1* leads to the production of UA intermediates lacking the acyl group, but UA derivatives lacking the distal glucose molecule (Teichmann *et al.*, 2011a) are also detected.
- deletion of *uhd1* on the other hand only results in the absence of the β-hydroxyl group of the short chain fatty acid attached to the distal glucose molecule (Teichmann *et al.*, 2011b).
- deletion of *uat2* leads to the production of UA derivatives lacking the acetyl group on the proximal glucose molecule.

The Atr1 transporter appeared to be quite unspecific, as many of the UA derivatives that are produced by the mutants were readily exported. UAs lacking the α-hydroxyl-group still show antibiotic activities (Teichmann *et al.*, 2011b) so this functionality might be superfluous, but no data about the possible lack of other characteristics is available yet. The gene products of *uhd1* and *fhd*1 seem to exhibit specific substrate specificity for short-chain fatty acids, which was suggested to make the enzymes suitable for practical applications in metabolic engineering. One could use them to specifically synthesize hydroxylated short-chain fatty acids, which have been reported to be valuable compounds in the chemical and medical industry (Teichmann *et al.*, 2011b).

These results demonstrate that genetic engineering of biosurfactant producers can aid in the specific production of certain compounds, which will be indispensable for the development of industrial standardized production processes. Unfortunately, *Pseudozyma flocculosa* does not respond easily to transformation, so transformants cannot be easily generated for this organism and are not available yet (Teichmann *et al.*, 2011a). One of the genes involved in cellobioselipid production of *U. maydis* has already been used for the tailored glycolipid production in *S. bombicola* (Roelants *et al.*, 2013).

I.4.1.5. *Sophorolipids*

SLs attracted industrial attention due to their excellent surface lowering properties, environmental friendly profile and good fermentation yields. The yeast *S. bombicola* is capable of producing up to more than 400 g/L of SLs (Daniel *et al.*, 1998). This efficient process resulted in commercialization of SLs for applications in e.g. cosmetics produced by Soliance (Sopholiance), the ecological cleaning products of Ecover and Wheatoleo (Sophoclean), dishwasher products of Saraya (Sophoron) and the filing of numerous patent applications. A review of 255 worldwide patents on bio-surfactants (Shete *et al.*, 2006) demonstrated that 24 % of those acted on SLs, clearly demonstrating the international commercial interest in SLs.

SLs could not only act as detergent, emulsifier or wetting agent in various conventional surfactant sectors (hard surface cleaners, degreasing agents, cosmetics, paints, food, shampoos,…), but due to their biological activity they also find potential applications as antimicrobial, immune modulating, antiviral and anticancer agents (Van Bogaert *et al.*, 2007b). Some of the earliest proposed applications of SLs were for cosmetics, deodorants and anti-dandruff agents. Additionally they are useful to treat skin disorders, as immunomodulators for Parkinsons's disease, Alzheimers's disease, psoriasis, AIDS treatment as well as for antiviral immunostimulation (Carr and Bisht, 2003). Moreover, they have several other pharmaceutical and medical applications. Other suggestions were the use of SLs as precursors for bioplastics and environmental applications (Maingault, 1999; Van Bogaert *et al.*, 2007b; Daverey and Pakshirajan, 2009b). Being biosurfactants a logical application lies, as mentioned above, in the formulation of detergents and cosmetics (skin care products) (Mager *et al.*, 1987), but they have also been suggested as stable consumable encapsulants for food and food oils.

For the successful industrial application of any surfactant, it is important that it remains stable at extreme conditions of temperature, pH and ionic strength. SLs are active at all tested salt concentrations (0-20 % NaCl), over a broad pH range of 2-10 and temperature (2 h heating time). Overall it could well be said that SLs are highly stable over a wide range of pH, ionic strength and temperature conditions and is therefore highly suitable for industrial applications.

Since small structural variations can have a significant influence on the biological activity or the physicochemical properties of a glycolipid, enzymatic or chemo-enzymatic synthesis of

customized SL derivatives has been the subject of several research papers (Gross *et al.*, 1999; Bisht *et al.*, 2000; Rau *et al.*, 2001; Carr and Bisht, 2003) and patent applications (Gross and Shah, 2004; Gross *et al.*, 2004; Gross and Shah, 2005). When the full SL biosynthetic pathway of *S. bombicola* was recently uncovered, the genetic modification approach in order to improve or customize its glycolipid production was thus also initiated:

- Deletion of the gene responsible for the first (*cyp52M1* gene) or second (*ugtA1* gene) step in SL biosynthesis results in a complete abolishment of SL production without any effect on the viability or growth rate of the yeast cells (Saerens *et al.*, 2011a; Van Bogaert *et al.*, 2013).

- Deletion of the second glucosyltransferase (*ugtb1*) results in the secretion of (acetylated) glucolipids (Saerens *et al.*, 2011c). Up until then glucolipids with a free carboxylic end were only produced by enzymatic conversion of acidic SLs obtained after alkaline hydrolysis of the crude SLs (Rau *et al.*, 1999; Saerens *et al.*, 2009) or by the microbial conversion of secondary alcohols or branched fatty alcohols into alkyl glucosides (Brakemeier *et al.*, 1998; Palme *et al.*, 2010). Acidic glucolipids (and SLs) especially attract attention because they are asymmetrical bola-amphiphiles that, in addition to the supramolecular structures they typically form, have increased chemical versatility as compared to the chemically synthesized symmetrical ones (Zhou *et al.*, 2004). The *ΔugtB1* deletion mutant was suggested to be an interesting strain as it offers *in vivo* production of these biomolecules starting from cheap renewable substrates. No lactonization of glucolipids has been observed. The appearance of acetylated glucolipids illustrates that the acetyltransferase, which normally decorates *de novo* SLs at their 6' and/or 6'' positions with acetyl groups, shows activity towards glucolipids as well. In this respect it is possible that acetylation of glucolipids occurs before the addition of the second glucosyl unit during *de novo* sophorolipid synthesis, without being necessary for this second glucosylation reaction. It has to be mentioned that these glucolipid molecules can also be produced in acetylated form by enzymatic conversion of the acidic acetylated SLs (Imura *et al.*, 2010).

- Deletion of the acetyltransferase (at) results in the production of exclusively non-acetylated SLs in both the acidic as lactonic conformation (Saerens *et al.*, 2011b). Such non-acetylated SLs have attracted attention for applications as antiviral drugs

(Shah *et al.*, 2005) or as starting molecules for the synthesis of dispersible nanoparticles (Kasture *et al.*, 2007) and glycolipid derivatives (Azim *et al.*, 2006).

- Deletion of the gene responsible for lactonisation of SLs (*lip*) results in the formation of exclusively acidic SLs with varying degrees of acetylation and this in yields comparable to the wild type (Ciesielska *et al.*, 2013) (cfr. chapter VI of this manuscript). This strain represents a great industrial opportunity as for certain applications one of the SL compounds is specifically required (Marchant and Banat, 2012), but the molecules are produced as a mixture by the wild type. Post-fermentative treatment is hence required when either the acidic or lactonic form is exclusively required. The abovementioned efficient producer offers a viable solution for the industrial production of acidic SLs, without the need of postfermentative modifications. An overexpression strain of this gene was also developed and is discussed in Chapter VI of this manuscript. Both strains are protected by a patent application that was filed in 2011, but is not public yet (Ciesielska *et al.*, 2011).

- Last but not least deletion of the SL transporter resulted in a significant drop of SL production (Van Bogaert *et al.*, 2013). However, as some SLs are still secreted (3 g/L in contrast to 30-40 g/L for the wild type), this suggests the involvement of additional active or passive SL transport systems (cfr. Chapter II of this manuscript). The engineering of the transport mechanisms of (modified) SL variants presents yet another opportunity for genetic engineering. A patent application was also filled for this transporter (Van Bogaert and Soetaert, 2009).

The fact that both (acetylated) glucolipids produced by the *ΔugtB1* deletion mutant and non-acetylated (lactonic and acidic) SLs produced by a *Δat* deletion mutant are secreted into the culture medium in addition to wild type SLs indicates a low substrate specificity of the SL transporter. However, the yields for these SL intermediates are a lot lower than those for the wild type being 3.9 g/L of glucolipids for the *Δugtb1* strain and 5 g/L of SLs for the *Δat* strain versus approximately 40 g/L for the wild type in shake flask experiments (Saerens, 2012). However, the secretion of these compounds could be the result of alternative secretion, as deletion of the transporter resulted in similar SL yields. For the *Δat* strain this lower yield wasn't attributed to toxic effects on *S. bombicola* cells as the possible toxicity of such molecules was investigated and this wasn't suspected to be true for the *Δugtb1* strain either. On the other hand, regulatory effects, which can be dual, could be causing these low yields. A

positive feedback loop exerted by the end products, which are now not formed, might exist. Alternatively, buildup of intermediates by suboptimal secretion of these intermediates could lead to negative feedback inhibition of the pathway.

The results obtained for the engineering of this highly productive SL producer were suggested to be of major biotechnological importance as new compounds can be obtained in a pure form and sometimes in presumably higher yields than the wild type (cfr. Chapter VI of this manuscript). However, for some of the mutants the selective production of SL intermediates results in a drop of the production, which should be overcome for industrial processes to be feasible.

I.5. Production and regulation of secondary metabolites in fungi

As already mentioned above, unraveling the genetics and regulation behind the biosynthesis of biosurfactants is indispensable for the tinkering, fine tuning and rearrangement of these biological pathways aiming for the efficient production of custom-made biosurfactants. Whereas the genetics and regulatory mechanisms for prokaryotic biosurfactants like rhamnolipids and surfactin have been extensively studied, the investigation of the genetics and (molecular) regulation behind eukaryotic biosurfactant production like cellobioselipids, MELS and SLs has only recently emerged. Although some insight has been obtained (cfr. I.2.2.), the exact regulatory mechanisms remain elusive. On the other hand, a lot of information about the genetics and regulation of other secondary metabolites produced by fungi is available. As several similarities with eukaryotic biosurfactant production were already observed, some general regulation mechanisms of fungal secondary metabolites will be shortly discussed below.

A preserved feature of genes encoding the biosynthetic pathways of secondary metabolites is the fact that they are usually clustered: in prokaryotes they are organized in one or more operons which can be co-regulated (Ruiz *et al.*, 2010), while in eukaryotes they are often arranged in co-regulated clusters (Keller and Hohn, 1997). Such clusters are found in the majority of filamentous fungi and may range from only a few to more than 20 genes. The reason for this clustering is not fully understood. However, both for prokaryotes and for eukaryotes, the clustering might reflect their introduction from another species by lateral (horizontal) gene transfer. This has for instance been proven for the cluster responsible for biosynthesis of penicillin (Landan *et al.*, 1990). It has been proposed that the organization of

secondary metabolites in gene clusters would confer an advantage for such spreading among different species (Walton, 2000). This “selfish cluster” hypothesis argues that only cluster organization guarantees the transfer of complete biosynthesis pathways upon horizontal gene transfer.

As stated above, generally spoken the genes in such clusters are co-regulated, with transcription of all the genes being activated or repressed simultaneously and independently of genes immediately outside the cluster. The transcription factor required for the regulation of the clustered genes can be either encoded by a gene of the cluster itself, or by a gene outside the cluster. In the first case, the gene cluster consists both of genes encoding enzymes that catalyze the steps in the biosynthetic process and of a transcription factor necessary for expression of these genes (Keller *et al.*, 2005; Hoffmeister and Keller, 2007). Such ‘narrow-domain transcription factors’ tend to bind only to promoter regions of the cluster genes and hence are considered pathway specific. A typical example of such a gene is *afl*R, found in the aflatoxin (AF) clusters of *Aspergillus flavus* and *A. parasiticus*, as well as in the sterigmatocystin (ST) cluster of *A. nidulans* (cfr. Figure I. 11; reviewed in (Shwab and Keller, 2008)).

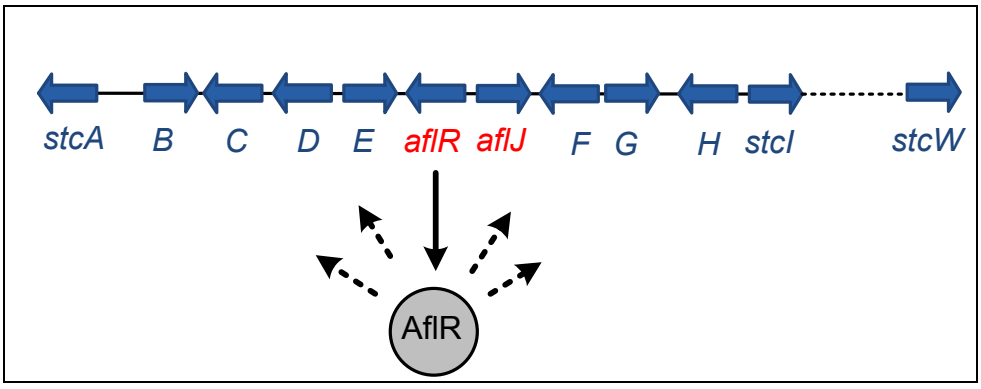


Figure I. 11 The sterigmatocystin gene cluster of *A. nidulans* is approximately 60 kb long and built up of biosynthetic genes *stc*A trough *stc*W and regulatory genes *afl*R and *afl*J. AflR is a positively acting transcription factor required for the expression of the *stc* genes adapted from (Shwab and Keller, 2008).

Both AF and ST are closely related carcinogenic mycotoxins derived from a similar gene cluster. Elimination of *afl*R results in loss of transcription of AF/ST biosynthetic genes and subsequent reduction in AF/ST production. This regulation is mainly limited to genes within the cluster, although several *afl*R-regulated genes were recently identified elsewhere in the

genome of *A. parasiticus*. The AF and ST clusters also contain the *afl*J gene of which the gene product interacts with AflR to obtain optimal expression of the cluster genes, although low levels of expression are observed in the absence of *afl*J in *A. parasiticus. Afl*R encodes a Zn_2Cys_6 zinc binuclear protein that activates transcription by binding palindromic DNA sequences in the promoters of the AF/ST cluster genes. Zinc binuclear proteins are the most common type of in-cluster pathway regulators in fungi. Other examples of this type of regulator are GliZ of *A. fumigatus*, required for expression of the gliotoxin cluster and MclR, required for compactin biosynthesis by *Penicillium citrinum*. Other in-cluster regulatory proteins include Cys_2His_2 zinc finger proteins such as Tri6 for trichothecene production by *Fusarium sporotrichiodes* and ankyrin-repeat proteins such as ToxE for production of HC-toxin by *Cochliobolus carbonum*, among others.

However, other biosynthetic gene clusters do not include a transcriptional regulator, which is the case for ergovaline and lolitrem gene clusters in the endophytes *Neotyphodium lolii* and *Epichloe festucae*. The level of transcription of these genes is very low in mycelia, but high in plants, suggesting that plant signaling pathways regulate the transcription (Fox and Howlett, 2008). The genes coding for the positive-acting PENR1 transcriptional complex required for penicillin production in a number of fungal species is also present outside the respective clusters (Litzka *et al.*, 1999).

In addition to pathway-specific regulators, fungal secondary metabolite (SM) production is also controlled at an upper hierarchic level by global transcription factors which regulate multiple physiological processes and generally respond to environmental cues such as carbon and nitrogen sources, pH, light, temperature and nutrition. In contrast to narrow-domain transcription factors, broad-domain transcription factors are encoded by genes located elsewhere in the genome and can bind to promoter regions in multiple SM clusters as well as to genes involved in other physiological processes. Thus, broad-domain factors are not pathway specific. It is likely that fungi turn the energetically costly process of secondary metabolite production off under certain environmental conditions, and on only when it would be advantageous. Signals generated in response to the environment are typically relayed through Cys2His2 zinc-finger proteins, including CreA for carbon signaling (Dowzer and Kelly, 1989), AreA for nitrogen signaling (Hynes, 1975), Zfr1 for sugar uptake (Flaherty and Woloshuk, 2004) and PacC for pH signaling (Tilburn *et al.*, 1995). These proteins may have either positive or negative regulatory effects on metabolite production. For example, penicillin

production is positively regulated by CreA and negatively regulated by PacC (Martin, 2000). Production of AF by *A. parasiticus* and *A. flavus* generally occurs only at acidic pH, though recently a South African isolate of *A. flavus* was identified in which low pH was inhibitory to AF production (Ehrlich *et al.*, 2005).

Nitrogen depletion is a usual suspect in the regulation of secondary metabolism. Possitive acting global regulatory genes like *are*A of *Aspergillus sp.* (Caddick, 1992) have been described in this respect. This so called GATA-type zinc finger transcription factor, contains a common DNA binding motif (Cys-X2-Cys-X17-Cys-X2-Cys) that recognizes a core 5'-GATA-3' sequence (Scazzocchio, 2000). It activates selective gene expression when a preferred nitrogen sources is lacking i.e. it mediates nitrogen catabolite derepression (Marzluf, 1997). A factor homologous to this GATA-factor for nitrogen control has also been identified for several other fungi (Fu and Marzluf, 1987; Minehart and Magasanik, 1991; Haas *et al.*, 1995). The involvement of such GATA factors has also been mentioned with respect to the regulation of glycolipid biosurfactant production (cfr. I.2.2.), but this remains to be investigated. Much still remains to be found out about nitrogen regulation in the well-established systems of *S. cerevisiae*, *A. nidulans*, and *N. crassa*, particularly regarding the nature of the signal(s) and the molecular details of the regulatory interactions (Wong *et al.*, 2008).

In addition to regulatory effects imposed by environmental cues, secondary metabolite production is also coordinated with the general development of the fungus. As with environmental conditions, it is advantageous for the fungus to produce certain secondary metabolites only at appropriate stages of its development. As such, a number of developmental pathways mediated by G-protein signaling also regulate secondary metabolite cluster expression. For example, signaling by the α-subunit of a heterotrimeric G-protein, FadA promotes vegetative growth and represses both sexual/asexual development and ST production in *A. nidulans* (cfr. Figure I. 12) as well as AF production in *A. flavus* and *A. parasiticus* (Calvo *et al.*, 2002). In contrast to ST, production of penicillin by *A. nidulans* is activated rather than repressed by FadA signaling, as is trichothecene production by *Fusarium sporotrichioides*, indicating that different secondary metabolites may also respond differently to developmental signals (Tag *et al.*, 2000).

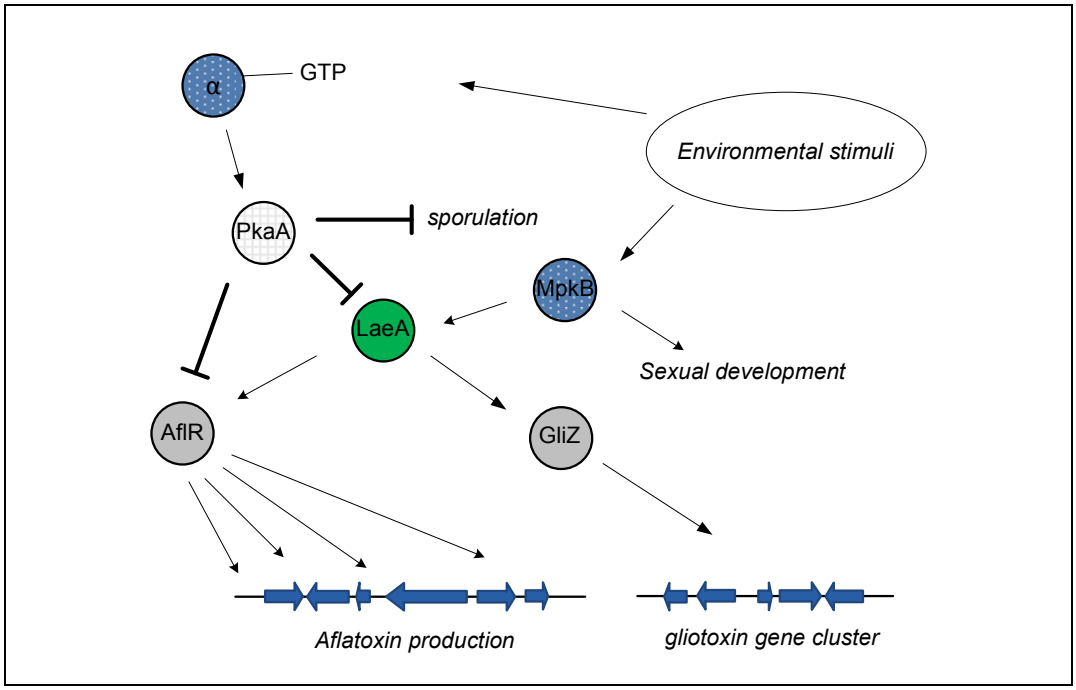


Figure I. 12 Signal transduction pathway connecting the general development of the fungus to secondary metabolite production in *Aspergillus nidulans*.

A global mechanism of secondary metabolite regulation in fungi was identified when the LaeA protein was discovered (Bok and Keller, 2004). LaeA is critical for expression, either positive or negative, of multiple SM gene clusters in *Aspergillus* species (including the ST cluster of *A. nidulans*) as well as in other fungi (Bok and Keller, 2004; Fox and Howlett, 2008; Shwab and Keller, 2008; Butchko *et al.*, 2012). The *lae*A gene is hence conserved in numerous filamentous fungi, except the yeast-like fungi. LaeA localizes to the nucleus and appears to be a methyltransferase, a category of enzymes known to play an important role in epigenetic gene regulation (Bok *et al.*, 2006b). In all clusters examined, regulation by LaeA is spatially limited to the genes within the cluster and does not extend to genes immediately adjacent (Bok *et al.*, 2006b; Fox and Howlett, 2008). Artificial introduction of additional genes in the cluster region results in a LaeA-dependent expression pattern, which confirms that LaeA exhibits an epigenetic control function (Bok *et al.*, 2006a). There is evidence that LaeA may alter the methylation state of certain nuclear proteins, possibly histones or other chromatin-associated proteins, and as such modify the chromatin structure at secondary metabolite cluster loci (Fox and Howlett, 2008) e.g. mutation of an s-adenosyl methionine binding site results in a Δ*lae*A phenotype (Bok *et al.*, 2006b). More specifically LaeA has been proposed to counteract H3K9 methylation in the sterigmatocystin gene cluster (Reyes-Dominguez *et al.*, 2010). Furthermore, loss of several other chromatin-modifying proteins in

A. nidulans results in up-regulation of a number of secondary metabolites (Shwab and Keller, 2008). Mutation of these genes also results in partial remediation of the loss of metabolite production in Δ*lae*A.

The prevalence of secondary metabolite clusters in subtelomeric areas also supports the possibility of epigenetic regulation as such regulation is commonly associated with certain subtelomeric genes in a wide variety of eukaryotes (Robyr *et al.,* 2002; De las Penas *et al.*, 2003; Halme *et al.*, 2004; Domergue *et al.*, 2005; Freitas-Junior *et al.*, 2005; Horn and Barry, 2005). Subtelomeric regions of fungal chromosomes are also found to be highly variable and often contain genes involved in niche specialization, which is in agreement with a role for secondary metabolites in adaptation to a specific environment (Fairhead and Dujon, 2006; Rehmeyer *et al.*, 2006; Perrin *et al.*, 2007). It is assumed that clustering confers some selective advantage to the fungus, and it is likely that one of these advantages may be related to efficiency of gene regulation. The chromatin-based regulation for these clusters is thus an attractive hypothesis. However, further experimentation is needed to determine whether or not epigenetic regulation provides sufficient explanation for the clustering of fungal secondary metabolite genes.

I.6. Conclusions

Biosurfactants produced by micro-organisms comprise a very diverse group of molecules with differing physicochemical and biological properties. Their biodegradability and the fact that they can be produced from renewable resources gives them an advantage over their chemical counterparts and may therefore make them suitable to partly replace chemicals. Two factors currently limiting the penetration of biosurfactants into the market are the limited structural variety and the rather high production price (due to suboptimal yields). These two factors could both be resolved by the genetic engineering of the natural producers, for which some examples have been provided above. The expanding knowledge of the genetics and molecular regulation of biosurfactant production will be of paramount importance as this knowledge represents the necessary base for the creation of good recombinant producers.

A lot of research has been dedicated to the elucidation of the biosynthetic pathways and corresponding regulatory networks of prokaryotic biosurfactants like surfactin and rhamnolipid. This is in contrast to eukaryotic biosurfactant biosynthesis and regulation thereof, which remained elusive for a long time. However, some major scientific breakthroughs have

been achieved in this field in the last seven years: the biosynthetic pathways and corresponding genes for glycolipid biosurfactants like mannosylerythritollipids, cellobioselipids (ustilagic acid and flocculosin) and sophorolipids have now been described. Moreover, part of the molecular regulation of cellobioselipid biosynthesis by *Ustilago maydis* has been described quite recently and some analogy with the regulation of biosynthesis of other secondary metabolites (SM) by fungi could be concluded. The biosynthetic genes were also often found to be clustered near the telomere, which was demonstrated to have a function in the regulation of the SM biosynthesis. Moreover, an 'in cluster' regulator was identified in the cellobioselipid gene cluster, which is also often the case for other SMs in fungi. Continued research is required to further uncover the underlying machinery of biosurfactant production in eukaryotes. The latter is indispensable for the development of enhanced recombinant strains, which may well become industrial strains for (tailored) glycolipid production.

A yeast strain of high industrial relevance is *Starmerella bombicola*, as it is capable of producing high yields of the glycolipid biosurfactant sophorolipids (SL) (over 400 g/L), which already resulted in commercialization of these molecules in ecological cleaning solutions (Ecover and Wheatoleo) and cosmetics (Soliance). The SL biosynthetic pathway and the corresponding gene cluster have been recently described and discovered and the biotechnological opportunities this offers will be investigated in this manuscript.

PART I:

UNRAVELING THE SOPHOROLIPID METABOLISM OF STARMERELLA BOMBICOLA


UNIVERSITEIT GENT

FACULTEIT BIO-INGENIEURSWETENSCHAPPEN

CHAPTER II:

THE BIOSYNTHETIC SOPHOROLIPID GENE CLUSTER

Part of this chapter has been accepted for publication:

Van Bogaert, I. N. A., Holvoet, K., Roelants, S. L. K. W., Li, B., Lin, Y., Van de Peer, Y., Soetaert, W., 2013. The biosynthetic gene cluster for sophorolipids: a biotechnological interesting biosurfactant produced by *Starmerella bombicola*. Molecular Microbiology (accepted and in production)

Chapter II.

THE BIOSYNTHETIC SOPHOROLIPID GENE CLUSTER

II.1. Introduction

Sophorolipids (SLs) are glycolipid biosurfactants consisting of a sophorose molecule linked O-glycosidically to a fatty acid (cfr. Figure I. 10). They were first identified as glyolipids produced by *Torulopsis* species (Gorin *et al.*, 1961). Structural variants of SLs are produced by several (un)related yeast species (cfr. Chapter I). Biosynthesis of such molecules by *R. bogoriensis* was first investigated by (Esders and Light, 1972b). They demonstrated the involvement of glucosyl- and acetyltransferases and were the first to propose a biosynthetic pathway consisting of the formation of a glucolipid from UDP-glucose and a hydroxylated fatty acid, followed by a second glycosylation (by the same or another protein), yielding non-acetylated SLs. The disaccharide moiety is subsequently acetylated by the action of an acetyltransferase(s) yielding mono- and di-acetylated SLs.

Although the involvement of glucosyl- and acetyltransferases in SL biosynthesis of *R. bogoriensis* was thus already proven in the early seventies, only in 2009 the discovery of a gene involved in SL biosynthesis by *S. bombicola* (Van Bogaert *et al.*, 2009a) was reported. In the years to follow, other genes involved in SL biosynthesis of this yeast were identified and the corresponding enzymes were characterized (Saerens *et al.*, 2011a; Saerens *et al.*, 2011b; Saerens *et al.*, 2011c). Although some inclarities still existed at the time, a review of the full hypothetical pathway was given by (Van Bogaert *et al.*, 2011c) (cfr. Figure II. 1). The first step in SL biosynthesis consists of the (sub)terminal hydroxylation of a fatty acid by the action of a cytochrome P450 monooxygenase (*cyp52M1*) (Van Bogaert *et al.*, 2009a). In contrast to CBL biosynthesis two, instead of one, glucosyltransferases are involved. The first one (*ugta1*) (Saerens *et al.*, 2011a) is responsible for the transfer of a glucose molecule from UDP-glucose to the hydroxylated fatty acid giving rise to a glucolipid and UDP, while a second glusocyltransferase (*ugtb1*) (Saerens *et al.*, 2011c) specifically transfers a second glucose molecule from UDP-glucose to the formed glucolipid (and not to the hydroxylated fatty acid). The resulting SLs are subsequently acetylated by the action of an acetyltransferase

(*at*) (Saerens *et al.*, 2011b). Although almost all of the genes responsible for SL biosynthesis were thus described, some information remained missing, like the process leading to lactonisation of SLs, the secretory pathway and the molecular regulation of this biosynthetic pathway. Lactonisation of SLs by the action of a (cell wall-bound) lipase (*lip*) was already suggested for *Candida apicola* by (Hommel *et al.*, 1994b). The responsible enzyme and respective gene (*lip*) was recently discovered for *S. bombicola* and proven to be fully responsible for lactonisation of SLs (Ciesielska *et al.*, 2013).

Genes encoding the biosynthetic pathways of secondary metabolites are often arranged in large co-regulated gene clusters (cfr. Chapter I). Expression of all the genes is often activated or repressed simultaneously and independently of genes immediately outside the cluster. In some cases the transcription factor required for the regulation of the clustered genes and specifically acting on the genes within the cluster, is also present within the gene cluster (Fox and Howlett, 2008). Two such biosynthetic gene clusters were recently identified for the glycolipid biosurfactants cellobioselipids, produced by two members of the *Ustilaginales*, namely *Ustilago maydis* and *Pseudozyma flocculosa*. They both contain a transcriptional activator (Rua1 and Rfl1 respectively) necessary for transcriptional activation of the cluster genes (Teichmann *et al.*, 2010; Teichmann *et al.*, 2011a). Secondary metabolite production is additionally also controlled at an upper hierarchic level by global transcription factors, which regulate multiple physiological processes and generally respond to environmental cues such as pH, temperature and nutrition. A biosynthetic gene cluster containing almost all the genes of the biosynthetic pathway was recently discovered in the genome of *S. bombicola* by analyzing whole genome sequencing data (cfr. Figure II. 1). The missing SL transporter was found to be present in this gene cluster (Van Bogaert *et al.*, 2013), similarly as is described for other secondary metabolites (cfr. Chapter I).

In this chapter the architecture and surrounding genomic regio of this gene cluster will be discussed. The abovementioned regulators of cellobioselipid biosynthesis were both found at the borders of the respective gene clusters, so the possibility existed that this was also true for the SL gene cluster of *S. bombicola.* Furthermore some *in silico* analyses of the now sequenced genome were done. Unraveling the genetics behind the biosynthesis of SLs is indispensable for further genetic engineering of *S. bombicola* and exploiting its efficient SL machinery.

II.2. Materials and methods

II.2.1. *Strains, media and culture conditions*

Starmerella bombicola ATCC 22214 was used as the parental or wild type strain. Yeast cells were grown on YPD medium (1 % yeast extract, 2 % peptone and 2 % glucose) (cfr. II.2.2) and 3C medium (10 % glucose, 1 % yeast extract, 0.1 % ureum and 15 % agar) was used to maintain the strains and to produce single colonies prior to performing liquid shake flask experiments for SL production. The latter were performed in the production medium described by (Lang *et al.*, 2000). Rapeseed oil (Sigma; 3.75 %) was added after 48 h when mentioned. Precultures were prepared from single colonies derived from 3C plates in 5 mL of the production medium in round bottomed culture tubes. These overnight grown precultures were inoculated (2 %) in shake flasks of 500 mL containing 100 mL of the production medium. The shake flask cultures were sampled regularly for SL composition, pH determination, glucose concentration.

Escherichia coli DH5α, XL10GOLD and Fusion Blue Ultracompetent cells were used for cloning experiments. *E. coli* cells were grown in Luria-Bertani (LB) medium (1 % trypton, 0.5 % yeast extract, 0.5 % sodium chloride (+15 % agar for plates) supplemented with 100 mg/L ampicillin. *E. coli* cultures were incubated at 37 °C on a rotary shaker (200 rpm).

II.2.2. *Molecular techniques*

II.2.2.1. *General techniques*

Yeast genomic DNA was isolated from overnight yeast cultures grown on YPD. The yeast cell wall was first removed enzymatically. This was done by incubation (90 min at 37°C) of the cell pellet derived from 1 mL of yeast culture with 0.80 g Yeast Lytic Enzyme (Sigma)/g wet cell weight in SCE buffer (1M sorbitol, 0.1M sodium acetate and 60 mM EDTA, pH 7.5) in presence of 3.75 µl β-mercapto-ethanol. Genomic DNA was isolated from the remaining protoplasts by means of the GenElute^{TM} Bacterial Genomic DNA kit (Sigma). Bacterial plasmid DNA was isolated using the QIAprep Spin Miniprep Kit (Qiagen) and the Qiaquick Gel extraction kit (Qiagen) was used for DNA purification from gel. Restriction nucleases were obtained from New England Biolabs (NEB) and restriction digests were performed as specified by the supplier. Primer design, sequence analysis and strategy design were performed using the Clone Manager Professional Suite software (Version 8.0). Primers were

ordered at Sigma Genosys and details can be found in Table II.1. All high fidelity PCR reactions were performed using the *Pfu* high fidelity polymerase unless stated otherwise. PCR products were purified using the SureClean purification kit (Bioline), cloned into derivatives of the pGEM-T® (Promega) vector and sent to LGC genomics (Germany) for sequence analysis. *S. bombicola* cells were transformed using a standard electroporation protocol (Saerens *et al.*, 2011a) and transformants were selected on YPD agar (15 %) plates supplemented with hygromycine (500 mg/mL). *E. coli* cells were transformed as described by (Sambrook and Russell, 2001) and selection occurred on LB agar plates supplemented with ampicillin. Colony PCRs both on *E. coli* and *S. bombicola* were performed using *Taq* polymerase (NEB).

II.2.2.2. *Knocking out of the two ORFs delimiting the gene cluster*.

Knock outs were created by replacing the largest part of both open reading frames (*orf1* and *orf2*) by a selective marker. Orf1 was predicted to be an alcoholdehydrogenase (most similar to an alcohol dehydrogenase of the prokaryote *Idiomarina baltica* (ZP_01042748.1) and will be named '*adh'*. Orf2 was termed '*regul*' because it could possibly encode a transcriptional regulator for the SL gene cluster. The selective marker that was used for both strategies was a hygromycine resistance gene under control of the strong constitutive *gapd* promoter (P*gapd*) and *Herpes simplex* tyrosine kinase (TK) terminator (Van Bogaert *et al.*, 2008a). The two orfs (with up- and downstream regions) flanking the SL biosynthetic gene cluster were picked up from genomic DNA of the *S. bombicola* wild type, with the respective primerpairs P69_FOR_*regul*up / P68_REV_*regul* and P70_FOR_*adh*up / P72_REV_*adh*down for the *regul* and *adh* ORF respectively. The two obtained PCR fragments were cloned into the pGEM-T® (Promega) vector, giving rise to pGEM-T_*regul* and pGEM-T_*adh*. These two vectors were digested with *Bam*HI / *Eco*NI and *Psh*A1 / *Hind*III respectively and the vector backbones were purified from gel. The obtained linear fragments were subsequently treated with Mung Bean Nuclease (NEB) and ligated with the selective marker, picked up from the vector pGEM-T_P*gapd_hygro* (Van Bogaert *et al.*, 2008a) with primerpair GAPDup_T7 / GHinf47Rev2. The resulting vectors (pGEM-T_*regul*KO_P*gapd_hygro* and pGEM-T_*adh*KO_P*gapd*_hygro) respectively were sent to LGC genomics (Germany) for sequence analysis. Both vectors subsequently served for picking up of the knock out cassettes by means of PCR using primerpairs P68/P69 for *regul* and P70/P72 for *adh*. The fragments were used to transform the *S. bombicola* wild type and selection occurred on YPD plates supplemented with hygromycine (500 mg/mL). Correct integration of the knock-out cassettes was controlled

by colony PCR using primerpairs P450 A21 down / *gapd*UP_SP6 for the *regul* orf and P70 / *gapd*UP_SP6 for the *adh* orf.

Table II. 1 Primers used for the creation of the knock out strains.

Primer names	**'5 Sequence 3'**
GAPDup_T7	GGATTAACCCGTGAGGAGAAG
GHinf47Rev2	TTTGAACAAACGACCCAACAC
P68_REV_*regul*	CTTTGTATCGGGTCTATGGG
P69_FOR_*regul*UP	GTTTCTTAGCCTCCCATGGAAG
P70_FOR_*adh*UP	ACGCAAGAGCTGTCTAACTG
P72_REV_*adh*DOWN	TTGCAACTGTGCCCTCCATC
P450 A21 down.n	CCAATCGATGGGAGAACTTGAAGCTAGAG
*gapd*UP_SP6	GGCTTCGATTGCTCGTATTG

II.2.3. *Sampling and analysis*

II.2.3.1. *Follow up of growth*

Optical density (OD) of cultures was measured at 600 nm using the Jasco V 630 bio spectrophotometer (Jasco Europe) of 1 mL of samples diluted with physiological solution (0.9 % NaCl in dH_2O) when required. Glucose concentration was determined using the 2700 select biochemistry analyser (YSI Inc.), samples were diluted below 7.5 g/L with dH_2O.

II.2.3.2. *Analytical/total extraction and analysis of sophorolipids*

Analytical SL samples were prepared as follows: 440 µL ethylacetate and 11 µL acetic acid were added to 1 mL of culture broth and shaken vigorously for 5 min. After centrifugation at 10 000 rpm for 5 min, the upper solvent layer was removed and translocated into a fresh eppendorf tube, containing 700 µL of ethanol. Although excellent extraction of SLs is obtained with this method, the possible formation of new more hydrophilic compounds could be missed so an alternative method for analysis of SL production was thus also used. This second method consisted of the addition of 2 volumes of ethanol to 1 volume of culture broth after which the mixture was shaken vigorously. The cell remains were subsequently allowed to settle down and the supernatant consisting of a $EtOH/H_20$/SL mixture was translocated into a HPLC vial. SL samples were analysed by HPLC.

HPLC was performed with a Varian Prostar HPLC system and a ChromolithR Performance RP-18e 100-4.6 mm column from Merck KGaA at 30 °C coupled to an Evaporative Light Scattering Detection (Alltech) at 40 °C. A gradient of two eluents, 0.5 % acetic acid aqueous

solution and 100 % acetonitrile (ACN), was used to separate the components. The gradient started at 5 % ACN and linearly increased to 95 % in 40 min. The mixture was kept like that for 10 min and was then brought back to 5 % ACN in 5 min. A flow rate of 1 mL/min was applied.

To quantify and compare the produced SLs, total SL extractions were performed. Three volumes of ethanol were added to the culture broth at the end of the incubation period and cell debris was removed by centrifugation at 1500 g during 10 min. For further gravimetric determination of the amount of produced SLs, the water-ethanol mixture was evaporated. Two volumes of ethanol were subsequently added to dissolve the SLs and the residual hydrophobic carbon source. The mixture was filtrated to remove the water-soluble compounds and was evaporated again. One volume of water was added and the resulting solution was set at pH 7, then one volume of hexane was added and the mixture was allowed to separate after vigorous shaking. The different fractions were collected, evaporated and the mass was determined. The hexane phase contains the residual oil, while the water phase contains the SLs.

II.2.4. *Bioinformatics*

The BOGAS genomic database of *S. bombicola* that was developed at the lab of Prof. Yves Van de Peer (Bioinformatics & Systems Biology, VIB) by PhDs Bing Li was used to perform searches in the *S. bombicola* genome. The sequenced and annotated genome of *S. bombicola* can be browsed with this database and for each orf a collection of data retrieved from public databases is summarized. All found protein domains are mapped using InterProScan. For each domain, the page provides the domain name, the database used to find this domain and a description of the domain. Protein homologs retrieved by Blast (NCBI) with and e-value lower than 1e- 5 are also shown. For each hit the gene name and description as well as the scores are presented.

The MEME algorithm (http://meme.nbcr.net/meme/cgi-bin/meme.cgi) (Bailey *et al.*, 2009) was used to identify potential regulatory sequences within the promoter regions of the cluster genes.

II.3. Results and discussion

II.3.1. *Architecture and regulation of the SL biosynthetic gene cluster*

Genome sequencing data revealed that, similarly as for other glycolipid biosurfactants, the genes that were already known to be involved in SL biosynthesis were located in one large biosynthetic gene cluster (± 11 kb) in the genome of *S. bombicola* (cfr. Figure II. 1). A gene that had not been described yet, but with high similarity to multidrug resistance proteins (MDR) of the ATP-binding cassette (ABC) superfamily was also found to be present in the gene cluster. Not entirely unexpectedly, knocking out this gene resulted in a major drop of SL titers to about 10 % of those obtained with the wild type (Van Bogaert *et al.*, 2013). These results demonstrated that this gene encodes the missing SL transporter, although some alternative transport mechanisms must also exist as knocking out the gene does not lead to a complete abolishment of SL production. The full biosynthetic pathway and cluster is depicted in Figure II. 1. The genes *cyp52M1*, *at* and the gene encoding the SL transporter are located on the same strand, whereas both glucosyltransferase genes, *ugtA1* and *ugtB1*, are located on the complementary strand and are hence transcribed in the opposite direction. Now that this SL biosynthetic gene cluster has been identified, the architecture, localization in the genome and regulation can be investigated.

II.3.1.1. *Defining the borders of the gene cluster*

To define the borders of the SL biosynthetic gene cluster, two knock out (KO) strains of the respective open reading frames (orf) delimiting the SL biosynthetic gene cluster were created. The first strain was knocked out in the orf delimiting the cluster at the left hand side (orf1), which encodes a putative alcoholdehydrogenase (*adh*) as was predicted by the BOGAS database. The nearest hit was an alcoholdehydrogenase of *Chromohalobacter salexigens* (52 % AA ID). In *Ustilago maydis*, an alcohol dehydrogenase, located at the rightmost border of the cellobioselipid gene cluster adjacent to the telomeric region (cfr. Chapter I, Figure I. 9), was found to be responsible for α-hydroxylation of the long chain length fatty acid of cellobioselipids (cfr. Chapter I, Figure I. 8 b) (Teichmann *et al.*, 2007). The second strain was knocked out in the orf delimiting the cluster at the right hand side (orf2). This orf was termed '*regul*', because it could possibly encode an 'in cluster' regulator of the other genes of the cluster, as is the case for *U. maydis* and *P. flocculosa* (cfr. Chapter I. 2.2.2).

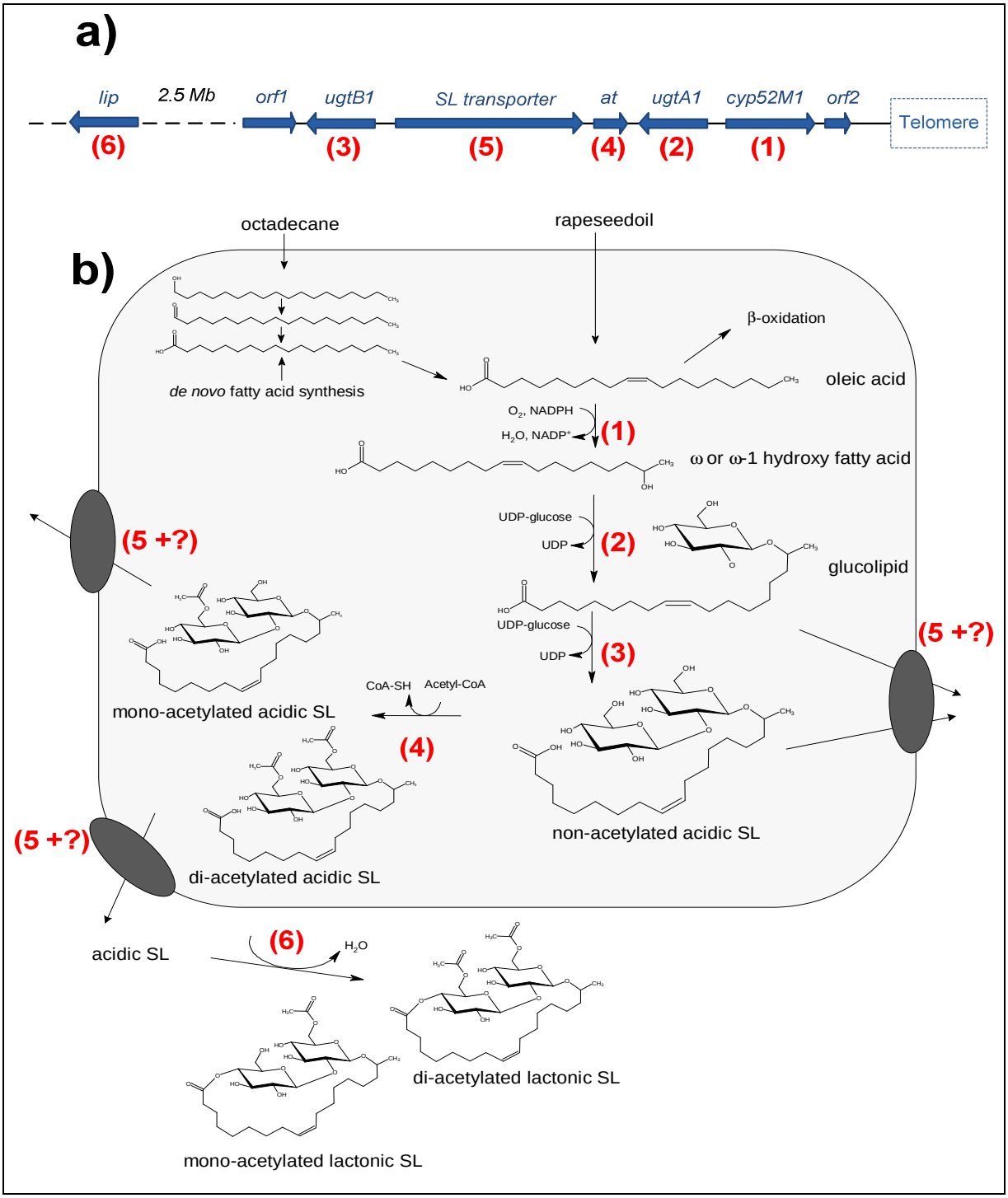


Figure II. 1 **(a)** Illustration of chromosome II of *S. bombicola* containing the full SL biosynthetic gene cluster (± 11 kb) and the gene responsible for lactonisation (*lip*) at the other side of the chromosome **(b)** The full SL biosynthetic pathway consisting of (1) hydroxylation of a fatty acid (C16 or C18) by a CYP52M1 monooxygenase (2) glucosylation of the FA-OH by the first glucosyltransferase UGTA1 (3) second glucosylation step of the formed glucolipid by a second glucosyltransferase UGTB1 giving rise to an acidic SL, which can be (4) acetylated by the action of an acetyltransferase AT. The different SLs are transported into the extracellular space by a multidrug transporter protein (5), but other transportsystems are also involved. Lactonisation (6) mainly occurs extracellularly as the responsible enzyme is secreted.

Both knock outs were created by replacing the largest part of the orfs by an antibiotic selective marker (hygromycine resistance gene) (cfr. Figure II. 2) under control of the strong constitutive *gapd* promotor (Van Bogaert *et al.*, 2008a).

- *Investigation of the potential role of orf 1 in SL biosynthesis*

The orf delimiting the leftmost border of the SL biosynthetic gene cluster (orf1) is as mentioned a putative alcohol dehydrogenase (*adh*). A knock out cassette for orf1 was created as described in II.2.2.2. and transformed into the *S. bombicola* wild type (cfr. Figure II. 2). Selection occurred on selective plates (YPD) containing hygromycine and colonies were controlled by colony PCR for the correct genotype. Of the forty six tranformants that were controlled by colony PCR, three that were confirmed to be successfully knocked out in orf1, were cultivated on the production medium to assess if deletion of this putative gene affected cell viability and/or SL machinery.

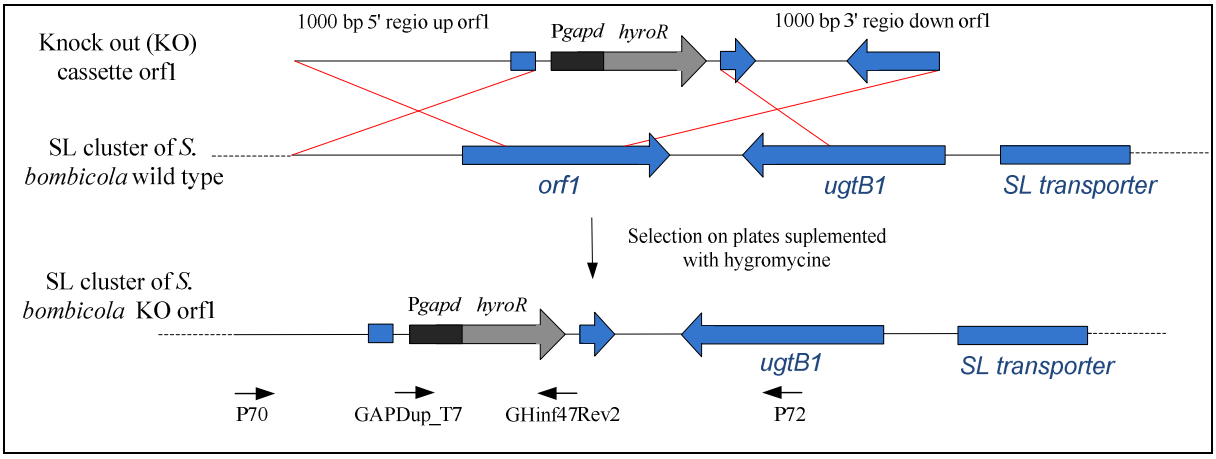


Figure II. 2 Schematic representation of the knock out cassette created for orf1 and integration thereof in the *S. bombicola* genome. The selective marker consisted of an hygromycine resistance gene (*hygroR*) under control of the strong constitutive promotor P*gapd*. The genotype of the correct transformants is depicted on the bottom of the figure and the most important used primers were also included. Deletion of orf2 occurred in exactly the same way and is hence not shown here.

The three mutants showed identical behaviour concerning growth and glucose consumption as compared to the wild type and to each other (data not shown). SL production and composition was also unaffected for the deletion strains as compared to the wild type. No clear fenotype could thus be assigned to the deletion of this orf and it has no clear function in SL biosynthesis on the production medium.

A possible role for this putative gene could be the involvement in growth and SL production of *S. bombicola* starting from n-alkanes, like was described in several papers (Ito and Inoue, 1982; Hommel, 1990). Aerobic conversion of n-alkanes to fatty acids namely requires alcoholdehydrogenase activity to convert alcohols formed by e.g. terminal hydroxylation of the n-alkane by the presumed activity of cytochrome P450 monooxygenases to an aldehyde (Wentzel *et al.*, 2007). Several candidate P450 genes of the *cyp52M1* family were found in the *S. bombicola* genome by genome walking experiments (Van Bogaert *et al.*, 2009a) and analysis of the now available genome reveals the presence of no less than seven such putative genes, besides the one involved in SL biosynthesis (*cyp52M1*). The presence of an alcoholdehydrogenase adjacent to the SL biosynthetic gene cluster might thus be indicative of the involvement of the corresponding enzyme in SL biosynthesis starting from n-alkanes.

- *Investigation of the potential role of orf2 in SL biosynthesis*

The small orf delimiting the rightmost border of the SL biosynthetic gene cluster is not similar to any other known proteins. As stated above, several gene clusters encoding secondary metabolites (SM) also encode a specific regulator involved in the biosynthesis of said SM inside the gene cluster. For the cellobioselipid gene clusters (both for *U. maydis* and *P. flocculosin*) such 'in cluster' transcription factors (*rua1* and *rfl1*) are located at one of the borders of the gene cluster (cfr. Chapter I. 2.2.2). The possibility thus exists that this small orf, delimiting the SL biosynthetic gene cluster, could possibly encode a specific regulator for SL biosynthesis. A knock out cassette for this orf was thus constructed as described in II.2.2.2 and transformed into the *S. bombicola* wild type similarly as depicted in Figure II.2. After selection on selective plates and subsequent colony PCR, three colonies for which orf2 had been correctly knocked out, were cultivated on SL production medium. Similarly as for the orf1 KO, growth for the three orf2 KOs under investigation was similar as compared to the wild type and amongst each other (data not shown). Glucose consumption for two of the three mutants was somewhat slower in the exponential growth phase, but SL composition then again was similar to the wild type, though suboptimal production could be argued for one of the mutants. However, shake flask production experiments are subject to great variation, as shake flask cultivation of the wild type also sometimes results in suboptimal SL production. Deletion of the abovementioned regulator (*rua1*) of ustilagic acid (UA) production resulted in the absolute abolishment of glycolipid production. This was thus clearly not the case for deletion of orf2, eliminating the possibility that this hypothetical protein has a similar function

as Rua1. No clear fenotype could thus be assigned to this orf2 either and it thus has no clear function in SL biosynthesis. No in-cluster specific regulator is thus involved in regulation of SL biosynthesis, which is not unusual as other eukaryotic biosynthetic gene clusters for secondary metabolites do not include a transcriptional regulator either e.g. the ergovaline and lolitrem gene clusters of the endophytes *Neotyphodium lolii* and *Epichloe festucae* (Fox and Howlett, 2008).

- *Regions up- and downstream of the SL biosynthetic gene cluster*

Further upstream of orf1, another putative cluster was found: six genes similar (42-64 % AA ID) to the bacterial *ent*A to -F genes responsible for formation of the siderophore enterobactin, combined with a putative transporter and a transmembrane ferric reductase. It seems like all required elements attributed to siderophore formation are there. Recent unpublished RNA sequencing data furthermore confirmed that this cluster is functional, as it was clearly upregulated in the stationary growth phase. Hence, the core SL biosynthetic gene cluster is not extended at the 5' side beyond the *ugtB1* gene.
At the right-hand side the cluster is delimited by subtelomeric and telomeric (TTAGGG)n repeats), which is also the case for the UA gene cluster of *U. maydis* (cfr. Chapter I.2.2.2). The fact that the biosynthetic gene cluster is located near the telomere is not coincidental as telomeric localisation of several SM gene clusters of yeasts and fungi near the telomere has been described and has been linked to the regulation of such gene clusters (cfr. II.3.1.2).

II.3.1.2. *Regulation of the SL biosynthetic gene cluster*

- *Transcription factor based regulation*

Since SL biosynthesis has been linked to nitrogen limitation (cfr. Chapter I.2.2.3), one might expect GATA-like regulatory sequences (cfr. Chapter I.5) in the upstream regions of genes involved in this pathway. Such sequences are recognized by GATA factors, which are responsible for regulating the expression of nitrogen-regulated genes (Magasanik and Kaiser, 2002). These factors are employed to globally control nitrogen metabolism in yeasts and fungi enabling them to utilize a wide variety of nitrogen sources. However, such global regulatory proteins have also been linked to regulation of secondary metabolite biosynthesis, as these are also often regulated by the nitrogen metabolism (cfr. Chapter I). This regulation by GATA-factors could occur in a direct manner, through binding of these global regulators to GATA

upstream activating sequences (UAS) (consensus: WGATAR) in the promotor region(s) of the SM genes or indirectly, by regulating the expression of a specific regulator required for expression of the genes of the SM pathway. The last option holds true for the cellobioselipid (UA) biosynthetic gene cluster of *U. maydis* (cfr. Chapter I.2.2.2). A conserved upstream activating sequence (UAS) (T/G)(G/T)CGCAT(A/T)(C/T)(C/T)(G/A) was found to be present in all the promoters of the clustered structural genes involved in UA biosynthesis and was shown to mediate Rua1 (an 'in cluster' regulator) dependent expression of these genes (Teichmann *et al.*, 2010). The authors suggested that regulatory proteins that sense global and specific nitrogen availability are in turn involved in Rua1 expression.

The same could hold true for SL biosynthesis by *S. bombicola* as a pathway specific regulator can be present elsewhere in the genome and/or direct regulation exerted by GATA factors might be involved. The intergenic regions of the clustered SL biosynthetic genes were thus investigated and several GATA consensus motifs: (A/T)GATA(A/G) were indeed found to be present in at least one copy in the 5' upstream regions of the clustered genes, sometimes only 50 bp of each other, which was proven to promote strong binding of GATA factors (Magasanik and Kaiser, 2002). The motifs are depicted as orange triangles in Figure II.3 c. The presence of these motifs might indicate the involvement of a global and/or more specific GATA factor regulating expression of these genes. When inspecting the *S. bombicola* genome several putative GATA factors presumably involved in such processes were found by homology with known GATA factors involved in nitrogen metabolism (cabom03g03750, cabom03g08130 (AreA/Nit2), cabom02g04710 (Nit2), cabom03g14570 (AreB/Nit2)).

Furthermore a bioinformatical approach was also done to assess whether a more specific consensus motif might be present in the promotor regions of the gene cluster. This was done using the MEME algorithm (http://meme.nbcr.net/meme/doc/overview.html). This resulted in the identification of a consensus motif that was present in one or more copies in all the promotor regions of the genes of the cluster, being CAG(C/G)(G/T/C)TGCAT (cfr. Figure II. 3a). The motif is represented by green triangles in the promotor regions of the SL biosynthetic pathway. Filled symbols represent putative motifs for which the first three and last five basepairs were strictly conserved being CAG and TGCAT respectively. The other two basepairs varied between G, C and T. Putative TATA boxes (consensus sequence: TATA(A/T)A(A/T)(A/G) are depicted as red boxes. Such TATA-boxes were identified in the 50-200 bp upstream of highly regulated or stress-induced genes in *S. cerevisiae* (Basehoar *et*

al., 2004). The possible regulatory function of such motifs in SL biosynthesis can be assessed by introducing point mutations and performing biological experiments, which should be executed in the future.

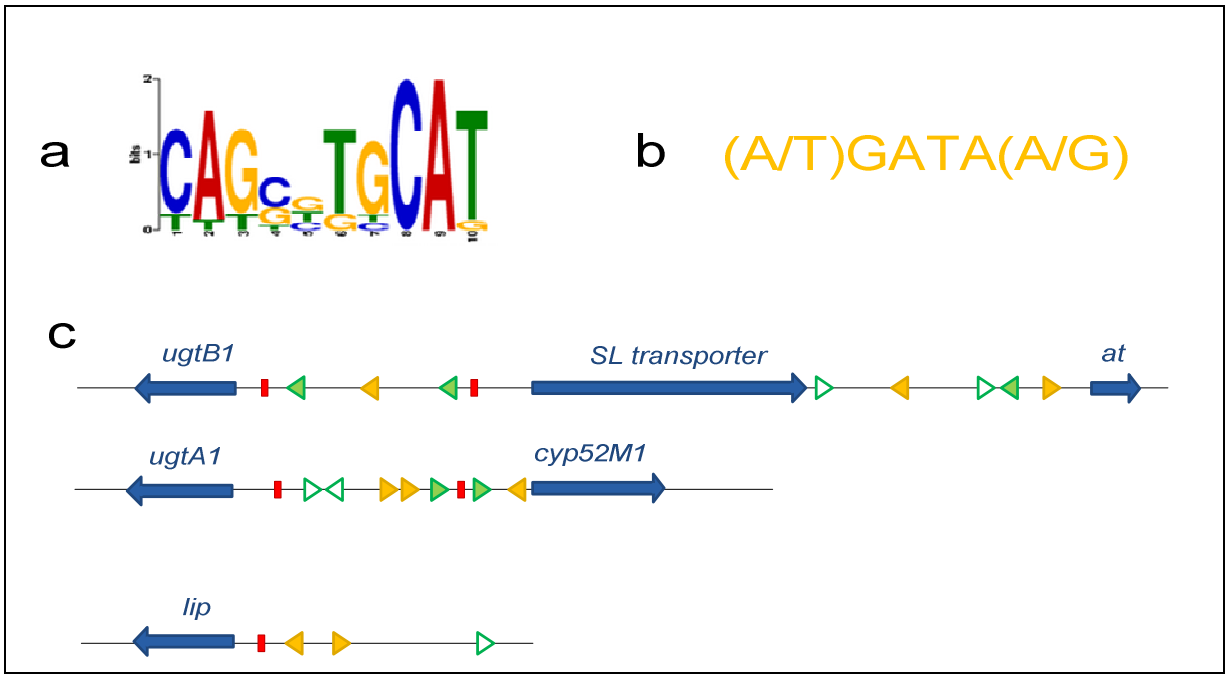


Figure II. 3 **(a)** consensus motif found in the promotor regions of the SL biosynthetic pathway using the MEME algorithm (green triangles in (c)) **(b)** GATA-consensus motif (orange triangles in (c)) **(c)** localization the two motifs a in the promotor regions: filled green triangles represent the motif depicted in (a) for which the first and last five basepairs are strictly conserved (CAG and TGCAT respectively). The open symbols represent motifs where one mismatch is allowed. The red boxes represent putative TATA-boxes.

When scanning the *S. bombicola* genome, more than 170 putative zinc finger proteins were found. It is possible that a specific regulator of SL biosynthesis can be found amongst them. However, the possibility exist that the presumed regulator is insufficient similar to known zinc finger proteins, as was the case for the Rua1 regulator of UA biosynthesis (Teichmann *et al.*, 2010), and that it can hence not be found by homology. Further in dept bioinformatic analysis of the *S. bombicola* genome aiming for the identification of possible regulatory proteins, in combination with expression data (RNA sequencing in combination with qPCR (cfr. Chapter V)) could result in the withholding of some candidates, which can then be selected for further research.

- *Other types of regulation*

Another important aspect of this SL biosynthetic gene cluster is the fact that is located near the telomere of the second chromosome of *S. bombicola*, which is characterized by telomeric and subtelomeric repeats. A predicted reverse transcriptase, an integrase and an RNase H were also found in this region when analyzing the sequencing data. The presence of such retrovirus integrase-related and reverse transcriptase gene sequences in eukaryotes is usually indicative of mobile elements (transposons). The telomeric localization of biosynthetic gene clusters containing repeated DNA sequences and transposable elements (TEs) or transposonlike (TL) elements (Pryde *et al.*, 1997) has been linked to chromatin based regulation and the action of histone deacetylases (HDAC) (Keller and Hohn, 1997). A hypothesis that was posed, is that chromatin mediated regulation of some of these SM clusters may hinge on the presence of such TL elements, interacting somehow with global regulators such as histone modifying complexes as was demonstrated for penicillin biosynthesis (Shaaban *et al.*, 2010). These elements are also hypothesized as potential driving forces in SM gene cluster evolution (Perrin *et al.*, 2007; McDonagh *et al.*, 2008). It could thus be assumed that the abovementioned processes could also hold true for the regulation of the SL biosynthetic gene cluster, as many similarities were identified by analyzing the surroundings of the SL biosynthetic gene cluster, but this remains to be investigated.

II.3.2. *Alternative transport mechanisms*

Despite the industrial importance of *S. bombicola* and its SLs, until recently it was not known how the SLs are secreted in such high amounts into the culture medium (up to more than 400 g/L). When the gene cluster was identified, one unknown putative protein was found in the middle of it. This putative protein similar to ABC multidrug resistance (MDR) proteins was found to be the missing SL transporter as deletion thereof led to a severe drop in SL production (Van Bogaert *et al.*, 2013). However, deletion of this gene did not result in a complete abolishment of SL production. This is in contrast to the observations made by (Hewald *et al.*, 2006), who reported a complete elimination of MEL production when knocking out the Mmf1 transporter protein of the MEL biosynthetic gene cluster (cfr. Chapter I.2.2.1). These observations indicate the involvement of (an) additional secretion system(s) of SLs in *S. bombicola*. Vesicles could be involved, but the involvement of additional active and/or passive transporters is also plausible.

The complete *S. bombicola* genome was scanned and 26 other putative ATP-binding cassette (ABC) transporters were identified. However, besides the ABC superfamily, there is another family that occurs ubiquitously in all classifications of living organisms: the major facilitator superfamily (MFS) or also the uniporter-symporter-antiporter family. 76 of such putative MFS proteins were found when searching the *S. bombicola* genome of which 25 were not assigned to the secretion of a specific class of compounds. Proteins assigned to this MFS superfamily were already described to be involved in transport of glycolipid biosurfactants (Hewald *et al.*, 2006) in addition to the involvement of the ABC type of transporters for other molecules (Teichmann *et al.*, 2007). Although well over 100 families of transporters have now been recognized and classified, the ABC superfamily and MFS account for nearly half of the solute transporters encoded within the genomes of microorganisms (Pao *et al.*, 1998). One can thus imagine that alternative transporters for SLs and/or derivatives thereof could be present amongst these putative proteins of the ABC and/or MFS superfamily.

Another putative MDR transporter, very similar to the SL transporter is also present in the *S. bombicola* genome, but the corresponding protein is substantially smaller than the SL transporter and possibly corresponds to only one half of the SL transporter. It is thus possible that this gene is responsible for secretion of the minor amounts of SLs still occurring in the knock out strains. Knocking out this gene in a *Δmdr* background is one of the first actions that should be undertaken to address the residual SL secretion in *Δmdr* strains.

II.4. Conclusion

In this chapter the architecture of the sophorolipid (SL) biosynthetic gene cluster of the yeast *Starmerella bombicola* was described. SL production by this organism is of industrial importance as several companies produce these biosurfactants and/or include them in their products and large international companies have initiated R&D projects focused on SL biosynthesis. The cluster was found to contain five genes and all of them are directly involved in the core biosynthetic pathway: a cytochrome P450 monooxygenase (*cyp52M1*), two glucosyltransferases (*ugtA1* and *ugtB1*), an acetyltransferase (*at*) and a SL transporter (*mdr*), which had not been described before the cluster was identified. Knocking out the two orfs delimiting the cluster at the 5' and 3' hand side resulted in strains demonstrating identical behavior as the wild type. Recent RNA sequencing data under conditions of SL production furthermore showed absolutely no (orf2) or very low (orf1) expression for these two orfs delimiting the SL biosynthetic gene cluster, which confirms the results obtained for the

deletion strains described in this chapter. The SL biosynthetic gene cluster was thus found to be delimited by the *cyp52M1* gene at the 3' hand side and further downstream by telomeric and subtelomeric repeats. At the 5' hand side the cluster is delimited by the glucosyltransferase *ugtB1* gene, responsible for addition of the second glucose moiety of SLs. The putative alcoholdehydrogenase (orf1) that lies upstream of the *ugtB1* gene as said does not have a (clear) function in SL biosynthesis, but could be involved in SL synthesis from n-alkanes as ADH activity is required for this (Wentzel *et al.*, 2007). This remains to be determined, and can be done by cultivating the respective KO in parallel to the wild type on n-alkanes and analyzing SL biosynthesis. No transcription factor was found in the SL biosynthetic gene cluster, nor was the enzyme taking care of lactonization of acidic SLs. As deletion of the SL transporter gene present in the SL gene cluster, still resulted in some residual SL production (Van Bogaert *et al.*, 2013), which is in contrast to the results obtained for other yeasts, where deletion of the glycolipid transporter resulted in a complete abolishment of glycolipid production (Hewald *et al.*, 2006), the involvement of other transport mechanisms was hypothesized. Multiple candidates, putative ABC and/or MFS transporters, for this residual secretion were found when searching the *S. bombicola* genome. One putative transporter highly similar to the MDR SL transporter was also identified. The possible involvement of the corresponding protein in SL secretion can be investigated by knocking out this gene in the Δ*mdr* strain described in (Van Bogaert *et al.*, 2013).

Putative GATA-regulatory sequences and several copies of a putative regulatory motif: CAG(C/G)(G/T/C)TGCAT were found in the promotor regions of the genes of the cluster using an *in silico* approach. Relevance of such putative cis regulatory sequences has to be investigated by performing biological experiments. However, as SL production is certainly regulated by pH, temperature, carbon source, C/N ratio, nitrogen source and/or depletion and possibly also by (positive and/or negative) feedback regulation exerted by the end- and/or intermediate products (suboptimal SL intermediate production by Δ*at* and Δ*ugtB1* mutants (Saerens *et al.*, 2011b; Saerens *et al.*, 2011c)), these cues must somehow be integrated at the promotor regions of the genes of the biosynthetic pathway. Detection of these stimuli may be shunted through regulatory cascades resulting in the transcriptional activation of genes required for SL biosynthesis by a specific regulator. Alternatively (or additionally) such regulation could occur through an effect called the telomere position effect (TPE), which involves regulation of subtelomeric SM biosynthetic gene clusters by chromatin based regulation and the action of histone deacetylases (HDAC; reviewed in (Yin and Keller, 2011)).

In filamentous fungi, a key global regulator (LaeA) (cfr. Chapter I.5) was found to regulate the expression of a variety of fungal natural product gene clusters including siderophore cluster expression (Bok and Keller, 2004). A siderophore gene cluster was also found to be induced under conditions leading to SL biosynthesis, which might be indicative for the existence of such global regulator for *S. bombicola*, but this remains to be further investigated.

The current understanding of the regulation of SM gene clusters encoding biosynthesis of glycolipid biosurfactants is a lot less advanced than that for other SMs produced by fungi. However, the recent discovery of several biosynthetic gene clusters involved in biosurfactant production by yeasts, and especially the industrial important yeast *S. bombicola*, might stimultate the further investigation of the regulatory effects at work. An in dept bioinformatic analysis of the genome searching for possible regulatory proteins, in combination with expression data could finally result in the withholding of some candidates, which could subsequently be knocked out for further investigation. Elucidation of the regulation of the SL biosynthetic pathway will not only result in a better comprehension of SL biosynthesis by *S. bombicola*, but will also create opportunities for the rational engineering of this interesting biosynthetic pathway. Overexpression of the abovementioned global regulator LaeA e.g. resulted in enhanced production of several SMs in *Aspergillus* (Bok and Keller, 2004). A similar effect was obtained when the Rua1 regulator of cellobioselipid biosynthesis in *U. maydis* was overexpressed (Teichmann *et al.*, 2010). These last examples thus demonstrate that such research is not only of fundamental nature, but will undeniably have consequences for the (industrial) production of such molecules, which is indispensable for the further development of the biobased economy.

The discovery of this SL biosynthetic gene cluster thus represents important opportunities for the genetic engineering of *S. bombicola* for the production of (new to nature) tailor-made biomolecules. This will especially be further elaborated in Chapters VI and VII of this manuscript.

CHAPTER III:

CATABOLISM OF SOPHOROLIPIDS BY *STARMERELLA BOMBICOLA*

Parts of this chapter are in preparation for publication:

Roelants, S. and Groeneboer, S., Ciesielska, K., Verweire, S., Van Bogaert, I., De Maeseneire, S. L., Devreese, B. and Soetaert, W. Catabolism of the biosurfactant sophorolipids by the yeast *Starmerella bombicola*.

Ciesielska, K. and Roelants**, S.**, Verweire, S., Van Bogaert, I., De Maeseneire, S. L., Devreese, B. and Soetaert, W. Combined techniques for the determination of protagonists involved in sophorolipid catabolism by its producer *Starmerella bombicola*. (some experimental work still in progress)

The proteomic results presented in this chapter were generated by PhDs Katarzyna Ciesielska at L-Probe.

Chapter III.

CATABOLISM OF SOPHOROLIPIDS BY *STARMERELLA BOMBICOLA*

III.1. Introduction

Glycolipids produced by microorganisms include a wide range of structures and hence properties. Different natural roles have been attributed to these natural surfactant molecules. Several biosurfactants have anti-bacterial and/or antifungal properties (Kulakovskaya *et al.*, 2004; Gross and Shah, 2005; Golubev *et al.*, 2008; Mimee *et al.*, 2009b) or can enhance growth of the producing microorganism on hydrophobic water-insoluble compounds. They are also involved in bacterial pathogenesis, quorum sensing and biofilm formation (Ron and Rosenberg, 2001). A possible natural role that has long been underexposed is the possibility that these molecules could also serve as a carbon sink produced by the microorganism when an excess of carbon is available. When the latter becomes limiting, the converted carbon is 'protected' against consumption by other competing microorganisms. Moreover the produced molecules can additionally even impose inhibitory effects against such competitors.

Microorganisms are capable of accumulating a wide array of energy and carbon storage compounds when other nutrient supplies are limiting. These intra- or extracellular carbon storage compounds can be mobilized again under carbon starvation conditions temporarily occurring in the environment. Well known examples include polyhydroxyalkanoates (PHA) (Chen, 2010) produced by a group of taxonomically different microorganisms and storage carbohydrates like glycogen (Wilson *et al.*, 2010), curdlan (Buller and Voepel, 1990) and trehalose (Arguelles, 2000). The ecological significance and regulation of these storage compounds has been thoroughly investigated. In addition these molecules often play a significant role in microbial fitness. Accumulation of PHA for example is more than just an intracellular carbon storage strategy, but additionally enhances the survival of several bacteria under environmental stress conditions (Zhao *et al.*, 2007). Furthermore, yeast cells that can accumulate glycogen have a growth advantage over cells that cannot, suggesting that glycogen makes a contribution toward overall fitness (Anderson and Tatchell, 2001).

On the contrary, biosynthesis of glycolipid biosurfactants has been linked to diverse range of possible natural roles, but conversely here, the possible function of these molecules as a kind of storage compound has long been underexposed. Yet already in the 70's the group of Light suggested the hydrolysis of sophorolipid (SL) molecules produced by *Rhodutorola bogoriensis* by enzymes found in its own cell extracts (Esders and Light, 1972a; Bucholtz and Light, 1976). They suggested the presence of a catabolic acetylesterase and possibly glycosyl hydrolases and succeeded in the isolation, purification and partial characterization of one enzyme having this acetylhydrolase activity (Bucholtz and Light, 1976). More recently metabolisation of the glycolipid flocculosin (cellobioselipid) by its producer *Pseudozyma flocculosa* has been described by Mimee et al (Mimee *et al.*, 2009a).

Production and secretion of SLs has mostly been described in relation with growth of the producing microorganisms on hydrophobic carbon sources and was considered as a way to aid the producing organisms assimilating these hydrophobic substrates like oils and n-alkanes (Ito *et al.*, 1980; Ito and Inoue, 1982; Hommel, 1990). However Hommel *et al.* later stated that biosynthesis of SLs by *Candida* yeasts can't simply be a prerequisite for the degradation of extracellular hydrocarbon, as SL production from glucose (and other sugars) alone was reported for *Candida apicola* and *Starmerella bombicola*, independent of the fact that hydrocarbons were present or not (Hommel and Huse, 1993). Additional indications can be found in the fact that *Pseudozyma antarctica* accumulates erythritol lipids up to 12 % of the dry weight when growing on glucose. The authors suggested that these strains accumulated the MEL intracellularly as storage material together with triglycerides (Kitamoto *et al.*, 1992b). In yet another publication Hommel *et al.* suggested SLs produced by *C. apicola* to act as extracellular carbon storage material (Hommel *et al.*, 1994b). More recently some authors have again started to suggest that glycolipid biosurfactants could be produced as storage compounds when one nutrient becomes limiting in combination with remaining high carbon concentrations. This would indicate that *S. bombicola* can use its own SLs as sole carbon source as was already suggested by (Garcia-Ochoa, 1996). Whereas it was later reported that SLs cannot be degraded by *S. bombicola* (Hu, 2000; Lo and Ju, 2009), we had some very clear indications of the contrary. Therefore, a series of experiments was set up to investigate the hypothesis that one of the roles of SL production by *Starmerella bombicola* is the creation of an extracellular 'secured' carbon sink to feed on when cells are starved for carbon and attempts were done to identify the responsible mechanisms/enzymes.

III.2. Materials and Methods

III.2.1. *Strains, media and culture conditions*

Starmerella bombicola ATCC 22214 was used as the parental or wild type strain. Additionally a *lipase* (cfr. Chapter VI) negative mutant (*Δlip*) knocked out in the gene responsible for lactonisation (Ciesielska *et al.*, 2013) was used for growth experiments and assays. Yeast cells were grown on YPD medium (1 % yeast extract, 2 % peptone and 2 % glucose) (cfr. III.2.7) and 3C medium (10 % glucose, 1 % yeast extract, 0.1 % ureum and 15 % agar) was used to maintain the strains. For degradation experiments, the production medium described by Lang (Lang *et al.*, 2000) was used to which rapeseed oil (Sigma; 3.75 %) was added after 48h when mentioned. Precultures were prepared from single colonies derived from 3C plates in 5 mL of the production medium in round bottomed culture tubes. These overnight grown precultures were inoculated (2 %) in shake flasks of 1 L containing 200 mL of the production medium. The shake flask cultures were sampled regularly for SL composition, pH determination, glucose concentration and supernatant collection for ammonium determination. Degradation intermediates started to accumulate in the supernatant of such cultures after 21 to 30 days of incubation. When the latter occurred, one set of cultures was left on the shaker for further follow up and a second set was harvested. These experiments were repeated to confirm the obtained results. The latter was done by centrifugation (10 min, 4000 x *g*, 4 °C). Supernatant and pellet were separated and further processed for performing assays (secretome and lysate) or incubation experiments.

Alternatively the yeast was grown on a modified SD medium (SD = 0.67 % yeast nitrogen base without amino acids (DIFCO) and 2 % glucose) or a modified production medium to which di-acetylated lactonic or non-acetylated acidic SLs were added as the sole C-source (cfr. Table III. 1). The SLs were added to the prepared media and the mixtures were subsequently filter sterilized. These cultures were sampled likewise as the ones described above. Two methods were used to start up such cultures. The first one consisted of 3 actions. First an inoculum of an ‘old’ culture on production medium (cultivated for > 20 days) was plated on modified SD agar plates. When growth was visible, a liquid preculture on modified SD medium was prepared (5 mL) from these plates. This preculture was subsequently used as inoculum to start up a shake flask experiment on modified SD medium (100 mL or 200 mL). Secretome and cellpellets of such cultures were harvested to perform proteomic analyses (cfr. III.3.5).

Table III. 1 Compostion of the used media

Compound [g/L]	Production medium	Modified Production medium	SD medium	Modified SD medium
Glucose	120	120	20	/
YNB w/o AA	/	4	6.7	6.7
Yeast extract	4	/	/	/
3Na-citraat $2H_2O$	5	/	/	/
NH_4Cl	1.5	1.5	/	/
KH_2PO_4	1	1	/	/
K_2HPO_4	0.16	0.16	/	/
$MgSO_4 . 7H_2O$	0.7	0.7	/	/
NaCl	0.5	0.5	/	/
$CaCl_2 . 2H_2O$	0.27	0.27	/	/
Sophorolipids	/	20		20

Due to growth difficulties on the plates an alternative method to start up liquid cultures was also used for the cultures on modified production medium: 5 mL of an old culture on production medium was spun down and the harvested cells were redissolved in 5 mL of modified production medium. This cell solution was incubated at 30 °C, 200 rpm for at least 24 h or until cells were clearly in exponential phase (budding) and subsequently used as an inoculum to prepare a second preculture on modified production medium (5 mL), which was in turn used as inoculum for a shake flask experiment (100 mL or 200 mL).

Escherichia coli DH5α, XL10GOLD and Fusion Blue Ultracompetent cells were used for cloning experiments. *E. coli* cells were grown in Luria- Bertani (LB) medium (1 % trypton, 0.5 % yeast extract, 0.5 % sodium chloride (+15 % agar for plates) supplemented with 100 mg/L ampicillin. *E. coli* cultures were incubated at 37 °C on a rotary shaker (200 rpm).

III.2.2. *Sampling and analysis*

III.2.2.1. *Follow up of growth*

Optical density (OD) and glucose concentration were monitored as described in Chapter II. 2.3.1. The viability of yeast cells in prolonged cultivation experiments was assessed by determining colony forming units (CFU): decimal dilutions in physiological solution were plated on 3C agar plates and incubated at 30 °C for three days.

III.2.2.2. *Analytical extraction and analysis of samples*

Preparation and HPLC analysis of analytical SL samples was done as described in Chapter II. 2.3.2. The same column and LC conditions were used for performing LC-MS analyses of said samples using an 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.

A second LC-MS method was used for sophorose detection. The used column was a Hypercarb PGC 100 x 4.6 mm column from Thermo. Three eluents were used; 4 % methanol (A), 100 % acetonitrile (B) and 15 % formic acid (C) and run time of one sample was 40 minutes (min). Analysis started at 100 % of eluent A for 4 min, after which a linear decrease of A to 78 % and respective increase of B to 22 % was executed in 5 minutes. This was followed by a further decrease of A to 41 % and respective increase of C to 37 % in the next 8 min after which eluent A decreased further to 3 % in the next 10 min, whereas C increased to 75 % and B was kept at 22 %. After this A was increased again to 50 % in 2 minutes as was B and C dropped back to 0 %. The next minute served to raise A back to 100 %, which was kept like that for the last 10 minutes of the run. A flow rate of 0.6 mL/min was applied. A sophorose standard of 250 mM was used.

III.2.2.3. *Ammonium determination*

The amount of free ammonium (NH_4^+-ions) in the culture medium was determined using the Ammonia Rapid Megazyme Kit. A standard solution was prepared using NH_4Cl.

III.2.2.4. *Effect of pH on sophorolipid solubility and composition*

Solubility of highly purified lactonic and acidic SLs was determined in buffered (KH_2PO_4/K_2HPO_4) aqueous solutions set at pH values of 3; 5; 7 and 8.5 with HCl and NaOH respectively. Minor amounts of the respective SLs were added until the products could not be solubilised anymore and precipitation occurred. When this occurred the solutions were placed at 30 °C and more SLs were added to saturation if this slight heating again resulted in disappearance of the precipitate. This was performed on an analytical balance in order to determine the amount of product that could be solubilised at a certain pH.

The possible effects of pH variations on SL composition during prolonged cultivations were investigated by incubation of di-acetylated lactonic SLs in buffered solutions (pH 3, 5, 7 and 8.5) at 30 °C, 200 rpm for one week. These solutions were sampled for pH determination and HPLC/LC-MS analysis as described in III.2.2.2.

III.2.3. *Preparation of purified sophorolipids*

A large amount of di-acetylated lactonic SLs was obtained from a bioreactor experiment performed in the past (cfr. chapter VI for details of such experiments) of which the harvested (precipitated) di-actylated lactonic SLs were crystallized by overnight incubation at 4 °C in water (150 rpm). The resulting solution containing the SL crystals was placed at -80 °C and lyophilized overnight after which a white powder consisting of very pure lactonic SLs was obtained. Part of these highly purified di-acetylated lactonic SLs were converted to acidic non-acetylated SLs as described by (Saerens *et al.*, 2009) and both compounds were used as such in the assays.

III.2.4. *Incubation of the cell free culture medium (unconcentrated secretome) with sophorolipids*

The collected supernatant of a set (with and without addition of rapeseed oil) of the abovementioned old cultures, which had been incubating for 21 days on production medium, was centrifuged twice to remove all insoluble particles. The resulting cleared supernatants derived from the wild type or *Δlip* mutant (both 200 mL), were each divided in three volumes. The first volume was supplemented with di-acetylated lactonic SLs (20 g/L), the second volume with non-acetylated acidic SLs (20 g/L) and the last one was supplemented with an equal amount of sterile dH_20. Highly pure lactonic and acidic SLs were obtained as described in III.2.3. The resulting solutions were filter sterilized using a 0.22 µm filter and incubated (30 °C) in sterile shake flasks on a rotary shaker (200 rpm). Two flasks with production medium w/o glucose and set at the pH of the collected secretomes were supplemented with both forms of SLs respectively and served as controls. The flasks were incubated for one month and sampled regularly for SL analysis, glucose and pH determination.

III.2.5. *Assays*

III.2.5.1. *Concentrated secretome preparation*

The collected supernatants (200 mL) were centrifuged twice to remove all insoluble particles (15 minutes, 4000 x *g*, 4 °C). Three tablets of protease inhibitor cocktail (Roche) were added and the secretomes were subsequently filtered using a 0.22 µm filter to completely eliminate remaining yeast cells. The proteins were concentrated using a stirring ultra filtration cell (Model 8200 Milipore) containing a 10 kDa cut off G membrane (Sartorius Stedim). After

concentrating to 30 mL, one startvolume of wash buffer (20 mM Tris, pH 7) was added. The stirring cell was put under pressure again to reconcentrate the secretome to 30 mL. The concentration of the protein solution was determined using a BCA Protein Assay Kit (Pierce). In case the concentration was lower than 1 mg/mL, the proteins were further concentrated using a Vivaspin 15R (Sartorius stedim). The obtained solutions were used to perform the assays.

III.2.5.2. *Lysate preparation*

Cell pellets of the described experiments were collected as described above (10 minutes, 4000 x *g*, 4°C). The pellets were dissolved in lysis buffer (20 mM Tris, pH 7; 0.5 mM $MgCl_2$; 5 % glycerol and 1 tablet of protease inhibitor cocktail (Roche) per 50 mL of buffer). This cell solution was added to a 2 mL tube of lysing matrix C (MP biomedicals) containing 1 mm silica spheres and subjected to 2 cycles of 6 m/s in a FastPrep Celldisrupter (MP biomedicals). Soluble fractions were used for the assays after centrifugation of the crude lysates (10 minutes, 4000 x *g*, 4°C).

III.2.5.3. *Assays*

The lysate or secretome preparations were incubated (30 °C, on a rotary wheel at 30 rpm) in lysis buffer or wash buffer (containing 0.5 mM $MgCl_2$) respectively, supplemented with 2 mM of di-acetylated lactonic SLs or non-acetylated acidic SLs. Unless stated otherwise the end volume of each reaction was 1 mL containing 1 mg of protein. Incubation times are mentioned in the results section. These assays were sampled to determine the possible release of glucose using the 2700 select biochemistry analyser (YSI Inc.), to monitor the pH and for HPLC and LC-MS analysis of the samples as described in III.2.2.2. Hydroxylated lauric acid (50 nmol) was added as an internal standard before extraction. For each assay a first control consisted of 2 mM of SLs solved in the respective buffer to control for potential spontaneous breakdown of SLs during incubation. A second one consisted of the lysate or secretome preparations supplemented with an equal amount of dH_20 (instead of SLs) to correct for background of remaining SLs and/or other compounds in the protein solutions.

III.2.6. *Proteomics*

Proteomic analysis of several experiments was performed at the lab of Professor Bart Devreese (L-Probe) in cooperation with PhDs Katarzyna Ciesielska and PhD Sara Groeneboer.

Several secretomes were prepared for proteomic analysis. Shortly said they were filtered using a 250 mL Filtropur V25 vacuum filter with a pore diameter of 0.2 μm (Sarstedt, Nümbrecht, Germany) and concentrated using an Amicon stirred cell (Millipore, Billerica, MU, USA) equipped with a Sartorius membrane with a 10 kDa cut-off (Sartorius, Goettingen, Germany) until a volume of 30 mL was reached. A second concentration step was performed using Vivaspin 15R columns with a 10 kDA cut-off (Sartorius, Goettingen, Germany). The proteins were subsequently separated on a 12.5 % SDS-PAGE gel ran at 110 V. Bands were subsequently cut from the gels, subjected to trypsin digestion and the peptide mixtures were analysed on nanoLC-LTQ-FT ICR.

The results were analysed using the Mascot Daemon software version 2.2.2 (Ions score cut-off 30, Significance threshold 0.05) against the *S. bombicola* database, which was received from Professor Yves Van de Peer (Bioinformatics & Systems Biology, VIB) and Phds Bing Li. Proteins with a minimum of two unique detected peptides were claimed as identified. Results were further investigated by performing BLAST searches on the NCBI website.

III.2.7. *Creation of the hsbA knock out strain*

The generally used molecular techniques are described in Chapter II. 2.2.1. To create a knock out strain of the *hsbA* gene, a similar strategy as the one described in Chapter II to knock out orf1 and orf2 was employed and the used primers can be found in Table III.2. A plasmid containing a hygromycine resistance gene preceded by the strong, constitutive *gapd* promoter of *S. bombicola* (Van Bogaert *et al.*, 2008a) was digested using unique restriction enzymes *Bse*RI and *Eco*NI. This fragment was purified and ligated with the digested and purified upstream region of the *hsbA* gene (up*hsbA*). The latter was picked up from genomic DNA of *S. bombicola* using primerpair P95_FOR_up*hsbA*_ext*Bse*RI / P96_REV_up*hsbA*_ext*Eco*NI and digested with the same restriction enzymes. The resulting vector was subsequently digested using restriction enzymes *Spe*I and *Not*I and subsequently ligated with the purified and digested downstream region of the *hsbA* gene (d*hsbA*). This fragment was also picked up from genomic DNA using primers P97_FOR_d*hsbA*_ext*Spe*I and P98_REV_d*hsbA*_ext*Not*I. The resulting vector (pGEM-T_up*hsbA*_P*gapdhygro*_d*hsbA*) was sequenced and subsequently used to create the knock out cassette with primers P99_FOR_KOcassette*hsbA* and P100_REV_KOcassette*hsbA* which was subsequently transformed into the *S. bombicola* wild type strain. Transformants were selected on YPD plates supplemented with hygromycine

(500 mg/mL). Colony PCR on transformants was performed using primerpairs P102_FOR_upP*hsba* / P101_REV_*hygro*Insert (1819 bp) and positive transformants were used to perform cultivation experiments on the production medium.

Table III. 2 Primers used to create the *hsbA* knock out strain

Primer names	**'5-Sequence-3'**
P95_FOR_pHsbA_extBseRI	GGGAGGAGAAGTCCTGCATCCGAGCTAGCCAAGCATAAGAAG
P96_REV_upHsbA_ext*Eco*NI	GGCCTGGGGAAGGCATTTTAAGAGAGGGTATGAAGATCTG
P97_FOR_dHsbA_ext*Spe*I	GGACTAGTTCGCCATGCTCATCTAAG
P98_REV_dHsbA_ext*Not*I	GGGCGGCCGCGTTGATTGATTGTTGAAGAAATTC
P99_FOR_KOcassetteHsbA	CCGAGCTAGCCAAGCATAAG
P100_REV_KOcassetteHsbA	GTTGATTGATTGTTGAAGAAATTC
P101_REV_HygroInsert	ACAGACGTCGCGGTGAGTTC
P102_FOR_UPpHsbA	GCGATGTCAGCATGACTACC

III.3. Results and Discussion

III.3.1. *The effect of pH on SL solubility and disassembly*

The effect of the pH on SL solubility was assessed because the pH changes dramatically during cultivation of *Starmerella bombicola* on production medium (cfr. Figure III. 2). The solubilities (g/L) of di-acetylated lactonic and non-acetylated acidic SLs in an aqueous solution buffered at different pH values are represented in Table III. 3.

Table III. 3 Influence of pH on solubility (g/L) of different forms of SLs

pH	3	5	7	8.5
Acidic non-acetylated	75	93	187	185
Lactonic di-acetylated	25	42	69	136

The solubility of the SLs diminishes dramatically as pH drops. Translated to culture conditions this means that a pH drop will result in precipitation of the (predominantly produced) di-acetylated lactonic SLs. This event can be considered as a removal of the end-product from the culture medium, which could be one of the reasons for the very high titers (up to more than 400 g/L) of the produced SLs as there will be no or limited end-product inhibition. When the pH of the culture medium starts rising again, these molecules will be solubilised and the precipitation will hence disappear, which fits with our observations (cfr. III.3.2). The structures of di-acetylated acidic and di-acetylated lactonic SLs (both C18:1, ω-1) are depicted in Figure III. 1 a and b respectively.

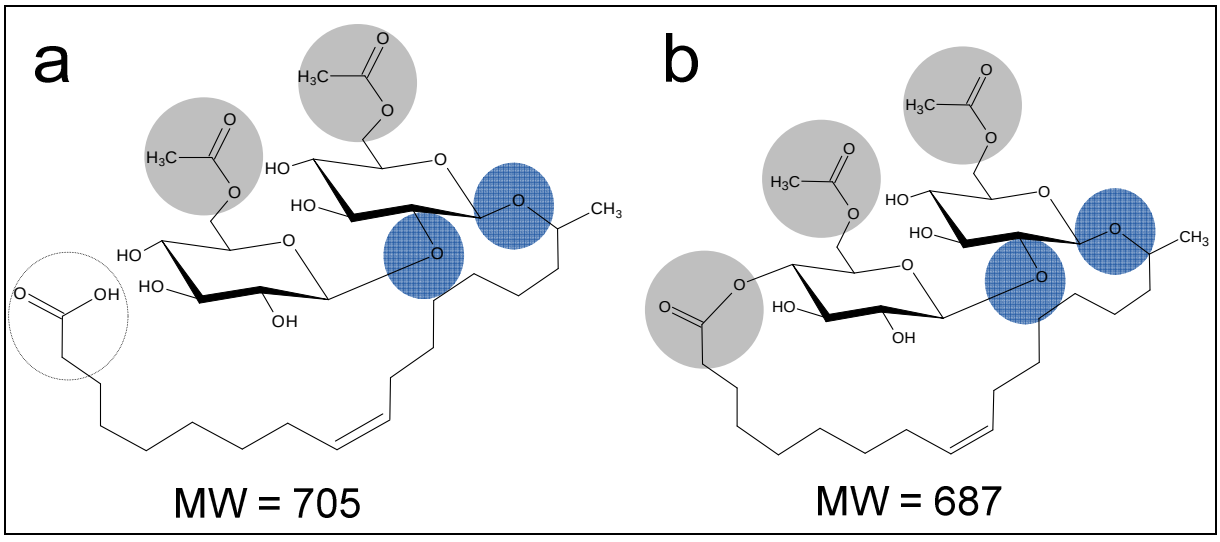


Figure III. 1 Di-acetylated **(a)** acidic and **(b)** lactonic SLs and their respective molecular weights (MW). The grey areas represent ester bonds, the blue areas the glycosidic bonds, characterizing these molecules.

As spontaneous hydrolysis of the ester bonds (lacton- and acetyl-) typical for these SLs was reported to occur at alkaline conditions (pH 9-10) at room temperature, eventually leading to non-acetylated acidic SLs (Inoue *et al.*, 1980), the di-acetylated lactonic SLs were incubated at several buffered pH values for one week (30 °C). The SL composition was subsequently analyzed on HPLC and LC-MS. Whereas no effects were observed for pH values of 7 and 5, some effects were observed for the incubation at pH 8.5 and pH 3. The high pH indeed resulted in the hydrolysis of the esterbonds and hence the appearance of di- and mono-acetylated acidic SLs and mono- and non-acetylated lactonic derivatives for the incubation at pH 8.5. However, such high values are never attained in cultivation experiments (cfr. Figure III. 2). Incubation at a pH value of 3 on the other hand resulted in a minor hydrolysis of the glycosidic bonds.

III.3.2. *Catabolism of sophorolipids*

III.3.2.1. *Growth on production medium*

When SL production is desired, *S. bombicola* is cultivated on the production medium described by (Lang *et al.*, 2000) (cfr. Table III. 1) and rapeseed oil is added after 48 hours of cultivation. Shake flask experiments are stopped when glucose is depleted, which occurs after 8 to 10 days of cultivation (Figure III. 2 a), and SLs are harvested. However, to investigate catabolism of the produced SLs, the cultures were left on the shaker beyond glucose exhaustion and sampling continued (Figure III. 2 b).

Growth is as always accompanied with a strong decrease of the pH resulting from secretion of organic acids such as citrate and isocitrate (up to 10 g/L) and utilization of the nitrogen source (Stüwer *et al.*, 1987). A steep pH and ammonium rise (cfr. Figure III. 2) are observed after glucose is depleted from the medium. The pH rise can (partly) be explained by the rise in free ammonium ions, which is the result of deamination of amino acids once glucose is depleted and the yeast shifts from carbohydrates to proteins as the main energy source. The produced citrate and isocitrate can in addition also be used as a carbon source as the concentrations of these acids were reported to decrease when the glucose in the culture medium was nearly consumed (Albrecht *et al.*, 1996). Concomitant with these changes the composition of extracted SLs also changes dramatically. Typically the dominant form of SLs produced when culturing *S. bombicola* on the production medium are the di-acetylated lactonic SLs (C18:1 and C18:0). These compounds start precipitating (cfr. Figure III. 3 a) during cultivation and production because their solubility dramatically decreases when pH drops (cfr. Table III. 3).

When no oil is added to the cultures, this precipitation occurs after 130 hours (± 5.5 days) of cultivation and the amount increases as cultivation proceeds.

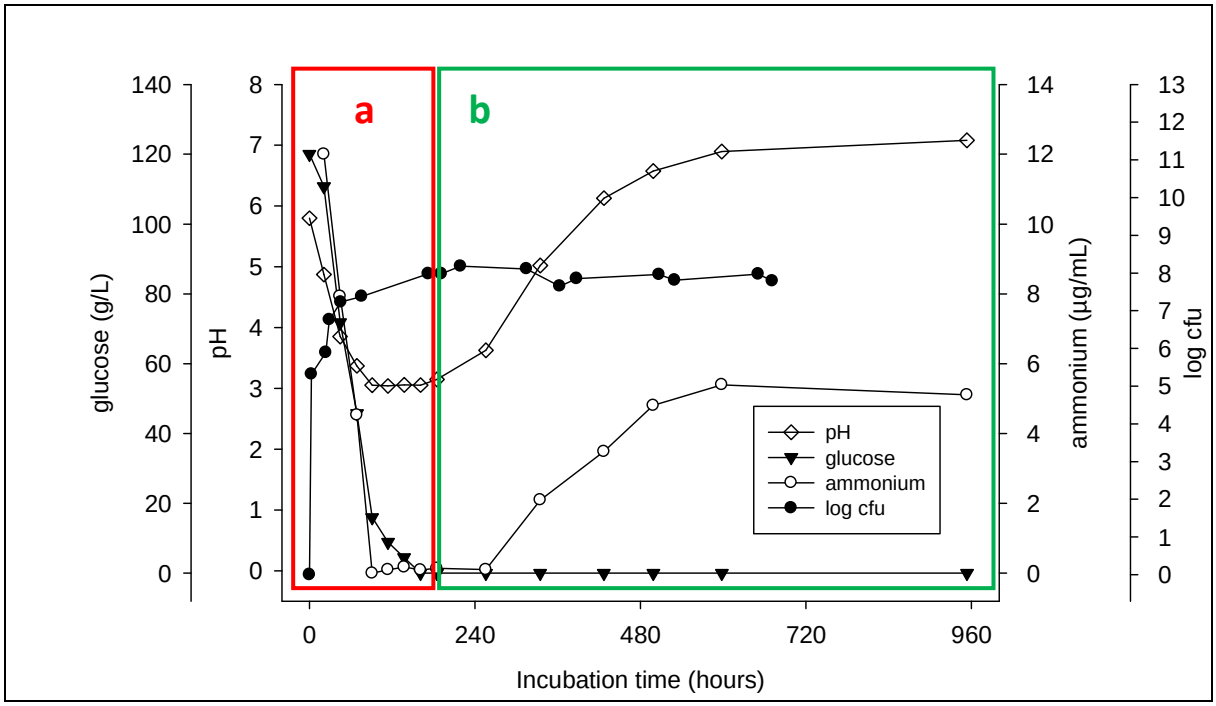


Figure III. 2 **(a)** SL biosynthesis and **(b)** SL catabolism phase of *Starmerella bombicola* cultivated on production medium. (◊) pH (▼) glucose concentrations (○) ammonium concentrations and (●) log cfu/mL values are depicted in function of time.

However, when these cultures are left on the shaker after glucose has been depleted, the precipitated SLs (di-acetylated lactonic form) start disappearing from the bottom of the flask again after 330 hours (± 14 days) of incubation to eventually completely vanish (cfr. Figure III. 3 b).

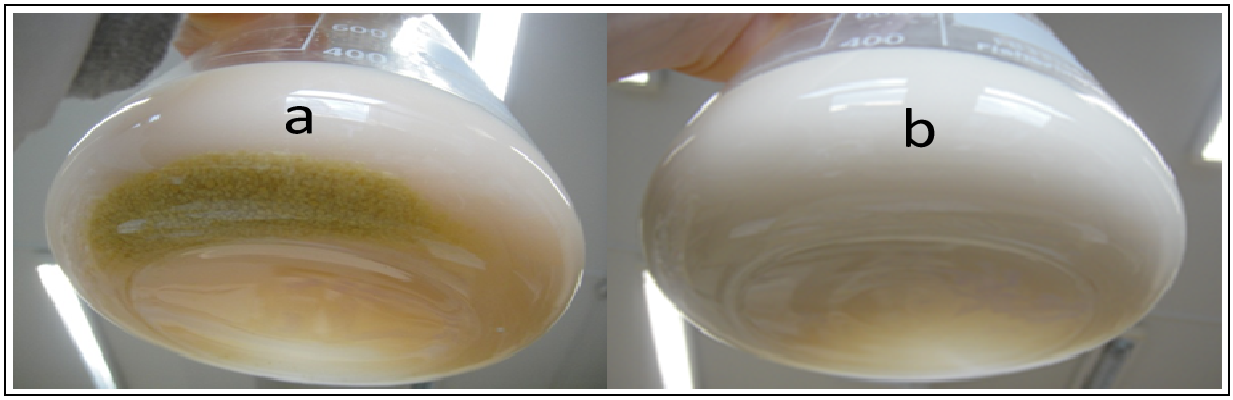


Figure III. 3 Shake flask culture of *Starmerella bombicola* on production medium (without addition of oil) after **(a)** 240 hours (10 days) of cultivation and **(b)** 960 hours (40 days) of cultivation.

This observation can be explained by the rising pH and subsequent better solubilisation of the produced SLs. There is more to this though, as analysis of the chromatograms of SL samples of such prolonged cultivations reveals that the composition of the SL mixture changes in function of time which can be concluded from the chromatograms depicted in Figure III. 4.

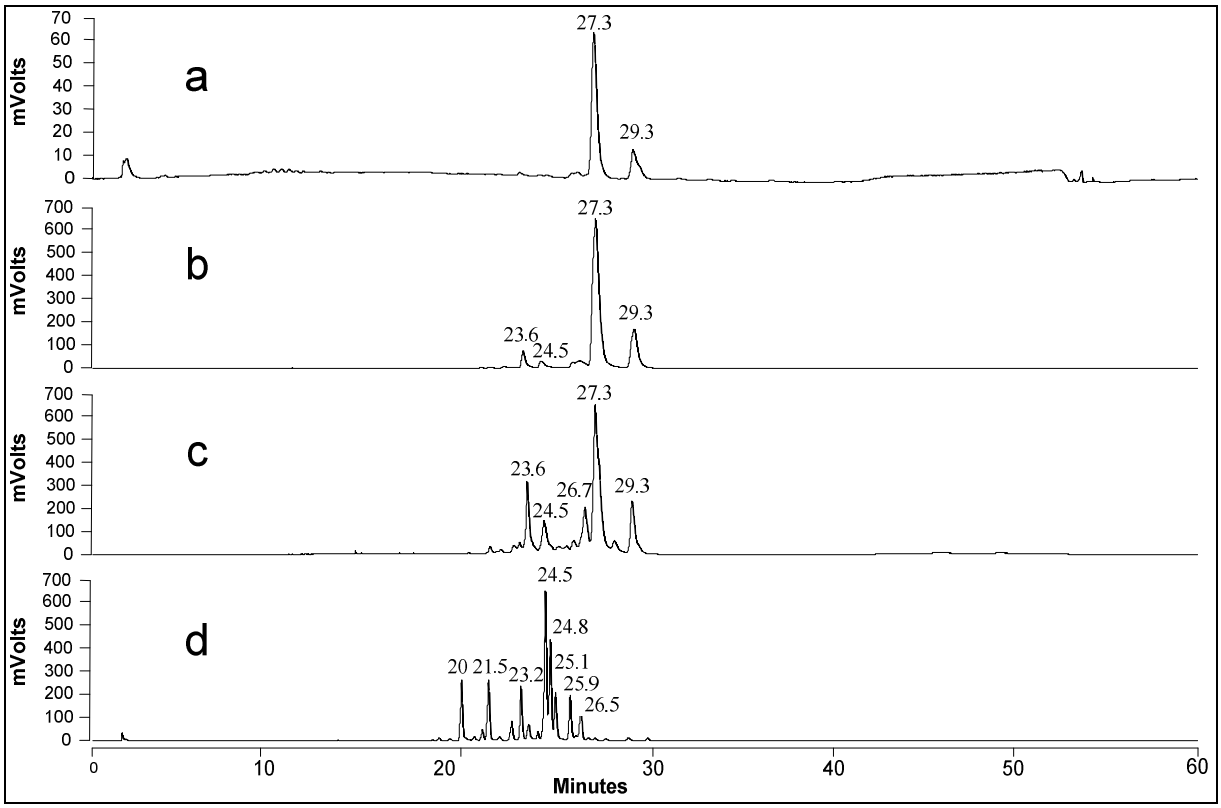


Figure III. 4 HPLC-ELSD chromatograms of extractions from a *S. bombicola* culture on production medium (without addition of oil) taken after **(a)** 8 days **(b)** 15 days **(c)** 40 days **(d)** 100 days of incubation. Peaks correspond to di-acetylated lactonic SLs C18:0 (29 min) and C18:1 (27 min), mono-acetylated lactonic SLs (23 and 24 min), di-acetylated acidic SLs (23 and 24 min), mono-acetylated acidic SLs C18:1 (21 min), non-acetylated acids SLs C18:1 (20 min) and hydroxylated fatty acids (25 and 26 min).

The main SL variants produced during the first fase of cultivation are the di-acetylated lactonic SLs eluting at 27.3 min (C18:1) and 29.3 min (C18:0). These compounds gain in concentration during this first phase of cultivation (cfr. Figure III. 4 a and b). However, after 40 days of cultivation (cfr. Figure III. 4 c), the SL mixture becomes enriched in more hydrophilic compounds eluting at 23 and 24 minutes corresponding to the di-acetylated acidic SLs and mono-acetylated lactonic SLs. After 3 months of cultivation (cfr. Figure III. 4 d) the peaks eluting at 27 and 29 minutes, corresponding to the di-acetylated lactonic SLs, have completely disappeared and a complex mixture of compounds is present in the extracellular

medium. When analyzing such samples on LC-MS, it turns out that this mixture is composed of mono- and di-acetylated lactonic SLs (still present in minor amounts); di-, mono- and non-acetylated acidic SLs; glucolipids and hydroxylated fatty acids.

These observations could thus indicate extracellular dissimilation of the produced SLs by *S. bombicola* in times of carbon starvation, which is also confirmed by rather constant log cfu values (cfr. Figure III. 2) and the detection of possible catabolic intermediates in the medium. Alternatively, one could argue these to be intermediates of the SL biosynthetic pathway leaking out of dead or damaged cells. However, this explanation does not account for the complete disappearance of the di-acetylated lactonic SLs (cfr. Figure III. 4 d). An explanation at least for the hydrolysis of the lacton- and acetylbonds has to be given. A possible explanation for the latter is that these molecules are taken up again by the cells and therefore disappear. This would then again be indicative of some kind of recycling of the produced molecules by its producer and point in the direction of SL dissimilation by *S. bombicola*. Although some spontaneous effects might also be argued (cfr. III.3.1), values as high as 8.5 are not obtained during shake flask experiments and incubation of SLs at pH values of 7 did not result in such spontaneous effects. The remainder of the observations is thus suggested to be attributed to the action of catabolic enzymes *i.e.* deacetylating and/or delactonizing enzyme(s) and glycosyl hydrolases responsible for hydrolyzation of the O-glycosidic bonds. The fact that a catabolic acetylesterase acting on the acetylgroups of (acidic) SLs produced by *Rhodotorula bogoriensis* was identified in cell lysates of this yeast (Bucholtz and Light, 1976) strengthens this assumption. In vitro enzymatic assays with acetylesterase, cutinase and lipase enzymes have furthermore been reported to result in deacetylation of di-acetylated lactonic SLs while keeping the lactone ring intact (Asmer *et al.*, 1988; Carr and Bisht, 2003). In contrast to the ester bonds, the glycosidic bonds (cfr. Figure III.1) connecting the sophorose residu with the fatty acid moiety and also the two glucose molecues is chemically stable and hydrolysis is not expected to occur without severe physicochemical or enzymatic action. If the abovementioned glucolipids and fatty acids are derived of the di-acetylated lactonic SLs instead of being intermediates of the biosynthetic pathway, then this can only result from the activity of one or more enzyme(s) responsible for cleaving of the glycosidic bond(s). Such enzymatic conversion of SLs to glucose lipids has also been reported using commercial glycosidases, releasing one glucose molecule from the dissacharide lipid (Rau *et al.*, 1999).

To further investigate these observations, a series of experiments was set up. First, the dissimilation of SLs as a sole carbon source by *S. bombicola* will be discussed (cfr. III. 3.2.2),

which will be followed by the results obtained for protein incubation assays with SLs (III.3.3 and III.3.4). Proteomic experiments (III.3.5) and knock out experiments (III.3.6) were finally performed in an attempt to identify possible protagonists in SL catabolism.

III.3.2.2. *Growth on sophorolipids as the sole carbon source*

If *S. bombicola* produces SLs amongst others as some kind of carbon sink, then the yeast must be able to use these molecules as the sole carbon source as was already stated by Garcia-Ochoa and Casas (Garcia-Ochoa, 1996). A modified SD and modified production medium (cfr. Table III. 1) with SLs as the sole carbon source were used for cultivation experiments. Growth on such media can only occur if *S. bombicola* is able to dissimilate its own SLs. Multipolar budding was observed after three to four days of cultivation, confirming that this is indeed the case. High cell densities weren't reached though.

The observed growth can clearly be linked to feeding of *S. bombicola* on the added SLs, which were being hydrolysed as can be concluded by comparing an early sample taken after 1 day of cultivation (cfr. Figure III. 5 a) with a sample taken after 8 days (cfr. Figure III. 5 b) and 18 days of cultivation (cfr. Figure III. 5 c). After 1 month of cultivation almost all the SLs and -derivatives have disappeared from the extracellular culture medium (cfr. Figure III. 5 d). These observations thus correspond well to the ones observed in Figure III. 4, although the time frame here is a lot smaller. An overview of the compounds corresponding to the observed peaks in Figure III. 5 is given in Table III. 4 and the observed effects are discussed below.

The first two activities that can be detected are ringopening and deacetylation (cfr. Figure III. 5 a) of the substrate resulting in di-acetylated acidic SLs and mono-acetylated lactonic SLs respectively. These compounds coelute at 24.3 and 24.5 & 23.6 minutes. Several peaks per unique MW are detected, because terminal (ω) and subterminal (ω-1) hydroxylated intermediates are both present, which correspond to the same MW. Moreover, mono-acetylated derivatives can carry the acetylgroups on the internal or external glucose moiety, which again results in two compounds with the same molecular weight eluting at different retention times. Continued cultivation (cfr. Figure III. 5 b) clearly leads to further dissimilation of SLs into mono-acetylated acidic SLs (21.0 and 21.6 min) and non-acetylated acidic SLs (19.0, 19.5 and 20.2 min). Corresponding C18:0 and C18:2 intermediates are also detected (peaks at 19.0, 20.2, 23.7 and 23.6 min). Minor amounts of non-acetylated lactonic SLs (C18:1) eluting at 22.8 minutes are now also detected. After 18 days of cultivation (cfr.

Figure III. 5 c) the non-acetylated intermediates (peaks at 19.0, 19.5 and 20.2 minutes) in the culture medium have diminished and after 30 days of cultivation almost all the SLs have disappeared (scale!) from the culture medium (cfr. Figure III. 5 d).

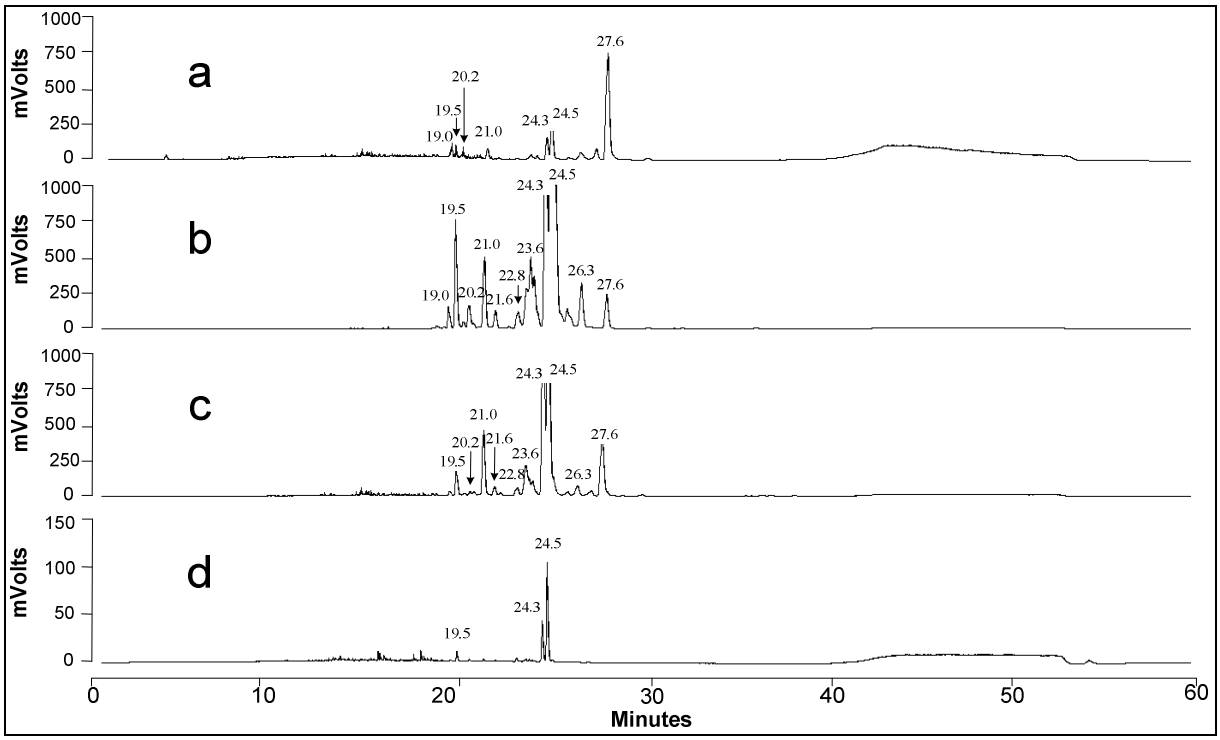


Figure III. 5 HPLC-ELSD chromatograms of samples from a *S. bombicola* culture on modified production medium with di-acetylated lactonic SLs (27.6 min) as the sole carbon source taken after **(a)** 1 day **(b)** 8 days **(c)** 18 days and **(d)** 30 days of cultivation.

During cultivation a pH drop from 6.06 to 5.60 (after 16 days) occurred, so in this case spontaneous hydrolysis of ester bonds at higher pH values cannot be argued. SLs present in the controls (cell free incubations of the culture medium with SLs) moreover remained quasi unaltered (not shown), so spontaneous effects can be ruled out anyway. Leakage of biosynthesis intermediates is also very unlikely here, as biosynthesis is not expected to occur under these conditions, at least not in such large amounts. Moreover, in that case one would expect to also detect hydroxylated fatty acids and/or glucolipids (as was the case for the experiment described in III.3.2.1 in the culture medium, which was not the case. Moreover, the cells were clearly budding, indicating growth, and after 30 days of incubation almost all the SLs had disappeared from the culture medium (cfr. Figure III. 5 d).

These results not only provide evidence for the dissimilation of SLs by *S. bombicola*, but also suggest the presence of extracellular enzyme(s) responsible for part of these catabolic

processes. At least an enzyme causing ring opening of the di-acetylated lactonic SLs and one or more catabolic acetyl hydrolases responsible for deacetylation of lactonic and acidic SLs are expected to be present in the extracellular culture medium. The existence of such enzyme was also proven for SLs produced by *R. bogoriensis* (Bucholtz and Light, 1976).

Table III. 4 Identification of the major peaks detected in Figure III. 5 after LC-MS analysis, of which the most predominant peaks are framed.

Elution time	Mass	Identity		Acetyl?
19.0	619	SL open	C18:2	non
19.5	621	SL open	C18:1	non
20.2	623	SL open	C18:0	non
21.0	663	SL open	C18:1	mono
21.6	663	SL open	C18:1	mono
22.8	603/663	SL closed/open	C18:1	non/mono
23.6	705	open	C18:1	di
23.7	643 and 703	SL closed/open	C18:2	mono/di
24.3	645 and 705	SL closed/open	C18:1	mono/di
24.5	645	SL closed	C18:1	mono
26.3	647 and 707	SL closed/open	C18:0	mono/di
27.0	687	SL closed	C18:1	di
27.6	687	SL closed	C18:1	di

As no glucolipids or hydroxylated fatty acids (nor glucose or sophorose) were detected in the culture medium, it is assumed that non-acetylated acidic SLs (eluting at 19.5 min) are taken up again by the starved cells and that further metabolisation happens intracellularly. This is strengthened by the fact that these compounds clearly disappear from the culture medium after they are being formed, thus without releasing the respective building blocks in the culture medium. However, because cells are present, one cannot exclude the possible further extracellular catabolism of the non-acetylated acidic SLs (peak at 19.5 min). In that case sugars and/or hydroxylated fatty acids would be released in the culture medium, which could immediately be taken up again by the starved cells, rendering it impossible for them to be detected in the extracellular culture medium. The presence of an extracellular glucosyl hydrolase can hence not be ruled out and this can only be investigated by executing cell-free experiments.

III.3.3. *Incubation of the cell free culture medium (unconcentrated secretome) with sophorolipids*

To further investigate the above, the cell free culture medium, containing all the extracellular proteins (secretome), of a culture of *S. bombicola* cultivated on production medium (21 days old) was separated from the cells by centrifugation and filtration. This unconcentrated secretome was subsequently incubated with di-acetylated lactonic and non-acetylated acidic SLs as described in III.2.4. . By doing so, an answer could possibly be given to the question if all these processes either happen extracellularly, with a complete disassembly of the SLs, or if at some point certain SL derivatives are taken up again by the cells with further catabolism occuring intracellularly.

At the time of the execution of these experiments an extracellular 'lipase' responsible for lactonisation of acidic SLs was identified (cfr. Chapter VI). The possibility that this enzyme is also responsible the opposite activity namely ringopening of lactonic SLs once cultures shift from production to catabolisation phase was hence also investigated. This was done by performing an additional incubation experiment with the di-acetylated lactonic SLs and the cell free culture medium obtained of a knock out strain of this gene, which was created at our laboratory (Ciesielska *et al.*, 2013).

- *Di-acetylated lactonic SLs (wild type)*

The prepared solution was sampled regularly and the results for SL composition can be found in Figure III. 6.

The pH of the solutions remained stable at about 5.5 for both the assays and the controls throughout the experiment. A conversion of the di-acetylated lactonic form of SLs (27.6 min) to the di-acetylated acidic form and to the mono-acetylated lactonic form (co-eluting at 23.4 min) was detectable after 7 days of incubation for the assays (not shown). After 10 days of incubation almost all of the added di-acetylated lactonic SL was converted to the di-acetylated acidic form and to a lesser extent to the co-eluting mono-acetylated lactonic form (cfr. Figure III. 6 b). The conversion was complete after 32 days of incubation (cfr. Figure III. 6 c). Some new peaks also appeared between 20 and 23 minutes. These correspond to mono- and non-acetylated acidic SLs. Some ringopening and deacetylation was also observed for the control incubation of SLs (cfr. Figure III. 6 right panels), but these spontaneous effects were a lot less

pronounced than those for the incubation with the cell free culture medium as can be concluded by comparing the left and right panel of Figure III. 6 c.

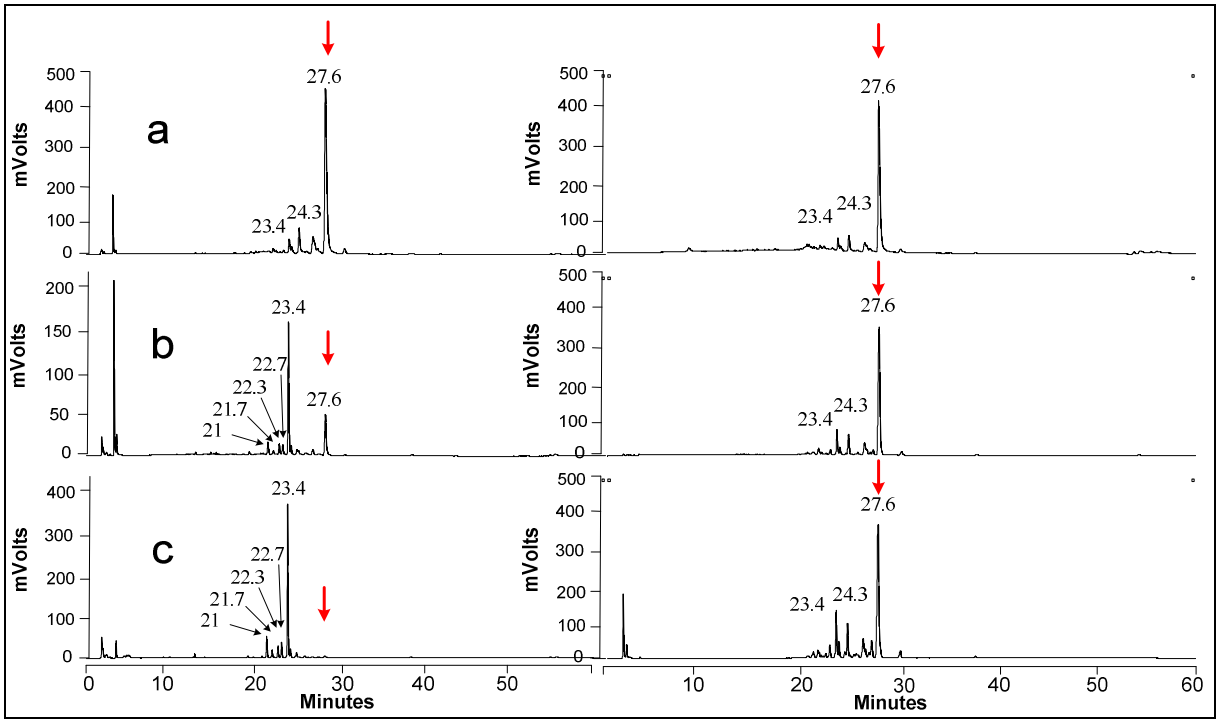


Figure III. 6 **(left)** HPLC-ELSD chromatograms of samples from cell free culture medium of the wild type (unconcentrated secretome) incubated (30°, 200rpm) with di-acetylated lactonic SLs (27.6 min, red arrow) after **(a)** 1 hour **(b)** 10 days and **(c)** 32 days of incubation. **(right)** controls for SLs at each time point.

- *Non-acetylated acidic SLs (wild type)*

The prepared solution was sampled regularly and the results for SL composition can be found in Figure III. 7.

The added non-acetylated acidic SLs remain unaltered when incubated with the unconcentrated secretome and behave identical as the control as can be concluded from the analyses of the samples depicted in Figure III. 7. Incubation of non-acetylated acidic SLs with the unconcentrated secretome thus did not lead to the formation of glucolipids and/or fatty acids, so acetylation does not seem to be a protective mechanism against attack by an extracellular glycosylhydrolase-like enzyme, which is the case for cellobioselipids produced by *P. flocculosa* (flocculosin) (cfr. Chapter I.3.2), where deac(et)ylation of the glycolipids at high pH values leads to a very fast extracellular metabolisation of the deac(et)ylated derivatives (Mimee *et al.*, 2009a). This indicates that these non-acetylated acidic molecules

are the end-products of the extracellular catabolism of SLs by *S. bombicola*. They are thus hypothesized to be taken up again by the starved cells to be further catabolised intracellularly, explaining their disappearance of cultures cultivated on SLs as the sole carbon source (cfr. III.3.2.2). However, acetylation could still be a protective mechanisms to prevent uptake of acetylated SL derivatives, as was suggested for other types of molecules by (Danchin, 2009).

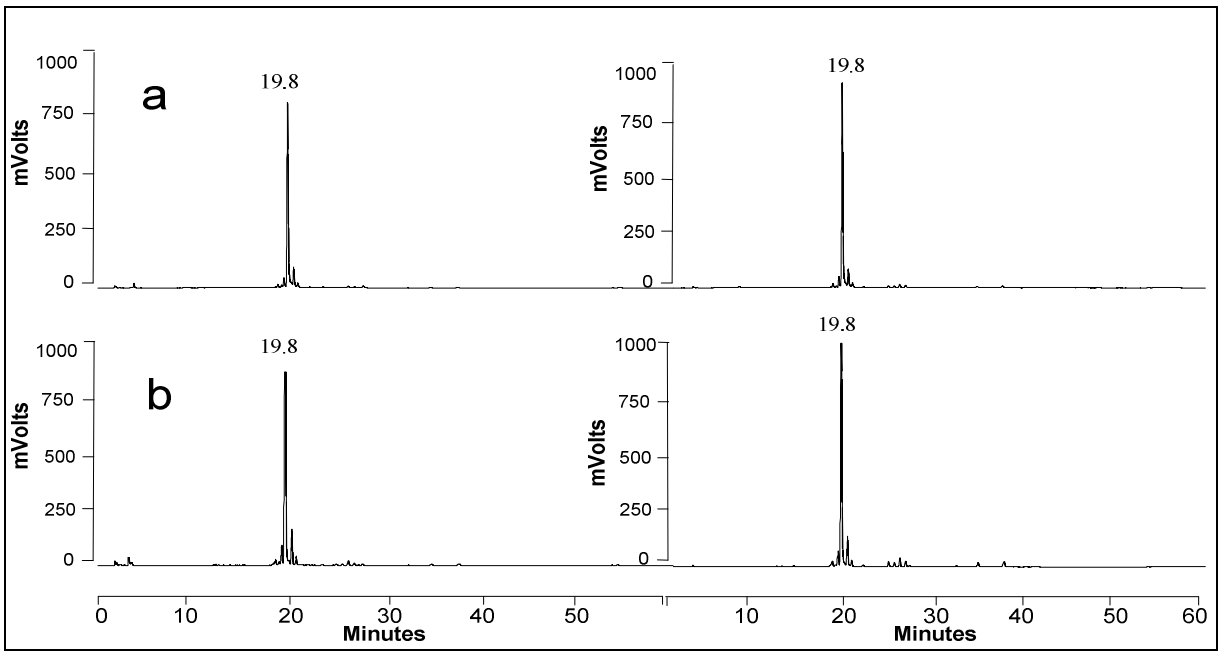


Figure III. 7 **(left)** HPLC-ELSD chromatograms of samples from cell free culture medium of the wild type incubated (30°C, 200rpm) with non-acetylated acidic SLs (19.8 min) after **(a)** 1 hour **(b)** 16 days of incubation. **(right)** Controls for SLs at the same time points.

The obtained results were confirmed by repeating these experiments. It was furthermore shown that the results were identical whether or not oil had been added to the cultures (after 48 h of cultivation) and that the obtained results for di-acetylated lactonic SLs were unaltered at pH values of 3.5, indicating that the responsible enzymes are at least active in a pH range between 3.5 and 5.5. Taken together, these results clearly confirm the hypothesis stated in III.3.2.2, that at least ringopening of lactonic and deacetylation of lactonic and/or acidic SLs are events effectuated by (an) extracellular enzyme(s).

- *Di-acetylated lactonic SLs (Δlip)*

The possibility that the lipase enzyme responsible for lactonisation of acidic SLs is also responsible for the opposite activity, namely ringopening of lactonic SLs which was detected in the abovementioned experiments, is investigated in this point. This was done by repeating the incubation experiment with the di-acetylated lactonic SLs, but this time with the cell free

culture medium (unconcentrated secretome) of a knock out strain of this gene (Δ*lip*) (Ciesielska *et al.*, 2013).

The prepared solution was sampled regularly and the results for SL composition can be found in Figure III. 8.

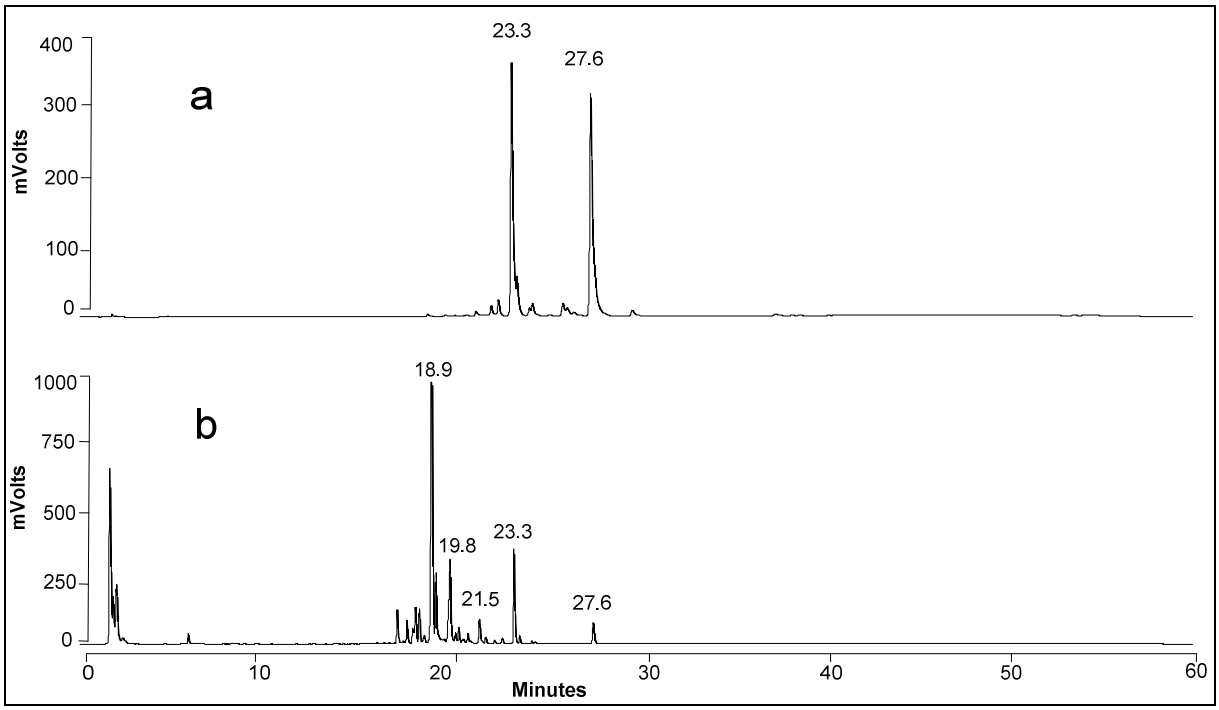


Figure III. 8 HPLC-ELSD chromatograms of samples from the cell free culture medium of a *Δlip* mutant incubated (30°C, 200 rpm) with di-acetylated lactonic SLs (27.6 min) after **(a)** 16 days and **(b)** 32 days of incubation. Controls for SLs for each time point can be found in Figure III. 6 on the right panel.

Because this *Δlip* mutant only produces (high amounts) of acidic SLs (peak at 23.3 min), which are well water soluble, removal of the cells from the culture medium does not lead to removal of these SLs. These were therefore present in the assay for the cell free culture medium from the beginning of the experiment. When comparing Figure III. 6 b with Figure III. 8 a it is hence hard to assess if ring opening of lactonic SLs also occurs in the absence of the lipase enzyme, as these molecules were already present in the assays from the beginning. However, after 32 days of incubation a clear effect is visible and ring opening can be concluded in a lipase- and cell free culture medium, as the ratio between the acidic (23.3 min) and lactonic (27.6 min) SLs clearly increased, which can be concluded when comparing Figure III. 8 a with Figure III. 8 b. It can therefore be concluded that the lipase is not (the only) enzyme responsible for ring opening.

However, another interesting observation was done when comparing these results with the analogue ones for the wild type (cfr. Figure III. 6). Whereas for the wild type the peak at 23.3 minutes is still the predominant one after 32 days of incubation (cfr. Figure III. 6 c), a partial conversion of this peak to the ones eluting at 18.9, 19.8 and 21.5 min corresponding to the non- (18.9 and 19.8 min) and mono-acetylated (21.5 min) acidic SLs can be observed for the Δ*lip* mutant (cfr. Figure III. 8 b). The initial presence of the di-acetylated acidic SLs (23.3 min) in rather high amounts in the assay is the most plausible explanation for this observation, as the mono- and non-acetylated acidic SLs are derived from deacetylation of the di-acetylated acidic ones. For the wild type experiment these 'substrate' molecules are only present in comparable amounts after 32 days of incubation, when the sampling was stopped. Alternatively this observation could indicate that for the wild type there is some kind of balance between lactonisation of di-acetylated acidic SLs and ring opening of the resulting lactonic SLs, which then wouldn't be the case anymore for a lipase knock out. Alternatively one could presume that this lipase could also be involved in (less efficient) acetylation of acidic non-acetylated SLs, but this presumption is rather farfetched as no acetyl-CoA was added. Enzyme assays with the purified lipase enzyme in the future will give more insight in this.

III.3.4. *Assays*

To further investigate intracellular degradation of SLs and to furthermore enable the correct (protein concentrations, pH etc.) comparison of several situations, two kinds of assays with concentrated and standardized protein solutions were subsequently set up; one for the extracellular (secretome) and one for the intracellular (cell pellets) proteins. The final aim was to perform fractionation experiments with these standardized assays, which could possibly lead to the identification of enzyme(s) involved in the catabolism of SLs.

These experiments were furthermore repeated with secretomes and lysates derived from younger cultures to assess if these catabolic processes are already active when the cells are in full biosynthetic production mode, which was reported for *R. bogoriensis* (Bucholtz and Light, 1976). The first one was cultivated on the production medium for only 96 hours (4 days) and the second one was stopped when glucose had been fully depleted from the culture medium for 24 h, which occurred after 240 hours of cultivation (10 days). Unfortunately some of these last experiments failed, amongst others because something went wrong with the di-acetylated

lactonic SL substrate. These experiments should thus be repeated to be able to answer all the research questions.

III.3.4.1. *Concentrated secretomes of wild type cultures*

The concentrated secretomes of wild type cultures, cultivated on production medium or on modified production medium were prepared as described in III.2.5.1 and the assays were performed as described in III.2.5.3.

- *Di-acetylated lactonic SLs*

When the di-acetylated lactonic SLs were used as the substrate with a secretome of cultures grown on production medium, identical effects as the ones described above (III.3.3) and depicted in Figure III. 6 were obtained and the results are hence not shown here.

However, when repeating this experiment with a secretome derived from a culture cultivated on SLs as the sole C-source, the demonstrated activity was a lot higher; all of the added substrate (di-acetylated lactonic SLs) was completely converted to di-acetylated acidic SLs (and to a lesser account to mono-acetylated lactonic SLs) after only 6 days of incubation, whereas this wasn't the case for the secretomes derived from the cultures on regular production medium. This indicates that higher relative concentrations of the responsible enzymes are present in such secretomes, as the used protein concentrations in the two assays were identical. These conditions thus apparently lead to upregulation of the responsible gene(s), which is a straightforward and useful observation for further investigation.

- *Non-acetylated acidic SLs*

In contrast with the assays of the unconcentrated secretome with non-acetylated acidic SLs described in III.3.3., assays performed with concentrated secretomes derived from a culture on normal production medium or from one that was cultivated on SLs, both gave rise to the detection of some catabolic activity (after 14 days of incubation). The observed conversions were the same as those observed for the lysates, which will be described below (III. 3.4.2), but a lot less prominent. Shortly said, non-acetylated acidic SLs were dissimilated into glucolipids and hydroxylated fatty acids. To avoid repetition the reader is referred to III.3.4.2 for the discussion of these results. Repeating the experiments with younger cultures of 4 days and 10 days old did not give rise to any detectable catabolic activity on the added non-

acetylated acidic SLs, not even after 14 days of incubation, so the responsible enzymes do not seem to be expressed yet at these time points.

These results thus clearly differ from those obtained for the untreated raw supernatant described in III.3.3 for which absolutely no extracellular catabolic activity on non-acetylated acidic SLs was detected even after 16 days of incubation. The pH could be a possible explanation, as the pH of the experiment described in III.3.3 remained equal to 3.5 throughout the incubation, whereas here the pH of the concentrated secretomes was set at 7. The responsible enzyme(s) for disassembly of non-acetylated acidic SLs are only present (in substantial amounts to detect activity) in the extracellular space of old cultures and of cultures that were cultivated on SLs as the sole C-source. This can either indicate that on production medium they are only produced/secreted after prolonged cultivation or alternatively that at earlier time points they are degraded by specific acidic proteases (cfr. III.3.5). Alternatively these enzymes have a very low k_{cat} (the measured activity here was also very low) or are not actively secreted into the culture medium (leading to very low concentrations in the experiment described in III.3.3), which would imply that their presence in the secretome of older cultures results from leakage of these (intracellular) enzymes from damaged/dead cells. Whatever the reason, the detected catabolic activity on non-acetylated acidic SLs was a lot higher intracellularly (cfr. III.3.4.2), so from all the experiments executed with the secretomes, one might conclude that the catabolic pathway of non-acetylated acidic SLs is preferably performed intracellularly. This could be a strategy of the yeast to avoid the release of compounds like glucose and fatty acids in the extracellular space, where they are also accessible to competitors, but instead leave the carbon 'secured' to further degrade it intracellularly.

III.3.4.2. *Lysates of wild type cultures*

Lysates of wild type cultures cultivated on production medium were prepared as described in III.2.5.2 and assays with di-acetylated lactonic and non-acetylated acidic SLs were performed as described in III.2.5.3. Parallel experiments were run, which allows the assessment of the formation of catabolic intermediates in function of time (one assay can maximally be sampled twice). The assays were sampled after 24h, 48h, 72h, 96h, 6 days and 9 days respectively and chromatograms after incubation times of 6 days for lactonic and 9 days for acidic SLs are depicted in Figure III. 9 and Figure III.10 respectively.

- *Di-acetylated lactonic SLs*

Incubation of the lysate with di-acetylated lactonic SLs (peak at 27.6 min) (cfr. Figure III. 9 a) gives rise to very similar results as those obtained for cultivation of *S. bombicola* on SLs as the sole C-source (cfr. Figure III. 5).

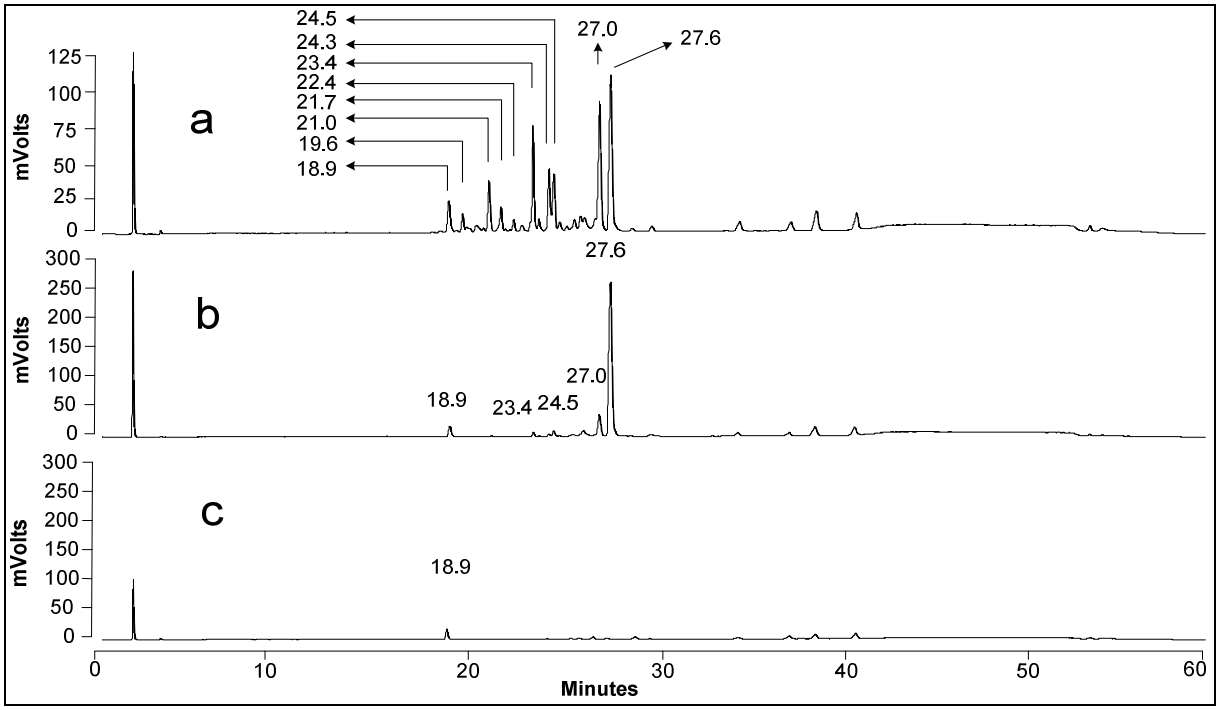


Figure III. 9 HPLC-ELSD chromatograms of samples from **(a)** an assay of a lysate from the wild type and di-acetylated lactonic SLs (27.6 min) after 6 days of incubation. **(b)** and **(c)** correspond to the controls for SLs and the lysate respectively. The peak at 18.9 minutes corresponds to an internal standard (C12:0-OH).

Actions like ring opening (23.4 and 24.3 min) and (subsequent) deacetylation of the resulting di-acetylated acidic SLs to mono- (21.0 and 21.7 min) and non-acetylated acidic intermediates (19.6 min) were detected after analysis of the samples on LC-MS (cfr. Table III. 5). Deacetylation of the di-acetylated lactonic SLs giving rise to mono-acetylated lactonic SLs was also detected (23.4, 24.3 and 24.5 min). All these intermediates were already detected after only 24 hours of incubation, but the effects became more pronounced upon longer incubation. All but one: the emergence of the non-acetylated acidic SLs eluting at 19.6 minutes became only evident after 6 days of incubation.

A preference for deacetylation at a specific position is suspected to exist, as there is a difference in the amounts of the respective mono-acetylated acidic derivates eluting at 21 and 21.7 minutes, which was also observed in Figure III. 5. Such preference was also suggested

by (Bucholtz and Light, 1976) for deacetylation of SLs produced by *Rhodotorula bogoriensis*, and was also demonstrated for the opposite action, namely biosynthetic acetylation of SLs by *S. bombicola* (Saerens, 2012). The latter is more efficient for the inner 6' position as compared to the 6'' position and the formation of di-acetylated SLs is preferably done in that order, but also occurs (at least *in vitro*) the other way around. This also explains the rather late occurrence of the non-acetylated acidic SLs.

Similarly as for the results described in III.3.2.2 and to a lesser account in III.3.3, deacetylation of the di-acetylated lactonic SLs was also detected. The resulting mono-acetylated lactonic SLs corresponding to one MW (MW 645) were found to be eluting at 4 different retention times in LC-MS chromatograms, as is also observed in Table III. 4 and Table III.5. This is caused by either the linkage of an ω- or ω-1 hydroxylated C18:1 fatty acid in combination with acetylation at either the 6' or 6'' position. The compound with MW of 645 eluting at 24.5 minutes is a lot more abundant than the other ones (concluded from analyzing LC-MS chromatograms), which could again indicate that one of the acetylgroups is less attainable for deacetylation by the same or a second deacetylase as the one acting on the acidic SLs.

Selective hydrolysis of the acetylgroup on the 6' position of di-acetylated lactonic SLs was described by (Dekoster *et al.*, 1995) after incubation of di-acetylated lactonic SLs with a cutinase derived from *Fusarium solani*. Full deacetylation of di-acetylated lactonic SLs leading to the formation of non-acetylated lactonic SLs was on the other hand shown with three commercial acetylesterases (Asmer *et al.*, 1988). Moreover, it has been proposed, based upon 3D models, that lactonic SLs fit well in the binding pocket of several lipases (Carr and Bisht, 2003). When this occurs, the 6'' acetylgroup is buried inside the macrolactone structure, however, the one on the 6' position is well accessible to the active site of the lipase leading to deacetylation of the 6' position. Surprisingly when these experiments were repeated with the non-lactonic SLs, these were not deacylated (Carr and Bisht, 2003). The authors suggested that the macrolactone provides a structural motif to the SLs that allow it to fit in the binding cavity of the lipase such that an acylated sophorose moiety is accessible to the active site. These data from the literature could thus point in the direction of the involvement of two separate acetyl hydrolases for the observed deacetylating activities on lactonic and acidic acetylesterases respectively.

Table III. 5 Identification of the major peaks detected in Figure III. 9 after LC-MS analysis, of which the most prominent ones are framed.

Elution time	mass	identity		acetyl?
18.9	216	hydroxylated fatty acid	C12:0	
19.6	621	SL open	C18:1	non
20.4	637	SL open	C16:0	mono
20.8	623	SL open	C18:0	non
21.0	663	SL open	C18:1	mono
21.7	663	SL open	C18:1	mono
22.4	665	SL open	C18:0	mono
23.4	645 and 705	SL closed/SL open	C18:1	mono/di
23.8	645	SL closed	C18:1	mono
24.3	645 and 705	SL closed/SL open	C18:1	mono/di
24.5	645	SL closed	C18:1	mono
26.24	647	SL closed	C18:0	mono
27.0	687	SL closed	C18:1	di
27.6	687	SL closed	C18:1	di
38.9	255	fatty acid	C16:0	
41.0	283	fatty acid	C18:0	

Further disassembly of SLs was not detected, possibly because the incubation time wasn't long enough to allow the build up of non-acetylated acidic SLs (peak at 19.6 min) of which catabolic intermediates could subsequently be derived and detected.

- *Non-acetylated acidic SLs*

Incubation of the lysate with the non-acetylated acidic SLs leads to a clear fragmentation of these compounds into glucolipids and hydroxylated fatty acids as can be concluded from Figure III. 10 and Table III.6. These effects were visible after 6 days of incubation, and weren't observed yet after 4 days of incubation, indicating a slower process then the one described above for di-acetylated lactonic SLs (activity detected after 24 h). However, the detected activities are a lot higher than those detected (identical total protein concentrations) in the extracellular space described in III.3.4.1, which as stated indicates that this dissassimilation of non-acetylated acidic SLs predominantly happens intracellularly. None of the abovementioned effects were observed when repeating this experiment for a culture that had been incubating for 96 hours, so again the responsible enzymes are probably not expressed yet at that moment.

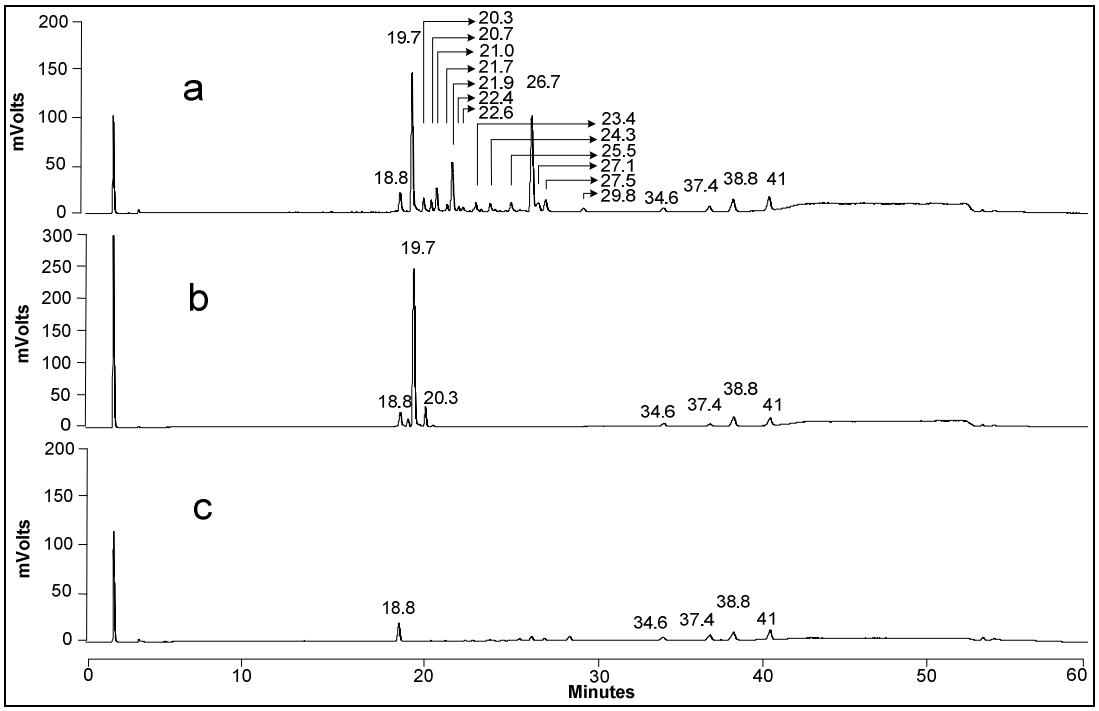


Figure III. 10 HPLC-ELSD chromatograms of samples from **(a)** an assay of a lysate from the wild type and non- acetylated acidic SLs (peak at 19.7 min) after 9 days of incubation. (**b)** and **(c)** correspond to the controls for SLs and the lysate respectively. The peak at 18.8 min corresponds to an internal standard (C12:0-OH).

The biggest peaks appearing correspond to C18:1 glucolipids (21.9 min) and C18:1 hydroxylated fatty acids (26.7 min) derived from the substrate (mainly non-acetylated acidic SLs (C18:1)) eluting at 19.7 minutes. Other compounds include other hydroxylated fatty acids derived from other SL variants present in the substrate, giving rise to C18:0, C18:2 and C16:0 hydroxylated fatty acids.

Hydroxylated fatty acids also appeared in the LC-MS chromatograms of the incubation of the control for the lysate after 7 days of incubation (SLs also still present/soluble), but to a much lesser extent (when compared to the standard) than the effect observed for the assays. These hydroxylated fatty acids are probably derived from the acidic SLs that are still present in the lysate in small amounts (also detectable on LC-MS). However, it can also be argued that these intermediates are derived from hydroxylation of the corresponding non-hydroxylated fatty acids present in the lysate (predominantly C16:0 and C18:0) by the action of CYP enzymes.

Table III. 6 Identification of the major peaks detected in Figure III. 10 after LC-MS analysis of which the most prominent ones are framed.

Elution time	Mass	Identity		Acetyl?
18.8	215	hydroxylated fatty acid	C12:0	
19.7	621	SL open	C18:1	non
20.3	623	SL open	C18:0	non
20.7	663	SL open	C18:1	mono
21	459	glucolipid	C18:1	non
21.7	665	SL open	C18:0	mono
21.9	459	glucolipid	C18:1	non
22.4	459	glucolipid	C18:1	non
23.4	705	SL open	C18:1	di
24.3	501	glucolipid	C18:1	mono
25.5	295	hydroxylated fatty acid	C18:2	
26	271	hydroxylated fatty acid	C16:0	
26.7	297	hydroxylated fatty acid	C18:1	
27.1	299	hydroxylated fatty acid	C18:0	
27.5	297	hydroxylated fatty acid	C18:1	
29.8	299	hydroxylated fatty acid	C18:0	
34.6	321+227	?		
38.8	255	fatty acid	C16:0	
41	283	fatty acid	C18:0	

Non-acetylated acidic SLs are thus disassembled intracellularly into glucolipids and hydroxylated fatty acids, suggesting that at least a stepwise removal of the glucose molecules is occurring. This was confirmed by measuring the glucoseconcentrations of the end samples (after 9 days of incubation). Whereas glucose was clearly released in the assays (an amount equal to 20 % of the total amount of glucose initially captured in the added SLs was found back in the final sample of the assay), no glucose release was detected in the control, so the glucose released in the assay was clearly derived from SLs. However, this result does not rule out the possibility that the sophorose molecule can (also) be removed as a whole of the fatty acid, which can subsequently be followed by hydrolysis of the disaccharide giving rise to two glucose molecules.

Indeed, minor amounts of a compound corresponding to the mass of sophorose were detected by performing LC-MS of these assays with the method described in III.2.2.2. The latter is in contrast with flocculosin (cellobioselipid) degradation, for which glucose was shown to be released by catabolic enzymes, whereas cellobiose was never detected (Mimee *et al.*, 2009a).

One last remark that should be made here is that this incubation of the lysate of the wild type with non-acetylated acidic SLs also resulted in the formation of acetylated intermediates (glucolipids and SLs) (cfr. Table III.6). These compounds weren't present in the controls, and can hence only arise from the action of the acetyltransferase responsible for acetylation of SLs (Saerens *et al.*, 2011b). This indicates that this enzyme must still be present and active in the cell lysates of cultures that were cultivated for 21 days. This acetylation was already detected after only 24 hours of incubation, clearly indicating a more efficient biosynthetic process as compared to the catabolic one hydrolysing the glycosidic bonds.

III.3.5. *Proteomics*

One of the main conclusions that can be drawn from the abovementioned experiments is the presence of secreted extracellular enzymes responsible for ringopening of lactonic SLs and deacetylation of lactonic and acidic SLs. The acidic SLs are likely to be taken up again by the starved cells, leading to further intracellular degradation. It was furthermore found that cultivating *S. bombicola* on a medium containing SLs as the sole C-source not unexpectedly results in an upregulation of the catabolic enzymes in the secretomes. A first attempt was done to fractionate the secretome (and lysate) with the aim of eventually identifying the responsible enzymes, but this still resulted in several fractions with the observed activities. An experiment was thus set up to perform proteomic experiments, both on secretomes of cultures with SLs as the sole carbon source as on secretomes of old cultures on normal production medium, as an alternative way of finding these enzymes. Proteomic analysis was performed by PhDs Katarzyna Ciesielska as described in section III.2.6.

First, the secretome of two cultures cultivated on SLs as the sole carbon source (modified SD medium) were subjected to proteome analysis. The first one was harvested at maximal budding, after 4 days of incubation. The second one, when budding had started to diminish, which occurred after 7 days of incubation. These two secretomes had to be combined to obtain protein concentrations high enough for proteomic analysis. Secondly, a culture of *S. bombicola* on production medium was harvested after 21 days of cultivation, when the precipitated SLs clearly started to disappear from the bottom of the flask (cfr. III.3.2.1). The combined results of proteins that were detected in both conditions are summarized in Table III. 7. All hits that were found for the combination of the first two experiments were also found for the second experiment with the exception of a predicted secreted lipase (cabom03g17600). Another predicted lipase (cabom02g1720) was found back in both proteomic experiments and

corresponds to the enzyme responsible for lactonisation of SLs. As SL biosynthesis is not expected to occur under these conditions (SL dissimilation), it is somewhat peculiar that this protein is detected in these experiments.

The other predicted secretory lipase (cabom03g17600) was, as mentioned above, only detected in the combined secretomes of the cultures cultivated on SLs as the sole C-source. As ringopening and/or deacetylation of lactonic SLs was found to be happening a lot faster for the secretome of such cultures in comparison with a degrading 'old' culture on production medium (cfr. III.3.4.1), this could indicate that this putative lipase could possibly be involved in the observed ringopening (and/or deacetylation). This was the first gene that was selected for further investigation.

Another protein that attracted our attention was a protein with homology to a hydrophobic protein A (HsbA) of *Aspergillus oryzae*. The latter protein binds to hydrophobic materials such as the biodegradable plastic PBSA in the presence of NaCl or $CaCl_2$, which leads to recruitment of a polyesterase cutinase (CutL1). When CutL1 is accumulated on the PBSA surface, it stimulates PBSA degradation (Ohtaki *et al.*, 2006). CutL1 more effectively degraded PBSA microparticles precoated with HsbA than those without HsbA. One could imagine similar interactions to also occur in a shake flask experiment with insoluble precipitated SLs. However, this needs further investigation. This was a second gene that was selected for further investigation (cfr. III. 3.6). An additional remark that has to be made here is that degradation of polycaprolactone (a synthetic plastic) by a cutinase produced by *Fusarium* strains was strongly dependent on the pH; pH drop in the culture medium resulted in inactivation of the cutinase by an extracellular aspartyl protease appearing during such pH drop (Murphy *et al.*, 1999). Such events could also prevent SL degradation by *S. bombicola* when carbon is still available and pH is low. Carbon starvation and subsequent pH rise could then possibly lead to full activity of the responsible enzyme(s) and dissimilation of SLs. Two such proteases (Cabom01g10760 and Cabom01g03660) were identified in the secretomes, so this could be a plausible assumption.

A lot of hypothetical glycosyl hydrolases were also detected. Most of them with predicted functions involved in cell wall degradation and/or -organization, but a few with no clear predicted function. Because some hydrolase activity was detected in the secretome of *Starmerella bombicola*, two secreted glycosyl hydrolases with no clear predicted function were selected for further investigation (Cabom01g03900 and Cabom02g09580). It is however

Table III. 7 Combined results of a proteomic experiment on the secretome of *S. bombicola* cultured on production medium and a medium with SLs as the sole C-source

Gene	Database hit	Predicted function
Cabom02g07380	Hydrophobic surface binding protein A (*A. oryzae*) (HsbA)	? Possibly recruiting SL degrading enzymes?
Cabom01g01190	Endo-1,3(4)-beta-glucanase (DSE1); Glycosyl hydrolase family 81	Daughter cell-specific secreted protein with similarity to glucanases, degrades the cell wall from the daughter side causing the daughter cell to separate
Cabom01g00590	Endochitinase (CTS1)	Required for cell separation after mitosis
Cabom02g02320	1,3-beta-glucanosyltransferase (GAS1)	Required for cell wall assembly + role in transcriptional silencing
Cabom03g08560	Beta-glucosidase (SUN family) (SIM1)	Cell wall organisation
Cabom03g02060	Probable glycosidase CRF2 (*A. fumigatus*), Glycosyl hydrolase family 16	Cleavage of β-1,4 or β-1,3 glycosidic bonds in various glucans and galactans
Cabom02g11140	Lysophospholipase (PLB3)	Phospholipid metabolism; hydrolyzes phosphatidylinositol and phosphatidylserine. Transacylase activity in vitro
Cabom02g08250	Cell wall protein (ECM33)	Required for proper cell wall integrity and for the correct assembly of the mannoprotein outer layer of the cell wall. Important for apical bud growth
Cabom03g05170	Probable 1,3-beta-glucanosyltransferase (GAS3)	Putative 1,3-beta-glucanosyltransferase, possible inactive member of the GAS family
Cabom01g03900	Uncharacterized lipoprotein ybbD (*B. subtillis*) Glycosyl hydrolase family3	? Involved in SL metabolism? (hydrolyzing O-glycosyl compounds)
Cabom02g11610	Catalase-peroxidase (KatG)	Protecting cells from H_2O_2 toxicity
Cabom02g10760	Null hit	? Involvement in SL degradation?
Cabom02g01580	Aspartic proteinase yapsin (YPS3)	Cell wall growth and maintenance; attached to plasma membrane via GPI anchor
Cabom02g06860	Purine nucleoside permease (*S. stipitis*)	Transmembrane transport of nucleosides
Cabom03g07360	Protein PRY2, Cysteine-rich secretory protein family	Sterol binding protein involved in the export of acetylated sterols.
Cabom01g05390	Transaldolase (*S. pombe*)	Enzyme of the non-oxidative phase of the pentose phosphate pathway
Cabom01g10760	Candidapepsin SAPT1 (*C. tropicalis*), Aspartate protease	Protein degradation, inactive proenzymes that activate autocatalytically at acidic pH
Cabom01g08280	Carboxypeptidase Y CPY1 (*C. albicans*), Serine Carboxypeptidase	Protein catabolism or maturation
Cabom01g03660	Candidapepsin SAPT1 (*C. tropicalis*), Aspartate protease	Protein degradation, inactive proenzymes that activate autocatalytically at acidic pH
Cabom01g14090	Ribonuclease T2 (*Y. lipolytica*)	Secretory Rnase. Possibly regulates membrane permeability or stability.
Cabom02g01720	Secretory Lipase (LIP1) (*C. albicans*)	Lactonisation of sophorolipids (proven)
Cabom01g03830	Protein AXL2, MID2 domain	Plasma membrane protein required for axial budding in haploid cells.
Cabom02g09580	Bardwin related endoglucanase	? Involved in SL degradation?
Cabom03g17600	Secretory Lipase (LIP1) (*C. albicans*)	? Involved in SL degradation?

not unlikely that hydrolases predicted to have a role in e.g. cell wall degradation and/or organization could also show glycosyl hydrolase activity towards SLs.

A predicted lysophospholipase was also selected for further investigation. Phospholipases B are characterized as enzymes that sequentially remove both fatty acyl groups from diacylglycerolphospholipids and therefore have both phospholipase A and lysophospholipase activities. This gene was also considered to be a candidate for involvement in SL catabolism.

For one detected protein (Cabom02g10760) no, or very bad database hits were found. This protein was expressed in both conditions and one of the 'best' hits was a hypothetical protein of *Ustilago maydis*, a fungus that also produces glycolipid biosurfactants and for that reason we also selected this protein for further investigation.

Finally the expression of a protein (Cabom01g03900) similar to a lipoprotein of *Bacillus subtilis* is interesting. *Bacillus subtillis* is known to produce more than two dozen different antibiotics and one of the most powerful biosurfactant, surfactin. Several of these antibiotic producers possess membrane-bound lipoproteins which exhibit a sequestering-like function that prevent high local concentrations of the produced antibiotics close to the cytoplasmatic membrane (or interferes with antibiotic pore formation). Self protection against these antimicrobial secondary metabolites is based upon this effect in addition to ABC transporter proteins that export them from the cytoplasmatic membrane into the extracellular space (Stein, 2005). *Starmerella bombicola* is insensitive to high titers of SLs so some form of immunity could also be attributed to the presence of such proteins. As the predicted protein was furthermore a putative glycosyl hydrolase, this gene was also selected for further investigation.

As several yeast and fungi are capable of degradation of glycolipids, a hydrolysis thus executed by different (related) enzymes, it is plausible that no specific enzyme(s) exist for the hydrolysis of the β-1,2 bond of sophorose or the O-glycosyl bond between the sophorose and the fatty acid moiety. This hypothesis is strengthened by the fact that non-acetylated acidic SLs can be hydrolyzed to glucolipids by a wide array of commercial enzyme (mixes) (Rau *et al.*, 1999). Genes could thus be redundant and knocking out a gene responsible for the observed action could still be rescued by another gene executing the same function. Genes responsible for cell wall degradation for example could possibly also be responsible for hydrolysis of the glycosidic bonds present in SLs. Deacetylation and ringopening seem to be more specific and it is hence more likely to see effects when knocking out these 'unique' genes. This statement is substantiated by the fact that deacetylation of di-acetylated SLs

produced by *Rhodotorula bogoriensis* was only effectuated by one unique acetylesterase of several enzymes with acetylesterase activity in cell extracts of this yeast (Bucholtz and Light, 1976).

III.3.6. *Knocking out genes derived from proteomic analysis*

One protein that was omnipresent in proteomic analyses was a protein that was predicted as 'hydrophobic protein A (HsbA)' (cfr. Table III. 7). Because this protein was also highly abundant in the secretome of cultures degrading SLs and because of the abovementioned reasons, a knock out strain of this gene was created.

For the *ΔhsbA* mutant no effect was observed as compared to the wild type when cultivating it on production medium for a prolonged period of time. CFUs, glucose consumption and SL catabolism were identical. This could mean that this protein is not involved in SL catabolism or alternatively that the effects exerted by this protein are marginal. Attempts will be done to assess if this mutant is able to grow on SLs as the sole C-source and assays like the ones described above will also be performed once more knock out strains are available.

Knock outs for the other mentioned candidates are currently in the pipeline, but unfortunately no further results are available yet.

III.4. Conclusion

In this chapter the catabolism of SLs produced by *S. bombicola* was investigated. We found that this glycolipid producing yeast is able to grow on SLs as the sole C-source and can thus effectuate catabolism of its own glycolipids in contrast to what was stated by (Lo and Ju, 2009). This catabolism for one thing consists of the action of secreted enzymes at least responsible for ringopening of di-acetylated lactonic SLs, deacetylation of the formed di-acetylated acidic SLs and deacetylation of di-acetylated lactonic SLs. These processes eventually give rise to the non-acetylated acidic SLs. Further catabolism of these compounds by hydrolysis of the glycosidic linkage between the two glucose moieties and between the sophorose- and fatty acid moiety was (in minor amounts) was shown to happen predominantly intracellularly. The major catabolic route for *S. bombicola* is thus suggested to consist of the extracellular ringopening and deacetylation eventually leading to non-acetylated acidic SLs. These molecules are subsequently taken up by the starved cells and further catabolised intracellularly into glucolipids, hydroxylated fatty acids, sophorose and glucose.

This could be a strategy to protect the extracellular storage material against other competing microorganisms. Full and efficient extracellular degradation of deacetylated acidic SLs would lead to the release of glucose and fatty acids in the extracellular space, also accessible to other (competing) microorganisms. This hypothesis could require the presence of a SL transporter other than the MDR transporter as this is an ABC transporter and these are exclusively reported to show export activity. For other biosurfactants, the involvement of transporters of the major facilitator family (MFS) was described (Hewald *et al.*, 2006). Different families of such MFS permeases exist, each effectuating transport (with inwardly and/or outwardly directed polarity) of a single class of compounds like simple sugars, oligosaccharides, inositols, drugs, amino acids, nucleosides, organophosphate esters, Krebs cycle metabolites, and a large variety of organic and inorganic anions and cations (Pao *et al.*, 1998). As stated in Chapter II 3.2 several of such candidates can be found in the *S. bombicola* genome. It also has to be mentioned that the responsible enzymes should be very stable, as activity is still detected after 14 days of incubation (30 °C) for the assays, and 30 days for experiments with crude secretomes.

When looking at flocculosin degradation by *P. flocculosa*, deac(et)ylation was reported to be an absolute prerequisite for further metabolisation of the non ac(et)ylated cellobioslipid into glucolipids and hydroxylated fatty acids (cfr. Chapter I 3.2), which also seems to be the case for SL catabolism by *S. bombicola*. However, the catabolic pathway for flocculosin was shown to completely exist extracellularly and to be very efficient, as complete disassembly of the added substrate (1 g/L) into glucolipids and hydroxylated fatty acids was effectuated after only 24 hours of incubation with a concentrated cell free culture medium. The authors suggested that in culture conditions the spontaneous deac(et)ylation of flocculosin, effectuated by the pH rise caused by the usage of proteins as the main energy source, allows the fungus to further metabolise the deac(et)ylated molecules to glucolipids and hydroxylated fatty acids and use these as an energy source during conditions of food shortages. The possible involvement of an acetyl hydrolase was not mentioned, whereas the involvement of such enzyme was shown for sophorolipid catabolism. For flocculosin catabolism no cellobiose was detected in any of the analyses, so only a stepwise release of the dissacharide from the fatty acid by the secreted proteins was suggested. For SL catabolism at least a stepwise removal of the glucose molecules also exists, as glucolipids and hydroxylated fatty acids were detected in the assays. Whereas (Garcia-Ochoa, 1996) reported the substantial release of sophorose from SLs, we did not quite share this observation as sophorose was only

detected in very small amounts. However, possible enzymatic cleavage of the released sophorose, could lead to an underestimation of the total release of this disaccharide in the assays. Sophorose is a rare sugar with industrial applications in the induction of cellulase production (Mandels *et al.*, 1962) and 10 mg costs about 300 $ (Sigma), so an enzyme selectively releasing sophorose from SLs could have some very interesting industrial applications.

When performing these 'degradation' experiments at least the biosynthetic activity of the acetyltransferase was also detected, which indicates an overlap between biosynthetic (acetyltransferase) and catabolic activities in these assays. This is in agreement with the detection of both acetyltransferase and -esterase activities in cell extracts of *Rhodutorola bogoriensis* (Bucholtz and Light, 1976) as was discussed in Chapter I. SL production was already described as a way of regulation the energy level of the cell (cfr. Chapter I). One can imagine that the overlap of these two distinct metabolisms could offer some advantages in quickly changing environmental conditions. Carbon excess in combination with nitrogen depletion would lead to SL production. When the carbon becomes depleted, cells can survive by using the produced and secured SLs as a carbon source. However, if a new pulse of carbon becomes available again, the produced acetyl-CoA can be easily used again for acetylation of deacetylated acidic SLs and *de novo* synthesis of fatty acids. As such the yeast is well adapted to the dynamic conditions occurring in the environment.

A first attempt was done to fractionate the secretome and lysate to be able to possibly identify the responsible enzymes, but this still resulted in several fractions with the observed activities. However, an upregulation of the responsible catabolic enzymes in the secretomes of cultures that were cultivated on SLs as the sole C-source was demonstrated. Proteomic experiments were thus performed on the secretomes of such cultures and of an old culture grown on general production medium and the results were compared as an alternative way of finding these enzymes. This resulted in the identification of a few possible candidates involved in the abovementioned catabolic pathway. Knock out strategies were elaborated and deletion strains are in the pipeline. One strain was already constructed for the gene *hsbA*, but this did not result in any detectable effects on SL metabolism.

The results described in this chapter should also be linked to susceptibility of some microorganisms to SL toxicity, whereas others aren't sensitive at all. These effects could be linked to possible degradation of these molecules by non sensitive organisms and could also

be an explanation for the insensitivity of the producing organisms for its own secondary products. This was already postulated for flocculosin, which did not show toxic effects at higher pH values, probably because of spontaneous deac(et)ylation of these molecules under these conditions (Mimee *et al.*, 2009a). Although some rare glycosidases exist that are able to degrade acetylated cellobioselipid (ustilagic acid) (Eveleigh *et al.*, 1964), generally spoken the ac(et)ylgroups protect the molecules against hydrolysis (Mimee *et al.*, 2009a) and are on the other hand also responsible for the antagonistic effects. It could thus be postulated that organisms producing some kind of extracellular hydrolase, capable of degrading the otherwise 'toxic' molecules would probably not be sensitive, as was suggested by (Eveleigh *et al.*, 1964). The catabolism of the produced glycolipids by the producing organism can thus not only be described as a way of securing carbon, but also as a possible way of coping with high concentrations of the biosurfactants.

Finally it has to be mentioned that a less efficient catabolism of SLs by *S. bombicola* as compared to cellobioselipid catabolism by *P. flocculosa* could possibly be (one) explanation for the very high SL titers obtained with *S. bombicola* as compared to cellobioslipid titers obtained with *P. flocculosa*. Identification of the responsible enzymes and creation of subsequent knock out strains in *S. bombicola* is not expected to result in (spectacular) changes in the obtained SL yields, as very high titers (up to more than 400 g/L) can already be obtained with this yeast. However, such strategies could be interesting for glycolipid producers with a very efficient catabolic pathway, possibly limiting the achievement of high yields of these molecules in the wild type strains. Furthermore, the development of *S. bombicola* as a platform organism might benefit from strains knocked out in the catabolic pathway as one can imagine that the expression of enzymes, transcription factors, transporters, etc. necessary for a fully active catabolic pathway require energy. Energy that is wasted if such pathway is not wanted anymore. Moreover, deletion of the catabolic enzymes in new industrial SL producing strains (cfr. Chapter VI) could guarantee the production of standardized products as the catabolic pathway is (partly) responsible for the occurrence of non- and mono-acetylated derivatives in the production process.

PART II:

DEVELOPMENT OF A MOLECULAR TOOLBOX FOR *STARMERELLA BOMBICOLA*

CHAPTER IV:

MOLECULAR TOOLS FOR ENGINEERING *STARMERELLA BOMBICOLA*

Part of this chapter has been published as a patent application:

Soetaert, W.; De Maeseneire, S. L.; Saerens, K., Roelants, S., and Van Bogaert, I. (2010). "Yeast strains modified in their sophorolipid production and uses thereof." UGent. WO2011154523 (A1).

Part of this chapter is in preparation for publication:

Roelants, S., De Maeseneire, L. S., Saey L., De Preester, K., Van Bogaert, I. and Soetaert, W. "Development of a molecular toolbox for the biosurfactant producing yeast *Starmerella bombicola*: towards the creation of an industrial platform strain."

Chapter IV.

MOLECULAR TOOLS FOR ENGINEERING *STARMERELLA BOMBICOLA*

IV.1. Introduction

One of the milestones of Biosurf, the SBO-research project to which this work contributed, was to transform *Starmerella bombicola* into a platform organism for the production of interesting biomolecules *e.g.* transforming this sophorolipid producing yeast into a rhamnolipid producing one. To do so, at least nine genes would have to be introduced and expressed in *S. bombicola*, requiring multiple promoters, terminators and integration sites, which will also be needed for the successful expression of other complex heterologous biosynthetic pathways in this yeast. To allow an elegant combinatorial assessment of these and other parameters on gene expression, a suitable reporter system is needed. Furthermore, the availability of an episomal expression system and several (reusable) selection markers would be a great advantage. At the onset of this project only one endogenous promotor (the glyceraldehyde 3-phosphate dehydrogenase promoter (P*gapd*)) (Van Bogaert *et al.*, 2008a) and one heterologous promotor (the viral tyrosine kinase promoter (P*tk*)) (Van Bogaert *et al.*, 2008b) had been used to express one heterologous gene (the hygromycinephosphotransferase) (Van Bogaert *et al.*, 2008b) in *S. bombicola*. In this chapter, the development of a suitable reporter system, along with the development of a reusable marker tool is described. Furthermore, the construction of an episomal vector for *S. bombicola* was investigated.

Conventionally used reporter systems are enzymes such as β-galactosidase (LacZ), chloramphenicol acetyltransferase (Cat) and luciferase (Lux). Their signal amplification is derived from substrate cleavage by the reporter protein. Another widely applied reporter protein is the green fluorescent protein (GFP) of the marine invertebrate *Aequorea victoria* (Chalfie, 1995). GFP is a single auto fluorescent polypeptide with a molecular mass of 26 kDA that emits green light (508 nm) when excited with ultraviolet light (395 nm). Because one GFP molecule represents only one fluorophore, the GFP output signal is smaller than that for enzymatic reporters. However, no cofactors or substrates have to be provided, making GFP an extremely useful marker for gene expression (Chalfie *et al.*, 1994; Cubitt *et al.*, 1995):

GFP fluorescence can be easily and continuously measured by fluorometry, a relatively inexpensive and rapid technique. The development of a reporter system using GFP expression in combination with fluorometry was investigated here, since activity will need to be determined in several strains, under varying conditions and at different stages of growth, which requires an easy and straightforward system.

The development of the reporter system was based upon an integrative GFP expression cassette, as the heterologous pathways will also be integrated in the genome and the parameters thus have to be assessed likewise. Although the use of a stably integrated chromosomal GFP copy allows accurate quantification of GFP expression in individual cells, visualization of a single chromosomally integrated GFP copy requires a bright fluorescent reporter molecule and highly sensitive detection equipment. To attain this, several GFP variants with shifted excitation maxima that fluoresce more intensely than the wild type GFP of *Aequorea victoria* were developed (Cormack *et al.*, 1996; Cormack *et al.*, 1997b). Even codon optimized GFP variants for mammalian cells or organisms with a non-canonical codon usage like the yeast *Candida albicans* are available (Cormack *et al.*, 1997a). Two *gfp* variants, *gfpmut3b* and *yegfp*3, were chosen for validation as reporter protein for *S. bombicola*. They were both placed under control of the endogenous strong and constitutive *gapd* promotor (P*gapd*) (Van Bogaert *et al.*, 2008a) and integrated at the *ura3* locus of *S. bombicola*. The latter allows the assessment of possible limits and drawbacks of the developed *gfp* expression system and furthermore enables the selection of the best of these two *gfp* variants for the construction of a promoter library.

As only two markers and no vectors are available for this yeast, some solutions for these shortcomings were also strived for to further equip the *S. bombicola* molecular toolbox.

IV.2. Materials and methods

IV.2.1. *Strains, plasmids, media and culture conditions*

Wild type *S. bombicola* ATCC 22214, an *ura3*- auxotrofic mutant thereof (strain PT36) and a derivative of strain PT36 knocked out in the gene responsible for the first step of SL biosynthesis (strain *Δura3Δcyp52M1*) and thus defective in SL biosynthesis were used as the parental strains. In the PT mutant the *ura3* coding sequence is removed by homologous recombination with a PT cassette. Hence, *ura3*PT mutants (*ura3*$^-$) are auxotrophic for uracil or uridine and can be transformed back to prototrophy with a functional *ura3* gene. SD medium (0.67 % yeast nitrogen base without amino acids (DIFCO), 2 % glucose and 15 % agar) was used to select for *ura3*$^+$ transformants. Other media and culture conditions can be found in Chapter II All 3 abovementioned strains were available in the strain collection of the Laboratory of Industrial Biotechnology and Biocatalysis. A last yeast strain, the *ura3* negative *S. cerevisiae* strain FY1679-01B (MATa, *Δura3*-52, *Δleu2*, *Δtrp1*, *Δhis3*, *Δgal2*) was obtained from Euroscarf. This strain can be reverted to prototrophy using the *ura3* marker of *S. bombicola* (Van Bogaert *et al.*, 2007a). *Escherichia coli* XL10GOLD ultracompetent cells were used for plasmid maintenance and for all cloning experiments. An *E. coli* strain harbouring an *E. coli* GPF expression cassette, *E. coli* BL21 pCX_P22_*gfp*mut*3b*-histag was a kind gift of dr. ir. Joeri Beauprez. *S. bombicola* and *E. coli* were cultured as described in Chapter II. *E. coli* BL21 pCX_P22_*gfp*mut*3b*-histag was grown on synthetic medium (Waegeman *et al.*, 2011).

Plasmid pGALyEGFPTU which harbours the *yegfp3* variant (Cormack *et al.*, 1997a) was obtained from the LMBP plasmid collection. This GFP variant contains two mutations (S65G and S72A) relative to the wild type GFP and is codon optimized for the yeast *Candida albicans*. Plasmid pCX_P22_*gfpmut3b*, which contains the *gfpmut3b* variant (Cormack *et al.*, 1996), was isolated from the abovementioned *E. coli* strain and only contains two mutations (S65G and S72A) relative to the wild type GFP.

IV.2.2. *Molecular techniques*

IV.2.2.1. *General techniques*

Yeast genomic DNA and plasmid DNA were isolated as described in previous chapters. RNA was extracted using the RNeasy® Plant Mini Kit (Qiagen). This RNA was converted to cDNA

using the QuantiTect RT® Kit (Qiagen). RnaseZAP (Sigma) was used to ensure a RNAse free environment.

All high fidelity PCR reactions (unless otherwise specified) were performed using the *Pfu* high fidelity DNA polymerase (Stratagene) according to the manufacturer's guidelines. For colony PCR and other control reactions, *Taq* DNA polymerase (NEB, Westburg BV) was used. All primers were obtained from Sigma Genosys and designed using the Clone Manager Professional Suite 8 program and can be found in Table IV. 2.

PCR reaction mixtures were cleaned up using the QiAquick PCR Purification Kit (Qiagen) or the column free Sure Clean Plus Kit (Bioline) for PCR fragments larger than 4kb. The latter was also used for clean-up of restriction/ligation mixtures. The Qiaquick Gel extraction kit (Qiagen) was used for DNA purification from gel. All restriction nucleases were obtained from New England Biolabs (NEB, Westburg BV) and restriction digests were performed as specified by the supplier. The pGEM®-T vector (Promega) was used for all cloning experiments. T4 DNA ligase (Fermentas) and the In Fusion Dry Down PCR cloning Kit (Clontech) were used for cloning and ligating. All DNA sequences were determined at Agowa, Germany.

S. bombicola PT mutants were transformed using an adapted version of the lithium acetate method (Gietz and Schiestl, 1995; Van Bogaert *et al.*, 2008b) and an elektroporation protocol described elsewhere (Saerens *et al.*, 2011a). Transformants were selected on SD plates and plated on YDP to check viability. Negative controls were obtained by transforming the cells with an equal amount of sterile MiQ water. Linear DNA for transformation was obtained by standard *Pfu* PCR using the primers P1_FOR_*ura3*v and P32_REV_cassette. *Escherichia coli* cells were transformed as described by (Sambrook and Russell, 2001) and selection occurred on LB agar plates supplemented with ampicillin (100 mg/L).

IV.2.2.2. *Construction of the yegfp3 and gfpmut3b expression cassettes*

Two promoterless platform vectors, each harbouring a different GFP gene, pGEM-T_cassette_*yegfp3* and pGEM-T_cassette_*gfpmut3b* (cfr. Figure. IV.2) were created in way that allows their use as a backbone for the creation of a promotor library. They respectively contain the GFP variants *yegfp3* and *gfpmut3b* flanked at both sides by approximately 1 kb long homologous regions for recombination at the genomic PT locus in *ura3* negative PT mutants. The vectors can be linearized with two unique restriction enzymes of which one

(*Sap*I), cuts just before the ATG 'start' codon of the respective *gfp* variant which allows us to clone promoter fragments in frame with the GFP variant. The second unique cutter can be chosen from a multiple cloning site (MCS3) of which the sequence was added as a non binding extension to primer P5_FOR_*yegfp*_extMCS3 (bold characters in Table IV. 2). This MCS3 (cfr. Figure IV. 2) contains 11 unique restriction sites of which two (*Sap*I and an arbitrary chosen second one) can be used to linearize the vectors.

The platform vector pGEM-T_cassette_*yegfp3* was constructed first. First the 5' UTR of the *ura3* gene (including the *ura3* promotor), the *ura3* coding sequence and *ura3* 3' UTR (including the *ura3* terminator) were amplified from genomic DNA of wild type *S. bombicola* ATCC 22214, using primers P1_FOR_*ura3*v and P2_REV_*ura3*v (PRODUCT1). This first PCR product, PRODUCT1, forms the left flank of the *gfp* expression cassettes, providing the final cassette with the 1000 base pairs needed for homologous recombination in PT mutants of *S. bombicola*, and at the same time providing the cassette with a selection marker. Since the pGEM-T vector was used for cloning, the High Fidelity PCR Master Kit (Roche) was used for amplification. The resulting PCR product (1970 bp) was ligated into pGEM-T using T4 DNA ligase, yielding pGEM-T_P1. This vector still contained a *Sap*I recognition site originally present in pGEM-T, which was removed to make *Sap*I a unique restriction enzyme of MCS3. The latter was done using the Quick Change Site Directed Mutagenesis kit (Stratagene) with primers P7_FOR_QCSapIpGEM-T and P8_REV_QCSapIpGEM-T and the resulting vector was called pGEM-T_P1_QC*Sap*I. The rest of the expression cassette, consisting of two parts, was subsequently cloned in the obtained vector. These two parts form the right flank of the expression cassette, i.e. the 3' UTR of the *ura3* gene amplified from genomic DNA using primers P3_FOR_*ura3*t_ext*yegfp3* and P4_REV_*ura3*t_ext*Not*I (PRODUCT2, 1055 bp) on the one hand and the *yegfp3* variant with the *C. albicans mal*2 terminator amplified from pGAL*yegfp*TU using primers P5_FOR_*yegfp*_extMCS3 and P6_REV_*yegfp*_ext*ura3*t (PRODUCT3, 1255 bp), on the other hand. The multi cloning site MCS3 and a *Not*I restriction site were added to PRODUCT3 and PRODUCT2 respectively by non binding extensions on primers P4 and P5, respectively. PRODUCT3 and PRODUCT2 were subsequently fused by overlap PCR. The necessary overlaps were already added as non binding extensions of primers P3 and P6, Primers, P5_FOR_*yegfp*_extMCS3 and P4_REV_*ura3*t_ext*Not*I were used for the amplification of the fusion product (2275 bp). The fusion PCR product and the vector backbone pGEM-T_P1_QC*Sap*I were both digested using

*Spe*I and *Not*I, purified and ligated. The resulting vector, pGEM-T_cassette_*yegfp*, was one of the two blanco platform vectors used for all further experiments.

The second platform vector, pGEM-T_cassette_*gfpmut3b* was derived from pGEM-T_cassette_*yegfp3* by removing the *yegfp3* variant with the unique restriction enzymes *Sap*I and *Bsr*GI (introduced with P6 in pGEM-T_cassette_*yegfp*) and replacing it with the *gfpmut3b* variant. The latter was amplified from plasmid pCX_P22_*gfp* using primers P15_FOR_*gfpmut3b*_InFu and P16_REV_*gfpmut3b*_InFu. These two primers provide the resulting PCR fragment, PRODUCT 4 (734 bp) with the necessary 15 base pairs for cloning it into the linearized vector, using the In Fusion Dry Down PCR cloning kit.

For the amplification of different lengths of the *gapd* promoter (194 bp, 492 bp and 1560 bp) the respective primers P25_FOR_P*gapd*194, P22_FOR_P*gapd*492 and P19_FOR_P*gapd* 1560 were combined with P18_REV_ P*gapd*. All promoter fragments were thus provided with the fifteen base pairs necessary for cloning with the In Fusion Dry Down PCR Cloning Kit in the *Sap*I-*Ava*I linearized pGEM-T_cassette_*gfpmut3b* and pGEM-T_cassette_*yegfp3* vectors, yielding pGEM-T_P*gapd*194_*yegfp*, pGEM-T_P*gapd*492_*yegfp3* and pGEM-T_ P*gapd*1560_*yegfp3* and the respective *gfpmut3b* variants. The expression cassettes were picked up from the final vectors with primers P1_FOR_cassette and P32_REV_cassette and transformed into *S. bombicola* PT mutants. The promotorless expression cassettes served as controls.

IV.2.2.3. *Construction of an episomal vector for* S. cerevisiae *containing the yegfp3 expression cassette*

The vector pGEM-T_ARS/CEN_P*gapdhygro* (cfr. 2.2.7) contains the ARS1 and CEN4 sequences of *S. cerevisiae* derived from pZ3bt (Branduardi *et al.*, 2004). This vector was used as a backbone for the insertion of the promotorless *yegfp3* expression cassette and the P*gapd*1560_*yegfp3* expression cassette. To do so, the pGEM-T_ARS/CEN_P*gapdhygro* backbone was digested using restriction enzymes *Psp*OMI and *Sbf*I and the largest fragment (4586 bp) was gel purified. The two expression cassettes were derived from the corresponding vectors pGEM-T_cassette_*yegfp3* and pGEM-T_P*gapd_yegfp3* as the largest fragments of a *Sal*I digest (4234 bp and 5773 bp respectively). All gel purified fragments were treated with Mung Bean nuclease (NEB) and the vector backbone was ligated with both expression cassettes yielding pGEM-T_cassette_*yegfp*_ARS/CEN and pGEM-

T_P*gapd*1560_*yegfp*_ARS/CEN. These vectors were transformed in *S. cerevisiae* and transformants were selected on YPD plates supplied with 200 µg/mL hygromycine.

IV.2.2.4. *Construction of* gfpco10g *expression cassettes*

When the genome sequencing results of *S. bombicola* became available, this allowed us to determine the codon usage of this yeast. The averaged codon usage was determined of eight *S. bombicola* homologs of genes described in literature to have a biased codon usage (cfr. Table IV. 1) and the protein sequence of GFPmut3b = yEGFP3 was back-translated using this averaged codon usage. The codon optimised genes were synthesised at GenScript (Piscataway, USA), with the MCS 3 cloning site added at the 5' side of the *gfp* coding sequence and the terminator sequences of the *cyp52M1* gene (Van Bogaert *et al.*, 2009a) added at the 3' side. Because initially the usage of 10 genes was aimed for, this terminology was maintained and the synthetic gene was named *gfpco10genes* or shorter *gfpco10g;* the sequence can be found in the appendices.

Table IV. 1. *S. bombicola* genes with a biased codon usage and their Effective Number of Codons (Nc).

Gene	Corresponding enzyme	(Nc)
tpi	Triose phosphate isomerase	35.12
tef	Transcription elongation factor	31.52
tef2	Transcription elongation factor 2	49.53
pgk	Phosphoglycerate kinase	32.67
gapd	Glyceraldehyde-3-phosphate dehydrogenase	31.09
eno	Enolase	33.99
adh	Alcohol dehydrogenase	54.35
adh2	Alcohol dehydrogenase 2	41.48

The *yegfp3* variant in the pGEM-T_cassette_*yegfp3* platform vector was replaced by the codon optimized *gfpco10g* variant. The abovementioned platform vector was digested with *Kas*I and *Bsr*GI, as was the *gfpco10g* fragment (748 bp) after amplification from the pUC vector using primers P107_FOR_*gfpco10g*_MCS3 and P108_REV_ *gfpco10g_Bsr*GI. The two fragments were ligated and the resulting vector pGEM-T_*gfpco10g*_T*gal_ura3*locus was used as the platform vector for further experiments.

Amplification of the *gapd* promoter with primers P19_FOR_P*gapd*1560 and P110_REV_P*gapd*1560_ext*gfpco10g* (1585 bp) introduced the 15 bp non-binding extensions necessary for placing the amplified promoter in frame with the *gfpco10g* variant of the *Ava*I and *Sap*I digested vector backbone pGEM-T_*gfpco10g*_T*gal_ura3*locus (7209 bp) with the In Fusion Dry Down PCR cloning Kit (Clontech). Exactly the same was done for the promotor

of the *cyp52M1* gene, which was picked up from genomic DNA using primers P76_FOR_P*cyp*_infu and P112_REV_P*cyp*_ext*gfpco10g* (517 bp). The resulting vectors, pGEM-T_P*gapd*1560_*gfpco10g*_T*gal_ura3*locus and pGEM-T_P*cyp_gfp*co10g_T*gal_ura3*locus in addition to the promotorless vector pGEM-T_*gfpco10g*_T*gal_ura3*locus were controlled by sequencing and subsequently used as templates to pick up the respective expression cassettes with primers P1_FOR_cassette and P32_REV_cassette. These were transformed into the *S. bombicola* PT Δ*ura3* strain and PT mutants additionally knocked out in the *cyp52M1* gene (Van Bogaert *et al.*, 2013). Selection occurred on SD plates and correct integration of the cassettes was controlled by performing colony PCR using primer pairs P30_FOR_checkPROMIN / *gapd*up_sp6 (1464 bp), P30_FOR_checkPROMIN / P77_REV_P*cyp*_Infu (744 bp) and P33_FOR_checkcassIN / P120_REV_*gfpco10g* (2258 bp) for upstream integration of the respective expression cassettes and for all three cassettes the primer pair P36_FOR_checkIN / P35_REV_checkcassIN_DOWN (1292 bp) and was used to check downstream integration.

IV.2.2.5. *Construction of an episomal vector for S. bombicola*

Initial experiments to test transformation of the *S. bombicola* wild type strain with an episomal vector were performed with plasmid pZ_3GFP (Branduardi *et al.*, 2004). Selection of the transformants occurred on plates supplemented with G418 (1200 µg/mL) (Sigma). For further experiments, an episomal vector for *S. bombicola* was created by inserting the ARS1/CEN4 sequences of *S. cerevisiae* in a vector harbouring the hygromycine resistance marker under control of the homologous *S. bombicola gapd* promotor (pGEM-T_P*gapd*1560_*hygro*) (Van Bogaert *et al.*, 2008b). The ARS1/CEN4 sequences were picked up from plasmid pZ_3GFP using primerpairs P61_FOR_ARS1 / CEN4_ext*Sal*I and P62_REV_ARS1 / CEN4_ext*Nsi*I (1695 bp) and the resulting PCR fragment was digested with the two restriction enzymes of which the recognition sites were added as non-binding extensions with primers P61 and P62 (*Sal*I and *Nsi*I respectively). The pGEM-T_P*gapd*1560_*hygro* vector was digested with the same restriction enzymes. Both fragments were purified and subsequently ligated, yielding vector pGEM-T_ARS/CEN_P*gapd*1560_*hygro*. This vector was transformed in *S. bombicola* wild type and *S. cerevisiae*. Transformants were selected on YPD plates supplemented with hygromycin (500 µg/mL for the former and 200 µg/mL for the latter). The resulting transformants were grown on selective medium and plasmids were isolated using a modified version of the QIAprep Spin Miniprep protocol (Qiagen), which consisted of lysing the cells using acid

washed glass beads (Saerens *et al.*, 2011c) prior to plasmid extraction as described in the Miniprep protocol. The resulting extractions were transformed in *E. coli* as described above and positive transformants were selected using colony PCR.

Table IV. 2 Primers used for the creation of the *gfp* expression vectors and -cassettes and for the episomal vector. Bold characters represent non-binding extensions. The ATG start codon of *yegfp3* (or *gapd*) is underlined.

Primer	**5'-sequence-3'**
P1_FOR_*ura3*v	AGAACAAGGCCGAGTATGTC
P2_REV_ *ura3*v	TGCCAGCAGATCATCATCAC
P3_FOR_ *ura3*t_ext*yegfp*	**GGATCCCCGCAGGGC**ATGCAACTTGCACATGAATACC
P4_REV_ *ura3*t_ext*Not*I	**TAGCGGCCGC**GTCAGATTAGCCTCCGACATAG
P5_FOR_*yegfp*_extMCS3	**GCACTAGTATACCCGGGCGCCTCAGCTCTTCGATG**TCTAAAGGTGAAGAAT
P6_REV_ *yegfp3* _ext*ura3*t	**TATTCATGTGCAAGTTGCAT**GCCCTGCGGGGATCCATACG
P7_FOR_QCSapIpGEM-t	CGTATTGGGCGCT**C**CTCCGCTTCCTCGCTCACTGACTC
P8_REV_QCSapIpGEM-t	GAGTCAGTGAGCGAGGAAGCGGAG**G**AGCGCCCAATACG
P15_FOR_ *gfpmut3b* _InFu	**CTCAGCTCTTCG**ATGAGTAAAGGAGAAGAACTTTTC
P16_REV_*gfpmut3b* _InFu	**CTGCATTATTTGTAC**AGTTCATCCATGCCATGTG
P18_REV_P*gapd*	**TTCACCTTTAGACAT**TTGTGTAGAGTTGTTTTTG
P19_FOR_ P*gapd* 1560	**CACTAGTATACCCGG**GACATCCGATGTGTAGTTA
P22_FOR_ P*gapd* 492	**CACTACTATACCCGG**TCAAGCACTGCGGCACTC
P25_FOR_ P*gapd* 194	**CACTAGTATACCCGG**CCGTCATTGCAGGGTGTG
P32_REV_cassette	GTCAGATTAGCCTCCGACATAG
P61_FOR_ARSCEN_ext*Sal*I	**GTCGG**TCGACCTTACATTTTATGTTAGCTGGTGGACTG
P62_REV_ARSCEN_ext*Nsi*I	**ACC**ATGCATACTAGTTGGTGTGGAAGTCCATA
P107_FOR_*gfpco10g*_MCS3	CACTAGTATACCCGGGCGCCTCAG
P108_REV_ *gfpco10g* _*Bsr*GI	TTATTTGTACAGCTCGTCCATTCC
P110_REV_P*gapd*1560_*gfpco10g*	**CTCTCCTTTCGACAT**TTGTGTAGAGTTGTTTTTG
P76_FOR_ P*cyp* _infu	**CACTAGTATACCCGG**TGAATATTCGTAGGGAGAAGCTAAATTG
P112_REV_P*cyp*_*gfpco10g*	**CTCTCCTTTCGACAT**ATATGTACTTTTCAATATGATAAACGGAGAAATAAC

IV.2.2.6. *Reuse of the ura3 marker*

Transformants harbouring a *phaC1co* expression cassette (cfr. Chapter VII) were used to determine if flanking the *ura3* gene by two identical repeats could result in homologous recombination between these repeats when placing such strains on medium selective for *ura3* negative strains. Overnight cultures on liquid SD medium (5 mL) were plated on SD medium containing uracil (0.3 g/L), uridine (0.3 g/L) and 5-FOA (1 g/L). These compounds were added by means of filter sterilisation prior to pouring the plates and the plates were stored in the dark, because 5-FOA is light sensitive. Three different volumes (50, 100 and 200 µL) of the overnight cultures were plated on separate selective plates.

IV.2.3. *Sampling and analysis*

IV.2.3.1. *Fluorescence*

Fluorescence was measured using a Spectramax Gemini XS device (Molecular Devices, St. Grégoire, France) with black 96-well plates (Nunc). Fluorescence emission was measured at 511 nm after excitation at 491 nm and was quantified in relative fluorescence units (RFUs). Cultures were grown on YPD or SD medium (50 mL). On regular time intervals, the OD was measured and samples were taken from the shake flasks with a volume to obtain an equal OD of 400 in 2 mL for all cultures. The cells were washed twice with PBS buffer (137 mMNaCl; 2.7 KCl; 10 mM $Na_2HPO_4.2\ H_2O$; 2 mM KH_2PO_4; pH 7.4) and finally the cell pellets were resuspended in 2 mL PBS buffer before measuring fluorescence. Alternatively, cells (brought to an OD value of 100) were lysed using acid washed glass beads as described by (Saerens *et al.*, 2011c) and fluorescence of the crude lysates was measured. In both cases the supernatants were also included in the measurements and the blancs consisted of the lysis buffer, the PBS buffer, the media and strains harbouring the promoterless expression cassettes in addition to the parental strains. PBS was chosen as the buffer, because it was proven to be the best choice for the construction of a promotorbank using GFP (De Mey *et al.*, 2007).

In a second type of experiments, the BioLector (m2p labs) device was also used to measure fluorescence. This device allows the online measurement of fluorescence during growth, without the need to sample the cultures. Both normal 48-well flower plates, allowing determination of biomass formation (via scattered light) and fluorescence and 48-well cell culture plates equipped with pH and pO_2 electrodes (cfr. Figure IV. 1 a), enabling additional online detection of these last two parameters during the orbital shaking process, were used in the experiments. Microplates containing 1 mL of culture medium were covered with gas-permeable membranes for monoseptical operation of the cultures. Humidity was set at 95 %, [O_2 %] at 20.95 % and the shaker frequency was set at 1300. Excitation at 486 nm and emission at 510 nm were used to determine fluorescence.

The last experiments were performed in 24-well growth plates (cfr. Figure IV. 1 b) in which 3 mL cultures were sampled for one week. Samples (200 µL) were pipetted into 96-well clear bottomed microtiter plates (Nunc) and the OD was measured using the FLUOstar OPTIMA device (BMG labtech). For OD values above 1, the samples were diluted with saline solution. The samples (or dilutions) were subsequently pipetted into black 96-well plates (Nunc) and fluorescence was measured using the same device. Excitation occurred at 485 nm and

emission was measured at 510 nm. Controls consisted of the wild type strain, the transformants with the promotorless expression cassettes and the (diluted) culture media.

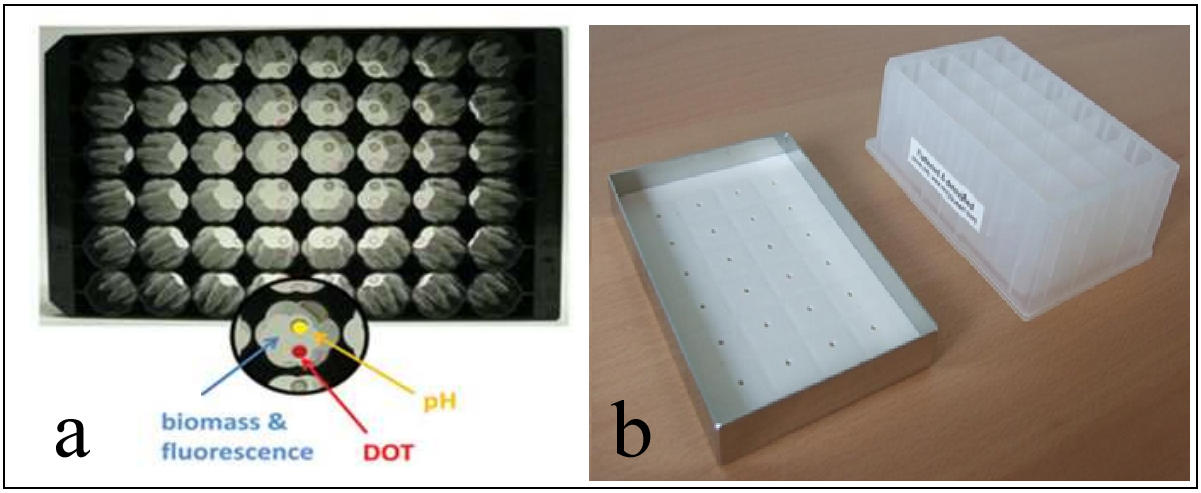


Figure IV. 1 **(a)** BioLector 48 well Flower plate with the BioLector measurement principle **(b)** 24-well culture plates allowing the coculturing of 3 mL cultures (24).

IV.2.3.2. *Assessing the effect of SLs on GFP*

Lysates of *E. coli* BL21 pCX_P22_*gfpmut3b*-histag were used as a source of purified His-tagged GFP protein by performing standard His-Tag purification under native conditions (Aerts *et al.*, 2011). Lysates of such cultures were also used to assess the possible effect of SLs on GFP by adding a dilution series obtaining final SL concentrations of 0, 5, 20 and 65 g/L in an end-volume of 200 µL of *E. coli* lysates or culture solution. For the dilution series, a solution containing a mixture of SLs in water (65 %) was used (Soliance).

IV.2.3.3. *Native PAGE and MSMS*

Protein separation and analysis was done as described by (Brandt *et al.*, 2008). Shortly said: crude lysates of *S. bombicola* transformants and *E. coli* BL21 pCX_P22_*gfpmut3b*-histag were separated on native PAGE (purified GFP was used as a positive control). The gels were scanned using a fluorescence scanner (Pharos FX Plus molecular imager (Biorad)) and subsequently stained by placing the gels overnight in a gently shaken container containing coomassie dye. Fragments in which GFP could possibly be located were subsequently cut from the gels and destained. The resulting protein mixtures were subjected to trypsin digest after which the resulting peptide mixtures were analyzed using the 4800 MALDI TOF/TOF (Applied biosystems). The spectra were analyzed using the search engine Mascot.

IV.3. Results and Discussion

IV.3.1. *Construction and validation of a yegfp3 and gfpmut3b platform*

Two platform vectors for the evaluation of different promoters at the *ura3* locus were developed as described in IV 2.2.3. These vectors allow the cloning of promotor (fragments) exactly in front of the *gfp* variants (cfr. Figure IV. 2): the start codon of the original gene is switched by the start codon of the *gfp* variant.

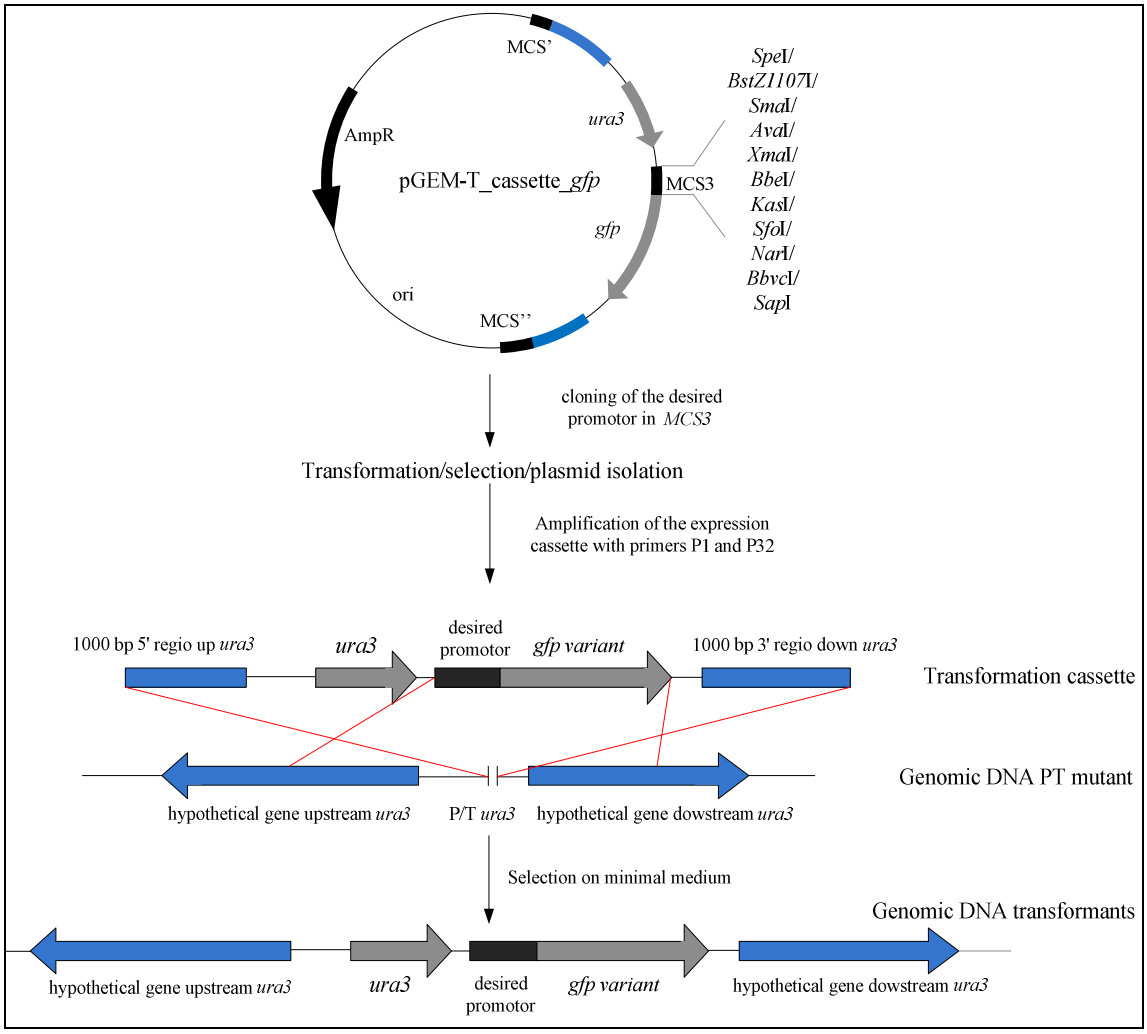


Figure IV. 2 Platform for the construction of a promotor library for *S. bombicola*. The multi cloning site (MCS3) contains 11 unique restriction sites. The last restriction site of MCS3 is the *Sap*I restriction site, which contains the start codon of *gfp* so that it can be used to clone promoter fragments exactly in front of the *gfp* coding region. Gfp expression occurs from a genomically integrated expression cassette at the *ura3* locus of *S. bombicola*.

These two vectors encode an identical protein, but the codon usage of the corresponding genes is different. The first *gfp* variant, *gfpmut3b*, contains two mutations (S65G and S72A) relative to the wild type (wt) *gfp* of *Aequorea Victoria* (Cormack *et al.*, 1996). The second one, *yegfp3*, contains the same mutations, but has additionally been codon optimized for *Candida albicans* (Cormack *et al.*, 1997a). Moreover, the non-canonical codon CTG (abnormally encoding serine in *C. albicans*) was replaced by the TTG codon, to correctly insert a leucine at that position in the protein. Two platform vectors were created to allow for the selection of the best *gfp* variant for the expression platform. The latter was done by evaluating the two platforms using the strong constitutive *gapd* promotor as this promotor had already been used for the expression of a heterologous protein in *S. bombicola* (Van Bogaert *et al.*, 2008a).

For extensive genetic engineering of *S. bombicola* the expression cassettes are preferably as short as possible. Because different fragment lengths of the promoter enable *S. bombicola* cells to grow on hygromycin selective medium (Van Bogaert, 2008), the idea was to immediately assess the activity of three different promotor lengths. Thus, per *gfp* variant, 3 expression cassettes were created: one with the shortest P*gapd* fragment (194 bp) only containing the TATA-box at -78 bp, the second with a 457 bp promoter fragment containing 2 CAAT-sequences located at positions -366 bp and -354 bp besides the TATA-box and the third with the largest promoter fragment consisting of the full intergenic region (1560 bp) between the *gapd* gene and the upstream open reading frame, thus containing the two CAAT-boxes and the TATA-box in addition to other possible upstream regulatory sequences.

The 6 expression cassettes in addition to the 2 promotorless control cassettes were transformed into the PT36 Δ*ura3* mutant of *S. bombicola* and transformants were selected on SD plates. Correct integration and architecture of the expression cassettes in the genomic DNA of the transformants was controlled by performing colony PCR followed by several control PCRs on gDNA of selected transformants (cfr. Figure IV. 3).

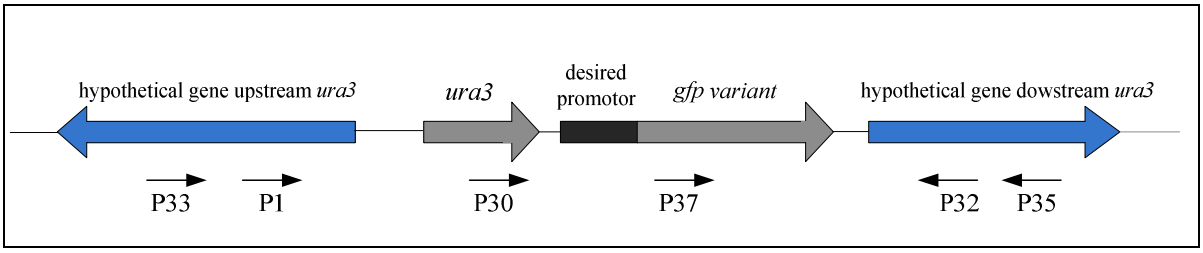


Figure IV. 3 *gfp* expression cassette integrated in the genomical DNA of *S. bombicola* and primers used to determine the correct genotype of the *gfp* transformants. Several combinations of these primers were used to perform control PCR on gDNA of the transformants

All transformants, with exception of the P*gapd*457*_gfpmut3b* expression strain, were obtained and the number-letter codes of the corresponding colonies are, for convenience, also used to designate the *gfp* expression strains in further experiments. The keys for linking those to the corresponding expression cassettes is depicted in Table IV. 3.

Table IV. 3. *S. bombicola gfp* transformants (codes) with their respective *gfp* expression cassettes.

Colony	Expression cassette	Inserted pGAPDpromotor elements
A1	*yegfp3*	None
A5	*gfpmut3b*	None
D2	P*gapd*194*_yegfp3*	TATA-box
D9	P*gapd* 457*_yegfp3*	TATA-, CAAT-,CAAT- boxes
E1	P*gapd* 1560*_yegfp3*	TATA-, CAAT-,CAAT- boxes
F11	P*gapd* 194*_gfpmut3b*	TATA-box
B11	P*gapd* 1560*_gfpmut3b*	TATA-, CAAT-,CAAT- boxes
WT	*S. bombicola* wild type	/

IV.3.1.1. *GFP expression in S. bombicola*

The transformants were grown in YPD medium (50 mL) in shake flasks and the OD was monitored (cfr. Figure IV. 4). The shake flasks were sampled in the mid exponential phase, after 21 hours of cultivation, and towards the beginning of the stationary phase, after 45 hours (cfr. blue arrows in Figure IV. 4). The wild type and the promotorless transformants (A1 and A5) were used as controls. The measured fluorescence values are depicted in Figure IV. 5.

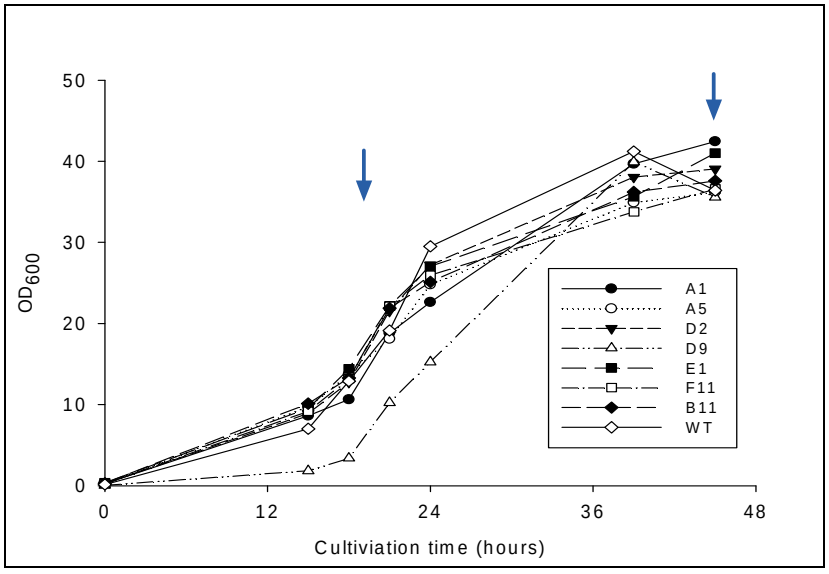


Figure IV. 4 ODs of the transformants and the wild type in function of time. Samples for measurement of fluorescence were taken after 21 and 45 hours of cultivation respectively (represented by blue arrows in the figure).

Based on their growth, the transformants behave similar both to each other and to the wild type (except for strain D9 (P*gapd*457_*yegfp3*)). Unfortunately, no fluorescence was detected in the cells nor in the supernatant after 21 or 45 hours of cultivation (cfr. Figure IV. 5 left and right). Only the values of the samples taken after 45 hours of cultivation are depicted, as the situation was identical for the samples after 21 hours of cultivation. Some background was obtained from the buffer, but especially the cells and YPD cause background fluorescence resulting in equal RFUs for the wild type, promotorless transformants and *gfp* expression strains.

These results were somewhat disturbing as completely no fluorescence could be detected above the background for biomass. A second growth experiment was started with the aim of lysing the cells prior to fluorescence measurements. This was done to assess if some interference with the cell wall of the yeast cells occurred, which could explain the absence of fluorescence. The experiment was thus repeated, and cells were lysed after 21 and 39 hours of cultivation. GFP fluorescence was measured in crude lysates; controls consisted of the lysis buffer and the culture medium. Lysing the cells gave rise to similar results as the ones depicted in Figure IV. 5 and so lysing the cells clearly did not solve the problem.

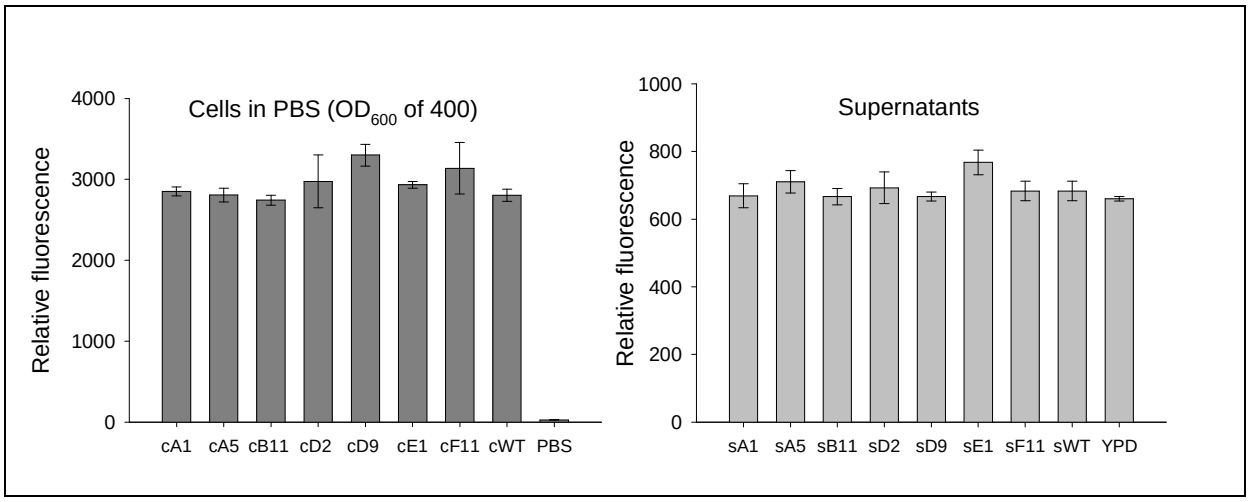


Figure IV. 5 Fluorescence of concentrated cells in PBS (OD of 400) and the corresponding supernatants after 48 hours of cultivation. Controls consisted of the sterile culture medium (YPD) and PBS buffer. Strains A1 and A5 correspond to the promotorless *gfp* expression strains.

Because the expression cassettes are integrated at the *ura3* locus and this gene does not need to be expressed on YPD medium (uracil is present in the medium), another hypothesis was that the *ura3* locus could be transcriptionally inactive on YPD and that expression of the adjacent *gfp* would thus not occur either. Although unlikely, this can easily be tested by

performing a growth experiment on SD medium. As no amino acids are present in this medium, the *ura3* gene has to be expressed for cells to be viable and in that case *gfp* expression should also occur. Unfortunately, no fluorescence could be detected for such SD grown cells either. Not after 21 hours of cultivation nor after 45 hours of cultivation.

Other factors that could influence GFP formation and/or fluorescence are pH, temperature and oxygenation. The pH of *S. bombicola* cultures drops to about 3.5, but the intracellular pH is considered to remain quite constant around 7. Moreover, GFP has already been successfully expressed in several other yeasts and fungi (Atkins and Izant, 1995; Spellig *et al.*, 1996; Neuveglise *et al.*, 1998; Staib *et al.*, 2000; Branduardi *et al.*, 2004) and acidification of the culture medium is a general characteristic for yeasts. Furthermore, GFP was reported to denature reversibly at lower pH values to re-nature at higher pH values. Since the pH of the PBS buffer was set at 7.4 and measurement of the concentrated cell solutions in PBS buffer after a few or 24 hours of incubation did not either result in fluorescence, it can be concluded that the pH was probably not the problem.

Another cause of the negative results could be the presence of sophorolipids, detergent like molecules, which could have an adverse effect on the (folding of) the heterologous GFP protein. Lysates of *E. coli* cells expressing GFP were thus incubated with several concentrations of SLs (up to 65 g/L final concentration), after which the fluorescence of the mixtures was determined. Although some inhibitory effects were detected for the highest concentrations of SLs, this could absolutely not account for the complete absence of fluorescence in the *gfp* transformants. Moreover, SLs are not produced when cultivating *S. bombicola* on YPD, at least not in detectable amounts, and intracellular amounts are never expected to be that high. To completely rule out this option, the expression cassettes were transformed in a Δ*ura*3Δ*cyp52M1* mutant of *S. bombicola* (not producing any SLs). Again, no fluorescence could be detected, ruling out SLs as possible reason for the failure of *gfp* expression. Finally, sequence errors in the expression cassettes were ruled out by sequencing amplified PCR fragments from gDNA of the transformants (cfr. IV.3.1.1.) containing the complete expression cassette (amplified with primer pair P33/P35 (cfr. Figure IV. 3). The problem thus had to be situated at another level and additional experiments were thus necessary.

IV.3.1.2. *GFP expression in S. cerevisiae*

Because expression of the *yegfp3* variant was proven to give rise to fluorescent *S. cerevisiae* cells (Cormack *et al.*, 1997a) and the *S. bombicola ura3* marker was proven to be able to complement a *ura3* negative *S. cerevisiae* strain, the P*gapd*1560*_yegfp3* expression cassette was cloned into an *S. cerevisiae* episomal vector. A vector harbouring the promotorless *yegfp3* cassette was also created to serve as a control. Both vectors were transformed into *S. cerevisiae* and the transformants were selected on SD plates. Positive transformants were selected by colony PCR. The keys for the *S. cerevisiae* and *S. bombicola* strains are depicted in Table. IV. 4.

Table IV. 4 *S. bombicola* and *S. cerevisiae* transformants with respective *gfp* expression cassettes.

Colony	Expression cassette	Inserted pGAPDpromotor elements
A1	*S. bombicola*_cassette_*yegfp3*	None
E1	*S. bombicola*_casssette_P*gapd*1560_*yegfp3*	TATA-, CAAT-,CAAT- boxes
A4	*S. cerevisiae*_casssette_*yegfp3*	None
H9	*S. cerevisiae* _casssette_P*gapd*1560_ *yegfp3*	TATA-, CAAT-,CAAT- boxes

These *S. cerevisiae* transformants were grown on SD medium in a BioLector device in parallel with the corresponding *S. bombicola* transformants harbouring the exact same *gfp* expression cassettes integrated in the genome. These strains were all evaluated in triplicate. Because biomass gives rise to a certain fluorescence background, the obtained values for fluorescence were divided by those obtained from biomass. The results are depicted in Figure IV. 6.

Surprisingly, exactly the same expression cassette that does not lead to fluorescence in *S. bombicola* (strain E1) does give rise to GFP expression in *S. cerevisiae* (strain H9). No fluorescence is observed with the promotorless expression cassettes. Episomal expression of *yegfp3* under control of the *S. bombicola gapd* promotor in *S. cerevisiae* results in maximum fluorescence after 44 hours of incubation followed by a steep decrease of fluorescence, which might be a result of protein degradation in combination with a decrease or cease in GFP expression. After 60 hours of cultivation no fluorescence can be detected anymore.

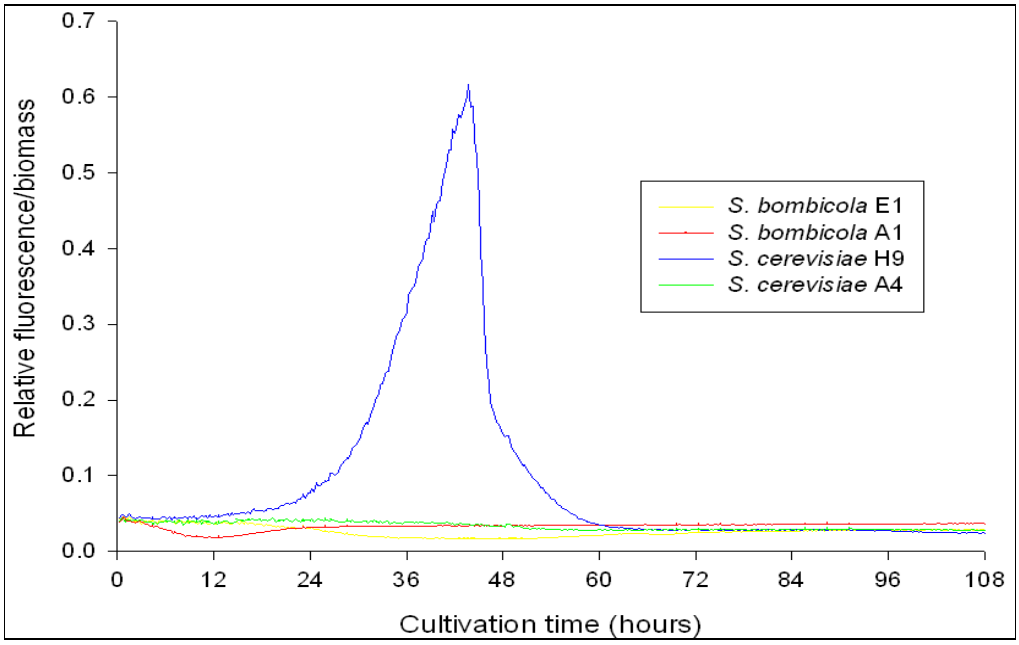


Figure IV. 6 Relative fluorescence (corrected for biomass) in function of cultivation time (hours) for *S. bombicola* and *S. cerevisiae* harbouring identical expression cassettes. Results are the average of three independent experiments (wells in the BioLector device).

These results indicate that nothing is wrong with the expression cassettes, but that the problem is thus most probably linked with expression thereof in *S. bombicola* and more specifically with transcription and/or translation of *gfp* in *S. bombicola*. These assumptions were confirmed by controlling expression of *gfp* at the RNA and protein level. Whereas transcription of the gene could be demonstrated for all strains by extracting RNA and performing PCR on the derived cDNA, no GFP protein was detected after MS/MS analysis of protein extracts of lysates of the E1 transformant (results not shown) nor were fluorescent bands detected when analysing the native page gels with a fluorescence scanner, in contrast to an *E. coli* strain expressing GFP which gave rise to fluorescent bands and the detection of GFP peptides on MS/MS (not shown).

IV.3.2. *Construction and validation of a codon optimised gfpco10g platform*

As the problem seemed to be situated at the translation level, a codon optimised *gfp* was constructed. The averaged codon usage of the *S. bombicola* homologs of eight genes (cfr. Table IV. 1) described in literature to be highly expressed in yeast due to codon bias, was determined and the protein sequence of *gfpmut3b* or *yegfp3* (idem) was back-translated using this averaged codon usage. The platform vector described in IV.3.1 was subsequently

equipped with this codon optimised *gfpco10genes* or shorter *gfpco10g* by replacing the coding sequence of *yegfp3* by that of *gfpco10g* as described in IV.2.2.5, thus creating a new platform vector.

The largest part (1555 bp) of the *gapd* promoter was subsequently cloned in front of the *gfpco10g* coding sequence to assess if this time the system could be used for expression experiments. The expression cassette and control cassette without promotor were picked up from the respective vectors with primerpair P1 / P32 and transformed in the *S. bombicola ΔuraΔcyp52M1* strain giving rise to transformant E1'' and A1'' respectively (cfr. keys in Table IV. 5).

Surprisingly very bright fluorescence was now detected for the E1'' transformant, whereas no fluorescence was detected for the A1'' control strain harbouring the promotorless *gfpco10g* expression cassette (cfr. Figure IV. 7). Suboptimal codon usage thus seemed to completely account for the lack of fluoresence in transformants equipped with the *yegfp3* and *gfpmut3b* genes. The problem could be resolved by simply adapting the codon usage, which makes this is a very important finding to be kept in mind for future genetic engineering of *S. bombicola*. A suitable reporter system is thus finally available.

This platform vector can now be used to construct a promotor library spanning various promotor strengths and -activities (inducible versus constitutive). The selected promotor sequences simply have to be amplified and cloned into the developed *gfpco10g* platform vector. GFP expression experiments with the thereof derived *S. bombicola gfp* expression strains as described in IV.3.1.2 can subsequently be performed and a chart of the respective promotor activities can be composed. The initial aim to test different promotor lengths of the P*gapd* as elaborated for the *yegfp3*and *gfpmut3b* variants (cfr. IV. 3.1) should also be excecuted as this can potentially significantly shorten the expression cassettes. Due to the time loss associated with the lack of fluorescence with the first two *gfp* variants, the construction of such library wasn't possible anymore within the time frame of this work, but will be continued in the future. However, this reporter system also provides the opportunity to investigate the biotechnological opportunities offered by the SL biosynthetic gene cluster which will be discussed in IV.3.2.1.

IV.3.2.1. *Using the gfpco10g reporter system to investigate regulation of the SL gene cluster*

Since the idea of genetic engineering of this host focuses on the heterologous production of (tailor-made) glycolipids, making use of the highly efficient SL biosynthetic pathway, the promoters, terminators and the corresponding genomic integration sites thereof should be studied in more detail.

The *cyp52M1* promotor was chosen for the first investigations as this gene is responsible for the first step of SL biosynthesis. The *cyp52M1* upstream region contains one putative TATA-box (TATATATT) corresponding to the consensus sequence described by (Basehoar *et al.*, 2004) (TATA(A/T)A(A/T)(A/G) at position -75. All strains that were constructed in the time course of this work are summarized in Table IV. 5. These were all constructed for integration at the *ura3* locus (*ura3i*). A strain defective in SL biosynthesis (Δ*ura3*Δ*cyp52M1*) was included as parental strains to assess the possible effects of SL presence/absence on the transcriptional activity of the promoter and/or on the reporter protein. GFP expression experiments were performed with these strains and the results are depicted in Figure IV. 7.

Table IV. 5 *S. bombicola* transformants with respective *gfp* expression cassettes.

Strain	Inserted expression cassette	Parental strain	Inserted promotor elements
E1’	P*gapd*1560*_yegfp3_*T*gal_ura3i*	Δ*ura3*Δ*cyp52M1*	P*gapd*1560 (TATA & 2x CAAT)
A1’’	*gfpco10g_* T*gal_ura3i*	Δ*ura3*Δ*cyp52M1*	None
E1’’	P*gapd*1560*_gfpco10g_*T*gal_ura3i*	Δ*ura3*Δ*cyp52M1*	P*gapd*1560 (TATA & 2x CAAT)
E1’’’	P*gapd*1560*_gfpco10g_*T*gal_ura3i*	Δ*ura3*	P*gapd*1560 (TATA & 2x CAAT)
I1’’	P*cyp_gfpco10g_* T*gal_ura3i*	Δ*ura3*Δ*cyp52M1*	P*cyp* (TATA)
I1’’’	P*cyp_gfpco10g_* T*gal_ura3i*	Δ*ura3*	P*cyp* (TATA)

The activity of the *cyp52M1* promotor at the *ura3* locus was determined by placing the *gfpco10g* reporter protein under its control with integration of the expression cassette at the *ura3* locus (=*ura3*i, strain I1’’). Expression was determined on the SL production medium and the E1’’ strain, harbouring the same expression cassette under control of the P*gapd,* and the original E1’ transformant (*yegfp3*) (cfr. IV 3.1) were cultivated in parallel.

The pronounced effect of the codon usage on *gfp* expression was confirmed: whereas no fluorescence is detected for E1’, a bright signal is detected for E1’’, with both transformants containing exactly the same expression cassette with the exception of the *gfp* coding region.

The result obtained for strain I1'' is somewhat surprising: although high GFP fluorescence was expected in the stationary phase, relatively low fluorescence was detected, starting already in the exponential growth phase. High SL productivity in the stationary phase, was expected to partly result from high activity of the promoters of the gene cluster in this phase. However, the promoter activity of the gene of the first step in SL biosynthesis (P*cyp52M1*) seemed to be very low as compared to the P*gapd* activity. It could be argued that the absence of SLs is responsible for the lack of strong activation of the *cyp52M1* promotor. However, when transforming the expression cassette in the *Δura3* strain (I1'''), a similar result was obtained (not shown). Another reason that could be argued is that it is the combination of the promoter, terminator and integration site that lies at the basis of efficient gene expression. This can be further investigated by integration of the *gfpco10g* expression system under control of the P*cyp52M1* and T*cyp52M1* at the *cyp52M1* locus. Different combinations of these parameters can furthermore be evaluated. The strategies to do so were elaborated and the construction of such expression cassettes was initiated, but unfortunately the evaluation thereof also wasn't possible anymore in the timeframe of this work and will also be the subject of future research.

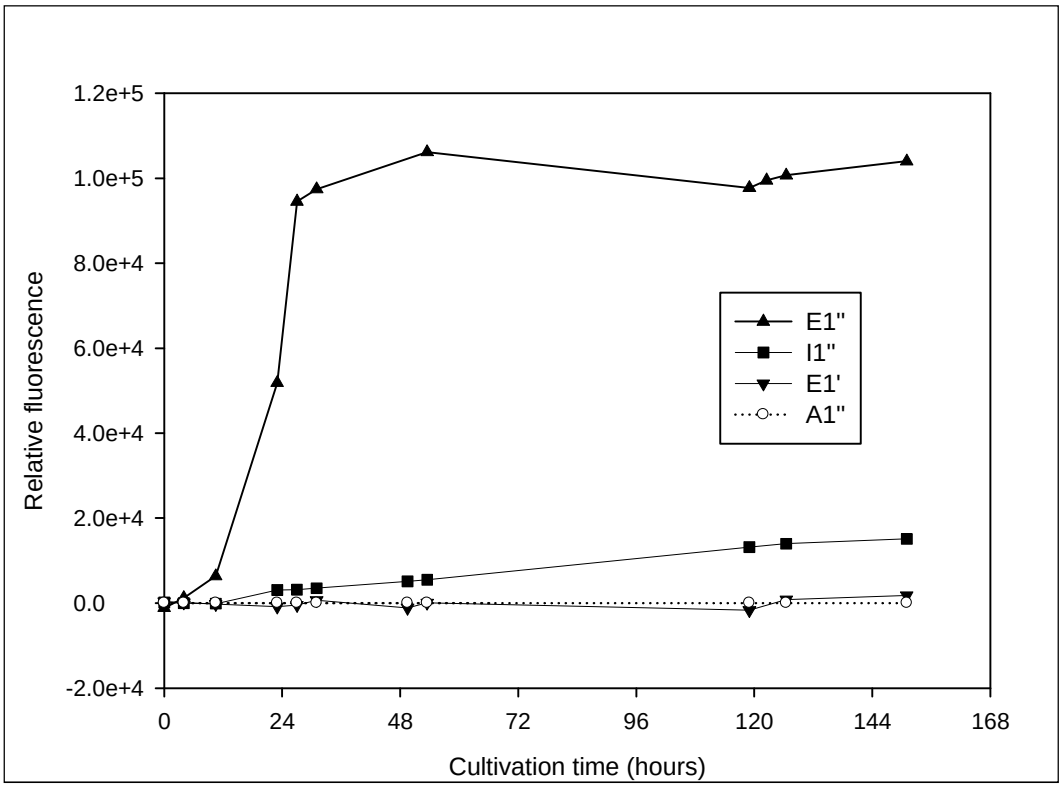


Figure IV. 7 Relative fluorescence in function of cultivation time (hours) for several *S. bombicola* transformants cultivated on the SL production medium. Keys can be found in Table IV. 5. The values obtained for the control strain (A1") were subtracted from all the measured fluorescence values including those from A1'' itself.

Recent RNA sequencing data enabled the comparison between expression of the *gapd* gene and the *cyp52M1* gene in the stationary and exponential growth phase respectively. These data confirmed the results obtained in this chapter. Gene expression of *gapd* was at least 7 times higher than that of *cyp52M1* in the stationary growth phase on production medium. As expression of the genes of the SL biosynthetic pathway was already considered to be high, the P*gapd* of *S. bombicola* can be considered as a very strong promoter. This again confirms that the developed reporter system represents a very reliable molecular tool useful for further research.

Repeating this experiment on YPD medium furthermore resulted in similar high promoter activity (high fluorescence) for the strains containing the *gfpco10g* gene under control of the P*gapd*, whereas absolutely no fluorescence was detected for the strains containing the *gfpco10g* under control of the P*cyp52M1*. The absence of SL biosynthesis on this medium can thus be explained by the absence of gene expression of the genes of the SL biosynthetic pathway on this medium. This is probably caused by the low (2 %) glucose concentrations in this medium as compared to those of the production medium (12 %), but this will be further discussed in Chapter VII of this manuscript.

IV.3.3. *Creation of an episomal vector for Starmerella bombicola*

The abovementioned efforts to create a reporter system for expression studies would have been a lot less laborious if an episomal vector for *S. bombicola* would have been available. Such vector is considered to be an indispensible tool for the efficient genetic engineering of any micro-organism.

As no plasmids were found in this yeast, an attempt was done to create an episomal vector that could be used in further experiments. The pZ_3GFP plasmid (Branduardi *et al.*, 2004), available at our lab, contains the ARS1 and CEN4 sequences of *S. cerevisiae* which enable plasmid maintenance in *S. cerevisiae* and in the yeast *Zygosaccharomyces bailii* (Branduardi *et al.*, 2004). The vector also contains a kanamycin resistance cassette, which allows G418 based selection of yeast transformants and an ampicilin resistance marker for selection of transformed *E. coli* cells. However, *S. bombicola* is highly resistant to G418 (Van Bogaert, 2008) and very high concentrations of G418 antibiotic thus have to be used to maintain the selection pressure with this plasmid. The contained ARS1/CEN4 sequences were thus used to create a new plasmid based on hygromycine resistance in *S. bombicola*, which is a more

suitable antibiotic for *S. bombicola*, although concentrations of 500 µg/mL are still necessary for selection. The created vector, pGEM-T_ARS1/CEN4_P*gapd*1560*hygro*, was transformed in *S. bombicola* and in *S. cerevisiae* as a positive control and transformants were selected on selective plates containing hygromycine. For *S. cerevisiae*, 28 colonies appeared on the selective plates, whereas 2 colonies appeared for *S. bombicola*. Colony PCR showed that 22 of the 28 *S. cerevisiae* colonies contained the plasmid, as did both *S. bombicola* colonies. The positive colonies were subcultivated 5 times on YPD medium supplemented with hygromycine. Plasmid isolation of the cultures of the last generation was followed by subsequent transformation of these extracted plasmids in *E. coli* to assess if we were dealing with stable episomal vectors and not chromosomally integrated parts thereof. The transformations with preps originating from *S. cerevisiae* resulted in the formation of *E. coli* colonies and colony PCR gave rise to the correct fragment (884 bp) for all the tested colonies, so they all contained the episomal vector derived from *S. cerevisiae*. This was in contrast to the preps originating from *S. bombicola* for which no *E. coli* colonies appeared. The pGEM-T_ARS/CEN_P*gapd*hygro vector (or part of it) was thus probably inserted in the genome of *S. bombicola* by illigimate recombination, conferring the cells with resistance to hygromycine. This experiment was repeated with identical results. It can hence be decided that the autonomous replicating sequences (ARS1) in combination with the centromeric (CEN4) sequences are unfortunately not suited for plasmid maintanance in *S. bombicola*.

IV.3.4. *Reuse of the ura3 marker*

Only two markers are currently available for selection of *S. bombicola* transformants: the auxotrophic *ura3* marker (Van Bogaert *et al.*, 2007a) and the hygromycine resistance marker (Van Bogaert *et al.*, 2008a). New selection markers are in the pipeline, but in the meantime a strategy was developed to reuse the *ura3* marker. Reuse of a marker is only possible if positive and negative selection is possible, which holds true for the auxotrophic *ura3* marker. *Ura3* negative strains will not be able to survive on medium without the addition of uracil, so selection of transformants resulting from the transformation of a functional *ura3* gene in such strains can occur on SD plates without uracil. These *ura3* positive strains are on the other hand sensitive to 5'-FOA (5'-fluoroorotic acid), which is converted to a toxic compound by the *ura3* gene product. *Ura3* negative strains can hence be selected on plates containing 5'-FOA. Uracil (and uridine) also have to supplied, as the *ura3* negative strains are not be viable on media w/o uracil.

In Chapter VII the development of a *phaC1co* expression strain is described. This strain contains the *ura3* marker flanked by two identical repeats (cfr. Figure IV. 8). The sequence of this repeat corresponds to the terminator of the *cyp52M1* gene (T_{cyp} = 196 bp). Possible intramolecular recombination between the two repeated sequences, resulting in an *ura3* negative strain, which can be transformed back to prototrophy, was assessed by plating the *phaC1co* transformant on selective medium for *ura3* negative strains.

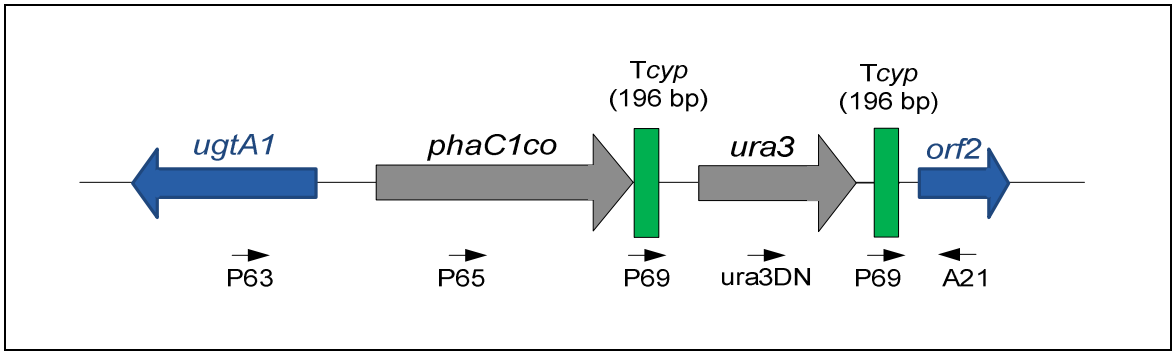


Figure IV. 8 Architecture of *phaC1co* transformants (cfr. Chapter VII) at the *cyp52M1* locus of the SL gene cluster. Two identical sequences (T_{cyp}), represented as blue boxes, are present in the genome of these transformants. Binding sites for used primers are also depicted in the figure.

Several volumes of an overnight liquid culture on SD medium of the A8 *phaC1co* expression strain were plated on selective plates as described in IV.2.2.6. Volumes of 50 and 100 µL gave rise to separate colonies. A first colony PCR was performed after 5 days of incubation, a second one after 7 days of incubation. Two primerpairs were used to select transformants where homologous recombination between the two T*cyp* regions (cfr. Figure IV. 9) had occurred. These were P65 / A21 and P69 / A21. These two PCRs allow for the selection of recombinants resulting from homologous recombination between the two T*cyp* regions. The fragments obtained with these primer pairs both for the parental strain (A8) as for the recombinants, are depicted in Table IV. 3

Table IV. 3 Primerpairs used to check for recombination between the two T*cyp* regions of the *phaC1co* expression cassette.

Primerpair	Fragments for the parental strain (bp)	Fragments for the recombinants (bp)
P65 / A21	3239	1472
P69 / A21	3181 and 707	707
*ura3*DN / A21	1193	/

Colonies displaying the correct fragments for the first two colony PCRs were subsequently controlled again for the absence of the *ura3* marker with a third primerpair (ura3DN / A21). The parental strain was used as a positive control for this last PCR. Three of 11 correct colonies were grown on YPD medium and subjected to genomic DNA extraction. A PCR fragment containing the complete *phaC1co* expression cassette was subsequently amplified from the extracted gDNA with primerpair P63 / A21 and controlled by sequencing. If no errors are present, this *ura3* negative *phaC1co* expression strain can be used for further genetic engineering of this strain again by using the *ura3* marker. Moreover, if so, the developed method will be considered feasible for reuse of the *ura3* marker, which will facilitate the further genetic engineering of *S. bombicola* until more markers become available. The sequencing results indeed confirmed the results obtained by colony PCR: homologous recombination between the two T*cyp* regions had occurred for all three transformants, which can thus also expected to be true for the other transformants.

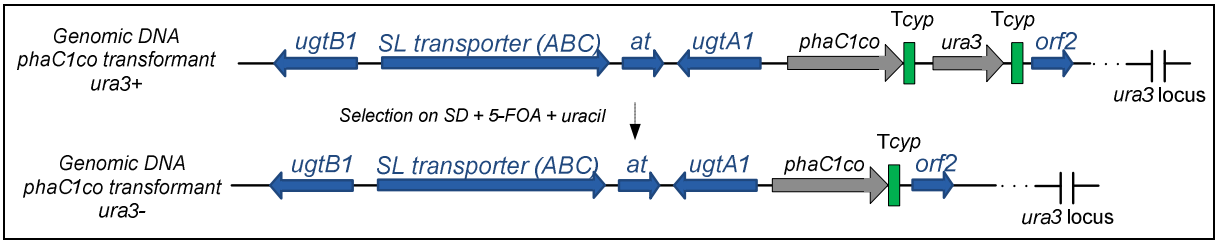


Figure IV. 9 Induced homologous recombination between two repeated sequences enabling the reuse of the *ura3* marker. The depicted example is a *phaC1co* expression strain that will be discussed in Chapter VII of this work.

A total of 96 colonies were tested of which 11 gave rise to the correct fenotypes after all three control PCRs. This corresponds to at least 11.5 % of the totally tested transformants and probably more, as several colonies did not give rise to a PCR fragment at all (due to failure of the PCR reaction). The developed method is thus considered to be feasible for reusing the *ura3* marker and is also considered to be very useful until new (auxotrofic) markers become available.

IV.4. Conclusion

The aim of this chapter was to develop certain molecular tools necessary for enhanced genetic engineering of *S. bombicola*. At the start of this work only two markers, two promoters and no episomal vectors were available for genetic modificiation of this industrially important yeast.

The first aim was thus to construct a promotor library using green fluorescent protein as a reporter system. Two variants, *yegfp3* and *gfpmut3b*, were evaluated for use in further expression studies using several fragments of the *gapd* promotor. However, no fluorescence could be detected for any of the resulting strains. Several approaches were elaborated to solve this, but they all failed. Yet, transformation of an episomal vector, harbouring the constructed *yegfp3* expression cassette under control of the *S. bombicola gapd* promotor in *S. cerevisiae* resulted in fluorescent *S. cerevisiae* cells. This clearly demonstrated that nothing was wrong with the expression cassettes.

A codon optimized version of *gfp*: *gfpco10g* was subsequently obtained and placed under control of the *gapd* promotor. Strains transformed with this expression cassette gave rise to a clear fluorescent signal; hence, suboptimal codon usage seemed to account for the complete lack of fluoresence in transformants equipped with the *yegfp3* and *gfpmut3b* genes. The problem could be resolved by simply changing the codon usage, which makes this is a very important finding to be kept in mind for future genetic engineering of *S. bombicola*. The system was validated using two endogenous promoters; P*cyp52M1* and P*gapd* and it can be concluded that the developed reporter system is suitable for assessing the activity of a wide range of promoters.

Now that a suitable reporter system is available, this should enable the construction of a promotor library. For efficient production of heterologous proteins/pathways in *S. bombicola*, a large variety of heterologous and homologous, either constitutive or inducible yeast promoters spanning the full range of activities should be available. Now that a reliable reporter system is available, a set of interesting endogenous and/or heterologous promoters can be selected and cloned into the platform vector. Subsequent *gfp* expression experiments with the respectively derived *S. bombicola* transformants will allow the creation of a chart of promotoractivities which can be used for further genetic engineering purposes.

Not only the promotor, but the influence of other parameters like the used terminator and the site of integration can be investigated with this reporter system. The aim for the future should be to assess the separate and combined effects of these factors with the developed *gfp* reporter system. Moreover, the regulation of the important SL biosynthetic gene cluster can be investigated using this reporter protein. This could give us some further insight in the influence of important parameters like the culture medium and environmental factors like pH and temperature on the biosynthetic gene cluster. The system can furthermore also be used to

investigate possible regulatory mechanisms like *e.g*. the proposed telomere positioning effect (TPE) of the SL gene cluster, which was described in Chapter I and Chapter II of this manuscript. Shortly said, placing the reporter protein within the SL cluster region and comparing *gfp* expression from a certain promoter with *gfp* expression elsewhere in the genome, using the same promoter, is expected to provide some initial insight in this.

The efforts to create some kind of reporter system would have been a lot less laborious if an episomal vector for *S. bombicola* would be available. An episomal vector could furthermore be of extreme use for the further genetic engineering of *S. bombicola* as it enables fast and easy tests, prior to integration of interesting functionalities in the *S. bombicola* genome. Unfortunately the *S. cerevisiae* autonomous replicating sequences (ARS1) in combination with the centromeric (CEN4) sequences are not suitable for plasmid replication/stability in *S. bombicola*. The 2µ plasmid of *S. cerevisiae* was also evaluated in the past as were the ARS1 sequences of *Y. lipolytica*, which also both failed. One will thus probably have to resort to the construction of genomic libraries in order to retrieve endogenous *S. bombicola* ARS sequences, which can subsequently be used for the development of an episomal vector. Centromeric sequences (CEN) can be searched for by means of synteny by comparing the genes adjacent to CEN sequences in other yeasts. However, this only works if the yeasts are closely enough related to each other.

Because only two markers (the auxotrophic *ura3* marker and the antibiotic hygromycine resistance marker) were available at the onset of this project, the reuse of the *ura3* marker was investigated. It was shown that intra-molecular homologous recombination can be triggered by flanking the *ura3* auxotrophic marker with two identical fragments of -in this case- 196 bp and placing transformants on selective medium for *ura3* negative strains. This method can thus be used to recycle the *ura3* marker several times and enables the combination of several genetic modifications in one strain. This tool will be indispensable for the further extensive genetic engineering of *S. bombicola* until more (auxotrophic) markers for *S. bombicola* will become available.

In this chapter the available molecular toolkit for *S. bombicola* has been further equiped, or at least the foundations therefore have been laid. Although a lot of progress has been made and important information has been acquired, a lot of work is still needed to transform *S. bombicola* into an easy ‘engineerable’ yeast.

CHAPTER V:

DEVELOPMENT OF A QPCR PLATFORM FOR *STARMERELLA BOMBICOLA*

Chapter V.

DEVELOPMENT OF A QPCR PLATFORM FOR *STARMERELLA BOMBICOLA*

V.1. Introduction

As was mentioned before, one of the aims of the project to which this research contributed was to genetically engineer *S. bombicola* with the aim of creating a platform organism for the production of tailor-made glycolipids and to make use of the highly productive SL machinery to do so. At least initially, as not enough promoters and sites of integration have been described or investigated and the SL biosynthetic gene cluster thus offers a solution for both problems. However, lots of parameters influence SL biosynthesis (cfr. Chapter I) and nothing is known about negative and/or positive feedback of the products themselves and/or intermediates on regulation of the pathway. Regulatory effects excerted on different levels can be expected, but this chapter strives for the investigation of effects on the transcriptome level. Therefore, the development of a qPCR platform enabling the investigation of expression of the SL biosynthetic gene cluster under varying conditions was aimed for.

Whereas qPCR is considered as a reliable method to compare gene expression, the normalization procedure to correct sample-to-sample variation remains a critical and challenging problem (Teste *et al.*, 2009). The most common practice is to normalize either to total RNA amount, ribosomal RNA or to a single internal reference gene termed HouSeKeeping gene (HSK). Several mathematical models have been developed that calculate the "relative" mRNA expression changes of a target gene with respect to a HSK. The "2ΔΔCt" approach is the most popular application in quantitative RT-PCR, but it assumes optimal and identical PCR efficiencies of target and reference genes. Violation of this rule results in a systematic bias that either underestimates or overestimates the initial copy numbers. This problem can be bypassed by adjusting for PCR efficiency. However, accurate relative quantification implies that the expression of the reference gene is perfectly stable in the sample set. In the past the most commonly used normalization strategy involved standardization to a single presumed constitutively expressed control gene. However, in recent years, it has become clear that no single gene is constitutively expressed under all experimental conditions. Since even small variations of an internal control could lead to non-

reliable expression data, it is critical to validate that the expression of reference genes is stable, prior to their use for normalization in qPCR analysis. To overcome the "circular problem" of evaluating the expression stability of a candidate gene if no reliable measure is available to normalize the candidate, (Vandesompele *et al.*, 2002) developed a statistical algorithm termed geNorm. Their strategy relied on (i) a careful selection of a set of genes that display minimal variation across different biological conditions, and (ii) normalization of the genes of interest to the geometric mean of a minimal, albeit optimal number of the selected genes.

Therefore , he purpose of this chapter was to test a carefully selected set of potential HSKs to be used for normalization of qPCR data from samples derived of a range of conditions. This was done for the *S. bombicola* wild type, but an auxotrophic lysine mutant (Δ*lys*) currently used for proteomic experiments (SILAC) on *S. bombicola*, was also included to enable validation of proteomic results with qPCR. The stability of 20 candidate housekeeping genes was investigated in a representative set of samples derived from varying conditions. Eleven of these potential HSKs were selected from literature, the remaining nine were found to be stably expressed in *S. bombicola* on the protein level between the exponential and stationary growth phase, which indicates that their gene expression should also be stable.

To validate the system as a tool for controlling gene expression in (transformants) of *S. bombicola*, the most stable subset of internal controls, which gave rise to a robust normalization factor, was applied to investigate expression of the genes of the SL gene cluster.

V.2. Materials and methods

V.2.1. *Strains, media and culture conditions*

Wild type *S. bombicola* ATCC 22214 and an auxotrofic mutant (Δ*lys*) of the former were used for the HSK validation experiments. The last strain was included, because the Δ*lys* strain is used for performing proteomic experiments (SILAC), and including it in this assessment will allow the usage of the qPCR expression platform for the validation of proteomic data obtained for this strain. Validation of qPCR as a tool for the evaluation of gene-expression in transformants of *S. bombicola* was performed for the abovementioned wild type.

The production medium described in previous chapters was used to perform growth experiments as was an adapted version of the former containing only 1 % of glucose (instead of 12 %). The addition of rapeseed oil is one of the variable parameters that could have an influence on gene expression of the SL gene cluster, so cultures with (w) and without (w/o) addition of rapeseed oil were included in the experiments.

V.2.2. *Sampling, RNA extraction and RNA quality*

V.2.2.1. *Sampling*

Yeast samples for real-time PCR analysis (1 mL in 2 mL eppendorf tubes (flat bottom)) were centrifuged (3,000 rpm, 4 °C, 3 min), and the cell pellets were flash frozen in liquid nitrogen and stored at -80 °C until RNA extraction.

Alternatively the cell pellet was washed with 1 mL of cold PBS (3,000 rpm, 4 °C, 3 min) and subsequently dissolved in 500 µL of cold RNAlater (Qiagen) combined with 500 µL of cold PBS and stored at -80°C.

V.2.2.2. *RNA extraction*

Because no extensive qPCR experiments have been performed yet with *S. bombicola*, as its genome was only recently sequenced at our lab, no information considering RNA extraction thereof is available in the literature. It was already demonstrated that the *S. bombicola* cell wall is more difficult to rupt as compared to model organisms like *S. cerevisiae*. This is especially true for stationary cells and sometimes results in very low concentrations of DNA when performing gDNA extractions of the yeast. RNA extraction was thus also expected not to be as straightforward as described for *S. cerevisae*. As a good and repeatable RNA

extraction method is required to enable the development of a robust qPCR platform, several sampling methods and RNA extraction methods were compared:

1) Mechanical disruption of the frozen cells using a spatula to crush the frozen pellets under constant addition of liquid nitrogen.
2) Lyophilizing the frozen pellets overnight and pulverizing the resulting lyophilized pellets under constant addition of liquid nitrogen.
3) Defrosting the cells solved in RNAlater and placing the cell suspensions it in a petri dish at -80 °C followed by lyophilisation (overnight).
4) Centrifuge the RNAlater cell suspensions (3,000 rpm, 4 °C, 3 min), resolve the pellets in RNase free water and place the resulting solution in a petri dish at -80 °C followed by lyophilisation (overnight).
5) A third method for the RNAlater samples consisted of centrifugation and flashfreezing with liquid nitrogen of the resulting pellets, which were subsequently treated as described in method 1).
6) Lysing the cells using the fast prep device (MP). Frozen pellets were thawed and dissolved in 1 mL of lysis buffer and shaken for 20 sec at 6 m/sec.

For these first four methods subsequent RNA extraction was performed with the RNeasy mini kit (Qiagen) using the yeast protocol as specified by the manufacturer. The first step consisted of adding 600 µL of buffer RLT, to which β-mercaptoethanol had been added, to the crushed material and using a syringe to dissolve the material in this lysis buffer, while applying shearing forces to the cell material. After that 600 µL of ethanol was added to the obtained suspension, mixed well and the protocol as provided by the supplier was subsequently followed. All the above described methods were thus performed to determine the best one for breaking open the yeast cells prior to RNA extraction. Methods were compared based on the RNA yield and quality (cfr. V.3.1). To eliminate contamination of genomic DNA, a DNase treatment was performed according to the RNeasy kit instruction with the RNase-free DNase set (Qiagen).

7) A fifth and last method consisted of RNA extraction using a solvent extraction in combination with the fast prep device. Cell pellets derived from flash frozen or RNAlater samples were solved in 0.5 mL extraction buffer (1 mM EDTA, 0.1 M LiCl, 0.1 M Tris-Cl pH 7.5), 0.5 mL PCI (phenol-chloroform-isoamylalcohol in a ratio 25:24:1) and 1 % SDS (final concentration; Sigma). This solution was pipetted in 2 mL tubes containing lysis matrix C (MP) and placed in the Fastprep Device (MP Biomedicals) and shaken for 20 sec at 6 m/sec. These tubes were subsequently

centrifuged (14000 t/min, 10 min, 4°C) and the PCI extraction was repeated to purify the RNA. Subsequently 1/20 volume of 40 % KAc, pH 5.5 and 2 volumes of 100 % EtOH (cold) were added, after which the samples were vortexed and placed at -20 °C for at least one hour. Samples were subsequently centrifuged (14,000 rpm, 10 min, 4 °C) after which the supernatants were removed. Cells were subsequently washed with 70 % EtOH (cold) to remove all remaining phenol. The dried pellet was solved in 50 µL of RNasefree water. The extracted RNA was quantified using the Nanodrop device (Thermo Scientific) and stored at -80 °C for further experiments.

V.2.2.3. *RNA quality*

RNA quality plays a major role in the generation of accurate quantitative results from gene expression analysis experiments. The RNA quality of the different extraction and sampling methods was analysed by running some representative samples of different RNA extraction methods on the Experion (Bio-Rad Laboratories, Inc.) system. This automated electrophoresis system provides an effective method for determining both the quality and quantity of RNA in gene expression analysis experiments. The calculated RNA quality indicator (RQI) automatically assesses the integrity of RNA samples.

The NanoDrop A260/A280 ratios were additionally used for quality control for further experiments as analysis using the Experion device is expensive and was hence only used to analyse a number of representative samples. Each of the systems was also used to measure the RNA yield. The quantification from the NanoDrop was used to measure the input concentration of RNA for cDNA synthesis. As another preliminary quality control assay, the absence of contaminant genomic DNA in RNA preparations was verified using RNA as a template in real-time PCR assays (i.e. RNA not reverse-transcribed to cDNA).

V.2.2.4. *SPUD assay*

Current quality assessment procedures are unable to reliably detect the presence of inhibitors in RNA samples, which demonstrates the need for an additional quality assessment parameter. The “SPUD” assay developed by (Nolan *et al.*, 2006) is a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations from all tissues with the exception of potato. An oligo derived from the *Solanum tuberosum* (potato) *phyB* gene (GenBank Y14572) is the only amplifiable target and gene specific primers are used for application in the SPUD assay, which consists of a qPCR reaction in the presence of 1 µL of RNA. Under the imposed conditions, a reference Ct value of between 23 and 24 is characteristic of an uninhibited assay.

Control qPCR assays were performed in the presence of water (negative control) or heparine (positive control). The reaction mixtures (5 µL final volume) consisted of 2 µL of the Eurogentec Mastermix, 0.25 µL of each primer (250 nM final concentration), 2 µL of SPUD (5000 molecules/µL) and 1 µL of H_2O (negative control) or 1 µL of heparine (1/4000 or 1/6000 or 1/8000) (positive control) or 1 µL of the respective RNA sample. When the difference between the Ct values of the H_2O samples and the RNA samples is > 1, some kind of inhibitory compound is concluded to be present in the RNA samples. A set of representative samples for the final gene expression experiments was chosen to perform the SPUD assay. The different conditions were the same as those to determine gene expression stability of candidate reference genes (cfr. V.2.1.2).

V.2.2.5. *cDNA synthesis*

cDNA synthesis was performed using the SuperScript III First-Strand synthesis kit (Invitrogen) using oligo$(dT)_{20}$ primers in 20 µL reactions containing 200 ng of total RNA according to the manufacturer's recommendations. The cDNA concentration was then determined using the Nanodrop device (Thermo Scientific).

V.2.1. *qPCR*

Oligonucleotides for real-time PCR (cfr. Table V. 1) were subjected to a BLAST analysis against the *S. bombicola* genome database for specificity confidence in addition to analysis using the M-fold server to avoid positioning on risky secondary structures. The SsoAdvanced SYBR Green Supermix (Bio-Rad) was used to perform the experiments described in V. *2.3.1* and V. *2.3.2* and reactions volumes consisted of 2.5 µL of mastemix, 0.25 µL of each primer (250 nM final concentration) and 2 µL of template. Negative controls (NTC) consisted of RNase free H_2O (Sigma) and two different interrun calibrators (IRCs) were also included: gDNA of *S. bombicola* (5 ng/µL) extracted like described in Chapter II. 2.2.1 and a mix of several cDNA samples (5 ng/µL) covering the complete range of conditions under investigation. A mastermix for gDNA and one for cDNA respectively were prepared; both with a concentration of 5 ng/µL, which were subsequently divided in aliquots of 60 µL and frozen at -20 °C. Each sample was run in duplicate in a 384-well plate (Bio-Rad, CA) with the LighCycler 480 Real-Time PCR device.

Table V. 1 Primers used for performing qPCR. HSK= house keeping gene; GOF = genes of interest

Primer	5'-sequence-3'	Primer	5'-sequence-3'
P417_FOR_HSK1_LSC2	TGACGAGGTTGAAGGTATCA	P445_FOR_HSK15_TFC1	TGCACTCGAAGACATAATCC
P418_REV_HSK1_LSC2	GAACGCGTCAAATTGTACTG	P446_REV_HSK15_TFC1	ATCGTCGACAAACTTGACAT
P419_FOR_HSK2_CPAI	ATCCCATCTCTCGACTTACA	P447_FOR_HSK16_ACT1	TAAGGAGCTCTACGGTAACA
P420_REV_HSK2_CPAI	CGAAGTAAGGCTTCCAATCA	P448_REV_HSK16_ACT1	CATCGAAGATGGAGCCATAG
P421_FOR_HSK3_MPP	ATGAGCTTCTCGACCTCTTA	P449_FOR_HSK17_G6PDH	TCAAATTGCCCGTAATGAGT
P422_REV_HSK3_MPP	AAATGAACCAGGTCGTCAAT	P450_REV_HSK17_G6PDH	CAATCCCGGGAATTTGGTAT
P423_FOR_HSK4_ERVP	GCTGTCTATGGTCATCACTG	P451_FOR_HSK18_UBC6	GCAGCGTGAAATGGAATATG
P424_REV_HSK4_ERVP	CATCGAAGCTTGTGTCCTTA	P452_REV_HSK18_UBC6	CACCGACTGCTTAATTTCCT
P425_FOR_HSK5_GTPBP	GACAACTCTTACACGGACTC	P453_FOR_HSK19_TAF10	AGACCACAGACCACAGAATA
P426_REV_HSK5_GTPBP	AACCGTTCGGATCTTGAAAT	P454_REV_HSK19_TAF10	GCTCTAATCCGGGAGAATTG
P427_FOR_HSK6_HSP	CGATGTTGCTAACCTTCTGA	P455_FOR_HSK20_PMAI	CCCTTGGTGTCTCTGTTAAG
P428_REV_HSK6_HSP	TGAAGTAGTAGCACCAATGC	P456_REV_HSK20_PMAI	CAACAATGGTCTGAGCAGTA
P429_FOR_HSK7_GADPH	TCAACGCTACCATCAAGAAG	P457_FOR_GOF1_CYP52M1	CAATCCTGGCAACCAAATTC
P430_REV_HSK7_GADPH	CGTCGTAGATAGAGGAGTGA	P458_REV_GOF1_CYP52M1	CGAGACTGTTTCCATCCATT
P431_FOR_HSK8_RIP	GCATGAAATTGACGAAGCAA	P459_FOR_GOF2_UGTA1	AGCAGCCATCAACTATGAAG
P432_REV_HSK8_RIP	CGATCATAACAAGCCACTCA	P460_REV_GOF2_UGTA1	CTTTGACGCCCAATATACCA
P433_FOR_HSK9_PGIA	CATGGTTTCCGAAGCTCTTA	P461_FOR_GOF3_UGTB1	CCGCAGTGATCATACCTTAG
P434_REV_HSK9_PGIA	CTTCAGGGTCTCAGCAATAC	P462_REV_GOF3_UGTB1	GTGGTTCATAGCGAGTTTCT
P435_FOR_HSK10_PDC	TACGACAATGTCTTCTTCGG	P463_FOR_GOF4_AT	CTGCGACTCATCCTATTAGC
P436_REV_HSK10_PDC	GAGTGAGCTGAGTGATCTTG	P464_REV_GOF4_AT	AATTGTTGAGCCATCTCCAA
P437_FOR_HSK11_SDH	TTGAAGGCTAATGCTGACTC	P465_FOR_GOF5_MDR	TGTATGGAGTGAGGAAGGTT
P438_REV_HSK11_SDH	GTCTTGGCATCAGAGAACAT	P466_REV_GOF5_MDR	GTACTTGAGGTCGAGTAGGA
P439_FOR_HSK12_COF	TTTACGACTTCGACTACGAG	P467_FOR_GOF6_LIP	CTTGTCGAGCAGTATGTTGA
P440_REV_HSK12_COF	GTAGATCATCTTGTTGCGGA	P468_REV_GOF6_LIP	TCATGAAGAAGACCCGGATA
P441_FOR_HSK13_ACT2	TTCCAGAGCATCATGAAGTG	P469_FOR_GOF7_ADH	TCACGCACTCAAACTTACAA
P442_REV_HSK13_ACT2	CATCGAAGATGGAGCCATAG	P470_REV_GOF7_ADH	GTGATACCAGCACAGAGAAG
P443_FOR_HSK14_PL	TAAAGATCTGAAGCTCACGC	P471_FOR_GOF8_REGUL	GGGCCATTATACCGACAAAT
P444_REV_HSK14_PL	GCGGATACGTGATTCCATTA	P472_REV_GOF8_REGUL	CACCATGACGAATCCCATAA

V.2.1.1. *Determination of primer efficiencies*

The PCR efficiency of each primer pair (Eff) was evaluated by the dilution series method using both a mix of cDNA samples and gDNA as the templates. Briefly, a 4x serial dilution using 6 points, with the highest concentration being 32 ng/µL was used and the efficiencies for each primerpair were determined from standard curves using the formula 10(-1/slope). This analysis was done to evaluate the primer efficiencies of all the primers that would subsequently be used for performing qPCR experiments. If a primerpair turns out to have a primer efficiency lower than 1.8 a new set of primers will be developed.

V.2.1.2. *Determination of the stability of 20 candidate reference genes*

Twenty candidate reference genes were selected from literature and proteomic results. For accurate qPCR experiments, normalization using at least 3 controlled housekeeping genes should be done. Because no expression data is available for *S. bombicola* and because one should first assess the expression stability of a certain potential housekeeping gene, prior to using it to normalize qPCR data, a set of 20 candidate housekeeping genes for *S. bombicola* was chosen. Eleven of them were selected from literature of qPCR experiments performed for other yeasts and/or fungi. Because the expression stability of many of these genes was never validated, but just assumed to be stable, additionally a set of nine genes derived from proteomic data generated by PhDs Katarzyna Ciesielska at the lab of Professor Bart Devreese were also included. These proteins were shown to have a stable protein expression ratio of 1 between two conditions under investigation (exponential versus stationary cells). Shortly said these results were obtained by analyzing data from SILAC experiments using the Maxquant software. This software automatically determines SILAC peptide and protein ratios. An overview of the candidate reference or housekeeping genes (HSK) is given in Table V.2. The reference gene candidates which were derived from proteomics are underlined. One exception was actin, which was found from the proteomic based approach, but is also used a lot for normalization in the literature. The table also includes the genes of interest, namely the genes of the SL biosynthetic gene cluster. The two open reading frames flanking the SL gene cluster (cfr. Chapter II) were also included.

The expression stability of the candidate reference genes was subsequently assessed in a set of conditions relevant for the final gene expression experiments. These conditions were:

- Exponential versus Stationary cells (6 hours, 24 hours, 74 hours and 213 hours)
- With oil/without oil
- *S. bombicola* wild type/*S. bombicola* Δ*lys* strain

All possible combinations were included in the assay. The stability of mRNA expression of the evaluated reference genes was evaluated by PhD Filip Pattyn at the lab of Professor Jo Vandesompele (Center for Medical Genetics) using the geNorm software. This program calculates the gene expression stability measure "M" for a potential reference gene as the average pair-wise variation for that gene with all other tested genes. Then it ranks genes considering that those with the lowest M value are the most stably expressed. The authors of

the geNorm software recommend the use of minimally three stable internal control genes for calculation of the normalization factor (NF3) (Vandesompele *et al.*, 2002).

Table V. 2 Overview of the candidate housekeeping genes (HSK) selected from literature and proteomics (underlined) and the genes of interest for this study.

Code	Short Genename	Corresponding enzyme
HSK1	LSC2	Succinyl-CoA synthetase
HSK2	CPAI	Carbomoyl phosphate synthetase
HSK3	MPP	Mitochondrial processing peptidase subunit
HSK4	ERVP	ER vesicle protein
HSK5	GTPBP	GTP binding protein (GTPase)
HSK6	HSP	Heat shock protein
HSK7	GADPH	Glyceraldehyde-3-phosphate dehydrogenase
HSK8	RIP	Ubiquitinol cytochrome C reductase
HSK9	PGIA	Glucose-6-phosphate isomerase
HSK10	PDC	Pyruvate decarboxylase
HSK11	SDH	Saccharopine dehydrogenase (lysine biosynthesis)
HSK12	COF	Cofilin (actin depolymerizing factor)
HSK13	ACT2	Actin
HSK14	PL	Pyridoxal reductase
HSK15	TFC1	RNA polymerase III TF
HSK16	ACT1	Actin
HSK17	G6PDH	Glucose-6-phosphate dehydrogenase (pf G6PD)
HSK18	UBC6	Ubiquitin-conjugating enzyme
HSK19	TAF10	Subunit of TFIID and SAGA complexes
HSK20	PMAI	Adenosin triphosphatase
GOF1	CYP52M1	Cytochrome P450 monooxygenase
GOF2	UGTA1	Glucosyltransferase A1
GOF3	UGTB1	Glucosyltransferase B1
GOF4	AT	Acetyltransferase
GOF5	MDR	Multidrug resistance proteins
GOF6	LIP	Lipase
GOF7	ADH	Alcoholdehydrogenase (= orf1)
GOF8	REGUL	Putative regulator SL cluster (= orf2)

V.2.1.3. *RT-PCR for expression analysis*

The reaction mix (20 µL final volume) consisted of 10 µL of 2x sensimix (BioLine), 2.5 µL of each primer (250 nM final concentration), 4 µL of H_2O, and 1 µL of a 1/10 dilution of the cDNA preparation. A blank (no template control=NTC) was also incorporated in each assay. The thermocycling program consisted of one hold at 95 °C for 10 min, followed by 40-50 cycles of 25 s at 95 °C, and 60 s at 58 °C and 20 s at 72 °C. After completion of these cycles,

melting-curve data were collected to verify PCR specificity, contamination and the absence of primer dimers. The gene expression data were analysed and normalized using the REST software (Pfaffl *et al.*, 2002). Relative expression levels were determined with efficiency correction, which considers differences in primer pair amplification efficiencies between target and reference genes.

V.3. Results

V.3.1. *Determining important parameters for performing qPCR experiments*

V.3.1.1. *Determination of the best method for RNA extraction*

Several methods for RNA extraction were tested and evaluated using different parameters. RNA quality was set as the most important factor and was determined using the Experion device at the lab of Professor Jo Vandesompele. Two representative samples are shown in Figure V. 1.

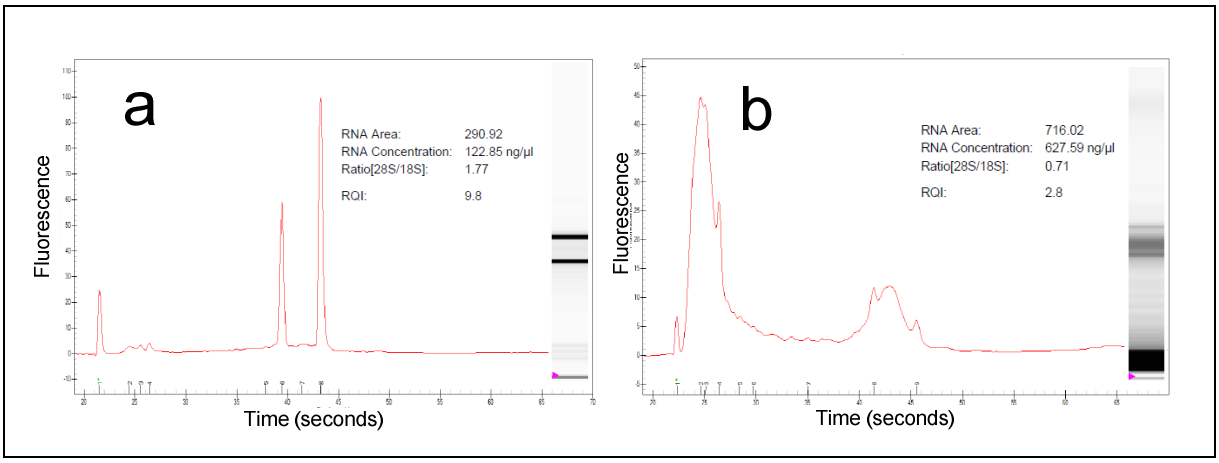


Figure V. 1 Typical outputs of experion analyses used to determine RNA quality. Two representative samples are shown: **(a)** RNA extracted using the third method described in Table V. 3 (good) **(b)** RNA extracted using the last method described in Table V. 3 (bad). The RNA quality indicator (RQI) value is a measure for RNA integrity scaled from 0 (very bad) to 10 (very good).

From the results depicted in Figure V. 1, one can clearly decide that the extraction method for which the experion analysis is depicted in Figure V. 1. a, gives rise to a considerable amount of highly qualitative RNA (122 ng/µL; RQI: 9.8). This is in sheer contrast with the second extraction method for which the experion analysis is depicted in Figure V.1 b. Although high amounts of RNA are obtained with this method (627 ng/µL), the quality thereof is extremely poor (RQI: 2.8) and can't be used to perform reliable qPCR experiments. When loading such RNA samples on an agarose gel the two clear separate bands that should be visible for the 18S and 28S rRNA are instead present as a smear, as can also be concluded from the experion output depicted on the right of the respective analyses. The ribosomal RNA is thus degraded,

which indicates that this is also true for the present mRNA and such samples should not be used to perform gene expression analysis.

A representative set of samples was run on the experion and furthermore controlled by running them on standard agarose gels and by using the nanodrop device to determine the concentration. An overview of the results for the seven tested RNA extraction methods described in V.2.2.2 is given in Table V. 3.

Table V. 3 Different parameters used for the selection of the best RNA extraction method and the obtained results for each method.

Method	Experion	Concentration	Gels	Time
1.Cell pellets (-80°C) + grinding + cte addition of liquid N_2	High	ok	ok	moderate
2.Grinding lyophilized cell pellets + cte addition of liquid N_2	High	ok	ok	very long
3.RNA later + lyophilizing + grinding + cte addition of liquid N_2	Very high	ok	ok	longest
4.RNA later + washing+ lyophilizing + grinding + cte add of N_2	High	ok	ok	longest
5.RNA later + pelleting + grinding + cte addition of liquid N_2	High	ok	ok	very long
6.FAST PREP	Bad	Very high	bad	fast
7.FAST PREP + solvent	Bad	Very high	bad	fast

From the results depicted in Table V. 3 one can conclude that the first method, namely grinding of frozen cell pellets (-80 °C) with liquid nitrogen is the best method, when making the balance between the needed processing time and the obtained results. Although this method takes longer than the methods based on lysing the cells using the Fast Prep device, this method gives rise to good and repeatable results and will hence always be used in the future for RNA extraction and subsequent qPCR experiments, which is thus be part of the developed qPCR platform.

V.3.1.2. *Spud assay and gDNA contamination*

To assess if inhibitory compounds for qPCR (SLs, oil,etc.) are present in the RNA samples a Spud assay was performed as described in V.2.2.4. A set of representative samples was chosen to perform this assay (the different conditions were the same as those to determine gene expression stability of candidate reference genes). All assays containing *S. bombicola* RNA generated the same Ct values as the water controls. However, there was a clear shift to higher Ct values and reduced amplification efficiency for the samples containing the heparine (positive controls). It can thus be concluded that no inhibitory compounds are present in the RNA samples and that qPCR analysis of such samples will be reliable.

In the same run the same RNA samples were subjected to standard qPCR to assess if contaminating genomic DNA was still present in the samples. No amplificiation or very high Ct values were obtained for these experiments, so it can be concluded that the used RNA extraction methods give rise to sufficiently pure RNA freed from contaminating DNA.

V.3.1.3. *Determining primer efficiencies*

Two dilution series of cDNA and gDNA were prepared to determine the primerefficiencies of the developed primers. These should ideally be equal to 2, but values as low as 1.8 are still accepted. The primer efficiencies of all the used primercombinations were higher than 1.93 and are thus all considered to meet the set standards for good qPCR primers. The developped primers can thus be used for performing reliable qPCR experiments.

Now that all these parameters have been controlled, the set of chosen candidate reference genes can be assessed for expression stability and if good, the best three can subsequently be used for the normalization of expression data.

V.3.2. *Assessing the stability of 20 potential reference genes*

For accurate qPCR experiments, normalization using at least 3 controlled housekeeping genes should be done. Because no expression data is available for *S. bombicola* and because one should first assess the expression stability of a certain potential housekeeping gene, prior to using it to normalize qPCR data, a set of 20 potential candidate housekeeping genes for *S. bombicola* was chosen. Eleven of them were selected from literature of qPCR experiments performed for other yeasts and/or fungi. Because the expression stability of many of these genes was never validated, but just assumed to be stable, a set of 9 additional genes derived from proteomic data were also included (cfr. V.2.1.2). These proteins were shown to have a stable protein expression ratio of 1 between two conditions under investigation (exponential versus stationary cells).

The expression stability of the 20 candidate reference genes was assessed in a set of samples representative for the situations that would be used for the expression analysis of the genes of interest as described in V.2.3.2. The resulting qPCR data were analysed using the geNorm software and the results are shown in Figure V. 2.

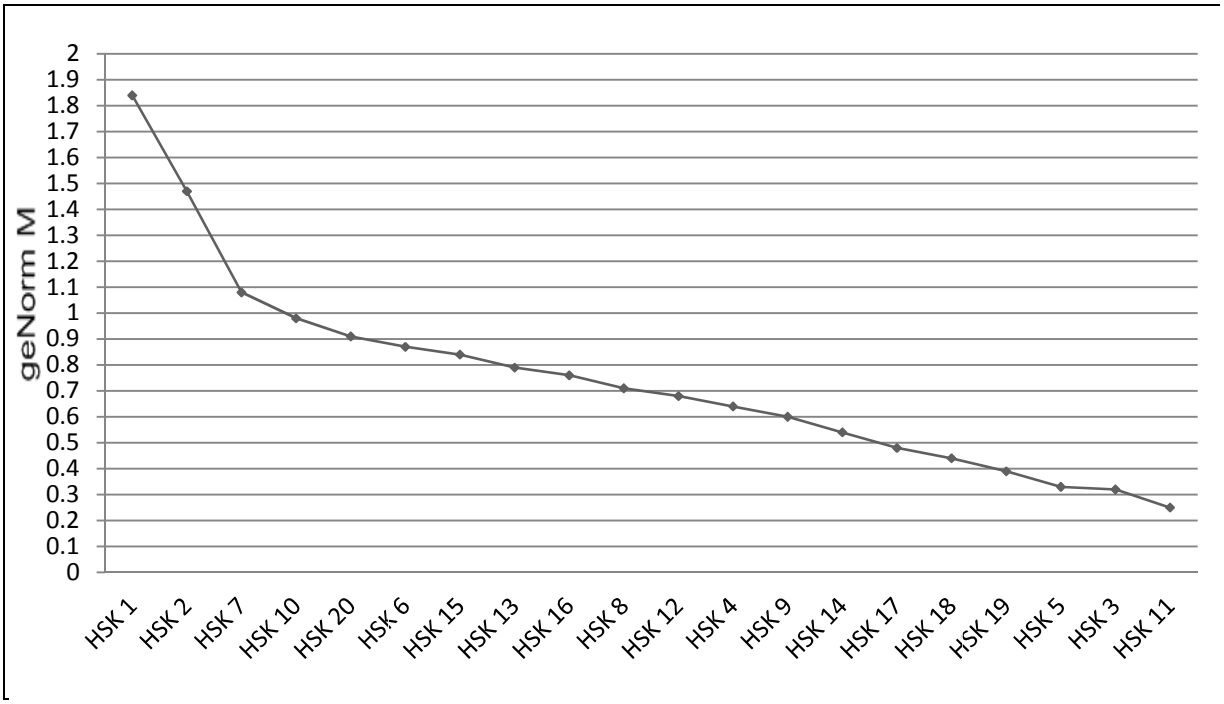


Figure V. 2 Output of the geNorm program ranking the set of candidate housekeeping genes according to their M value. Low M-values correspond to the most stable genes.

The geNorm software calculates an M-value for each candidate reference gene and the lower the M-value the more stable the expression of the gene. It can thus be concluded from Figure V. 2 that the three best reference genes under the imposed conditions are HSK 3, HSK 5 and HSK 11 (cfr. Table V. 1 for corresponding genenames). Strikingly, these were all derived from the proteomic data (stably expressed genes between exponential and stationary conditions), so the gene expression in this case is well related to protein translation. It can also be concluded that genes like *actin* (HSK 13 and HSK 16) and *gapd* (HSK 7) which are generally considered to be good housekeeping genes indeed give rise to acceptable M-values under the imposed conditions as M values below 1.5 are considered to be ok. However, they are far from being the most stable genes and it is thus recommended to assess the stability of a candidate reference gene before using it to normalize qPCR data. Furthermore it should be mentioned that almost all of the selected candidate reference genes show M-values below 1.5 and that the chosen set can hence be considered as a good collection of reference genes, which can be used again to reassess stability if one wants to include new sample conditions.

For performing gene expression experiments, qPCR assays corresponding to the selected three most stably expressed genes will thus be included for all the samples under investigation and the results thereof will be used for subsequent normalization of the obtained qPCR data.

V.3.3. *Validation of the qPCR platform*

Now that a qPCR platform is available for *S. bombicola* it can be used to perform gene expression experiments. This will *e.g.* enable the further investigation of the influence of factors like pH, temperature, used lipogenic substrate, etc. which have been described to affect the SL productivity and/or composition (Cfr. Chapter I), on the level of the expression of the genes of the SL biosynthetic pathway.

As the conditions imposed in the production medium (cfr. Chapter I, II and III) result in high SL production in the stationary growth phase, the genes of the SL biosynthetic pathway are expected to be upregulated under such conditions. Furthermore, addition of oil to the cultures results in higher SL titers as compared to experiments to which no oil has been added. For this last observation two possible causes can exist. Firstly this could be attributed to a higher precursor availabiltity for SL biosynthesis, as the yeast can incorporate the deduced fatty acids as such in the SL molecules. Alternatively the addition of oil could result in a further upregulation of the genes of the SL biosynthetic pathway as compared to stationary cultures to which no oil was added, resulting in higher SL productivity. A combination of both effects cannot be ruled out either.

Two shake flasks experiments with the wild type cultivated on production medium were thus initiated. One with (w) addition and one without (w/o) addition of oil. Two biological replicas were included for each situation. The results of the gene expression analyses are depicted in Table V. 4. Normalization was done using HSK 3, HSK 5 and HSK 17 as in the first runs it became clear that something was wrong with gene-expression of HSK 11.

Table V. 4 Results of a qPCR gene expression experiment of which the data were analysed using the REST software. When genes are significantly regulated, they are termed with UP or DOWN, if no significant effects were observed, this is marked with '/'

Gene of interest (GOF)	Stationary w oil versus stationary w/o oil	Stationary versus Exponential
cyp52M1	/	UP
ugtA1	Not tested	UP
ugtB1	Not tested	UP
at	Not tested	UP
mdr	Not tested	/
lipase	/	/

The first thing that can be concluded from Table V.4 is that the addition of oil did not result in the upregulation of the first gene (*cyp52M1*) of the biosynthetic pathway nor of the last one (*lipase*). Although the presence of lipogenic substrates in the culture medium was described to influence the amount of lactonisation (Zhou *et al.*, 1992; Davila *et al.*, 1994), the effects thereof could not be demonstrated on the transcriptional level. These observations are thus more likely attributed to effects on the enzyme level, which is plausible, as lactonisation involves the release of a water molecule and will thus be stimulated in a hydrophobic surrounding, for which the lipogenic substrate might be held responsible.
The other genes weren't included yet in this analysis, but these experiments will be executed in the future. However, as the first gene of the SL biosynthesis (*cyp52M1*) wasn't significantly upregulated in the presence of oil, one might expect that the same effects will be observed for the other genes. In that case the higher SL titers obtained when cultivating *S. bombicola* on the production medium w the addition of oil as compared to an identical medium w/o the addition of oil, are purely attributed to a higher precursor availability for SL biosynthesis in the first situation. The opposite was demonstrated to be true for the presence of high glucose concentrations in the culture medium, but this will be discussed in Chapter VII of this manuscript.

A second conclusion that can be drawn from Table V.4 is that almost all the structural genes of the SL biosynthetic pathway are significantly upregulated in the stationary growth phase. This was expected, as SL production is mainly described to occur in the stationary growth phase. The lipase gene surprisingly wasn't significantly upregulated in the stationary phase and thus seems to be differently regulated as the other genes of the SL biosynthetic pathway. This could be attributed to the fact that the gene is located elsewhere in the genome and is thus subject to other regulatory mechanisms. However, this was also true for the SL transporter. These data were confirmed by recent RNA sequencing data derived from exponential and stationary cells. In one of the two replicas the transporter was also not significantly upregulated in the stationary growth phase, so this gene is most likely to a lesser extent upregulated as compared to the other genes of the SL biosynthetic pathway.

As the results derived from RNA sequencing thus gave rise to the same conclusions (for the second experiment), this confirms that the developed qPCR platform is suitable for performing gene expression experiments and that the regulatory effects of a range of parameters on gene expression of the SL biosynthetic pathway can now be investigated.

V.4. Conclusions

In this chapter a qPCR platform for *Starmerella bombicola* was developed.

Seven methods for RNA extraction were compared, which resulted in the selection of one method, which will be used for all future qPCR experiments.

The expression stability of twenty candidate reference genes in several conditions was assessed using the software package geNorm. This resulted in the selection of the three most stable reference genes, which can be used for the normalization of future gene expression experiments derived from similar conditions as the ones that were evaluated. The complete set of candidate reference genes gave rise to reasonable M-values so the chosen collection of HSKs is considered as a good set of reference genes of which stability can be reassessed if one wants to include new conditions.

The qPCR platform was subsequently validated by performing expression analysis experiments with the wild type. This gave rise to repeatable and logical results as all the genes of the SL biosynthetic cluster were clearly upregulated under conditions of SL biosynthesis. The addition of oil furthermore did not result in an additional upregulation of two genes of the SL pathway as compared to the situation were no oil was added. Higher SL productivity when oil is added to the culture medium can thus be completely attributed to the higher precursor availability for SL production. Furthermore, the described effect of the presence of oil in the culture medium on lactonisation is suspected to be caused by the creation of a more hydrophobic surrounding for the lipase enzyme, which will stimulate lactonisation. The effects of parameters like pH, nitrogen limitation, temperature, carbon source, etc. on expression of the genes of the SL biosynthetic pathway can now be easily investigated, which will be the subject of future research.

This qPCR platform is furthermore considered to be a useful tool for future genetic engineering of *S. bombicola*. A strain knocked out in the acetyltransferase of the SL biosynthetic pathway *e.g.* gives rises to the production of non-acetylated SLs, which is a desirable feature (Saerens *et al.*, 2011b). However, yields drop to about 10 % as compared to those of the wild type, which is too low for an industrial production process for such molecules to be feasible. However, an additional deletion of the lipase responsible for lactonisation (*ΔatΔlip)* results in a strain producing almost equal amounts of SLs as the wild

type (unpublished results). In the first case it was assumed that a positive feedback of the di-acetylated acidic SLs on the pathway existed, which wasn't occurring anymore in the *Δat* strain, or alternatively a negative feedback of a buildup of non-acetylated molecules existed in the *Δat* strain. However, this is in contradiction with the results obtained for the recently developped *ΔatΔlip* strain. The question thus raises what kind of effects cause these observations. The developed qPCR platform can thus be used to investigate if this can be attributed to regulatory effects *e.g.* negative feedback of the intermediates on gene expression or positive feedback of the end-products or that alternatively some other effects are at work *e.g.* transport related issues. This should also be done for the strain knocked out in the second glucosyltransferase (*ugtB1*) for which comparable results were obtained (Saerens *et al.*, 2011c).

PART III:

STARMERELLA BOMBICOLA AS A PLATFORM ORGANISM FOR INTERESTING BIOMOLECULES

CHAPTER VI:

PRODUCTION OF 100 % LACTONIC SLS

Part of this chapter was submitted as a patent application:

Ciesielska, K.; Roelants, S. L. K. W.; Van Bogaert, I. N. A.; B., D., *et al.* (2011). A lactonase derived from *Candida bombicola* and uses thereof. UGent. PCT/EP2012/075571 (public july 2013).

Part of this chapter is in preparation for publication:

Roelants, S., Ciesielska, K., L, Moens, H., Kuen Chow, K., Verweire, S., De Maeseneire, S. L., Van Bogaert, I., Devreese, B. and Soetaert W. Towards the industrialisation of new biosurfactant producers: microbial production of 100 % lactonic sophorolipids.

Chapter VI.

PRODUCTION OF 100 % LACTONIC SLS

VI.1. Introduction

Sophorolipid (SL) production by *Starmerella* species e.g. *S. bombicola* has one great advantage over other biosurfactant producers; its SL machinery is highly efficient, resulting in titers of 400 g/L and higher (Davila *et al.*, 1997; Daniel *et al.*, 1998). This feature allowed the development of industrial production processes with this yeast and applications of the produced biosurfactants in *e.g.* the ecological cleaning solutions of Ecover. However, the efficient production process has the disadvantage that a mixture of as many as 23 homologs (8 major and up to 15 minor) (Davila *et al.*, 1993), containing both acidic and lactonic forms (Asmer *et al.*, 1988) is formed. Moreover the ratio of these homologs varies as a function of their substrate, *e.g.* a hydrocarbon source, as well as the fermentation conditions (cfr. Chapter I). This hampers product development, since the varying structures of the molecules translate in different physicochemical properties and function. Depending on the required function of the biosurfactant in the final product formulation, there will thus be a requirement for either a single form of the molecule to be available or for the naturally produced mixture to be manipulated to give changed ratios of the components (Marchant and Banat, 2012). Acidic SLs for example have been found therapeutically active for fibrinolysis, wound healing, desquamation, depigmenting and activation of macrophages (Maingault, 1999). The lactone form of SLs has been documented to be the most active SL type, both with regard to lowering surface tension and to antimicrobial activity. The acetylated lactonic SLs have found applications in cosmetics as component of deodorants and moisturizers, as antidandruff or bacteriostatic agents and in the treatment of skin diseases (Mager *et al.*, 1987) and were also reported to have anti-cancer activity (Scholz *et al.*, 1998; Gross *et al.*, 1999). The produced SL mixture thus has to be fractionated and purified to be suitable for such specific applications. A lot of effort has therefore been directed towards the elaboration of the optimal cultivation conditions (Stüwer *et al.*, 1987; Lang *et al.*, 2000) and processes for the isolation of one of the main forms (both acid and lactone) have been described by *e.g.* using solvents. However, the use of solvents is not desirable in view of the use of SLs as biosurfactants and it is furthermore undesirable that the product contains traces of organic solvents.

As stated above, and discussed in Chapter I, a lot of effort has been put in the investigation of the influence of the medium composition and culture conditions on the produced SL mixture. A medium component that has been described to have a profound effect on the SL composition, and especially the lacton/acidic ratio, is citrate. Lactonization of the acidic forms of SLs produced by *C. apicola* was shown to proceed only in the presence of whole cells, which were grown in medium supplemented with citrate (Stüwer *et al.*, 1987). Furthermore, a cell-wall bound lipase activity in *C. apicola* with synthetic capacities in apolar media was demonstrated (Baldé, 1989). When lactonic SLs were found to be formed and deposited as crystals on the cell wall of the yeast, this incited researchers to suggest the involvement of a cell wall-bound lipase in lactone formation (Hommel *et al.*, 1994b). Indeed, the enzyme responsible for lactonisation of SLs in *S. bombicola* was recently identified in the secretome of *S. bombicola* by performing proteomic experiments (Ciesielska *et al.*, 2013). The enzyme was indeed predicted to be a 'lipase' and was most similar to a lipase of *Kurtzmanomyces sp*. I-11 (Kakugawa *et al.*, 2002a) and lipase A from *Candida Antarctica*, although lactonisation clearly cannot be defined as typical 'lipase' activity. Surprisingly this enzyme was found to be fully responsible for lactonisation, as a Δ*lip* deletion strain was exclusively capable of producing 100 % acidic SLs (Ciesielska *et al.*, 2013).

As the involvement of an enzyme in lactone formation of SLs is now proven, the possibility exists that the observed influences of the medium composition and/or culture conditions on the ratio lacton/acidic SLs could be attributed (directly or indirectly) to some kind of regulatory effect on the lipase expression or activity. Overexpression of this gene with a strong constitutive promotor elsewhere in the genome could in this case circumvent such regulatory events. This could possibly give rise to a strain producing very pure lactonic SLs, which would certainly represent an important opportunity for the development of a simplified industrial production process for lactonic SLs. An overexpression strategy for the lipase gene was thus elaborated and the resulting strains were evaluated for SL production and composition under several cultivation conditions.

VI.2. Materials and Methods

VI.2.1. *Strains, media and culture conditions*

Starmerella bombicola ATCC 22214 was used as the parental or wild type strain. *S. bombicola* PT36, an *ura3* auxotrophic mutant, was used to construct the overexpression strains. The obtained transformants and wild type were cultivated in 100 mL of the production medium described in previous chapters. This medium contains sodium citrate dihydrate (5 g/L). The presence of this compound in the culture medium strongly affects the composition of the SLs (ratio of acidic/lactonic). When no citrate is present, this strongly favours the formation of acidic SLs. SL composition was hence assessed for both the wild type and overexpression transformants in an adapted production medium that did not contain citrate. Yeast cultures were incubated at 30 °C on a rotary shaker (200 rpm) and cultivation was stopped when glucose was depleted from the medium.

Bioreactor experiments were carried out in a Biostat® B 3 L culture vessel (Sartorius-BBI Systems) with a working volume of 1.5 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 vessels. For maintaining pH at 3.5, 5 N NaOH was used. The pH of the production medium at the beginning of the experiment was set at 5.8 and there was no correction for a too alkaline pH, so the pH was allowed to drop spontaneously to 3.5. Feeding of the hydrophobic carbon source was initiated 48 hours after inoculation, and from that time point on 25 g of rapeseed oil was added every 24 hours. Additional glucose was added 110 hours after inoculation. The stirring rate had to be lowered to 600 rpm after 5 days of cultivation to control foam formation for the wild type fermentations as the addition of the hydrophobic carbon source did not help anymore. Bioreactors were stopped when the second batch of glucose was depleted, which occurred after 7 days of cultivation. SLs were harvested by placing the culture medium overnight at 70 °C which resulted in precipitation of the SLs as an oily substance heavier then water, which was easily separated from the broth.

Escherichia coli DH5α cells were used in all cloning experiments and were grown in Luria-Bertani (LB) medium (1 % trypton, 0.5 % yeast extract and 0.5 % sodium chloride) supplemented with 100 mg/L ampicillin. Liquid *E. coli* cultures were cultivated at 37 °C on a rotary shaker (200 rpm).

VI.2.2. *Molecular techniques*

VI.2.2.1. *General techniques*

General molecular techniques can be found in previous chapters. The used primers can be found in Table VI. 1.

Table VI. 1 Primers used for the creation of an overexpression strain of the *S. bombicola* lipase gene

Primernames	**'5 Sequence 3'**
P1_FOR_*ura3*v	AGAACAAGGCCGAGTATGTC
P32_REV_cassette	GTCAGATTAGCCTCCGACATAG
P30_FOR_checkpromIN	AAGGCGGGCTGGAATGCATATCTGAG
P33_FOR_checkcassIN	CCATACTCAAGCGCGAACAC
P35_REV_checkcassIN	GAGCTCAAGACGCGTTTACTCAATGC
P124_FOR_P*gki*_ext*Spe*I	CTGGCAAATCACTAGGTGCTTAGGGTGCGTGTG
P125_REV_P*gki*_extlipase	GAAAAAAACAGAGCCAGCATTTTTTCTGGTTTGGAGGACCTTGGGTAG
P126_FOR_*lip*_extP*gki*	GGTCCTCCAAACCAGAAAAAATGCTGGCTCTGTTTTTTTCG
P127_REV_*lip*_ext*Bam*HI	TGCCCTGCGGGGATCTTCACTCTAAGAAATCCTCCGAGGAAATC

VI.2.2.2. *Creation of the overexpression cassette*

Overexpression of the lipase gene was obtained by placing it under the regulatory control of the phosphoglycerate kinase promoter sequence (P*gki*) and the terminator sequences of the lipase itself. The complete *gki* promotor sequence and lipase gene sequence were both amplified from genomic DNA of *Starmerella bombicola* using primerpairs P124_FOR_P*gki*_extinfu*Spe*I / P125_REV_P*gki*_extlipase and P126_FOR_lipase_extP*gki* / P127_REV_termlac_extInfu*Bam*HI respectively. Both fragments were subsequently fused using overlap PCR. The pGEM-t_cassette_*yegfp3* vector (cfr. Chapter IV), which contains the *ura3* auxotrophic marker under control of its own terminator -and up and downstream sequences for homologous recombination at the *ura3* locus- was cut with the enzymes *Spe*I and *Bam*HI. The abovementioned linear PCR fragment (P*gki*-lipase) was subsequently inserted into the cut vector using the Infusion Advantage PCR cloning kit (Clontech). The resulting vector of 7896 bp was used as a template to generate the lipase overexpression cassette using the primerpair P1_FOR_*ura3*v / P32_REV_cassette. The fragment of 4904 bp was used to transform the *ura3* negative *S. bombicola* PT36 strain and integration occurred at the *ura3* locus. The resulting strain thus harboured two copies of the lipase gene; one under its own regulatory sequences, unaltered in the genome, and a second one under control of the strong constitutive *gki* promoter integrated at the *ura3* locus. Correct integration of the cassette at the *ura3* locus was controlled by performing colony PCR with primerpair

P33_FOR_checkcassIN / P125_REV_P*gki*_extlipase at the left side of the insertion cassette and primerpair P30_FOR_checkpromIN / P35_REV_checkcassIN_DOWN at the right hand side of the insertion cassette.

VI.2.3. *Total SL extraction*

This protocol for SL extraction is based on the one described in (Fleurackers *et al.*, 2010). Wild type cultures of *S. bombicola* were cultivated for 8 days as described above. At the end of the incubation period, 3 volumes of ethanol were added to the culture broth and yeast cells were removed by centrifugation (4000 x g, 10 min). The water/ethanol mixture was evaporated using a rotary evaporator. Two volumes of ethanol were subsequently added, which results in solubilisation of the SLs and residual hydrophobic carbon sources. The resulting solution was passed through a Whatman filter to remove insolubles. To remove remaining hydrophobic substrates, a last step was performed consisting of evaporation of the ethanol followed by the addition of ½ volume of water. The pH of the solution was subsequently set at 6.5 and ½ volume of hexane was added. The resulting mixture was shaken vigorously and brought into a separatory funnel. The water fraction contains the SLs and hexane fraction the residual hydrophobic substrate.

VI.2.4. *Analytical techniques*

VI.2.4.1. *Follow up of growth*

Cell growth of the cultures was monitored in different ways: the cell density (OD) was measured as described before. This method is reliable in the exponential growth phase, but becomes less valid in the stationary phase when the lipophilic substrate is added and SLs are produced. Both substances interfere with optical density measurements and cell cry weight (CDW) was hence also determined as a measure of growth in bioreactor experiments. Colony forming units (CFU) and glucose concentrations were determined as described before.

VI.2.4.2. *Analytical extraction and analysis of sophorolipids*

Samples for SL analysis were extracted with ethylacete (EtAc) and/or ethanol (EtOH) and were analyzed by HPLC and LC-MS analysis as described in Chapter III.

VI.3. Results and discussion

VI.3.1. *Construction and characterization of overexpression transformants*

The lipase overexpression cassette was constructed as described in VI.2.2.2 and transformed in an *ura3*- auxotrophic *S. bombicola* mutant (PT36) (cfr. Figure VI. 1). After selection of *ura3*+ colonies on selective plates (SD), correct integration at both sides of the overexpression cassette was controlled by performing colony PCR using two primer combinations. Sixteen colonies were controlled of which three showed the two correct fragments after colony PCR and these were selected for further characterization. One of the three transformants did not produce SLs, so the discussed results only apply to the two remaining ones and the wild type. When cultivated on production medium, the two transformants behaved similar as the wild type: the pH for both the transformants and the wild type dropped from 5.8 to 3.1 in the first 96 hours of cultivation and remained stable at this value until glucose was depleted from the medium after about 216 hours of cultivation and the experiments were ended (cfr. Chapter III Figure III.2). CFU's were determined at four time points during the cultivations and this parameter wasn't affected either in the overexpression transformants as compared to the wild type. SL production was followed up during cultivation and the results thereof are described in VI.3.2.2.

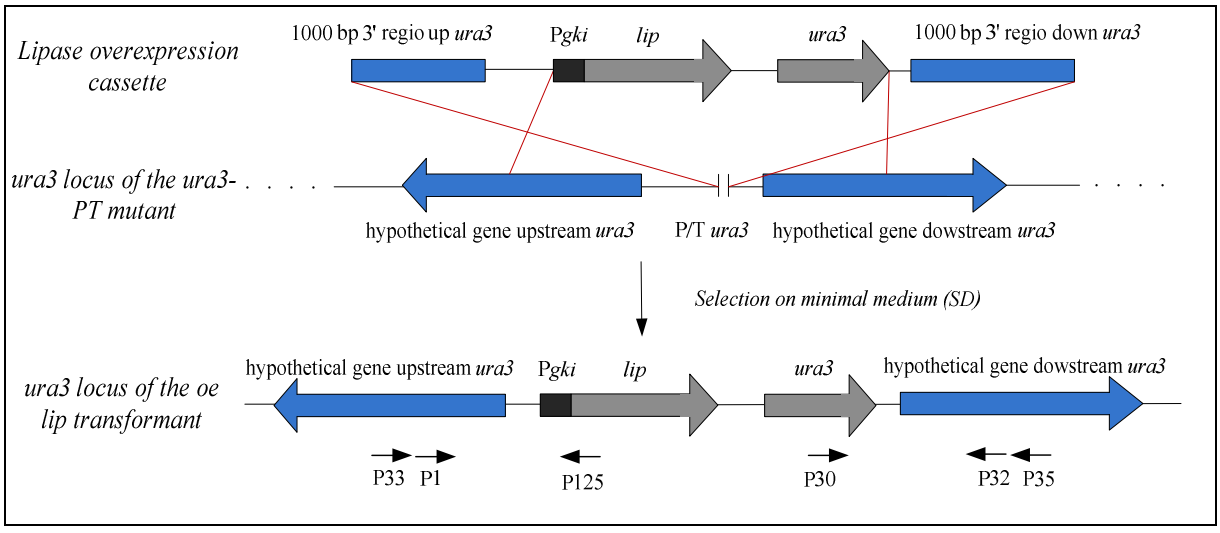


Figure VI. 1 Schematic representation of the lipase overexpression cassette and integration thereof at the *ura3* locus of an *ura3* negative auxotrophic mutant of *S. bombicola*. The strong *gki* (phosphoglycerate kinase) promotor was used to obtain overexpression of the lipase gene. The primers used for determining correct integration of the overexpression cassette in the genome of the obtained overexpression strain (oe lip) are depicted at the bottom of the figure.

VI.3.2. *Sophorolipid production by a lipase overexpression mutant*

VI.3.2.1. *On production medium without citrate*

Lactonisation is, as was stated, described to be highly dependent on the presence of citrate in the culture medium. The production medium used in this work (cfr. Chapter III, Table III.1), was developed to produce high amounts of lactonic di-acetylated SLs and contains citrate. Now that an overexpression strain of the lipase gene is available, the nature of these effects can be further investigated. By omitting citrate from the culture medium, one can f.e. conclude if citrate is needed to obtain optimal activity of the lipase enzyme or alternatively if regulation at the expression level exists. It was suggested by (Stüwer *et al.*, 1987) that the effects caused by citrate are just a consequence of the buffering action of the conjugate base of a weak acid. They showed that low pH values of the culture medium resulted in production of acidic SLs, thus the buffering effect of citrate was suggested to favour the formation of lactonic SLs merely because of higher pH values. Citrate on the other hand is a chelating agent, and metal ions like zinc or calcium could be necessary for the action of a hypothetical enzyme (lactonase) responsible for ringopening of lactonic SLs. The presence of citrate could thus prevent the action of a lactonase by sequestering its cofactors and as such favour the predominance of lactonic SLs.

To investigate this, a modified production medium without citrate was used for cultivation of the wild type and the two overexpression transformants. Important parameters like pH, glucose consumption and OD were followed and are depicted in Figure VI. 2. The results of the wild type and only one of overexpression transformants (the other one behaved identical) are shown to not overload the figure.

The observed pH drop was equal for both the wild type and the overexpression transformants. If differences in SL composition between the wild type and the overexpression strain thus exist, then the presumed pH effect on lactonisation can only be explained by a regulatory effect excerted by the pH integrated at the native lipase promoter, as all other factors are identical for both experiments/strains. pH values as low as 2.0 were reached, whereas the reached minimum was 3.1 for the normal production medium (cfr. VI.3.1). Some differences in the measured OD values were observed, but this is caused by interference of SLs and oil with the absorbance. Log CFUs were also determined before the cultures were stopped and these were equal to 8.1 ± 0.3 and 8.0 ± 0.2 logCFU/mL for the wild type and the

overexpression transformant respectively, so it can be decided that the wild type and overexpression transformants behave identical on this modified production medium.

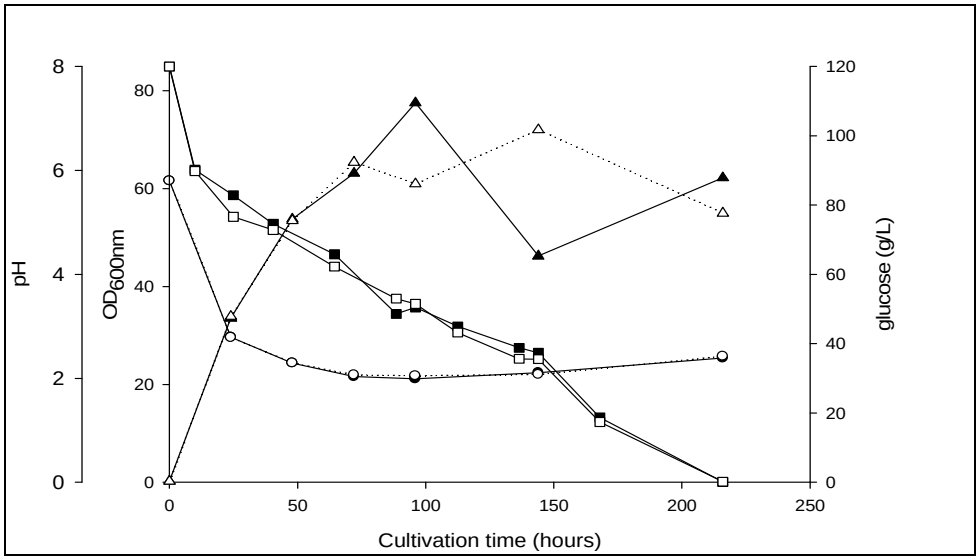


Figure VI. 2 Important parameters for growth and SL production of the S. *bombicola* wild type (filled symbols) and a lipase overexpression transformant (open symbols) cultivated on production medium w/o citrate. pH (○ and ●) glucose concentration (□ and ■) and optical density (Δ and ▲) are depicted in function of time.

SL biosynthesis was obviously also monitored and HPLC-ELSD chromatograms of samples taken after 6 days of incubation are depicted in Figure VI. 3. The samples were subsequently analysed on LC-MS to undeniably link the eluting peaks to specific compounds.

Overexpression of the lipase gene clearly leads to lactonisation of SLs in the absence of citrate: (di-acetylated lactonic SLs eluting at 27.6 minutes (C18:1) and 30 minutes (C18:0) and mono-acetylated lactonic SLs eluting at 24.3 minutes (C18:1)), whereas the wild type mainly produces acidic SLs (peaks eluting around 18 minutes) under these conditions. Some acidic forms are still present for the overexpression strain, though in marginal amounts as compared to the wild type strain.

These results clearly show that the influence of citrate on the ratio of lacton/acidic SLs is resulting from regulatory effects on the level of gene expression. Integration of an extra copy of the the lipase gene under control of the strong, constitutive P*gki* promotor namely leads to the production of lactonic SLs by the transformant in the absence of citrate. It could be argued that in the wild type only one copy of the lipase gene is present, whereas in the transformant

there are now two copies present in the genome. However, the observed effects are not expected to be the result of a mere extra copy of the gene in the genome, but rather of another copy under control of another promotor free from the imposed regulatory restrictions of the wild type lipase expression. These regulatory mechanisms could thus be avoided by simply placing an extra copy of the lipase gene under control of a strong constitutive promotor in the genome.

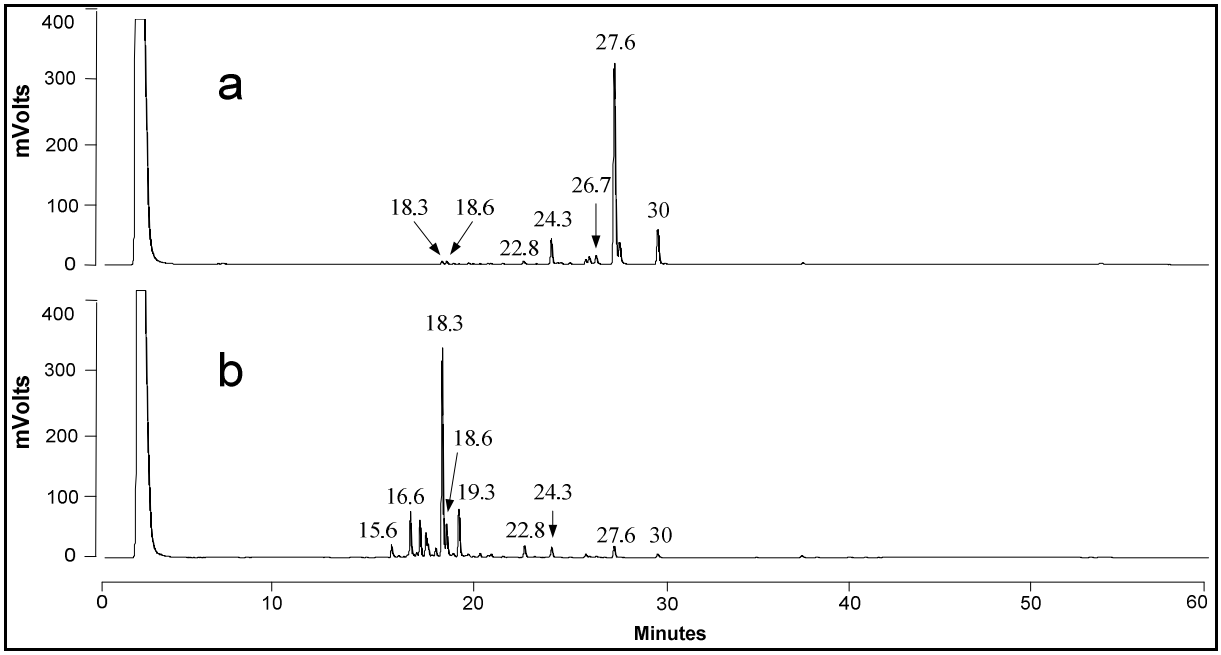


Figure VI. 3 HPLC-ELSD chromatograms of samples obtained from shake flask experiments of **(a)** a lipase overexpression transformant and **(b)** the wild type both cultivated on production medium w/o citrate after 6 days of cultivation. Samples were extracted with ethanol to not discriminate for certain kinds of SLs. Peaks around 18 minutes correspond to acidic SLs, peaks around 27 minutes to lactonic SLs

Results for total extractions of SLs of the shake flasks experiments are summarized in Table VI. 2. These results, averaged for three biological replicas each, suggest higher yields of SLs (different forms) for the overexpression transformant as compared to the wild type. The remaining amount of oil was substantially higher for the wild type, also suggesting better metabolisation of the added oil by the overexpression strain, which could subsequently also lead to higher yields of SLs. However, enzyme assays with the lipase enzyme did not demonstrate typical lipase activity of this enzyme (unpublished results). Alternatively a more efficient SL production for the overexpression strain (possibly because of lower end product inhibition caused by lower solubility of the lactonic SLs (cfr. Chapter III, Table III.3) could

result in a certain 'pull' on the biosynthetic pathway and as such indirectly cause enhanced oil metabolisation.

Table VI. 2 SL yields and residual oil after cultivation of the wild type and overexpression strain on the production medium w/o citrate. These results are the average of three independent biological replicas.

Strain	SL yield (g/L)	Residual oil (g/L)
Wild type	17 ± 5	9.0 ± 4.0
Overexpression strain	35 ± 13	0.1 ± 0.0

The variations on SL yields and residual oil are pretty high though. Shake flask experiments are known to be subject to high variability, so no real conclusions can be drawn here. Another remark that can be made is that the two end products are different for the two strains. A second experiment was thus set up with the normal production medium, optimized for production of lactonic SLs. This was also done to asses if overexpression of the enzyme responsible for lactonisation could possibly even lead to higher ratios of lactonic/acidic SLs in this optimized medium. It was also interesting to assess if the overexpression strain would indeed show higher SL productivity as compared to the wild type on this medium.

VI.3.2.2. *On regular production medium containing citrate*

The lipase enzyme is present in substantial amounts in the secretome of the wild type strain cultivated on production medium. Furthermore, the majority of the produced SLs under these conditions are the lactonic ones. Thus, overexpression of the gene could possibly not have any effect on the ratio lactonic/acidic SLs, when citrate is present in the medium and expression of the lipase gene is thus not hampered. Three biological replicas of both the wild type and a selected overexpression transformant were cultivated on the normal production medium and samples were extracted with ethanol and ethylacetate as described in the materials and methods section. HPLC-ELSD chromatograms are depicted in Figure VI. 4.

The peak eluting at 18.3 minutes in Figure VI. 4b corresponds to the non-acetylated acidic SLs, whereas the peak eluting at 27.6 minutes corresponds to the di-acetylated lactonic SLs. Surprisingly, overexpression of the lipase, in addition to the effects described in VI.3.2.1, also leads to a very clear effect on the SL composition on regular production medium as can be concluded when comparing Figure VI. 4 a and b. Whereas acidic SLs are still present in substantial amounts in the sample derived of the wild type, these compounds are not detected

for the overexpression transformant. The total SL yields for these experiments were again subject to very high variation and it was thus decided to perform a bioreactor experiment for which all cultivation parameters can be controlled, and more reliable results are expected to be obtained.

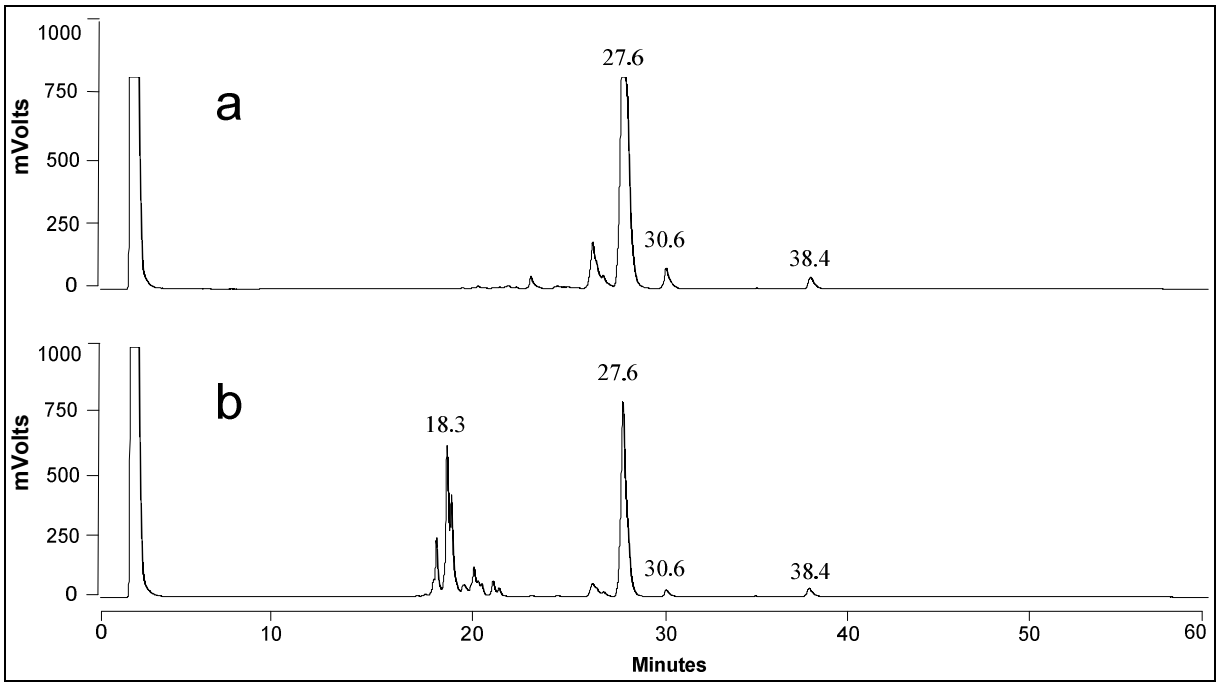


Figure VI. 4 HPLC-ELSD chromatograms of samples from a shake flask experiment with **(a)** a lipase overexpression strain and **(b)** the wild type both cultivated on normal production medium after 6 days of cultivation. Samples were extracted with ethanol to not discriminate for certain kinds of SLs.

VI.3.2.3. *Bioreactor experiment on standard production medium*

Four bioreactor experiments were run in parallel, providing two biological replicas for both the wild type and the lipase overexpression strain. All reactors were inoculated from separate shake flasks, which were in their turn inoculated from separate precultures. Typical parameters were monitored and are depicted in Figure VI. 5. Unfortunately, two bioreactor experiments behaved suboptimal, thus unfortunately no good biological replicas are available for these experiments. The results for the two succeeded experiments are shown in Figure VI. 5.

The pO_2 in the vessels dropped a lot faster for the overexpression strains as compared to the wild type (cfr. Figure VI. 5 c). This could be indicative of higher SL productivity as it was

demonstrated that higher oxygenation leads to higher SL production (oxygen is needed for the first step of SL biosynthesis) (Casas and Garcia-Ochoa, 1999; Guilmanov *et al.*, 2002; Kurtzman *et al.*, 2010). A higher oxygen 'consumption' could hence be indicative of higher productivity. This hypothesis is strengthened by the fact that typical indicators for growth like glucose consumption, pH drop (cfr. Figure VI. 5 a and b) and CFU values (not shown) were very similar for the wild type and the overexpression mutant, which indicates that the extra oxygen was not used for growth related processes. The initial pO_2 drop was also a lot more pronounced for the overexpression strain, which in combination with an earliand of SL production as SLs were indeed detected earlier detection of SLs in the culture medium, could indicate that SL biosynthesis is also inititiated earlier in the overexpression strain as compared to the wild type.

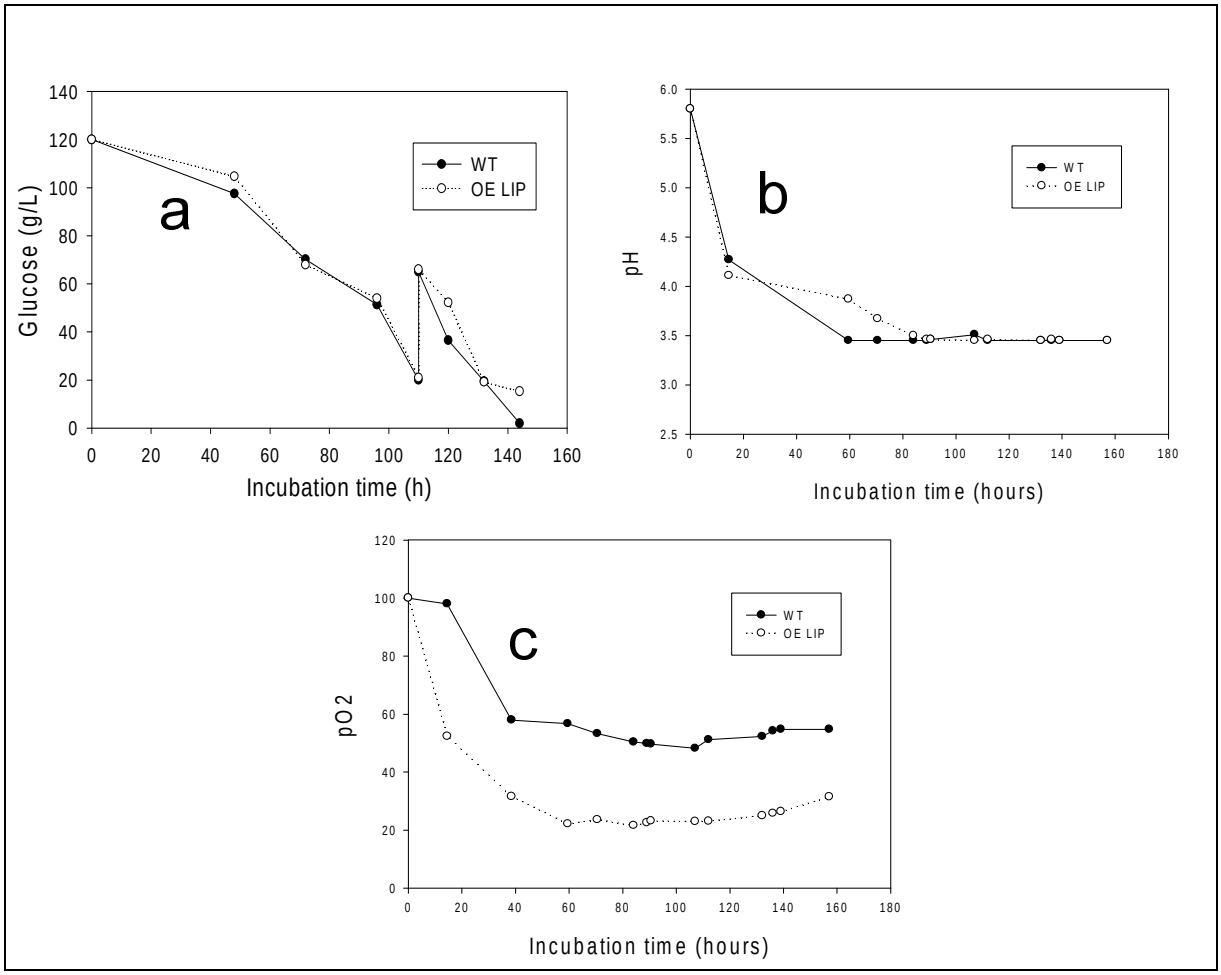


Figure VI. 5 Parameters of a bioreactor experiment with the *S. bombicola* wild type (filled symbols) and overexpression strain (open symbols) on production medium **(a)** glucose concentration **(b)** pH and **(c)** pO_2 in function of time.

The total SL yields of the succeeded wild type and overexpression strain cultivations respectively accounted up to 188.74 g (= 126 g/L) versus 213.42 g (= 142 g/L). This experiment should be repeated to confirm these results, but one can expect the productivity of the overexpression strain to be at least as high as that of the wild type. The abovementioned higher oxygen usage could even indicate that the productivity for this overexpression strain is indeed higher than this for the wild type. Sampling for SL extractions furthermore demonstrated that SLs were detected earlier for the overexpression strains and that the SLs produced by the latter were strongly enriched in lactonic SLs from the start of the cultivation (normally more towards the end (cfr. Chapter I)), which confirms the results obtained for the shake flask experiments (cfr. VI.3.2.2). This was also visible in the vessels as a faster appearing precipitate, corresponding to lactonic SLs. The chromatograms of two samples taken after five days of cultivation are depicted in Figure VI. 6.

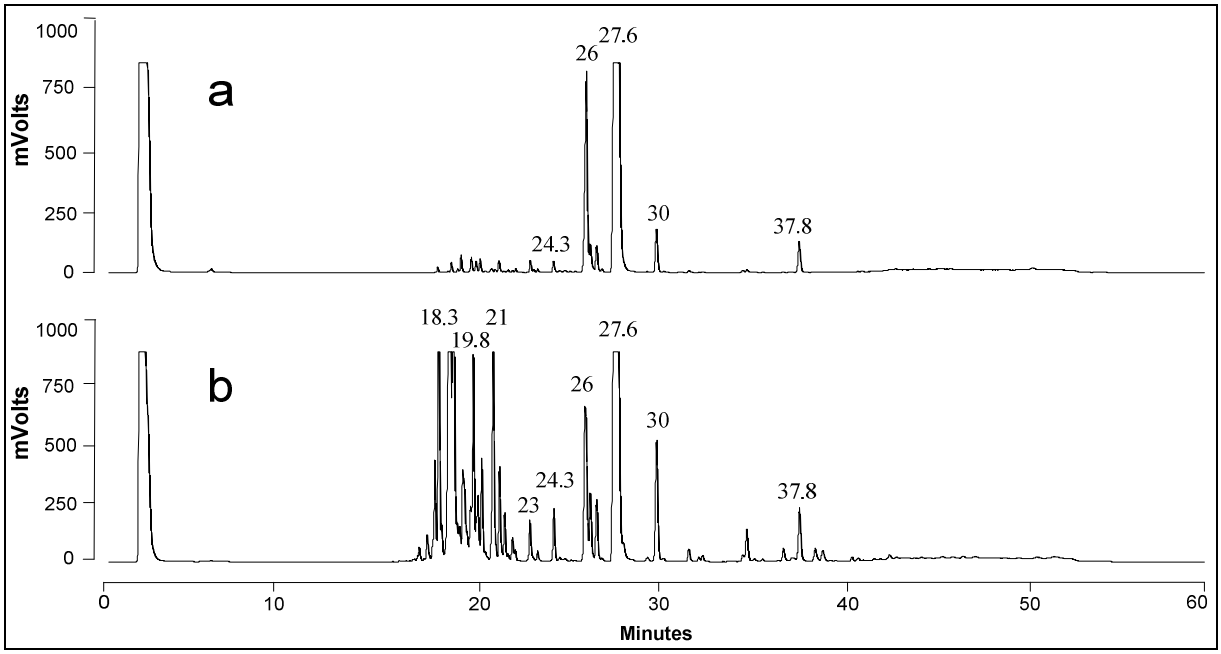


Figure VI. 6 HPLC-ELSD chromatograms of samples from a bioreactor experiment with **(a)** a lipase overexpression strain and **(b)** the wild type both cultivated on normal production medium after 5 days of cultivation. Samples were extracted with ethanol to not discriminate for certain kinds of SLs.

Once again it is clear that the overexpression strain produces a product that is a lot less complex than the one produced by the wild type, which is obvious when comparing Figure VI. 6 a with b.

Another last interesting phenomenon that has to be mentioned, was the absence of foam formation for the overexpression strains. Whereas for the wild type, considerable foam formation caused by the presence of solubilised acidic SLs, resulted in a constant threat of overfoaming, the overexpression strains never posed the threath to overfoam. This could be an additional advantage for industrial production processes where foam formation can be a real burden.

VI.4. Conclusions

The fact that there is a strong interest in industrial production and applications of SLs can be deduced from the strong patenting activity in the field (Shete *et al.*, 2006). The last years large companies like BASF, Evonik-Degussa, Unilever, Henkel, Cargill, etc. have initiated R&D projects focused on green alternatives of chemical surfactants. Numererous patent applications deal with applications of SLs in e.g. ecological cleaning solutions, dishwasher products, cosmetics, etc. It has been reported that the lacton form is necessary, or at least preferable, for many of these applications (Hu and Ju, 2001; Develter and Renkin, 2012). Moreover, separation of the lactonic forms from the acidic forms is necessary to obtain maximum effectiveness of the first compounds, which is mostly associated with long and expensive procedures involving the usage of solvents (Develter and Renkin, 2012). Alternative methods to obtain pure lactonic SLs were also developed and involve the use of buffers to alter the pH of aqueous systems, preferentially crystallizing the lactones (Hu and Ju, 2001). In this chapter the development of a very interesting strain exclusively producing these lactonic SLs was discussed. The respective strain carries an extra copy of the gene responsible for lactonisation, under control of a strong constitutive promotor (P*gki*). This strain furthermore allowed us to investigate the nature of certain regulatory effects on lactonisation in the wild type.

A medium component that has been described to have a profound effect on lactonisation in the wild type, is citrate. The effect of citrate was suggested to be a result of the buffering effect caused by this medium component by (Stüwer *et al.*, 1987). Low pH values, caused by the absence of citrate were stated to favor the formation of acidic SLs. When citrate wasn't supplied in the production medium of *S. bombicola*, this indeed resulted in pH values as low as 2 in contrast to values of 3.1 in the presence of citrate (cfr. Figure VI. 2). For the wild type this indeed resulted in the almost exclusive formation of acidic SLs (cfr. Figure VI. 3 b). However, for the overexpression strain the complete opposite was true as almost exclusively

lactonic SLs were detected (cfr. Figure VI. 3 a). Citrate must thus exert some effect on the native promotor of the lipase gene, as lactonic SLs were very efficiently formed in the absence of citrate by a strain containing an extra copy of the lipase gene under control of a strong constitutive promotor (P*gki*). These results thus reject possible regulatory effect of citrate at the enzyme level. Such effects could have included the sequestering of cofactors needed for the activity of an enzyme with lactonase activity and/or by allosteric effects exerted by citrate on the lipase enzyme. The regulatory effect of citrate on gene expression in the wild type can be direct, although an indirect effect e.g. exerted by the pH (as suggested by Stüwer et al. (1987)) is not unlikely as low pH values were shown to inhibit lactonisation in the wild type. The pH is furthermore known to influence the complete SL machinery as pH values of 5 significantly hamper high SL formation. In the light of these results, these effects can be suggested to be integrated at the promoters of the genes of the SL biosynthetic pathway. Alternatively end-product inhibition at these higher pH values (better solubility of SLs) could be the reason for this.

The development of a strain capable of producing almost 100 % lactonic SLs opens perspectives for direct production of lactonic SLs by fermentation, without the need for additional and expensive separation procedures. Such strain is of high industrial relevance as the SL machinery of the wild type was reported (and shown in this chapter) to be highly sensitive to changes of the medium composition and/or the cultivation conditions. The optimization effort required to obtain a product as pure as possible after cultivation, especially when *e.g.* using waste streams (which can vary in quality) can be estimated to be a lot lower when working with strains like the one described in this chapter. Such strain thus represents a realistic opportunity for the development of more robust production systems, which is expected to result in industrial production processes.

The higher lacton versus acidic ratio of SLs furthermore resulted in the absence of foam formation in bioreactor experiments. This represents an additional opportunity for the development of industrial production processes, where foam formation can be a real burden. Higher SL yields were obtained for the overexpression strain as compared to the wild type in several cultivation experiments. The possibility exist that better end-product removal due to precipitation of the lactonic SLs, results in a certain pull on the biosynthetic pathway, which gives rise to higher productivities. Moreover, SL production seemed to be initiated earlier, as SLs were detected earlier during the cultivation. The tight regulation of SL biosynthesis in the

wild type results in a sub-optimal production process, as 25% of the cultivation time is dedicated to growth without (or very low) production. The presumed earlier onset of SL biosynthesis in the overexpression strain could be an additional benefit resulting in a more economical production process.

In conclusion we have created a new SL producing *S. bombicola* strain, which is capable of producing highly pure lactonic SLs, without the need of further purification. Such excellent producer, giving rise to the efficient, robust and easy production of a specific compound of very high purity, with the additional technical benefit of low foam formation is expected to be a major advantage for industrial SL production processes and applications.

CHAPTER VII:

PRODUCTION OF THE BIOPLASTIC PHA WITH *STARMERELLA BOMBICOLA*

Part of this chapter has been accepted for publication:

Roelants, S., Saerens, K., Derycke, T., Li, B., Lin, Y.-C., Van de Peer, Y., De Maeseneire, S. L., Van Bogaert, I., Soetaert, W., 2013. *Candida bombicola* as a platform organism for the production of tailor-made biomolecules. Biotechnology and Bioengineering. (accepted and in production)

Chapter VII.

PRODUCTION OF THE BIOPLASTIC PHA WITH *STARMERELLA BOMBICOLA*

VII.1. Introduction

As mentioned before, one of the goals of the project to which this thesis contributed was to genetically engineer the highly productive sophorolipid producing yeast *S. bombicola* with the aim of transforming it into a platform organism for the efficient production of tailor-made (glycolipid) biomolecules. One could for example try to substitute the fatty acid tail with shorter/longer/branched variants and/or replace the hydrophilic sophorose head with another sugar molecule. Attempts have been done to alter the structural variety of the produced SLs by *S. bombicola* and *C. apicola* by feeding different sugars (Gobbert *et al.*, 1984; Klekner *et al.*, 1991) or fatty acids with a shorter or longer fatty acid tail (Van Bogaert *et al.*, 2010). This did not result in the production of altered structures though. This is partly attributed to the fact that *S. bombicola* produces the sophorose moiety of SLs *de novo* (Tulloch and Spencer, 1968) in combination with the high donor and acceptor specificities of the two glucosyltransferases involved in SL biosynthesis (Saerens, 2012). When aiming for the synthesis of such tailored glycolipid molecules, the introduction of heterologous enzymes with a differing substrate specificity or enzyme engineering of the native ones will hence be indispensable. The most obvious choice are the enzymes involved in the biosynthesis of cellobioselipids (CBLs) or rhamnolipids (RLs) (cfr. Chapter I), as the biosynthetic pathways for these glycolipids are described in detail and the responsible gene sequences are available.

RLs have, like SLs, been described as one of the biosurfactants which represent significant opportunities to replace chemical surfactants with sustainable biologically produced alternatives (Marchant and Banat, 2012). However, efficient production suitable for industrial production processes has only been obtained with the opportunistic human pathogen *Pseudomonas aeruginosa*. This limits penetration of RLs in certain sectors and brings forth safety issues for industrial RL production. These issues in combination with the fact that the regulation of RL production is tightly controlled already resulted in several attempts for the heterologous production of RLs (cfr. Chapter I). These attempts either did not succeed or alternatively resulted in low titers.

The heterologous production of RLs by *S. bombicola* was thus considered, as this yeast is a highly productive glycolipid producer, which is moreover resistant to high titers of such molecules (SLs). However, to obtain this goal, extensive genetic engineering of *S. bombicola* will be required. Since only a limited amount of molecular tools were available at the start of this project, it was considered wise to ensure the fact that *S. bombicola* has the potential to be transformed into a platform organism for the heterologous production of interesting biomolecules.

RLs are build up out a rhamnose sugar head coupled to a dimer of β-hydroxyfatty acids (cfr. Figure VII. 1 a). The latter are also the building blocks of polyhydroxyalkanoates (PHAs) (cfr. Figure VII. 1 b), which are bioplastics produced by a wide variety of microorganisms, mainly prokaryotes (Doi *et al.*, 1992; McInerney, 1992; Steinbuchel and Wiese, 1992; Follner *et al.*, 1995; Lee *et al.*, 1995; Lee, 1996; Verlinden *et al.*, 2007). The universal term 'PHA' refers to several types of polymers with varying properties. Short chain length (SCL) PHAs have thermoplastic properties and generally find applications as the ecological alternative for classic commodity plastics. Medium chain length (MCL) and long chain length (LCL) PHAs have elastomeric properties. These last molecules are not commercialized yet, but may find applications in the biomedical and/or pharmaceutical industry. Heterologous production of SCL and MCL PHA has already been achieved in other yeasts such as *Saccharomyces cerevisiae* (Leaf *et al.*, 1996), *Yarrowia lipoliptica* (Haddouche *et al.*, 2010; Haddouche *et al.*, 2011), *Pichia pastoris* (Poirier *et al.*, 2002) and *Arxula adeninivorans* (Terentiev *et al.*, 2004).

We aimed for the production of PHA in the nonconventional yeast *S. bombicola* as one example of the possible genetic engineering of this yeast into a platform organism. Because the production of LCL PHA would have additional scientific and industrial relevance and because the yeast grows well on LCL lipogenic substrates such as rapeseed oil, this was set as an additional goal.

RL biosynthesis by *S. bombicola* was still kept in mind, so toxicity tests with RLs were performed to ensure that production of these molecules by this yeast would not be hampered by toxic effects imposed by the end-products.

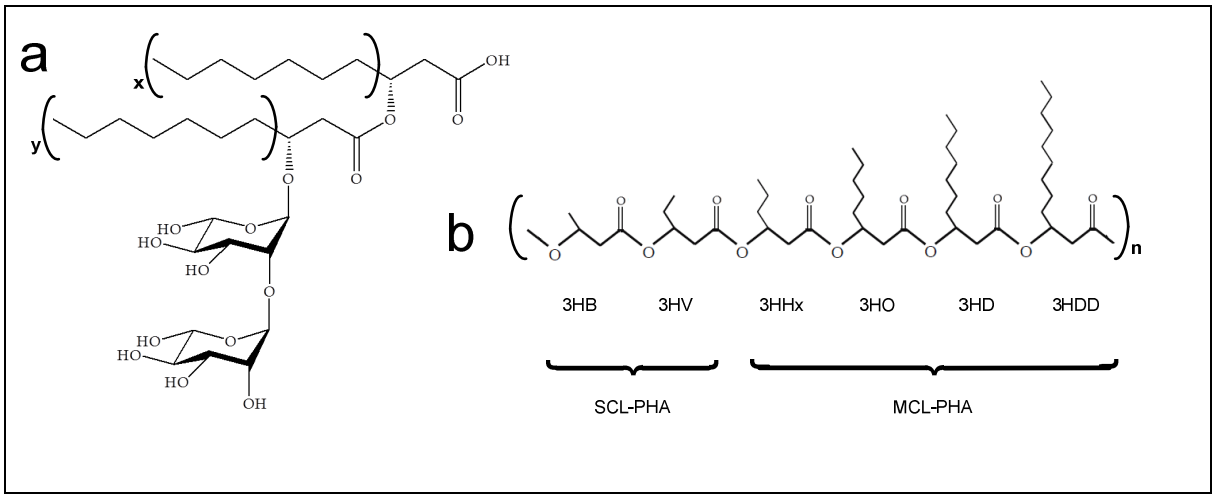


Figure VII. 1 Structures of **(a)** dirhamnolipid consisting of a dimer of fatty acids, the chain length of x and y varies between 5 and 9 C-atoms and **(b)** polyhodroxyalkanoates (PHA), n = 10^3-10^4, SCL = short chain length, MCL = medium chain

VII.2. Materials and methods

VII.2.1. *Strains and culture conditions*

Starmerella bombicola ATCC 22214 and a Δ*cyp52*M1 deletion strain defective in SL biosynthesis (Van Bogaert *et al.*, 2013) were used as the reference strains. The *ura3* negative auxotrophic PT36 strain (cfr. Chapter IV) was used as the parental strain to transform the *phaC1co* expression cassette. Yeast cells were grown on SD medium (0.67 % yeast nitrogen base without amino acids (DIFCO) and 2% glucose), YPD medium (1 % yeast extract, 2 % peptone and 2 % glucose), 3C medium (10 % glucose, 1 % yeast extract, 0.1 % ureum and 15 % agar) or the production medium described by Lang (Lang *et al.*, 2000). Lipogenic substrates were added to the culture medium after 48h. A modified version of this production medium with only 0.2 % or 2 % of glucose (instead of 12 %) was also used to perform growth experiments. Yeast cultures were incubated at 30 °C on a rotary shaker (200 rpm). Precultures were prepared from single colonies derived from 3C plates in 5 mL of the production medium in round bottomed culture tubes. These overnight grown precultures were inoculated (2 %) in 100 mL of production medium. The used lipogenic substrates were rapeseed oil (Sigma; 3.75 %), palm oil (Sigma; 3.75 %) and coconut oil (Supelco; 1.0 %).

VII.2.2. *Molecular techniques*

Standard molecular techniques are described in Chapter II. 2.2.1.

VII.2.2.1. *Creation of the transformants*

The *phaC1co* expression cassette was integrated in the *S. bombicola* genome using a knock out/knock in strategy at the locus of the gene responsible for the first step of the SL biosynthetic pathway (*cyp52M1*) (Van Bogaert *et al.*, 2013) (cfr. Figure VII. 2). Expression cassettes were transformed into an auxotrophic *S. bombicola* strain, so the expression cassette contained a functional *ura*3 gene (Van Bogaert *et al.*, 2008a) as a selectable marker. Transformants were selected on SD plates and correct integration was verified by yeast colony PCR.

VII.2.2.2. *Synthetic constructs*

The PhaC1 protein sequence from *Pseudomonas resinovorans* (ENA accession number AAD26365.2) was backtranslated using the averaged codon usage of the genes of the SL biosynthetic pathway (Genbank Accession numbers HM440974.1 (UGTB1), HQ670751 (AT), HM440973 (UGTA1), EU552419 (CYP52M1) and HQ660581 (MDR SL TRANSPORTER)), which was determined using an online tool (Stothard, 2000). An SKL (TCTAAGCTG) peroxisomal target sequence (PTS) found in at least one peroxisomal enzyme of *S. bombicola* (Van Bogaert *et al.*, 2009b) was added to the 3' terminus of the codon-optimised *phaC1*co sequence. Additionally the up- and downstream regulatory regions of the *cyp52M1* gene were added to the 5' (488 bp) and 3' (190 bp) side of the gene respectively. The 5' UTR regio (488 bp) was extended to 1098 bp to obtain the 1000 bp of homology needed for homologous recombination at the *cyp52M1* locus. The construct was ordered at GenScript (Piscataway, USA) and obtained cloned into a pUC vector. The sequence can be found in the appendices. The *phaC1*co construct was amplified from the obtained vector using primerpair P55_FOR_UP*cyp*_ext*Nhe*I / P58_REV_*phaC1co*T*cyp*_ext*EcoR*I yielding a fragment of 2986 bp. The primers respectively contained *Nhe*I and *Eco*RI extensions, allowing subsequent digestion with said restriction enzymes for further subcloning of the synthetic construct (cfr. VII.2.2.3).

VII.2.2.3. *Creation of the PHAC1co expression cassette*

All primers used for generation of the construct can be found in Table VII. 1. Construction of the cassette occurred in three steps:

First the *ura3* marker gene with its 5' and 3' regulatory sequences (1970 bp) was amplified from genomical DNA of *S. bombicola* ATCC 22214, using primers P1_FOR_*ura3*v and P2_REV_*ura3*v. This amplicon was subsequently cloned into the pGEM-T ® vector (Promega) yielding pGEM-T_*ura3*. Secondly, the region for homologous recombination at the 3' side of the *cyp52M1* locus (down*cyp*) was amplified from genomic DNA of *S. bombicola* using primers P53_FOR_down*cyp*_ext*Spe*I and P54_REV_down*cyp*_ext*Not*I. The resulting amplicon was digested with the unique cutters *Spe*I and *Not*I as was the pGEM-T_*ura3* vector and the restriction fragments were ligated after purification, yielding pGEM-T_*ura3*_down*cyp*. A third and last step consisted of introduction of the synthetic construct into the expression cassette. pGEM-T_*ura3*_down*cyp* was thus digested using the unique cutters *Nhe*I and *Eco*RI. This double restriction digest yielded two fragments (5644 bp and 358 bp) of which the largest one was gel purified and subsequently ligated with the digested, amplified synthetic construct (cfr. VII.2.2.2). The *phaC1co* expression cassette was amplified from the resulting vector, pGEM-T_cassette_*phaC1co* using primers P63_FOR_cass*phaC1co* and P64_REV_cass*phaC1co*. The resulting linear PCR fragment was purified and used for transformation of the *S. bombicola* PT36 strain (cfr. Chapter IV). Transformants were selected on SD plates. Correct integration of the *phaC1co* expression cassette into the genome was controlled by colony PCR with primerpairs UDPGTA1R/P58 and P9/A21TotRev. Additionally, gDNA was isolated from three of the obtained transformants and PCR was performed on this gDNA with primers further up- and downstream than the primers used for amplification of the expression cassette: UDPGTA1R and A21TotRev respectively. The resulting PCR product was purified and sent to LGC genomics (Germany) for sequence analysis.

Table VII. 1 Primers used for the creation of the *phaC1co* expression strains

Primer names	**'5 Sequence 3'**
P1_FOR_*ura3*v	AGAACAAGGCCGAGTATGTC
P2_REV_ *ura3*v	TGCCAGCAGATCATCATCAC
P53_FOR_down*cyp*_ext*Spe*I	TTACTAGTGTTTCTTAGCCTCCCATGGAAGAAACG
P54_REV_down*cyp*_ext*Not*I	AATTGGCCTTGCGGCCGCGGTGTCGACTCGCCAAATTCCATC
P55_FOR_up*cyp*_ext*Nhe*I	GTTGCTAGCTCTCGGCAGATTTCCTTG
P58_REV_PHAC1coT*cyp*_ext*Ecor*I	AGAATTCGTCGGTTAAACGCACTCCTTCA
P63_FOR_cass*phaC1co*	CTCTCGGCAGATTTCCTTGTG
P64_REV_cass*phaC1co*	GGTGTCGACTCGCCAAATTC
P9_FOR_seqQC*SapI_ura3*down	GCACACTTCAACCTTCCTAC
A21TotRev	GCTCTTGTTCGGTACTCTTATTG
UDPGTA1R	CCTACCTCTCTTCCCTGATCT

VII.2.3. *Sampling and analysis*

Cell growth and SL synthesis were monitored as described in previous chapters. For PHA analysis cell pellets were harvested (4000 rpm, 4°C) after 12 days of cultivation, frozen at -80 °C and subsequently lyophilized (24 h). 100 mg of the resulting freeze-dried material was washed several times with hot methanol (65 °C) to remove oil, free fatty acids, and acyl-CoA, including 3-hydroxyacyl-CoA. PHA hydrolysis and fatty acid methyl ester (FAME) formation were performed using acidic methanolysis of 100 mg of freeze-dried cell material in a 4 ml 1:1 chloroform/methanol + 3 % H_2SO_4 mixture at 95 °C for 4h. After methanolysis 4 ml of 0.9 % (wt/vol) NaCl was added to the tubes and the resulting solution was centrifuged (4000 rpm, 4 °C) after vigorous shaking. The organic phase was subsequently collected for analysis on GC-MS. 1 mg of internal standard (2-hydroxyhexanoic acid) was added to the samples before methanolysis to allow quantification of the produced PHA. Analysis of the obtained FAMES was performed using a GC (TraceGC ultra, Interscience) operated with a 0.25 mm RxiR ml column (Restek) connected to a DSQ MS (Interscience). The used carrier gas was helium. The following temperature profile was used: 2 minutes at 64 °C followed by a linear increase of 30 °C/min to 200 °C. When the column reached 200 °C a second linear increase of 50 °C/min to 310 °C took place. Identification of compounds was done using the Xcalibur software which was coupled to the NIST MS Search 2.0 bibliotheca.

VII.2.4. *RNA sequencing*

RNA was extracted from samples taken after 98 hours of cultivation of the *S. bombicola* wild type on production medium (12 % glucose versus 0.1 % of glucose) using the RNA maxiprep kit (Qiagen).

Strand-specific sequencing was performed by GATC Biotech AG (Konstanz, Germany) on an Illumina HiSeq 2000. Eight tagged libraries were combined resulting in 180.000.000 single reads of about 50 bp. Mapping of the sequences to the genomic sequence and data analysis was performed at the laboratory of Professor Yves Van de Peer (Bioinformatics & Systems Biology, VIB) by Bing Li. Expression of the genes is presented as Isoform-level relative abundance in Fragments Per Kilobase of exon model per Million mapped fragments (FPKM). To compare the expression between two conditions, the log2 ratio FPKM1/FPKM2 was calculated with the Cufflinks software. The q-value is the corrected p-value of the statistic test. Two biological replicas and respective RNA sequencing results were analyzed.

VII.2.5. *Toxicity of rhamnolipids towards Starmerella bombicola*

Toxicity tests were performed with the *S. bombicola* wild type on 3C agar plates supplemented with 20 g/L of RLs by plating 20 µL of two different dilution series (10^{-1} to 10^{-8}) derived from either a culture on production medium in the exponential growth phase (30 h) or the stationary growth phase (96 h). Controls consisted of plating these dilution series on normal 3C plates. The toxicity was also assessed in liquid cultures on 3C medium to which RLs were added in two end-concentrations (5 g/L and 20 g/L). The effects were again tested for exponentially growing cells (RLs present form the start) and stationary cells (RLs added after 96 hours of cultivation).

To prepare the media and the plates, the RLs were added to culture media and the resulting solutions were filter sterilized. Colony forming units were determined as a measure of cell viability throughout the cultivation as described in previous Chapters and the pH was also monitored.

A second set of experiments was set up to assess possible uptake/catabolism of RLs by *S. bombicola*. This was done with the *S. bombicola* wild type and the Δ*cyp52M1* deletion strain. Samples were taken for HPLC-ELSD and LC-MS analysis and the standard techniques for SL extraction and analysis were used for RL analysis (cfr. Chapter VI).

The used rhamnolipids were a kind gift from Professor I. Banat.

VII.3. Results and discussion

VII.3.1. *Production of the bioplastic PHA with Starmerella bombicola*

VII.3.1.1. *Creation of a phaC1co synthase expression strain*

Few publications mention incorporation of long chain length (LCL) intermediates in the microbially produced PHA. Incorporation of C16 monomers into PHA was reported for *Pseudomonas resinovorans* grown on soy bean oil (Ashby and Foglia, 1998), thus indicating broad(er) substrate specificity of the involved PHA synthase. Since heterologous expression of a non-codonoptimized *gfp* gene in *S. bombicola* (cfr. Chapter IV) did not result in expression of the protein, the expression of the PHA synthase of *P. resinovorans* was preceded by performing codon optimization of the corresponding gene. Furthermore, peroxisomal targeting of the PHAC1co synthase, which is necessary because this protein uses intermediates of the peroxisomal β-oxidation for PHA synthesis (cfr. Figure VII. 5), was assured by adding the peroxisomal targeting signal to the 3' terminus of the gene. Reuse of the marker (cfr. Chapter IV) enabling further genetic engineering was attained by placing the *ura3* marker between two repeated sequences (T*cyp*). Since production of PHA was intended instead of production of the naturally produced SLs, a knock out/knock in strategy was developed, resulting in production of PHA whilst disabling SL production by a single recombination event at the *cyp52M1* locus (cfr. Figure VII. 2). The former encodes a P450 monooxygenase responsible for the first step of the SL biosynthetic pathway (Van Bogaert *et al.*, 2009a). Knocking out this gene results in a strain fully deficient in SL production (Van Bogaert *et al.*, 2013). Homologous recombination of the expression cassette at the *cyp52M1* locus puts the codon optimized *phaC1co* gene under control of the 5' and 3' regulatory sequences of the *cyp52M1* gene. Expression and PHA production should thus occur under conditions when SL production occurs in the wild type. As this predominantly happens in the stationary growth phase, the averaged codon usage of the genes of the SL pathway was used to backtranslate the PhaC1 synthase protein sequence from *P. resinovorans*.

The *phaC1co* expression cassette was constructed as described in VII.2.2.3. In total 16 colonies appeared on the selective SD plates after 4 to 11 days of incubation. The genotype of these 16 transformants was assessed by yeast colony PCR. 10 out of 16 displayed the correct genotype after this first colony PCR. Two additional colony PCRs were performed on the positive colonies to control for correct integration of the cassette at both sides of the

integration site. Genomic DNA of three of the positive transformants (A2, A5 and A8) was subsequently isolated and a PCR fragment containing the complete expression cassette was amplified from it and sent for sequence analysis. Sequence analysis revealed that the *phaC1co* expression cassette was correctly and completely integrated at the *cyp52M1* locus for all three selected transformants.

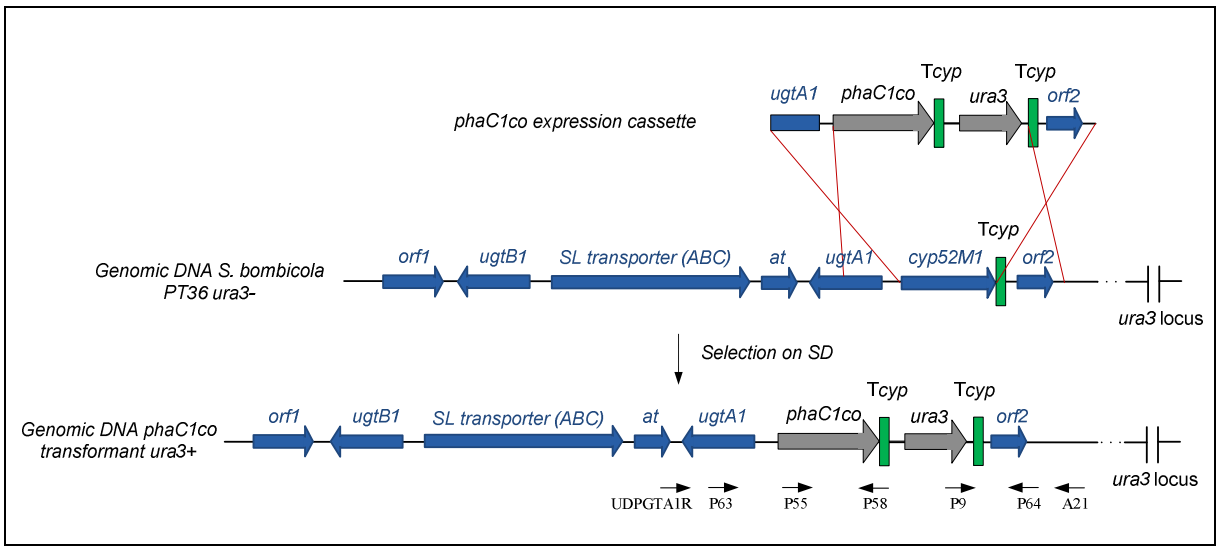


Figure VII. 2 Integration of the linear *phaC1co* expression cassette at the *cyp52M1* locus of the SL biosynthetic gene cluster (cfr. Chapter II) of an *ura3- S. bombicola* strain (PT36). The *ura3* marker is placed between two repeated sequences consisting of the *cyp52M1* terminator (T*cyp*) enabling reuse of the *ura3* marker (cfr. Chapter IV). The most important used primers are depicted at the bottom of the figure.

VII.3.1.2. *Characterization of the phaC1co transformants*

To assess if these obtained transformants behave similar as the wild type, they were grown on the production medium described by (Lang *et al.*, 2000). This experiment was performed with and without the addition of rapeseed oil after 48 h of cultivation. The wild type *S. bombicola* strain and *Δcyp52M1* mutant were grown in parallel as controls and treated identically. Samples for SL extraction and pH measurement were taken throughout the cultivation as were samples for glucose consumption and CFU determination and results are depicted in Figure VII. 3. The three *phaC1co* transformants (only the results for transformant A8 are shown), showed identical behavior as the *Δcyp52M1* deletion strain (not shown). However, likewise as for this last deletion strain, glucose consumption in the stationary phase was slower for the *phaC1co* expression transformant as compared to the wild type. Glucose was depleted from the medium after nine days of cultivation for the wild type versus twelve days for the

phaC1co transformant and Δ*cyp52M1* mutant, which is a typical observation for strains knocked out in one of the SL biosynthetic genes and can be explained by the lack of SL production (Saerens *et al.*, 2011a). Expression of a PHA synthase in a *Δcyp52M1* background thus does not trigger glucose utilization.

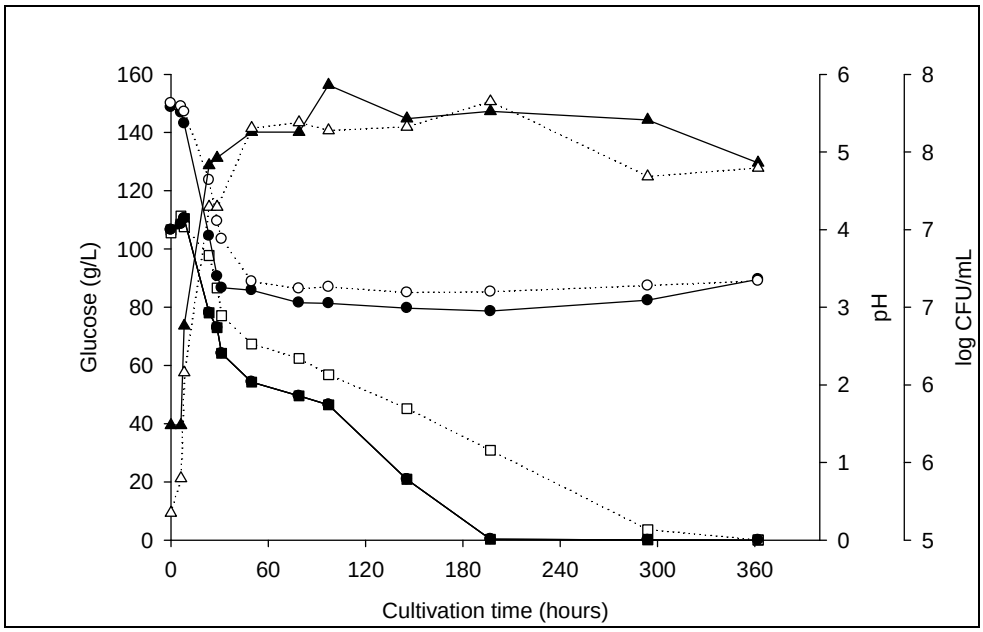


Figure VII. 3 Important parameters for growth of the S. *bombicola* wild type (open symbols) and a *phaC1co* expression transformant (filled symbols) cultivated on production medium. pH (○ and ●) glucose concentration (□ and ■) and log/CFU value's (Δ and ▲) are depicted in function of time.

VII.3.2. *PHA synthesis on production medium*

Next, we assessed if the transformants were capable of producing PHA instead of SLs. Whereas there clearly was SL production for the wild type, no SLs were detected for the *phaC1co* expression transformants, whether oil had been added to the cultures or not. GC-MS analysis of the FAMES derived from the *S. bombicola phaC1co* A8 transformant, to which rapeseed oil was added, indeed led to the identification of monomers derived from MCL-PHA. The transformant thus successfully produced PHA instead of SLs. The five compounds derived from PHA, which were detected in three biological replicas and their mol % of total PHA are: 3-methylhydroxyhexanoate (26 mol %), 3-methylhydroxyoctanoate (33 mol %), 3-methylhydroxydecanoate (24 mol %) 3-methylhydroxydodecanoate (10 mol %) and 3-methylhydroxytetradecanoate (8 mol %) with a total amount of 2.0 % ± 0.3 % wt/dwt PHA (cfr. Table VII. 2 and Figure VII.4).

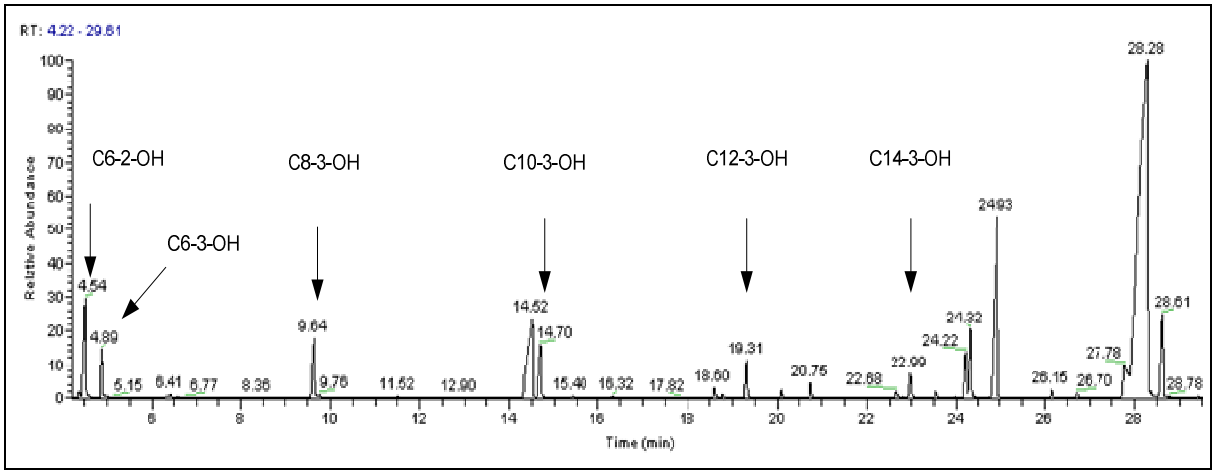


Figure VII. 4 GC-MS analysis of a sample derived from a *phaC1co* transformant (A8) grown on production medium with addition of rapeseed oil after 48h of cultivation. 3-OH monomers derived from MCL-PHA were identified and are visible in the chromatogram: 4.89 min (C6-3OH), 9.64 min (C8-3OH)), 14.70 (C10-3OH), 19.31 min (C12-3OH) and 22.99 (C14-OH). The peak at 4.54 min is derived from the internal standard (C6-2OH), which was used to calculate the amount of produced PHA.

These compounds were not detected in the samples derived from the wild type or Δ*cyp52M1* cultures, nor in the samples derived from the *phaC1co* A8 transformant to which no rapeseed oil was added. It can therefore be concluded that the introduced codon optimized *phaC1co* synthase gene under control of the *cyp52M1* regulatory sequences gives rise to a functional Pha synthase, effectuating PHA synthesis in *S. bombicola*. It can also be concluded that rapeseed oil is required for PHA production in *S. bombicola*. This is most likely attributed to the fact that a substantially productive β-oxidation is required to provide enough β-hydroxylated fatty acid building blocks for the peroxisomally located PhaC1 synthase. Although, futile cycling from fatty acid biosynthesis to β-oxidation was reported for yeast (Marchesini and Poirier, 2003), these marginal amounts are most likely not sufficient for (detectable) PHA production to occur in the absence of oil in the culture medium. It could also be argued that regulatory effects excerted by the oil and/or liberated fatty acids on the *cyp52M1* regulatory sequences are needed for substantial promotor activity and thus synthesis of the PhaC1 protein. Low amounts of PhaC1 could then explain the lack of PHA detection when no oil was added to the culture medium. However, SL production in the wild type (requiring *cyp52M1* transcription) does occur when no oil is added to the medium, hence the presence of oil is not an absolute requirement for transcriptional activity. Indeed, recent qPCR expression data confirm that the presence of oil does not lead to an additional upregulation of

cyp52M1 gene expression in the stationary phase in the wild type (cfr. Chapter V). It can thus be conluded that the first hypothesis explains the lack of PHA formation

PHA production and composition thereof for two other lipogenic substrates; palm oil and coconut oil, were evaluated in a second experiment of which the results are shown in Table VII. 2.

Table VII. 2 PHA yields (wt/cdw %) after cultivation of the *phaC1co* transformant (A8) on the production medium with addition of several lipogenic substrates and mol % of the respective PHA monomers.

	Rapeseed oil	**Palm oil**	**Coconut oil**
PHA monomer	(mol %)	(mol %)	(mol %)
C6-3OH	26	23	19
C8-3OH	32	29	41
C10-3OH	24	29	25
C12-3OH	10	12	14
C14-3OH	8	7	1
TOTAL	**2.0 ± 0.3**	**2.2 ± 0.2**	**1.3 ± 0.1**

The obtained yields are higher than those reported for other yeasts equipped with a peroxisomal PHA synthase, without further metabolic engineering of the strains; these reported yields are equal to or lower than 1 % wt/dwt (Poirier *et al.*, 2002; Marchesini *et al.*, 2003; Haddouche *et al.*, 2010).

The presented results on one hand reflect the specificity of the PhaC1co synthase. This can be concluded when comparing the fatty acid composition of the used substrates (cfr. Table VII. 3) with the composition of the PHA produced by the *phaC1co* transformant cultivated in the presence of these substrates (cfr. Table VII. 2). As no LCL fatty acids are detected in the produced PHA, one can conclude that LCL fatty acids entering β-oxidation are only considered as substrate for the introduced PhaC1co synthase when they have undergone several rounds of β-oxidation. The composition of the produced PHA can on the other hand be controlled by smart selection of the used substrate(s), as can be concluded by comparing the composition of PHA derived from coconut oil with the PHA composition derived from the LCL substrates (palm oil and rapeseed oil) (cfr. Table VII. 2). Whereas C14 monomers are still present in considerable amounts in PHA derived from growth experiments with LCL substrates, they are only present in marginal amounts in PHA derived from the cultures fed with coconut oil. Unfortunately, no C16 and/or C18 monomers could be detected even though the composition of the rapeseed and palm oil is enriched in LCL fatty acids and C16 PHA

monomers were detected by Ashby *et al*. (1998) when growing *P. resinovorans* on soy bean oil (comparable fatty acid composition as palm oil). The substrate specificity of the used PhaC1co synthase of *P. resinovorans* is apparently either not broad enough to incorporate these 3-OH fatty acid monomers or the introduced amounts in *S. bombicola* are below the detection limit. The results of Ashby *et al.* could hence not be transferred to *S. bombicola* by expressing the Pha synthase of *P. resinovorans* in this host.

Table VII. 3 Fatty acid composition (%) of used oils.

Fatty acid	Rapeseed oil	Palm oil	Coconut oil
C6	/	/	0 - 8
C8	/	/	5 - 9
C10	/	/	6 - 10
C12	/	/	44 - 52
C14	0 - 0.1	0.5 - 2	13 - 19
C16	3 - 5	32 - 45	8 - 11
C18:0	1 - 7	2 - 7	1 - 3
C18:1	54 - 63	38 - 52	5 - 8
C18:2	20 - 30	5 - 11	0 - 1
C18:3	7 - 9	/	/

VII.3.3. *PHA synthesis on production medium with modified glucose concentrations*

- PHA production experiments

The production medium used for the abovementioned growth experiments was optimized for high SL production and therefore contains a lot of glucose (12 %). However, this high glucose concentration can have a repressing effect on the enzymes of the β-oxidation (Kolkman *et al.*, 2006), which on the other hand are needed to supply the building blocks for PHA production (cfr. Figure VII. 5). It was thus expected that PHA yields could be elevated if β-oxidation would not be repressed. A second experiment was thus set up in conditions with modified glucose concentrations (2 % and 0.2 %) and addition of rapeseed oil as this was needed for PHA production. Cells were harvested 10 days after addition of oil.

The result was somewhat surprising since PHA yields were substantially lower, more specifically 0.35 % wt/dwt and 0.05 % wt/dwt for 2 % and 0.2 % glucose respectively in comparison with 2.0 % wt/dwt for the original production medium (12 % glucose).

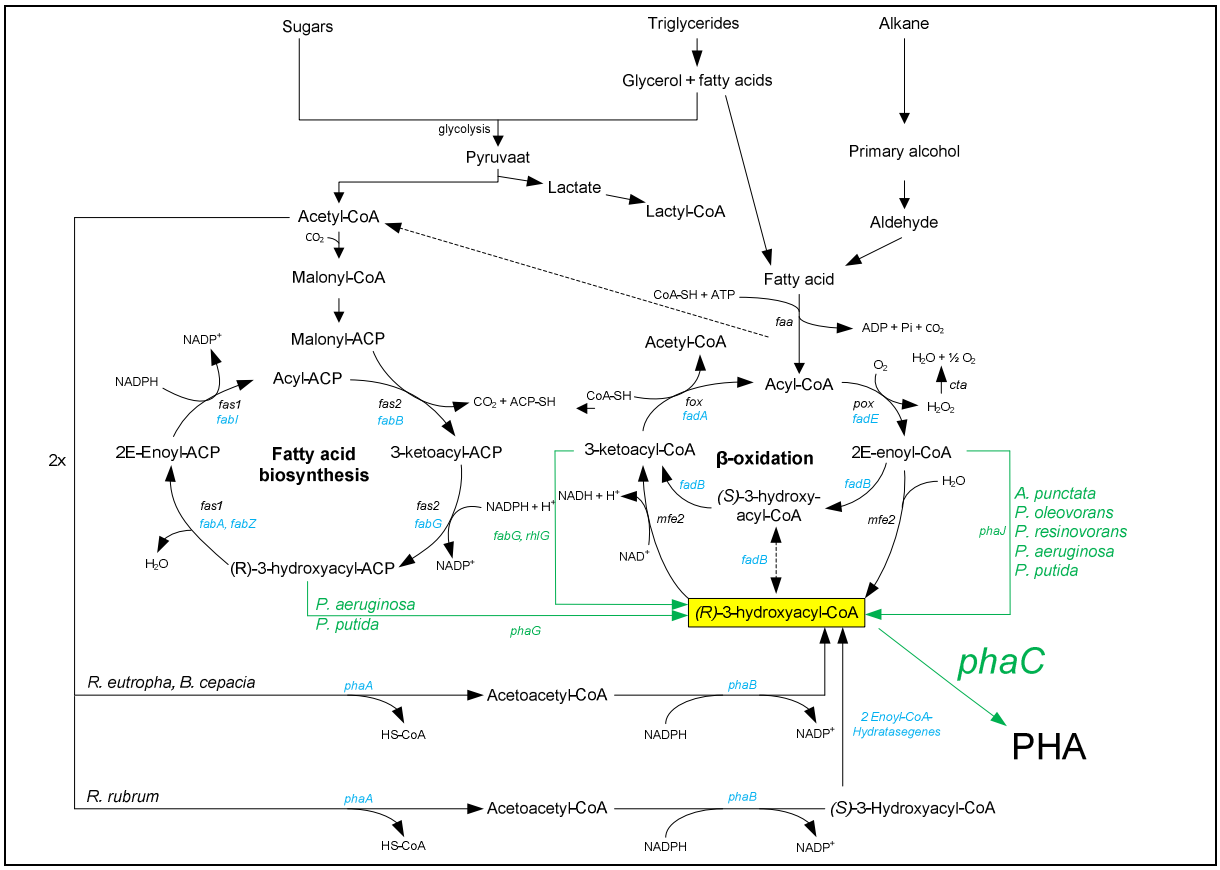


Figure VII. 5 Representation of the fatty acid metabolism in prokaryotes (blue genes names) and yeast (black gene names) and possible routes (of the respective organisms) that could be introduced for enhanced PHA production in *S. bombicola* using genetic engineering (green gene names).

Since the *cyp52M1* regulatory sequences drive expression of the *phaC1co* gene in the transformant, substantial high C/N and/or C/P ratios are thus probably needed for optimal activation of the regulatory sequences. It was stated by (Hommel *et al.*, 1994a) that it is not the C/N ratio that affects SL production (of *C. apicola*), but rather the absolute amount of N. This is thus in contrast with our results as the latter remains constant in our experiments. However, the C/N ratio, and not the absolute amount of carbon and/or nitrogen was proven to regulate SL biosynthesis for *R. bogoriensis* (Cutler and Light, 1979). Likewise, high SL production by *S. bombicola* cultivated on the production medium (12 % glucose) will thus probably not only be attributed to a high supply of building blocks for SL production, but the effect on the regulatory level on the genes of the SL pathway is probably equally important. A possible remark that could be made here, is that the lower amounts of glucose in the medium could also affect PHA yields at the level of precursor synthesis. It could indeed be stated that glucose can be channeled to β-oxidation via futile cycling of intermediates of fatty acid

biosynthesis (Marchesini and Poirier, 2003) and that higher glucose concentrations (12 % versus 2 and 0.2 % respectively) could hence lead to higher PHA amounts for the higher glucose concentrations. However, as no PHA was detected in cells grown on the production medium with high glucose concentrations (12 %) without addition of oil, this reasoning is unlikely and the observed results are very likely caused by effects on the regulatory level.

Such regulatory effects are of paramount importance for the further genetic engineering of *S. bombicola.* Especially if one wants to make use of its efficient SL machinery for the production of other interesting biomolecules *i.e.* using the promoters and loci of the SL biosynthetic gene cluster as was done for PHA production. As RNA sequencing experiments were run for the IWT-SBO project (Biosurf) to which this work contributed, a few samples were included to further investigate the abovementioned effects.

- RNA sequencing experiments

RNA sequencing was performed for the *S. bombicola* wild type grown on the production medium with 12 % versus 0.1 % of glucose, and the expression of the genes of the SL biosynthetic pathway in the stationary phase for these two situations was compared. The data were analysed at the lab of Professor Yves Van de Peer (Bioinformatics & Systems Biology, VIB) by PhDs Bing Li. The data for the SL biosynthetic gene cluster and more specifically the *cyp52M1* gene are shown in Table VII. 4.

RNAseq results indeed support the abovementioned hypothesis as a clear and significant repression (or lack of upregulation) of the genes of the biosynthetic SL gene cluster was observed for low glucose conditions (0.1 %) as compared to high glucose concentrations (12 %). Of the 4625 predicted genes for which data was collected, 79 were significantly up- or downregulated in conditions of low (0.1 %) initial glucose concentrations as compared to high (12 %) initial glucose concentrations. Of those 79, the 5 genes responsible for SL biosynthesis (Van Bogaert *et al.,* 2013) and more specifically the *cyp52M1* gene, were significantly downregulated in conditions of low (0.1 %) glucose concentration, which proves the regulatory effect of glucose on the expression of these genes. It has to be mentioned that the absolute log2 value (hence the difference of expression of the genes under these two conditions) was significantly higher for these 5 genes as compared to the obtained values for all the other regulated genes, for which absolute log2 values of around 4 were observed. This

illustrates the distinct and significant effect of glucose on the regulation of the genes of the SL biosynthetic pathway. The effect was most pronounced for the *cyp52M1* gene and hence for the *cyp* promotor, clearly confirming the results obtained for PHA production in this yeast. It also has to be mentioned that both experiments were performed in the absence of oil and that high expression driven by the P*cyp52M1* (and the other genes of the SL biosynthetic pathway) is thus clearly not dependent on the presence of oil as was already described above.

Table VII. 4 Quantitative expression results for the genes of the SL biosynthetic pathway in conditions of high (12 %) versus low (0.1 %) glucose concentrations. The results for the first step of SL biosynthesis (*cyp52M1*) are depicted in bold as the promotor of this gene was used for *phaC1co* expression.

	Expression levels (FPKM)		**Log2 (FPKM1/FPKM2)**	**q-value**	**Significant?**
Gene	0.1% glucose w/o oil	12% glucose w/o oil			
lipase	79.5	388.4	2.29	0.33	no
ugtB1	1.6	606.3	8.59	1.55E-08	yes
SL transporter	0.7	124.8	7.47	3.27E-07	yes
at	1.3	408.6	8.26	1.58 E-08	yes
ugtA1	97.1	6106.7	5.98	3.48 E-03	yes
cyp52M1	**3.4**	**1786.2**	**9.04**	**5.22E-08**	yes

High glucose concentrations are thus needed for optimal *cyp52M1* promotor activity, which on the other hand represses the pathway needed for Pha synthase substrate supply (β-oxidation), this situation thus results in a contradiction in terminis.

VII.3.4. *Toxicity of rhamnolipids (RLs) towards Starmerella bombicola*

Because production of RLs by *S. bombicola* is one of the future goals of the research focused on *S. bombicola*, one has to ensure that these compounds do not impose toxic effects on *S. bombicola*. Toxicity of RLs towards this yeast was thus assessed at different concentrations, growth phases and types of cultivation. The experiments were performed under conditions promoting SL production, as this would most probably be the prevailing situation when producing (tailor-made) glycolipids (like RLs) with this yeast.

When plating dilution series of either exponential (30 hours of cultivation) or stationary cultures (96 hours of cultivation) of the wild type on solid 3C agar plates supplied with RLs

(20 g/L), an inhibitory effect of RLs on cell growth can be observed as almost no cell growth could be detected anymore for dilutions beyond 10^{-2}, both for cells derived from exponential as stationary cultures (cfr. Figure VII. 6). Only the results for the stationary culture (96 h) are shown as results for the exponential one (30 h) were similar.

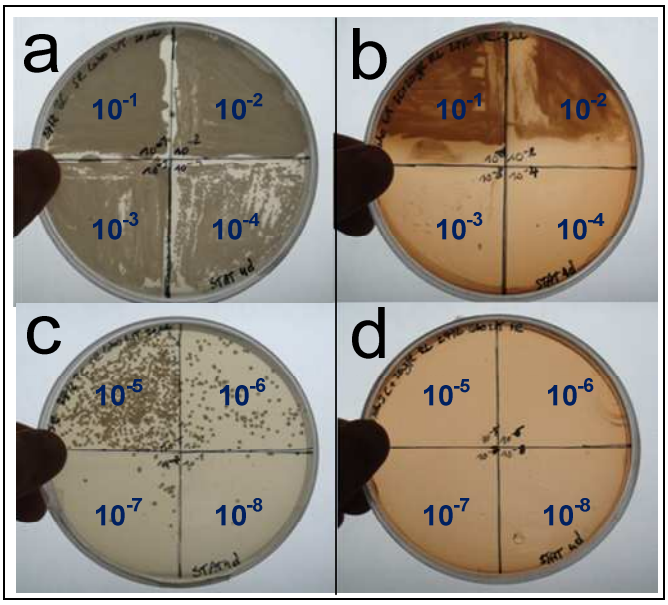


Figure VII. 6 Solid agar plates containing normal 3C medium (**a** and **c**) and 3C medium supplemented with RLs (**b** and **d**) inoculated with a dilution series derived from a *S. bombicola* wild type culture on production medium (96 h). Similar results were obtained for an exponential culture (30 h).

Another experiment was set up to examine the influence of the presence of RLs in liquid *S. bombicola* cultures. RLs were thus added in two concentrations (5 g/L and 20 g/L) to liquid cultures of wild type *S. bombicola* cells cultivated on 3C medium. Again the effect both on exponential as stationary cells was assessed (for exponential addition, the RLs were present in the shake flasks from the beginning of the experiments). Log CFU/mL values were determined and the results are depicted in Figure VII. 7. Only the results for the exponential addition of RLs are shown not to overload the figure. These results are similar, though more pronounced, as the ones obtained for the stationary addition of RLs.

In contrast to the experiments on solid culture media, no adverse effects of the addition of RLs to liquid cultures of *S. bombicola* were observed. On the contrary, the cultures to which RLs in a concentration of 20 g/L had been added seemingly even showed higher colony

forming units then the ones to which no or less RLs had been added. However, the pH values of these first cultures also remained higher during the stationary phase as compared to those of the cultures to which no RLs had been added (not shown). This pH effect is caused by the fact that the RL preparation is alkalic. These higher pH values could be an explanation for the higher CFU values as the pH dropped to 3 (and below) for the culture to which no RLs had been added in contrast to a minimum of 5 for the culture to which 20 g/L of RLs had been added. Alternatively the RLs could exert a more direct positive effect on growth of *S. bombicola*, which would be an interesting finding.

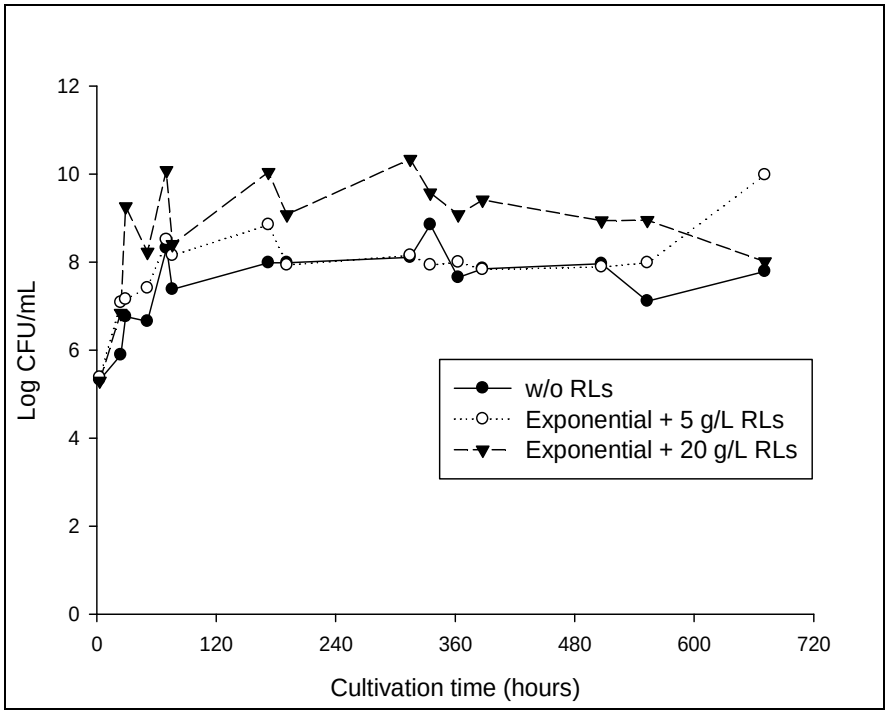


Figure VII. 7 Log CFU/mL values for exponential addition of RLs to *S. bombicola* cultures on 3C medium with (5 g/L and 20 g/L) and without (w/o) addition of RLs (control).

As catabolism of SLs by this yeast was shown in Chapter III of this manuscript, the possibility exists that the yeast can also use other glycolipids as a source of carbon. This hypothesis was examined by analyzing samples of the abovementioned experiments on HPLC. When doing so it became clear that both SLs and RLs were present in these samples and that the different SL and RL variants show overlapping elution times, which made it difficult to draw conclusions. However, as SL production thus apparently wasn't hampered by the presence of RLs, two conclusions can be drawn from the above: firstly RLs do not hamper growth and/or cell viability of *S. bombicola* in liquid cultures and they moreover do not

negatively affect the SL machinery of *S. bombicola.* No negative indications for engineering RL production in *S. bombicola* can thus be concluded.

To further investigate possible catabolism of RLs by *S. bombicola*, the experiment was repeated with the *Δcyp52M1* deletion strain (Van Bogaert *et al.*, 2013). The former is deficient in SL synthesis, which allows to solely observe the evolution of the RL composition in the culture medium without 'contamination' of SLs in the chromatograms. The *Δcyp52M1* deletion strain was cultivated on 3C medium in the presence of 5 g/L of RLs. Cultures were sampled throughout cultivation and HPLC chromatograms of an early (3 days of cultivation) and late sample (43 days of cultivation) are depicted in Figure VII. 8

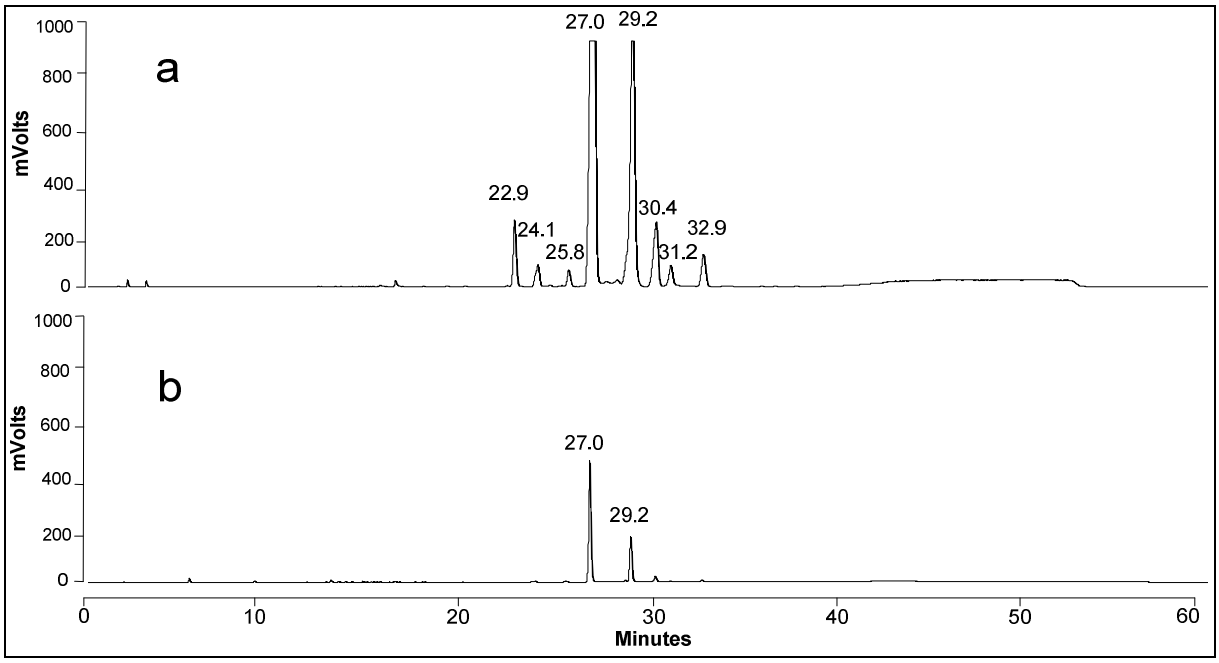


Figure VII. 8 HPLC-ELSD chromatograms of extractions from the *S. bombicola Δcyp52M1* strain cultivated on 3C medium supplied with RLs (5 g/L) after **(a)** 3 days **(b)** 43 days of cultivation. Similar results were obtained for a biological replica of the experiment.

It seems, as was already suspected, that RLs are taken up from the culture medium by the *Δcyp52M1* deletion strain. In contrast to SL catabolism (cfr. Chapter III) no extracellular dissimilation in the building blocks, but rather an uptake of the (non-acetylated) RL molecules is observed when analyzing HPLC-ELSD chromatograms. If this is indeed the case, then this would strengthen the hypotheses made for SL catabolism in Chapter III (deacetylation preceding uptake). However, when analyzing the samples on LC/MS more peaks were detected then on HPLC-ELSD, which corresponded to rhamnose and mono RL with only one β-hydroxylated fatty acid (C_{10}) attached to it. These compounds weren't present in the used

RL preparation. However, the presence of a considerable amount of free β-hydroxylated C_{10} fatty acid was detected in the substrate. This compound wasn't present anymore in the 'degradation' samples, which thus indicates that the cells metabolize these compounds derived from RLs. It thus seems that *S. bombicola* can catabolise RLs if incubated long enough after glucose has been depleted from the medium. This is an important finding for the further genetic engineering of *S. bombicola* aiming for the production of RLs. A remark that should be made here is that a cell free incubation of RLs, to control for spontaneous effects, wasn't included in these experiments due to the limited availability of RLs to perform experiments. Although similar experiments with SLs (cfr. Chapter III) did not result in spontaneous degradation of the SL molecules, such control experiment should be performed to be absolutely sure of RL catabolism by *S. bombicola*. The presumed uptake/metabolism of RLs only occurred after glucose was depleted from the culture medium, so these findings aren't suspected to limit RL production by *S. bombicola* either.

VII.4. Conclusion

In this chapter (and the preceding one) the first steps to transform *S. bombicola* into a platform organism for the production of (tailor-made) biomolecules were successfully taken. A codon optimized Pha synthase gene (*phaC1co*) introduced at the *cyp52M1* locus of the SL gene cluster gave rise to a strain defective in SL production, while instead producing the bioplastic polyhydroxyalkanoate (PHA) when cultivated on a combined substrate consisting of glucose (12 %) and a lipogenic substrate. The obtained PHA yields (2.0 % ± 0.3 % wt/dwt) were higher than those reported for other yeasts equipped with a peroxisomal Pha synthase without further metabolic engineering of the strains (< 1 % wt/dwt). The composition of the produced MCL-PHA on one hand reflects the fatty acid profile of the used lipogenic substrate, but on the other hand also reflects the substrate specificity of the Pha synthase as feeding the strain with LCL substrates did not lead to the desired incorporation of LCL-PHA monomers (C16 and higher) in the produced PHA. Production experiments using lower glucose concentrations (2 % and 0.2 %) resulted in very low or no PHA production, whereas higher yields were expected, as glucose is known to represses the β-oxidation, which is needed to supply the building blocks for PHA production. As no PHA was detected either for production experiments under high glucose concentrations (12 %) when no oil was added to the culture medium, the obtained results for the low glucose concentrations were not expected to result from a lower supply of PHA building blocks from glucose (trough *de novo* fatty acid

synthesis), but rather from regulatory effects excerted on the *cyp52M1* regulatory regions. Indeed RNA sequencing results demonstrated the absolute importance of high glucose concentrations for a transcriptionally active SL biosynthetic gene cluster. The SL biosynthetic gene cluster is thus proposed to be regulated by the C/N ratio or by the absolute amount of glucose present in the culture medium. This can be further investigated by using the developed qPCR platform described in Chapter V. This necessity for high glucose concentrations for *cyp52M1* promotor activity on one hand in combination with the repressing effect of such high glucose concentrations on the β-oxidation on the other hand, indicates that elevated PHA synthesis in *S. bombicola* will require the use of another promotor. A promotor repressed by glucose and incuced by fatty acids and/or oils, like the icocitrate lyase promotor (P*icl*) from the glyoxylate pathway or the catalase promotor (P*cat*) for instance, could be good candidates. It is important to take note of the fact that the *cyp52M1* regulatory sequences are still active without the presence of SLs, as PHA production was accomplished in a strain unable to produce SLs. Some positive feedback of the end-products (SLs) could have been expected, which would limit the efficient production of new biomolecules using the SL regulatory sequences in SL negative strains. For this example this did not seem to pose a problem, but if high(er) yields are aimed for, this can be a factor that has to be kept in mind.

Because the production of RLs by *S. bombicola* is aimed for in the future, the toxicity of these molecules towards this yeast was also examined. In contrast to inhibitory effects observed when growing *S. bombicola* on solid growth medium supplied with RLs (20 g/L), no inhibitory effects could be attributed to the presence of RLs (5 g/L and 20 g/L) in liquid cultures (exponential and stationary). On the contrary, some positive growth effects were observed. This could be a mere pH effect as the added RL solution is alkalic and has a buffering effect as the pH drop associated with exponentially growing yeast cells (cfr. Chapter III) is less pronounced when these RLs were added to the cultures. Alternatively a more direct positive effect might exist, e.g. metabolisation of RLs (or building blocks thereof) by *S. bombicola*. In Chapter III of this manuscripts, the catabolism of SLs by *S. bombicola* was demonstrated and the results suggested that acetylation of the glycolipids might possibly protect SLs from being taken up again by the cells. RLs do not contain acetylgroups, which could lead to easy uptake of these molecules. Indications for the latter were provided in this chapter, as supplied RLs disappeared from the culture medium of a *Δcyp52M1* mutant after glucose depletion. When analyzing these samples on LC-MS, the RLs also appeared to be digested in the respective building blocks like rhamnose and mono-RL with only one β-

hydroxylated fatty acid attached to it, which weren't present in the substrate. These results should be further investigated, as they are not only interesting but also important in the light of further genetic engineering of *S. bombicola* for the production of RLs. In conclusion RLs are not toxic for *S. bombicola* when they are present in liquid culture medium and they also do not negatively affect the SL machinery of the yeast, as SL production still occurs when RLs are added to the culture medium. This is also important if one wants to make use of the SL regulatory machinery for RL biosynthesis.

These results are of great importance for the further genetic engineering of *S. bombicola*. A possible strategy to engineer RL production in *S. bombicola* involves the replacement of (some) of the functional genes of the SL pathway by these of the RL biosynthetic pathway, as such using the regulatory network for SL biosynthesis for RL production, which thus seems plausible. However, for molecules requiring lower glucose input, *e.g*. PHA production, alternative strategies will have to be elaborated.

The combination of the results presented in this and the preceding chapter, illustrates that *S. bombicola* can be considered as a new production host for RLs and more generally for the production of tailor-made (glycolipid) biomolecules. It can also be concluded that codon optimization (using the codon usage of the genes of the SL biosynthetic pathway) together with gene expression controlled by the regulatory sequences (5' and 3' UTRs) of the gene(s) of the SL biosynthetic pathway or a combination of both, is considered a successful strategy for further engineering of the yeast. The presence of SLs was not necessary for promotor activity of the *cyp52M1* gene, nor did the presence of RLs inhibit the SL machinery. It is thus feasible to integrate new functionalities at genomic loci necessary for SL production to on one hand safeguard the new biomolecules from SL contamination, while on the other hand maintaining the advantages of a clear separation between growth and production phase. Although promising, a lot of work still remains to be done to uncover the full regulatory network for SL biosynthesis and more importantly to close the gap between such interesting findings on lab scale and possible industrial applications.

CHAPTER VIII:

CONCLUSIONS AND PERSPECTIVES


Chapter VIII.

CONCLUSIONS AND PERSPECTIVES

Secondary metabolites derived from plants and micro-organism play a role of paramount importance in modern daily life. Biosurfactants are a group of secondary metabolites produced by micro-organisms, which are suspected to gain industrial interest in the coming years. A general overview of biosurfactants produced by micro-organisms was given in Chapter I of this dissertation and the biosynthesis and regulation thereof was particularly discussed for the glycolipid biosurfactants produced by yeasts and fungi.

Surfactants are used for a wide variety of applications in household and personal care products, food processing industry, agriculture and others (cfr. Figure VIII. 1). Consequently, they are amongst the most important classes of industrial bulk chemicals, with a total world production exceeding 13 million tons per year (Levinson, 2008), which is predicted to reach 20 million tons worth $ 36 518 million by 2017 (Rohan, 2012).

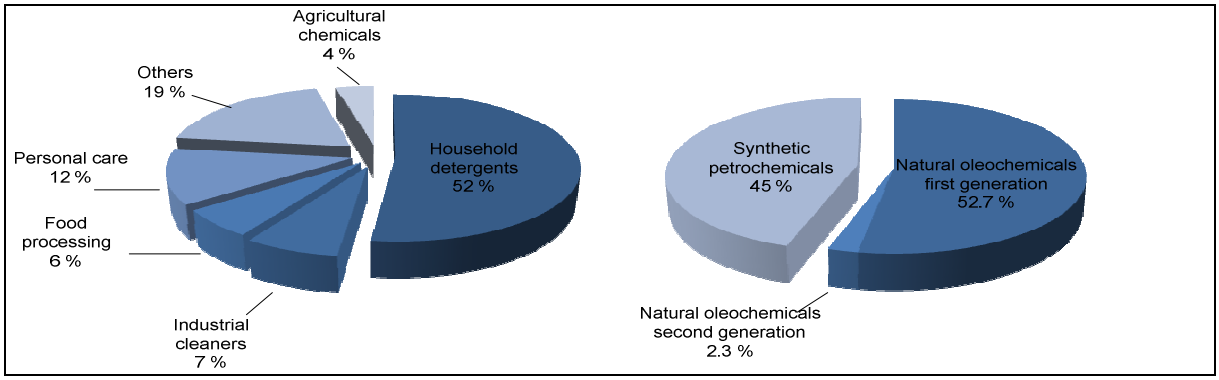


Figure VIII. 1 **(a)** Schematic depiction of the global surfactant market in respect to applications according to Levinson (2008) which equalled 13 M ton in 2006 **(b)** estimated division of the total surfactant market in 2020 based on several reports (anonymous, 2007; Rust and Wildes, 2008).

About half of the produced surfactants find applications as household and laundry detergents, which can result in environmental problems if they end up in the environment after use. Moreover, traditional surfactants were in the past mostly produced using chemical production processes starting from petrochemical resources such as ethylene, benzene, kerosene and n-paraffins. Growing environmental awareness during the last ten years in combination with diminishing oil reserves has resulted in a shift from these pure synthetic products to (semi-)

bio-based alternatives. Currently more than half of the surfactant market is reported to consist of “natural” oleochemicals. A minor fraction thereof are biosurfactants of the second generation and an even smaller fraction thereof are represented by biosurfactants produced by micro-organisms. Their biodegradability and low ecotoxicity and the fact that they can be produced by a natural fermentation process from renewable feedstocks or even waste streams in combination with good functional properties gives them an advantage over their chemical counterparts (Banat *et al.*, 2000). Bio-surfactants are thus a promising target for the biobased economy. Especially glycolypid biosurfactants like sophorolipids (SLs), rhamnolipids (RLs), mannosylerythritol lipids (MELs) and cellobioselipids (CBLs) have been described to provide significant opportunities to replace chemical surfactants in several sectors (Marchant and Banat, 2012).

Unfortunately, the productivity of natural bio-surfactant producing strains is often too low for commercialization. Moreover, microbial derived bio-surfactants typically have a defined structure, which is to a certain degree limited by the biosynthetic capacities of the producing strain. This is in sharp contrast with the chemically derived surfactants where one can easily introduce variation by simply changing the building blocks. As a result of these two limitations, up until now the industrial production of only a narrow range of microbial bio-surfactants is feasible and only sophorolipids, rhamnolipids and mannosylerythritol lipids are currently commercialized. One of the challenges for the further elaboration of the biobased economy thus involves the tinkering, fine tuning and rearrangement of these biological pathways with the aim of obtaining higher yields in combination with a more extensive structural variety, preferably by using only a few platform organism(s) allowing the development of standardized procedures.

A highly productive bio-surfactant producer is the yeast *Starmerella bombicola*, which is capable of producing up to more than 400 g/L of sophorolipids (SLs). This efficient process resulted in commercialization of SLs for applications in cosmetics produced by Soliance (Sopholiance), applications of these biomolecules in the ecological cleaning products of Ecover and Wheatoleo (Sophoclean), dishwasher products of Saraya (Sophoron) and the filing of numerous patent applications. The full biosynthetic pathway of SLs was shown in Figure II. 1 and almost all the responsible genes were found to be encoded in one large genomic gene cluster. Only the gene responsible for lactonisation of SLs was found to be encoded elsewhere in the genome.

Although the structure of the produced SL molecules can to a certain extent be varied by smart feeding strategies (Van Bogaert *et al.*, 2011b), one will always be delimited by the biosynthetic capacity of the yeast, which is characterized by the specificity of the corresponding enzymes of the biosynthetic pathway and thus by the information encoded in the genome. The potential of genetic engineering with the aim of altering the structure of the produced molecules was already proven in the past and resulted in two patent applications for the production of new-to nature SLs (Develter and Fleurackers, 2007; Van Bogaert *et al.*, 2011a). This stimulated further research with the aim of transforming *S. bombicola* into a platform organism for tailor-made biomolecules. An interesting target was identified in rhamnolipids. These biosurfactants are described to have numerous applications, but industrial production is hampered by the complex regulation mechanisms and safety issues associated with the pathogenic rhamnolipid producer *Pseudomonas aeruginosa*. To our knowledge they are only commercially produced by an American company (Jeneil) and a Chinese company (Urumqi). Heterologous production of such molecules by a non-pathogenic yeast strain capable of producing large amounts of glycolipid molecules, without experiencing any adverse affects of high titers in the culture medium thus seems a plausible research goal.

In order to genetically engineer this sophorolipid producing yeast into a rhamnolipid producing one, at least nine heterologous genes should be transformed into *Starmerlla bombicola* and possibly even more. The successful expression of such heterologous biosynthetic pathway in this non-conventional yeast thus requires the development of a molecular toolbox, which was investigated in PART II of this manuscript.

Multiple promoters, terminators, sites of integration and an elegant combinatorial assessment of all these parameters should be available. To do so a suitable reporter system is needed. In chapter IV of this dissertation the development of a reporter system, using a codon optimized *gfpco10g* (green fluorescent protein) was described. Codon optimization was an absolute requirement to obtain GFP expression, which is an important fact to be kept in mind to proceed to further genetic engineering of this yeast. The created platform was proven to be functional and useful for the determination of promotor activity. This validation was done by placing the gene under control of the endogenous *S. bombicola* promotors of the glyceraldehyde-3-phosphate dehydrogenase gene of the glycolysis (P*gapd*) and the cytochrome P450 monooxygenase gene of the SL biosynthetic pathway (P*cyp*). The activity of the latter was shown to be upregulated in the stationary growth phase, which is consistent

with the occurrence of SL production, which predominantly occurs in the stationary growth phase. Its activity was rather low as compared to that of P*gapd*, an effect that was not really expected, seen the very efficient SL machinery of *S. bombicola*. However, recent semi-quantitave RNA sequencing data confirmed that expression of the *gapd* gene in the early stationary growth phase was indeed 7 times higher than that of the *cyp* gene, thus confirming the obtained results. The availability of these RNA sequencing data will now allow for the selection of a set of potential interesting endogenous promoters, which in combination with the created reporter system will allow for the construction of a promotorbank, harboring inducible and constitutive (endogenous) promoters covering a wide range of activities, which can subsequently be used for genetic engineering purposes. Moreover, the reporter system can be used to screen for optimal sites of integration for new functionalities. Furthermore the reporter system will also serve as a method to perform research on the more fundamental level. More specifically this expression system will allow the further investigation of the regulatory network affecting regulation of the SL biosynthetic gene cluster of *S. bombicola*. The subtelomeric localization of this gene cluster could namely indicate the existence of a regulatory effect called the telomere position effect (TPE), which involves the regulation of subtelomeric SM biosynthetic gene clusters by chromatin based regulation and the action of histone deacetylases (HDAC). Such effects influence gene expression trough the mere localization thereof within the gene cluster, independently of the used promotor. GFP expression experiments *e.g.* by placing the *gfpco10g* under control of a certain promotor integrated inside the SL cluster compared to the situation where this expression cassette is located outside the gene cluster, could thus serve for the further elucidation of the regulation of this biosynthetic gene cluster. This will not only extend the fundamental knowledge of SL biosynthesis, but furthermore create opportunities for the genetic engineering of this yeast by using its efficient SL machinery.

The further investigation of the regulation of the SL biosynthetic gene cluster will also be possible by using the developed qPCR platform, which was described in Chapter V of this manuscript. The transcriptional stability of a set of 20 candidate reference genes was controlled and the platform was validated by performing expression analysis experiments on the genes of the SL biosynthetic pathway. A clear upregulation of almost all of the genes of the SL biosynthetic pathway in the stationary growth phase was demonstrated by using this platform. These results were consistent with RNA sequencing and proteomic data that were obtained by other project partners, so the qPCR platform is considered to be a reliable method

to further investigate the influence of parameters like pH, temperature, carbon source, etc. on the transcription of the genes of SL biosynthetic pathway. More importantly it can be used to monitor gene expression of production strains derived from the wild type: a strain knocked out in the acetyltransferase of the SL biosynthetic pathway *e.g.* gives rises to the production of non-acetylated SLs, which is a desirable feature (Saerens *et al.*, 2011b). However, yields drop to about 10 % as compared to those of the wild type, which also holds true for a strain knocked out in the second glucosyltransferase (*ugtB1*) (Saerens *et al.*, 2011c). When these results were obtained it was assumed that a possible positive feedback of the endproduct on the pathway existed, which wasn't occurring anymore in the *Δat* and *ΔugtB1* strains, or alternatively that a negative feedback loop, excerted by a buildup of SL intermediates, existed in these strains. However, an additional deletion of the *lipase* gene responsible for lactonisation in the *Δat* strain results in a strain (*ΔatΔlip)* producing almost equal amounts of SLs as the wild type (unpublished results). The question thus raises what kind of effects cause these observations. The developed qPCR platform can thus be used to investigate if this can be attributed to regulatory effects *e.g.* negative feedback of the intermediates on gene expression or positive feedback of the end-products or that alternatively some other effects are at work *e.g.* transport related issues. The qPCR platform was already used to investigate the possible regulatory effect of the addition of oil to the culture medium on two of the genes of the SL biosynthetic pathway. Both genes (*cyp52M1* and *lipase*) were not further upregulated in the stationary phase when oil was added to the culture medium. Higher SL titers in the presence of oil as comparared to those in the absence of oil can thus be attributed to a high lipgenic precursor availability and not to inducing effects of the lipogenic substrate on the level of transcription. The opposite holds true for the glucose as high glucoseconcentrations were demonstrated to be absolutely necessary for high transcriptional activity of the genes of the SL biosynthetic pathway. This was described in chapter VII of this manuscript and will be further discussed below. A higher amount of lactonisation in the presence of oil can on the other hand be attributed to effects on the enzyme level as lactonisation involves the release of a water molecule and a more hydrophobic surrounding will probably stimulate the action of the enzyme.

The availability of an episomal expression system would be a great advantage for transforming Starmerella bombicola into a platform organism. The availability of such episomal vector circumvents the process of creating large linear expression cassettes containing long (1000 bp at each side) regions needed for homologous recombination in the

genome. Possible (industrial) production strains of *S. bombicola* will always contain the new functionalities inserted in the genome, as such avoiding the necessity of selective media and possible risk of losing the episomal expression vector. However, for research purposes such episomal expression vector would represent a substantial simplification of the molecular activities needed to test new activities in this yeast. The creation of an episomal vector was investigated in Chapter IV of this manuscript. The usage of autonomous replication (ARS) and centromeric (CEN) sequences derived from other yeasts (*S. cerevisiae* and *Y. lipolytica*) did not result in the maintenance of an episomal vector in *S. bombicola*, nor was the 2µ plasmid of *S. cerevisiae* maintained (previous research). The creation of an episomal vector will thus probably require the identification of the centromeric (CEN) and/or ARS sequences of the yeast itself. This can be attained by the creation of genomic libraries and screening for stable transformants (ARS) and/or by applying synteny with related yeast species with known CEN sequences. However, the latter will require some luck, as the existence of sequencing data of yeasts with known CEN sequences that are closely enough related to *S. bombicola* is required to be able to apply this technique.

The extensive genetic engineering of *S. bombicola* will furthermore require the availability of more than the two selective markers that are currently available for *S. bombicola* (*ura3* and *hygroR*). Now that the genome is available it should be pretty straightforward to select a set of genes that can be used to construct auxotrophic strains and preferably those allowing both positive and negative selection. Alternatively one can try to reuse the existing markers. This can only be done if positive and negative selection is possible, which is only the case for the auxotrophic *ura3* gene marker. Its reuse was demonstrated in Chapter IV by placing it between two repeats of 196 bp in a transformation cassette. Intramolecular homologous recombination of these repeats, resulting in the removal of the *ura3* marker, was induced by placing the obtained *ura3+* transformants on selective medium for *ura3-* strains. The obtained transformants were subsequently checked for homologous recombination by performing colony PCR as spontaneous *ura3-* mutants are also selected by this method. The resulting strain can then again be re-used for introduction and/or deletion of genes. This process can be repeated several times enabling the repeated reuse of this marker.

The possible conversion of *S. bombicola* into an industrial production platform was further investigated in PART III of this dissertation.

As stated above, the further genetic engineering of this yeast aiming for the production of rhamnolipids (RLs) was set as one of the goals for future research. One of the proposed strategies, simply said, involves the replacement of the genes of the SL biosynthetic pathway of *S. bombicola* by these of the RL biosynthetic pathway, as such making use of the regulatory mechanisms normally leading to SL biosynthesis, while at the same time eliminating SL production and all this in a single recombination event. As unfortunately not a lot is known about the regulation of this gene cluster (yet), it was considered wise to first assess if this is a feasible strategy and that the highly efficient SL biosynthesis is for instance not only possible due to the existence of some kind of positive feedback loop.

Therefore the production of the bioplastic polyhydroxyalkanoate (PHA) by *S. bombicola* was chosen as a test case to evaluate the abovementioned strategy, which was discussed in Chapter VII. PHA production was already demonstrated for other yeasts and only requires the introduction of one gene (*pha synthase*). Because *S. bombicola* grows well on LCL substrates and also preferentially incorporates these in the produced SLs, the production of LCL-PHA (with potential interesting applications) was set as an additional goal. Therefore the Pha synthase of *Pseudomonas resinovorans* was chosen, as the PHA produced by this strain was reported to contain C16 monomers, which indicates a broader substrate specificity of the Pha synthase of *P. resinovorans* as compared to other *Pseudomonas* species, only reported to produce MCL-PHA. The gene was codon optimized using the codon usage of the genes of the SL biosynthetic gene cluster and integrated at the *cyp52M1* locus. By doing so, the expression of the *phaC1co* gene was as good as possible fine-tuned with the regulatory effects normally leading to SL biosynthesis, while the latter was inhibited by replacing the *cyp52M1* gene with the *phaC1co* gene. This was thus done to investigate the efficacy of using the efficient SL machinery for the production of a structurally different interesting biomolecule in a SL deficient strain. Indeed, cultivating the obtained transformant on the SL production medium containing 12 % glucose with addition of a lipogenic substrate resulted in the production of the bioplastic polyhydroxyalkanoate (PHA) in a SL defective *S. bombicola* strain. The obtained PHA yields (2.0 % ± 0.3 % wt/dwt) were higher than those reported for other yeasts equipped with a peroxisomal Pha synthase, without further metabolic engineering of the strains (< 1 % wt/dwt). The addition of a lipogenic substrate was an absolute prerequisite for PHA production, as were the high glucose concentrations, as cultivation on a medium containing only 2 % or 0.2 % of glucose gave rise to only very low PHA yields. The observed effects of the glucose concentration were proven to result from suboptimal P*cyp52M1* activity,

as RNA sequencing showed that expression of all the genes of the SL biosynthetic pathway are significantly downregulated on such low (0.1 %) glucose medium. This again is a very interesting finding for the further genetic engineering of *S. bombicola*, when making use of the regulation of the SL cluster for the production of (new to nature) biomolecules. Moreover, this example proved that the presence of SLs wasn't necessary to obtain P*cyp52M1* activity. Some positive feedback loops might still exist, but at least a considerable residual activity is maintained. It can also be concluded that codon optimization (using the codon usage of the genes of the SL biosynthetic pathway) in combination with gene expression controlled by the regulatory sequences (5' and 3' UTR's) of the (first) gene(s) of the SL biosynthetic pathway or a combination of both, is considered to be a successful strategy for further engineering of the yeast. It is furthermore feasible to integrate new functionalities at genomic loci necessary for SL production to on one hand safeguard the new biomolecules from SL contamination, while on the other hand maintaining the advantages of the SL regulatory network.

Because the heterologous production of RLs is set as one of the future goals for the continued engineering of *S. bombicola*, the toxicity of these molecules towards this yeast was also investigated and described in Chapter VII. It was concluded that RLs are not toxic towards *S. bombicola* when they are present in a liquid culture medium and that their presence does not either negatively affect the SL machinery of the yeast, as SL production still occurs when RLs are present. The combination of all these findings is very important for the future engineering of *S. bombicola* into a new production host for RLs and more generally for the production of tailor-made (glycolipid) molecules by this non pathogenic yeast.

Although the last four years a lot of progress has been made in the understanding of SL biosynthesis by *S. bombicola*, a lot of work still remains to done to uncover the full regulatory network. Several gene clusters encoding secondary metabolites contain an 'in cluster' regulator, which is responsible for transcriptional activation of the other genes of the cluster. Such 'in cluster' regulator was recently also found at the borders of the biosynthetic gene clusters responsible for cellobioselipid biosynthesis in *U. maydis* and *P. flocculosa*. Because the full genome of *S. bombicola* became available during this research, the two open reading frames (orfs) delimiting the SL gene cluster were deleted to investigate a possible involvement of these putative genes in SL biosynthesis and/or regulation. The results thereof are described in Chapter II of this manuscript. The two resulting knock out strains did not behave differently as compared to the wild type and SL biosynthesis was unaffected. A global

regulator for the SL biosynthetic pathway was thus not found at the borders or encoded within the gene cluster. This is rather regrettable as such specific (absolute) regulator of the SL biosynthetic pathway would represent the holy grail for the conversion of *S. bombicola* into an industrial platform strain. Such regulator could namely be placed under the control of any promotor with the desirable (inducible) activity (as demonstrated with the reporter platform) needed for the production of a certain compound, which would enable very efficient rational pathway design in *S. bombicola*. However, such specific regulator (a putative zinc finger protein) could still be identified by combining an in dept *in silico* analysis of the genome of *S. bombicola* with the now available RNA sequencing data. This should enable the withhold of a few candidates of which the possible biological relevance can be easily investigated by performing biological experiments.

The effects of the now expanding knowledge of this industrially important SL producer should not be delimited to the conversion of this yeast into a platform organism for the production of new biomolecules. On the contrary, although this efficient SL producer is already used for the industrial production of SL with producers, distributers and applicants such as Soliance (France), Ecover (Belgium), Wheatoleo (France), Saraya (Japan), MG Intobio (Korea), SyntheZyme (USA) and multinationals such as Henkel, some major drawbacks are linked to SL production with *S. bombicola*. The extension of our genetic knowledge of this yeast might create opportunities for the development of new industrial production strains for SLs. A very nice example hereof was described in Chapter VI of this dissertation *e.g.* the development of an overexpression strain for the gene responsible for lactonisation of SLs (oe *lip*):

One of the abovementioned drawbacks is the fact that the natural SL mixture produced by the wild type *S. bombicola* yeast consists of a mixture of different SL variants. The produced SL mixture thus has to be fractionated and purified to be suitable for specific applications, which require e.g. only the lactonic form. A lot of effort has therefore been invested in the elaboration of the optimal cultivation conditions and subsequent processes for the isolation of one of the main forms (both acid and lactone). Solvents can be used, but are often not desirable in view of the use of SLs as biosurfactants and it is furthermore undesirable that the product contains traces of organic solvents. The enzyme responsible for lactonisation of SLs was identified at the lab of one of our projectpartners by analyzing proteomic results and the overexpression of this gene is described in Chapter VI. Surprisingly the introduction of one extra copy of this ‘lipase’ gene in the genome under control of a strong constitutive promotor

(P*gki*), gave rise to a very clear effect on SL biosynthesis. Whereas for the wild type a mixture of acidic and lactonic SLs was produced, this overexpression strain gave rise to the almost exclusive production of lactonic SLs when cultivated on production medium. More remarkably even, under conditions normally favouring the almost exclusive formation of acidic SLs (absence of citrate from the culture medium), the efficient formation of almost 100 % lactonic SLs was observed in conctrast to the wild type which produced only acidic SLs. The development of the oe *lip* strain capable of producing almost 100 % lactonic SLs opens perspectives for direct production of lactonic SLs by fermentation without the need for additional and expensive separation procedures. The higher lacton versus acidic ratio of SLs furthermore resulted in the absence of foam formation in bioreactor experiments in contrast to the wild type. This represents an additional opportunity for the development of industrial production processes, where foam formation can be a real burden.

A second drawback of SL production by the wild type, is the fact that the SL machinery is highly sensitive to changes in the medium composition and/or the cultivation conditions. This was also demonstrated by the experiments described above, where the ommitment of citrate from the culture medium results in a severe effect on the composition of the produced SLs. Moreover, other environmental parameters like pH, temperature, used lipogenic carbon source, etc. also have an influence on the composition of the produced SL mixture and on the SL productivity. Such regulatory effects are undesirable for robust industrial production processes as they may hamper the guaranteed stability and quality of the end-product. However, the results discussed in chapter VI demonstrate that such effects can completely be avoided by placing one of the genes of the biosynthetic pathway under control of another promotor (in this case the strong, constitutive P*gki* promotor), which should be considered as a major opportunity for industrial production processes for lactonic SLs in addition to the already mentioned advantage of the the high purity of the produced lactonic SLs and absence of foam formation.

A last and third drawback is the fact that the tight regulation of SL biosynthesis in the wild type results in a sub-optimal production process, as 25 % of the cultivation time is dedicated to growth without (or very low) production. Onset of SL biosynthesis with the developed overexpression strain was detected earlier during cultivation as compared to the wild type. The overexpression strain furthermore seemed to produce more SLs as compared to the wild type. This was attributed to the fact that overexpression of the last step in the biosynthetic pathway results in some kind of ‘pull’ on the pathway, possibly leading to higher

concentrations of the end-product (di-acetylated lactonic SLs), which furthermore precipitates in the culture medium, because of low solubility at the low pH values, possibly leading to lower end-product inhibition.

In conclusion in Chapter VI a new SL producing *S. bombicola* strain with high industrial relevance was created. This strain is capable of producing highly pure lactonic SLs, without the need of further purification. Such excellent producer, giving rise to the efficient, robust and easy production of a specific compound of very high purity, with the additional technical benefit of low foam formation is expected to be a major advantage for industrial production and application of SLs.

Also of importance for the industrial production of SLs and their derivatives, are the results described in Chapter III of this dissertation in which *S. bombicola* was proven to catabolize its own SLs under certain conditions. A proposed natural function is the creation of a carbon sink under conditions of carbon excess (high glucose was an absolute requirement for the transcriptional activity of the SL biosynthetic gene cluster). Some specific activities of *S. bombicola* catabolic enzymes on different SL derivatives e.g. ringopening without deacetylating activity or selective de-acetylating activity were detected. Such enzymes with specific activity might be an alternative way of producing specific compounds like di-acetylated acidic SLs or mono-acetylated lactonic SLs respectively. The first ones can also be produced by cultivation of a strain knocked out in the lipase gene responsible for lactonisation (Ciesielska *et al.*, 2013), but then the non-acetylated derivatives are also present and the ratio of these compounds is highly susceptible to variation, which is a drawback for industrial production of such molecules. The identification and production of some of the responsible enzymes, might thus have some industrial relevance for possible post-fermentative modifications of SLs produced by *S. bombicola*. Using the end-product of the lipase overexpression strain as the substrate would furthermore ensure the purity and quality of the derived compounds. A short list of possible conversions linked to the detected activities in degradation experiments is given below:

- Conversion of di-acetylated lactonic SLs to di-acetylated acidic SLs. The production therof can also be obtained by using a strain knocked out in the abovementioned gene responsible for lactonisation of acidic SLs (*lip*). However, it this case also the non-acetylated acidic SLs are formed and the ratio of the two forms is moreover highly susceptible to variation, which could hamper industrial production processes with such

strain, as in the industry standardized products with certain specifications are required. SL production with this strain is furthermore associated with high foam formation, which can be quite a burden.

- Conversion of di-acetylated lactonic SLs to mono-acetylated lactonic SLs.
- Conversion of non-acetylated acidic SLs to sophorose and hydroxylated fatty acid. This last activity was only detected in very minor amounts. However, sophorose is a rare sugar with described applications in the industry *e.g.* cellulase induction in fungi. Sophorose is furthermore extremely expensive, so the discovery and isolation of such enzyme might again have some industrial relevance.

In this PhD thesis some of the aspects of the SL metabolism of *S. bombicola* became elucidated, whereas the largest part still remains unknown. The first steps towards the conversion of this yeast into an industrial production platform were taken. Several molecular tools were developed, which can now be used for the further extensive genetic engineering of *S. bombicola*. The possible relevance of this yeast as a production platform was furthermore demonstrated by to examples described in the last part of this dissertation.

Some of the results described in this manuscript are expected to have implications for the industrial production of SLs, for which the filing of two patent applications linked to this work are also indicative (Soetaert *et al.*, 2010; Ciesielska *et al.*, 2013).

REFERENCES

REFERENCES

Abalos, A.; Pinazo, A.; Infante, M. R.; Casals, M., *et al.* (2001). Physicochemical and antimicrobial properties of new rhamnolipids produced by *Pseudomonas aeruginosa* AT10 from soybean oil refinery wastes. *Langmuir,* 17: 1367-1371.

Abdel-Mawgoud, A. M.; Aboulwafa, M. M. and Hassouna, N. A. H. (2008). Optimization of surfactin production by *Bacillus subtilis* isolate BS5. *Applied Biochemistry and Biotechnology*, 150 (3): 305-325.

Abdel-Mawgoud, A. M.; Lépine, R. and Déziel, E. (2010). Rhamnolipids: diversity of structures, microbial origins and roles. *Applied Microbiology and Biotechnology,* 86: 1323-1336.

Aerts, D.; Verhaeghe, T. F.; Roman, B. I.; Stevens, C. V., *et al.* (2011). Transglucosylation potential of six sucrose phosphorylases toward different classes of acceptors. *Carbohydrate Research*, 346 (13): 1860-1867.

Al-Tahhan, R. A.; Sandrin, T. R.; Bodour, A. A. and Maier, R. M. (2000). Rhamnolipid-induced removal of lipopolysaccharide from *Pseudomonas aeruginosa*: Effect on cell surface properties and interaction with hydrophobic substrates. *Applied and Environmental Microbiology*, 66 (8): 3262-3268.

Albrecht, A.; Rau, U. and Wagner, F. (1996). Initial steps of sophoroselipid biosynthesis by *Candida bombicola* ATCC 22214 grown on glucose. *Applied Microbiology and Biotechnology*, 46 (1): 67-73.

Albus, A. M.; Pesci, E. C.; RunyenJanecky, L. J.; West, S. E. H., *et al.* (1997). Vfr controls quorum sensing in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 179 (12): 3928-3935.

Anderson, C. and Tatchell, K. (2001). Hyperactive glycogen synthase mutants of *Saccharomyces cerevisiae* suppress the glc7-1 protein phosphatase mutant. *Journal of Bacteriology*, 183 (3): 821-829.

anonymous. (2007). The Lead market Task force on Bio-Based products: Accelerating the development of the market for bio-based products in Europe: 24.

Arguelles, J. C. (2000). Physiological roles of trehalose in bacteria and yeasts: a comparative analysis. *Archives of Microbiology,* 174 (4): 217-224.

Arima, K.; Kakinuma, A. and Tamura, G. (1968). Surfactin a crystalline peptidelipid surfactant produced by *Bacillus subtilis*-isolation, characterization and its inhibition on fibrin clot formation. *Biochemical and Biophysical Research Communications,* 31 (3): 488-&.

Ashby, R. D. and Foglia, T. A. (1998). Poly(hydroxyalkanoate) biosynthesis from triglyceride substrates. *Applied Microbiology and Biotechnology,* 49 (4): 431-437.

Ashby, R. D.; Nunez, A.; Solaiman, D. K. Y. and Foglia, T. A. (2005). Sophorolipid biosynthesis from a biodiesel co-product stream. *Journal of the American Oil Chemists Society*, 82 (9): 625-630.

Ashby, R. D.; Solaiman, D. K. Y. and Foglia, T. A. (2006). The use of fatty acid esters to enhance free acid sophorolipid synthesis. *Biotechnology Letters*, 28 (4): 253-260.

Asmer, H.-J.; Lang, S.; Wagner, F. and Wray, V. (1988). Microbial production, structure elucidation and bioconversion of sophorose lipids. *Journal of the American Oil Chemists Society*, 65 (9): 1460-1466.

Atkins, D. and Izant, J. G. (1995). Expression and analysis of the green fluorescent protein gene in the fission yeast *Schizosaccharomyces pombe*. *Current Genetics*, 28 (6): 585-588.

Awai, K. and Wolk, C. P. (2007). Identification of the glycosyl transferase required for synthesis of the principal glycolipid characteristic of heterocysts of *Anabaena sp* strain PCC 7120. *Fems Microbiology Letters*, 266 (1): 98-102.

Azim, A.; Shah, V.; Doncel, G. F.; Peterson, N., *et al.* (2006). Amino acid conjugated sophorolipids: A new family of biologically active functionalized glycolipids. *Bioconjugate Chemistry*, 17 (6): 1523-1529.

Bailey, T. L.; Boden, M.; Buske, F. A.; Frith, M., *et al.* (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, 37: W202-W208.

Baldé, I. (1989). Exoenzymmuster von Biotensid-positiven und -negativen Hefen der Gattung *Torulopsis*. Leipzig, Karl Marx University. PhD.

Banat, I. M.; Franzetti, A.; Gandolfi, I.; Bestetti, G., *et al.* (2010). Microbial biosurfactants production, applications and future potential. *Applied Microbiology and Biotechnology*, 87: 427-444.

Banat, I. M.; Makkar, R. S. and Cameotra, S. S. (2000). Potential commercial applications of microbial surfactants. *Applied Microbiology and Biotechnology*, 53 (5): 495-508.

Barros, B. A.; de Oliveira, M. C. F.; Mafezoli, J.; Barbosa, F. G., *et al.* (2012). Secondary Metabolite Production by the Basidiomycete, *Lentinus strigellus*, under Different Culture Conditions. *Natural Product Communications*, 7 (6): 771-773.

Basehoar, A. D.; Zanton, S. J. and Pugh, F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. *Cell*, 116: 699-709.

Beal, R. and Betts, W. B. (2000). Role of rhamnolipid biosurfactants in the uptake and mineralization of hexadecane in *Pseudomonas aeruginosa*. *Journal of Applied Microbiology*, 89 (1): 158-168.

Belanger, R. R. and Deacon, J. W. (1996). Interaction specificity of the biocontrol agent *Sporothrix flocculosa*: A video microscopy study. *Phytopathology*, 86 (12): 1317-1323.

Bisht, K. S.; Gao, W. and Gross, R. A. (2000). Glycolipids from *Candida bombicola*: Polymerization of a 6-O-acryloyl sophorolipid derivative. *Macromolecules*, 33 (17): 6208-6210.

Black, T. A.; Cai, Y. P. and Wolk, C. P. (1993). Spatial expression and autoregulation of HetR, a gene involved in the control of heterocyst development in *Anabaena*. *Molecular Microbiology*, 9 (1): 77-84.

Black, T. A. and Wolk, C. P. (1994). Analysis of a Het- mutation in *Anabaena sp*. strain PCC7120 implicates a secondary metabolite in the regulation of heterocyst spacing. *Journal of Bacteriology*, 176 (8): 2282-2292.

Bok, J. W.; Hoffmeister, D.; Maggio-Hall, L. A.; Murillo, R., *et al.* (2006a). Genomic mining for *Aspergillus* natural products. *Chemistry & Biology*, 13 (1): 31-37.

Bok, J. W. and Keller, N. P. (2004). LaeA, a regulator of secondary metabolism in *Aspergillus spp. Eukaryotic Cell*, 3 (2): 527-535.

Bok, J. W.; Noordermeer, D.; Kale, S. P. and Keller, N. P. (2006b). Secondary metabolic gene cluster silencing in *Aspergillus nidulans. Molecular Microbiology*, 61 (6): 1636-1645.

Boles, B. R.; Thoendel, M. and Singh, P. K. (2005). Rhamnolipids mediate detachment of *Pseudomonas aeruginosa* from biofilms. *Molecular Microbiology*, 57 (5): 1210-1223.

Boothroyd, B.; Thorn, J. A. and Haskins, R. H. (1956). Biochemistry of the ustilaginales. 12. Characterization of extracellular glycolipids produced by *Ustilago sp. Canadian Journal of Biochemistry and Physiology*, 34 (1): 10-14.

Brakemeier, A.; Wullbrandt, D. and Lang, S. (1998). Microbial alkyl-sophorosides based on 1-dodecanol or 2-, 3- or 4-dodecanones. *Biotechnology Letters*, 20 (3): 215-218.

Brandt, I.; Gerard, M.; Sergeant, K.; Devreese, B., *et al.* (2008). Prolyl oligopeptidase stimulates the aggregation of alpha-synuclein. *Peptides*, 29 (9): 1472-1478.

Branduardi, P.; Valli, M.; Brambilla, L.; Sauer, M., *et al.* (2004). The yeast *Zygosaccharomyces bailii:* a new host for heterologous protein production, secretion and for metabolic engineering applications. *Fems Yeast Research*, 4 (4-5): 493-504.

Bucholtz, M. L. and Light, R. J. (1976). Hydrolysis of 13-sophorosyloxydocosanoic acid esters by acetyl- and carboxylesterases isolated from *Candida bogoriensis*. *The journal of biological chemistry*, 251 (2): 431-437.

Buller, C. S. and Voepel, K. C. (1990). Production and purification of an extracellular polyglucan produced by *Cellulomonas flavigena* strain KU. *Journal of Industrial Microbiology*, 5 (2-3): 139-145.

Burger, M. M.; Glaser, L. and Burton, R. M. (1963). The enzymatic synthesis of a rhamnose-containing glycolipid by extracts of *Pseudomonas aeruginosa*. *The journal of biological chemistry*, 238: 2595-2601.

Butchko, R. A. E.; Brown, D. W.; Busman, M.; Tudzynski, B., *et al.* (2012). Lae1 regulates expression of multiple secondary metabolite gene clusters in *Fusarium verticillioides. Fungal Genetics and Biology*, 49 (8): 602-612.

Cabrera-Valladares, N.; Richardson, A. P.; Olvera, C.; Trevino, L. G., *et al.* (2006). Monorhamnolipids and 3-(3-hydroxyalkanoyloxy)alkanoic acids (HAAs) production using *Escherichia coli* as a heterologous host. *Applied Microbiology and Biotechnology*, 73 (1): 187-194.

Caddick, M. X. (1992). *Characterization of a major Aspergillus regulatory gene, AreA.*

Caddick, M. X.; Jones, M. G.; van Tonder, J. M.; Le Cordier, H., *et al.* (2006). Opposing signals differentially regulate transcript stability in *Aspergillus nidulans*. *Molecular Microbiology*, 62 (2): 509-519.

Calvo, A. M.; Wilson, R. A.; Bok, J. W. and Keller, N. P. (2002). Relationship between secondary metabolism and fungal development. *Microbiology and Molecular Biology Reviews*, 66 (3): 447-+.

Camargo, S.; Valladares, A.; Flores, E. and Herrero, A. (2012). Transcription Activation by NtcA in the Absence of Consensus NtcA-Binding Sites in an *Anabaena* Heterocyst Differentiation Gene Promoter. *Journal of Bacteriology*, 194 (11): 2939-2948.

Cameotra, S. S. and Makkar, R. S. (2004). Recent applications of biosurfactants as biological and immunological molecules. *Current Opinion in Microbiology*, 7 (3): 262-266.

Carr, J. A. and Bisht, K. S. (2003). Enzyme-catalyzed regioselective transesterification of peracylated sophorolipids. *Tetrahedron*, 59 (39): 7713-7724.

Casas, J. A. and Garcia-Ochoa, F. (1999). Sophorolipid production by *Candida bombicola*: Medium composition and culture methods. *Journal of Bioscience and Bioengineering*, 88 (5): 488-494.

Cavalero, D. A. and Cooper, D. G. (2003). The effect of medium composition on the structure and physical state of sophorolipids produced by *Candida bombicola* ATCC 22214. *Journal of Biotechnology*, 103 (1): 31-41.

Cha, M.; Lee, N.; Kim, M. and Lee, S. (2008). Heterologous production of Pseudomonas aeruginosa EMS1 biosurfactant in *Pseudomonas putida*. *Bioresource Technology*, 99 (7): 2192-2199.

Chalfie, M. (1995). Green Fluorescent Protein. *Photochemistry and Photobiology*, 62 (4): 651-656.

Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W., *et al.* (1994). Green Fluorescent Protein as a Marker for Gene-Expression. *Science*, 263 (5148): 802-805.

Chen, G. Q. (2010). *Plastics from Bacteria*. Berlin Heidelberg, Springer-Verlag.

Chen, J.; Song, X.; Zhang, H.; Qu, Y. B., *et al.* (2006). Sophorolipid produced from the new yeast strain *Wickerhamiella domercqiae* induces apoptosis in H7402 human liver cancer cells. *Applied Microbiology and Biotechnology*, 72 (1): 52-59.

Cheng, Y. L.; McNally, D. J.; Labbe, C.; Voyer, N., *et al.* (2003). Insertional mutagenesis of a fungal biocontrol agent led to discovery of a rare cellobiose lipid with antifungal activity. *Applied and Environmental Microbiology*, 69 (5): 2595-2602.

Ciesielska, K.; Roelants, S. L. K. W.; Van Bogaert, I. N. A.; B., D., *et al.* (2011). A lactonase derived from *Candida bombicola* and uses thereof. UGent. PCT/EP2012/075571.

Ciesielska, K.; Roelants, S. L. K. W.; Van Bogaert, I. N. A.; B., D., *et al.* (2013). A lactonase derived from *Candida bombicola* and uses thereof. PCT/EP2012/075571.

Cooper, D. G. and Paddock, D. A. (1984). Production of a biosurfactant from *Torulopsis bombicola*. *Applied and Environmental Microbiology*, 47 (1): 173-176.

Cormack, B. P.; Bertram, G.; Egerton, M.; Gow, N. A. R., *et al.* (1997a). Yeast-enhanced green fluorescent protein (yEGFP): A reporter of gene expression in *Candida albicans*. *Microbiology-Uk*, 143: 303-311.

Cormack, B. P.; Valdivia, R. H. and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, 173 (1): 33-38.

Cormack, B. P.; Valdivia, R. H. and Falkow, S. (1997b). Mutants of green fluorescent protein (GFP) with enhanced fluorescence characteristics. *Bioluminescence and Chemiluminescence - Molecular Reporting with Photons*: 387-390.

Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E., *et al.* (1995). Understanding, Improving and Using Green Fluorescent Proteins. *Trends in Biochemical Sciences*, 20 (11): 448-455.

Cutler, A. J. and Light, R. J. (1979). Regulation of hydroxydocosanoic acid sophoroside production in *Candida bogoriensis* by the levels of glucose and yeast extract in the growth medium. *Journal of Biological Chemistry*, 254 (6): 1944-1950.

Danchin, A. (2009). Cells need safety valves. *BioEssays*, 31: 769-773.

Daniel, H. J.; Otto, R. T.; Binder, M.; Reuss, M., *et al.* (1999). Production of sophorolipids from whey: development of a two-stage process with *Cryptococcus curvatus* ATCC 20509 and *Candida bombicola* ATCC 22214 using deproteinized whey concentrates as substrates. *Applied Microbiology and Biotechnology*, 51 (1): 40-45.

Daniel, H. J.; Reuss, M. and Syldatk, C. (1998). 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. *Biotechnology Letters*, 20 (12): 1153-1156.

Das, P.; Mukherjee, S. and Sen, R. (2008). Genetic Regulations of the Biosynthesis of Microbial Surfactants: An Overview. *Biotechnology and Genetic Engineering Reviews, Vol 25*. Harding, S. E. and Tombs, M. P. 25: 165-185.

Daverey, A. and Pakshirajan, K. (2009a). Production of sophorolipids by the yeast *Candida bombicola* using simple and low cost fermentative media. *Food Research International*, 42 (4): 499-504.

Daverey, A. and Pakshirajan, K. (2009b). Production, Characterization, and Properties of Sophorolipids from the Yeast *Candida bombicola* using a Low-cost Fermentative Medium. *Applied Biochemistry and Biotechnology*, 158 (3): 663-674.

Daverey, A. and Pakshirajan, K. (2010). Kinetics of growth and enhanced Sophorolipids Production by *Candida bombicola* Using a Low-Cost Fermentative Medium. *Applied Biochemistry and Biotechnology*, 160 (7): 2090-2101.

Daverey, A., Pakshirajan, K., Sumalatha, S. (2011). Sophorolipids production by *Candida bombicola* using dairy industry wastewater. *Clean Technologies and Environmental Policy*, 13 (3): 481-488.

Davey, M. E.; Caiazza, N. C. and O'Toole, G. A. (2003). Rhamnolipid surfactant production affects biofilm architecture in *Pseudomonas aeruginosa* PAO1. *Journal of Bacteriology*, 185 (3): 1027-1036.

Davila, A. M.; Marchal, R.; Monin, N. and Vandecasteele, J. P. (1993). Identification and determination of individual sophorolipids in fermenatation products by gradient elution high performance liquid chromatography with ELSD detection. *Journal of Chromatography*, 648 (1): 139-149.

Davila, A. M.; Marchal, R. and Vandecasteele, J. P. (1992). Kinetics and balance of a fermentation free from product inhibition - sophorose lipid production by *Candida bombicola Applied Microbiology and Biotechnology*, 38 (1): 6-11.

Davila, A. M.; Marchal, R. and Vandecasteele, J. P. (1994). Sophorose Lipid Production from Lipidic Precursors - Predictive Evaluation of Industrial Substrates. *Journal of Industrial Microbiology*, 13 (4): 249-257.

Davila, A. M.; Marchal, R. and Vandecasteele, J. P. (1997). Sophorose lipid fermentation with differentiated substrate supply for growth and production phases. *Applied Microbiology and Biotechnology*, 47 (5): 496-501.

De las Penas, A.; Pan, S. J.; Castano, I.; Alder, J., *et al.* (2003). Virulence-related surface glycoproteins in the yeast pathogen *Candida glabrata* are encoded in subtelomeric clusters and subject to RAP1- and SIR-dependent transcriptional silencing. *Genes & Development*, 17 (18): 2245-2258.

De Mey, M.; Maertens, J.; Lequeux, G. J.; Soetaert, W. K., *et al.* (2007). Construction and model-based analysis of a promoter library for *E-coli*: an indispensable tool for metabolic engineering. *Bmc Biotechnology*, 7: 14.

deFerra, F.; Rodriguez, F.; Tortora, O.; Tosi, C., *et al.* (1997). Engineering of peptide synthetases - Key role of the thioesterase-like domain for efficient production of recombinant peptides. *Journal of Biological Chemistry*, 272 (40): 25304-25309.

Deinema, M. H. (1961). Intra- en extracellular lipid production by yeasts. Wageningen, Landbouwhogeschool Wageningen.

Dekoster, C. G.; Heerma, W.; Pepermans, H. A. M.; Groenewegen, A., *et al.* (1995). Tandem mass spcetrometry and nuclear magnetic resonance spectroscopy studies of *Candida bombicola* sophorolipids and product formed on hydrolysis by cutinase. *Analytical Biochemistry*, 230 (1): 135-148.

Deleu, M.; Bouffioux, O.; Razafindralambo, H.; Paquot, M., *et al.* (2003). Interaction of surfactin with membranes: A computational approach. *Langmuir*, 19 (8): 3377-3385.

Desai, J. D. and Banat, I. M. (1997). Microbial production of surfactants and their commercial potential. *Microbiology and Molecular Biology Reviews*, 61 (1): 47-64.

Develter, D. and Fleurackers, S. (2007). A method for the production of short chained glycolipids. Ecover. EP1953237.

Develter, D. and Renkin, M. (2012). Sophorolactone compositions and uses thereof. CT, E. C.

Deziel, E.; Lepine, F.; Milot, S. and Villemur, R. (2003). RhlA is required for the production of a novel biosurfactant promoting swarming motility in *Pseudomonas aeruginosa*: 3-(3-hydroxyalkanoyloxy)alkanoic acids (HAAs), the precursors of rhamnolipids. *Microbiology-Sgm*, 149: 2005-2013.

Doi, Y.; Kawaguchi, Y.; Koyama, N.; Nakamura, S., *et al.* (1992). Synthesis and degradation of polyhydroxyalkanoates in *Alcaligenes eutrophus*. *Fems Microbiology Reviews*, 9 (2-4): 103-108.

Domergue, R.; Castano, I.; De las Penas, A.; Zupancic, M., *et al.* (2005). Nicotinic acid limitation regulates silecing of *Candida* adhesins during UTI. *Science*, 308 (5723): 866-870.

Dowzer, C. E. A. and Kelly, J. M. (1989). Cloning of the CreA gene from *Aspergillus nidulans*- A gene involved in carbon catabolite repression. *Current Genetics*, 15 (6): 457-459.

Dreja, M.; Vockenroth, I. and Plath, N. (2012). Biosurfactants - Exotic Specialties or Ready for Application? *Tenside Surfactants Detergents*, 49 (1): 10-17.

Dsouza, C.; Nakano, M. M. and Zuber, P. (1994). Identification of ComS, a gene of the SrfA operon that regulates the establishment of genetic competence in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America*, 91 (20): 9397-9401.

Dubeau, D.; Deziel, E.; Woods, D. E. and Lepine, F. (2009). *Burkholderia thailandensis* harbors two identical rhl gene clusters responsible for the biosynthesis of rhamnolipids. *Bmc Microbiology*, 9: 12.

Dunlap, C. A.; Bowman, M. J. and Schisler, D. A. (2012). Genomic analysis and secondary metabolite production in *Bacillus amyloliquefaciens* AS 43.3: a biocontrol antagonist of Fusarium Head Blight. *Biological Control*.

Edwards, J. R. and Hayashi, J. A. (1965). Structure of a rhamnolipid from *Pseudomonas aeruginosa*. *Archives for Biochemistry and Biophysics*, 111: 415-421.

Ehira, S.; Ohmori, M. and Sato, N. (2003). Genome-wide expression analysis of the responses to nitrogen deprivation in the heterocyst-forming cyanobacterium *Anabacna sp* strain PCC 7120. *DNA Research*, 10 (3): 97-113.

Ehrlich, K. C.; Montalbano, B. G. and Cotty, P. J. (2005). Divergent regulation of aflatoxin production at acidic pH by two *Aspergillus* strains. *Mycopathologia*, 159 (4): 579-581.

Esders, T. W. and Light, R. J. (1972a). Characterization and in-vivo production of three glycolipids from *Candida bogoriensis* - 13-glucopyranosylglucopyranosyloxydocosanoic acid and its monoacetylated and diacetylated derivatives. *Journal of Lipid Research*, 13 (5): 663-&.

Esders, T. W. and Light, R. J. (1972b). Glucosyltransferases and acetyltransferases involved in biosynthesis of glycolipids from *Candida bogoriensis*. *Journal of Biological Chemistry*, 247 (5): 1375-&.

Evans, C. T. and Ratledge, C. (1985). Possible regulatory roles of ATP-citrate lyase, malic enzyme and AMP deaminase in Lipid accumulation by *Rhodosporidium toruloides* CBS-14. *Canadian Journal of Microbiology*, 31 (11): 1000-1005.

Eveleigh, D. E.; Dateo, G. P. and Reese, E. T. (1964). Fungal metabolism of complex glycosides - Ustilagic acid. *Journal of Biological Chemistry*, 239 (3): 839-&.

Fairhead, C. and Dujon, B. (2006). Structure of *Kluyveromyces lactis* subtelomeres: duplications and gene content. *Fems Yeast Research*, 6 (3): 428-441.

Fan, Q.; Huang, G. C.; Lechno-Yossef, S.; Wolk, C. P., *et al.* (2005). Clustered genes required for synthesis and deposition of envelope glycolipids in *Anabaena sp* strain PCC 7120. *Molecular Microbiology*, 58 (1): 227-243.

Felse, A. P.; Shah, V.; Chan, J.; Rao, K. J., *et al.* (2007). Sophorolipid biosynthesis by *Candida bombicola* from industrial fatty acid residues. *Enzyme and Microbial Technology*, 40: 316-323.

Fiedler, G.; Arnold, M.; Hannus, S. and Maldener, I. (1998). The DevBCA exporter is essential for envelope formation in heterocysts of the cyanobacterium *Anabaena sp.* strain PCC 7120. *Molecular Microbiology*, 27 (6): 1193-1202.

Finking, R. and Marahiel, M. A. (2004). Biosynthesis of nonribosomal peptides. *Annual Review of Microbiology*, 58: 453-488.

Flaherty, J. E. and Woloshuk, C. P. (2004). Regulation of fumonisin biosynthesis in Fusarium verticillioides by a zinc binuclear cluster-type gene, ZFR1. *Applied and Environmental Microbiology*, 70 (5): 2653-2659.

Fleurackers, S. J. J. (2006). On the use of waste frying oil in the synthesis of sophorolipids. *European Journal of Lipid Science and Technology*, 108 (1): 5-12.

Fleurackers, S. J. J.; Van Bogaert, I. N. A. and Develter, D. (2010). On the production and identification of medium-chained sophorolipids. *European Journal of Lipid Science and Technology*, 112 (6): 655-662.

Flores, E. and Herrero, A. (2010). Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nature Reviews Microbiology*, 8 (1): 39-50.

Follner, C. G.; Muller, S.; Steinbuchel, A. and Babel, W. (1995). Biosynthesis of poly-3-hydroxybutyric acid by the facultatively methanol-assimilating bacterium *Mycoplana rubru* B346 and recombinant strains. *Journal of Basic Microbiology*, 35 (3): 179-188.

Fox, E. M. and Howlett, B. J. (2008). Secondary metabolism: regulation and role in fungal biology. *Current Opinion in Microbiology*, 11 (6): 481-487.

Freitas-Junior, L. H.; Hernandez-Rivas, R.; Ralph, S. A.; Montiel-Condado, D., *et al.* (2005). Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell*, 121 (1): 25-36.

Fu, Y. H. and Marzluf, G. A. (1987). Characterization of Nit-2, the major nitrogen regulatory gene of *Neurospora crass*a. *Molecular and Cellular Biology*, 7 (5): 1691-1696.

Galli, G.; Rodriguez, F.; Cosmina, P.; Pratesi, C., *et al.* (1994). Characterization of the surfaction synthetase multienzyme complex. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology*, 1205 (1): 19-28.

Gambacorta, A.; Pagnotta, E.; Romano, I.; Sodano, G., *et al.* (1998). Heterocyst glycolipids from nitrogen-fixing *cyanobacteria* other than *Nostocaceae. Phytochemistry*, 48 (5): 801-805.

Gambacorta, A.; Romano, I.; Trincone, A.; Soriente, A., *et al.* (1996). The heterocyst glycolipids of cyanobacteria. 5. Heterocyst glycolipids from five nitrogen-fixing cyanobacteria. *Gazzetta Chimica Italiana*, 126 (10): 653-656.

Garcia-Ochoa, F. a. C., J.P. (1996). Procedimiento para la produccion de sophorosa por *Candida bombicola*. marcas, O. e. d. p. y. Spain: 4.

Gietz, R. D. and Schiestl, R. H. (1995). Transforming yeast with DNA. *Methods in Molecular and Cellular Biology*, 5 (5): 255-269.

Gobbert, U.; Lang, S. and Wagner, F. (1984). Sophorose lipid formation by resting cells of *Torulopsis bombicola*. *Biotechnology Letters*, 6 (4): 225-230.

Golubev, W. I.; Kulakovskaya, T. V.; Shashkov, A. S.; Kulakovskaya, E. V., *et al.* (2008). Antifungal cellobiose lipid secreted by the epiphytic yeast *Pseudozyma graminicola*. *Microbiology*, 77 (2): 171-175.

Gorin, P. A. J.; Spencer, J. F. T. and Tulloch, A. P. (1961). Hydroxy fatty acid glycosides of sophorose from *Torulopsis magnoliae*. *Canadian Journal of Chemistry*, 39: 846-855.

Grangemard, I.; Peypoux, F.; Wallach, J.; Das, B. C., *et al.* (1997). Lipopeptides with improved properties: Structure by NMR, purification by HPLC and structure-activity relationships of new isoleucyl-rich surfactins. *Journal of Peptide Science*, 3 (2): 145-154.

Grangemard, I.; Wallach, J. and Peypoux, F. (1999). Evidence of surfactin hydrolysis by a bacterial endoprotease. *Biotechnology Letters*, 21 (3): 241-244.

Gross, R. A.; Guilmanov, V. and Scholz, C. (1999). Glycolipids from *Torulopsis bombicola*: Biosynthesis, lipase-selective modifications and anti-cancer activity. *Abstracts of Papers of the American Chemical Society*, 217: 181-BIOT.

Gross, R. A. and Shah, V. (2004). Antimicrobial properties of various forms of sophorolipids. World patent WO2004/044216 A1.

Gross, R. A. and Shah, V. (2005). Antifungal properties of various forms of sophorolipids. US Patent US2005/0164955 A1.

Gross, R. A.; Shah, V. and Doncel, G. F. (2004). Spermicidal and virucidal properties of various forms of sophorolipids. Technoprop Colton, L. L. C., P O: 9.

Grossman, A. D. (1995). Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. *Annual Review of Genetics*, 29: 477-508.

Guerra-Santos, L.; Kappeli, O. and Fiechter, A. (1984). *Pseudomonas-Aeruginosa* Biosurfactant Production in Continuous Culture with Glucose as Carbon Source. *Applied and Environmental Microbiology*, 48 (2): 301-305.

Guilmanov, V.; Ballistreri, A.; Impallomeni, G. and Gross, R. A. (2002). Oxygen transfer rate and sophorose lipid production by *Candida bombicola*. *Biotechnology and Bioengineering*, 77 (5): 489-494.

Gutierrez, R. M. P.; Gonzalez, A. M. N. and Ramirez, A. M. (2012). Compounds Derived from Endophytes: A Review of Phytochemistry and Pharmacology. *Current Medicinal Chemistry*, 19 (18): 2992-3030.

Haas, H.; Bauer, B.; Redl, B.; Stoffler, G., *et al.* (1995). Molecular cloning and analysis of nre, the major nitrogen regulatory gene of *Penicillium chrysogenum*. *Current Genetics*, 27 (2): 150-158.

Haba, E.; Pinazo, A.; Jauregui, O.; Espuny, M. J., *et al.* (2003). Physicochemical characterization and antimicrobial properties of rhamnolipids produced by *Pseudomonas aeruginosa* 47T2 NCBIM 40044. *Biotechnology and Bioengineering*, 81 (3): 316-322.

Haddouche, R.; Delessert, S.; Sabirova, J.; Neuveglise, C., *et al.* (2010). Roles of multiple acyl-CoA oxidases in the routing of carbon flow towards beta-oxidation and polyhydroxyalkanoate biosynthesis in *Yarrowia lipolytica*. *Fems Yeast Research*, 10 (7): 917-927.

Haddouche, R.; Poirier, Y.; Delessert, S.; Sabirova, J., *et al.* (2011). Engineering polyhydroxyalkanoate content and monomer composition in the oleaginous yeast *Yarrowia lipolytica* by modifying the beta-oxidation multifunctional protein. *Applied Microbiology and Biotechnology*, 91 (5): 1327-1340.

Halme, A.; Bumgarner, S.; Styles, C. and Fink, G. R. (2004). Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast. *Cell*, 116 (3): 405-415.

Halsall, J. R.; Milner, M. J. and Casselton, L. A. (2000). Three subfamilies of pheromone and receptor genes generate multiple B mating specificities in the mushroom *Coprinus cinereus*. *Genetics*, 154 (3): 1115-1123.

Hammami, W.; Labbe, C.; Chain, F.; Mimee, B., *et al.* (2008). Nutritional regulation and kinetics of flocculosin synthesis by *Pseudozyma flocculosa*. *Applied Microbiology and Biotechnology*, 80 (2): 307-315.

Hamoen, L. W.; Eshuis, H.; Jongbloed, J.; Venema, G., *et al.* (1995). A small gene, designated ComS, located within the coding region of the 4th amino acid activation domain of SrfA is required for competence development in *Bacillus subtilis*. *Molecular Microbiology*, 15 (1): 55-63.

Haskins, R. H. (1950). Biochemistry of the ustilaginales: I. Preliminary cultural studies of *Ustilago zeae*. *Canadian Journal of Research*, 28c: 213-223.

Haury, J. F. and Wolk, C. P. (1978). Classes of *Anabaena variabilis* mutants with oxygen sensitive nitrogenase activity. *Journal of Bacteriology*, 136 (2): 688-692.

Hauser, G. and Karnovsky, M. (1954). Studies on the production of glycolipide by *Pseudomonas aeruginosa*. *Journal of bacteriology*, 68 (6): 645-654.

Heinz, E.; Tulloch, A. P. and Spencer, J. F. T. (1969). Stereospecific hydroxylation of long chain compounds by a species of *Torulopsis*. *Journal of Biological Chemistry*, 244 (4): 882-&.

Herrero, A.; Muro-Pastor, A. M. and Flores, E. (2001). Nitrogen control in *cyanobacteria*. *Journal of Bacteriology*, 183 (2): 411-425.

Hewald, S.; Josephs, K. and Bolker, M. (2005). Genetic analysis of biosurfactant production in *Ustilago maydis*. *Applied and Environmental Microbiology*, 71 (6): 3033-3040.

Hewald, S.; Linne, U.; Scherer, M.; Marahiel, M. A., *et al.* (2006). Identification of a gene cluster for biosynthesis of mannosylerythritol lipids in the basidiomycetous fungus *Ustilago maydis*. *Applied and Environmental Microbiology*, 72 (8): 5469-5477.

Higa, K. C.; Rajagopalan, R.; Risser, D. D.; Rivers, O. S., *et al.* (2012). The RGSGR amino acid motif of the intercellular signalling protein, HetN, is required for patterning of heterocysts in *Anabaena sp* strain PCC 7120. *Molecular Microbiology*, 83 (4): 682-693.

Hoffmeister, D. and Keller, N. P. (2007). Natural products of filamentous fungi: enzymes, genes, and their regulation. *Natural Product Reports*, 24 (2): 393-416.

Hommel, R. K. (1990). Formation and physiological role of biosurfactants produced by hydrocarbon-utilizing microorganisms. *Biodegradation*, 1: 107-119.

Hommel, R. K. and Huse, K. (1993). Regulation of sophorose lipid production by *Candida (Torulopsis) Apicola*. *Biotechnology Letters*, 15 (8): 853-858.

Hommel, R. K.; Stegner, S.; Weber, L. and Kleber, H. P. (1994a). Effect of ammonium ions on glycolipid production by *Candida (Torulopsis) apicola*. *Applied Microbiology and Biotechnology*, 42 (2-3): 192-197.

Hommel, R. K.; Weber, L.; Weiss, A.; Himmelreich, U., *et al.* (1994b). Production of Sophorose Lipid by *Candida (Torulopsis) Apicola* Grown on Glucose. *Journal of Biotechnology*, 33 (2): 147-155.

Horn, D. and Barry, J. D. (2005). The central roles of telomeres and subtelomeres in antigenic variation in African *trypanosomes*. *Chromosome Research*, 13 (5): 525-533.

Hu, Y. (2000). Production, purification and enzymatic polymerization of sophorolipids. *Department of Chemical Engineering*. Akron, University of Akron: 159.

Hu, Y. M. and Ju, L. K. (2001). Purification of lactonic sophorolipids by crystallization. *Journal of Biotechnology*, 87 (3): 263-272.

Hynes, M. J. (1975). Studies on the role of AreA gene in regulation of nitrogen catabolism in *Aspergillus nidulans*. *Australian Journal of Biological Sciences*, 28 (3): 301-313.

Imura, T.; Masuda, Y.; Minamikawa, H.; Fukuoka, T., *et al.* (2010). Enzymatic conversion of diacetylated sophoroselipid into acetylated glucoselipid: surface-active properties of novel bolaform biosurfactants. *Journal of Oleo Science*, 59 (9): 495-501.

Inoue, S.; Kimura, Y. and Kinta, M. (1980). Dehydrating purification process for a fermentation product. US.

Iqbal, S.; Khalid, Z. M. and Malik, K. A. (1995). Enhanced biodegradation and emulsification of crude oil and hyperproduction of biosurfactants by a gamma ray induced mutant of *Pseudomonas aeruginosa*. *Letters in Applied Microbiology*, 21 (3): 176-179.

Irie, Y.; O'Toole, G. A. and Yuk, M. H. (2005). *Pseudomonas aeruginosa* rhamnolipids disperse *Bordetella bronchiseptica* biofilms. *Fems Microbiology Letters*, 250 (2): 237-243.

Ito, S. and Inoue, S. (1982). Sophorolipids from *Torulopsis bombicola*- Possible relation to alkane uptake. *Applied and Environmental Microbiology*, 43 (6): 1278-1283.

Ito, S.; Kinta, M. and Inoue, S. (1980). Growth of yeasts on normal-alkanes-Inhibition by a lactonic sophorolipid produced by *Torulopsis bombicola*. *Agricultural and Biological Chemistry*, 44 (9): 2221-2223.

Itoh, S.; Honda, H.; Tomita, F. and Suzuki, T. (1971). Rhamnolipids produced by *Pseudomonas aeruginosa* grown on paraffin (mixture of C12, C13 and C14 fractions). *Journal of Antibiotics*, 24 (12): 855-&.

Jarvis, F. G. and Johnson, M. J. (1949). A glyco-lipid produced by *Pseudomonas aeruginosa*. *Journal of the American Oil Chemists Society*, 71: 4124-4126.

Jones, D. F. (1967). Novel macrocyclic glycolipids from *Torulopsis gropengiesseri*. *Journal of the Chemical Society C-Organic* (6): 479-&.

Jung, J.; Yu, K. O.; Ramzi, A. B.; Choe, S. H., *et al.* (2012). Improvement of surfactin production in *Bacillus subtilis* using synthetic wastewater by overexpression of specific extracellular signaling peptides, comX and phrC. *Biotechnology and Bioengineering*, 109 (9): 2349-2356.

Kakugawa, K.; Shobayashi, M.; Suzuki, O. and Miyakawa, T. (2002a). Cloning, characterization, and expression of cDNA encoding a lipase from Kurtzmanomyces sp I-11. *Bioscience Biotechnology and Biochemistry*, 66 (6): 1328-1336.

Kakugawa, K.; Tamai, M.; Imamura, K.; Miyamoto, K., *et al.* (2002b). Isolation of yeast Kurtzmanomyces sp I-11, novel producer of mannosylerythritol lipid. *Bioscience Biotechnology and Biochemistry*, 66 (1): 188-191.

Kasture, M.; Singh, S.; Patel, P.; Joy, P. A., *et al.* (2007). Multiutility sophorolipids as nanoparticle capping agents: Synthesis of stable and water dispersible Co nanoparticles. *Langmuir*, 23 (23): 11409-11412.

Keller, N. P. and Hohn, T. M. (1997). Metabolic pathway gene clusters in filamentous fungi. *Fungal Genetics and Biology*, 21 (1): 17-29.

Keller, N. P.; Turner, G. and Bennett, J. W. (2005). Fungal secondary metabolism - From biochemistry to genomics. *Nature Reviews Microbiology*, 3 (12): 937-947.

Kim, H. S.; Kim, Y. B.; Lee, B. S. and Kim, E. K. (2005). Sophorolipid production by *Candida bombicola* ATCC 22214 from a corn-oil processing byproduct. *Journal of Microbiology and Biotechnology*, 15 (1): 55-58.

Kiss, L. (2003). A review of fungal antagonists of powdery mildews and their potential as biocontrol agents. *Pest Management Science*, 59 (4): 475-483.

Kitamoto, D.; Akiba, S.; Hioki, C. and Tabuchi, T. (1990). Extracellular accumulation of mannosylerythritol lipids by a strain of *Candida antarctica. Agricultural and Biological Chemistry*, 54 (1): 31-36.

Kitamoto, D.; Fuzishiro, T.; Yanagishita, H.; Nakane, T., *et al.* (1992a). Production of mannosylerythritol lipids as biosurfactants by resting cells of *Candida antarctica*. *Biotechnology Letters*, 14 (4): 305-310.

Kitamoto, D.; Isoda, H. and Nakahara, T. (2002). Functions and potential applications of glycolipid biosurfactants - from energy-saving materials to gene delivery carriers. *Journal of Bioscience and Bioengineering*, 94 (3): 187-201.

Kitamoto, D.; Morita, T.; Fukuoka, T.; Konishi, M., *et al.* (2009). Self-assembling properties of glycolipid biosurfactants and their potential applications. *Current Opinion in Colloid & Interface Science*, 14: 315-328.

Kitamoto, D.; Nakane, T.; Nakao, N.; Nakahara, T., *et al.* (1992b). Intracellular accumulation of mannosylerythritol lipids as storage materials by *Candida antarctica*. *Applied Microbiology and Biotechnology*, 36 (6): 768-772.

Klekner, V.; Kosaric, N. and Zhou, Q. H. (1991). Sophorose lipids produced from sucrose. *Biotechnology Letters*, 13 (5): 345-348.

Koch, A. K.; Kappeli, O.; Fiechter, A. and Reiser, J. (1991). Hydrocarbon assimilation and biosurfactant production in *Pseudomonas aeruginosa* mutants. *Journal of Bacteriology*, 173 (13): 4212-4219.

Kolkman, A.; Daran-Lapujade, P.; Fullaondo, A.; Olsthoorn, M. M. A., *et al.* (2006). Proteome analysis of yeast response to various nutrient limitations. *Molecular Systems Biology*, 2: 16.

Konishi, M.; Fukuoka, T.; Morita, T.; Imura, T., *et al.* (2008). Production of new types of sophorolipids by *Candida batistae*. *Journal of Oleo Science*, 57 (6): 359-369.

Koponen, O.; Takala, T. M.; Saarela, U.; Qiao, M. Q., *et al.* (2004). Distribution of the NisI immunity protein and enhancement of nisin activity by the lipid-free NisI. *Fems Microbiology Letters*, 231 (1): 85-90.

Kraas, F. I.; Helmetag, V.; Wittmann, M.; Strieker, M., *et al.* (2010). Functional Dissection of Surfactin Synthetase Initiation Module Reveals Insights into the Mechanism of Lipoinitiation. *Chemistry & Biology*, 17 (8): 872-880.

Kroken, S.; Glass, N. L.; Taylor, J. W.; Yoder, O. C., *et al.* (2003). Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proceedings of the National Academy of Sciences of the United States of America*, 100 (26): 15670-15675.

Kulakovskaya, E. V.; Kulakovskaya, T. V.; Golubev, V. I.; Shashkov, A. S., *et al.* (2007). Fungicidal activity of cellobiose lipids from culture broth of yeast *Cryptococcus humicola* and *Pseudozyma fusiformata*. *Russian Journal of Bioorganic Chemistry*, 33 (1): 156-160.

Kulakovskaya, T.; Shashkov, A.; Kulakovska, E.; Golubev, W., *et al.* (2009). Extracellular Cellobiose Lipid from Yeast and Their Analogues: Structures and Fungicidal Activities. *Journal of Oleo Science*, 58 (3): 133-140.

Kulakovskaya, T. V.; Golubev, W. I.; Tomashevskaya, M. A.; Kulakovskaya, E. V., *et al.* (2010). Production of Antifungal Cellobiose Lipids by *Trichosporon porosum*. *Mycopathologia*, 169 (2): 117-123.

Kulakovskaya, T. V.; Kulakovskaya, E. V. and Golubev, W. I. (2003). ATP leakage from yeast cells treated by extracellular glycolipids of *Pseudozyma fusiformata*. *Fems Yeast Research*, 3 (4): 401-404.

Kulakovskaya, T. V.; Shashkov, A. S.; Kulakovskaya, E. V. and Golubev, W. I. (2004). Characterization of an antifungal glycolipid secreted by the yeast *Sympodiomycopsis paphiopedili*. *Fems Yeast Research*, 5: 247-252.

Kulakovskaya, T. V.; Shashkov, A. S.; Kulakovskaya, E. V. and Golubev, W. I. (2005). Ustilagic acid secretion by *Pseudozyma fusiformata* strains. *Fems Yeast Research*, 5: 919-923.

Kurtzman, C. P.; Price, N. P. J.; Ray, K. J. and Kuo, T. M. (2010). Production of sophorolipid biosurfactants by multiple species of the *Starmerella (Candida) bombicola* yeast clade. *Fems Microbiology Letters*, 311: 140-146.

Kurz, M.; Eder, C.; Isert, D.; Li, Z. Y., *et al.* (2003). Ustilipids, acylated beta-D-mannopyranosyl D-erythritols from *Ustilago maydis* and *Geotrichum candidum*. *Journal of Antibiotics*, 56 (2): 91-101.

Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P., *et al.* (1996). A new enzyme superfamily - The phosphopantetheinyl transferases. *Chemistry & Biology*, 3 (11): 923-936.

Landan, G.; Cohen, G.; Aharonowitz, Y.; Shuali, Y., *et al.* (1990). Evolution of isopenicillin-N synthase genes may have involved horizontal gene transfer. *Molecular Biology and Evolution*, 7 (5): 399-406.

Lang, S.; Brakemeier, A.; Heckmann, R.; Spockner, S., *et al.* (2000). Production of native and modified sophorose lipids. *Chimica Oggi-Chemistry Today*, 18 (10): 76-79.

Lang, S. and Wullbrandt, D. (1999). Rhamnose lipids - biosynthesis, microbial production and application potential. *Applied Microbiology and Biotechnology*, 51 (1): 22-32.

Latifi, A.; Winson, M. K.; Foglino, M.; Bycroft, B. W., *et al.* (1995). Multiple homologs of LuxR and LuxL control expression of virulence determinants and secondary metabolites through quorum sensing in *Pseudomonas aeruginosa* PAO1. *Molecular Microbiology*, 17 (2): 333-343.

Leaf, T. A.; Peterson, M. S.; Stoup, S. K.; Somers, D., *et al.* (1996). *Saccharomyces cerevisiae* expressing bacterial polyhydroxybutyrate synthase produces poly-3-hydroxybutyrate. *Microbiology-Uk*, 142: 1169-1180.

Lee, K. T.; Kim, J. Y.; Rhee, Y. H.; Bae, K. S., *et al.* (1995). Biosynthesis of poly-beta-hydroxyalkanoates by *Bacillus thuringiensis* R-510. *Journal of Microbiology*, 33 (1): 59-65.

Lee, S. Y. (1996). Bacterial polyhydroxyalkanoates. *Biotechnology and Bioengineering*, 49 (1): 1-14.

Lepine, F.; Deziel, E.; Milot, S. and Villemur, R. (2002). Liquid chromatographic/mass spectrometric detection of the 3-(3-hydroxyalkanoyloxy)alkanoic acid precursors of

rhamnolipids in *Pseudomonas aeruginosa* cultures. *Journal of Mass Spectrometry*, 37 (1): 41-46.

Levinson, M. I. (2008). Surfactant production: present realities and future perspectives. *Handbook of detergents*. Zoller, I. and Sosis, P., CRC Press. 142: 1-37.

Li, Y. M.; Yang, S. Z. and Mu, B. Z. (2010). The surfactin and lichenysin isoforms produced by *Bacillus licheniformis* HSN 221. *Analytical Letters*, 43 (6): 929-940.

Litzka, O.; Bergh, K. T.; Van den Brulle, J.; Steidl, S., *et al.* (1999). Transcriptional control of expression of fungal beta-lactam biosynthesis genes. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology*, 75 (1-2): 95-105.

Liu, J. and Chen, W. L. (2009). Characterization of HetN, a protein involved in heterocyst differentiation in the *cyanobacterium Anabaena sp* strain PCC 7120. *Fems Microbiology Letters*, 297 (1): 17-23.

Lo, C.-M. and Ju, L.-K. (2009). Sophorolipids-induced cellulase production in cocultures of Hypocrea jecorina Rut C30 and *Candida bombicola*. *Enzyme and Microbial Technology*, 44 (2): 107-111.

Luque, I.; Flores, E. and Herrero, A. (1994). Molecular mechanism for the operation of nitrogen control in *Cyanobacteria. Embo Journal*, 13 (23): 5794-5794.

Magasanik, B. and Kaiser, C. A. (2002). Nitrogen regulation in *Saccharomyces cerevisiae*. *Gene*, 290 (1-2): 1-18.

Mager, H.; Röthlisberger, R. and Wagner, F. (1987). Use of sophorose-lipid lactone for the treatment of dandruffs and body odor. European Patent 0209783.

Maier, R. M. (2003). Biosurfactants: Evolution and diversity in bacteria. *Advances in Applied Microbiology, Vol 52*, 52: 101-121.

Maingault, M. (1999). Utilization of sophorolipids as therapeutically active substances or cosmetic products, in particular for the treatment of the skin. France, Institut Francais du Petrole, Rueil-Malmaison Cedex (FR): 7.

Mandels, M.; Reese, E. T. and Parrish, F. W. (1962). Sophorose as an inducer of cellulase in *Trichoderma viride*. *Journal of Bacteriology*, 83 (2): 400-&.

Marchand, G.; Remus-Borel, W.; Chain, F.; Hammami, W., *et al.* (2009). Identification of Genes Potentially Involved in the Biocontrol Activity of *Pseudozyma flocculosa*. *Phytopathology*, 99 (10): 1142-1149.

Marchant, R. and Banat, I. (2012). Biosurfactants: a sustainable replacement for chemical surfactants? *Biotechnology Letters*, 34 (9): 1597-1605.

Marchesini, S.; Erard, N.; Glumoff, T.; Hiltunen, J. K., *et al.* (2003). Modification of the monomer composition of polyhydroxyalkanoate synthesized in *Saccharomyces cerevisiae*

expressing variants of the beta-oxidation-associated multifunctional enzyme. *Applied and Environmental Microbiology*, 69 (11): 6495-6499.

Marchesini, S. and Poirier, Y. (2003). Futile cycling of intermediates of fatty acid biosynthesis toward peroxisomal beta-oxidation in *Saccharomyces cerevisiae. Journal of Biological Chemistry*, 278 (35): 32596-32601.

Martin, J. F. (2000). Molecular control of expression of penicillin biosynthesis genes in fungi: Regulatory proteins interact with a bidirectional promoter region. *Journal of Bacteriology*, 182 (9): 2355-2362.

Marzluf, G. A. (1997). Genetic regulation of nitrogen metabolism in the fungi. *Microbiology and Molecular Biology Reviews*, 61 (1): 17-&.

Matsuyama, T. and Nakagawa, Y. (1996). Bacterial wetting agents working in colonization of bacteria on surface environments. *Colloids and Surfaces B-Biointerfaces*, 7 (5-6): 207-214.

McDonagh, A.; Fedorova, N. D.; Crabtree, J.; Yu, Y., *et al.* (2008). Sub-telomere directed gene expression during initiation of invasive aspergillosis. *Plos Pathogens*, 4 (9): 21.

McGrath, S.; Wade, D. S. and Pesci, E. C. (2004). Dueling quorum sensing systems in *Pseudomonas aeruginosa* control the production of the *Pseudomonas* quinolone signal (PQS). *Fems Microbiology Letters*, 230 (1): 27-34.

McInerney, M. J., Amos, D. A., Kealy, K. S., Palmer, J. A. (1992). Synthesis and function of polyhydroxyalkanoates in anaerobic syntrophic bacteria. *Fems Microbiology Reviews*, 9 (2-4): 195-205.

Menkhaus, M.; Ullrich, C.; Kluge, B.; Vater, J., *et al.* (1993). Structural and functional organization of the surfactin synthetase multienzyme system. *Journal of Biological Chemistry*, 268 (11): 7678-7684.

Mimee, B.; Labbe, B.; Pelletier, R. and Belanger, R. R. (2005). Antifungal activity of flocculosin, a novel glycolipid isolated from *Pseudozyma flocculosa*. *Antimicrobial Agents and Chemotherapy*, 49 (4): 1597-1599.

Mimee, B.; Labbe, C. and Belanger, R. R. (2009a). Catabolism of flocculosin, an antimicrobial metabolite produced by *Pseudozyma flocculosa*. *Glycobiology*, 19 (9): 995-1001.

Mimee, B.; Pelletier, R. and Belanger, R. R. (2009b). In vitro antibacterial activity and antifungal mode of action of flocculosin, a membrane-active cellobiose lipid. *Journal of Applied Microbiology*, 107 (3): 989-996.

Minehart, P. L. and Magasanik, B. (1991). Sequence and expression of Gln3, a positive nitrogen regulatory gene of *Saccharomyces cerevisiae* encoding a protein with a putative zinc finger DNA-binding domain. *Molecular and Cellular Biology*, 11 (12): 6216-6228.

Moslavac, S.; Nicolaisen, K.; Mirus, O.; Al Dehni, F., *et al.* (2007). A TolC-Like protein is required for heterocyst development in *Anabaena sp* strain PCC 7120. *Journal of Bacteriology*, 189 (21): 7887-7895.

Mukherjee, A. K. and Das, K. (2005). Correlation between diverse cyclic lipopeptides production and regulation of growth and substrate utilization by Bacillus subtilis strains in a particular habitat. *Fems Microbiology Ecology*, 54 (3): 479-489.

Muller, M. M. and Hausmann, R. (2011). Regulatory and metabolic network of rhamnolipid biosynthesis: Traditional and advanced engineering towards biotechnological production. *Applied Microbiology and Biotechnology*, 91 (2): 251-264.

Mulligan, C. N. (2005). Environmental applications for biosurfactants. *Environmental Pollution*, 133 (2): 183-198.

Mulligan, C. N. (2009). Recent advances in the environmental applications of biosurfactants. *Current Opinion in Colloid & Interface Science*, 14: 372-378.

Muro-Pastor, A. M.; Valladares, A.; Flores, E. and Herrero, A. (2002). Mutual dependence of the expression of the cell differentiation regulatory protein HetR and the global nitrogen regulator NtcA during heterocyst development. *Molecular Microbiology*, 44 (5): 1377-1385.

Murphy, C. A.; Cameron, J. A.; Huang, S. J. and Vinopal, R. T. (1999). Inactivation of polycaprolactone depolymerase (cutinase) in *Fusarium* cultures by an extracellular protease. *Journal of Industrial Microbiology & Biotechnology*, 22 (2): 71-77.

Murry, M. A. and Wolk, C. P. (1989). Evidence that the barrier to the penetration of oxygen into heterocysts depends upon 2 layers of the cell envelope. *Archives of Microbiology*, 151 (6): 469-474.

Muthusamy, K.; Gopalakrishnan, S.; Ravi, T. K. and Sivachidambaram, P. (2008). Biosurfactants: Properties, commercial production and application. *Current Science*, 94 (6): 736-747.

Nagai, S.; Okimura, K.; Kaizawa, N.; Ohki, K., *et al.* (1996). Study on surfactin, a cyclic depsipeptide .2. Synthesis of surfactin B-2 produced by *Bacillus natto* KMD 2311. *Chemical & Pharmaceutical Bulletin*, 44 (1): 5-10.

Nerurkar, A. S. (2010). Structural and Molecular Characteristics of Lichenysin and Its Relationship with Surface Activity. *Biosurfactants*. Sen, R. Berlin, Springer-Verlag Berlin. 672: 304-315.

Neuveglise, C.; Nicaud, J. M.; Ross-Macdonald, P. and Gaillardin, C. (1998). A shuttle mutagenesis system for tagging genes in the yeast *Yarrowia lipolytica*. *Gene*, 213 (1-2): 37-46.

Nguyen, T. T. and Sabatini, D. A. (2011). Characterization and Emulsification Properties of Rhamnolipid and Sophorolipid Biosurfactants and Their Applications. *International Journal of Molecular Sciences*, 12 (2): 1232-1244.

Nolan, T.; Hands, R. E.; Ogunkolade, W. and Bustin, S. A. (2006). SPUD: A quantitative PCR assay for the detection of inhibitors in nucleic acid preparations. *Analytical Biochemistry*, 351 (2): 308-310.

Noordman, W. H. and Janssen, D. B. (2002). Rhamnolipid stimulates uptake of hydrophobic compounds by *Pseudomonas aeruginosa*. *Applied and Environmental Microbiology*, 68 (9): 4502-4508.

Ochsner, U. A.; Fiechter, A. and Reiser, J. (1994a). Isolation, characterization, and expression in *Escherichia coli* of the *Pseudomonas aeruginosa* rhlAB genes encoding a rhamnosyltransferse involved in rhamnolipid biosurfactant synthesis. *The journal of biological chemistry*, 269: 19787-19795.

Ochsner, U. A.; Koch, A. K.; Fiechter, A. and Reiser, J. (1994b). Isolation and characterization of a regulatory gene affecting rhamnolipid biosurfactant synthesis in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 176: 2044-2054.

Ochsner, U. A. and Reiser, J. (1995). Autoinducer-mediated regulation of rhamnolipid biosurfactant synthesis in *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America*, 92: 6424-6428.

Ochsner, U. A.; Reiser, J.; Fiechter, A. and Witholt, B. (1995). Production of *Pseudomonas aeruginosa* rhamnolipid biosurfactants in heterelogous hosts. *Applied and Environmental Microbiology*, 61 (9): 3503-3506.

Ohtaki, S.; Maeda, H.; Takahashi, T.; Yamagata, Y., *et al.* (2006). Novel hydrophobic surface binding protein, HsbA, produced by *Aspergillus oryzae*. *Applied and Environmental Microbiology*, 72 (4): 2407-2413.

Olmedo-Verd, E.; Muro-Pastor, A. M.; Flores, E. and Herrero, A. (2006). Localized induction of the ntcA regulatory gene in developing heterocysts of *Anabaena sp* strain PCC 7120. *Journal of Bacteriology*, 188 (18): 6694-6699.

Olmedo-Verd, E.; Valladares, A.; Flores, E.; Herrero, A., *et al.* (2008). Role of Two NtcA-Binding Sites in the Complex ntcA Gene Promoter of the Heterocyst-Forming *Cyanobacterium Anabaena* sp Strain PCC 7120. *Journal of Bacteriology*, 190 (22): 7584-7590.

Pacwa-Plociniczak, M.; Plaza, G. A.; Piotrowska-Seget, Z. and Cameotra, S. S. (2011). Environmental applications of biosurfactants: recent advances. *International Journal of Molecular Sciences*, 12: 633-654.

Palme, O.; Comanescu, G.; Stoineva, I.; Radel, S., *et al.* (2010). Sophorolipids from *Candida bombicola*: Cell separation by ultrasonic particle manipulation. *European Journal of Lipid Science and Technology*, 112 (6): 663-673.

Panek, A. (1963). Function of trehalose in bakers yeast (*Saccharomyces cerevisiae*). *Archives of Biochemistry and Biophysics*, 100 (3): 422-&.

Pao, S. S.; Paulsen, I. T. and Saier, M. H. (1998). Major facilitator superfamily. *Microbiology and Molecular Biology Reviews*, 62 (1): 1-+.

Pearson, J. P.; Gray, K. M.; Passador, L.; Tucker, K. D., *et al.* (1994). Structure of the autoinducer required for expression of *Pseudomonas aeruginosa* virulence genes. *Proceedings of the National Academy of Sciences of the United States of America*, 91 (1): 197-201.

Pearson, J. P.; Passador, L.; Iglewski, B. H. and Greenberg, E. P. (1995). A second N-acylhomoserine lactone signal produced by *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America*, 92 (5): 1490-1494.

Pearson, J. P.; Pesci, E. C. and Iglewski, B. H. (1997). Roles of *Pseudomonas aeruginosa* las and rhl quorum-sensing systems in control of elastase and rhamnolipid biosynthesis genes. *Journal of Bacteriology*, 179 (18): 5756-5767.

Pekin, G.; Vardar-Sukan, F. and Kosaric, N. (2005). Production of sophorolipids from *Candida bombicola* ATCC 22214 using Turkish corn oil and honey. *Engineering in Life Sciences*, 5 (4): 357-362.

Perfumo, A.; Rancich, I. and Banat, I. M. (2010). Possibilities and challenges for biosurfactants use in petroleum industry. *Biosurfactants* Sen, R. New York, Springer Science+Business Media. 672: 135-145.

Perrin, R. M.; Fedorova, N. D.; Bok, J. W.; Cramer, R. A., *et al.* (2007). Transcriptional regulation of chemical diversity in *Aspergillus fumigatus* by LaeA. *Plos Pathogens*, 3 (4): 508-517.

Pesci, E. C. and Iglewski, B. H. (1997). The chain of command in *Pseudomonas* quorum sensing. *Trends in Microbiology*, 5 (4): 132-134.

Pesci, E. C.; Pearson, J. P.; Seed, P. C. and Iglewski, B. H. (1997). Regulation of las and rhl quorum sensing in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 179 (10): 3127-3132.

Peypoux, F.; Bonmatin, J. M.; Labbe, H.; Grangemard, I., *et al.* (1994). [ALA4] Surfactin, a novel isoform from *Bacillus subtilis* studies by MASS and NMR spectroscopies. *European Journal of Biochemistry*, 224 (1): 89-96.

Peypoux, F.; Bonmatin, J. M. and Wallach, J. (1999). Recent trends in the biochemistry of surfactin. *Applied Microbiology and Biotechnology*, 51 (5): 553-563.

Pfaffl, M. W.; Horgan, G. W. and Dempfle, L. (2002). Relative expression software tool (REST (c)) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Research*, 30 (9): 10.

Poirier, Y.; Erard, N. and Petetot, J. M. C. (2002). Synthesis of polyhydroxyalkanoate in the peroxisome of *Pichia pastoris*. *Fems Microbiology Letters*, 207 (1): 97-102.

Price, N. P. J.; Ray, K. J.; Vermillion, K. E.; Dunlap, C. A., *et al.* (2012). Structural characterization of novel sophorolipid biosurfactants from a newly identified species of *Candida* yeast. *Carbohydrate Research*, 348: 33-41.

Pryde, F. E.; Gorham, H. C. and Louis, E. J. (1997). Chromosome ends: all the same under their caps. *Current Opinion in Genetics & Development*, 7 (6): 822-828.

Puchkov, E. O.; Zähringer, U.; Lindner, B.; Kulakovskaya, T. V., *et al.* (2002). The mycocidal, membrane-active complex of *Cryptotoccus humicola* is a new type of cellobiose lipid with detergent features. *Biochimica Et Biophysica Acta*, 1558: 161-170.

Rahim, R.; Ochsner, U. A.; Olvera, C.; Graninger, M., *et al.* (2001). Cloning and functional characterization of the *Pseudomonas aeruginosa* rhlC gene that encodes rhamnosyltransferase 2, an enzyme responsible for di-rhamnolipid biosynthesis. *Molecular Microbiology*, 40 (3): 708-718.

Rau, U.; Hammen, S.; Heckmann, R.; Wray, V., *et al.* (2001). Sophorolipids: a source for novel compounds. *Industrial Crops and Products*, 13: 85-92.

Rau, U.; Heckmann, R.; Wray, V. and Lang, S. (1999). Enzymatic conversion of a sophorolipid into a glucose lipid. *Biotechnology Letters*, 21: 973-977.

Rau, U.; Nguyen, L. A.; Schulz, S.; Wray, V., *et al.* (2005). Formation and analysis of mannosylerythritol lipids secreted by *Pseudozyma aphidis*. *Applied Microbiology and Biotechnology*, 66 (5): 551-559.

Reddy, B. V. B.; Kallifidas, D.; Kim, J. H.; Charlop-Powers, Z., *et al.* (2012). Natural Product Biosynthetic Gene Diversity in Geographically Distinct Soil Microbiomes. *Applied and Environmental Microbiology*, 78 (10): 3744-3752.

Rehm, B. H. A.; Kruger, N. and Steinbuchel, A. (1998). A new metabolic link between fatty acid de novo synthesis and polyhydroxyalkanoic acid synthesis - The phaG gene from *Pseudomonas putida* KT2440 encodes a 3-hydroxyacyl-acyl carrier protein coenzyme A transferase. *Journal of Biological Chemistry*, 273 (37): 24044-24051.

Rehmeyer, C.; Li, W. X.; Kusaba, M.; Kim, Y. S., *et al.* (2006). Organization of chromosome ends in the rice blast fungus, *Magnaporthe oryzae*. *Nucleic Acids Research*, 34 (17): 4685-4701.

Reyes-Dominguez, Y.; Bok, J. W.; Berger, H.; Shwab, E. K., *et al.* (2010). Heterochromatic marks are associated with the repression of secondary metabolism clusters in *Aspergillus nidulans*. *Molecular Microbiology*, 76 (6): 1376-1386.

Risser, D. D. and Callahan, S. M. (2007). Mutagenesis of hetR reveals amino acids necessary for HetR function in the heterocystous *cyanobacterium Anabaena* sp strain PCC 7120. *Journal of Bacteriology*, 189 (6): 2460-2467.

Robyr, D.; Suka, Y.; Xenarios, I.; Kurdistani, S. K., *et al.* (2002). Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell*, 109 (4): 437-446.

Rodrigues, L. R. and Teixeira, J. A. (2010). Biomedical and therapeutic applications of biosurfactants. *Biosurfactants*. Sen, R. New York, Springer Science+Business Media. 672: 75-87.

Roelants, S. L. K. W.; Saerens, K.; Derycke, T.; Li, B., *et al.* (2013). *Candida bombicola* as a platform organism for the production of tailor-made biomolecules. *Biotechnology and Bioengineering*, accepted.

Rohan. (2012). Surfactants Market by product types, substrates, geopraphy & applications: global industry trends & forecasts to 2017. *MarketsandMarkets*. Dallas.

Rohlfs, M.; Albert, M.; Keller, N. P. and Kempken, F. (2007). Secondary chemicals protect mould from fungivory. *Biology Letters*, 3 (5): 523-525.

Ron, E. Z. and Rosenberg, E. (2001). Natural roles of biosurfactants. *Environmental Microbiology*, 3 (4): 229-236.

Rosa, C. A. and Lachance, M. A. (1998). The yeast genus *Starmerella* gen. nov. and *Starmerella bombicola* sp. nov., the teleomorph of *Candida bombicola* (Spencer, Gorin & Tullock) Meyer & Yarrow. *International Journal of Systematic Bacteriology*, 48: 1413-1417.

Ruiz, B.; Chavez, A.; Forero, A.; Garcia-Huante, Y., *et al.* (2010). Production of microbial secondary metabolites: Regulation by the carbon source. *Critical Reviews in Microbiology*, 36 (2): 146-167.

Rust, D. and Wildes, S. (2008). Surfactants: A Market opportunity Study Update.

Saerens, K. (2012). Synthesis of glycolipids by *Candida bombicola*. . *Faculty of Bioscience Engineering*. Ghent, Ghent University. PhD: 262.

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

Saerens, K. M. J.; Roelants, S. L. K. W.; Van Bogaert, I. N. A. and Soetaert, W. (2011a). Identification of the UDP-glucosyltransferase gene UGTA1, responsible for the first glucosylation step in the sophorolipid biosynthetic pathway of *Candida bombicola* ATCC 22214. *Fems Yeast Research*, 11 (1): 123-132.

Saerens, K. M. J.; Saey, L. and Soetaert, W. (2011b). One-Step Production of Unacetylated Sophorolipids by an Acetyltransferase Negative *Candida bombicola*. *Biotechnology and Bioengineering*, 108 (12): 2923-2931.

Saerens, K. M. J.; Zhang, J.; Van Bogaert, I. and Soetaert, W. (2011c). Cloning and functional characterization of the UDP-glucosyltransferase UgtB1 involved in sophorolipid production by *Candida bombicola* and creation of a glucolipid-producing yeast strain. *Yeast*, 28: 279-292.

Sambrook, J. and Russell, D. W., Eds. (2001). *Molecular cloning: a laboratory manual, 3rd edition.* New York, Cold Spring Harbor Laboratoy Press.

Scazzocchio, C. (2000). The fungal GATA factors. *Current Opinion in Microbiology*, 3 (2): 126-131.

Schneider, A.; Stachelhaus, T. and Marahiel, M. A. (1998). Targeted alteration of the substrate specificity of peptide synthetases by rational module swapping. *Molecular and General Genetics*, 257 (3): 308-318.

Scholz, C.; Mehta, S.; Nicolosi, R.; Bisht, K., *et al.* (1998). Bioactivity of extracellular glycolipids - Investigation of potential anti-cancer activity of sophorolipids and sophorolipid-derivatives. *Abstracts of Papers of the American Chemical Society*, 216: U4-U4.

Schramm, L. L.; Stasiuk, E. N. and Marangoni, D. G. (2003). Surfactants and their applications. *Annual Reports on the Progress of Chemistry Section C*, 99: 3-48.

Seydlova, G. and Svobodova, J. (2008). Review of surfactin chemical properties and the potential biomedical applications. *Central European Journal of Medicine*, 3 (2): 123-133.

Shaaban, M.; Palmer, J. M.; El-Naggar, W. A.; El-Sokkary, M. A., *et al.* (2010). Involvement of transposon-like elements in penicillin gene cluster regulation. *Fungal Genetics and Biology*, 47 (5): 423-432.

Shah, V.; Badia, D. and Ratsep, P. (2007). Sophorolipids having enhanced antibacterial activity. *Antimicrobial Agents and Chemotherapy*, 51: 397-400.

Shah, V.; Doncel, G. F.; Seyoum, T.; Eaton, K. M., *et al.* (2005). Sophorolipids, microbial glycolipids with anti-human immunodeficiency virus and sperm-immobilizing activities. *Antimicrobial Agents and Chemotherapy*, 49 (10): 4093-4100.

Shete, A. M.; Wadhawa, G.; Banat, I. M. and Chopade, B. A. (2006). Mapping of patents on bioemulsifier and biosurfactant: A review. *Journal of Scientific & Industrial Research*, 65 (2): 91-115.

Shwab, E. K. and Keller, N. P. (2008). Regulation of secondary metabolite production in filamentous *ascomycetes*. *Mycological Research*, 112: 225-230.

Soberon-Chavez, G.; Lepine, F. and Deziel, E. (2005). Production of rhamnolipids by *Pseudomonas aeruginosa*. *Applied Microbiology and Biotechnology*, 68 (6): 718-725.

Soetaert, W.; De Maeseneire, S. L.; Saerens, K.; Roelants, S., *et al.* (2010). Yeast strains modified in their sophorolipid production and uses thereof. UGent. WO2011154523 (A1).

Solaiman, D. K. Y.; Ashby, R. D.; Nunez, A. and Foglia, T. A. (2004). Production of sophorolipids by *Candida bombicola* grown on soy molasses as substrate. *Biotechnology Letters*, 26 (15): 1241-1245.

Solaiman, D. K. Y.; Ashby, R. D.; Zerkowski, J. A. and Foglia, T. A. (2007). Simplified soy molasses-based medium for reduced-cost production of sophorolipids by *Candida bombicola*. *Biotechnology Letters*, 29 (9): 1341-1347.

Solomon, J. M.; Magnuson, R.; Srivastava, A. and Grossman, A. D. (1995). Convergent sensing pathways mediate response to two extracellular competence factors in *Bacillus subtilis*. *Genes & Development*, 9 (5): 547-558.

Spellig, T.; Bottin, A. and Kahmann, R. (1996). Green fluorescent protein (GFP) as a new vital marker in the phytopathogenic fungus *Ustilago maydis*. *Molecular & General Genetics*, 252 (5): 503-509.

Spoeckner, S.; Wray, V.; Nimtz, M. and Lang, S. (1999). Glycolipids of the smut fungus *Ustilago maydis* from cultivation on renewable resources. *Applied Microbiology and Biotechnology*, 51: 33-39.

Stachelhaus, T.; Schneider, A. and Marahiel, M. A. (1995). Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science*, 269 (5220): 69-72.

Staib, P.; Michel, S.; Kohler, G. and Morschhauser, J. (2000). A molecular genetic system for the pathogenic yeast Candida dubliniensis. *Gene*, 242 (1-2): 393-398.

Stanghellini, M. E., Miller, R.M., Rasmussen, S.L., Kim, D.H., Zhang, Y. (1998). Microbially produced rhamnolipids (biosurfactants) for the control of plant pathogenic zoosporic fungi. US, Arizona Board of Regents, on behalf of the University of Arizona, Tucson AZ.

Staron, P.; Forchhammer, K. and Maldener, I. (2011). Novel ATP-driven Pathway of Glycolipid Export Involving TolC Protein. *Journal of Biological Chemistry*, 286 (44): 38202-38210.

Stein, T. (2005). *Bacillus subtilis* antibiotics: structures, syntheses and specific functions. *Molecular Microbiology*, 56 (4): 845-857.

Stein, T.; Heinzmann, S.; Kiesau, P.; Himmel, B., *et al.* (2003). The spa-box for transcriptional activation of subtilin biosynthesis and immunity in *Bacillus subtilis*. *Molecular Microbiology*, 47 (6): 1627-1636.

Steinbuchel, A. and Wiese, S. (1992). A *Pseudomonas* strain accumulating polyesters of 3-hydroxybutyric acid and medium-chain-length 3-hydroxyalkanoic acids. *Applied Microbiology and Biotechnology*, 37 (6): 691-697.

Steller, S.; Sokoll, A.; Wilde, C.; Bernhard, F., *et al.* (2004). Initiation of surfactin biosynthesis and the role of the SrfD-thioesterase protein. *Biochemistry*, 43 (35): 11331-11343.

Stüwer, O.; Hommel, R.; Haferburg, D. and Kleber, H. P. (1987). Production of crystalline surface-active glycolipids by a strain of *Torulopsis apicola*. *Journal of Biotechnology*, 6 (4): 259-269.

Sudha, S.; Kumanan, R. and Muthusamy, K. (2010a). Optimisation of medium composition for the production of sophorolipids from *Candida tropicalis*. *Journal of Chemical and Pharmaceutical Sciences*, 3: 129-130.

Sudha, S.; Kumanan, R.; Vijayashree, P. and Muthusamy, K. (2010b). Comparative study for the production, characterization and antimicrobial studies of sophorolipids using *Candida tropicalis*. *Asian Journal of Pharmaceutical and Clinical Research*, 3: 69-72.

Sullivan, E. R. (1998). Molecular genetics of biosurfactant production. *Current Opinion in Biotechnology*, 9: 263-269.

Syldatk, C.; Lang, S.; Wagner, F.; Wray, V., *et al.* (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 (1-2): 51-60.

Tag, A.; Hicks, J.; Garifullina, G.; Ake, C., *et al.* (2000). G-protein signalling mediates differential production of toxic secondary metabolites. *Molecular Microbiology*, 38 (3): 658-665.

Takahashi, M.; Morita, T.; Wada, K.; Hirose, N., *et al.* (2011). Production of Sophorolipid Glycolipid Biosurfactants from Sugarcane Molasses Using *Starmerella bombicola* NBRC 10243. *Journal of Oleo Science*, 60 (5): 267-273.

Tanigawa, R.; Shirokane, M.; Maeda, S.; Omata, T., *et al.* (2002). Transcriptional activation of NtcA-dependent promoters of *Synechococcus* sp PCC 7942 by 2-oxoglutarate in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 99 (7): 4251-4255.

Teichmann, B.; Labbe, C.; Lefebvre, F.; Bolker, M., *et al.* (2011a). Identification of a biosynthesis gene cluster for flocculosin a cellobiose lipid produced by the biocontrol agent *Pseudozyma flocculosa*. *Molecular Microbiology*, 79 (6): 1483-1495.

Teichmann, B.; Lefebvre, F.; Labbe, C.; Bolker, M., *et al.* (2011b). Beta Hydroxylation of Glycolipids from *Ustilago maydis* and *Pseudozyma flocculosa* by an NADPH-Dependent beta-Hydroxylase. *Applied and Environmental Microbiology*, 77 (21): 7823-7829.

Teichmann, B.; Linne, U.; Hewald, S.; Marahiel, M. A., *et al.* (2007). A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from *Ustilago maydis*. *Molecular Microbiology*, 66 (2): 525-533.

Teichmann, B.; Liu, L. D.; Schink, K. O. and Bolker, M. (2010). Activation of the ustilagic acid biosynthesis gene cluster in *Ustilago maydis* by the C2H2 Zinc Finger Transcription Factor Rua1. *Applied and Environmental Microbiology*, 76 (8): 2633-2640.

Terentiev, Y.; Breuer, U.; Babel, W. and Kunze, G. (2004). Non-conventional yeasts as producers of polyhydroxyalkanoates - genetic engineering of *Arxula adeninivorans*. *Applied Microbiology and Biotechnology*, 64 (3): 376-381.

Teste, M. A.; Duquenne, M.; Francois, J. M. and Parrou, J. L. (2009). Validation of reference genes for quantitative expression analysis by real-time RT-PCR in *Saccharomyces cerevisiae*. *Bmc Molecular Biology*, 10: 15.

Tilburn, J.; Sarkar, S.; Widdick, D. A.; Espeso, E. A., *et al.* (1995). The *Aspergillus* Pacc Zinc finger transcription factor mediates regulation of both acid-expressed and alkaline-expressed genes by ambient pH. *Embo Journal*, 14 (4): 779-790.

Tran, H. G. (2012). Development of a general glycosylation technology by molecular engineering of carbohydrate active enzyme. *Department of Biochemical and Microbial Technology*. Ghent, Ghent University. PhD.

Tsuge, K.; Ohata, Y. and Shoda, M. (2001). Gene yerP, involved in surfactin self-resistance in *Bacillus subtilis*. *Antimicrobial Agents and Chemotherapy*, 45 (12): 3566-3573.

Tulloch, A. P.; Hill, A. and Spencer, J. F. T. (1967). A new type of macrocycli lactone from *Torulopsis apicola*. *Chemical Communications* (12): 584-&.

Tulloch, A. P. and Spencer, J. F. T. (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. *Canadian Journal of Chemistry*, 46: 1523-1528.

Tulloch, A. P.; Spencer, J. F. T. and Deinema, M. H. (1968). A new hydroxy fatty acid sophoroside from *Candida bogoriensis*. *Canadian Journal of Chemistry*, 46: 345-348.

Tulloch, A. P.; Spencer, J. F. T. and Gorin, P. A. J. (1962). The fermentation of long-chain compounds by *Torulopsis magnoliae*. *Canadian Journal of Chemistry*, 40: 1326-1338.

Van Bogaert, I.; De Maeseneire, S. L.; Develter, D.; Soetaert, W., *et al.* (2008a). Cloning and characterization of the glyceraldehyde 3-phosphate dehydrogenase gene of *Candida bombicola* and use of its promotor. *Journal of Industrial Microbiology and Biotechnology*, 35: 1085-1092.

Van Bogaert, I.; Develter, D. and Fleurackers, S. (2011a). Method for the production of medium-chain sophorolipids. Ecover. US2011136110.

Van Bogaert, I.; Fleurackers, S.; Van Kerrebroeck, S.; Develter, D., *et al.* (2011b). Production of new-to-nature sophorolipids by cultivating the yeast *Candida bombicola* on unconventional hydrophobic substrates. *Biotechnology and Bioengineering*, 108 (4): 734-741.

Van Bogaert, I.; Roelants, S.; Develter, D. and Soetaert, W. (2010). Sophorolipid production by *Candida bombicola* on oils with a special fatty acid composition and their consequences on cell viability. *Biotechnology Letters*, 32: 1509-1514.

Van Bogaert, I. and Soetaert, W. (2009). Sophorolipid transporter protein. UGent. US2012311741 (A1).

Van Bogaert, I. N. A. (2008). Microbial synthesis of sophorolipids by the yeast *Candida bombicola*. *Department of Biochemical and Microbial Technology* Ghent, Belgium, Ghent University: 239.

Van Bogaert, I. N. A.; De Maeseneire, S. L.; De Schamphelaire, W.; Develter, D., *et al.* (2007a). Cloning, characterization and functionality of the orotidine-5 '-phosphate

decarboxylase gene (URA3) of the glycolipid-producing yeast *Candida bombicola*. *Yeast*, 24 (3): 201-208.

Van Bogaert, I. N. A.; De Maeseneire, S. L.; Develter, D.; Soetaert, W., *et al.* (2008b). Development of a transformation and selection system for the glycolipid producing yeast *Candida bombicola*. *Yeast*, 25: 272-278.

Van Bogaert, I. N. A.; Demey, M.; Develter, D.; Soetaert, W., *et al.* (2009a). Importance of the cytochrome P450 monooxygenase CYP52 family for the sophorolipid-producing yeast *Candida bombicola*. *Fems Yeast Research*, 9 (1): 87-94.

Van Bogaert, I. N. A.; Holvoet, K.; Roelants, S. L. K. W.; Li, B., *et al.* (2013). The biosynthetic gene cluster for sophorolipids: a biotechnological interesting biosurfactant produced by *Starmerella bombicola*. *Molecular Microbiology*.

Van Bogaert, I. N. A.; Sabirova, J.; Develter, D.; Soetaert, W., *et al.* (2009b). Knocking out the MFE-2 gene of *Candida bombicola* leads to improved medium-chain sophorolipid production. *Fems Yeast Research*, 9 (4): 610-617.

Van Bogaert, I. N. A.; Saerens, K.; De Muynck, C.; Develter, D., *et al.* (2007b). Microbial production and application of sophorolipids. *Applied Microbiology and Biotechnology*, 76 (1): 23-34.

Van Bogaert, I. N. A.; Zhang, J. X. and Soetaert, W. (2011c). Microbial synthesis of sophorolipids. *Process Biochemistry*, 46 (4): 821-833.

Vandesompele, J.; De Preter, K.; Pattyn, F.; Poppe, B., *et al.* (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, 3 (7): 12.

Vazquez-Bermudez, M. F.; Flores, E. and Herrero, A. (2002). Analysis of binding sites for the nitrogen-control transcription factor NtcA in the promoters of *Synechococcus* nitrogen-regulated genes. *Biochimica Et Biophysica Acta-Gene Structure and Expression*, 1578 (1-3): 95-98.

Verlinden, R. A. J.; Hill, D. J.; Kenward, M. A.; Williams, C. D., *et al.* (2007). Bacterial synthesis of biodegradable polyhydroxyalkanoates. *Journal of Applied Microbiology*, 102 (6): 1437-1449.

Vertesy, L.; Kurz, M.; Noelken, G. and Wink, J. (1998). Ustilipides, method for the production and the use thereof US6472158 (B1)

Waegeman, H.; Beauprez, J.; Moens, H.; Maertens, J., *et al.* (2011). Effect of iclR and arcA knockouts on biomass formation and metabolic fluxes in Escherichia coli K12 and its implications on understanding the metabolism of *Escherichia coli* BL21 (DE3). *Bmc Microbiology*, 11.

Walsby, A. E. (1985). The permeability of heterocysts to the gases nitrogen and oxygen. *Proceedings of the Royal Society B-Biological Sciences*, 226 (1244): 366-+.

Walton, J. D. (2000). Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: An hypothesis. *Fungal Genetics and Biology*, 30 (3): 167-171.

Wentzel, A.; Ellingsen, T. E.; Kotlar, H. K.; Zotchev, S. B., *et al.* (2007). Bacterial metabolism of long-chain n-alkanes. *Applied Microbiology and Biotechnology*, 76 (6): 1209-1221.

Whang, L.-M.; Liu, P.-W. G.; Ma, C.-C. and Cheng, S.-S. (2009). Application of rhamnolipid and surfactin for enhanced diesel biodegradation-Effects of pH and ammonium addition. *Journal of Hazardous Materials*, 164 (2-3): 1045-1050.

Wilhelm, S.; Gdynia, A.; Tielen, P.; Rosenau, F., *et al.* (2007). The autotransporter esterase EstA of *Pseudomonas aeruginosa* is required for rhamnolipid production, cell motility, and biofilm formation. *Journal of Bacteriology*, 189 (18): 6695-6703.

Wilson, W. A.; Roach, P. J.; Montero, M.; Baroja-Fernandez, E., *et al.* (2010). Regulation of glycogen metabolism in yeast and bacteria. *Fems Microbiology Reviews*, 34 (6): 952-985.

Winson, M. K.; Camara, M.; Latifi, A.; Foglino, M., *et al.* (1995). Multiple N-acyl-homoserine lactone signal molecules regulate production of virulence determinants and secondary metabolites in *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America*, 92 (20): 9427-9431.

Wolk, C. P.; Ernst, A. and Elhai, J. (1994). *The Molecular Biology of Cyanobacteria*. Dudrecht, The Netherlands, Kluwer Academic.

Wong, K. H.; Hynes, M. J. and Davis, M. A. (2008). Recent advances in nitrogen regulation: a comparison between *Saccharomyces cerevisiae* and filamentous fungi. *Eukaryotic Cell*, 7: 917-925.

Xu, Q.; Nakajima, M.; Liu, Z. and Shiina, T. (2011). Biosurfactants for microbubble preparation and application. *International Journal of Molecular Sciences*, 12: 462-475.

Yakimov, M. M.; Kroger, A.; Slepak, T.; Giuliano, L., *et al.* (1998). A putative lichenysin A synthetase operon in *Bacillus licheniformis*: initial characterization. *Biochimica Et Biophysica Acta-Gene Structure and Expression*, 1399 (2-3): 141-153.

Yin, W. B. and Keller, N. P. (2011). Transcriptional Regulatory Elements in Fungal Secondary Metabolism. *Journal of Microbiology*, 49 (3): 329-339.

Youssef, N. H.; Duncan, K. E. and McInerney, M. J. (2005). Importance of 3-hydroxy fatty acid composition of lipopeptides for biosurfactant activity. *Applied and Environmental Microbiology*, 71 (12): 7690-7695.

Youssef, N. H.; Wofford, N. and McInerney, M. J. (2011). Importance of the Long-Chain Fatty Acid Beta-Hydroxylating Cytochrome P450 Enzyme YbdT for Lipopeptide Biosynthesis in *Bacillus subtilis* Strain OKB105. *International Journal of Molecular Sciences*, 12 (3): 1767-1786.

Zhang, J.; Saerens, K. M. J.; Van Bogaert, I. N. A. and Soetaert, W. (2011). Vegetable oil enhances sophorolipid production by *Rhodotorula bogoriensis*. *Biotechnology Letters*, 33: 2417-2423.

Zhang, Y. M. and Miller, R. M. (1994). Effect of a *Pseudomonas* rhamnolipid biosurfactant on cell hydrofobicity and biodegradation of octadecane. *Applied and Environmental Microbiology*, 60 (6): 2101-2106.

Zhao, M. X.; Jiang, Y. L.; He, Y. X.; Chen, Y. F., *et al.* Structural basis for the allosteric control of the global transcription factor NtcA by the nitrogen starvation signal 2-oxoglutarate. *Proceedings of the National Academy of Sciences of the United States of America*, 107 (28): 12487-12492.

Zhao, Y. H.; Li, H. M.; Qin, L. F.; Wang, H. H., *et al.* (2007). Disruption of the polyhydroxyalkanoate synthase gene in *Aeromonas hydrophila* reduces its survival ability under stress conditions. *Fems Microbiology Letters*, 276 (1): 34-41.

Zhou, Q. H.; Klekner, V. and Kosaric, N. (1992). Production of sophorose lipids by *Torulopsis bombicola* from safflower oil and glucose. *Journal of the American Oil Chemists Society*, 69 (1): 89-91.

Zhou, Q. H. and Kosaric, N. (1993). Effect of lactose and olive oil on intracellular and extracellular lipids of *Torulopsis bombicola*. *Biotechnology Letters*, 15 (5): 477-482.

Zhou, Q. H. and Kosaric, N. (1995). Utilization of canola oil and lactose to produce biosurfactant with *Candida bombicola*. *Journal of the American Oil Chemists Society*, 72 (1): 67-71.

Zhou, S. Q.; Xu, C.; Wang, J.; Gao, W., *et al.* (2004). Supramolecular assemblies of a naturally derived sophorolipid. *Langmuir*, 20 (19): 7926-7932.

Zhu, K. and Rock, C. O. (2008). RhlA converts beta-hydroxyacyl-acyl carrier protein intermediates in fatty acid synthesis to the beta-hydroxydecanoyl-beta-hydroxydecanoate component of rhamnolipids in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 190: 3147-3154.

SUMMARY

SUMMARY

Biosurfactants are a family of very diverse secondary metabolites produced by a wide variety of microorganisms. They have been described to find (potential) applications in the medical world, personal care sector, mining processes, food industry, cosmetics, crop protection, pharmaceuticals, bio-remediation, household detergents, paper and pulp industry, textiles, paint industries, etc. (Banat *et al.*, 2000). Their biodegradability and the fact that they can be produced from renewable resources gives them an advantage over their chemical counterparts and may therefore make them suitable to partly replace chemicals. Especially glycolipid biosurfactants, like sophorolipids (SLs), have been described to provide significant opportunities to replace chemical surfactants in several sectors (Marchant and Banat, 2012). Growing environmental awareness and rising oil prices, combined with the fact that our dependency on this finite source of energy and chemical compounds could lead to a catastrophy once we run out of oil, has resulted in renewed interest in these molecules in the last ten years. The major two factors currently limiting the penetration of biosurfactants into the market are firstly the limited structural variety and secondly the high production price due to rather low productivities. A solution for both drawbacks can be identified in the genetic engineering of natural producers. Hence, one of the critical factors for the application of biosurfactants in industry is a detailed knowledge of their genetics, as this may hold the key to their future (enhanced) economical production using enhanced microbial strains.

This work focuses on the efficient natural glycolipid biosurfactant producer *Starmerella bombicola*, and in particular on the biological opportunities this yeast offers. Due to its very efficient sophorolipid machinery it has already found several applications in the industry and is the subject of various patent applications. This yeast thus represents an ideal ‘template strain’ to be transformed into a platform organism for the production of tailor-made biomolecules and this challenge was set as the research goal for the project to which this work contributed.

The first part of this dissertation thus focusses on the further unraveling of the sophorolipid metabolism of *S. bombicola* as this is of paramount importance when one wants to make use of its efficient SL machinery for the expression of heterologous biosynthetic pathways. The borders of the SL biosynthetic gene cluster were defined and possible regulatory effects affecting SL biosynthesis were discussed. Unfortunately a specific regulator of SL

biosynthesis was not found within the gene cluster, which would represent a very versatile tool for the further genetic engineering of *S. bombicola*. The identification of such a regulator should be persued in the future, which will be facilitated by the fully annotated genome that is now available in combination with the results of RNA sequencing experiments. In the first part of this manuscript the existence of a SL catabolic pathway in *S. bombicola* besides the biosynthetic pathway was also described. One of the natural roles of SL production by *S. bombicola* can thus be described as the creation of a carbon sink when an excess of carbon is combined with a deficiency of other yet to be identified nutrients. The corresponding catabolic enzymes might represent an opportunity to selectively produce certain SL derivatives.

In the second part of this work the further elaboration of the molecular toolbox for *S. bombicola* was described, which will facilitate the extensive genetic engineering of this yeast. A first tool consists of the development of a *gfp* reporter system, which will serve for the creation of a promotorbank and will furthermore enable an elegant combinatorial assessment of different parameters on gene expression. Moreover, this reporter system can be used to investigate the regulatory network affecting gene expression of the SL biosynthetic gene cluster. The latter can also be investigated using a second developped tool, consisting of a qPCR platform. A set of 20 candidate reference genes was screened, which resulted in the selection of a subset of reference genes that can be used to normalize and compare qPCR data derived from an array of conditions. This qPCR platform is also expected to be a usefull tool for the evaluation of future (SL) production strains. A third and last tool consists of a method for reusing the *ura3* marker, which will be indispensable for the further extensive genetic engineering of *S. bombicola,* requiring the combination of several modifications in one strain, until more markers become available.

In the third part of this manuscript the possible relevance of this yeast as an industrial production platform was demonstrated by providing two examples. The first example consists of the production of the bioplastic PHA by *S. bombicola* by replacing the first gene of the SL biosynthetic pathway (*cyp52M1*) with a codon optimized PHA synthase (*phaC1co*), thus utilising the SL regulatory network. This resulted in the production of a reasonable amount of PHA (2.0 % wt/dwt) in an SL deficient strain, and thus proves that this is a feasible strategy for the engineering of *S. bombicola*. Furthermore, rhamnolipids (RL) did not give rise to any toxic effects on the cells and SL biosynthesis still occurred in the presence of RLs in the

culture medium. Combined, these results are considered as proof that one can proceed with the engineering of the SL producer *S. bombicola* in an RL producing one. A second example consists of a strain producing 100 % lactonic SLs, even under conditions favoring the formation of acidic SLs. This strain is considered to represent a realistic industrial opportunity, as purification of the naturally produced SL mixture is currently needed to obtain the lactonic or acidic SLs in pure form, which is needed for certain applications. Bioreactor experiments with this new strain furthermore gave rise to a production process free of foam formation and higher SL yields were obtained as compared to the wild type. Taken together it is obvious that the last strain will probably have an industrial finality. The industrial relevance contained within these last two examples can be derived from the fact that they were part of two separate patent applications (Soetaert *et al.*, 2010; Ciesielska *et al.*, 2011).

In conclusion in this work the transformation of *S. bombicola* into an industrial production platform organism for interesting biomolecules was aimed for. The feasibility thereof was proven by two examples. Several molecular tools were developed to enable extensive genetic engineering of this yeast and furthermore enable the eludidation of the regulatory network influencing SL biosynthesis. Extended research and engineering are obviously needed to reach the set goal, but some significant steps in the right direction have been taken with the results presented in this work.

SAMENVATTING

Biosurfactants zijn een familie van zeer uiteenlopende secundaire metabolieten geproduceerd door een grote verscheidenheid van micro-organismen. Deze moleculen hebben (potentiële) toepassingen in de medische wereld, persoonlijke hygiëne, mijnbouw processen, voedingsindustrie, cosmetica, gewasbescherming, farmaceutica, bio-sanering, schoonmaakmiddelen, papier- en pulp-, textiel-, verf-industrie, etc. (Banat *et al.*, 2000). Hun biologische afbreekbaarheid en het feit dat ze kunnen worden geproduceerd uit hernieuwbare bronnen geeft hen een voordeel ten opzichte van de chemische tegenhangers waardoor ze geschikt zijn om chemische stoffen deels te gaan vervangen in de toekomst. Vooral glycolipide biosurfactants, zoals sophorolipiden (SL), werden beschreven als mogelijke vervanger van chemische surfactants in diverse sectoren (Marchant en Banat, 2012). Het groeiende publieke ecologische bewustzijn, de stijgende olieprijzen en het feit dat afhankelijkheid van een eindige bron van energie en chemische verbindingen (petroleum) tot een catastrofe kan leiden eens deze uitgeput geraakt, heeft geleid tot hernieuwde belangstelling voor deze moleculen. Momenteel zijn er twee factoren die de penetratie van biosurfactants in de markt verhinderen; de eerste is de beperkte structurele variatie en de tweede is de hoge productie prijs, veroorzaakt door een te lage productiviteit. Een oplossing voor beide nadelen kan geïdentificeerd worden in de genetische manipulatie van natuurlijke producenten. Eén van de belangrijkste factoren voor de toepassing van biosurfactants in de industrie is bijgevolg een gedetailleerde kennis van hun genetica.

Dit werk handelt over de efficiënte surfactant producent *Starmerella bombicola*, en meer specifiek over de biotechnologische mogelijkheden die deze gist biedt. De bijzonder efficiënte sophorolipiden productie van deze gist heeft reeds geleid tot verscheidene industriële toepassingen en octrooiaanvragen. Daarom is *S. bombicola* een ideale 'sjabloon stam' die kan worden omgevormd tot een platform organisme voor de productie van nieuwe biomoleculen. Deze uitdaging was één van de doelen van het overkoepelend project waartoe dit werk heeft bijgedragen.

Het eerste deel van dit proefschrift richt zich op de verdere ontrafeling van het sophorolipide metabolisme van *S. bombicola* omdat dit enorm belangrijk is als men gebruik wil maken van de efficiënte SL machinerie voor de expressie van heterologe genen. De grenzen van de biosynthetische SL gencluster werden gedefinieerd en mogelijke regulatiemechanismen van de biosynthese werden besproken. Een regulator van SL biosynthese werd spijtig genoeg niet gevonden in de gencluster. Zulk een specifieke regulator zou een veelzijdig instrument zijn voor de verdere genetische manipulatie van *S. bombicola*. De identificatie van dergelijke

regulator zal worden nagestreefd in de toekomst. Deze zoektocht zal vergemakkelijkt worden door het volledig geannoteerde genoom dat nu beschikbaar is in combinatie met de resultaten van RNA sequencing experimenten. In het eerste deel van dit manuscript wordt ook het optreden van SL degradatie door *S. bombicola* beschreven. Een van de natuurlijke functies van SL productie door *S. bombicola* kan derhalve beschreven worden als de vorming van een koolstofreserve wanneer een teveel aan koolstof gecombineerd wordt met een tekort aan andere nog niet geïdentificeerd voedingsstoffen.

In het tweede deel van dit werk wordt de verdere uitwerking van de moleculaire 'toolbox' voor *S. bombicola* beschreven, welke de uitgebreide genetische manipulatie van deze gist zal vergemakkelijken. Een eerste instrument omvat de ontwikkeling van een GFP reporter systeem, dat dient voor het creëren van een promotorbank. Bovendien maakt dit systeem een elegante combinatorische beoordeling van verschillende parameters op genexpressie mogelijk. Bovendien kan dit reporter systeem gebruikt worden om de regulatie van de SL gencluster te onderzoeken. Dit laatste kan ook worden bewerkstelligd door gebruik te maken van het ontwikkelde qPCR platform. Een set van 20 kandidaat referentiegenen werd onderzocht, waarna een subset van referentiegenen gedefinieerd werd die kan gebruikt worden voor het normaliseren van qPCR gegevens van een reeks van omstandigheden. Dit qPCR platform zal ook een nuttig instrument zijn voor de beoordeling van toekomstige (SL) productiestammen. Een derde en laatste hulpmiddel bestaat uit de ontwikkeling van een methode voor het hergebruik van de *ura3* merker. Deze laatste tool zal onmisbaar zijn voor de verdere uitgebreide genetische manipulatie van *S. bombicola* tot er meer markers beschikbaar zijn.

In het derde deel van dit manuscript werd de mogelijke relevantie van deze gist als een industrieel productie platform aangetoond door middel van twee voorbeelden. Het eerste voorbeeld bestaat uit het produceren van de bioplastic PHA door *S. bombicola* door vervanging van het eerste gen van de SL biosynthetische pathway (*cyp52M1*) met een codon geoptimaliseerd PHA synthase (*phaC1co*). Dit resulteerde in de productie van een redelijke hoeveelheid PHA (2.0% wt / dwt) in een SL deficiënte stam. Bovendien hebben rhamnolipiden (RL) geen toxisch effect op de cellen en gaat SL biosynthese gewoon door in de aanwezigheid van RL in het kweekmedium. Gecombineerd worden deze resultaten beschouwd als bewijs dat men de SL producent *S. bombicola* kan omvormen in een RL producent. Een tweede voorbeeld omvat een stam die 100% lacton SL produceert, zelfs onder omstandigheden die de vorming van open SLs bevoordeelt. Deze stam heeft industriële relevantie daar fractionatie van het natuurlijk geproduceerde SL mengsel momenteel nodig is

om lacton of zure SL in zuivere vorm, te verkrijgen. Bioreactor experimenten met deze nieuwe stam resulteerden in een productieproces zonder schuimvorming en bovendien werden hogere SL opbrengsten verkregen in vergelijking met het wild type. Samengevat is het duidelijk dat de laatste stam waarschijnlijk een industriële finaliteit heeft. Het industriële belang van deze laatste twee voorbeelden kan worden afgeleid uit het feit dat ze een deel uitmaken van twee afzonderlijke octrooiaanvragen (Soetaert et al., 2010, Ciesielska et al., 2011).

Om te concluderen werd in dit werk de transformatie van *S. bombicola* naar een industrieel productie platform nagestreefd. De haalbaarheid hiervan werd bewezen aan de hand van twee voorbeelden. Verschillende moleculaire ‘tools’ werden ontwikkeld om uitgebreide genetische manipulatie van deze gist te bewerkstelligen. Deze tools maken het bovendien mogelijk de regulatie van SL biosynthese te onderzoeken. Uitgebreide research en verdere genetische manipulatie zijn uiteraard nodig zijn om het gestelde doel te bereiken, maar een aantal belangrijke stappen in de goede richting zijn reeds gezet met de resultaten in dit werk.

CURRICULUM VITAE

UNIVERSITEIT GENT

FACULTEIT **BIO-INGENIEURSWETENSCHAPPEN**

CURRICULUM VITAE

Personalia

Name	Sophie L.K.W. Roelants
Date of birth	1st of June 1983
Place of birth	Herentals
Nationality	Belgian
Professional address	Centre of Expertise-Industrial Biotechnology and Biocatalysis Faculty of Bioscience Engineering Ghent University Coupure Links 653 B-9000 Gent
Email	Sophie.roelants@ugent.be
Phone	+32 474 21 22 91

Education

2009-2013	Ghent University, Centre of Expertise-Industrial Biotechnology and Biocatalysis (InBio.be), Faculty of Bioscience Engineering, Ghent, Belgium: **PhD research** • PhD thesis: *Starmerella bombicola* as a platform organism for the production of biobased compounds • Promotor: Prof. dr. ir. Wim Soetaert • co-promotor: dr. ir. Inge Van Bogaert
2004-2007	Ghent University, Faculty of Bioscience Engineering, Ghent, Belgium: **Master of Bioscience Engineering option Cell and Gene Biotechnology** Erasmus exchange program: Universidad de Sevilla, Spain, 6 months • Master thesis: Model-based optimalisation of succinate production by *Escherichia coli* • Promotor: Prof. dr. ir. Wim Soetaert • Tutor: dr. ir. Joeri Beauprez
2001-2004	University of Antwerp, Faculty of Sciences, Antwerp, Belgium
1995-2001	Instituut Dames van het Christelijk Onderwijs, Antwerp, Maths-Science

Courses

Introductory course on Technology Transfer Skills, April-May 2013, Ghent, Belgium.

Courses followed in the framework of the Doctoral School program (Bioscience Engineering):

Biotechnology: Biosafety, GMP and intellectual Property (2011)

Advanced Academic English: Writing Skills (2010) and Conference Skills (2011)

Leadership and Efficiency: Personal Effectiveness (2010) and Communication Skills (2010)

Biotechnology and Society (2010)

Entrepreneurship and Innovation (2009)

Publications

Published:

Roelants, S., Saerens, K., Derycke, T., Li, B., Lin, Y.-C., Van de Peer, Y., De Maeseneire, S. L., Van Bogaert, I., Soetaert, W., 2013. *Candida bombicola* as a platform organism for the production of tailor-made biomolecules. Biotechnology and Bioengineering. (accepted and in production)

Van Bogaert, I. N. A., Holvoet, K., **Roelants,** S. L. K. W., Li, B., Lin, Y., Van de Peer, Y., Soetaert, W., 2013. The biosynthetic gene cluster for sophorolipids: a biotechnological interesting biosurfactant produced by *Starmerella bombicola*. Molecular Microbiology (accepted and in production)

Van Bogaert, I., **Roelants, S.,** Develter, D., Soetaert, W., 2010. Sophorolipid production by *Candida bombicola* on oils with a special fatty acid composition and their consequences on cell viability. Biotechnol. Lett. 32**,** 1509-1514.

Saerens, K. M. J., **Roelants, S.** L. K. W**.**, Van Bogaert, I. N. A., Soetaert, W., 2011. Identification of the UDP-glucosyltransferase gene UGTA1, responsible for the first glucosylation step in the sophorolipid biosynthetic pathway of *Candida bombicola* ATCC 22214. FEMS Yeast Res. 11**,** 123-132.

In preparation:

<u>Mini-reviews (proposal accepted by the journal of 'Applied Microbiology and Biotechnology').</u>

Roelants, S., De Maeseneire L. S., Ciesielska, K., Saerens, K., Van Bogaert, I. and Soetaert, W. Biosurfactant gene clusters in eukaryotes: regulation and biotechnological potential.

Roelants, S., De Maeseneire L. S., Gheys R., Van Bogaert, I. and Soetaert, W. Biosurfactant gene clusters in prokaryotes: regulation and biotechnological potential.

<u>Research papers:</u>

Roelants, S. and Groeneboer, S., Ciesielska, K., Verweire, S., De Maeseneire, S., Van Bogaert, I., Devreese, B. and Soetaert, W. Catabolism of the biosurfactant sophorolipids by the yeast *Starmerella bombicola*.

Ciesielska, K. and **Roelants, S**., Verweire, S., Groeneboer, S., Verweire, S., De Maeseneire, S., Van Bogaert, I., Devreese, B. and Soetaert, W. Combined techniques for the determination of protagonists involved in sophorolipid catabolism by its producer *Starmerella bombicola*. (some experimental work still in progress).

Roelants, S., and De Maeseneire, L. S., Saey L., De Preester, K., Van Bogaert, I. and Soetaert, W. Development of a molecular toolbox for the biosurfactant producing yeast *Starmerella bombicola*: towards the creation of an industrial platform strain.

Roelants, S., Ciesielska, K., L, Moens, H., Kuen Chow, K., Verweire, S., De Maeseneire, S. L., Van Bogaert, I., Devreese, B. and Soetaert W. Towards the industrialisation of new biosurfactant producers: microbial production of 100 % lactonic sophorolipids.

Patent Applications

Soetaert, W.; De Maeseneire, S. L.; Saerens, K.; **Roelants, S.,** *et al.* (2010). Yeast strains modified in their sophorolipid production and uses thereof. UGent. WO2011154523 (A1).

Ciesielska, K.; **Roelants, S.** L. K. W.; Van Bogaert, I. N. A.; B., D., *et al.* (2011). A lactonase derived from *Candida bombicola* and uses thereof. UGent. PCT/EP2012/075571. (public july 2013).

Attended Conferences and Workshops

RRB-9: Renewable Resources and Biorefineries IV, 5-7th of June 2013, Antwerp, Belgium, **Oral presentation**, "Sophorolipids produced by the yeast *Starmerella bombicola*: biotechnological opportunities*."***Roelants, S**., Ciesielska, K., Gheys, R., Verweire S., Chow, K.K., Van Bogaert, I. and Soetaert, W.

Knowledge For Growth: 30th of May 2013, Ghent, Belgium, **Poster Presentation**, "*Starmerella bombicola* as a platform organism." **Roelants, S**., Saerens, K., Derycke, T., Van Bogaert, I. and Soetaert, W.

Yeasterday 2013: 24th of May 2013, Ghent, Belgium, **Poster presentation**: "Closing the gap between the lab and industry: upscaling of biosurfactant production with *Starmerella bombicola*." **Roelants, S**., Moens, H., Van Bogaert, I., Vanlerberghe, B. and Soetaert W.

CINBIOS: Workshop valorisation of organic waste streams oils: oils and fats, 31th of January 2013, Leuven, Belgium, **Passive participation.**

CINBIOS: Forum voor Industriële Biotechnologie, 23th of October 2012, Mechelen, Belgium, **Oral presentation**, "Transforming *Candida bombicola* into a platform organism for the production of interesting biomolecules.", **Roelants, S.,** Saerens, K., Derycke, T., Van Bogaert, I. and Soetaert, W.

ICY: 13th International Congress on Yeasts: Yeasts for a sustainable future, 26-30th of August 2012, Madison, Wisconsin, USA, **Poster presentation**, "Transforming *Candida bombicola* into a platform organism for the production of interesting biomolecules.", **Roelants, S**., Saerens, K., Derycke, T., Van Bogaert, I. and Soetaert, W.

First Ghent Bio-Economy's International summer school on the integration of green and white biotechnology, 7-10 august 2012, Ghent, Belgium, **Poster Presentation**, "Transforming *Candida bombicola* into a platform organism for the production of interesting biomolecules." **Roelants, S**., Saerens, K., Derycke, T., Van Bogaert, I. and Soetaert, W.

RRB-8: Renewable Resources and Biorefineries VIII, 4-6th of June 2012, Toulouse, France, **Oral presentation**, "Production of the bioplastic PHA by the yeast *Candida bombicola."* **Roelants, S.**, Derycke, T., De Vos, W., Van Bogaert, I. and Soetaert, W. **and Poster presentation,** "Biotechnological production platform for tailor-made glycolipid Bio-surfactants". Roelants, S., Ciesielska, K., Li, B., Zhang, J., Chow, K.K., Lin, Y.-C., Groeneboer, S., Van Bogaert, I., Devreese, B., Van de Peer, Y. and Soetaert, W.

Yeasterday 2012: 16th of May 2012, Groningen, Netherlands, **Poster presentation**, "Development of a reporter system for the platform organism *Candida bombicola*". **Roelants, S**. and Demaeseneire, S., Saey, L., De Preester, K., Van Bogaert, I. and Soetaert, W.

YCGMB: 25th International Conference on Yeast Genetics and Molecular Biology, 11-16th of July 2011, Kortowo-Olsztyn, Poland, **Poster presentation**, "Production of the bioplastic PHA by the yeast Candida *bombicola*". **Roelants S.**, Derycke, T., Van Bogaert, I. and Soetaert, W.

Yeasterday 2011: 20th of May 2011, Leiden, The Netherlands, **Poster presentation**, "Production of the bioplastic PHA by the yeast *Candida bombicola*"**., S.**, Derycke, T., Van Bogaert, I. and Soetaert, W., **Roelants, S**., Derycke, T., Van Bogaert, I. and Soetaert, W.

16th PhD Symposium on Applied Biological Sciences, 20 December 2010, Ghent, Belgium, **Poster presentation**, "Biotechnological Production Platform for new Tailor-made Glycolipid Biosurfactants". **Roelants, S.**, Saerens, K., Ciesielska, K., Groeneboer, S., Van Bogaert, I., Devreese, B., Van de Peer, Y. and Soetaert, W.

VIBes in Biosciences, 2nd International VIB PhD student symposium 13-15th of October 2010, Leuven, Belgium, Passive Participation +workshops.

Second meeting on Yeast and Filamentous Fungi, 28-29 June 2009, Strasbourg, France, Passive Participation.

Student guidance

Practical excercices: 'General Microbiology (2009-2012)

Tutor of master and bachelor theses

Wullaert Shana, M.Sc. thesis (2013)

Verweire Stijn, Bachelor thesis (2012)

De Preester Krystle, Bachelor thesis (2012)

De Vos Wim, M. Sc. Thesis (2012) not finished

Derycke Thibaut, M. Sc. Thesis (2011)

Debreuck Justin, Bachelor thesis (2010)

Others

Member of the organising committee of the **Yeasterday** conference (24th of May 2013).

APPENDICES

APPENDICES

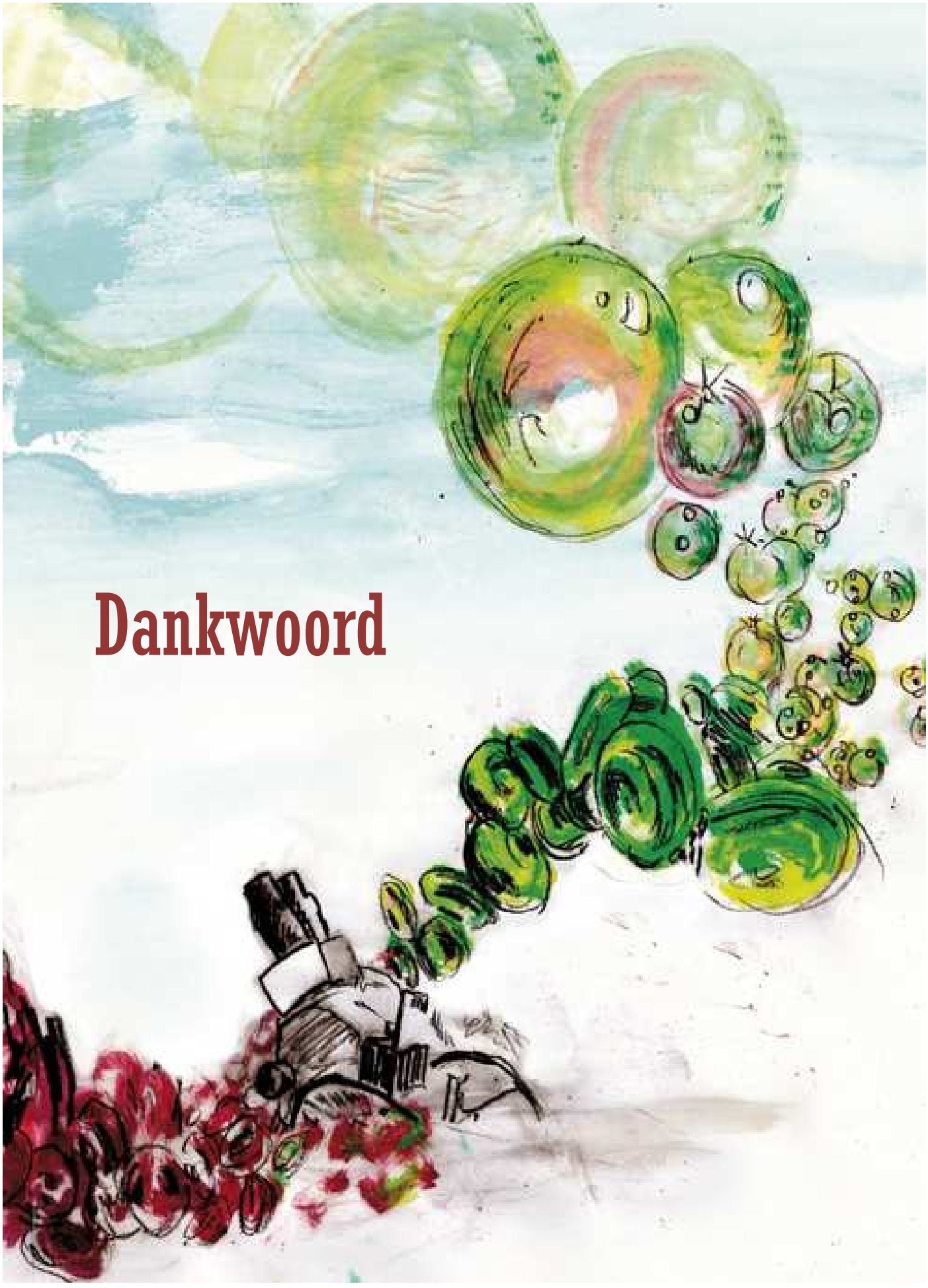
APPENDIX A: codon optimized gene sequences

Codon optimized DNA sequence of the GFPmut3B protein. The average codon usage of 8 genes with a codon bias, selected from the *S. bombicola* genome was used to perform codon optimization. An added multicloning site (MCS3) and *cyp52M1* terminator sequences are underlined. The resulting gene sequence was termed *gfpco10g*.

CACTAGTATACCCGGGCGCCTCAGCTCTTCGATGTCGAAAGGAGAGGAGCTGTTC
ACTGGCGTCGTCCCAATCCTGGTCGAACTTGATGGCGACGTTAACGGTCACAAGT
TTAGCGTGAGCGGTGAAGGTGAGGGTGATGCTACCTACGGCAAGCTGACTCTTA
AGTTCATCTGCACCACTGGCAAGCTTCCAGTTCCATGGCCCACACTCGTTACAAC
CTTCGGCTATGGCGTTCAGTGCTTTGCCCGTTACCCCGACCACATGAAGCAGCAC
GACTTCTTCAAGTCGGCCATGCCTGAGGGCTATGTTCAGGAACGTACTATCTTCT
TCAAGGACGACGGCAACTACAAAACTCGTGCTGAAGTCAAGTTTGAGGGCGATA
CTTTGGTAAACCGTATCGAGCTCAAGGGCATTGACTTCAAGGAGGACGGTAACA
TTCTTGGTCATAAGCTTGAGTATAATTACAACTCTCACAACGTCTACATTATGGCT
GACAAGCAGAAGAATGGAATTAAAGTCAACTTCAAGATTCGCCATAACATCGAG
GACGGCAGCGTACAGCTCGCTGACCACTACCAACAGAACACTCCCATTGGCGAC
GGCCCTGTTCTTCTCCCTGACAACCACTATCTATCTACCCAGTCCGCCTTGTCCAA
AGACCCCAACGAGAAGCGCGACCACATGGTTCTTTTGGAGTTCGTTACTGCCGCT
GGCATCACGCACGGAATGGACGAGCTGTACAAATAAGTTTCTTAGCCTCCCATG
GAAGAAACGTTCCCTCCTTAATTGGTTCAATCAACCCATGTTAACTCAACCTGTG
GCAACCTTTTTATTATTCTGAGCACTTTACTCAAGAATGGCTGCAGTTCTTTATTT
ATGCTCTTTATGAAGAATCATGTAAATCATGTAAATCTTGAAGGAGTGCGTTTAA
CCGACGGATCC

Codon optimized DNA sequence of the PHAC1 synthase protein of *P. resionvorans*. The averaged codon usage of the 5 genes of the biosynthetic pathway was used for codon optimization. The peroxisomal target sequence (PTS1) is underlined.

ATGTCTAACAAGAATAATGAGGATCTCCAGCGCCAAGCGTCCGATAATACGTTG
AATCTAAACCCAGTGATTGGCATAAGGGGTAAGGATCTCCTCTCGTCGGCTCGGA
TGGTACTCCTTCAGGCAATCAAGCAGCCTTTTCATTCCGCCAAACACGTAGCTCA
TTTCGGTTTGGAATTGAAAAATGTGTTACTAGGACAAAGTGGTCTCCAGCCAGAG
GCCGACGACAGAAGGTTCAATGACCCAGCCTGGTCCCAAAATCCTCTGTATAAG
AGGTATTTGCAGACATATTTGGCATGGCGAAAGGAGCTGCACAGCTGGATTGAT
GAGTCTAATCTATCTTCGCAAGATGCTTCTCGGGGCCACTTTGTTATCAATCTAAT
GACGGAAGCAATGGCACCTACAAACAGTATGGCCAACCCCGCAGCTGTGAAGAG
GTTTTTCGAAACAGGTGGCAAGAGTTTATTAGATGGGCTCTCACACCTGGCCAAA
GACATGGTTAACAATGGAGGTATGCCCAGTCAAGTTAATATGGACGCCTTTGAG
GTGGGTCAGAACCTAGCAACCACCGAAGGAGCTGTCGTTTTCAGGAATGATGTA
CTGGAGCTAATACAGTACAAGCCGATTACAGAGTCAGTGTACGAGAGACCACTG
TTGGTTGTCCCCCCGCAGATAAACAAATTCTACGTTTTCGATCTGTCACCAGAGA
AATCGCTCGCAAGATTCTGTCTTCGTTCAAATCTACAGACTTTTATTGTGAGCTGG
AGGAACCCAACGAAAGCCCAAAGGGAGTGGGGATTGTCTACTTATATTGAGGCT
CTTAAAGAGGCTATTGACGTGATTCTAAAGATCACCGGGGCAAAGGATCTTAAC
ATCTTGGGAGCGTGCTCAGGCGGAATCACTACTGTTGCACTCTTAGGACATTATC
AGGCTATCGGCGAGACAAAGGTCAATGCATTTACACAAATGGTGAGCGTCTTGG
ATTTCAACCTTGATTCACAAGTGGCCCTCTTTGCGGACGAACAAACACTAGAGGC
TGCTAAACGACGCTCATACCAAGCAGGGGTGCTGGAAGGGAAGGATATGGCTAA
AGTCTTTGCTTGGATGAGACCTAACGATCTTATTTGGAATTACTGGGTAAATAAC
TATCTGCTCGGGAACGAGCCGCCCGCGTTCGACATATTATATTGGAATAACGACA
CTACAAGACTGCCCGCTGCCTTCCATGGGGAACTCGTAGAAATGTTCAAAACGA
ATGCACTTACGAGACCGAACGCACTCGAAGTATGCGGCACCCCTATCGATTTGA
AACAAGTCACGTCAGACTTTTTCTGCCTAGCTGGAACTACTGATCATATAACACC
ATGGGAAGCGTGTTACCGAAGCGCGCTCCTACTGGGAGGTAAGTGCGAGTTCGT
TTTGAGCAACTCAGGTCATATACAGAGTATCCTCAATCCTCCTGGGAATCCGAAG
GCACGGTTTAGCACGGGATCCGAAATGCCAAAAGATCCAAAGGCGTGGTTGGAG
AATGCGACCAAGCACGCCGATTCATGGTGGTTGCACTGGCAGCAATGGATCGGA
GAAAGATCGGGGAAAACGAAGAAGGCCTCTTTCACACTCGGGAATAAGGCGTTC
CCGGCAGGAGAGGCCAGCCCTGGAACTTACGTGCACGAGCGA<u>TCTAAGCTG</u>TA



Dankwoord

Slechts een fractie van een procent van het werk dat kruipt in een doctoraat betreft de receptie en het voorwoord. Nochtans wordt hier door 95% van de aanwezigen het meeste belang aan gehecht: -SLIK-. Iedereen verwerft maar al te graag een plaatsje in het befaamde dankwoord en nu mijn beurt gekomen is om dat veelbesproken stukje tekst neer te pennen bekruipt me een zekere stress. Geen reden voor writer's block echter, want bij het maken van een lijstje van de mensen die ik hier ZEKER niet mag vergeten, krijg ik zowaar een krop in de keel. Zoveel mensen hebben direct of indirect een steentje bijgedragen aan dit werk dat het me niet (meer) verwondert dat mijn lijst van uitgenodigden voor de verdediging grotere proporties aanneemt dan ik gedacht had. Nu ik echter aan die laatste -ZWARE- loodjes bezig ben, ben ik écht blij dat ik na de vele mondelinge en gemailde 'merci's' mijn oprechte dank zwart op wit kan vereeuwigen. Ik zal hier dus inkt noch pagina's sparen, U weze gewaarschuwd! Daar gaan we dan:

Vooreerst wens ik de leden van de examencommissie Professor Van der Meeren, Professor Höfte, Professor Banat, Professor DeVreese and dr. De Maeseneire te bedanken voor hun grondige evaluatie van dit werk.

Een persoon die een bijzonder woordje van dank verdient, is mijn promotor die dit allemaal mogelijk maakte. Na een reis rond de wereld wekte Professor Soetaert mijn interesse met een zeker gistje met een 'gevaarlijke' naam dat biosurfactants produceerde. Een goede maand later was mijn 4-year ticket geboekt en was ik officieel deel van het BioSurf team. Wim, bedankt dat ik die ongelofelijke koppige gist op mijn eigen manier mocht temmen (or at least try), ook al geloofde je er niet (altijd helemaal) in; fluoresceren zouden ze! Ook bedankt om me de kans te geven gelijkgestemde wetenschappers te ontmoeten op (grote) internationale congressen. Ik heb er ongelofelijk veel van geleerd. Onlangs vertelde iemand op een congres dat er twee soorten wetenschappers bestaan. Wel, ik kon je meteen plaatsen onder de categorie van 'disruptive thinkers'. Je wil de dingen veranderen en dat doe je dan ook, of er daar nu mee akkoord gegaan wordt of niet. Ik ben trots deel te mogen uitmaken van het radarwerk dat jij uittekende en ik hoop oprecht dat eens alle schakeltjes op hun plaats vallen, het geheel zich zal ontwikkelen tot een goed draaiende en geoliede machine (op groene energie uiteraard!). Yes we can! ☺

Inge, mijn co-promoter, als reisleider van het BioSurf-team moest jij aan het begin van de reis uitzoeken hoe deze groep van zeer verschillende karakters te gidsen op deze grote ontdekkingsreis met een gemeenschappelijk doel. Het was soms een bumpy ride, maar ik denk dat er een mooie symbiose ontstaan is. Je mag trots zijn dat het werk beschreven in jouw doctoraat de kiem is geweest voor zeer veelbelovend en divers onderzoek. Bedankt voor je hulp bij het nalezen van mijn schrijfsels en vooral het vermaken van het jasje van mijn artikel toen de nood het hoogst was. Ik hoop dat het BioSurfing team zijn weg nog lang verderzet op de avontuurlijke paden en dat je je reisleider skills verder kan verfijnen.

Ons gistje was niet altijd een even aangename reisgenoot. Ik heb er meer dan eens FERM op gevloekt. Een aantal mensen hebben een onmisbare bijdrage geleverd om de gedeelde reis met deze koppigaard ondanks alles tot een goed einde te brengen. Ten eerste zou ik Sofie willen bedanken. Sofie, jij gidste mij bij mijn eerste stappen in het labo en het uittekenen van de blauwdruk van de eerste van veeeeeeeeele strategieën (ik denk zelfs dat ik het verfomfaaide papiertje nog heb!). Je was op mijn reis in onbekend gebied als het geruststellende gewicht van die alwetende reisgids in mijn rugzak. Ik heb ongelofelijk veel aan jou gehad en te danken en wil je hier dan ook nog eens oprecht en zeer hartelijk bedanken voor alle kleintjes en grootjes. Merci! Dan waren er die twee andere blondines van het BioSurf team. Karen als metgezel met dezelfde geesten te geselen. Het voelde als een geruststelling om niet alleen te zijn op dat hobbelige pad. Bedankt om er te zijn en om me mee te nemen op mijn

allereerste congres, ik heb veel fijne herinneringen aan dat tripje naar Strasbourg. Veel succes met die twee nieuwe uitdagingen, of allez drie feitelijk! ☺ En dan blondie Lientje. Wat heb ik op jou kunnen bouwen! Jouw reis startte ongeveer gelijktijdig met de mijne en onze wegen versmeltten zich algauw tot een heuse 'joy ride'. Onze zonnelunchen en shopmomentjes luisterden niet alleen de vaak lange werkdagen op, maar brachten ook licht wanneer dat nodig was. Bedankt voor je luisterend oor, betrokkenheid, onmisbare hulp in het onderzoek en daarbuiten, vriendschap, enzoverder enzovoort! Je stapte af de BioSurf bus, maar ik ben zó blij dat je nu terug meelift op die andere dubbeldekker! Once you go InBio, you never go back?! ☺ 'Elk nadeel heb zen voordeel': jouw 'adieu' maakte plaats voor allrounder SV. Stijn, het was een waar genoegen om jou als stagestudent te mogen begeleiden, al moest ik de flauwe grapjes er wel bijnemen ;-p. Ik was enorm blij dat Liens gouden handen zo snel waardige opvolging kregen, en dat is een zéér mooi compliment hoor, maar ook terecht. Je bezit een kostbaar pakketje bagage en hebt zin voor ondernemen (als je zin hebt tenminste ☺). Ik hoop dat je deze skills naar waarde inschat en je verder ten volle ontplooit. Kijk vooral naar jezelf! Ik kan je niet genoeg bedanken voor AL het werk dat je het afgelopen jaar voor mij verzet hebt. Niet alleen in het labo, maar ook voor de Yeasterday website en toebehoren en natuurlijk ook om nog om last minute nog een cover uit je mouw te schudden ☺. EEN WELGEMEENDE en DIKKE MERCI! Dat brengt me bij de laatste twee passagiers op de BioSurf bus: Kwok en Robin. Kwok, ge zijt een specialleke (mijn favoriete frietsaus overigens ☺). Je hebt me op cruciale momenten uit de nood geholpen, echter soms ook fameus laten vloeken, maar je hebt bovenal een hart van goud, al besef je het zelf soms niet. Bedankt voor je hulp op en buiten het werk, het wordt ten zeerste geapprecieerd. Last but not least: Robin, de nieuwste aanwinst van BioSurf. Vers van de pers en vol fris enthousiasme en nieuwe ideeën. Bedankt om verder te breien aan het begin van die zuurverdiende promotorbank. Het doet me ongelofelijk veel plezier dat één van de grootste 'troublechilds' van dit werk (hopelijk) de pubertijd zal ontgroeien in jouw handen. Veel succes op jouw reis, hopelijk ben je iets met onze reisverhalen: keep your eyes on the road and your hands upon the wheel! Een aantal enthousiastelingen liftten ook een stukje mee en leverden elk een onmisbare bijdrage aan dit werk: Justin, Thibaut, Wim, Krystle, Stijn en Shana: bedankt voor jullie inzet en enthousiasme! Het ga jullie goed op de ex-InBio wegen!
Parallel met de Inbio-BioSurf bus reden twee andere reisgezelschappen mee. Hulpstukken werden vlot uitgewisseld en mijns insziens was de samenwerking met L-probe, onder leiding van Professor Bart Devreese, en de Bioinformatics groep, onder leiding van Professor Yves Van de Peer, een zeer mooi voorbeeld van grensoverschrijdend wetenschappelijk onderzoek. Ik wens hier dan ook deze twee groepen bijzonder te bedanken voor de enorm leuke samenwerking! Bing, Yao Chen and Professor Van de Peer: thanks for translating that mountain of data into our own personal easy browsable database. We are still only able to see the tip of that large iceberg, but I'm sure a lot more interesting stuff will come out of your work, if we could only buy some time!
Sara, Kasia and Professor Devreese: thanks for providing me with shelter to repeat that cursed growth experiment for one last but FOURTH time! It didn't bring me what I was looking for, but gave birth to our shared 'degradation' story. Sara and Kasia, your hands and brains were irreplaceable in the third chapter of this work. THANKS! I really enjoyed working with you and discussing about those sometimes weird, but intriguing results. Sara, I wish you a lot of luck with your new 'space project'. Kasia, your work was and is of paramount importance for Biosurf(ing), and parts of my results (e.g. Chapter VI and Chapter V) would not exist without the input of yours. We will (hopefully) celebrate this soon with some well deserved publications! I hope you will stay a bit longer on that Biosurfing bus, so we can finally find some answers to these ancient mysteries! ☺ Thanks for everything and good luck with your final sprint! My thoughts are with you!

Natuurlijk was onze bus niet de enige 'on the road'. Joeri, Jo, Gaspard en Marjan (of moet ik Professor De Mey zeggen? ☺): we go way back! De begeleider-student band blijft toch altijd een beetje bestaan.

Jullie waren bij me ‘along the way’. Bedankt om proviand aan steun en hulp te geven wanneer ik die nodig had, vanaf het prille begin tot bij het begin van mogelijk een nieuwe start. Veel succes met InBiose! Ik geloof in jullie! Veel andere beatniks te vinden op de InBio-road. Isa, merci voor al het lekkers en de grappige conversaties. Je hebt me meer dan eens opgemonterd wanneer er WÉÉR maar eens iets mislukt was. Ook ex collega’s als Nele (OPNIEEEEEEEEEEUW!!!), Silvia en Niiiiicolaaaaaaas slaagden daarin als geen ander; Jef met zijn ongelofelijk droge opmerkingen (vooral wanneer het aan de ogen prikkende sarcasme, waarachtig genoeg NIET werd opgemerkt: HILARITEIT!), Hendrik en Joke op ontspannende fietstochtjes en ienie wienie Giangy: sweetie pie you make magic smiles appear wherever you go! Thank you for those you conjured on my face: serotonin (besides the coffee to stay awake) is my favorite drug to fight the despair of another failed experiment. Liesbetje, je was een beetje de InBio mascotte en we missen je nog steeds, maar écht TOPPI dat je nu volledig gaat voor die nieuwe uitdaging en dat je ons steeds bent blijven verblijden met bezoekjes. Bedankt voor al de hulp en plezier, binnen en buiten deze muren en nog veel succes hé…primus! Schilleke, jouw stripfiguur allures maken mij gewoonweg BLIJ! ☺ Als labmanager, en soms ook psycholoog van dienst, hielp je me meer dan eens uit de penarie en dat eeuwige relativeringsvermogen van jou ROCKS AROUND THE CLOCK! Catherine, bedankt voor de onmisbare hulp in het analytisch labo. Hopelijk vind ik vanaf nu de tijd om te ‘oefenen’ zodat je me er in de toekomst niet meer afrijdt! ☺
InBio heeft de reputatie een echt feestlabo te zijn, let wel onder het motto: work hard, play hard! Het InBio effect transformeert het ongezelligste -niet nader genoemde- café in een gezellige kroeg. De InBio party crew maakte dan ook menig feestje onveilig om tenslotte als laatsten van de dansvloer gekeerd te worden. Bedankt voor deze ontspannende momenten: komisch duo Tommeke & Papa Pie, Blondie, mijn favoriete Hollander Erik, Wouter alles bijhalve kabouter aka Mr. DJ aka de grootste weldoener op aarde, café-voyeur Joeri, Whispering Koen, Don Juan Dirk, Eierkop, Mopper Smurf Kwok, Da White Negro, Druggy Dries, nieuwe aanwinsten Magali and stand up comedian Stijn, funky dancemoves Marieke en Jo en natuurlijk niet te vergeten de LEEUW van de dansvloer Le Floche, maar daar kom ik later op terug. Naast de rauwe beats kloppen er ook heel wat ontfermende ‘mama-harten’ bij InBio: die lieverds van het secretariaat staken voor elk administratief, organisatorisch, taalkundig al dan niet last minute probleem en/of uitdaging vlotjes een helpende hand uit de mouw. Dominique en Hilde jullie hebben me écht een paar keer fameus uit de nood geholpen het afgelopen half jaar, dus nogmaals een welgemeende MERCI! En Anneleen en Barbara: bedankt voor de hulp bij het plannen, verbeteren en hoofdpijn verhelpen ☺. Van de Glycodirect groep wil ik graag ook nog Karel bedanken voor het afstaan van precious HPLC-time en Barbara voor de warme momenten. Professor Tom De Smet wil ik hier ook graag bedanken voor het nalezen van het verwarrendste hoofdstuk van dit werk. Bedankt voor je nuttige opmerkingen! Walter, Sofie en ex-collega Tom, jullie zitten wat afgezonderd, maar ook jullie wens ik te bedanken voor de hulp al dan niet in de vorm van bemoedigende woorden.

Dan mijn allerliefste friendjes en friendinnetjes uit ‘Entwerpen’ gelijk ze het hier nabootsend trachten uit te spreken ☺ en het nog verdere ‘Hertals’. Na die wereldreis, dan wel terug in het Belgenland, maar mijn trip in gistland verhinderde mij vaak deel te nemen aan andere ‘gist-gerelateerde-activiteiten’ ☺ Jullie zijn er echter steeds een onvoorwaardelijke baken van steun en vriendschap geweest tijdens mijn doctoraat, maar ook tijdens het daarmee vervlochten leven dat de afgelopen twee jaar niet altijd even vlekkenloos verliep. Ik was oprecht ontroerd door al jullie positieve energie in die iets moeilijkere periodes. Als ik rond mij kijk dan voel ik mij écht gezegend dat ik velen onder jullie al meer als de helft van mijn leven als mijn beste vrienden mag beschouwen. We worden oud! Binnenkort zullen jullie met weemoed terugblikken op dit hoofdstuk, dat dan tot het verleden zal behoren, wanneer mijn beruchte katers terug de kop zullen opsteken! HAHA! ☺ Hier dan nog eens bij

naam een bonte mix en ik hoop écht dat ik niemand vergeet: Nikita, Veerle, Filip, Vochtje (tandenborstel!), Wouter, Diana, Toefke, Kristien, Rattekop, Molleke (bedankt voor de prachtige tekening!), Muri, Herre, Klojee en Jack ☺, Nai, Vanessa (jaaa ge krijgt een boekske ja!), Vlien, Sneertje, Cawoline, Dr. Hannibal, Hellen, Stijn on Fire, Kirsten, Stefan, Tanita, Gino, Betty, Martini, Lise, Charlie (Lucky bastard!), Katha, Carotje, Bart, Annelies, Elza, Renée, Jef, Ellen, Carole, Wim Kosterke, Mieke, Kristien en last but not least Bicky: onze reis eindigde tijdens dit avontuur, maar we hebben samen zeer leuke tijden beleefd en ik hoop onze reis, wanneer alle putten in de weg gerepareerd zijn, verder te kunnen zetten in een andere hoedanigheid.
Bedankt allemaal voor jullie vriendschap en steun, ik zie jullie graag!

Ik had jullie gewaarschuwd. Kort en bondig ging het niet worden, maar dat zijn jullie natuurlijk ook niet van mij gewoon hé, dus nog even doorbijten ☺ Eén van de laatste alinea's brengt me bij mijn familie. De Roelantsen en de Nicolaïs meerbepaald. Bedankt voor jullie oprechte interesse in mijn onderzoek en voor de vele 'lekkere' familiebijeenkomsten. De stamnaam 'bombicola' mag dan wel niet van Bompa Nicolai Olen komen, zoals Tante Yvette zeer snugger afleidde, maar de wetenschapsmicrobe komt ongetwijfeld vanuit die hoek! Bedankt daarvoor dus bompa en bomma op jou kan ik toch altijd rekenen om wat lekkers uit je mouw te toveren hé. Bedankt voor de lekkere taarten en cakes, er zal goed geschmikkeld worden! Overigens reken ik erop dat mijn meterke vanuit de hemel zorgt voor een voorspoedige verdediging, dus Moeke: doe eens een efforreke voor mij hé! ☺ Saar, merci om je last minute mee op die andere 'Road' te begeven. Het was het laatste moment van 'ontspanning' ☺ voordat deze marathon die nu afloopt inging, en ik heb er nog lang op kunnen teren! En dan natuurlijk mama, papa en broer Fille: merci voor het warme nest en mam en pap voor de onvoorwaardelijke steun (op vele vlakken), geboden kansen en geloof tijdens mijn studies, erasmus, thesis, reizen, verbouwingen en uiteindelijk dit doctoraat…tot het bittere einde. Jullie hebben vaak geklaagd mij te weinig te zien, maar nu jullie dit boekje in handen hebben, zullen jullie wel beseffen dat dit toch wel een grondige reden had hé. Vanaf nu beloof ik terug vaker te komen genieten van lekker mama-eten en papa-wijn en vooral aangenaam gezelschap in de ver van stille Kempen (als jullie drukke agenda het toelaat tenminste :-p)! ☺ Bedankt voor ALLES; jullie zijn ongelofelijke schatten!!!! Ik zou ook heel graag mijn schoonfamilie willen bedanken: op korte tijd hebben jullie mij op zeer hartelijke wijze opgenomen in jullie warme nest en mij ook ongelofelijk gesteund tijdens die laatste zware loodjes, traiteur Yo en courier Marc in actie! MERCI! MERCI! MERCI!

Nog een laatste alineaatje te wijden aan een bijzonder persoontje alvorens de finish in zicht komt: Flochke aka de LEEUW van de dansvloer ☺. Hoewel we elkaar als je het rationeel beschouwt zinsverbijsterend kort kennen, hebben we omgekeerd evenredigsgewijs ☺ (wat zegt de wordnerd DAAR van? :-p) al een karrevracht aan ervaringen, dromen en gekoesterde momenten in onze gezamenlijke koffer verzameld. Je eeuwige geduld met warrie fie, vanzelfsprekend voor jou; een gekoesterd kleinood voor mij. Meer dan de helft van ons duet situeerde zich op dat laatste zeer hobbelige en met obstakels bezaaide stuk van dat doctoraatspad zonder afrit. Gaan! Gaan! en nog eens gaan! Je kan niet geloven hoeveel steun en energie ik geput heb uit je loutere aanwezigheid en schouder om op te steunen…en tsjiepen natuurlijk ☺. Je liet je overigens gewillig 'bezigen' voor het maken van tabellen en figuren en liet voor een aantal van mijn deadlines zelf ook kostbare slaap en dat tot op het àllerlaatste moment…. Ik ben je daar ENORM dankbaar voor! Ik plan, als je het goed vindt, nog vele internationale wegen, paden en wateren met jouw onveilig te maken, te starten met die heerlijke Italiaanse verrassing, maar nu eerst nog even een keer of vijf keer sterven :-s

Nog één laaste keer voor iedereen uit VOLLE BORST: ne welgemeende THANK YOU!!!!!

PIF POEF PAF…'t is AF!!!