Ghent University

Faculty of Bioscience Engineering

Department of Biochemical and Microbial Technology

Engineering the metabolism of *Starmerella bombicola* for the production of tailor-made glycolipids

ir. Robin Geys

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

Doctor (PhD) in Applied Biological Sciences

Academic year: 2016-2017

Examination committee: Prof. dr. ir. Mieke Uyttendaele (Ghent University, Belgium)
Prof. dr. ir. Monica Höfte (Ghent University, Belgium)
Prof. dr. Vlada Urlacher (Heinrich-Heine Universität, Germany)
Dr. ir. Manu De Groeve (Ablynx NV, Belgium)
Dr. ir. Sophie Roelants (Ghent University, Belgium)

Promotors: Prof. dr. ir. Wim Soetaert (promotor)
Prof. dr. ir. Inge Van Bogaert (co-promotor)
Centre of Expertise-Biotechnology and Biocatalysis
Department of Biochemical and Microbial Technology
Ghent University

Dean: Prof. dr. ir. Marc Van Meirvenne

Rector: Prof. dr. Anne De Paepe

Robin Geys was supported by Flanders Innovation & Entrepreneurship, formerly known as the institute Innovation by Science and Technology. Grant number 121129.

Titel van het doctoraatsproefschrift in het Nederlands:

Engineering van het metabolisme van *Starmerella bombicola* voor de productie van nieuwe glycolipiden.

Cover illustration: Shutterstock ID 256289095

To refer to this thesis:

Geys, R. (2016) Engineering the metabolism of *Starmerella bombicola* for the production of tailor-made glycolipids. PhD-thesis, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium.

ISBN-number: 978-90-5989-968-1

List of Abbreviations

5-FOA	5-fluoroorotic acid
ACP	Acyl carrier protein
AMP	Adenosine monophosphate
AOC	Allene oxide cyclase
AOS	Allene oxide synthase
APGs	Alkyl polyglucosides
AT	Acetyltransferase
ARS	Autonomously replicating sequence
bp	Base pair
CBL	Cellobiose lipid
CMC	Critical micelle concentration
CPR	Cytochrome P450 reductase
CTAB	Cetrimonium bromide
DES	Divinyl ether synthase
dTDP	Deoxythymidine diphosphate
EDTA	Ethylenediaminetetraacetic acid
ELSD	Evaporative light scattering detector
ER	Endoplasmatic reticulum
FAD	Flavin adenine dinucleotide
FAMEs	Fatty acid methyl esters
FAS	Fatty acid synthase
FMN	Flavin mononucleotide
GDP	Guanosine diphosphate
GFP	Green fluorescent protein
HAA	3-(3-hydroxyalkanoyloxy)alkanoic acids
HPL	Hydroperoxide lyase
HPLC	High-performance liquid chromatography
HSL	Homoserine lactone
kb	Kilobases
LE	Lactone esterase
LB	Luria-Bertani medium
LC-MS	Liquid chromatography–mass spectrometry
MDR	Multi-drug resistant ABC transporter
MEL	Mannosylerythritol lipid
MEOR	Microbial enhanced oil recovery
MFE-2	Multifunctional enzyme type 2
MIC	Minimum inhibitory concentration
MRM	Liquid chromatography–multiple reaction monitoring
NaCl	Sodium chloride
NADH	Reduced Nicotinamide adenine dinucleotide
NADPH	Reduced Nicotinamide adenine dinucleotide phosphate
nat	Nourseothricin acetyltransferase gene
NDP	Nucleoside diphosphate

Ni-NTA	Nickel-Nitrilotriacetic acid affinity resin
NRPS	Nonribosomal peptides synthetases
P450	Cytochrome P450 monooxygenase
PBS	Phosphate buffer saline solution
PCR	Polymerase chain reaction
pGAPD	Glyceraldehyde 3’-phosphate dehydrogenase promotor
pGKI	Phosphoglycerate kinase promotor
PGM	Phosphoglucomutase
PHA	Polyhydroxyalkanoate
PQS	Pseudomonas quinolone signal
QS	Quorum sensing
RhlA	Rhamnosyltransferase chain A
RhlB	Rhamnosyltransferase chain B
RhlC	Rhamnosyltransferase chain C
RmlA	Glucose-1-phosphate thymidylyltransferase
RmlB	dTDP-D-glucose 4,6-dehydratase
RmlC	dTDP-4-keto-6-deoxy-d-glucose 3,5-epimerase
RmlD	dTDP-4-keto-L-rhamnose reductase
RPLC	Reverse phase liquid chromatography
RPM	Rotations per minute
SD	Synthetic dextrose medium
SDS	Sodium dodecyl sulphate
TLC	Thin layer chromatography
UDP	Uridine diphosphate
UGTA1	UDP-glucosyltransferase A1
UGTB1	UDP-glucosyltransferase B1
ura3	Orotidine-5'-phosphate decarboxylase
YPD	Yeast peptone dextrose medium

Acknowledgments

“Moest je opnieuw kunnen kiezen, zou je het dan nog doen?”

Dit is waarschijnlijk één van de meest gestelde vragen aan een doctoraatsstudent gedurende de 4 zware jaren van onderzoek. Velen zullen zeggen, zeker tijdens de beruchte 2-jaarsdip, dat ze het niet weten of zullen volmondig “nee” antwoorden. De mensen die me kennen kunnen het antwoord waarschijnlijk al wel raden. Voor zij die me niet zo goed kennen, het is een resolute “ja”. Hoewel er soms periodes waren waarin het onderzoek eerder leek op een beerkar die vastgereden was is een zompige akker kan ik niet ontkennen dat alle goede, mooie, leuke en vooral interessante momenten veruit de mindere overschaduwen.

Natuurlijk heb ik dit niet alleen kunnen doen, vele mensen hebben me geholpen, bijgestaan of soms het bloed onder mijn nagels weten te krijgen in de aanloop naar dit doctoraat. Allereerst wil ik graag mijn ouders bedanken. Zonder hen zou ik hier nu niet staan met een vers doctoraat in de hand. Zij hebben me de afgelopen jaren altijd gesteund, me het volste vertrouwen gegeven en vooral altijd de vrijheid gegeven om te doen wat ik graag wou doen. Soms zullen ze zich wel afgevraagd hebben waarom ik godsnaam bepaalde zaken heb gedaan (of niet gedaan) maar zoals het goede ouders betaamt lieten ze de ruimte om zelf fouten te maken om je dan bij te staan indien nodig. In het huiselijke nest waren niet enkel mijn ouders een steun maar ook mijn zus had haar bijdrage. Iedereen met een broer of zus kan het wel beamen dat er gewoon eens zijn, eens onnozel doen of zelfs ruzie maken net die momenten zijn die blijven hangen.

Hoewel familie je stuurt en steunt doorheen je leven zijn er speciale momenten waarop je mensen leert kennen die evengoed een grote impact hebben. Tijdens de lagere school en alle jaren erna zijn er twee personen die ik speciaal wil vermelden, Ian en Tim. Merci gasten voor de vele namiddagen, avonden en nachten. Verder denk ik ook aan het Sint-Michielscollege. Hoewel het niet direct de favoriete school van mijn ouders was ben ik hier toch begonnen aan mijn middelbare carrière. Spreken over een carrière is misschien vreemd voor de eerste drie jaar, de term is echter wel van toepassing op de laatste vier. Hoewel ik vele mensen heb leren kennen in die periode zijn er slechts enkelen die ik af en toe nog eens hoor. De eerste namen die me zo te binnen schieten zijn Laura, Toon, Michael, Damien maar nog vele meer verdienen aandacht. Naast mijn medestudenten hebben ook enkele leerkrachten me weten te “sturen” tot waar ik hier nu sta. Merci Cuypers en Merciny. Het meeste hoor ik echter nog van mevrouw Bwarompy, Markske a.k.a. Marcia Bwarody. Altijd klaar voor een voos nieuw avontuur tijdens een nachtwandeling, altijd paraat om te babbelen of te luisteren (naar mijn soms non-stop geratel) en vooral altijd bereidt om haar ongezouten mening te geven. Merci Marcia voor de ontelbare keren dat ik jouw mening nodig had (en vooral kreeg) over soms iets pietluttigs zoals welke koptelefoon zal ik kopen ☺.

Naast het middelbaar zijn de hogere studies een andere tijdsperiode waarin veel nieuwe ervaringen worden opgedaan. De start van mijn loopbaan aldaar was in de Biomedische Wetenschappen. Het moet gezegd worden dat deze start eerder een soort valse start was aangezien een 18-tal maanden later ik ingeschreven was als aspirant bachelor in de bio-ingenieurswetenschappen. Het aantal mensen dat ik heb leren kennen in die zes jaar is ontelbaar. Gelukkig is het aantal vrienden wel makkelijker te overschouwen. Ze allemaal opnoemen zou te lang duren dus een kleine selectie dient zich aan. Vier jaar heb ik

doorgebracht op het studentenhome van campus Middelheim van de Universiteit Antwerpen. Hoewel ze eerder een rem op de studies te noemen waren wil ik toch mijn buren en kotgenoten vermelden. Merci Wouter, Karen, Ruben, Bart, Sam (zowel Puls als Verboven), Yannick, Alfons, Eef, Jens, Tim en Caroline voor de prettige avonden in de Konijnenpijp, Hagar, Buis, Profke, op de gang of op iemands kot. Daarnaast moet een student ook studeren. Hoewel bijna iedereen in de richting me pas heeft leren kennen tijdens de Zuid-Afrikareis (halverwege jaar 2) kan ik terugblikken op enkele mooie tijden met mijn studiegenoten. Vele van hen zie ik nog regelmatig en zijn waarschijnlijk ook aanwezig tijdens de verdediging. Merci Sven, Emilie, Elke, Laurent, Margo, Anouck en Daan. Ik ga me nu even verantwoorden, degene die langer in Gent zijn blijven hangen komen later.

In ieder geval, een deel van de studie is uiteraard de masterthesis. Dit was mijn eerste echte aanraking met degelijk wetenschappelijk onderzoek (de amateurastronomie daarbuiten gelaten). Het is tijdens deze periode dat het zaadje van het doctoraat zich pas wist te planten. Hoewel de argumenten om me binnen te halen bij Inbio tijdens de thesis slechts deels vervuld zijn geraakt (je maakt enkele nieuwe giststammen; je zal dan fermenteren met deze stammen; je kan dan naar de piloot gaan als alles goed gaat; je bent welkom op de barbecue op de piloot) blik ik terug op een bewogen jaar met enkele maanden verplicht niet pipetteren. De sfeer, het onderzoek en de mensen waren ruimschoots voldoende om me te overtuigen om nog enkele jaren te blijven hangen. Gelukkig was ik niet alleen in Gent. Enkele studiegenoten zijn ook blijven plakken voor enkele jaren. Met hun kon ik regelmatig eens op restaurant of gewoon bij iemand gezellig gaan koken. Magali, Sanne, Timothy, Marjolein en Thomas, wanneer is de volgende “Komen Eten”?

De 4,75 jaar na de masterthesis zijn de reden waarom dit dankwoord nu gelezen wordt. Tijdens die jaren heb ik veel studenten en medewerkers zien passeren. Wederom kan ik niet iedereen bedanken en/of vermelden. Ik ga dan ook beginnen bij het begin en ik zal wel zien waar ik uitkom. Merci Wim en Inge om me het vertrouwen te geven om te mogen starten aan een doctoraat. Zonder jullie inbreng, van ver en van dichtbij, zou ik nu niet deze tekst kunnen schrijven. Maar een doctoraat doe je niet alleen, je moet samenwerken met andere mensen, leren van hun fouten en ervaringen en vooral zelf proberen bij te dragen. Merci aan iedereen van het labo die mijn (onnozele) vragen wisten te beantwoorden of me op z’n minst me op een goed pad hebben gezet, die me iets hebben bijgeleerd. In mijn directe omgeving denk ik aan (voormalig) Biosurf, de mensen die mee bijdragen aan het creëren van nieuwe types glycolipiden. Merci Isabelle, Sofie DM, Sophie, Kasia, Kwok, Mieke H, Lien, Stijn, Lisa, Sofie L, Sylwia en Marilyn. Het labo was echter niet enkel Biosurf. Vele mensen van MEMO en Glyco hebben allen hun steentje bijgedragen. Ik denk bijvoorbeeld aan de vele klimpartijen die dienden als welkome afleiding (Dries, Griet, Martijn, Kwok, …), het leveren van essentiële figuren voor het doctoraat (Silke), het organiseren van reisjes en uitstapjes (Frederik, Margo, Dries, Maarten D, …), snel en veel fietsen (Magali, Gert, Thomas, Maarten D), marginalen leren kennen in een parkje (Bob), fitnessen (Maarten DM, Jorick, Stevie, Maarten D, Kwok) of door de vele schrale avonden in de Koe (Wouter, Eric, iedereen hierboven en zeker vele anderen). Een speciale vermelding verdienen mijn nalezers, namelijk Sofie DM, Sophie en Inge. Bedankt voor de tijd en moeite om door het manuscript te gaan en mijn soms beknopte en compacte schrijfstijl te masseren richting een iets meer verstaanbaar geheel.

Tot slot moet ik nog enkele mensen vermelden vooraleer ik kan afsluiten. Allereerst mijn studenten, zonder wie ik al dit werk niet had kunnen uitvoeren (Yaïka, Anouk, Julius, Margaux, Sylwia, Bélen, Christophe, Gwen, Niels, Jeroen en Jelle). Vervolgens mijn juryleden (Manu, Monica, Vlada en Sophie) en voorzitster (Mieke U.) voor de tijd en de moeite die het evalueren van een thesis vereisen. Tot slot, maar daarom niet minder belangrijk, is Chantal. Merci ☺.

Zoals jullie allen kunnen zien is het afleggen van een doctoraat meer dan de vier jaar die in het labo worden doorgebracht. Het is niet omdat iemands naam hier niet staat dat deze persoon niet het vermelden waard was, integendeel. Maar alle mensen opnoemen die iets hebben bijgedragen aan mijn tocht en het onderzoek dat eraan verbonden was zou me teveel doen uitweiden. Mijn excuses indien je hoop en verwachtingen niet voldaan zijn.

Om kort even samen te vatten: merci ma & pa, zus, Inbio, vrienden, collega's, verre kennissen en soms complete onbekenden voor de afgelopen jaren.

Robin Geys

13/02/2017

Table of Contents

Outline

Biotechnology has come a long way. Micro-organisms and enzymes have been used for thousands of years for preparing beer, cheese and even medicines. Though no fundamental understanding was present during those times, biotechnology was an everyday aspect of human life. With the advent of genetic engineering, new possibilities arose to steer the strength of this technology to create new kinds of molecules, enzymes or to optimize the micro-organisms used. Step by step, the knowledge was expanded and the possibilities increased alongside.

Thanks to a growing environmental awareness, alternatives for petrochemical derived molecules such as fuels, plastics or surfactants were researched starting from the capabilities of biotechnology for their production. In the case of fuels, biodiesel and biogasoline are both accepted renewable alternatives for their petrochemical rivals. Bioplastics such as polylactic acid and polyhydroxyalkanoates are used in both disposable and non-disposable applications such as wrapping material or medical implants. For the surfactants, several interesting molecule classes have been discovered, but their commercialization is hampered by high costs and a low molecule diversity combined with a minimal understanding of the wild type producers. The most promising class of these green surfactants, called biosurfactants, are the glycolipids. These molecules are composed out of a carbohydrate hydrophilic group attached to a lipidic hydrophobic group. Several types of them are already in use by industry. One of these are the sophorolipids, already applied in cosmetics and healthcare industry. The main producer of these sophorolipids is *Starmerella bombicola*, a yeast which in nature is associated with bumblebees. The molecules produced by this yeast are a mixture of different congeners. Though a mixture, certain structural similarities are conserved such as the preferred fatty acid tail. Attempts have been made during recent years to modify this yeast to steer the sophorolipid production towards a more uniform product.

Another successful class of glycolipids are the rhamnolipids. These molecules have interesting applications in for example pharmaceuticals, cosmetics and bioremediation. They are produced by bacteria belonging to the *Pseudomonas* and *Burkholderia* genera. Several species of these genera are known pathogens of humans, such as *Pseudomonas aeruginosa* and *Burkholderia pseudomallei*. Some of these species are capable of producing rhamnolipids in high amounts, making them an interesting candidates for further production purposes. Still, the pathogenic nature of several of the producers implies special measures to ensure employee safety during the production process and can have an impact on their applicability in certain industries. Attempts have been made to produce rhamnolipids in non-pathogenic bacteria such as *Escherichia coli* and *Pseudomonas putida,* but the overall production remains low.

The initial goal of this PhD was to engineer *Starmerella bombicola* to enable the production of rhamnolipids and from there on creating a production platform for other glycolipids as well. The benefits of this are multiple. First of all, *S. bombicola* is a non-pathogenic yeast. This already makes the production process less complex compared to the wild type producers of rhamnolipids. Furthermore, *S. bombicola* is resistant towards high levels of (bio)surfactants. Genetic engineering is possible and has led to several new sophorolipid producing strains already used in large scale production experiments. To produce rhamnolipids with *S. bombicola*, several aspects of their production need to be addressed. First of all, there is a

need for suitable precursors in the yeast cell. One of these precursors, the β-hydroxy fatty acids, are present though not accessible for the biosynthetic enzymes involved in rhamnolipid synthesis. The other precursor, activated rhamnose, is not present at all. Feeding all of these components is possible, but wouldn't make the production process economically interesting due to their high cost. Engineering *S. bombicola* for the production of these precursors was chosen to counteract these problems.

During the PhD, several obstacles became apparent when engineering *S. bombicola*. Though molecular tools for genetic engineering and screening purposes were present, other question remained. For fast and efficient engineering, plasmids are an indispensable tool. *S. bombicola* does not have a native plasmid so designing one was necessary. During Chapter II, an attempt has been made to create a functional, and preferably stable, plasmid for this yeast. After engineering and screening, pathways or genes are often integrated in the genome to ensure that the resulting production strains are stable. For efficient genetic engineering, multiple genes should be introduced during one single event. This was necessary as well for rhamnolipid engineering since up to 14 genes are required for the simplest molecule. Still, turning *S. bombicola* in a production platform for glycolipids in general also requires understanding where the synthesis happens inside the cell. This allows for more efficient targeting of biosynthetic enzymes when creating new glycolipid production strains.

Besides the more fundamental problems of creating a production platform, more practical hurdles have to be taken as well. As glycolipids are composed out of a carbohydrate and a lipid moiety, being able to modify both groups is a necessity. During the PhD, understanding both carbohydrate and lipid modifications was improved using the sophorolipid biosynthesis as a starting point. In Chapter III, new kinds of sophorolipids were created by changing the first step in their production pathway. Though several obstacles remain for large scale production, is was proven that even small modifications can have a significant impact on the structure of the molecules produced. In Chapter IV, carbohydrate engineering and strain optimisation resulted in strains producing glucolipids and cellobiose lipids. Though producer strains for both molecules already existed, the original engineered strains showed low productivity. Redesigning the molecular work and starting from strains with a higher overall fitness provided the necessary means for further enhancing production.

The final chapter focusses on the initial goal of producing rhamnolipids with *S. bombicola*. The engineering strategy was split up in three smaller subjects, one for the activated rhamnose, another for production of the hydrophobic moiety and a final one for coupling both into the final rhamnolipid molecule. For each subject, a suitable proof-of-concept was chosen to validate each part. At the end, no rhamnolipid molecules were produced, but some small successes were achieved while trying.

As a whole, this PhD points towards the applicability of *S. bombicola* as a production platform for different kinds of glycolipids. Several problems involving their production were investigated and solved. Albeit the final goal of rhamnolipids was not achieved, new knowledge on the production of other kinds of glycolipids might provide the necessary tools in the future to achieve this.

Chapter I

Literature Review

Biosynthesis of prokaryotic and eukaryotic biosurfactants

Parts of this chapter were published as:

Geys, R., Soetaert, W., Van Bogaert I., 2014.

Biotechnological opportunities in biosurfactant production

Current Opinion in Biotechnology, Volume 30, p66-72

Chapter I - Literature Review

I.1. Introduction

Surfactants, the major constituents of soap, are known since ancient times. The earliest evidence of soap-making dates back to the Babylonians 2800 years BC. The word soap is derived from the Latin word *sapo*, first mentioned by Pliny the Elder in his Historia Naturalis[1]. Throughout history, more applications, molecule types and resources were discovered and by 2014, the global surfactant market generated a revenue of 33 billion dollars, accounting for over 13 million tonnes. This market is expected to rise annually with 2.5 %[2]. The major interest in surfactants originates from their interesting properties. Their amphiphilic structure, combining both hydrophilic and lipophilic domains, allows them to interact with a broad variety of products and molecules. Surfactants locate themselves at the surface or interface between two phases like air and water or oil and water. They lower the surface tension of both phases, resulting in better contact between them. Besides lowering surface tension, surfactants are known for the formation of so-called micelles, structures made by the surfactant molecules when present above a certain concentration (critical micelle concentration or CMC). These micelles can have different forms and structures, depending on the structure and the concentration of the surfactant molecule. This is why they are extensively used in for example cleaning applications, food production, health care and many others. Most of these surfactants are obtained by chemical processes from petrochemical and/or oleochemical resources like earth oil or palm oil. They are divided in four main classes, describing the hydrophilic head group: ionics (either cationic or anionic), nonionics and amphoterics (containing both cationic and anionic parts).

During recent years, a shift towards more environmentally friendly surfactants is observed, caused by a growing awareness of both consumers and companies for the adverse effects that surfactants can have on the environment. For example, the majority of surfactants are produced from non-renewable petrochemical resources. Several companies have invested in producing surfactants based on renewable resources like glucose or oleochemicals. Examples of these ecological surfactants are alkyl polyglycosides (APGs), fatty acid methyl esters (FAMEs) and methyl sulphonates. These biobased molecules are often called biosurfactants while their production isn't in anyway more environmentally friendly than their classic petroleum derived cousins. Surfactants also have an impact on the environment after they have been used. Most surfactants are very resistant towards degradation which can lead towards accumulation of the surfactant in the environment. Furthermore, they have been proven to be toxic towards several marine organisms[3–5]. It has to be noted that a recent life cycle analysis of biosurfactants shows that their burden on the environment still can be significant. This is mostly due to the usage of first-generation renewable resources. Production of materials such as vegetable oils and glucose requires a massive amount of water and is in direct competition with food and feed production. Using second-generation materials such as agricultural waste streams will certainly lower the environmental impact[6].

Chemical production of glycosides was first reported by Emil Fischer in 1893[7]. Two main production processes can be identified for chemical glycosylations. The first one is acid-catalysed and can be seen as oligomerisation equilibria reactions. The Fischer synthesis uses a glucose donor such as starch and fatty alcohols to produce alkyl polyglucosides using an

acid catalyst and temperatures of approximately 110 °C. Though the carbohydrate moiety tends to polymerise during the production process, adjusting the ratio of the glucose donor and the fatty alcohol provides good control on the degree of polymerisation. The second production strategy are stereoselective substitutions on activated carbohydrates[8]. An example of such a production process is the Koenigs-Knorr reaction[9,10]. In this type of reaction, a glycosyl halide is condensed with a fatty alcohol producing alkyl glycosides. Though both production strategies have their own benefits and drawbacks, in general the production of alkyl polyglycosides by direct condensation reactions is a cost-effective way to generate chemical biosurfactants. If however a more pure compound is needed, the more exhaustive stereoselective substitutions can be used. This however implies more elaborate production schemes using expensive substrates and/or special catalysts, adding to the overall higher cost of molecules produced this way.

True biosurfactants, surfactants produced by prokaryotic and eukaryotic micro-organisms, are a worthy green alternative for the classic petrochemical surfactants. Because of their interesting properties like low ecotoxicity[11], easy biodegradability[5], mild production conditions and high effectiveness in various conditions[12,13], biosurfactants have numerous potential applications in a wide variety of industrial sectors like agriculture, pharmacy and medicine[14]. Biosurfactants can be divided in two main groups, the low and high molecular weight classes. Major examples of the former are fatty acids, phospholipids, lipopeptides and glycolipids. These amphiphilic molecules lower the surface tension between two phases, whereas the high molecular weight biosurfactants are often polyphilic and are rather stabilizers of emulsions, without significantly lowering the surface tension. The high molecular weight group are therefore also called bioemulsifiers and comprise, for example, polysaccharides, lipopolysaccharides and lipoproteins[15]. An overview of the different classes of biosurfactants as well as some producing organisms can be found in Table 1.1.

Table 1.1: Overview of common biosurfactant classes with their producing organism and reported yields. Yields are given in the same order as the producing organisms when more than one is listed. Nd = Not determined

Class	Biosurfactant	Producing organism	Titers (g/L)
Glycolipids	Sophorolipids	*Starmerella bombicola, Rhodotorula bogoriensis, Candida batistae*	>300[16]; 1.26[17]; 6[18]
	Rhamnolipids	*Pseudomonas aeruginosa, Burkholderia thailandensis*	>100[19]; 1.47[20]
	Mannosylerythritol lipids	*Ustilago maydis, Pseudozyma aphidis*	27[21], 75[22]
	Trehalose lipids	*Rhodococcus erythropolis*	40[23]
	Cellobiose lipids	*Ustilago maydis, Pseudozyma flocculosa*	1.7[21]; 15[24]
	Rubiwettin RG1	*Serratia rubidaea*	Nd[25]
	Xylolipid	*Pichia caribbica*	7.48[26]
Lipopeptides	Surfactin	*Bacillus subtilis*	6.45[27]
	Serrawettin	*Serratia marcescens*	Nd[28]
	Lichenysin	*Bacillus licheniformis*	0.025[29]
	Daptomycin	*Streptococcus roseosporus*	0.812[30]
	Gramicidins	*Bacillus brevis*	0.220[31]
Lipids	Oligomeric fatty acids	*Serratia rubidaea, P. aeruginosa*	Nd[25], 0.598[32]
	Spiculisporic acid	*Penicillium spiculisporum*	110[33]
Polymeric surfactants	Emulsan	*Acinetobacter calcoaceticus*	0.097[34]
	Liposan	*Yarrowia lipolytica*	8[35]
	Alasan	*Acinetobacter radioresistens*	2.2[36]
	Biodispersan	*A. calcoaceticus*	12[37]

The natural role of biosurfactants is still heavily debated. A generalization of their role is difficult due to their structural diversity and appearance in different ecological niches[38]. Their surface tension lowering capacities can enhance the availability of hydrophobic substrates for the organism and thus promote growth on these substrates. The biological activity of some biosurfactants might serve as a defence mechanism in ecological competition. It has been suggested that the production of sophorolipids provides an advantage to *S. bombicola* in niche protection as an extracellular carbon storage mechanism, while also having a mild antimicrobial effect[39]. Another natural role assigned to biosurfactants is the regulation of biofilm formation. Cell-bound surfactants can form an interface between a surface and the cell for easier attachment or other organisms can be inhibited in their biofilm formation. An extreme example of niche formation is the pathogenicity of *Pseudomonas aeruginosa*. This bacteria is dependent on the production of rhamnolipids for efficient colonization of their host organism[40]. Some biosurfactants have been shown to bind heavy metals, reducing their cell toxicity.

For the industrial production of biosurfactants, some obstacles still remain. Only a few are currently produced on an industrial scale and are commercially available, namely rhamnolipids, sophorolipids, mannosylerythritol lipids and surfactin. Sophorolipids and rhamnolipids are produced by companies like MG Intobio Co Ltd and Jeneil Biotech Inc (Table 1.2). These biosurfactants are already in use in cosmetics and cleaning products and future applications involving pharmaceuticals and bioremediation are nearby. Often issues like low yields combined with high purification costs, resulting in overall high costs, hamper market penetration and limited structural variation hinders further integration in existing application domains or the development of new ones (see Table 1.1 and Table 1.3).

Table 1.2: Overview of several companies producing biosurfactants[41].

Company	Biosurfactants	Applications
TeeGene Biotech	Rhamnolipids, Lipopeptides	Pharmaceuticals, cosmetics, antimicrobials
AGAE Technologies	Rhamnolipids	Pharmaceuticals, bioremediation, EOR
Jeneil Biosurfactant Co. LLC	Rhamnolipids	Cleaning products, EOR
Rhamnolipid Companies, Inc.	Rhamnolipids	Agriculture, EOR, pharmaceuticals
Fraunhofer IGB	Cellobiose lipids, MELs	Cleaning products, pharmaceuticals
Cognis Care Chemicals	Alkyl polyglucosides	Household cleaners, agrochemicals
Saraya CO. Ltd	Sophorolipids	Cleaning and hygiene products
Ecover Belgium	Sophorolipids	Cleaning products, cosmetics, bioremediation
Groupe Soliance	Sophorolipids	Cosmetics
MG Intobio Co. Ltd.	Sophorolipids	Cosmetics, personal hygiene
Synthezyme LLC	Sophorolipids	Cleaning products, cosmetics
Allied Carbon Solutions Ltd.	Sophorolipids	Agricultural products
Henkel	Sophorolipids, Rhamnolipids	Glass cleaning, cosmetics
Lion Corporation	Methyl ester sulfonates	Cleansing products
Kaneka Co.	Sophorolipids	Cosmetics and toiletry products
Evonik	Sophorolipids	Household cleaners

To counter these problems, a lot of effort has been put into engineering the production process using either cheap waste streams[42] or optimizing the fermentation parameters and product recovery[43]. With the rise of biotechnological engineering, new ways for improving the production of biosurfactants became available. Both the engineering of wild type producers and the development of heterologous production systems are viable approaches towards increased yields and enlarged molecular diversity.

Table 1.3: Prices for different kinds of (bio)surfactants for bulk orders. Prices are in $/kg unless otherwise stated. (Source: www.zauba.com).

Surfactant	Price per kg (bulk price)
Triton® X-100	$16.00
Sodium Stearate	$2.09
Cocamidopropyl Betaine	$1.92
Decyl Glucoside 55%	$3.29
Sodium Lauryl Sulfoacetate	$26.45
Sodium Laureth Sulfate 27%	$1.47
Sodium Dodecyl Sulphate	$1.28
Cetyl Trimethyl Ammonium Bromide	$9.66
First generation biosurfactants	
APG	$1.60
Methyl ester sulfonates	$0.95
Fatty acid methyl esters	$0.75
Second generation biosurfactants	
Di-rhamnolipid solid 97%	$270.27
R90L (90% pure rhamnolipid mixture)	$100 for 1 liter
Sophorolipids	$34.26
Surfactin	$39.89
Mannosylerythritol lipid	$252.12

In the following sections the industrially available rhamnolipids, sophorolipids and mannosylerythritol lipids will be discussed, as well as cellobiose lipids and lipopeptides, which are also considered of industrial and economical relevance. These glycolipids and lipopeptides are both low-molecular weight biosurfactants.

I.2. Rhamnolipids

I.2.1. Introduction

Rhamnolipids, perhaps the most studied glycolipids to date, are composed of one or two L-rhamnose molecules coupled to a mono- or dimer of β-hydroxy fatty acids[44] (Figure 1.1). The production of rhamnolipids depends on two pathways delivering the necessary intermediate molecules. The first one produces dTDP-L-rhamnose by the *rmlABCD* operon. The enzymes coded by the genes of this operon convert glucose-1-phosphate to the activated rhamnose that serves as a donor for the hydrophilic part of the rhamnolipids. The hydrophobic part is synthesized by dimerization of β-hydroxy fatty acids coupled to an acyl carrier protein, originating from the fatty acid synthesis which will be discussed later on. Two of these molecules are coupled by rhamnosyltransferase A (RhlA) to form 3-(3-hydroxyalkanoyloxy)alkanoic acids (HAA). Rhamnosyltransferase B (RhlB) will then transfer the rhamnose from dTDP-L-rhamnose to the HAA to form a monorhamnolipid. Rhamnosyltransferase C (RhlC) can use this monorhamnolipid to create dirhamnolipids[44]. Besides the rhamnolipids, HAA molecules can be detected outside the cell as well. Initially it was thought that they originated from degradation of the secreted rhamnolipids, but addition of deuterium-labelled rhamnolipids did not result in labelling of the free HAAs. Their existence might be explained by the fact that rhamnosyltransferases have a higher specificity towards longer HAAs with chain lengths of 10 or 12 carbon atoms while the secreted ones are mainly 8 or 10 carbon atoms long[32].

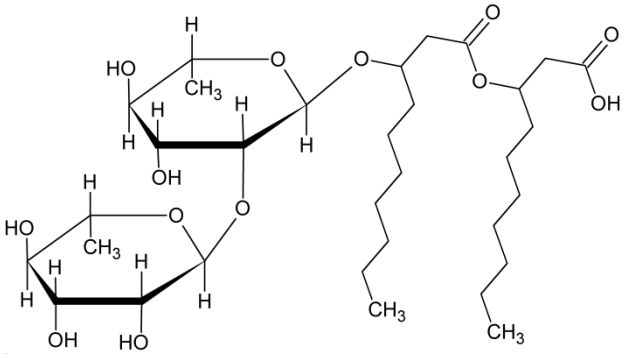


Figure 1.1: Structure of a rhamnolipid with two rhamnose moieties and two β-hydroxydecanoic acids.

Rhamnolipids are produced by several species from the *Pseudomonas*[45] and *Burkholderia*[46] genera with *Pseudomonas aeruginosa* being the top producer with titers over 100 g/L. They have the capability to reduce the surface tension of water from 72 mN/m to 24 mN/m[47], which is better than sodium dodecyl sulfate (SDS)[48] or cetrimonium bromide (CTAB)[49], two common chemical surfactants. Depending on the type of rhamnolipid tested, either a pure compound or a mixture, their critical micelle concentration varies between 10 and 234 mg/L. During degradation studies, it was found that compared to CTAB, SDS and Triton X-100 the rhamnolipids showed nearly full degradation in a timespan of several days[4]. Rhamnolipids are hence extensively tested for their properties in bioremediation of contaminated soils. They tend to form complexes with metal ions like cadmium, lead, nickel and mercury[50]. In the case of cadmium, up to 71.9 % is removed from kaolin, a typical soil component[51]. In view of their biodegradability, it is stated that rhamnolipids are stable enough to sequester metal cations, but unstable enough to not form a threat to the ecosystem. Solubilization followed by microbial degradation of complex hydrocarbons like diesel fuel[52], pentachlorophenol[53] or phenanthrene[54] is elevated when using rhamnolipids either as sole surfactant or in a mixture with other (bio)surfactants. Besides remediation they can also be used for microbial enhanced oil recovery (MEOR). Conventional oil extraction techniques only remove 30-45 % of the original oil in place. Using the excellent emulsification properties of rhamnolipids combined with their biodegradability make them interesting alternatives for chemical surfactants. MEOR can be applied *ex situ* by injecting culture broth or surfactant solutions or *in situ* by introducing the producing organism together with sufficient substrates. It has to be noted that oil fields contain high levels of H_2S, an inhibitory molecule for both growth of *P. aeruginosa* and production of rhamnolipids. By co-culturing sulphide-removing strains like *Pseudomonas stutzeri*, *in situ* MEOR remains possible[55].

Rhamnolipids are commercialized by a couple of companies like Jeneil Biotech Inc. and Agea Technologies. Unfortunately, the higher price compared to chemical surfactants (Table 1.3) combined with the pathogenicity of *P. aeruginosa* and less than optimal uniformity of the product still hampers further market penetration.

A part of this PhD (Chapter 5) focuses on the production of rhamnolipids and engineering of its biosynthesis. Since all the necessary building blocks for these rhamnolipids are derived from the central metabolism of the bacterial cell, a more detailed discussion is presented in the following sections. In a first part the synthesis of the hydrophobic part of the rhamnolipid is discussed (I.2.2), in a second one the biosynthesis of the hydrophilic part (I.2.3) and in a third part their coupling (I.2.4). Further on the regulation of the pathway is explained (I.2.5), and finally some engineering possibilities are discussed (I.2.6).

I.2.2. Bacterial fatty acid synthesis

Producing fatty acids is one of the key elements any prokaryotic and eukaryotic organism has to accomplish to be able to grow, maintain its membrane structures or modify biomolecules like oligosaccharides or proteins. In most eukaryotes, except plants, all the reactions are performed by the type I fatty acid synthase system (FAS I)[56]. This system is a big polyfunctional multizyme consisting of all necessary domains to synthesize fatty acids starting from acetyl-CoA. The bacterial type II system (FAS II) is a dissociated system with each catalytic step being carried out by a different enzyme. Even though the structure of both systems varies extremely, four steps can be identified that are shared between them: (1) a condensation, (2) a reduction, (3) a dehydration and a (4) final reduction. Another similarity of both systems is that they perform a cyclic reaction, using the product of one cycle as a substrate for the next until the desired fatty acid is produced. These steps as well as the enzymes responsible will be discussed below. A scheme of the cyclic reaction mechanism of the pathway can be found in Figure 1.2.

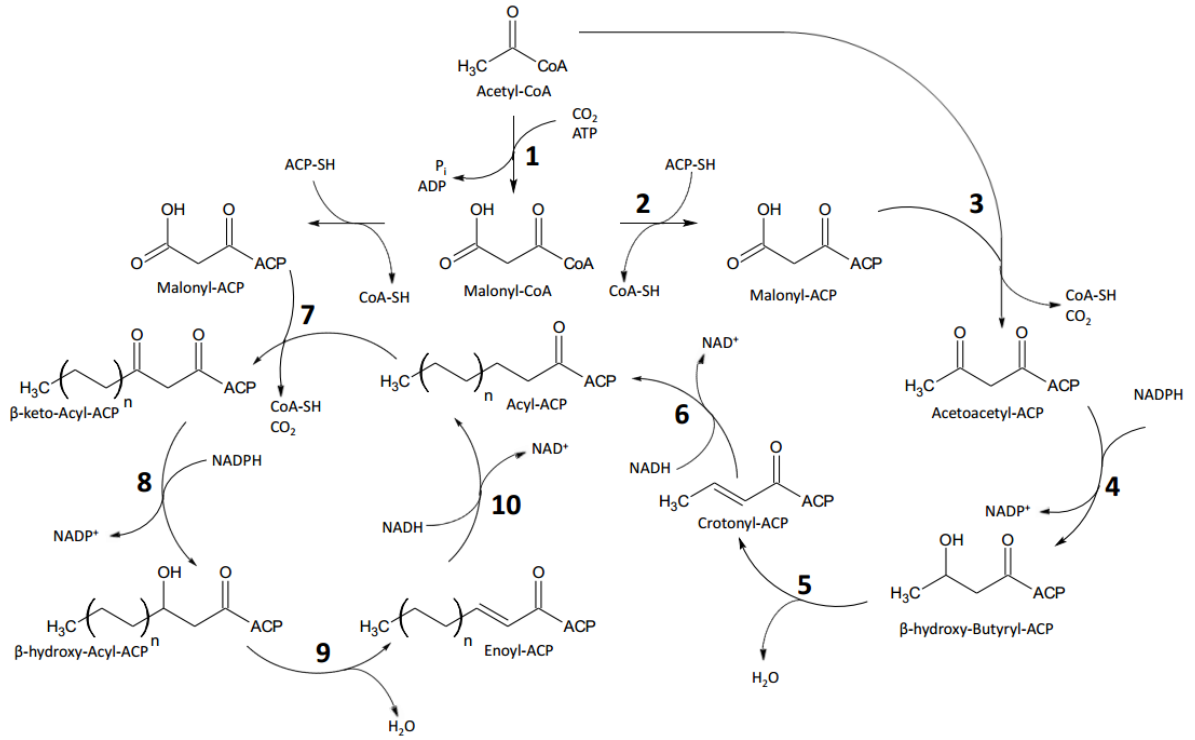


Figure 1.2: Fatty acid synthesis in prokaryotic organisms. Initiation of the synthesis starts at the right hand side while elongation of the acyl chain is shown at the left side. (1): Acetyl Carboxylase (ACCase), (2): Malonyl-CoA:ACP Transacylase (FabD), (3): β-Ketoacyl-ACP Synthase III (FabH), (4) & (8): β-Ketoacyl-ACP Reductase (FabG), (5): β-Hydroxyacyl-ACP Dehydratase (FabZ) and β-Hydroxydecanoyl-ACP Dehydratase (FabA), (6) & (10): Enoyl-ACP Reductase I (FabI), II (FabK) and III (FabL), (7): β-Ketoacyl Synthase I (FabB) and II (FabF), (9): β-Hydroxyacyl-ACP Dehydratase (FabZ) and β-Hydroxydecanoyl-ACP Dehydratase (FabA). n = [0, 7]. When the fatty acid has reached its final length, a thioesterase will hydrolyse the bond between the ACP and the fatty acid. In *P. aeruginosa*, step 3 is catalysed by FabY instead of FabH.

I.2.2.1. Acyl Carrier Protein

A central factor of the FAS II system is the acyl carrier protein (ACP), a small 10 kDa protein consisting of 78 amino acids. The ACP carries the growing acyl chain during the different catalytic steps. In *P. aeruginosa*, the ACP is encoded by *acpP*[57] and lies within the cluster of fatty acid biosynthetic genes. Similar genomic configurations can be found in other species like *Escherichia coli*[58], *Vibrio harveyi*[59] or *Bacillus subtilis*[60]. It has been proven that three different ACPs are present in *P. aeruginosa*. One of these is for example necessary for the production of N-butyrylhomoserine lactone molecules, essential for quorum-sensing[61]. The sequence and structure of the ACP is highly conserved between these and other species with identities over 80% being common. To be able to carry the growing acyl chain, the ACP needs to be converted from its inactive apo-ACP state to the active holo-ACP (Figure 1.3). This activation is done by a 4'-phosphopantetheinyl transferase, an enzyme transferring a pantetheine 4'-phosphate from Coenzyme A to a serine residue contained in the DSLD consensus amino acid sequence[62]. The intermediates in the synthesis are coupled to the terminal thiol group of the pantetheine by a thioester bound.

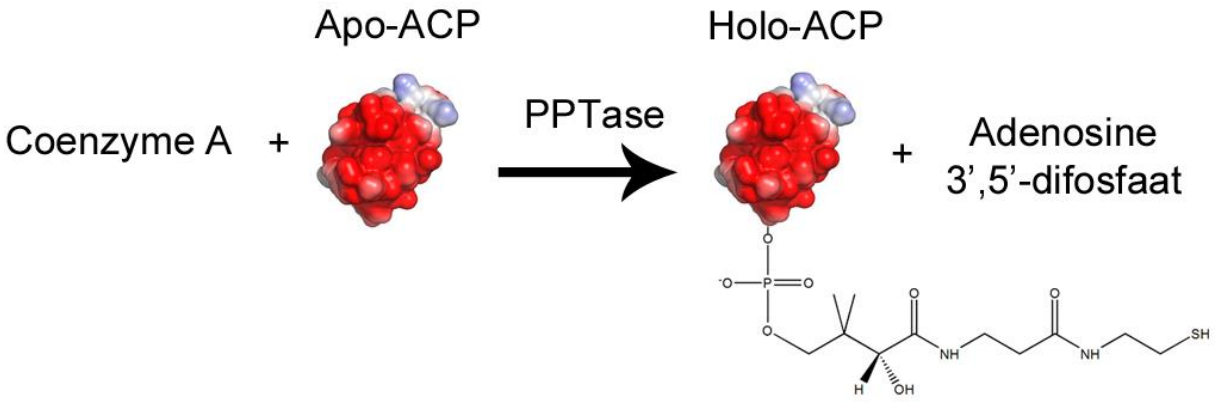


Figure 1.3: Schematic representation of ACP activation by the 4'-phosphopantetheinyl transferase (PPTase)

During synthesis, the growing acyl chain is buried within the hydrophobic core of the ACP between helix II and III[63]. It has been proposed that hiding the cargo protects the thioester bound or acts as a ruler to control the size of the fatty acids. It has been noted in *E. coli* that an ACP carrying an acyl chain has a smaller diameter compared to a non-acylated variant and is more resilient towards denaturation[64]. This stabilizing effect has been found to be proportional to the length of the acyl chain up to chain lengths of eight carbon atoms[65]. During docking of the ACP with one of the reaction partners, the acyl chain inside the ACP is transferred by a chain-flipping mechanism[66]. The exact way how this flipping mechanism is triggered is still to be discovered. Even though the ACP interacts with a wide variety of reaction partners, no clear binding sequence has been detected[67,68]. Still, structural research has shown that electrostatic interactions exist between the ACP and its reaction partner. Along helix II of the ACP, several electronegative residues can be detected that interact with electropositive residues in the ACP-binding sites of the reaction partners. Depending on the reaction partner, specific amino acid residues are responsible for interaction[69].

I.2.2.2. Acetyl-CoA carboxylase (Acc)

The first step in the synthesis of fatty acids is the generation of the essential precursor molecule malonyl-CoA (step 1 in Figure 1.2). This is governed by the biotin-dependent enzyme Acetyl-CoA carboxylase. In *E. coli*, Acc is composed out of four different subunits[70] which are carboxyltransferase α[71], carboxyltransferase β[71], biotin carboxyl carrier protein (BCCP) and biotin carboxylase[72]. Two half-reactions can be identified, the carboxylation of biotin by the biotin carboxylase using ATP and bicarbonate and the transfer of this carboxyl group onto acetyl-CoA to produce malonyl-CoA[73,74]. The biotin used in this process needs to be covalently bound to the BCCP. It has been shown that overproducing these subunits in *E. coli* can elevate the production of fatty acids by enlarging the pool of malonyl-CoA inside the cell[70]. Therefore, it has been suggested that Acc is a key regulatory enzyme of the fatty acid biosynthesis.

I.2.2.3. Malonyl-CoA:ACP transacylase (FabD)

In order to be available for fatty acid synthesis, the malonyl-CoA precursor needs to be transferred to the ACP (step 2 in Figure 1.2). This is done by malonyl-CoA:ACP transacylase, also known as FabD or MCAT using a ping-pong kinetic mechanism. The transfer of the malonyl-group is mediated by an intermediate malonyl-enzyme complex. In this state, the malonyl-group is covalently bound to a highly reactive serine residue[75]. This serine residue is part of a highly conserved pentapeptide GHSLG, strongly resembling the serine-dependent acyl hydrolases motif GXSXG[76]. It is worth mentioning that in a typical serine hydrolase, the intermediate acyl-enzyme is not stable and hydrolyses quickly in the presence of water. FabD however efficiently stabilizes this bond in a watery environment[77].

I.2.2.4. β-Ketoacyl-ACP Synthase III (FabH, FabY)

The first condensation reaction involves malonyl-ACP and acetyl-CoA (step 3 in Figure 1.2). This reaction is catalysed by FabH in *E. coli*[78]. Like FabD, FabH uses a ping-pong kinetic mechanism with an acetyl-enzyme intermediate[79]. The substrate specificity of FabH in *E. coli* is strictly limited to short chains like acetyl and propionyl-CoA while the other condensation enzymes discussed below easily utilize chain lengths differing from four to 16 carbon atoms. Another remarkable aspect is that FabH prefers CoA thioesters instead of ACP-bound acyl chains[78,80]. Since FabH initiates the cyclic fatty acid production, it is an important regulator of the pathway. Long-chain acyl-ACPs are the most efficient inhibitors of FabH. Decreasing the concentration of these molecules resulted in an increase of short-chain acyl-ACP types[80–82]. Strains overexpressing thioesterases, enzymes capable of degrading long-chain acyl-ACPs into a fatty acid and a holo-ACP, show constitutive production of fatty acids under otherwise limiting conditions[83,84].

Most bacteria only have one copy of the *fabH* gene in their genome. Noteworthy exceptions are *B. subtilis*[85] and *P. aeruginosa*[86]. In *B. subtilis*, the two FabH homologs are involved in the production of branched-chain fatty acids using iso- and anteiso-branched-chain acyl-CoA substrates. Expression in *E. coli* of these enzymes resulted in the production of a branched-chain, odd-chain fatty acid[85]. In *P. aeruginosa*, the story becomes even more complicated. Three potential genes have been identified as FabH while none of these actually are involved in initiating the fatty acid biosynthesis. Analysis of one of these *fabH* homologs, PA3286 or FabH3, showed that it was involved in shunting products, more specifically octanoyl-CoA, from the β-oxidation into the FASII pathway[87]. Potentially, this is done to conserve energy since

many different molecules are derived from the next few intermediates in the FAS II pathway. One of them is for example 3-hydroxydecanoyl-ACP, used in rhamnolipid and polyhydroxyalkanoate synthesis[86]. Most likely, the real initiator of the biosynthesis is FabY, another β-ketoacyl synthases distinct from FabH. Deletion of this gene resulted in slower growth while complementation with the *E. coli* FabH restored the wild type phenotype[88].

I.2.2.5. β-Ketoacyl-ACP reductase (FabG)

The next step (steps 4 and 8 in Figure 1.2) in the production of fatty acids is the reduction of the generated β-ketoacyl-ACP. This is done by FabG using NADPH in the process. This enzyme was discovered thanks to its high sequence identity to several acetoacetyl-CoA reductases[58]. Yet, it belongs to the short-chain dehydrogenase/reductase (SDR) family of enzymes, so correct annotation can be problematic. Up until now, only one isozyme has been detected that is capable of reducing the β-keto group. In *E. coli* and other bacteria*, fabG* is an essential gene making it an interesting target for antimicrobial drug development[89]. However, up until now, no inhibitors have been discovered.

In *P. aeruginosa*, a second β-ketoacyl-ACP reductase is present, also discovered based on its high sequence identity[90]. It has been hypothesized that the product of this *rhlG* gene uncouples the rhamnolipid production from the fatty acid synthesis. Recent data however states that RhlG doesn't have the required ACP docking site near its active centre and *in vitro* assays show that it's activity is 2000-fold lower than the one of FabG[91]. It has been theorized that RhlG has evolved into a single substrate enzyme requiring different enzymes and reaction partners than the regular FAS II system present in *P. aeruginosa*.

I.2.2.6. β-Hydroxyacyl-ACP-dehydratase (FabZ) and β-Hydroxydecanoyl-ACP-dehydratase (FabA)

Dehydration of the β-hydroxyacyl-ACPs is achieved by two different dehydratases (steps 5 and 9 in Figure 1.2). FabZ can only catalyse the dehydration while FabA can dehydrate and isomerize the double bond in the acyl chain, a necessary step for the production of unsaturated fatty acids. FabA has a very high substrate specificity towards β-hydroxy-decanoate ACPs[92], but it can accept shorter hydroxyl-acyl-ACPs as well[93]. It has been found that FabA stays inactive throughout the biggest part of the unsaturated fatty acid synthesis[93] due to its structure; the active site lays at the end of a 0.6 nm straight hydrophobic tunnel. Hence, the bend tails of the cis unsaturated fatty acids are not able to enter this tube effectively excluding them from further catalytic steps by FabA[94]. Another peculiar property of this enzyme is that the reactions are freely reversible. This means that a part of the cis-enoyl-ACPs is converted back to trans-enoyl-molecules and further processed to saturated fatty acids. Because of this interconvertability, only a part of the products are the cis-enoyl-ACPs[95]. Ratios range from equal[81,96] to 75:22:3[97] for respectively β-hydroxy, trans-2 and cis-3 products. In *P. aeruginosa*, *fabA* is clustered together with *fabB* in one operon[98]. FabB encodes an β-ketoacyl synthase, essential to the first step in the further elongation of the fatty acid intermediates. Specific interplay between FabA and FabB is needed for the efficient production of unsaturated fatty acids. FabB effectively pulls the reaction in favour of the unsaturated fatty acids, thereby competing with FabA for the cis-enoyl-ACPs. This in part explains that strains lacking FabB cannot produce unsaturated fatty acids. Overexpression of both *fabA* and *fabB* results in strains with higher ratios of unsaturated fatty acids[99].

FabZ is the dehydratase that is active in the further processing of unsaturated fatty acid intermediates and is more active than FabA on longer hydroxyl-acyl-ACPs. It has been suggested that variations in the FabA:FabZ-ratio are in fact one of the regulatory actions controlling the balance between saturated and unsaturated fatty acids[93]. The catalytic reactions of this enzyme are as easily reversible as the ones catalyzed by FabA. Kinetic characterization of *Plasmodium falciparum* FabZ clearly shows that the product of the reaction is again a good substrate for the reverse reaction. Though it has to be noted that these experiments were conducted with CoA-coupled molecules instead of ACP-coupled ones[100], similar results were reported for other species as well[101]. Structural analysis of *E. coli* FabA and *P. aeruginosa* FabZ revealed that both enzymes are nearly imposable. Both enzymes have the necessary key residues to perform dehydration and isomerization, but only FabA is capable of performing this reaction. A more close analysis revealed that in FabZ, the active site tunnel has a slightly different conformation effectively shielding the trans-enoyl-ACP from isomerization[102]. Nevertheless, it has to be noted that dividing FabZ and FabA in different classes is difficult due to the high structural similarities. *Enterococcus faecalis* utilizes for example a FabZ enzyme capable of replacing FabA in *E. coli*[103]. In *Streptococcus pneumoniae* even a third mechanism has been identified[104]. In this organism, the point where saturated and unsaturated synthesis separate lays behind the dehydration step. FabM is responsible for converting *trans*-2-decenoyl-ACP to*cis*-3-decenoyl-ACP.

I.2.2.7. Enoyl-ACP reductase I (FabI), II (FabK) and III (FabL)

Understanding the mechanism on how antibiotics like triclosan influence microbial metabolic steps has led to the characterization of enoyl-ACP reductases. In the special case of diazaborines, a single point mutation in the allele encoding FabI caused resisitence[105]. FabI is one of few enzymes that can reduce the trans-enoyl-ACP molecules to the corresponding acyl-ACPs, utilizing NAD(P)H as a cofactor for electron donation (steps 6 and 10 in Figure 1.2). Depending on which cofactor is used, the resistance of FabI towards antibiotics or overall stability can be altered. When using NADPH, an increase can be seen in antibiotic resistance, but the enzyme becomes less stable. It has been hypothesized that this is due the difference in charges between NADH and NADPH resulting in conformational changes. FabI also acts as an important control point for fatty acid biosynthesis since it's easily inhibited by palmitoyl-CoA, a molecule present when sufficient fatty acids are available[96,106].

Besides FabI, a few other enzymes exist that are able to reduce the enoyl-ACP intermediates. In species like *Enterococcus faecalis* and *Pseudomonas aeruginosa,* other classes of reductases can be detected[107,108]. FabL, specific to *Bacillus subtilis* was discovered by identification of the characteristic active site dyad tyrosine-X_6-lysine[109] also present in *E. coli* FabI. Reductases FabI and FabL are both members of the Short Chain Dehydrogenase/Reductase superfamily, sharing several properties even though sequence homology is rather low[110]. Shared properties are for example folding of the enzymes, presence of a Rossman fold for cofactor binding, the already mentioned conserved active site and size of the enzymes. Besides the already mentioned SDR-family members, two enoyl-reductases have been found that differ significantly even though catalysing the same reactions. These are FabK and FabV. In *Streptococcus pneumoniae*, the flavoprotein FabK is the only reductase present[111]. Interestingly, FabK has been shown to utilize only NADH, even in absence of its substrate. It is suggested that besides its role in fatty acid synthesis, it is also involved in oxygen-dependent anaerobic energy metabolism. FabV was initially found in *Vibrio cholera*

and proved to be essential for triclosan resistance[108,112]. Similar to FabK, it has a high preference for NADH over NADPH. It has been hypothesized that the high structural diversity of enoyl-reductases, compared to the other fatty acid synthesis reactions, is due to the existence of compounds inhibiting one or several enoyl-reductases[107,113,114] and their involvement in other metabolic processes than the production of fatty acids.

I.2.2.8. β-Ketoacyl-ACP Synthase I (FabB) and II (FabF)

Elongation of the fatty acid is performed by FabB and FabF (step 7 in Figure 1.2). FabB was discovered first in temperature sensitive mutants[115]. Though both enzymes are capable of condensing malonyl-ACP with the growing acyl chain and have a similar structure, they play different roles in the central metabolism of the cell. Deletion mutants of *fabB* are viable when supplemented with an exogenous source of unsaturated fatty acids. This proves the role of FabB in elongating the *cis*-3-decanoyl-ACP produced by FabA[116]. *In vitro* studies showed that FabF can also utilize the products of FabA. FabB is also known for its capacity to begin fatty acid synthesis by using a side reaction of malonyl-ACP decarboxylase in the absence of acetyl-ACP[117]. *In vitro* studies showed that both FabF and FabB can utilize a wide range of acetyl and ACP coupled fatty acids with a length up to 22 carbon atoms. Temperature adaptation in *E. coli* is controlled by FabF. At lower temperatures, FabF is far more capable than FabB to elongate palmitoleic (C16:1) molecules to cis-vaccenic acid (C18:1), a common membrane fatty acid in cold-adapted *E. coli*. Regarding antibiotic resistance, both enzymes are susceptible to cerulenin which forms an covalent bond with the active site cysteine[118].

I.2.3. Production of activated rhamnose

Rhamnose is a deoxy sugar commonly found in plants and bacteria. It is an essential component of the bacterial cell wall and can be found in the O-antigen part of the lipopolysaccharides of several microorganisms like *Salmonella enterica*[119] and in the rhamnolipids produced by *P. aeruginosa*. The bacterial system uses deoxy-thymidine-diphosphate-L-rhamnose as rhamnose source for rhamnolipid production. Production of this activated sugar starts from glucose-1-phosphate and by means of four catalytic steps it is converted to dTDP-L-rhamnose[120,121] (Figure 1.4). The enzymes show high sequence similarities between different species making it easy to generalize conclusions[122,123].

The first step in the conversion is carried out by a glucose-1-phosphate thymidylyltransferase called RmlA. In *S. enterica*[124], it catalyses a reversible group transfer using dTTP or UTP to create respectively dTDP- or UDP-D-glucose. During the reaction, the phosphate group of glucose-1-phosphate performs a nucleophilic substitution on the α-phosphate of dTTP. By this reaction, pyrophosphate is released and the stereochemistry of this phosphate is inversed[125]. RmlA is regulated by both allosterical and competitive inhibition by dTDP-L-rhamnose[126]. dTDP-L-rhamnose binds in the active site required for the production of dTDP-D-glucose. Due to its different structure, dTDP-L-rhamnose is protected from the pyrophosphorylase activity of RmlA[125].

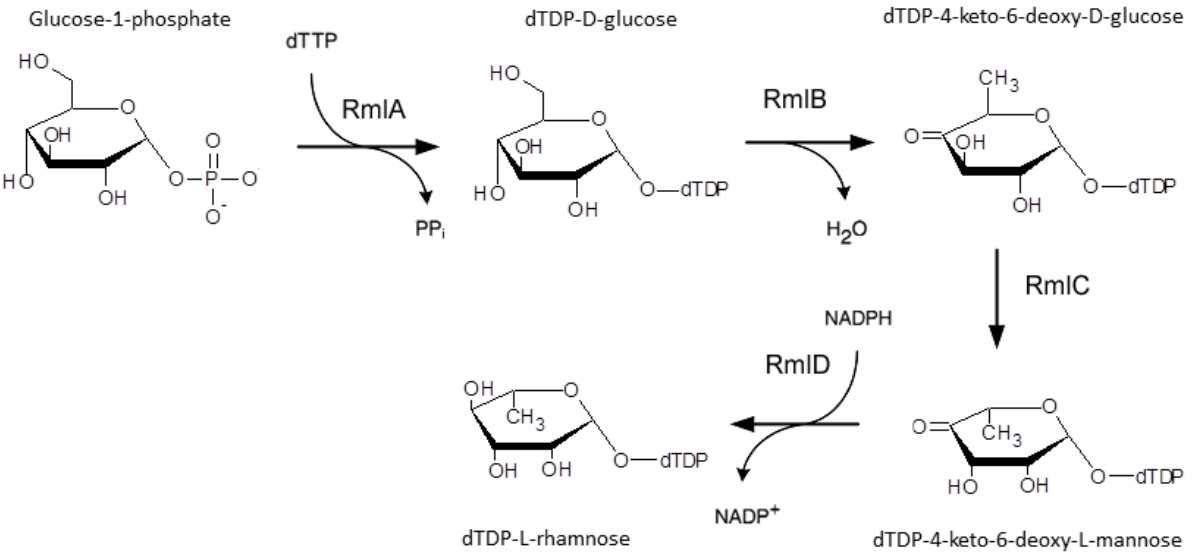


Figure 1.4: Production of dTDP-L-rhamnose starting from glucose-1-phosphate. First, thymidylyltransferase RmlA couples a thymidine group to the glucose-1-phosphate, creating dTDP-D-glucose. The next step is a dehydration by RmlB, followed by epimerization of the dTDP-4-keto-6-deoxy-D-glucose intermediate by RmlC. The final reduction step is carried out by RmlD resulting in dTDP-L-rhamnose.

RmlB catalyses the dehydration of dTDP-D-glucose to dTDP-4-keto-6-deoxy-D-glucose. It consists out of two domains, one for binding the cofactor NAD^+ and one binding the sugar-nucleotide. The first step in the mechanism is transferring the hydrogen from C4 from the glucose moiety to the NAD^+. This results in the dTDP-4-ketoglucose intermediate. Secondly, an essential glutamate residue removes a proton from C5. In the next step, a water molecule is removed creating the 4-keto-5,6-glucosene intermediate. This is followed by a hydride transfer from the NADH and inversion of the configuration of the glucose ring at the C6 position. Finally, a glutamate residue protonates the C5 position[127,128].

Epimerization of the dTDP-4-keto-6-deoxy-D-glucose is governed by RmlC. The configuration at positions C3 and C5 is inverted and the final product of this step is dTDP-4-keto-6-deoxy-D-mannose. The structure of RmlC does not resemble other known epimerases, suggesting it constitutes a different class[129].

The final step is the reduction of the C4 keto group by RmlD using NAD(P)H as the reducing cofactor. This step requires efficient catalysis as the intermediate keto-product shows enhanced reactivity. One essential element is Mg^{2+}, required for correct dimerization of RmlD. Sequence analysis has shown that RmlD might be part of the single domain reductase/epimerase/dehydrogenase (RED) protein superfamily[130]. Both a Rossman fold as several conserved catalytic residues have been identified.

I.2.4. Final steps in rhamnolipid biosynthesis

Coupling of β-hydroxyacyl-ACPs and dTDP-L-rhamnose is catalysed by the rhamnosyltransferases RhlA, RhlB and RhlC (Figure 1.5). In *P. aeruginosa*, transferases RhlA and RhlB are organized in the *rhlAB* operon[131]. In *Burkholderia* species, two identical gene clusters can be identified with all three transferases included[20,46]. It has been suggested that grouping *rhlC* together with *rhlA* and *rhlB* ensures equal stoichiometric ratios, thereby shifting the production to mainly di-rhamnolipids in the *Burkholderia* species.

The first rhamnosyltransferase RhlA doesn't transfer rhamnose as it name suggests, but instead it couples β-hydroxyacyl-ACPs to form β-D-(β-D-hydroxyalkanoyloxy)alkanoic acids (HAAs). Though mostly dimers are produced, the occasional trimer is also possible. In *P. aeruginosa* PA14, its most preferred substrate is β-hydroxydecanoyl-ACP[132]. When comparing substrates differing in only two carbon atoms in length like β-hydroxyoctanoyl-ACP and β-hydroxydodecanoyl-ACP, activities dropped to 9.8 % and 4.7 % respectively. This in part already explains the overall abundance of β-hydroxydecanoate in the *in vivo* produced rhamnolipids[133]. In *Burkholderia glumae*, the affinity of RhlA is highest for β-hydroxytetradecanoyl-ACP since the most abundant hydroxyl fatty acid of its rhamnolipids produced contains 14 carbon atoms[134]. The N-terminus of the transferase contains a putative signal peptide for potential targeting towards the membrane[131].

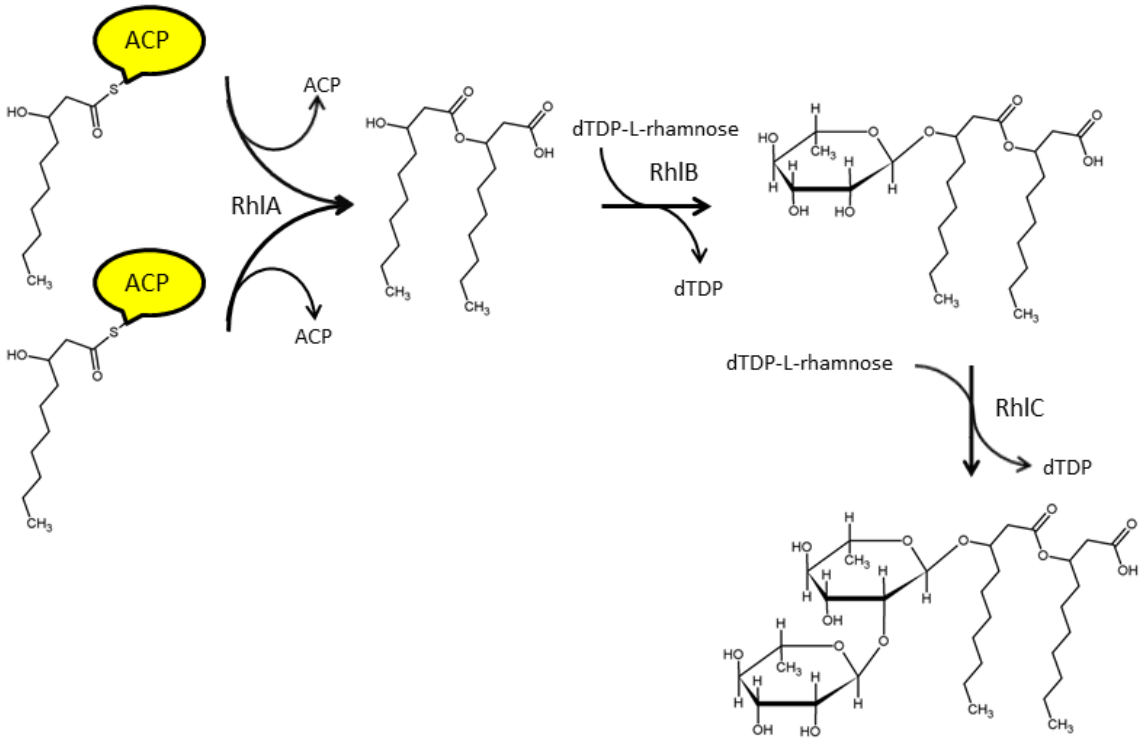


Figure 1.5: Final steps in the production of rhamnolipids. The first rhamnosyltransferase RhlA couples two β-hydroxy-decanoic acids to form a dimer of these molecules. The second transferase RhlB couples the first rhamnose moiety to form a monorhamnolipid. The second rhamnose necessary for dirhamnolipids is coupled by RhlC.

The second rhamnosyltransfere RhlB is essential for the production of mono-rhamnolipids[131]. In the process it uses the earlier produced HAA-molecules and dTDP-L-rhamnose. Sequence analysis has shown that the transferase contains two large hydrophobic regions, potentially transmembrane helices, suggesting that the transferase is membrane bound or at least associated. The final step in the biosynthesis is the addition of a second rhamnose moiety on the already produced mono-rhamnolipids by RhlC consuming again one dTDP-L-rhamnose as precursor[135]. Like RhlB, RhlC possesses a large hydrophobic region again suggesting its location inside the cell near the membranes. Interestingly, in *P. aeruginosa*, in the same operon upstream of the gene coding for RhlC lies PA1131. Though it has not been characterized yet, *in silico* structural analysis has shown that it belongs to the major facilitator superfamily of transporters. A similar genomic organization exists in *B. thailandensis* where up- and downstream of *rhlC* several genes potentially coding for transporters have been found[20].

I.2.5. Regulation

Production of secondary metabolites is often regulated by direct and indirect factors. In the case of rhamnolipids produced by *P. aeruginosa*, quorum sensing (QS) is an important part of the regulatory network controlling the biosynthesis of these molecules[136,137] (Figure 1.6). Quorum sensing can be explained as a way to sense population density by the local concentration of certain signal molecules. Various attempts have been made to understand and interfere with these regulatory mechanisms to uncouple production from the various factors influencing this system.

In *P. aeruginosa*, two major QS systems are operational[138]: *las* and *rhl*. Both systems produce a specific kind of QS molecule with a shared core structure. This core structure is a homoserine lactone (HSL), which is further processed in respectively N-3-oxo-dodecanoyl HSL by LasI and N-butanoyl HSL for RhlI. Both systems belong to the so-called LuxR-LuxI family regulators[139] and in both systems, the synthase of these molecules as well as the receptor protein of the mechanism are located in the same operon. The current understanding of the QS regulation is that the *las* operon acts as a global integrator of several signal cascades. Expression of the *las* operon is often followed by expression of the *rhl* operon.

Expression of the *rhl* operon results in the production of N-butanoyl HSL by RhlI, which in turn will activate RhlR. After activation, RhlR will bind on the *rhlAB* operon and express the first two transferases of the rhamnolipid biosynthesis. The role of RhlR is however subtler than just initiate expression of the *rhlAB* operon. When no N-butanoyl HSL is present, RhlR still binds to the promotor of the *rhlAB* operon, but effectively inhibits transcription[140]. Though it appears that the expression of the *las* operon regulates the expression of the *rhl* operon, experiments show that the expression of the latter operon is not necessarily a case of population density and that the *rhl* operon can also be expressed prior to the *las* operon. In mutant strains lacking the *las* operon, rhamnolipid production is still detectable, but productivity is often reduced. It has been found that the different regulatory systems can circumvent each other in case of a deficiency[141].

A third quorum sensing system is built around 2-heptyl-3-hydroxy-4-quinolone (Pseudomonas quinolone signal PQS). Production of this molecules is initiated by the *pqsABCD* operon and finalized by *pqsH* producing PQS. PQS will bind after a certain minimum level is reached to its receptor PqsR (also known as MvfR), which is expressed under control of LasR. The PQS-PqsR complex will further activate the *pqsABCD* operon resulting in elevated production of these molecules as well in elevated expression of PqsE, an enzyme elevating the activity if RhlR, again resulting in (higher) expression of the *rhlAB* operon.

Other regulators have a direct influence on the *rhlAB* operon as well. AlgR, a regulator involved in several pathways like alginate production[142], has a direct role in the expression of the *rhlAB* operon by binding the promotor[143] and inhibiting its activity. Deletion of this regulator shows an increased rhamnolipid production in biofilms compared to the wild type PAO1 strain[144]. Unfortunately, no differences could be observed when the knock-out strains were grown in liquid medium. Simple overexpression of AlgR did not guarantee *rhlAB* operon inhibition since AlgR requires activation by phosphorylation.

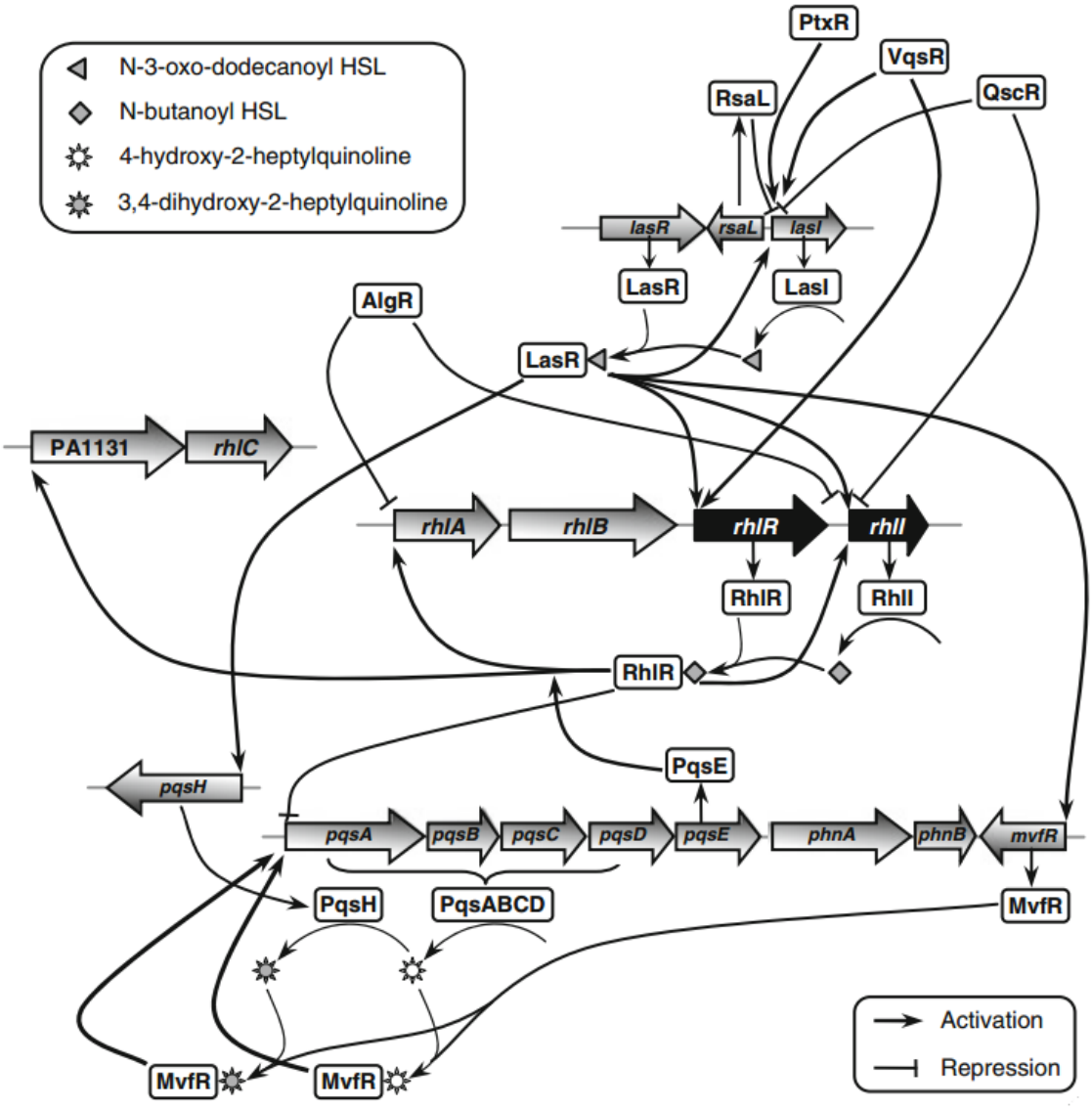


Figure 1.6: Part of the regulation controlling the production of rhamnolipids in *P. aeruginosa*[44].

Regulators like PtxR, RsaL, RsmA, DksA and several sigma factors like RpoN and RpoS[138,145,146] also influence the expression levels of the *rhlAB* operon. This is either by activating or inhibiting expression of either the *rhlAB* operon or its regulatory network. This explains why even without activation of the *las* system, production of rhamnolipids can occur.

I.2.6. Engineering possibilities

Great potential for obtaining more rhamnolipids lies in uncoupling their production from its complex regulatory system. This should ensure a stable and continuous production of the rhamnolipids. Another strategy for enhancing rhamnolipid production proved to be the introduction of the *Vitreoscilla* hemoglobin gene *vgb* in *P. aeruginosa*. Production for the engineered strain was higher than for the wild type NRRL B-771 strain[147]. This might be correlated to the higher oxygen uptake and availability, which enhances cell density and hence production. It has to be noted that more significant improvements were reported in the same

article by process engineering than by engineering of the strains themselves, resulting in yields being improved from 969 mg/L to 7543 mg/L. Though this increase in product titers is significant, it is still lower than the titers achieved by the wild type *P. aeruginosa* PAO1, the top producing strain.

Since the substrates used by RhlA are β-hydroxyacyl-ACPs, engineering the fatty acid synthesis to enlarge this pool of molecules has been proven successful as well. When inactivating either FabA or FabI, increased productivity by respectively 55 % and 36 % could be detected[132].

Production of rhamnolipids results in a mixture of both mono- and dirhamnolipids. By knocking-out *rhlC*, encoding a rhamnosyltransferase, only monorhamnolipids are produced without a big loss in production titers[135]. It has to be noted that in the research paper concentrations are expressed as milligrams of rhamnose per milliliter of broth. Since the wild type rhamnolipids are a mixture of mono- and dirhamnolipids, a direct comparison is difficult. Still, the strain lacking the second rhamnosyltransferase produces 120 ± 21 mg of rhamnose while the wild type producer PAO1 reaches 150 ± 15 mg per milliliter.

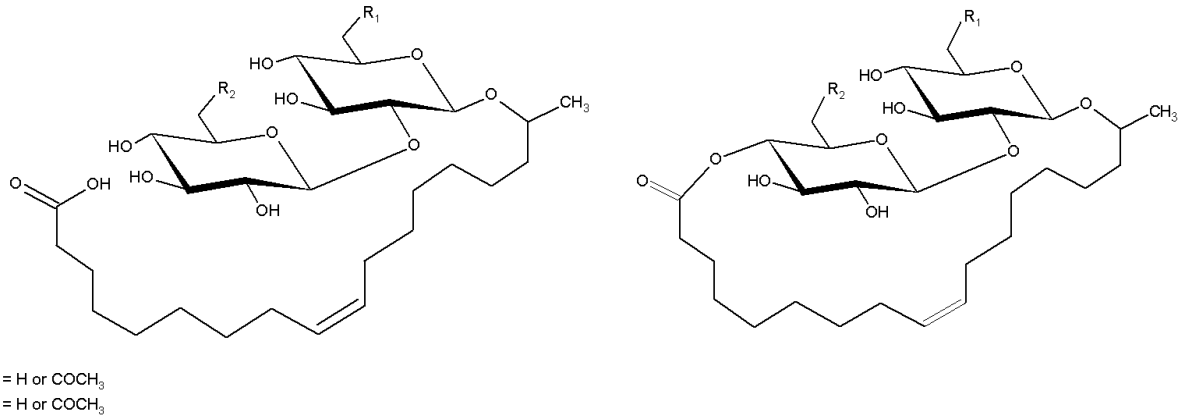
Engineering of species other than *P. aeruginosa* has been reported as well. Using this strategy, the opportunistic pathogenic nature of *P. aeruginosa*, which is a serious burden both for the production and application, can be circumvented. Usually, the *P. aeruginosa rhlAB* operon is expressed in a strain that is resistant towards high biosurfactant concentrations[148]. Using several *Pseudomonas putida* strains like KT2440[148], KT2442[149] and KCTC1067[150], titers ranging from 0.6 g/L[149] to 7 g/L[150] were obtained. The lower production compared to *P. aeruginosa* might be explained by insufficient production of the dTDP-L-rhamnose precursor, a necessary substrate for the rhamnosyltransferases[150].

Other production hosts were evaluated as well, such as *Escherichia coli* strains W2190[151] and BL21[152]. When using the same strategy as for *P. putida*, maximum production titers of 0.18 g/L were obtained. Again, it is suggested that the flux through the dTDP-L-rhamnose synthesis pathway is too low to guarantee higher production titers[151]. By introducing the *rhlAB* operon from *P. aeruginosa* PAO1 in *Burkholderia kururiensis* KP23, a 6-fold increase from 0.78 g/L to 5.67 g/L was detected[153].

I.3. Sophorolipids

I.3.1. Introduction

Sophorolipids (Figure 1.7), another well studied type of glycolipids, are produced by certain yeasts, most of them belonging to the *Starmerella* clade, like *Starmerella bombicola*[154], *Wickerhamiella domercqiae*[155] and *Candida batistae*[18]. In *S. bombicola*, the molecules are constituted out of a sophorose molecule (2-O-β-D-glucopyranosyl-D-glucopyranose) linked to a terminal or subterminal hydroxylated fatty acid with a chain length of 18 or 16 carbon atoms. These can undergo further modifications like acetylation by a specific acetyltransferase at position 6’ and 6” and lactonisation between the free carboxylic end and the C4” of the sophorose unit by an extracellular lactone esterase[39,156,157].



R_1 = H or $COCH_3$
R_2 = H or $COCH_3$

Figure 1.7: General structure of acidic (left) and lactonic (right) sophorolipids produced by *S. bombicola*. The fatty acid tail is made from oleic acid. Further acetylations are possible at R_1 and R_2.

Like all biosurfactants, sophorolipids lower the surface tension of water. Depending on the sophorolipids tested, most values average around 35 mN/m after addition to water[12]. The critical micelle concentration varies, depending on the type of sophorolipid and its uniformity, between 35 and 200 mg/L[158]. Interestingly, these molecules tend to maintain their properties in a wide range of salt concentrations, pH and temperatures. Still, at pH values above 7.5, hydrolysis of the ester bonds occurs. Similar to the rhamnolipids, sophorolipids are promising for soil remediation since they efficiently enhance microbial degradation of complex hydrocarbons[159,160] while being easily biodegradable themselves[161]. Other more specialized applications exist in cosmetics[162], medicine[163], pharmacy[164,165] and nano-industry[166–169]. Besides acting as a final product, sophorolipids can be used as a substrate as well, for the production of hydroxy fatty acids[16], glucolipids[16], sophorose[170] and novel chemically derived surfactants[171]. Sophorolipids are already used in several products like dish-washer and laundry products from companies like Ecover and Saraya. *S. bombicola*, known for the efficient production of sophorolipids up to 400 g/L[154], is the best studied producer so far.

I.3.2. Biosynthetic pathway

The production of sophorolipids is associated with the stationary phase and expression of the necessary enzymes can be detected when the yeast is transitioning from the exponential to the stationary phase[172]. The real trigger is still undefined, but several options have been proposed such as low C/N-ratio[156], phosphate limitation[173] and a low pH[174]. As is typical for secondary metabolites in fungi and yeast, in S. *bombicola* all the genes necessary for sophorolipid production, except the lactone esterase, are located inside a gene cluster situated on chromosome 2.

The first step in the biosynthesis of sophorolipids (Figure 1.8) by *S. bombicola* is the terminal or subterminal hydroxylation of a fatty acid by the cytochrome P450 monooxygenase enzyme CYP52M1. This class of enzymes is further discussed in I.3.3. Strains knocked-out in this P450 are incapable of producing sophorolipids[39] confirming its role. Next at play is UGTA1 that glucosylates the hydroxy fatty acid[175]. The main glucose donor for UGTA1 is UDP-glucose though small residual activity can be detected for UDP-galactose and UDP-glucuronic acid. When looking at hydroxy fatty acid preference, the highest activity is seen with 17-hydroxy-oleic acid. Terminally hydroxylated palmitic acid shows only 20 % of the activity compared to 17-hydroxy-oleic acid[176]. Sequence analysis of UGTA1 revealed that it belongs to the GT-B

type superfamily of glycosyltransferases. The second glucosyltransferase UGTB1 adds a second glucose moiety to the glucolipid produced by the first transferase. It shows the same preference for the glucose donor as UGTA1, but is more promiscuous when it comes to the glucolipid being glucosylated. When using alkyl glucosides as acceptor, still 83 % activity was detected for decylglucoside compared to the natural 17-O-glucopyranosyloctadecenoic acid substrate. It has become evident that the UDP-glucose dependent soluble glucosyltransferases UGTA1 and UGTB1 add the two glucose units in a stepwise manner to form a non-acetylated acidic sophorolipid.

The acetyltransferase is another soluble enzyme from the pathway. As a member of the maltose-*O*-acetyltransferase (MAT) and galactoside-*O*-acetyltransferase (GAT) subfamilies, it transfers an acetyl group from acetyl-CoA to the 6th carbon atom of the sugar moiety[177]. Transport of the sophorolipids to the extracellular environment is attributed to a specific transporter. The protein contains twelve transmembrane helices and two nucleotide binding sites[39] necessary for ATP-binding. Knocking out the transporter results in production levels lower than 10% of those of the wild type. These sophorolipids might be exported by another transporter-like protein or by diffusion through the plasma membrane of the yeast. It has to be noted that knocking-out the acetyltransferase also resulted in lowered production. It might be that the main substrate for the transporter are di-acetylated sophorolipids, the most abundant type found in the natural culture broth.

The final step in the production is the optional extracellular lactonisation of the sophorolipids by condensing the carboxyl group of the fatty acid to the 4” position of the sophorose. As mentioned before, the gene coding for the lactone esterase is not situated inside the sophorolipid cluster, but approximately 2.5 megabases upstream of the cluster on chromosome 2. The occurrence of lactonic sophorolipids pointed however towards the existence of such enzyme. Analysis of the exoproteome gathered in various growth phases identified four proteins showing lipid binding motifs[157]. One of them showed high similarities towards the *Candida antarctica* lipase A, a lipase with high specificity towards longer fatty acids and alcohols[178]. Further studies of this enzyme of *S. bombicola* confirmed its function as the missing sophorolipid lactone esterase. It has been hypothesized that the regulation of the lactone esterase is influenced by citrate concentrations[174].

Besides product titers, small variations in congener ratios exist among the different species producing sophorolipids. While in *S. bombicola* and *Candida apicola* the most abundant forms are the lactonic sophorolipids, *Candida batistae* and *Candida NRRL Y-27208* mainly produce free acidic variants[18,179]. Another difference is the position of the hydroxyl group on the fatty acid. Again a clear distinction can be made between the different yeast species. *S. bombicola* and *C. apicola* mainly incorporate ω-1 hydroxy fatty acids while *C. batistae* and *Candida NRRL Y-27208* prefer the terminal hydroxylated ones. Variations in acetylation pattern can also occur. *S. bombicola*, *C. batistae* and *Candida NRRL Y-27208* mainly produce di-acetylated lactonic sophorolipids in comparison to *C. apicola*, which produces non-, mono- and di-acetylated lactonic sophorolipids. Interesting outsider is the basidiomycete *Rhodotorula bogoriensis*[180]. The molecules produced by this yeast are composed of a 22 carbon atom fatty acid tail with the sophorose group connected by a O-glycosidic to the 13th carbon atom[181]. Minor amounts of C24 sophorolipids can be found as well. Using hydrophobic substrates like meadowfoam or rapeseed oil, higher productivities can be achieved[17].

I.3.3. P450 enzymes

As described above, the first step of the sophorolipid biosynthesis pathway is performed by a cytochrome P450 monooxygenase enzyme. Cytochrome P450 enzymes (P450s) have been studied for several decades because of their interesting biocatalytic properties like the capability to perform oxygenations, nitrations, reductions or even non-redox related reactions[182]. First described in the sixties as the carbon monoxide-binding pigment with a characteristic absorption at 450nm[183–185] when reduced with carbon monoxide, they are now one of the biggest enzyme families found in all kingdoms of nature[186]. They are arranged in families and subfamilies based on sequence homology[187,188]. Most P450s rely on a reaction partner for donating the electrons to perform their catalytic reaction. Depending on the number and type of partners, up to ten different classes can be distinguished.

In chapter 3, P450 enzymes are evaluated utilizing their interesting biocatalytic properties for the production of biosurfactants. These enzymes and their different classes are discussed in more detail below.

I.3.3.1. Structure and reaction mechanism

Though P450s differ in their amino acid sequence, reaction partners used or type of reaction catalysed, most P450s have a common structural organization (Figure 1.9). In general, it can be stated that P450s are composed out of thirteen α-helices (A, B, B', C-L) and five β-sheets (β_1 – β_5)[189]. As expected, a strong conservation can be seen around the central heme group, which includes α-helices E, I, J, K, L, B[190] and the β-sheets. Small variations between the core helices D, E, I, L[191] and the other helices (A, B, B', F, G, H and a part of K) that are less conserved result in the wide variety of substrates that can be accommodated. In general, six substrate recognition/binding sequences can be detected lining the active site[192].

Binding of the heme group happens at a cysteine residue accommodated in a region called the Cys-pocket on the distal end of the helix L. It is locked inside a highly conserved consensus sequence FXXXXGXRXCXG. During the beginning of the catalytic reaction, the electrons from this cysteine residue are donated and believed to significantly contribute to the heterolytic cleavage of molecular oxygen. Other residues may interact with the heme group as well. Mutating these or similar residues can result in altered reduction potential[193–195] or altered stability of the heme group[196–198]. On Helix I, a conserved motif with sequence (A/G)GX(E/D)T is found containing a highly conserved threonine residue, believed to be involved in catalysis[199–201]. Another essential motif is on the C-terminus of the K-helix. This EXXR-motif is well conserved and plays a role in formation of the tertiary structure of the P450 and in heme insertion.

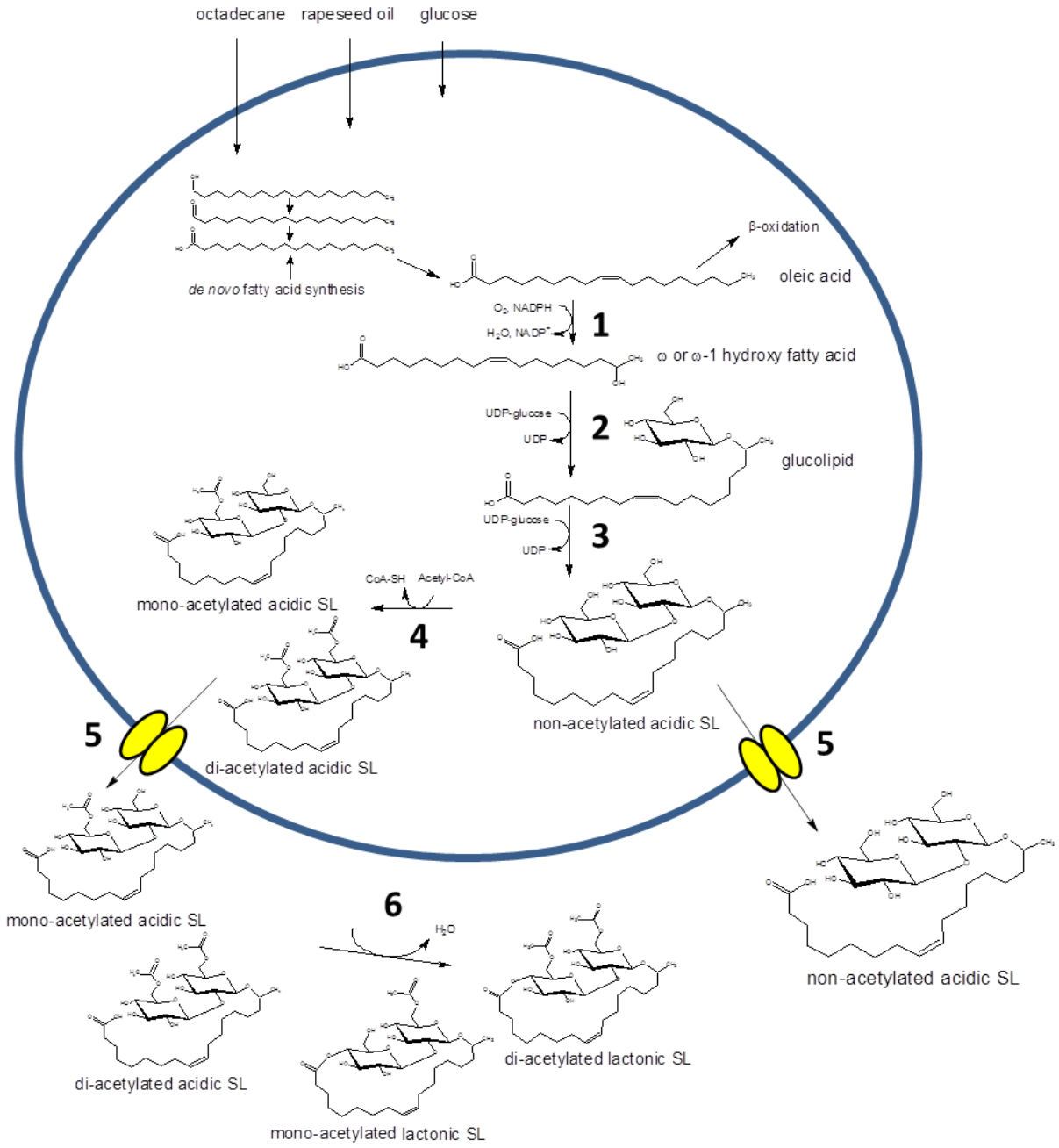


Figure 1.8: Biosynthetic pathway of the sophorolipids produced by *S. bombicola*. (1): Cytochrome P450 monooxygenase CYP52M1, (2): Glucosyltransferase UGTA1, (3): Glucosyltransferase UGTB1, (4): Acetyltransferase AT, (5): Multi-drug resistant ABC transporter MDR, (6): Lactone esterase LE

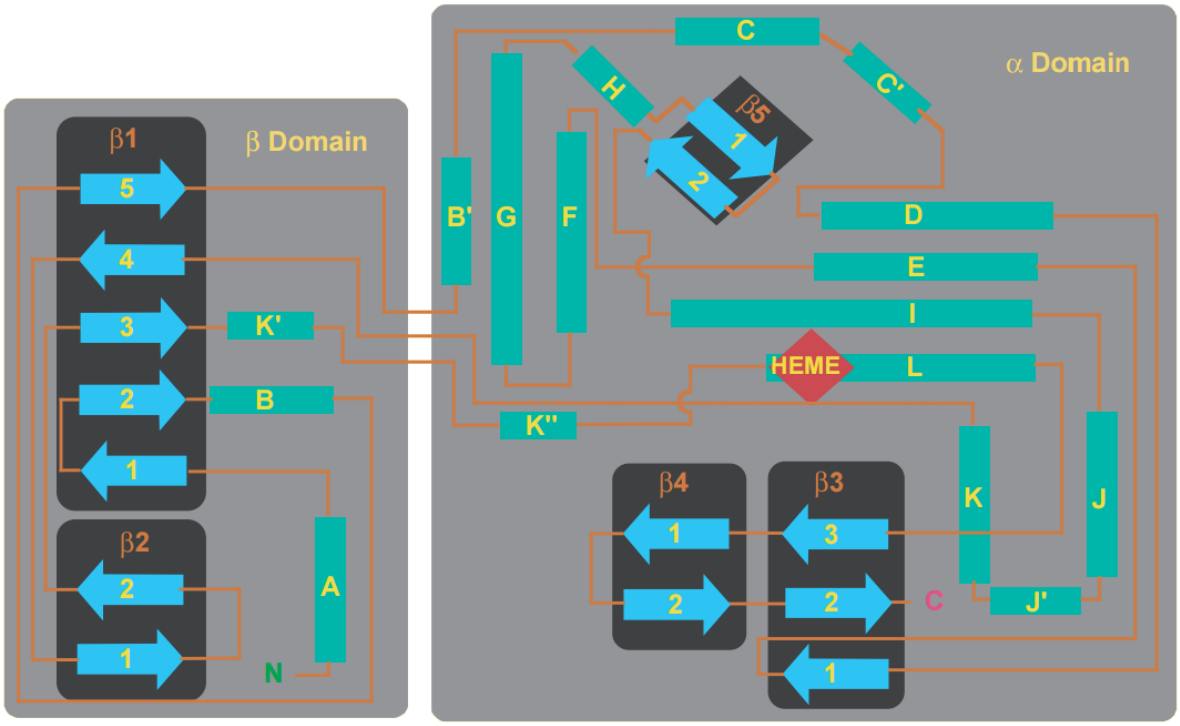


Figure 1.9: General topology of a P450 enzyme. Helices are depicted by the green boxes carrying a letter. Dark coloured shapes with blue arrows are β-sheets[202].

During its catalytic cycle (Figure 1.10), the heme iron group of the P450 can occur in eight distinct states. At rest **(1)** the ferric enzyme is in its low-spin state with a distal water ligand. When a substrate enters the active site **(2)** this will displace the water molecule and induce a shift in the d-orbitals of the iron atom shifting it to a high spin state[203,204]. This high spin state benefits in most cases the reduction of the ferric Fe^{3+} enzyme to the ferrous Fe^{2+} state by **(3)** electrons donated from NAD(P)H or another suitable donor. The next step **(4)** is binding of molecular oxygen resulting in an ferric superoxo complex[205,206]. Following a second reduction and first protonation **(5)**, respectively a peroxo-ferric and a hydroperoxo-ferric intermediate state are obtained[207]. Following the second protonation **(6)**, the molecular oxygen is heterolytically cleaved and a water molecule is detached and a ferryl oxo porphyrin radical cation remains[208]. This radical withdraws **(7)** a hydrogen atom from the substrate inducing the rebound of the oxygen to oxidize the substrate. Finally **(8)**, the product is released and the distal water ligand returns to stabilize the heme iron. A more in depth review was written by Ilia Denisov[209].

During the cycle, several abortive side reactions are possible resulting in a loss of electrons and generating radicals. These reactions are known as shunts. Three of them are reported. The autoxidation shunt releases a superoxide anion from the ferric superoxo intermediate, the peroxide shunt creates hydrogen peroxide from the hydroperoxo-ferric intermediate and the oxidase shunt produces a water molecule using an additional pair of electrons and protons[210,211].

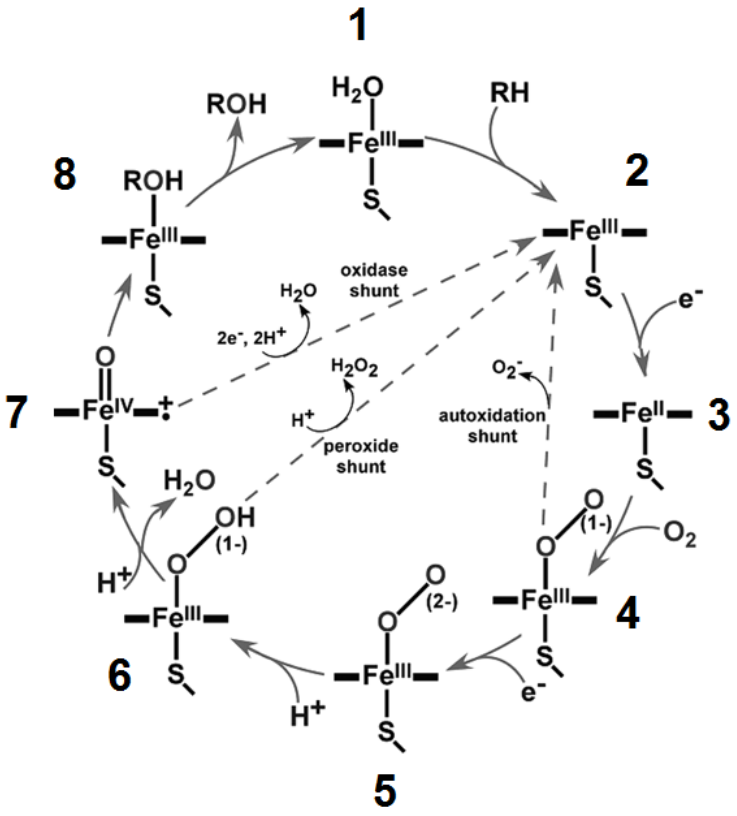


Figure 1.10: Catalytic cycle showing all the different intermediate compounds of the heme iron as well as all the different possible abortive side reactions[209].

I.3.3.2. Class I

The class I P450 systems are characterized by being composed out of three reaction partners (Figure 1.11). Even though they can be found in bacteria and mitochondria, this configuration is conserved. The first enzyme in the cascade is a FAD-containing reductase, which accepts the electrons from NAD(P)H. They are then transferred to the ferredoxin, which finally donates them to the P450. One small difference is that in bacteria all proteins are cytosolic while in the mitochondria both the reductase and the P450 are bound to the inner membrane[212,213].

The best studied bacterial example is CYP101A1 ($P450_{cam}$) from *P. putida* catalysing the 5-exo-hydroxylation of camphor to 5-hydroxycamphor. It utilizes putidaredoxin and putidadoxin to transfer electrons from NADH[214]. Putidaredoxin is an iron sulfur protein belonging to the [2Fe-2S] group[215]. During each turnover of the electrons, two binding events have to take place. These events are induced by conformational changes resulting from substrate and oxygen binding in $P450_{cam}$[214–216]. Other types of ferredoxins can be used as well like [3Fe-4S], [4Fe-4S][217] and combinations[218].

Eukaryotic systems have been found in mammals, insects and nematodes, but not in plants. Examples of human class I P450s are CYP11A1, CYP11B2 and CYP27B1[219]. These enzymes are active in steroid hormone or vitamin D production. Up to date, only one type of ferredoxin has been reported, namely the [2Fe-2S] type. Interesting to mention is that interchanging ferredoxins or reductases can lead to suboptimal results. Switching the ferredoxins from CYP101A1 and CYP11A1 only results in a partial electron transfer[214], while CYP2B4 is as efficient with the putidaredoxine as with its own reductase[220].

I.3.3.3. Class II

Class II enzyme systems are defined by having one or several electron transfer enzymes, either a cytochrome P450 reductase (CPR) or a cytochrome b5 reductase, and one P450, all attached to the membrane of the endoplasmatic reticulum (hence microsomal) (Figure 1.11). A noteworthy exception to this is CYP105A3 from *Streptomyces carbophilus,* which is completely soluble, but uses a separate soluble reductase containing the same domains as the eukaryotic systems[221]. Up to date it is the only bacterial class II system reported. The CPR is an essential electron transfer enzyme necessary in the microsomal I and II families of class II systems. While P450s of family I only interact with the CPR, the microsomal II family uses cytochrome b5 as an additional electron transfer partner and can use either a CPR or a cytochrome b5 reductase as an electron donor. Family III doesn't use a CPR at all, but only a specific cytochrome b5 reductase[222–224]. In total three different types of P450 interactions can be distinguished for cytochrome b5. The first one is transferring both electrons needed for the P450 from cytochrome b5 reductase and the cofactor NADH[225,226]. The second interaction is the delivery of only the second electron to the P450. The third interaction is noncatalytic. Several P450s show enhanced properties due to allosteric stimulation of cytochrome b5[227]. In the case of CYP17A1, cytochrome b5 even has an influence on the type of reaction the P450 catalyses. When no cytochrome b5 is present, CYP17A1 will perform as a hydroxylase of pregnenolone. The product from this reaction is a precursor for the glucocorticoids. When cytochrome b5 is present, CYP17A1 will also display 17,20-lyase activity, generating dehydroepiandrosterone, a precursor for the sex steroids[228]. Due to the high identity with bacterial flavodoxines and FAD-containing proteins, it has been suggested that the CPR is the result of a gene fusion[229].

The structure of the CPR consists out of five distinct domains. These are an N-terminal membrane anchor, an NADPH binding domain, a flavine adenine dinucleotide (FAD) binding domain, a flavine mononucleotide (FMN) binding domain and a linker domain in between the FAD and FMN domains. Binding of NADPH is initiated by electrostatic interactions between the positively charged binding site residues and the negatively charged phosphor-AMP from NADPH[230]. Next, the electrons are sequentially transferred to the FAD to form $FADH_2$. Without interaction of the amino acid residues of the CPR, the electrons are then transferred to the FMN, creating an $FMNH_2$. During the electron transfer, $FADH_2$ can be completely oxidized while FMN_2 forms a stable semiquinone intermediate. CPRs will always shuffle between a total of one to three electrons bound to their flavine domains[231–235]. Control over the amount of reducing elements inside the CPR is done by a second non-donating NADPH-binding site[236]. During the reduction of the CPR and its domains, the conformation of the enzyme changes from a compact to a more open form. The FMN-binding domain remains close to the membrane while the FAD-binding domain rises. This leads to a more exposed FMN-binding domain capable of interacting with the P450[237].

Well studied examples of the Class II system are the vertebrate families CYP1, CYP2 and CYP4. All of them are involved in fatty acid hydroxylation or epoxidation. Human CYP1A1 and CYP1B1 both act on arachidonic acid. The former performs mainly hydroxylations at the subterminal ω-1 position and epoxidations between carbon atoms 14 and 15[238]. The latter performs a wide selection of terminal and subterminal hydroxylations and epoxidations on arachidonic acid as well as several other substrates like 17β-estradiol[239,240]. The CYP2-family seems to show high affinity for fatty acids ranging from eight to 20 carbon atoms. CYP2E1

accepts mainly lauric, myristic, palmitic and stearic acids and hydroxylation is very strict towards the ω-1 position[241,242]. This high product specificity can be broadened by targeted mutations to include hydroxylations from ω-1 to ω-7 as well[243].

Another well studied group is the CYP52 family found in fungi and yeast. These P450s of microsomal family I act on alkanes, fatty acids and their derivatives. In *Candida tropicalis* seven CYP52 P450s have been identified[244–246]. Not all enzymes respond the same when alkanes, alcohols or fatty acids are present in the culture medium, but six out of seven have an aliphatic inducer and are present when glucose is available. All of these P450s might play a specific role to convert alkanes and fatty acids to α,ω-dicarboxylic acids. Selective engineering of the genes encoding these enzymes proved to be an interesting strategy for increasing production of these fatty acid derivatives[247]. In *Candida maltosa* up to nine different P450s of the CYP52 family are found and were partially characterised[248–250]. They are induced by alkanes and fatty acids, explaining the alkane utilization as source for production of cellular lipids and dicarboxylic acids[251]. Also worth mentioning are *Candida apicola*[252] and *Starmerella bombicola*[253,254]. Both yeasts produce sophorolipids which harbour a palmitic, stearic or oleic acid tail. In *S. bombicola*, CYP52M1 is responsible for the hydroxylation of these fatty acids. As the same type of fatty acids tails can be identified in the *C. apicola* mix[255], it is not unlikely that a CYP52M1 homologue is present in *C. apicola* performing the same reactions in the sophorolipid biosynthesis. Comparing the *C. apicola* CYP52E1 and CYP52E2 enzymes with *S. bombicola* CYP52E3, high identities can be found as well[254].

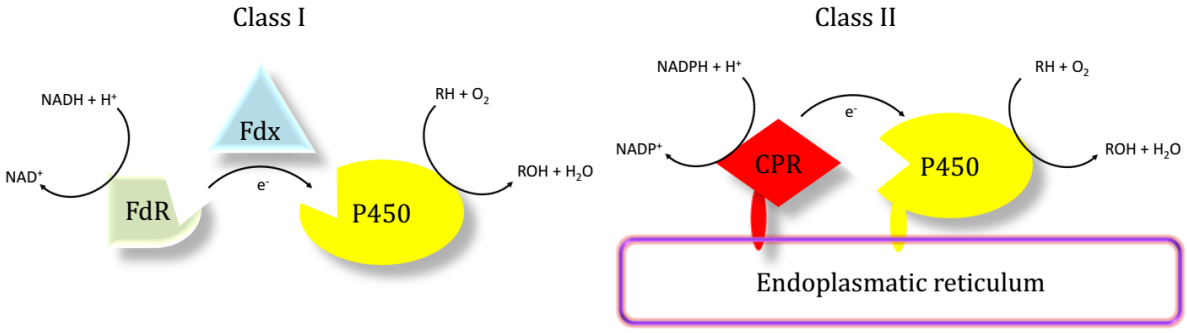


Figure 1.11: Prokaryotic class I and eukaryotic class II of the microsomal I family.

I.3.3.4. Class III

The class III systems highly resemble the ones from class I except that the second reaction partner is not an iron-sulfur protein, but a flavodoxin instead (Figure 1.12). In this way, it shares all the domains from class II like FAD and FMN, but in separate enzymes. A classic example of the class III enzymes is CYP176A1, a P450 from *Citrobacter braaki*, enabling this microorganism to grow on cineole[256,257].

I.3.3.5. Class IV

Class IV systems are best characterized in thermophilic species like the archaea *Sulfolobus solfataricus*. They again highly resemble the class I bacterial system except that the reducing elements are not donated by NAD(P)H, but from a 2-oxo acid like pyruvate (Figure 1.12) by a 2-oxoacid-ferredoxin oxidoreductase. CYP119 is the first P450 discovered to utilize such a non-NADPH-consuming flavoprotein[258–260]. It has to be noted that in the case of CYP119, the exact electron donor has not been discovered yet. Several other P450 have been discovered in thermophilic species, but it has to be said that more information about redox-partners and electron donors is scars.

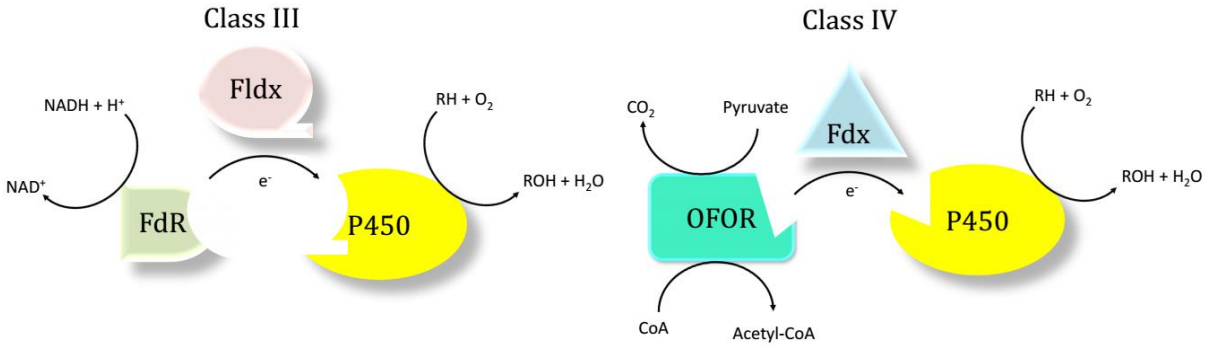


Figure 1.12: Class III and the oxo-acid utilizing class IV. OFOR = 2-oxoacid-ferredoxin oxidoreductase

I.3.3.6. Class V

Class V is characterized by the fusion of the ferredoxin reaction partner to the P450 (Figure 1.13). Only one member of this family has been characterized, but other candidates have been identified in other species. The enzyme MCCYP51FX from *Methylococcus capsulatus* is a 14α-sterol demethylase[261]. Structural analysis revealed that the ferredoxin domain is linked to the P450 by an alanine linker to the C-terminus of the enzyme. A homologue has been discovered in *Mycobacterium tuberculosis* though this P450 is not a fusion product[262,263]. However, a [3Fe-4S] ferredoxin gene is found next to the P450. Further research on the properties of the *M. capsulatus* system in comparison with the class I systems is ongoing. Potentially, by fusing both reaction partners, a more efficient electron transfer could be obtained.

I.3.3.7. Class VI

Another class consisting out of fusion enzymes is class VI (Figure 1.13). Here, a flavodoxin domain is fused to the P450, strongly resembling the class III system. The first P450 of this class to be discovered was XplA, an enzyme from *Rhodococcus rhodochrous*[264]. Special interest goes to this enzyme since it can convert hexahydro-1,3,5-trinitro-1,3,5-triazine, an explosive compound with widespread contamination[265]. Using *Rhodococcus* species or organisms heterologously expressing XplA might be interesting for bioremediation purposes[264,266]. The reductase of this system is XplB[266]. Genetic analysis of other RDX-degrading strains suggests that this property is the result of lateral gene transfers[267] of the P450s involved.

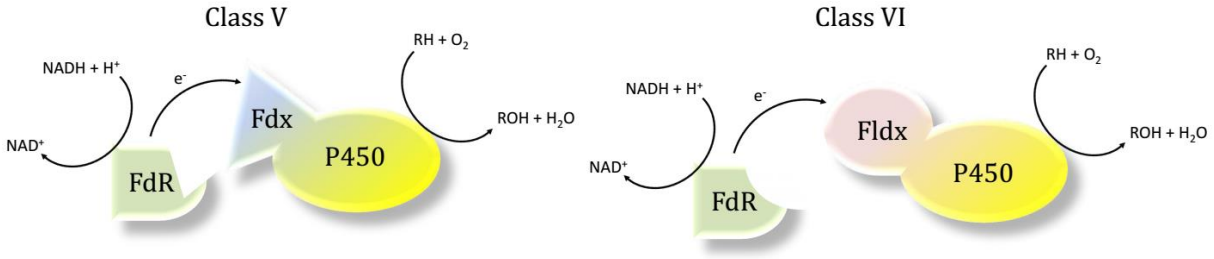


Figure 1.13: Fusion P450s of class V and class VI

I.3.3.8. Class VII

The class VII systems are fusion proteins containing a unique reductase domain with high homology to phthalate dioxygenase reductases of type I[268], commonly referred to as PFOR[269]. The reductase is fused to the C-terminus of the P450 (Figure 1.14). The reductase domain is constituted out of three subdomains, namely an FMN-binding domain, an NADH-binding domain and a [2Fe-2S] ferredoxin domain[270]. Indeed, in P450 PFOR enzymes, the electrons are donated to the FMN instead to an FAD like in most P450 classes. After the hydride transfer, NAD^+ is released from the NADH-binding domain. Small structural changes induced by the release control the transfer of the electrons to the [2Fe-2S]-domain. P450 Rhf, also known as CYP116B2, from *Rhodococcus* sp. NCIMB9784 was the first enzyme of this class to be discovered[270,271]. Other members have been identified as well in several *Burkholderia* species.[272] Sequence analysis of the P450 domain revealed high similarities with the class I CYP116A1 from *Rhodococcus erythropolis* NI86/21, an enzyme involved in the degradation of thiocarbamate herbicides[273]. Recently, CYP116B1 from *Cupriavidus metallidurans* was found to be a class VII P450 as well, with similar thiocarbamate degrading properties[274].

I.3.3.9. Class VIII

Class VIII enzymes are again fusion proteins. The reductase domain however has high similarity to the eukaryotic diflavin reductases (for example the CPR enzyme from class II) (Figure 1.14). The first discovered and model enzyme for this class is fatty acid hydroxylase CYP102A1 (BM3) from *Bacillus megaterium* ATTC 13368[275]. CYP102A1 harbours all the necessary domains to perform its catalytic reactions. The reductase domain contains a FAD and FMN-binding domain as well as a NADPH-binding pocket. The electrons are transferred from the NADPH to FAD and sequentially FMN almost the same way as in the eukaryotic cytochrome P450 reductase. In the eukaryotic reductase, the electrons come from the double reduced FMN group creating a FMNH semiquinone, while in contrast, the electrons in BM3 are donated from the FMNH semiquinone[276]. Studies have shown that electron transfer through the reductase domain, dioxygen binding and substrate binding are all not rate-limiting. Depending on the substrate bound, electron transfer rate can vary[277]. This control is believed to be due to conformational changes in the flavoprotein induced by the fatty acids. It is believed that the transfer of electrons from FMNH semiquinone towards the heme group is rate-limiting, together with other factors like product release[278].

Other compounds have been studied besides the natural substrates of CYP102A1, other compounds have been studied as well with promising results for biocatalytic applications[279–281]. The natural enzyme can deliver high region- and stereospecificity depending on the used substrate and conditions[282–284]. The high catalytic activity of BM3 combined with enlarged

substrate promiscuity under varying conditions also promises interesting opportunities for engineering the enzyme to produce high value molecules like novel potential drugs or for improving the understanding of drug intermediate toxicity. A first example of engineering BM3 is the substitution of F87V for producing high purity (14*S*,15*R*) epoxyeicosatrienoic acid.

In the wild type enzyme, the Phe[87] residue acts as a steric gate, controlling substrate specificity due to its bulky side chain protruding in the active site channel. By switching between two states, either the final methyl group or its 14,15-olefin are positioned near the heme group. Replacement by valine effectively expands the active site allowing better positioning of the 14,15-olefin[285]. It has to be noted that mutating Phe[87] has a distinct effect on coupling efficiencies[285,286] of other non-natural substrates of BM3. Another example is the production of piceatannol (3,5,3′,4′-tetrahydroxystilben), a hydroxylation product of resveratrol (3,4′,5-trihydroxystilbene). It shows interesting antitumor properties in several cell lines since it is known as a protein kinase inhibitor working on several cellular targets[287–289]. It can be produced in the liver by CYP1A2, but a side product is formed. Wild type and mutant variants of CYP102A1 all produce a single product, piceannatol. Activities higher than for the wild type BM3 enzyme and CYP1A2 were found for several mutated BM3 enzymes[290].

All enzymes of the CYP102A family are known as fatty acid hydroxylases. All of them perform subterminal ω-1, ω-2 and ω-3 hydroxylations of medium and long chain fatty acids like lauric, myristic and palmitic acid[291–293]. In *Fusarium oxysporum*, a similar self-sufficient P450 called CYP505A1 or $P450_{foxy}$ has been found[294,295]. In contrast to CYP102A1, $P450_{foxy}$ appears to be membrane bound or at least associated.

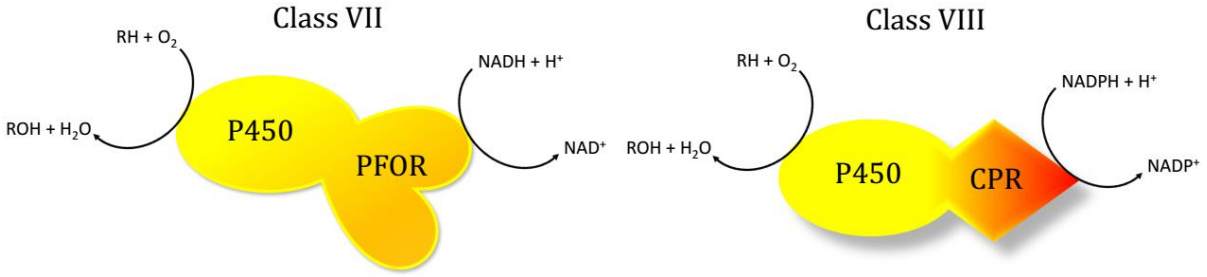


Figure 1.14: Self-sufficient P450s of class VII and class VIII

I.3.3.10. Class IX

The class IX system is occupied by a single family of P450 enzymes, CYP55A. It is the only P450 family known in eukaryotes that is completely soluble (Figure 1.15). All enzymes of this class are known as nitric oxide reductases ($P450_{nor}$). They can be found in both the cytosol and the mitochondria. They are necessary for clearing nitrogen oxide when dioxygen is limited by reducing it to nitrous oxide[296,297], thereby preventing nitrogen oxide inhibition of the mitochondria. The first enzyme discovered of this class was $P450_{nor}$ or CYP55A1 from *F. oxysporum*[298]. Upon binding of an electron donor, in the case of CYP55A1 the pyridine nucleotide NADH, a global conformational change enables efficient transfer of two electrons and one proton towards the heme group by a channel connecting the NADH binding site and the heme group[299]. Further analysis of fungi and yeast like *Cylindrocarpon tonkinense*[300], *Aspergillus oryzae*[301] or *Trichosporon cutaneum*[302] provided more isozymes of cytochrome $P450_{nor}$. Depending on the species, either NADH, NADPH or both can be used by their P450.

I.3.3.11. Class X

The class X enzymes can be called atypical since they don't depend on an external electron source for their reactions. Instead they use an intramolecular transfer system and don't require molecular oxygen. Their substrates are acyl hydroperoxides and by rearranging the bonds new products are generated (Figure 1.15). All members originate from the CYP74 family and are only detected in plant chloroplasts[303,304], fungi[305] and some bacteria[306]. The class members can be divided in three subfamilies, allene oxide synthase (AOS) CYP74A, hydroperoxide lyase (HPL) CYP74B and CYP74C and finally divinyl ether synthase (DES) CYP74D.

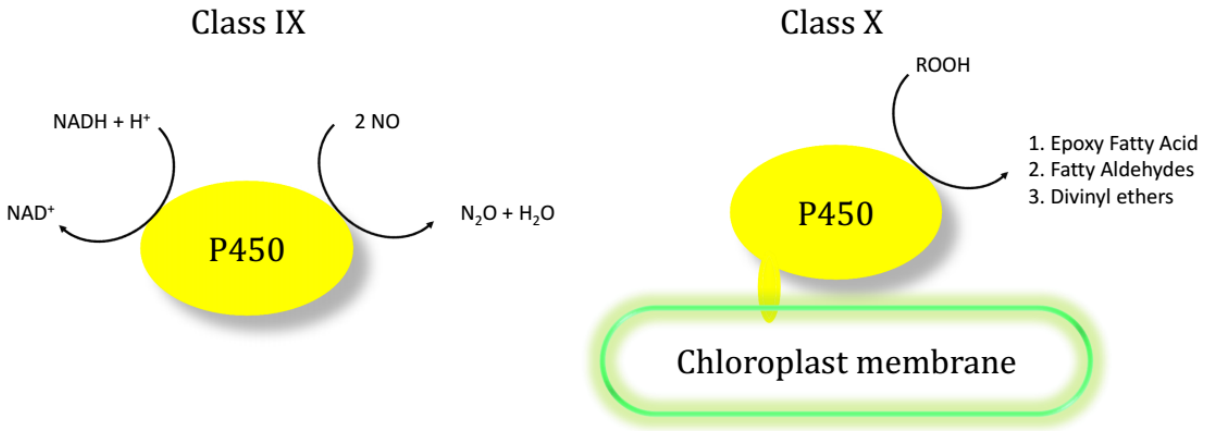


Figure 1.15: Class IX and class X P450 systems.

As mentioned, the catalytic mechanism of these P450s differs from traditional ones. An overview of all three reactions can be found in Figure 1.16. For AOS, the mechanism was elucidated using 13(*S*)-hydroperoxide linolenic acid (HPOT)[307]. The peroxide bond located at C13 is evenly broken by using a transition phase leaving an oxygen radical on the substrate and binding the leaving hydroxyl group in the heme iron. In a next step, the oxygen radical interacts with the nearby C12 methyl-group creating the epoxide. The radical is stabilized in this stage since the conjugated double bonds nearby delocalize it. In the final step, a proton from C12 is extracted and transferred to the heme group of the P450. This reacts with the hydroxyl group and is lost as water. The C12 methyl group rearranges to finally form the allene oxide 13-hydroperoxy-octadecatrienoic. This product can then be further processed either by enzymatic and non-enzymatic ways. The allene oxide can be used as a precursor for jasmonate formation by conversion to 12-oxo phytodienoic acid, catalysed by allene oxide cyclase (AOC). It has been hypothesized that due to the instable nature of the allene oxide, AOS and AOC are linked[308]. Non-enzymatic reactions include the conversion of the allene oxide in α- or γ-ketols. Though most AOS enzymes are specific towards this substrate, the AOS from barley is capable of accepting 9-hydroperoxides as well while others are only specific for 9-hydroperoxies.

The HPL mechanism is very similar to that of AOS (Figure 1.16) with specificities towards one or several kinds of hydroperoxides. The peroxide group is again evenly split with the hydroxyl group associating with the heme iron and leaving an oxygen radical on the substrate. This radical will transform into the epoxide by interacting with the adjacent methyl group. Next is the cleavage of the C:C bond of the oxirane ring followed by a nucleophilic attack by water on the remaining carbocation. The resulting hemiacetal degrades spontaneously in the corresponding aldehydes and aldoacids[309].

The final DES mechanism (Figure 1.16) corresponds to that of divinyl ether synthase. It again uses an epoxide intermediate which will undergo cleavage at the C:C bond. Next follows a proton abstraction from the nearby methyl group resulting in the ether group being flanked by double bonds.

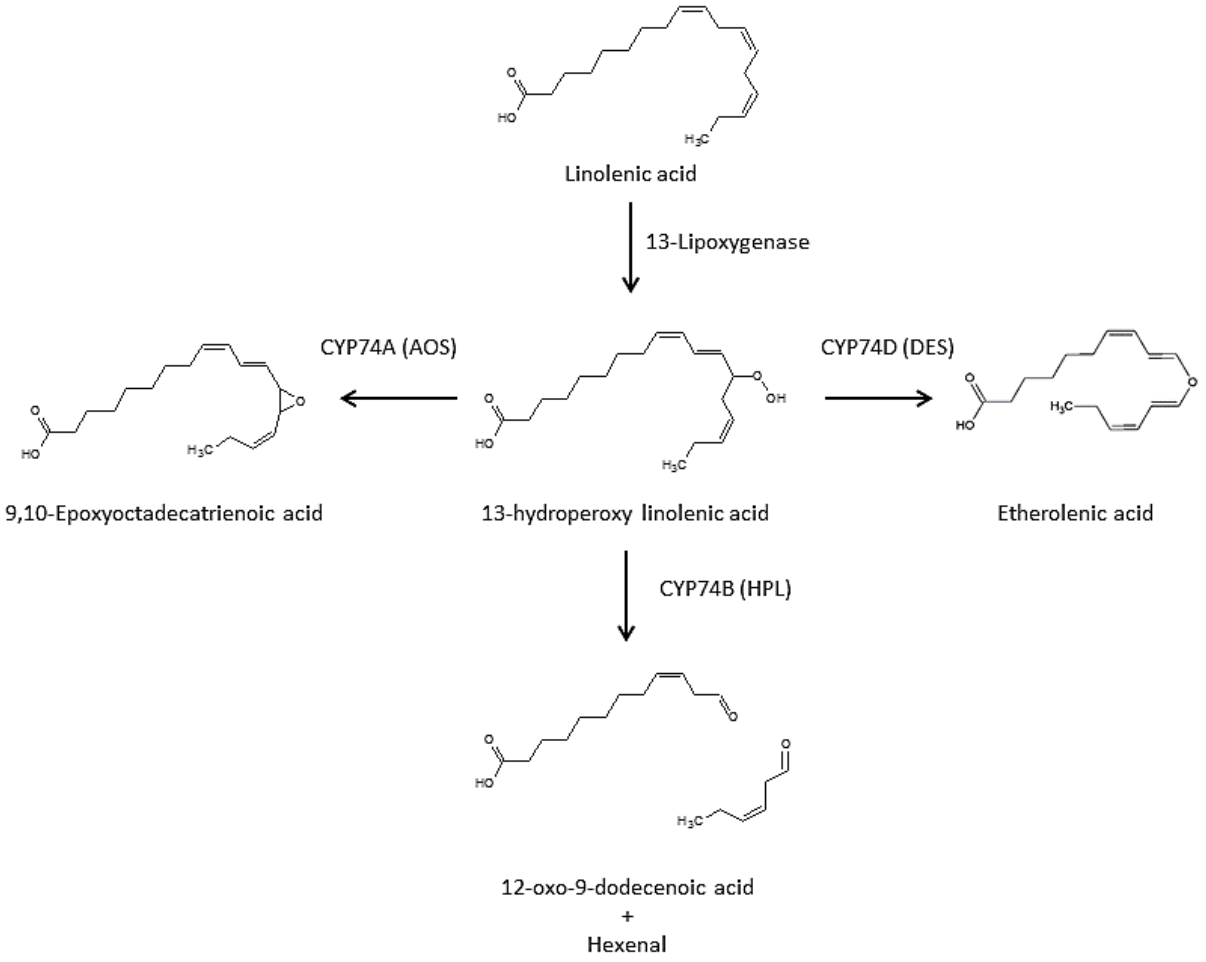


Figure 1.16: Different reactions catalysed by the class X P450 enzymes with 13-hydroperoxy linoleic acid as a substrate. The products can be further converted by enzymatic or spontaneous reactions.

I.3.4. Engineering possibilities

Development of an efficient transformation and selection system made engineering of *S. bombicola* possible[310–312]. For the moment, three different selection markers are used, namely the *ura3* auxotrophic marker and antibiotic resistance genes for hygromycine and nourseothricin. Due to the fact that no plasmid is available, all modifications are done by homologous recombination in the yeast genome.

Several knock-out and overexpression strains were created for the production of modified sophorolipids. One example involves the acetyltransferase, necessary for acetylation of the sophorose group. A knock-out strain of this gene produces purely unacetylated[177] as well as bolaamphiphilic sophorolipids[313]. The latter ones are molecules consisting of two non-acetylated sophorose hydrophilic groups with the newly introduced one coupled to the carboxylic group of the fatty acid (Figure 1.17).

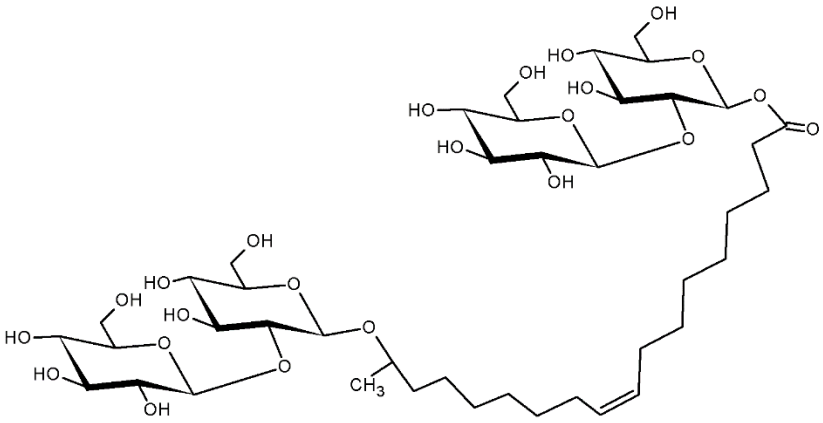


Figure 1.17: Example of a bolaamphiphilic sophorolipid with an oleic acid linker between both sophorose groups.

Another engineering example is that of the lactone esterase, the enzyme catalysing lactonisation of the sophorolipids. Knock-out strains are capable of producing the acidic form in concentrations comparable to the wild-type strains[157]. Overexpression of the lactone esterase results in producing lactonic sophorolipids at enriched ratios compared to the wild type profile.

Production of medium-chain sophorolipids is challenging as the required special substrates such as primary alcohols and diols are readily metabolized in β-oxidation. Knocking out the multifunctional enzyme type 2 (MFE-2), catalysing the second and third step of the β-oxidation, enhances the production of sophorolipids with shorter chain lengths up to 3 times compared to the wild-type strain[314]. Still, the dependence on special substrates instead of vegetable oil greatly affects the production cost. Another example of using the MFE-2 deletion strain is the production of 20-hydroxyeicosatetraenoic acid by *S. bombicola*[315]. Although the natural substrates for CYP52M1 are stearic and especially oleic acid, the double bonds of arachidonic acid effectively shorter the total chain length making it possible for CYP52M1 to hydroxylate the ω and ω-1 carbon atoms. Another way of engineering the hydrophobic tail involves the *aox1* gene which encodes and aldehyde oxidase. Knock-out strains of this gene produce sophorolipids with an alkyl chain instead of the traditional fatty acid tail when grown on primary alcohols[316].

Other successful engineering resulted in the production of biosurfactants differing from the traditional sophorolipid structures. A successful knock-out involves *ugtB1*, the gene responsible for the addition of a second molecule of glucose in the sophorolipid biosynthesis. After a prolonged fermentation, final concentrations of 3.9 g/L were obtained of both acetylated and non-acetylated glucolipids[317,318]. Introduction of the codon optimized *ugt1* gene from *U. maydis* in this knock-out strain was used for the heterologous production of cellobiose lipids in *S. bombicola*[318]. Even though 70 % of the obtained molecules are glucolipids, a diverse population of cellobiose lipids was observed as well. Most of these cellobiose lipids are new-to-nature, consist of a C18 hydroxy fatty acid tail and have no acylations of the cellobiose group. Also no lactonisation is observed. Still, optimization is required because of the glucolipid contamination and the less than optimal product concentrations that often remain below 1 g/L.

A final engineering example involves bioplastics. Several microorganisms like *P. aeruginosa* are known to produce polyhydroxyalkanoates (PHA), a type of bioplastic polymer build up from β-hydroxy fatty acids. Polymerization of the fatty acids is done by poly(3-hydroxyalkanoate)

synthases using acyl-CoA as a substrate. In the past, production of these molecules was proven to be successful in other yeast species like *Sacchoromyces cerevisiae*[319], *Pichia pastoris*[320] and *Yarrowia lipolytica*[321].

Introduction of the codon-optimized *P. resinovorans phaC1* gene coupled to a peroxisomal targeting sequence resulted in production of PHA molecules in *S. bombicola*. Depending on the substrates used, yields up to 2.2 % wt/dwt could be obtained[318]. These engineering examples show that it is possible to obtain very specific types of molecules with *S. bombicola*.

I.4. Cellobiose lipids

I.4.1. Production and regulation

Cellobiose lipids (CBLs) are produced by several fungi like *Ustilago maydis* and *Pseudozyma flocculosa*. They are composed of cellobiose (4-O-β-D-glucopyranosyl-D-glucopyranose) attached to a hydroxylated palmitic acid. Further modifications to this molecule such as acetylations, acylations and hydroxylations are possible[322–325] (Figure 1.18). The enzymes necessary for production of cellobiose lipids in *U. maydis* and *P. flocculosa* are grouped inside a gene cluster[326,327]. When comparing both clusters, it becomes clear that though they are used for the same kind of molecules, their overall organization and size differs.

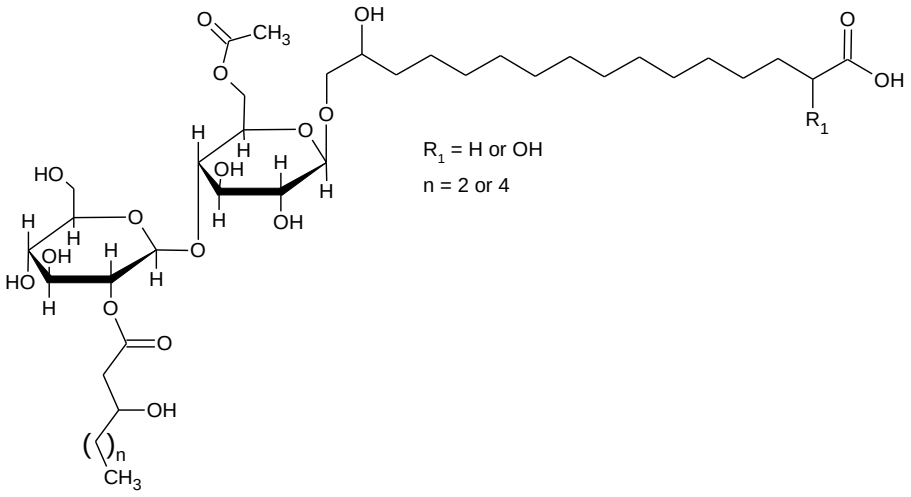


Figure 1.18: General structure of cellobiose lipids produced by *U. maydis*.

The synthesis of CBLs is initiated by CYP1, a cytochrome P450 enzyme necessary for terminal hydroxylation of palmitic acid[328]. A second hydroxyl group is introduced by CYP2 at the ω-1 position, but it was observed that the second hydroxylation is not necessary for further synthesis of the CBLs. The cellobiose moiety is coupled to the 15,16-dihydroxy palmitic acid by UGT1 in *U. maydis* or FGT1 in *P. flocculosa*. The lack of a second glucosyltransferase in the cluster points to the capability of sequential glucosylation by one transferase in both organisms.

Further modifications are carried out by acetyl/acyl transferases and hydroxylases[326]. The acyl transferase couples a β-hydroxy fatty acid with six or eight carbon atoms to the cellobiose group at position 2". These acylation reactions are carried out by UAT1 and FAT1. Prior to this acylation, a β-hydroxylation is carried out by UHD1 and FHD1. Both enzymes belong to the

short-chain dehydrogenase-reductases (SDR) and contain several conserved motifs like TGXXGXX and HxAs for cofactor binding[329]. The short-chain fatty acids which are coupled by the acyltransferases might be specifically produced by FAS2, a part of the fatty acid synthesis complex present in both organisms. Besides the acylations, acetylations by UAT2, FAT2 and Fat3 can be performed respectively on positions 5', 5' and 3". A final modification is performed by Ahd1 in *U. maydis*. Here, an α-hydroxy group is introduced on the palmitic acid chain. In *P. flocculosa*, such an enzyme hasn't been discovered: it might be that the precursors for the pathway are β-hydroxy-palmitic acids originating from the *de novo* synthesis of fatty acids.

In *U. maydis*, the CBL production is initiated under nitrogen starvation. Regulation is governed by Rua1, a transcription factor interacting with the cluster genes and located herein[330]. Overexpression of this regulator promoted biosurfactant production already in the exponential growth phase. In *P. flocullosa* N-limited induction nor a cluster-embedded regulator were reported[24]. It has been hypothesized that flocculosin is produced more as an energy overflow compound when the cells stop growing. This theory is supported by the ability of the fungus to catabolize its own glycolipids[331] in energy deprived situations.

I.4.2. Engineering possibilities

It has to be said that *U. maydis* produces two classes of glycolipids, cellobiose lipids and mannosylerythritol lipids (MELs, see I.5), with titers of total glycolipids up to 30 g/L with varying ratios of MEL's and cellobiose lipids[21]. Deletion of either the mannosyltransferase *emt1* or the cytochrome P450 monooxygenase *cyp1* genes results in total blockage of respectively the MELs or CBLs biosynthesis, allowing recovery of one type of glycolipid.

Other engineering possibilities have been reported besides deletion of either *cyp1* or *emt1*, both involved in respectively cellobiose lipid and mannosylerythritol lipid production. Deletion mutants of the *cyp2*, *ahd1*, *uat1* and *uhd1* genes in *U. maydis* produce altered cellobiose lipids[327]. These genes are respectively responsible for the ω-1 hydroxylation (*cyp2*) and α-hydroxylation (*ahd1*) of the fatty acid tail, the acylation and acetylation of the cellobiose group (*uat1*) and the hydroxylation of the short-chain fatty acid coupled to the cellobiose group (*uhd1*). Unfortunately, no data are available on the cellobiose lipids titers of these knock-out strains.

I.5. Mannosylerythritol lipids

Mannosylerythritol lipids (MELs) are interesting compounds consisting of a 4-O-β-D-mannopyranosyl-meso-erythritol with various acyl- and acetylation patterns, depending on the production organism and substrates used (Figure 1.19). Depending on the acetylation degree, MELs are divided into four groups. MEL A corresponds to di-acetylated variants, MEL B and C refer to mono-acetylated at C6 and C4 respectively, and finally MEL D are non-acetylated molecules[332]. Caution has to be made since some types of MELs differ from these four groups. For example the MELs from *Pseudozyma tsukubaensis* contain a 1-O-β-D-mannopyranosyl-meso-erythritol carbohydrate structure[333]. Though not a MEL, the fungus *Pseudozyma parantarctica* produces a similar molecule with a mannitol instead of erythritol group[334]. The titers of these molecules can easily reach 100 g/L or more[335].

Their properties are equivalent to other glycolipids like rhamnolipids and sophorolipids in terms of lowering surface tension and critical micelle concentration. Special interest goes out to these molecules because of their self-assembling properties[336], potential pharmaceutical applications[337] and their usage in cosmetics[338]. The Japanese companies Toyobo and Kanebo Cosmetics commercialize MEL's.

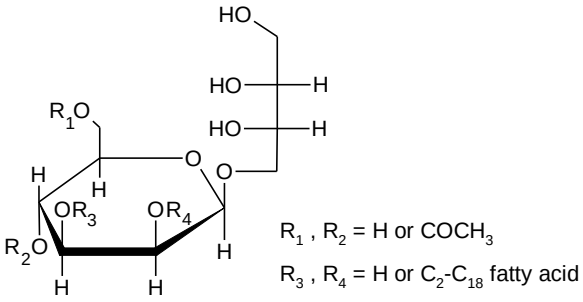


Figure 1.19: Main structure of mannosylerythritol lipids.

Biosynthesis in *U. maydis* starts with EMT1, which couples a mannose, donated by GDP-mannose, to the terminal C4 hydroxyl group of erythritol. In the next step, MAC1 and MAC2 add acyl groups to the mannose C2 and C3. Until now, it is not known which transferase works on which position. Structural analysis of the MELs however points out that C2 is mostly acylated with short chains (C2 – C8) while C3 harbours longer chains (C10- C18). A final modification is done by MAT1, transferring acetyl groups from acetyl-CoA to C4 and C5 of the mannose group[328]. Export of the MELs is governed by MMF1, a major facilitator type of transporter. Deletion of its coding sequence results in complete abolishment of production.

As mentioned in section I.4.2, some engineering is done to optimize production in *U. maydis.* Deletion of *emt1*, the gene encoding the erythritol/mannose transferase, abolished production while overexpression did not lead to higher production. The deletion of MAT1, the acetyltransferase responsible for both acetylation reactions, proved that it is possible to produce non acetylated MEL's without severely influencing the production capacity[328].

I.6. Lipopeptides

Several species belonging to the *Aspergillus*, *Bacillus* and *Pseudomonas* genera are capable of synthesizing lipopeptides. These molecules are characterized by an oligopeptide, either linear or circular, coupled to a fatty acid. The oligopeptide varies both in number and types of amino acids[339] (Figure 1.20). Unusual or non-proteiogenic amino acids are incorporated as well. For example, in surfactin, synthesized by *Bacillus subtilis*, two D-leucine molecules are present[340]. In the case of surfactin, titers of 6.45 g/L are possible[27].

Lipopeptides have drawn attention because of their potential as antimicrobial agents[341]. The minimum inhibitory concentration (MIC) of daptomycin for example towards species like *Enterococcus faecalis* ranges from 2 to 4 mg/L[342]. It has been used successfully in clinical trials for *S. aureus* bacteremia and right-sided endocarditis[343].

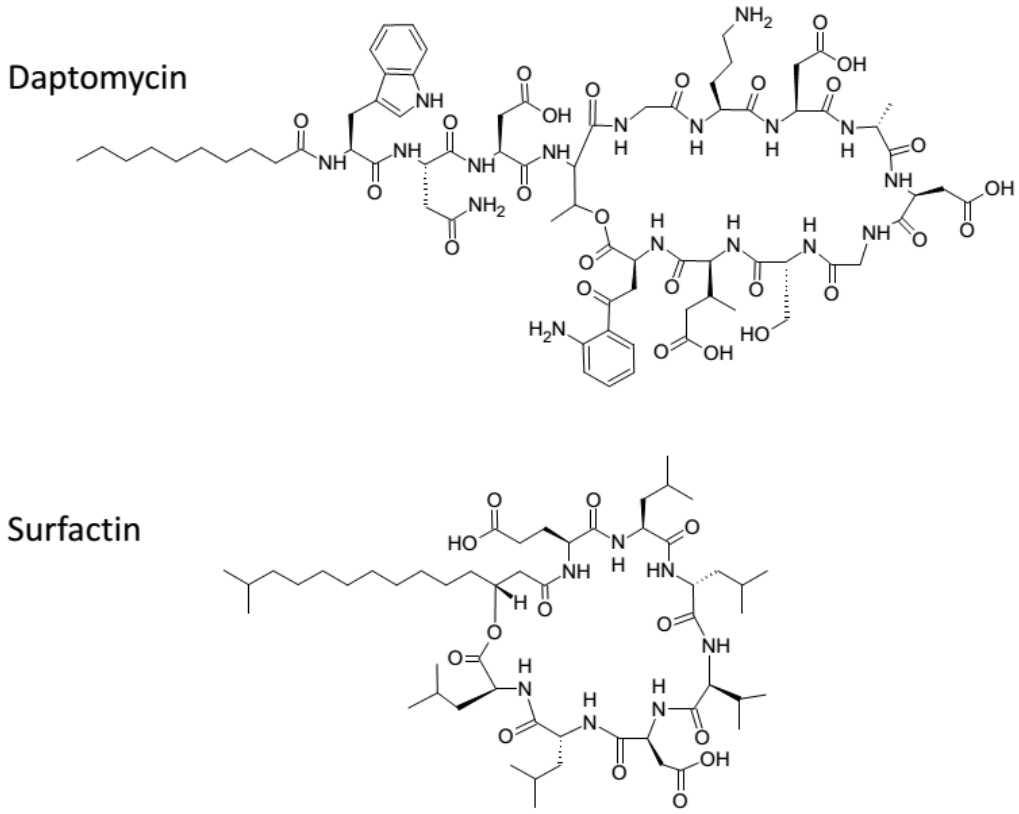


Figure 1.20: Structures of daptomycin and surfactin, two circular lipopeptides.

The hydrophilic oligopeptide structure of lipopeptides is synthesized by a non-ribosomal peptide synthetase (NRPS). This is a single enzyme in many fungi yet a multizyme complex in bacteria. Both types of NRPS can be further divided in modules consisting of several domains (Figure 1.21). Each module is responsible for the incorporation of at least one amino acid and contain the necessary domains for the elongation and modification of the peptidyl chain. Several modules can be grouped together in a subunit, a monomer being part of the multizyme complex coded by a single gene.

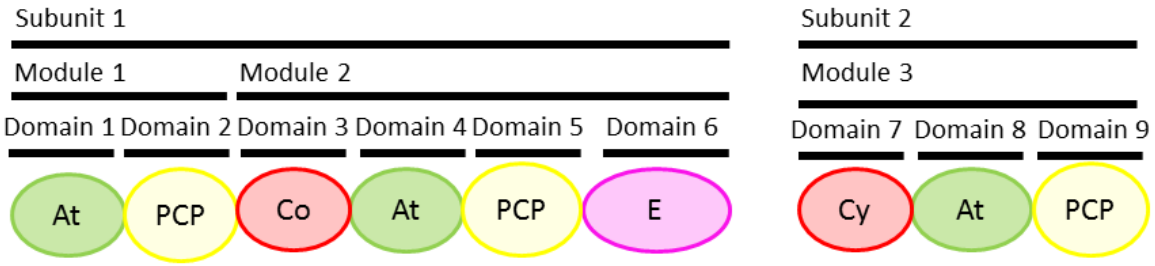


Figure 1.21: Simplified structure of a NRPS containing multiple subunits. Different domains can be present inside a single module. PCP = peptide carrier protein; At = adenylation domain required for amino acid thioesterification; Co = condensation domain required for amide bond formation of the growing peptide chain from the previous module with the amino acid group of the current module; Cy = Condensation and cyclization of serine, threonine or cysteine can be performed by certain condensation modules.; E = epimerization into D-amino acids.

The first module of the NRPS is called an initiation module, the following are extending modules and the final one is a releasing module. Termination can be performed by a thioesterase, which is mostly present at the end of the releasing module (type I) or can form an independent subunit (type II). These thioesterases also cyclize the peptide chain prior to release. Release of a linear fragment is catalysed by a reducing domain resulting in a terminal aldehyde or alcohol function. Further processing of the peptides by glycosylation or acylation is done by separate enzymes, which are often associated with the NRPS and are commonly situated near or inside the operons or gene clusters of the NRPS.

Significant effort has been put into engineering of the hydrophilic oligopeptide structure. In a first engineering strategy subunits of the NRPS are exchanged. In the case of daptomycin, a lipopeptide produced by *Streptomyces roseosporus*, replacing the final subunit DptD by introducing the genes encoding lptD and cdaPS3 from *Streptomyces fradiae* and *Streptomyces coelicolor* results in compounds differing in the last two amino acid of the peptide core[344]. A second option is to exchange or delete modules[345] or domains within these modules[346]. This leads to peptide cores differing in number and types of amino acids. By targeting the genes responsible for modification of the amino acids[347], even greater diversity can be created. Again in the case of daptomycin, a library of molecules differing in MICs against several kinds of microorganisms was derived using such combinatorial biosynthesis[348].

Besides by changing the size and amino acid sequence of the oligopeptide core, structural diversity has also been achieved by “decorating” reactions like halogenation, glycosylation and sulfation. Though these modifications not necessarily enhance antimicrobial activity, they enhance aqueous solubility and hydrolytic stability. Attempts to modify the fatty acid tail have been successful for the lipopeptides from *Streptomyces coelicolor* where a shorter tail was introduced[349], as well as for daptomycin derivative CB-183,315 where a longer, unsaturated branched fatty acid was coupled to the peptide ring[350]. A more comprehensive review on lipopeptide engineering was written by Winn *et al*[351].

A couple of examples exist where engineering of the production strain leads to significant higher production levels. Random mutagenesis strategies resulted in mutant strains with production capacities up to twelve times higher compared to the original wild-type producers (change from 33 mg/L to 391 mg/L)[352]. Genome shuffling was used for enhancing the surfactin production in *Bacillus amyloliquefaciens*. The recombinant strains are capable of producing up to 10 times more compared to the wild-type (maximum surfactin concentration was 350.1 mg/L) due to higher expressions levels of the genes involved[353]. More targeted methods like overexpression of certain genes or ribosome engineering are possible as well. In the case of *Streptomyces roseosporus*, three target genes where identified by *in silico* modeling. Two of them, *dptI* and *dptJ* govern modifications of amino acids, while the third one is involved in glucose catabolism. Co-overexpression resulted in a production increase of 43.2 %[354]. An example of ribosome engineering can again be found for *Streptomyces roseosporus*. When using streptomycin, an aminoglycoside protein synthesis inhibitor, resistant mutants showed a 2.2-fold increase in production of daptomycin from approximately 55 mg/L to 120 mg/L[355]. The higher productivity arises from a single mutation in the *rpsL* gene. Though the mode of action is still not fully understood, it might be due to higher protein synthesis during the late growth phase affecting the regulation and expression of the NRPS genes. Similar observation were made for *Streptomyces albus*[356], *Streptomyces avermitilis*[357] and *Saccharopolyspora erythraea*[357,358].

I.7. Conclusion

It is clear that a lot of research has already been dedicated to unravelling and understanding the biosynthetic pathways involving the production of biosurfactants. Though the commercialization of certain molecules was mainly due to process optimization, an era of tailor-made biosurfactants seems near. Nevertheless, several big obstacles remain that prevent further expansion of the molecular diversity and further optimisation of the production strains, and thereby usage in many industries and applications.

When looking at the economics of biosurfactants, it is impossible to ignore that their overall cost is significantly higher than that of the chemical derived detergents. Even biosurfactants produced through chemical means, but from renewable resources are considerably cheaper than microbial produced biosurfactants. High end applications are not hampered by this problem, but for example 'bulk applications' like household detergents are. As mentioned before, the cost aspect is mostly governed by yield and downstream processing expenses. Methods like resin or carbon adsorption, solvent extraction or precipitation already offer ways to attain high purity without extreme financial investments.

Still, lowering the overall cost of these remarkable molecules points towards the creation of robust, easy to cultivate, hyperproducing strains resulting in preferentially high product uniformity. In the case of *S. bombicola*, this was clearly achieved for certain types of molecules. Titers up to 140 g/L were recently reported for 99 % pure lactonic and 100 % acidic sophorolipids[174]. Unfortunately though, this is often not the case. Often yields remain south of the 30 g/L mark and products are composed out of many different congeners. To counter these problems, a more fundamental understanding is necessary of biosurfactant biosynthesis and its regulation.

Indeed, the regulatory networks controlling the production still need to be unravelled in most cases. For example, in the case of *P. aeruginosa*, the regulatory systems governing rhamnolipid production were extensively explored, but no clear advance of deregulation could be noted. In the case of *S. bombicola* even greater uncertainties remain since no true regulator has been found to date. The only example where engineering the regulation was successful was reported for *U. maydis* known for cellobiose lipid production. A simple overexpression of the cluster regulator Rua1 resulted in enhanced production even when using media that were not optimized for production. Regulation can also be circumvented by using well studied platform organisms like *E. coli* for heterologous production. However, this can result in losing the benefit of the wild type producer resistance towards these molecules leading to lower yields. A compromise would be using naturally robust strains to synthesize molecules similar to the ones they already produce. An example of this strategy is the heterologous production of cellobiose lipids in *S. bombicola*. It has to be mentioned though that the productivity of these molecules is low so further metabolic optimization and investigation is certainly required. Still, harbouring the capacity of efficient producers for the production of similar molecules might be a valid strategy for further cost reduction.

A second focus for strain engineering lies on the genes governing the biosynthesis of precursors or biosurfactants themselves. When looking again at glycolipids, variation in hydrophobic or hydrophilic moieties remain limited to a few carbohydrates and fatty acids. Understanding and/or engineering the specific enzymes offer interesting possibilities to

address either uniformity problems and the demand for tailor-made molecules. In the case of sophorolipids, this resulted in strains producing only lactonic or acidic sophorolipids, as mentioned above. For several classes of glycolipids, a P450 is necessary for activation of fatty acids. The engineering of P450s to alter their substrate specificity and product formation can already be a big step forward in the modification of glycolipids like sophorolipids and cellobiose lipids. Still it has to be noted that the specificity of the enzymes downstream of the P450 has to be compatible with the engineered P450. In the case of UGTA1 of *S. bombicola*, no other substrate than UDP-glucose can be used for glycosylation of the hydroxy fatty acid. Protein engineering can alter the specificities towards other kinds of activated sugars or broaden the range of accepted hydroxy fatty acids. On the other hand, interesting enzymes for heterologous expression can be identified using for example glycan array data, a screening method to investigate protein-carbohydrate interactions[359].

Molecular diversification can also be obtained by utilizing the promiscuity of the biosynthetic enzymes. New kinds of molecules have been synthesized using non-conventional substrates like alcohols and ketones, but the cost of the substrates counteracts the need for cheaper production. Recently, novel kinds of sophorolipids have been produced using petroselinic acid as a substrate[360]. This fatty acid is an isomer of oleic acid, the natural preferred substrate of CYP52M1. Though similar in structure, the double bond is closer towards the carboxylic group of the fatty acid moiety. This small difference in molecular structure resulted in different physico-chemical properties like lowered CMC compared to the molecules with an oleic acid as their hydrophobic group.

Finally, chemical or biocatalytic derivatization of molecules can be used to obtain new kinds of glycolipids with the possibility of new functional groups added to their structure. Recently, several examples have been reported where chemical derivatization of sophorolipids resulted in novel molecules with interesting properties[361–363]. For rhamnolipids[364] and mannosylerythritol lipids[365] similar strategies exist.

To conclude, it is evident that biosurfactant production is maturing, yet several obstacles remain, structural diversification being the major one. The creation of platform organisms like *S. bombicola* is contributing to attain cheap, eco-friendly molecules and the platform will prove a valuable tool in the next few years. In this thesis, attempts are made to better understand *S. bombicola* on a fundamental level and to combine improved knowledge with more production oriented engineering strategies to broaden *S. bombicola*'s product palette.

Chapter II

Expanding the molecular toolkit of S. bombicola

Part of this chapter is being prepared for publication as:

Roelants, S., Ciesielska, K., De Maeseneire, S., Lodens, S., Coussement, P., Pattyn, F., Geys, R., Saerens, K., Saey, L., Verweire, S., Van Bogaert, I., Devreese, B., Soetaert, W., 2016. Development and application of a molecular toolkit for the industrial yeast *Starmerella bombicola* (submitted Metabolic Engineering)

Chapter II - Expanding the molecular toolkit of *S. bombicola*

II.1. Introduction

Ever since the 1970s when the first genetic engineered organisms were created new techniques have been developed to ensure faster, more targeted engineering for a wide array of purposes. A major step forward in yeast engineering was the development of a hybrid plasmid for usage in *E. coli* and in *S. cerevisiae* in 1978[366]. Meanwhile, plasmid systems have been developed for other yeasts as well. Such plasmids can for example help screening for efficient enzymes or be used to determine optimal promotor strength and to balance pathways. Using a plasmid also solves potential location effects when integrating in the genome. Further, since no steps like linearization of the integration cassette are required and no incorporation of the DNA in the genome is necessary, engineering times can be shortened significantly.

S. bombicola, a yeast known for its efficient sophorolipid synthesis, is the subject of extensive research aiming at the development of a platform organism for the production of novel kinds of biosurfactants[318]. A transformation and selection procedure[310] and targeted genome engineering methods have been established, and these methods have been successfully used for the deletion and insertion of genes. Unfortunately, though, a plasmid system for this industrially interesting yeast has not been described. Yet, further expansion of the molecular toolkit of *S. bombicola* with such a plasmid system is essential for speeding up screening and engineering. Hence, in this thesis the development of an episomal vector which can be stably maintained in *S. bombicola* was aimed at.

A functional plasmid requires several types of sequences for maintenance and correct expression of the plasmids and its genes. One of those essential sequences is the origin of replication, called autonomous replicating sequence (ARS) in yeast. These sequences allow efficient replication of chromosomes in every eukaryotic cell and have been used successfully to maintain plasmid based systems in yeasts. Most of them are AT-rich and they often contain a preserved motif highly specific for each species[367]. These features can be used to scan chromosomes for unknown replication origins, a strategy also followed here. Other strategies investigated are the use of heterologous ARS sequences and the screening of genomic libraries.

Once a plasmid system is developed and plasmid based screening efforts have enabled the determination of the most beneficial combination of genes, promotors and terminators, the final combined (heterologous) pathway(s) needs to be introduced in the genome to ensure stability of the resulting strain and because no constant selection pressure would be needed. Hence, it has to be determined which is the largest size of recombination cassette that can be used. The larger this size, the more genes one can introduce in one single transformation. The size of the cassette to be introduced is strongly correlated with the transformation efficiency, which thus needs to be addressed as well. To explore the opportunities in *S. bombicola,* constructs with increasing size were evaluated, and in a final experiment the entire sophorolipid biosynthetic cluster was transformed in *S. bombicola* and integrated at a second location in the genome.

Another tool, which was missing for *S. bombicola* until recently[368], is a robust reporter system enabling the investigation of promotor, terminator and integration sites, but also external stimuli like medium components, temperature, oxygen levels, etc and the location of biosynthetic enzymes. The latter enables to direct engineered proteins towards specific organelles or to cluster a biosynthetic pathway for more efficient production of interesting molecules. The most famous reporter protein is green fluorescent protein (GFP)[369]. This protein, originating from a jellyfish, has been used for *in vivo* localization of proteins, screening promotor libraries or pathway assemblies, and as a biological sensor. Besides the traditional green variant, other emission spectra have been developed as well, ranging from violet to infrared. This allows the simultaneous use of several reporter proteins in one cell, enabling the study of colocalisation and interactions of different proteins or creating libraries of different regulatory elements like promotors and terminators. An expansion of the reporter set suited for *S. bombicola* is described in this chapter: two new fluorescent proteins were evaluated, fused to the sophorolipid biosynthetic genes.

II.2. Materials and Methods

II.2.1. Strains, plasmids and culture conditions

Wild type *Starmerella bombicola* ATCC22214 and a derived *ura3* auxotrophic PT strain[311] were used. Prior to transformation, the yeast cells are cultured in liquid YPD medium composed out of 2 % (w/v) glucose (Cargill; USA), 1 % (w/v) yeast extract (Brenntag; Belgium) and 2 % (w/v) bactopepton (Becton, Dickinson and Company; USA). Selection of *ura3* positive transformants is done on synthetic drop-out (SD) medium lacking uracil. The medium is composed out of Yeast Nitrogen Base without amino acids 0.67 % (w/v) (Sigma-Aldrich; USA), 2 % (w/v) glucose, 2 % (w/v) Agar Noble (Becton, Dickinson and Company; USA) and 0.77 % (w/v) Single Dropout Complete Supplement Mixture without uracil (MP Biomedicals; USA). Evaluation of mutant strains for their capacity to produce sophorolipids was done on medium described by Lang *et al*[370]. The cultures were supplemented with 35 g/L rapeseed oil. Cultivation occurred at 30 °C at 200 rpm unless otherwise stated.

Standard cloning procedures were carried out in *Escherichia coli* Dh5α unless otherwise stated. *E. coli* was cultured in Luria Bertani medium constituted out of 1 % (w/v) trypton (Biokar; Belgium), 0.5 % (w/v) yeast extract, 1 % (w/v) NaCl (Brenntag; Belgium) and if necessary 15 g/L A4 agar (Biokar; Belgium) for plates. Ampicillin (Sigma-Aldrich; USA) was added at a final concentration of 100 mg/L. Incubation of *E. coli* happened at 37 °C and 200 rpm unless stated otherwise.

II.2.2. Molecular techniques

II.2.2.1. General techniques

Primers, plasmids and final strains were designed using the Clone Manager Basic Version 9 software suite from Sci-ED Software. Primers were ordered at Integrated DNA Technologies (USA) and diluted upon arrival to a 10 mM stock solution. Working solutions are derived from these by a 100x dilution to 10 µM. All plasmids and cassettes were confirmed by sequencing at LGC Genomics (Germany). Bacterial plasmid DNA was isolated with the innuPREP Plasmid Mini Kit (Analytik Jena AG). Purification of linear DNA fragments was performed with the innuPREP PCRpure Kit (Analytik Jena AG). All PCR fragments were analysed for their presence and length by agarose gel electrophorese (1 % agarose in 40 mM Tris, 20 mM acetic acid, 1 mM EDTA). A 2-Log DNA Ladder is used as a reference for fragment length. The electrophoresis is run at 120 V for 25 minutes.

Transformation of *S. bombicola* was done following the protocol of Van Bogaert *et al*.[310]. Unless otherwise stated, cells were incubated on the previously described SD medium. Transformation of *E. coli* was performed as described by Sambrook and Russel[371]. Colony PCR on both yeast and *E. coli* were performed with TAQ polymerase (New England Biolabs; USA).

II.2.2.2. Isolation of genomic DNA from *S. bombicola*

Small scale (1~10 µg)

Isolation of small amounts of genomic DNA was performed on 5 mL cultures incubated overnight at 30°C in YPD. From these cultures, 2 mL were spun down and washed with 1 mL mQ water. The cell pellet was dissolved in 540 µL SCE buffer made out of 1 M sorbitol (Sigma-Aldrich; USA), 0.1 M sodium acetate (Sigma-Aldrich; USA) and 60 mM EDTA (Sigma-Aldrich; USA) at pH 7. To lyse the cell wall, 200 U of yeast lytic enzyme (MP Biomedicals; USA) was added. After incubation at 37 °C for 90 minutes, cells were centrifuged for 5 minutes at 14.000 rpm and 200 µL fresh SCE buffer and 20 µL of RNAse A (Sigma-Aldrich; USA) were added. After a two-minute incubation at room temperature, the protocol was switched to the GenElute Bacterial Genomic DNA kit (Sigma-Aldrich; USA) and the kit protocol was followed from step 5 onward.

Large scale (>10µg)

Large scale isolation commences with inoculating the desired strain in 50 mL of YPD. A fully stationary culture produces 100 µL of packed cell volume per mL used. For every 100 µL of packed cells 200 µL of TEAS, a buffer made out of 50 mM Tris-HCl (Sigma-Aldrich; USA), 1 mM ethylenediaminetetraacetic acid (EDTA) (Sigma-Aldrich; USA), 0.1 M lithium acetate (Sigma-Aldrich; USA) and 1% sodium dodecyl sulphate (Sigma-Aldrich; USA) at pH 7.4 is used for resuspension of the pellet. An equal volume compared to the TEAS of glass beads is added, as well as a double volume compared to the TEAS of PCI (Phenol:Chloroform:Isoamyl alcohol 25:24:1 saturated with 10 mM Tris, pH 8.0, 1 mM EDTA) (Sigma-Aldrich; USA). Cells are lysed by using the FastPrep®-24 Instrument (MP Biomedicals; USA) for 20 seconds at 4 m/s. After centrifugation for 10 minutes at 13.000 rpm the upper aqueous layer is collected and 1 mL of absolute molecular grade ethanol (VWR; Belgium) is added. After 5 minutes at -80 °C the mixture is centrifuged at 13.000 rpm at 4 °C and the pellet is recovered. The pellet is suspended in 200 µL TE-buffer (50 mM Tris-HCl, pH 7.4; 1 mM EDTA) and RNAse A is added. The whole is incubated at room temperature for 10 minutes and 500 µL of absolute ethanol is added followed by an incubation of 5 minutes at -80 °C. After centrifugation at 4 °C and 13.000 rpm the pellet is washed with 75 % ethanol:dH_2O followed by a final centrifugation step at 13.000 rpm and 4 °C. The pellet air dried and suspended in 100 µL of TE buffer. If necessary, the pellet is washed several times to obtain higher purities of the isolated genomic DNA.

II.2.3. Vector construction

II.2.3.1. Development of an ARS screening vector for *S. bombicola* and *S. cerevisiae*

Creation of the screening vector

As a start the pJET vector from the Thermo Scientific CloneJET PCR Cloning Kit was used. The pJET vector is not a yeast shuttle vector and hence it does not contain a yeast origin of replication. To keep the screening vector as small as possible, the decision was made to remove the endonuclease Eco47RI since it has no function further down the experiment. Primers P728_pJETinternFOR and P729_pJETinternRev result in a vector backbone of 2013

bp. The *ura3* marker was picked up from *S. bombicola* genomic DNA as well as the terminator of *cabom03g11120*. This terminator was chosen to minimize possible recombination effect at the *ura3* locus when the original terminator of the *ura3* gene would be used. Its functionality was proven during other experiments. The *ura3* marker and its promotor were amplified with primers P551_ura3endrev and P731_pUraUratRegpJETRev resulting in a fragment of 1188 bp. The terminator was amplified using primers P730_pUraUratRegpJETFor and P550_cabom03g11120downfor (336 bp). The entire vector was assembled using In-Fusion HD Cloning Kit and is called BS_EC_0073 (Figure 2.1). The vector is linearized by either using the primers P1316_tUraRev and P729_pJETinternRev (3577 bp) or by restriction digest with *Bam*HI depending on the strategy used for creation of the genomic library.

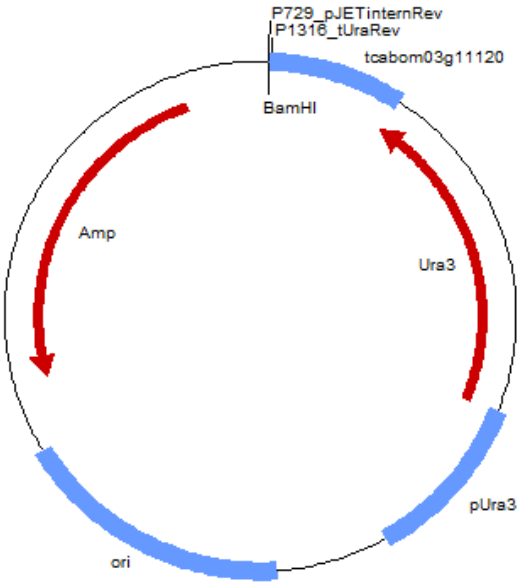


Figure 2.1: Overview of the constructed vector used for ARS screening.

Development of different episomal vectors

Several autonomously replicating sequences and potential candidates were chosen to be screened in the screening vector. In Table 2.1 gives an overview of the sequences that were ligated in the by PCR screening vector by Gibson Assembly (Table 2.2). Sequences of these inserts can be found in the appendices. The Cen4 and 2micron sequences were amplified from plasmids suited for use in *S. cerevisiae* (www.euroscarf.de). The *K. lactis* NRRL Y-1140 ARS was ordered as a gBlock from Integrated DNA Technology and the Repeatchrom1 and Intergenicchrom1 were amplified from genomic DNA of *S. bombicola* ATCC22214. The gBlock of the *K. lactis* ARS was ligated directly in the screening vector thanks to the integrated overlap regions. The final plasmids are named BS_EC_0085 for the *K. lactis* ARS, BS_EC_0142 and BS_EC_0144 for respectively the *S. cerevisiae* CEN4 and 2µ sequences and finally BS_EC_0236 and BS_EC_0263 for the *S. bombicola in silico* discovered sequences.

Table 2.1: Sequences used for ARS screening in *S. bombicola* and/or *S. cerevisiae*

Sequence name	Length (bp)	Forward Primer	Reverse Primer
CEN4	854	P1074_Cen4GibFor	P1075_Cen4GibRev
2micron	1162	P1076_2muGibFor	P1077_2muGibRev
KLARS	452	/	/
Repeatchrom1	1000	P1542b_Repeat760000GibFor	P1543_Repeat760000GibRev
Intergenicchrom1	3001	P1679_776779Kchrom1gibpJETFor	P1680_776779Kchrom1gibpJETRev

Table 2.2: Primers used for the ligation of candidate ARS sequences in the screening plasmid. When overhanging primers were used, a dash represents the start of the binding region of the primer.

Primer name	Primer sequence
P728_pJETinternFOR	CCAATTGCTTTCCAGTCGGGAAACC
P729_pJETinternRev	ATAATTCGGCTGCAGGGGCGGCCTC
P551_ura3endrev	TCATCTTGACTGAACTTTTCTCAGATA
P731_pUraUratRegpJETRev	CTGGAAAGCAATTGG-AAAACAGCCCGCTCTGCTGG
P730_pUraUratRegpJETFor	CTGCAGCCGAATTAT-GGATCCCGATAAACAGCTGACTCCGAGAATATGC
P550_cabom03g11120downfor	GTTCAGTCAAGATGAGCGTTATCGACCTGAATCAGTT
P1316_tUraRev	CCGATAAACAGCTGACTCCGAGAATATGC
P1679_776779Kchrom1gibpJETFor	GAGGCCGCCCCTGCAGCCGAATTAT-AACAATTAACAGATTCAAAGTCGTGACAT
P1680_776779Kchrom1gibpJETRev	ATTCTCGGAGTCAGCTGTTTATCGG-TCCTCAGAAGATTATTTGCAAGGCTA
P1542b_Repeat760000GibFor	ATCACGAGGCCGCCCCTGCAGCCGAATTAT-TAACACGTGCAGATAACATGTC
P1543_Repeat760000GibRev	TGCATATTCTCGGAGTCAGCTGTTTATCGG-TAATCCAGGCACAGGTTACT
P1074_Cen4GibFor	GCGTATCACGAGGCCGCCCCTGCAGCCGAATTATG-CGCTGGGCCATTCTCATGAA
P1075_Cen4GibRev	GCGGCTGCATATTCTCGGAGTCAGCTGTTTATCGG-ACTCTAAGAGGTGATACTTATTTACTG
P1076_2muGibFor	GCGTATCACGAGGCCGCCCCTGCAGCCGAATTATG-AACGAAGCATCTGTGCTTCATTTTG
P1077_2muGibRev	GCGGCTGCATATTCTCGGAGTCAGCTGTTTATCGG-ACGCATTTAAGCATAAACACGCACTA

Creation and screening of a genomic DNA library

Isolated genomic DNA was partially digested for creation of an ARS screening library. The vector BS_EC_0073 was digested with *Bam*HI and the genome with *Sau*3AI. The genomic fragments are purified by gel extraction using the Analytik Jena (Germany) InnuPREP Gel Extraction Kit to obtain certain length classes. The linearized vector is treated with Antarctic phosphatase (NEB) to remove the 5' phosphate to minimize self-ligation. After ligation with T4 ligase (NEB) of the vector and the genomic fragments, a second digest with *Bam*HI was performed to remove any self-ligated vector. The ligation mixture was then transformed in *E. coli* and the cells were incubated overnight with ampicillin. The day after, the library was isolated and transformed in *S. bombicola* PT36. After incubation on SD medium for 5 days, the colonies are screened for potential plasmid possession.

In another strategy the vector was linearized by PCR with primers P1316_tUraRev and P729_pJETinternRev (3577 bp) to ensure that self-ligation is minimized due to the lack of the 5' and 3' phosphate groups. The genomic DNA was digested by the blunt cutting enzyme *Pvu*II. Ligation and transformation were identical to the above described method.

II.2.3.2. Construction of enlarged cassettes for homologous recombination

Construction of the enlarged integration cassettes

The previous constructed plasmid BS_EC_0360 for recombination at the *cyp52m1* locus was used for constructing the enlarged recombination cassettes (Figure 2.2). The plasmid contains a pGEMT vector backbone for cloning in *E. coli* and homologous regions for incorporation in the genome of *S. bombicola*. The 5' homologous region encompasses 1000 base pairs (bp) upstream of the start codon of *cyp102a7*, the gene present inside the plasmid. Downstream of the locus are the *ura3* marker gene with its promotor and terminator followed by the 3' homologues region also encompassing 1000 bp.

Starting from this plasmid, the integration cassette was subsequently enlarged by inserting random pieces of *E. coli* LMG12566 genomic fragments in the *cyp102a7* coding sequence. In total three different cassettes were created by inserting pieces of 2050 bp, 3021 bp and 4082 bp. All inserts were picked up by high fidelity PCR from genomic DNA of *E. coli* MG1655 using the forward primer P271_insertTestFor combined with P272_Rev2KB, P498_Rev3kb and P618_Rev4kb (see Table 2.3). Linearization of the plasmid prior to insertion of the *E. coli* DNA fragments was performed by restriction digest with *Afl*II and *Cla*I (NEB; USA). The vector backbone and insert were ligated using the In-Fusion HD Cloning Kit (Takara; Japan). The final integration cassette was amplified from these vectors with the Phusion High Fidelity DNA Polymerase (NEB; USA) and primers P847_Cm2ExCasLongFor and P887_Cm2ExCasRev.

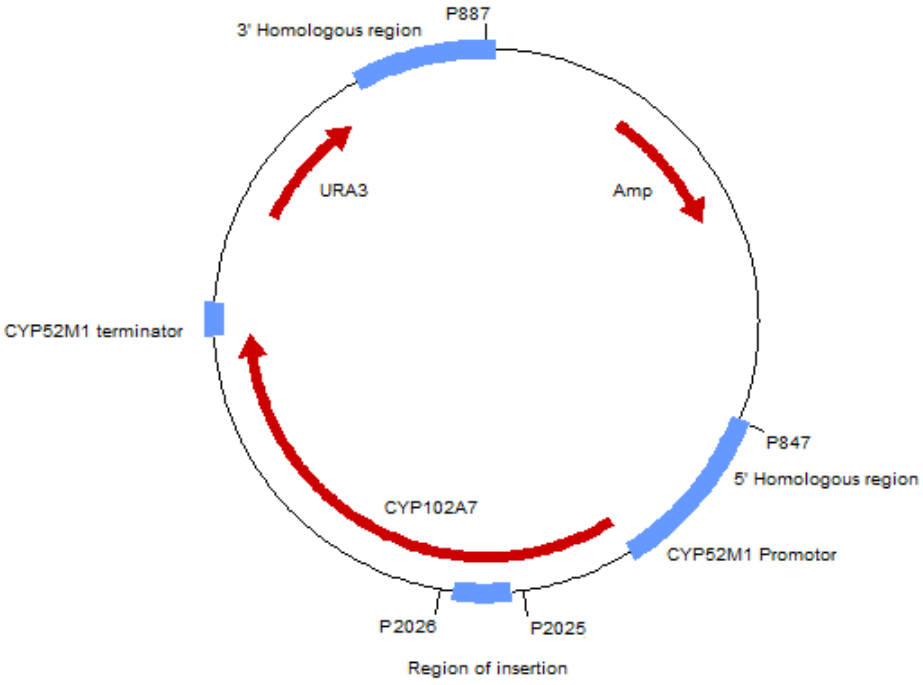


Figure 2.2: Overview of the plasmid BS_EC_0360 used for recombination at the *cyp52m1* locus.

The integration of the cassettes was checked by PCR at the 5' and 3' region of the *cyp52m1* locus by respectively primers P954_Cm2checkFor and P272_Rev2KB (2465 bp) and primers P953_ura3outendfor and P952_A21TotRev (1172 bp). Whole genome extraction followed by PCR was used to validate that the entire expression cassette was integrated. Primers used were P954_Cm2checkFor and P952_A21TotRev resulting in fragments of 7604 bp, 9654 bp, 10625 bp and 11868 bp for the different sizes. During efficiency testing, primers

P2025_CYP1021A7insertcheckFor and P2026_CYP102A7insertcheckRev were used instead. This resulted in a PCR-product with specific length for each cassette (482 bp, 2404 bp, 3375 bp and 4436 bp).

Construction of the sophorolipid cluster knock-in cassette

Construction of the cassette for a knock-in of the sophorolipid biosynthetic cluster at the *ura3* locus started from the available plasmid BS_EC_0251 designed for recombination at the *ura3* locus. The backbone of the vector was amplified with primers P647_FOR_vector pGAPD_ura3_ura3locus and P824_ura3cdsrev (4781 bp). The benefit of the *ura3* marker upstream of the inserted cluster is that this configuration enables using the double terminator of the *ugtb1* gene and thereby saving space on the plasmid (Figure 2.3). The entire sophorolipid cluster was amplified from genomic DNA of *S. bombicola* ATCC22214 with primers P1040_ClusterGibURA3For and P1041_ClusterGibURA3rev (11057bp). Both PCRs were carried out with Primestar GXL DNA Polymerase (Takara; Japan). The fragments were ligated by Gibson Assembly[372]. The final cassette was amplified from this plasmid using primers P49_FOR_URA3upextNheI and P1013_GAPDAT_FusionF (14121 bp) with the previous mentioned high fidelity polymerase. Transformants were checked for integration at the *ura3* locus with primers P33_FOR_checkcassIN and P72_REV_ADHdown for the 5' region (2728 bp) and P843_A21knockhygroCasFor and P35_REV_checkcassIN_DOWN for the 3' region (3187bp). To ensure that the entire cluster was integrated, a final PCR was performed on the correct mutant strains with P33_FOR_checkcassIN and P35_REV_checkcassIN_DOWN (14329 bp). A similar PCR was done at the level of the original sophorolipid cluster location with primers P469_FOR_GOF7_ADH and P64_REV_cassPHAC1 (12712 bp).

Table 2.3: Primers used for the construction of the CYP102A7 base cassette and subsequent enlarged cassettes. Restriction sites are marked by an underline. When overhanging primers were used, a dash represents the start of the binding region of the primer.

Primer name	Primer sequence
P33_FOR_checkcassIN	CCATACTCAAGCGCGAACAC
P35_REV_checkcassIN_DOWN	GAGCTCAAGACGCGTTTACTCAATGC
P49_FOR_URA3upextNheI	CGCTAGCC-ATGCTGAAGACAGCACCGCTGCTATC
P64_REV_cassPHAC1	GGTGTCGACTCGCCAAATTC
P72_REV_ADHdown	TTGCAACTGTGCCCTCCATC
P271_insertTestFor	TCCGGATAAGCTTAA-GCGTGGTCGTTAGGCATTCC
P272_Rev2KB	GCCAAAGCCTTATCG-CCAGCCGGATACACTTCCAG
P469_FOR_GOF7_ADH	TCACGCACTCAAACTTACAA
P498_Rev3kb	GCCAAAGCCTTATCG-CGTCGTCCATGCTTTCAG
P618_Rev4kb	GCCAAAGCCTTATCG-GCGGTACCAATCTCCAGTCC
P647_FOR_vector_pGAPD_ura3_ura3locus	CTGCAGTTTTAAAAATTTGCAGCTCTATAG
P824_ura3cdsrev	TCATCTTGACTGAACTTTTCTCAGATA
P843_A21knockhygroCasFor	GAGTCGGGCGTTATTTCTCC
P847_Cm2ExCasLongFor	CGTTGTCAAGTCCTAAGGTAT
P887_Cm2ExCasRev	AAGCGTGAAGCTCCTCTGACAATC
P952_A21TotRev	GCTCTTGTTCGGTACTCTTATT
P953_ura3outendfor	TAAAGAAACGAAGGGCCCAGCAGTC
P954_Cm2checkFor	TCATAGCGAGTTTCTTTGCATGTG
P1013_GAPDAT_FusionF	GCCTCGTCAACCATCTTATC
P1040_ClusterGibURA3For	GGGCTGGAATGCATATCTGAGAAAAGTTCAGTCAAGATGA-TTCTGCTCTCAACACCGAGT
P1041_ClusterGibURA3rev	AACTAAATTACTATAGAGCTGCAAAT-TTTTAAAACTGCAGTTAAGAAAACCGCACAACC
P2025_CYP1021A7insertcheckFor	CTTGTTGACCGGATCATCG
P2026_CYP102A7insertcheckRev	AATGCCTCGGCATCTTCTC

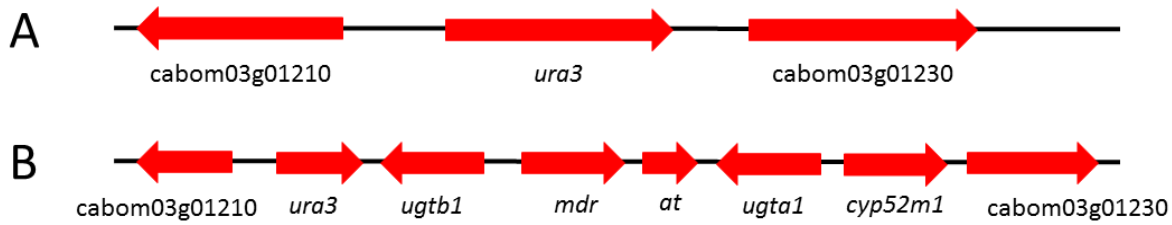


Figure 2.3: Genomic organization of the *ura3* locus and surrounding sequences before (A) and after (B) knock-in of the sophorolipid gene cluster.

II.2.3.3. Constructs for the localization of the sophorolipid biosynthetic enzymes

Construction of the CYP52M1-FP plasmids

The starting plasmid for the construction of the different fusion construct is derived from plasmid BS_EC_0081. The vector was linearized with P69_FOR_RegulUP and P1439_BMRLinkerRev (8228 bp) and Primestar GXL High Fidelity polymerase. The inserts are the coding sequences of *gfpco10genes*[368], *bfpcostatgenes* and *rfpcostatgenes* (see Appendix B). The sequences of the blue and red fluorescent proteins are synthetic constructs. The BFPcoStatgenes (BFP) (Genbank AAB16959.1) and RFPcoStatgenes (RFP) (Swiss-Prot Q9U6Y8.1) coding sequences were back translated from protein sequences using the codon usage of *S. bombicola* during stationary phase (see Appendix A). This phase was chosen since expression of *cyp52m1* and the other genes of the cluster only occurs in this phase. The back translated coding sequences were finally ordered as a gBlock at Integrated DNA Technologies. DNA sequences can be found in the appendices. The coding sequences of the fluorescent proteins were amplified (Table 2.4) and subsequently assembled by Gibson Assembly at the C-terminus of CYP52M1. The linker used between CYP52M1 and the fluorescent proteins belongs to CYP102A1 and was already present on the starting plasmid BS_EC_0081. The final plasmids are BS_EC_0218_pGEMT_CYP52M1GFP, BS_EC_0219_pGEMT_CYP52M1BFP and BS_EC_0216_pGEMT_CYP52M1RFP. Amplification of the different fluorescent constructs was done with primers P847_Cm2ExCasLongFor and P887_Cm2ExCasRev. Screening of the *S. bombicola* mutants was performed by colony PCR with primers P954_Cm2checkFor and P1439_BMRlinkerRev (2716 bp) for the 5' region and with primers P953_ura3outendfor and P952_A21TotRev (1172 bp) for the 3' region.

Construction of the CYP52M1-GFP and UGTA1-RFP fusion plasmids

Assembly of plasmids where RFP is fused to the first glucosyltransferase UGTA1 was done by using the previous constructed BS_EC_0218 plasmid (see III.2.2.2) as a vector backbone donor and integrating the new 5' homologues region for integration at both the *cyp52m1* locus and the *ugta1* locus together with the *rfp* coding sequence fused to the *ugta1* sequence. To test whether there is an effect on transferase activity by the RFP, both N- and C-terminally fused constructs are made for UGTA1. All pieces were assembled in one Gibson assembly. The final plasmid is called BS_EC_0353 (5' *ugta1-rfp* 3') and BS_EC_0354 (5' *rfp-ugta1* 3').

Table 2.4: Primers used for the construction of the localization plasmids When overhanging primers were used, a dash represents the start of the binding region of the primer.

Primer name	Primer sequence
P52_REV_upCYP	ATATGTACTTTTCAATATGATAAACGGAGAAATAACG
P69_FOR_RegulUP	GTTTCTTAGCCTCCCATGGAAG
P722_BM3CPRLinkerInfFor	AAAATTCCGCTTGGCGGTATTCCTTCACC
P952_A21TotRev	GCTCTTGTTCGGTACTCTTATT
P953_ura3outendfor	TAAAGAAACGAAGGGCCCAGCAGTC
P954_Cm2checkFor	TCATAGCGAGTTTCTTTGCATGTG
P1014_URA3_Cb_gen_FOR	GTTGAACCATGATGGCAGTGTTCG
P1394_URA3RevGib	GAACACTGCCATCATGGTTCAACTTCTGCTCTCAACACCGA GTGTAGTTCTTGGTTT
P1439_BMRlinkerRev	GTTTTCTGCCTTTTTGCGTAC
P1529_BFPGibCYP52M1For	GCTAAAAAAGTACGCAAAAAGGCAGAAAAC-TCGAAAGGAGAGGAGCTGTTCACTGG
P1530_BFPGibCYP52M1Rev	GAACGTTTCTTCCATGGGAGGCTAAGAAAC-TTATTTGTACAGCTCGTCCA
P1531_RFPGibCYP52M1For	GCTAAAAAAGTACGCAAAAAGGCAGAAAAC-AGGTCTTCCAAGAACGTTATCAAGGAATT
P1532_RFPGibCYP52M1Rev	GAACGTTTCTTCCATGGGAGGCTAAGAAAC-CTAGAGGAAGAGGTGGTGAC
P1533_GFPGibCYP52M1For	GTCTGCTAAAAAAGTACGCAAAAAGGCAGAAAAC-TCGAAAGGAGAGGAGCTGTTCACTGG
P1534_GFPGibCYP52M1Rev	GAACGTTTCTTCCATGGGAGGCTAAGAAAC-TTATTTGTACAGCTCGTCCA
P1535_UGTARFP_CYP52M1GFP_BBRev	TATGGGAGAGCTCCCAACGCGTTGGATG
P1536_UGTARFP_CYP52M1GFP_BBFor	TATATGGCCTGAAAGAGGCAAGC
P1537_UGTARFP_CYP52M1GFP_5homGibFor	TGCATCCAACGCGTTGGGAGCTCTCCCATA-GACTCCTAGAAAAGAAATTGAC
P1538_UGTARFP_CYP52M1GFP_5homGibRev	CACCTCTTCCTCTAGAATC-GTACGATCAAATCAGA
P1539_UGTARFP_CYP52M1GFP_RFPGibFor	TTTGATCGTACGATT-CTAGAGGAAGAGGTGGTGAC
P1540_UGTARFP_CYP52M1GFP_RFPGibRev	TTAGCGGTGAGTTCT-AGGTCTTCCAAGAACGTTAT
P1541_UGTARFP_CYP52M1GFP_UGTAGibRev	GTTCTTGGAAGACCT-AGAACTCACCGCTAAGGC
P1542_UGTARFP_CYP52M1GFP_UGTAGibFor	GCTTGCGGCTTGCCTCTTTCAGGCCATATA-ATGCCCGTTTTGAGTCTGGTA
P2184_RFPGibPromFor	GCTTGCGGCTTGCCTCTTTCAGGCCATATA-ATGAGGTCTTCCAAGAACG
P2185_RFPGibUGTARev	GACTCAAAACGGGCAT-GAGGAAGAGGTGGTGACG
P2186_UGTA1GibRFPFor	CACCACCTCTTCCTC-ATGCCCGTTTTGAGTCTG
P2187_UGTA1GibTermRev	CTCTTCCCTGATCTGATTTGATCGTACGATTCTAA-GAACTCACCGCTAA

II.2.4. Sampling and analysis

II.2.4.1. Follow up of growth and glucose consumption

Biomass formation was followed by measuring the optical density at 600 nm using the FLUOstar OPTIMA fluorometer (BMG LABTECH GmbH, Germany). Samples were diluted in 0.9 % NaCl in dH_2O solution if necessary. Follow up of the glucose concentration was done with the 2700 Select Biochemistry Analyzer (YSI inc.; USA). Samples were diluted in a range between 0.1 g/L and 7.5 g/L to ensure correct measurement.

II.2.4.2. Analysis of sophorolipid production

Sophorolipid production was followed by sampling every two days. The samples were extracted by adding three volumes of ethanol to the culture broth. After shaking vigorously, the samples were spun down at 14.000 g for 5 minutes. The supernatant was analysed by HPLC.

The HPLC system used is an Agilent (USA) Varian Prostar HPLC system with Evaporative Light Scattering Detection (Grace; USA) at 40°C. The column used is a Chromolith® Performance RP-18 endcapped 100-4.6 HPLC from Merck KGaA (Germany). The column is kept at 30 °C during the analysis. A gradient of two eluents, 0.5 % acetic acid (Sigma-Aldrich; USA) aqueous and 100 % acetonitrile (Sigma-Aldrich; USA), is used to separate the different molecules. The analysis starts at 95 % acetic acid aqueous and changes over a period of 40 minutes to 100 % acetonitrile. This condition is kept for 10 minutes to flush the column. Afterwards, the eluent composition is switched back to 95 % acetic acid aqueous in a 5-minute period. The flow rate was 1 mL/minute during the entire analysis. Quantification of sophorolipids was done using a calibration curve of purified acidic and lactonic sophorolipids. If necessary, total sophorolipid extraction and purification was performed followed by gravimetrical determination of the obtained sophorolipids.

Liquid chromatography mass spectroscopy (LC-MS) analysis was performed with an Intertek ASG (Manchester, UK) with a Micromass Quattro Ultima LIMS 1107 (Waters). The detection range was set at m/z 200 to 800 and the negative ion mode was applied. The same column and LC conditions as for the HPLC analysis were used.

UPLC-ELSD was done on an Acquity UPLC H-Class CM with an Acquity UPLC ELS detector (Waters). The column used is an Acquity UPLC® CSH™ C18 1.7 µm, 2.1x50 mm. The temperature of the drift tube, nebulizer and column is set at 50, 12 and 30 °C respectively. The same gradient conditions were applied as for the HPLC method, only differing in the time required to obtain 100 % acetonitrile (6.72 minutes). The flow rate used was 0.6 mL/minute. Total time of analysis is 10.12 minutes.

Analysis by thin layer chromatography (TLC) is based on the method developed by Asmer at al[156]. The samples were spotted in volumes of 2 µL on VWR TLC silica gel 60 F254 25 aluminium sheets. The eluens used is chloroform:methanol:water in a 65:15:2 ratio. The TLC plate was air dried and sprayed with a 10% H_2SO_4 solution and developed by applying heat.

II.2.5. Fluorescence imaging

Imaging of the fluorescent *S. bombicola* strains was done using a Nikon A1+ with a Plan Apo VC 60x oil DIC N_2 filled objective. Excitation wavelengths of 405.3 nm, 488.0 nm and 561.7 nm for respectively blue, green and red fluorescent protein were used. Emission was observed at 450 nm, 525 nm and 595 nm. Imaging was done with a Nikon Galvano scanner with a DU4 detector.

II.2.6. qPCR and Multi Reaction Monitoring

II.2.6.1. qPCR sample preparation

Cells were harvested at the time point of interest and centrifuged for 5 minutes at 14.000 g. The supernatant was removed and the remaining cell pellet flash frozen in liquid nitrogen. RNA isolation was performed with the RNeasy Mini Kit (Qiagen; Netherlands). The harvested cells were mechanically crushed using liquid nitrogen and a spatula. When the cell pellet was sufficiently disrupted, 350 µL of RTL buffer was added and the cell suspension was further disrupted by a syringe and MICROLANCE 21 G2 green needle. The lysate was then centrifuged for 5 minutes at 14.000 g and the supernatant further processed by adding 1

volume of 70 %. This mixture was applied to the RNeasy Mini Spin Column. The column was centrifuged for 15 seconds at 15.000 rpm and the flow through was discarded. DNase digest on column was performed by adding 350 µL RW1 buffer to the column and centrifuging for 10 seconds at 10.000 rpm. The DNase incubation mix was prepared by adding 10 µL DNase stock solution (QIAGEN RNase-Free DNase 1500 Kunitz dissolved in 550 µL mQ) to 70 µL RDD buffer. The resulting 80 µL mix was added to the column and incubated for 15 minutes at room temperature. Before centrifugation for 15 seconds at 15.000 rpm, 350 µL RW1 buffer was added to the column. The resulting flow through was discarded. Two washing steps were performed with 500 µL RPE buffer. Elution of the DNA was performed with RNase free water. cDNA synthesis was performed with the Superscript III First-Strand synthesis kit (Invitrogen; USA) with oligo$(dT)_{20}$ primers. Concentrations and quality of the RNA and cDNA were measured by the Nanodrop (Thermo Scientific; USA).

II.2.6.2. qPCR analysis

qPCR analysis was performed in 72-well plates. For each strain, three housekeeping genes and the genes involved in sophorolipid production were selected (Table 2.5) and three technical replicates were taken for each strain. The SensiMix™ SYBR® & Fluorescein kit was used in all qPCR analyses. The reaction mixture was composed out of 1 µL forward and reverse primer (10 µM), 9 µL of mQ, 10 µL of Sensimix 2x which contains SYBR Green for detection. The amount of cDNA added to each sample was 1 µL. cDNA concentrations were normalised before addition. The reactions were performed in the Rotor-Gene 3000 (Corbett Life Science) using Rotor Discs (Qiagen). The qPCR was performed with one initial hold for 10 minutes at 95 °C followed by 45 cycles of 25 seconds at 95 °C, 60 s at 58 °C and 20 seconds at 72 °C. After the PCR, a melting curve was generated by increasing the temperature to 95 °C to assess the PCR specificity, potential contamination and primer dimers. Data analysis was performed with the qBase Plus software package developed by Biogazelle.

Table 2.5: Primers used for the qPCR analysis of the wild type and double cluster *S. bombicola*. HSK3_MPP = Mitochondrial processing peptidase subunit; HSK11_SDH = Saccharopine dehydrogenase; HSK17_G6PDH = Glucose-6-phosphate dehydrogenase.

Primer name	Primer sequence
P421_FOR_HSK3_MPP	ATGAGCTTCTCGACCTCTTA
P422_REV_HSK3_MPP	AAATGAACCAGGTCGTCAAT
P437_FOR_HSK11_SDH	TTGAAGGCTAATGCTGACTC
P438_REV_HSK11_SDH	GTCTTGGCATCAGAGAACAT
P449_FOR_HSK17_G6PDH	TCAAATTGCCCGTAATGAGT
P450_REV_HSK17_G6PDH	CAATCCCGGGAATTTGGTAT
P457_FOR_GOF1_CYP52M1	CAATCCTGGCAACCAAATTC
P458_REV_GOF1_CYP52M1	CGAGACTGTTTCCATCCATT
P459_FOR_GOF2_UGTA1	AGCAGCCATCAACTATGAAG
P460_REV_GOF2_UGTA1	CTTTGACGCCCAATATACCA
P463_FOR_GOF4_AT	CTGCGACTCATCCTATTAGC
P464_REV_GOF4_AT	AATTGTTGAGCCATCTCCAA
P465_FOR_GOF5_MDR	TGTATGGAGTGAGGAAGGTT
P466_REV_GOF5_MDR	GTACTTGAGGTCGAGTAGGA
P467_FOR_GOF6_LIP	CTTGTCGAGCAGTATGTTGA
P468_REV_GOF6_LIP	TCATGAAGAAGACCCGGATA

II.2.6.3. Protein extraction for MRM

The *S. bombicola* wild type (WT) and double cluster strains were both cultivated on the Lang medium until sophorolipid production was detected (72 hours). Proteins were extracted from cell pellets derived from 2 mL of culture using the Thermo Scientific TM Mem-PERTM Eukaryotic Membrane Protein Extraction Reagent kit. The hydrophobic and hydrophilic proteins were kept together. Next, proteins were precipitated with trichloroacetic acid (Sigma-Aldrich; USA) and washed with acetone. Eventually pellets were solubilized in 2 M urea (Sigma-Aldrich; USA) in 50 mM ammonium bicarbonate. 50 µg of protein solutions were spiked with 200 ng BSA (MS grade protein standard), reduced and alkylated, and digested with 0.5 µg trypsin. Digested samples were dried and dissolved in 100 µl 0.1% formic acid in water for LC-MRM analysis (5 µL injection).

II.2.6.4. LC-MRM Analysis

The digested samples were first separated by RP-LC on a U3000- RSLC system (Thermo). The sample was loaded onto an Acclaim PepMap100 pre-concentration column (L × ID 2 cm × 100 µm, C18, 5 µm, 100 Å) at a flow rate of 5 µl/min, and flushed for 5 min with 0.1% HCOOH/2% ACN. The sample was then separated on a Thermo Acclaim PepMap100 analytical column (L×ID 15 cm×75 µm; C18; 3 µm, 100 Å) at a flowrate of 300 nl/min, with mobile phases 0.1 % HCOOH in water (solvent A) and 0.1 % HCOOH in ACN (solvent B). Peptides were separated with a 30 min gradient, going from 2 to 40 % solvent B, and eluting peptides were sprayed directly in a 4000 QTRAP mass spectrometer (AB Sciex, Framingham, MA) with a NanoSpray II ESI source (AB Sciex), using a PicoTip Emitter (uncoated SilicaTipTM 10 ± 1 µm). The ion spray voltage was set at 3.5 kV, curtain gas at 10 (arbitrary units), nebulizing gas at 5, and interface heater temperature at 60 °C.

Target peptides (two for each protein; Table 2.6) were measured in multiple reaction monitoring (MRM) acquisition mode, with the Q1/Q3 resolution set at low and with a maximum total cycle time of 3 s. The double charged peptide was selected as precursor (Q1), and for each precursor three fragment ions were selected from the y-ions (Q3). The MRM data was imported in Skyline v2.5[373] The total area under curve (AUC) of each target peptide was calculated, and normalized to the spiked BSA standard.

Table 2.6: Sequences of the target peptides used during the MRM experiments.

Protein	Target peptide 1	Target peptide 2
UGTB1	AIPEQYDALQTALK	EVLATPSYHEK
MDR	GLTAASILNEAIDR	NNTPGALTSILAK
AT	EFNTIASESR	TVVGGVPAR
UGTA1	TGLPTVEQIK	LPDDVVVPENAR
CYP52M1	FNDFGLGAR	LAPVLPLNFR

II.3. Results and Discussion

II.3.1. Development of a plasmid for *S. bombicola*

Extra-chromosomal entities can be found both in prokaryotes and eukaryotes. Among these, plasmids are the most studied and used in biotechnological applications. Engineering of strains is facilitated and sped-up by these plasmids and hence they are essential tools in the state-of-the-art engineering strategies nowadays commonly used. Unfortunately, *S. bombicola* does not harbour a plasmid system like for example the *Saccharomycetaceae,* a shortcoming aimed to be solved in this PhD. Development of a plasmid system can be done by identifying sequences that are able to autonomously initiate replication inside the cell. Several strategies include *in silico* analysis of the genome, using heterologous replicating sequences from other organisms with known functionality or screening genomic libraries of *S. bombicola*.

II.3.1.1. Heterologous ARS screening

As described by Liachko[374], the ARS sequence located at position 781040–781491 on chromosome F of *Kluyveromyces lactis* NRRL Y-1140 can be used as a relatively stable ARS (also called panARS) for several budding yeasts like *S. cerevisiae* and *P. pastoris*. The screening vector BS_EC_0073 containing the *S. bombicola ura3* selection marker under control of its native promotor (II.2.3.1) was used to assess the possibility of using this sequence for plasmid replication in *S. bombicola*. As a control, the same plasmid was transformed in an *ura3* negative strain of *S. cerevisiae*. After 10 to 12 days colonies appeared on the selection plates for *S. bombicola,* but after screening by PCR these colonies turned out to have integrated the *ura3* marker in their genome, allowing them to grow. This was in sharp contrast to the *S. cerevisiae* mutants that appeared after 3 days. From these the plasmid was easily recovered and transformed in *E. coli*. Analysis by sequencing confirmed that no other replicating sequence was present in vector BS_EC_0085 (BS_EC_0073 with the *K. lactis* panARS) than the *K. lactis* panARS, further proving activity in *S. cerevisiae*. Similar results were observed for the 2µ and CEN4 sequences: illegitimate or non-homologues integration occurred in *S. bombicola,* while stable plasmid replication was confirmed in *S. cerevisiae*. Similar results were described by Roelants et al. for the CEN4 sequence[368].

II.3.1.2. Genomic libraries for ARS screening

As it became clear from the results of the heterologous ARS screening that *S. bombicola* is unable to recognize such ARS sequences, it was decided to scan its own genome. Several libraries were constructed from genomic material of *S. bombicola* using different restriction enzymes as described in II.2.3.1. Vector BS_EC_0073 contains a *Bam*HI restriction site which can be used to ligate genomic DNA partially digested with *Sau3*AI. Both enzymes create compatible sticky end DNA sequences to enable the efficient creation of a library. Another enzyme tested is *Pvu*II, a blunt cutting enzyme. After transformation of the genome library vectors, several colonies appeared on the selective plates. All were capable of growing on minimal selection medium, yet plasmid recovery in *E. coli* failed. Most likely, the strains integrated the *ura3* marker in their genome again. No useful ARS sequence for *S. bombicola* was found by following this strategy.

II.3.1.3. *In silico* search for potential ARS sequences

In silico searching for ARS sequences when no real knowledge is available of the organism of interest is similar to finding a needle in an Earth sized haystack. Yet, by using tools developed for or observations made in other organisms, interesting regions were searched for. We used for example GC-skew[375], AT-content, DNA curvature[376], motif searching, blasting known ARS sequences or any combination of the above. To limit the search space, only chromosome 1 of *S. bombicola* was used.

Using these methods, a region was detected on chromosome 1 that exhibits interesting properties. Position 777464 and the surrounding base pairs showed very high AT content, high bendability and curvature of the DNA strand. All these elements are known to be present in ARS sequences in *S. cerevisiae*. Unfortunately, after being tested in the screening vector (II.2.3.1 Table 2.1 Repeatchrom), no stable plasmid was maintained in *S. bombicola*.

Another *in silico* strategy used was screening for a potential centromere sequence. These sequences are often situated in regions where low to no active genes are present. When looking at RNA data, these regions also have lower expression levels due to their compact DNA structure. Such a region was detected for chromosome 1 of *S. bombicola,* but again after testing in the screening vector only mutants that had incorporated the vector were discovered (II.2.3.1 Table 2.1 Intergenicchrom).

Even though all efforts to develop a plasmid system for *S. bombicola* failed, it was decided to determine the maximum size of integrative cassettes for *S. bombicola* anyway, as this remains interesting for engineering purposes regardless of the plasmid system.

II.3.2. Maximum size of homology based genome integrating cassettes

Different integrative cassettes for recombination at the *cyp52m1* locus were created as described in II.2.3.2. This resulted in four different expression cassettes with lengths of 7604 bp, 9654 bp, 10625 bp and 11686 bp. The gene present on the smallest integration cassette codes for CYP102A7, an enzyme proven not to be harmful for *S. bombicola* when being expressed (see Chapter III). The larger integration cassettes were created by inserting a random DNA sequence originating from *E. coli* MG1655 in the coding sequence of *cyp102a7*. When transforming the different linear fragments, an equimolar mixture of the different cassettes was used to ensure equal transformation conditions for all the sizes tested. The mixture was transformed 5-fold into a fresh batch of competent *S. bombicola* cells. The resulting plates were incubated for one week. In total 176 colonies appeared on the selective plates. On average, 35.2 ± 8.58 colonies were counted per transformation. By colony PCR, all colonies were tested to determine the exact cassette being incorporated by amplifying over the *E. coli* insert (Table 2.7). This resulted in a PCR-product with specific length for each cassette (482 bp, 2404 bp, 3375 bp, 4436 bp). As depicted in Table 2.7, no significant differences in the number of obtained colonies was observed for the different cassette lengths. Even for the relatively short cassette of 7.6 kb, only 9 colonies carried the correct cassette. This is in sharp contrast with the 148 colonies which appeared on the selective plates, but showed a band of approximately 900 base pairs in the PCRs. These colonies appeared simultaneously with the correct mutants. Scanning the genome of the wild type *S. bombicola*

resulted in no additional binding sites for both primers, so this 900 bp band could not be explained. However, control PCR on wild type *S. bombicola* did produce a similar PCR fragment. This is not expected since the primers used for the screening are specific towards the *cyp102a7* coding sequence in the integration cassette, which is absent in the genome of wild type *S. bombicola*.

Overall, 15.9% of the mutants had successfully integrated one of the four different expression cassettes. Each transformation was performed with 3 µL of equimolar composition and a total amount of 520 ng per transformation. This translate to roughly 4.9 x 10^7 molecules of DNA per transformation or in total 2.5 x 10^8 molecules used. Comparing these values to the ones of Van Bogaert *et al*[310], some interesting parallels can be seen. In their research, a total amount of 2.5 µg was used during two transformations, equalling 1.1 x 10^9 DNA fragments. From all transformants, they selected 68 at random and screened them by colony PCR. In total, 12 correct mutants were identified, which corresponds to 17.6 % of the total sample population, similar to the results obtained here. When considering the number of colonies per amount of DNA, the enlarged cassettes have 10 positive transformants per µg of DNA. In the research of Van Bogaert *et al*.[377], similar values are obtained. This suggests that the incorporation of longer DNA fragments is as efficient as shorter ones and that transformation efficiency is not necessarily dependent on the size of the integrative cassette. Other parameters clearly also play a significant role in overall efficiency, which is low when for example compared to *S. cerevisia* where efficiency ratings up to 75% and higher have been reported[378]. Still, it is possible to construct large integrative DNA fragments for a specific genomic region and successfully integrate them.

Table 2.7: Analysis of the colonies after transformation of the different integration cassettes. # = number

		# integrating				
Repeat	**Total #**	**7.6 kb**	**9.6 kb**	**10.6 kb**	**11.6 kb**	**# without cassette**
1	43	6	2	1	0	34
2	28	1	0	1	0	26
3	44	1	3	1	1	38
4	25	0	1	2	2	20
5	36	1	1	2	2	30
Total	**176**	**9**	**7**	**7**	**5**	**148**

II.3.3. Doubling the sophorolipid biosynthetic cluster

II.3.3.1. Strain preparation and sophorolipid production analysis

Overexpression of certain genes essential in the production of secondary metabolites has been proven to be a valuable technique for enhanced production[174,379]. These overexpression experiments are often restricted to certain bottleneck steps in the pathway and aim at generating higher fluxes. In the case of *S. bombicola*, a real bottleneck hasn't been found in terms of sophorolipid biosynthesis. Therefore, it has been opted to double the cluster encoding the different enzymes to assess whether this leads to higher productivity, higher expression levels or altered regulation of sophorolipid biosynthesis. This also acts as a proof of concept for the integration length analysis since the cluster is already over 11000 bp in length. The cassette developed in this experiment has a total length of 14121 bp designed to integrate at the *ura3* locus (Figure 2.3). Several differences with the experiment described in II.3.2 have to

be pointed out. First of all, the cassette is 2500 bp longer than the longest one tested for integration at the *cyp52m1* locus in II.3.2. Another big difference is the genomic location for recombination. The *ura3* locus is situated on chromosome 3 around the 226 kb mark while the original cluster location is only 40 kb from the telomere of chromosome 2. Furthermore, the cassettes tested for the different sizes mainly contained *E. coli* sequences: the only homology present except for the recombination regions are the *ura3* promoter and terminator. In contrast, the double cluster integration cassette has an 11 kb long region homologues to the sophorolipid biosynthesis cluster since an extra copy of this cluster is being integrated without any modifications to it.

After transformation of the double cluster integration cassette, three colonies appeared on the selective plates. No additional transformations were done with the double cluster integration cassette. The colonies were screened for incorporation of the biosynthetic cluster at the *ura3* locus by colony PCR. As a control, a wild type *S. bombicola* was used (Figure 2.4). The wild type, which has only the *ura3* gene at the *ura3* locus, shows a band of 2944 bp while the cluster mutants clearly have a larger fragment at that position, representing the correct integration of the cassette. A similar PCR was conducted at the original cluster location to investigate potential genomic alterations. As can be seen for this position, the mutant strains show an identical result as the wild type. Though no sequencing has been done on the mutant strains, at first sight no difference is visible besides the integration of an extra copy of the sophorolipid cluster.

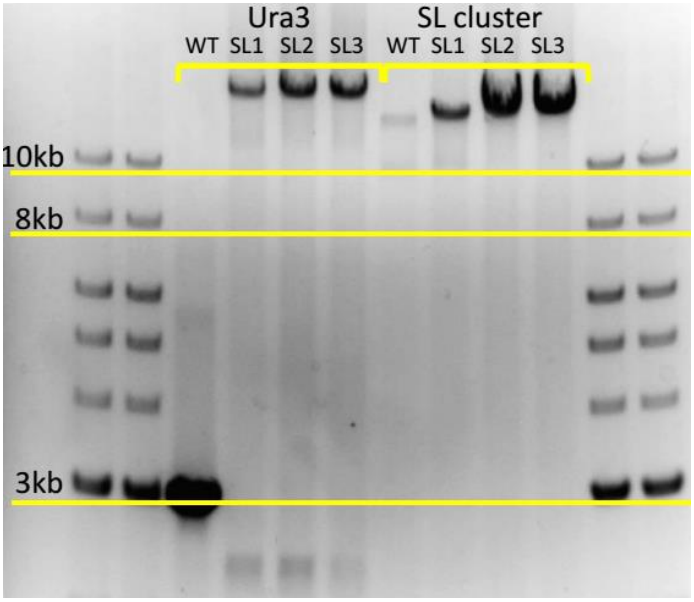


Figure 2.4: Result of the colony PCR on a wild type *S. bombicola* and the three mutants with a double SL biosynthetic cluster. On the left, the results for the ura3 locus are visible while on the right those for the original SL cluster position are visualized.

Next, a production experiment was conducted to test the new strains for their production capabilities (as described in II.2.1). All three mutants and a wild type were inoculated *in duplo* and sampled during the exponential and stationary phase to assess if the sophorolipid synthesis benefits from the doubled cluster on a different location. Since it is no longer situated at the end of chromosome 2, chromatin-level regulation[380] of the cluster might be reduced or even abolished, resulting in exponential production of sophorolipids. To reduce product variety, the strains were grown on oleic acid. This way, only one fatty acid type will be available for

CYP52M1 to hydroxylate instead of several different types when using vegetable oils like rapeseed oil. Analysis on HPLC-ELSD made it clear that there is no difference in sophorolipid mixture composition for the wild type *S. bombicola* and the double cluster mutants Unfortunately, there was big variety in production levels between the different replicates, which made a second production trial necessary.

The wild type and double cluster mutant 1 were inoculated in four-fold and oleic acid was used as a substrate. In Figure 2.5, follow-up of the production can be seen until 5 days after inoculation. Some small differences are visible at the beginning of the stationary phase (Figure 2.6).

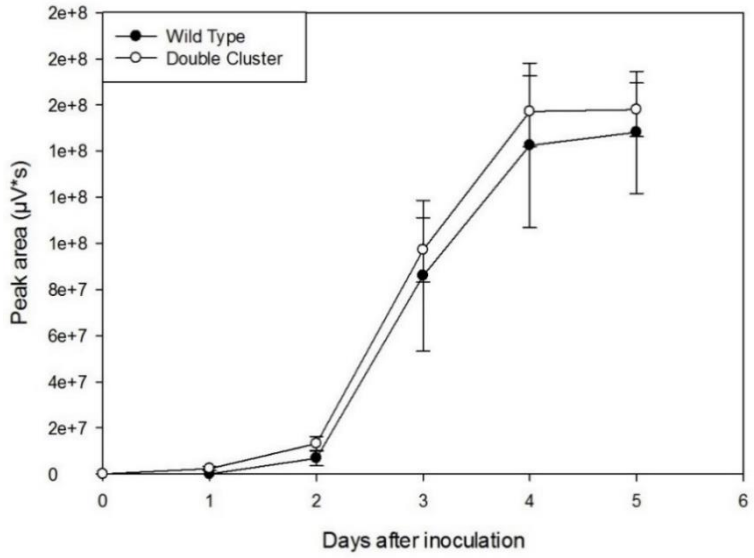


Figure 2.5: Production of sophorolipids by the wild type *S. bombicola* and the double cluster strain. For each strain four replicates were taken. The error bars represent standard deviation.

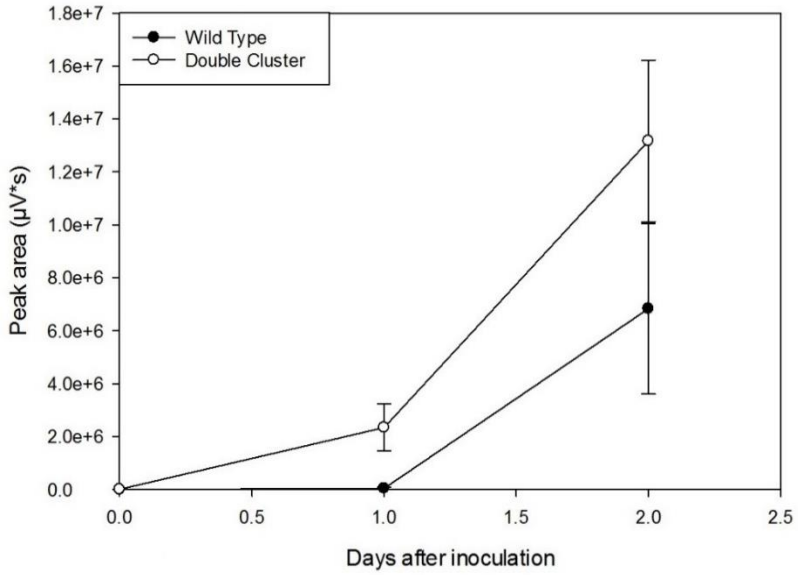


Figure 2.6: Production of sophorolipids during the exponential and early stationary phase for both wild type *S. bombicola* and the new double cluster mutant. For each strain four replicates were taken. The error bars represent standard deviation.

Statistical analysis of the HPLC chromatogram surface values indicated that this difference is significant ($p < 0.05$), yet after the 2nd day the difference is not significant anymore. It seems that the double cluster initiates production earlier or faster than the wild type, but the benefit of this earlier production is lost quickly as the rise in sophorolipid concentration is very similar between the wild type and double cluster mutant. If the extra cluster has an effect on production, it seems that this is only marginal. To see if the additional copies of the genes results in higher transcriptome or proteome levels, a follow-up experiment was performed for qPCR and MRM analysis.

II.3.3.2. Comparison of gene expression and protein levels between the wild type and double cluster strain by qPCR and MRM experiments

Transcriptome and proteome analyses were conducted to verify if the minor differences in production reflect minor differences in gene expression or protein level. This way the effect of the extra integrated cluster could be measured more directly. The samples were prepared as described in II.2.6.

The results from the qPCR experiment can be found in Figure 2.7 and those for the MRM experiment in Figure 2.8. Curiously, both experiments result in opposing data for some of the genes. In the qPCR experiment, the wild type *S. bombicola* has significantly higher relative amounts of mRNA for the acetyltransferase and the sophorolipid transporter. For *cyp52m1* and *ugta1*, no significant differences were detected. When looking to the protein level, the double cluster strain shows higher values for both the acetyltransferase and UGTA1, yet not for the transporter. This is an interesting observation since the samples for both experiments were taken at the same time.

Though no clear explanation is available, it might be that one of the methods used is not suitable yet for analysis of their respective omics domain. The extraction of both RNA and proteins relies on the efficient lysis of the yeast cells and subsequent purification of the RNA or proteins. For RNA extraction, small contaminations of ribonucleases or genomic DNA can lead to skewed results, either higher or lower. The results of the MRM experiment rely on the digestion of the proteins by trypsin. Though the sample is spiked with BSA to enable normalisation for the differences in digestion, experimental variability resulting in uneven protein amounts cannot be ruled out completely. Therefore, it is necessary to further investigate and compare the power of both analytical methods. Several strains validated in production trails that can be used for further validation of these techniques are already available at the lab. Examples are the lactone esterase overexpression strain or the *ugtb1* knock-out strain.

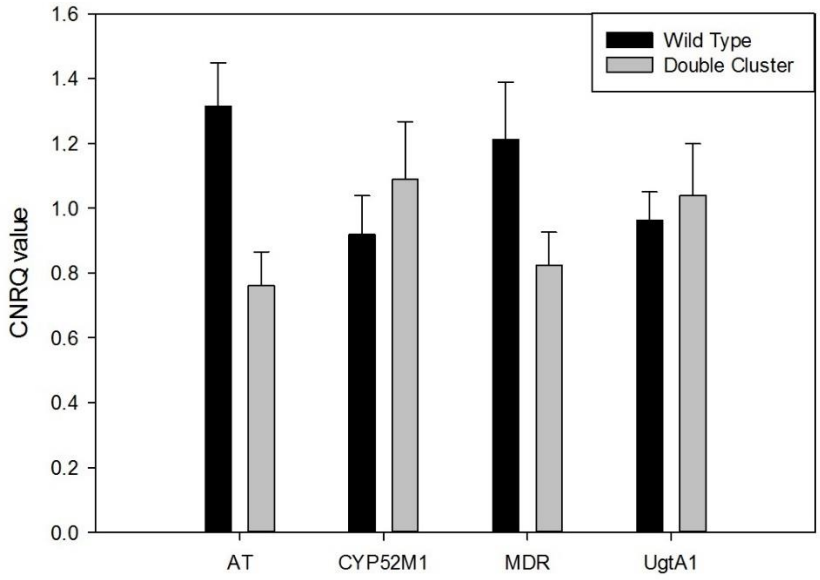


Figure 2.7: qPCR data for the cluster genes. For both the wild type and double cluster strain, the Calibrated Normalized Relative Quantities (CNRQ) values are calculated and plotted. AT = acetyltransferase, MDR = multi-drug resistant (MDR) ABC transporter involved in sophorolipid export, UGTA1 = glucosyltransferase A1 involved in sophorolipid synthesis.

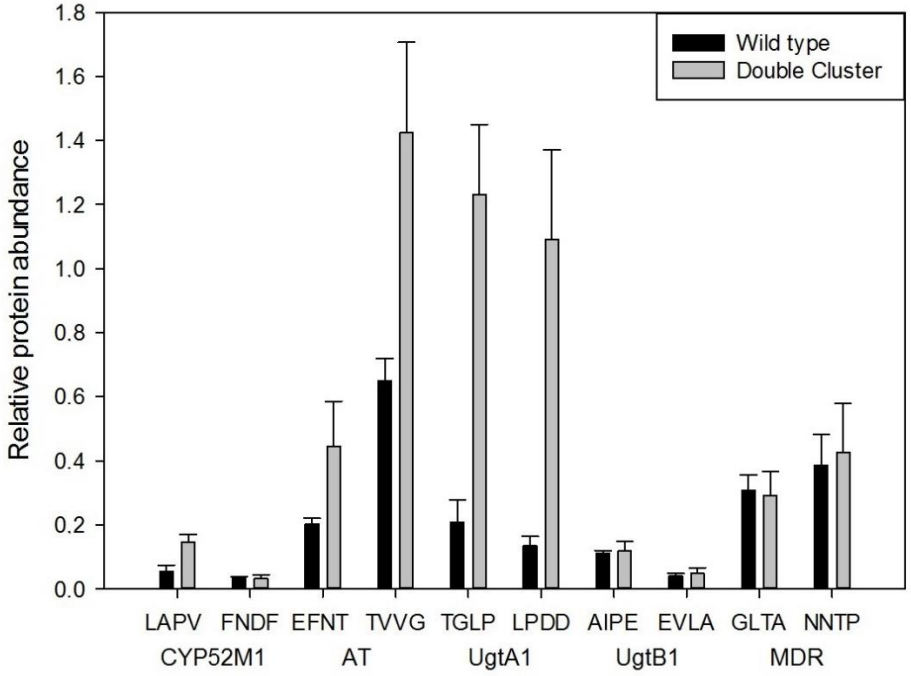


Figure 2.8: MRM data for the sophorolipid biosynthetic enzymes in both the wild type *S. bombicola* and double cluster strain. Full peptide sequences can be found in II.2.6.4 Table 2.6. AT = acetyltransferase, MDR = multi-drug resistant ABC transporter involved in sophorolipid export, UGTA1 = glucosyltransferase A1 involved in sophorolipid synthesis, UGTB1 = glucosyltransferase B1 involved in sophorolipid synthesis.

II.3.4. Localization of the sophorolipid biosynthetic enzymes

Strains expressing GFP have been developed for *S. bombicola* for some time[368]. They allowed the screening of promotors to investigate their strength and regulation. However, expanding the number of fluorescent proteins available is interesting when several targets need to be examined at the same time. This can for example be in a study of bidirectional promotors or for the localization of enzymes involved in a certain biosynthetic pathway.

In the case of the sophorolipid biosynthetic enzymes, some information is available on their location inside the cell. The P450 enzyme CYP52M1, catalysing the first step in the pathway, contains a transmembrane helix at its N-terminus. This is typical for class II P450 enzymes and necessary for anchoring them to the membrane of the endoplasmatic reticulum where their reaction partner resides. This sounds logical, since the substrate for the first transferase UGTA1 are hydroxylated fatty acids, hydrophobic compounds residing in the lipid bilayer of the membranes. Though somewhat more hydrophilic, the glucolipids used as a substrate by UGTB1 are still relatively hydrophobic and thereby potentially associated with the membranes as well. Still, *in silico* analysis done by Saerens *et al*.[381] for both transferases did not result in the discovery of signal peptides or significant hydrophobic surface regions.

As the localization of the enzymes of the sophorolipid pathway till now was only based on *in silico* analyses, it was decided to set up *in vivo* localisation experiments for these important partners in glycolipid production. Indeed, knowing the true location of activity for these enzymes is crucial in the further development of *S. bombicola* as a platform production host for tailor-made glycolipids or other interesting compounds.

II.3.4.1. CYP52M1 localization by fluorescent protein fusion.

As a first test, CYP52M1 was coupled to the already available GFPco10genes[368] and to the new blue fluorescent protein (BFP) and red fluorescent protein (RFP) (described in II.3.4.1). The chimeric enzymes consisted of the P450 coding sequence linked by the natural linker from CYP102A1 and the fluorescent protein of interest. Since the N-terminus of CYP52M1 is anchored in the membrane of the ER, only C-terminal constructs were developed.

After transformation of the different expression cassettes and screening of the mutants, microscopy and growth and production trials were conducted to localize the enzyme inside the cell and to evaluate the influence of coupling an extra domain to the CYP52M1 on its function or activity. A production experiment was conducted with the three different fluorescent strains in comparison with the wild type *S. bombicola*.

In terms of glucose consumption there is no clear difference to be detected between the fluorescent strains and the wild type yeast (Figure 2.9). For sophorolipid production, some differences can be seen between the wild type and the different fusion strains (Figure 2.10), yet the enzyme activity of CYP52M1 is at least not hindered by the presence of an extra domain at its C-terminus and its native localisation is most likely conserved. From the data it is appears that the GFP and RFP strains produce more than the wild type *S. bombicola* while the BFP produces similar levels of sophorolipids, but statistical analysis did not show a significant difference. Most likely this is due to the small sample size (the strains were only tested *in duplo*) resulting in low power of the analysis. A production trial with a higher number of replicates is needed to verify if the observed differences are indeed significant. Still, it can be concluded

that the addition of an extra domain to CYP52M1 did not result in severely impaired production. These results open the possibility of coupling other domains to CYP52M1 in order to investigate for example increasing productivity.

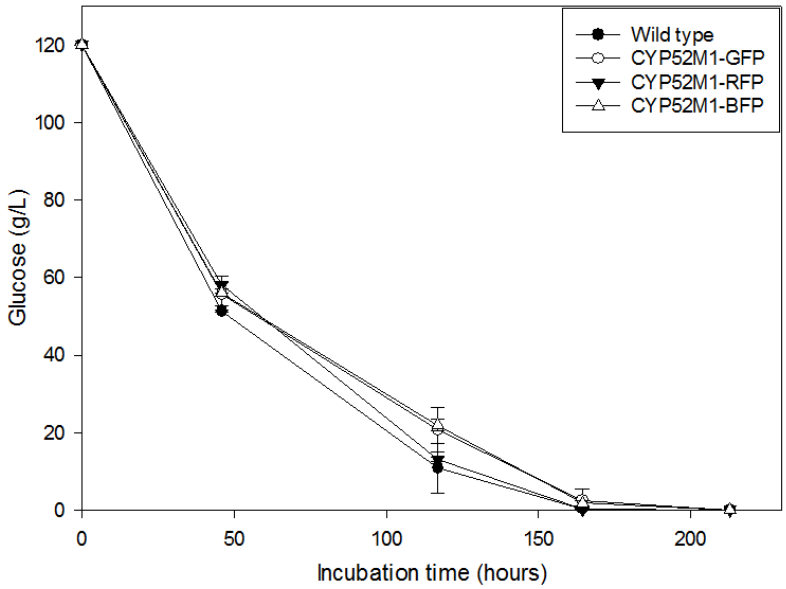


Figure 2.9: Glucose consumption of the wild type yeast and the fluorescent strains.

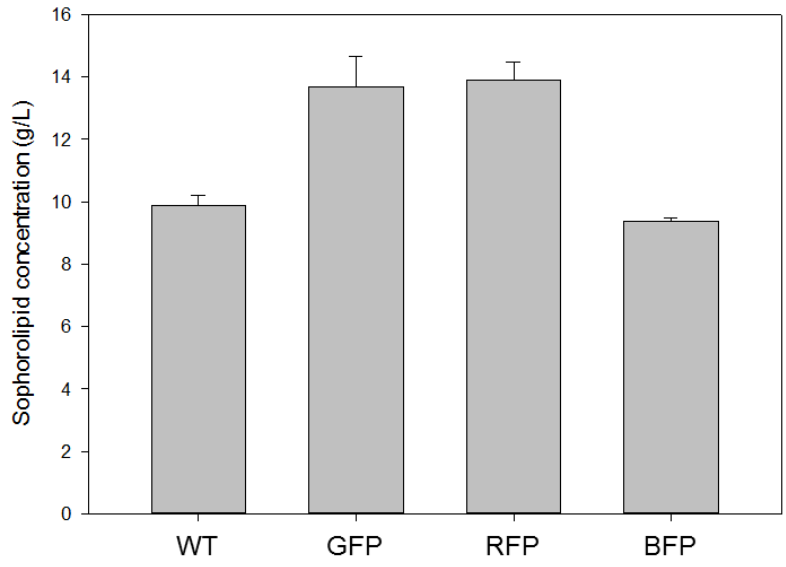


Figure 2.10: Sophorolipid production of the wild type yeast and the fluorescent strains 6 days after inoculation.

A striking visual observation is the strong colour change in the CYP52M1-RFP mutant in comparison with the other two constructs. As can be seen in Figure 2.11, this culture has a distinctive red colour. This can be explained by potentially higher emission strength of the RFP since all mutants have the same promotor controlling the fusion construct[382].

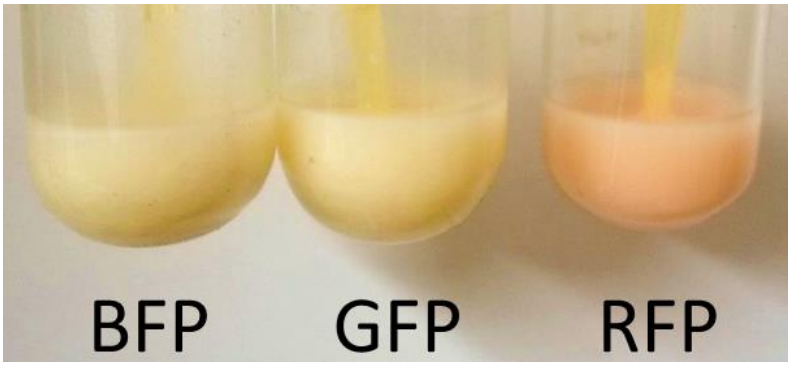


Figure 2.11: Visual difference between the CYP52M1-BFP, CYP52M1-GFP and CYP52M1-RFP strain

Using confocal microscopy, the location of the fluorescent chimeric proteins was visualized. The small size of *S. bombicola*, approximately 4 µm, limited detailed information about organelles or other structures, but as can clearly be seen in Figure 2.12, localization of the P450 is possible. The dark dot represents the nucleus. For each construct, a bright area near the nucleus can be seen. This corresponds to the endoplasmatic reticulum, which is often associated with the nucleus or at least near it[383]. Class II P450s like CYP52M1 are anchored to its membrane. These images clearly prove the localization of the P450 on these membrane structures instead of a more homogenous distribution throughout the cell. They also prove that the fluorescent proteins can be used for localization purposes in *S. bombicola*. From the three fluorescent proteins tested, RFP showed the highest signal when studied by a confocal microscope. As already mentioned, this might be due to the higher efficiency of the protein in regard to conversion of excitation energy to emitted light compared to the other fluorescent proteins tested[382].

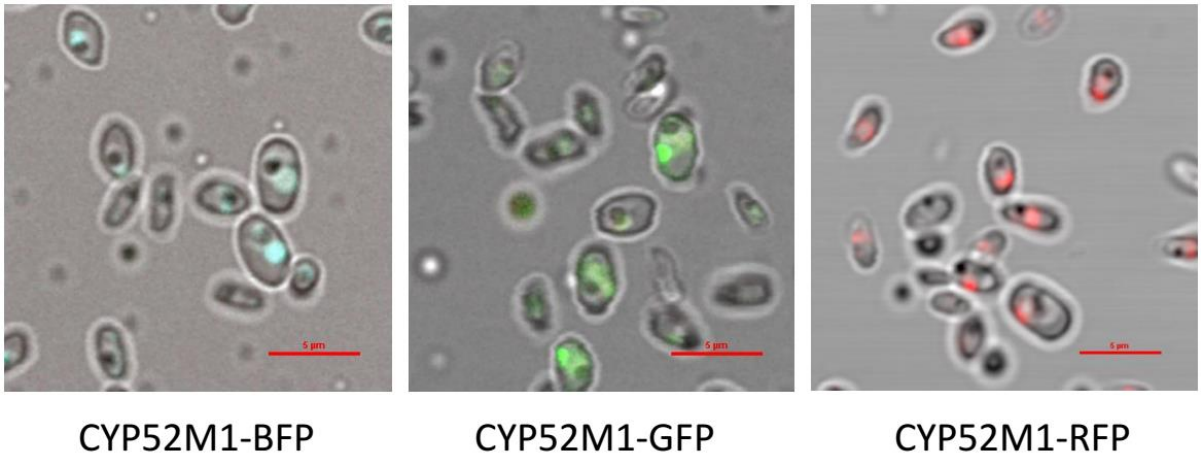


Figure 2.12: Confocal fluorescent microscopy images of the three fluorescent CYP52M1 constructs.

II.3.4.2. Localisation of Ugta1-RFP and CYP52M1-GFP

A next step in the localisation of the sophorolipid enzymes is the coupling of a fluorescent protein to the first glucosyltransferase UGTA1. Though it has been noted that the UGTA1 is present in the soluble cell lysate fractions of *S. bombicola*[175], its exact location in the cell is still unknown. To counter any potential effects on expression or enzyme activity, both C- and N-terminal fused constructs were generated. Both UGTA1-RFP fusion enzymes were introduced together with the CYP52M1-GFP chimeric construct in *S. bombicola* to investigate potential colocalisation of these enzymes (Figure 2.13).

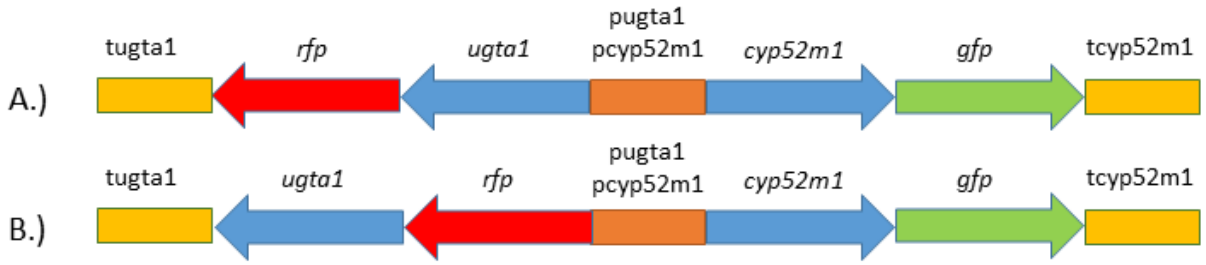


Figure 2.13: Visual representation of the two different expression cassettes for UGTA1 and CYP52M1 localization. A.) C-terminal UGTA1-RFP fusion enzyme. B.) N-terminal UGTA1-RFP fusion enzyme.

When the cultures were visually inspected, a clear difference was seen between the wild type yeast and the fusion constructs (Figure 2.14). Striking is the pink colour of the C-terminally fused RFP compared to the N-terminal RFP construct. While the C-terminal construct shows a strong pink colour compared to the wild type, the pink colour is only marginal for the N-terminal strain. Another difference between the two fusion constructs is the presence of a brown precipitate in the pink C-terminal mutant culture, probably caused by sedimenting sophorolipids. In order to assess the impact of two fusion enzymes on the sophorolipid pathway, a production trial was conducted.

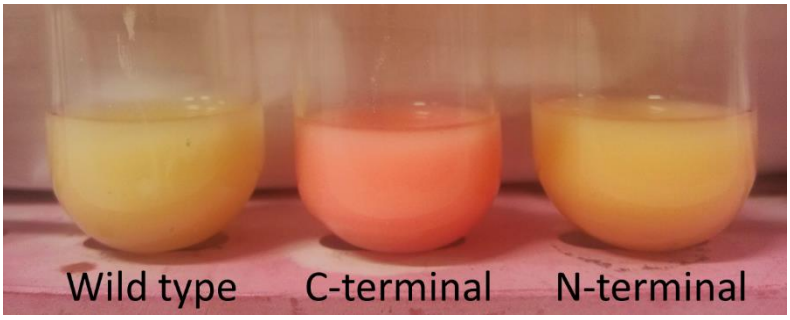


Figure 2.14: Visual observation of RFP-fusion constructs expression. From left to right wild type *S. bombicola*, C-terminal UGTA1-RFP and N-terminal RFP-UGTA1. Both fluorescent strains also had CYP52M1-GFP expressed.

As can be clearly seen in Figure 2.15, the two different fluorescent strains clearly produce different amounts of sophorolipids. While the C-terminally coupled RFP has a productivity better than the wild type *S. bombicola* ($p = 0.01982$), the N-terminal RFP strain showed no production at all. Between the wild type and the C-terminal RFP mutant, no difference in sophorolipid composition could be observed. This proves that both the GFP and RFP do not interfere in the production of sophorolipids when simultaneously coupled to CYP52M1 and the C-terminus of UGTA1, respectively.

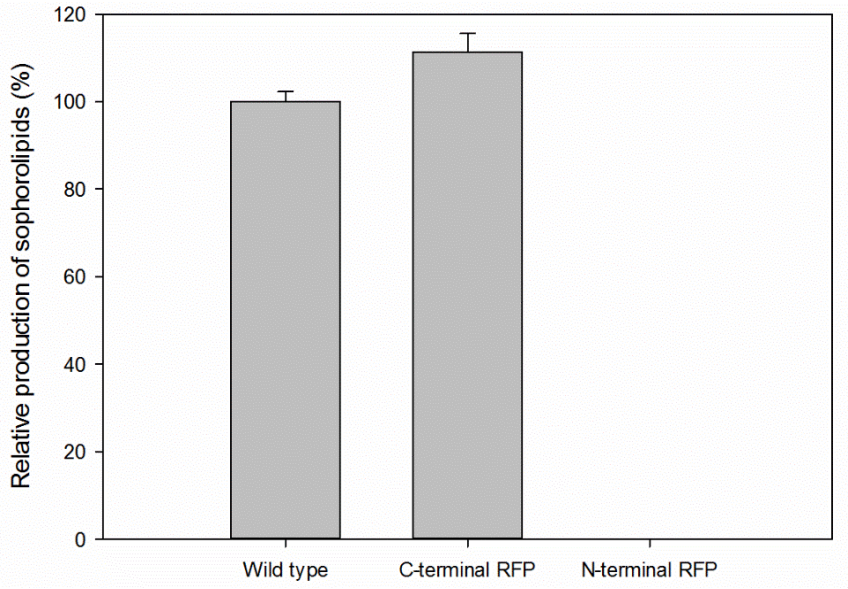


Figure 2.15: Relative production of sophorolipids for the wild type *S. bombicola*, the C-terminal RFP and the N-terminal RFP mutants.

While inspecting the different strains using a confocal microscope, the difference in fluorescence becomes very clear (Figure 2.16). In the C-terminal fused RFP strain, both the red and green fluorescent chimeric proteins are easily detected while for the N-terminal one only marginal detection of both is possible. When combining both green and red channels of the image, the resulting yellow colour suggests a colocalisation of the proteins.

When looking at the N-terminal RFP fusion construct it exhibits lower fluorescence values compared to the CYP52M1-GFP while the brightness of the RFP is higher than GFP (as could be seen for the CYP52M1-FP constructs both visually and in the imaging data). Extreme contrast adjustments show that there is expression, but this remains just above noise levels for most cells in the sample. As a control, a wild type *S. bombicola* was imaged as well with both green and red channels adjusted to the same levels as the N-terminal RFP image. No real fluorescence could be detected, confirming the observed fluorescence for the N-terminal construct being the result of protein expression.

Using the C-terminally coupled UGTA1-RFP strain, three samples were taken and imaged in both green and red fluorescence channels. Next, the resulting data was analysed using JACoP, a colocalisation plugin for ImageJ. In each analysis, the green channel was taken as channel 1 and the red channel as channel 2. Threshold determination was done automatically by the software to determine areas of interest and background. In Table 2.8 an overview of the resulting parameters can be found. Calculated colocalisation values are given both with and without using these threshold values.

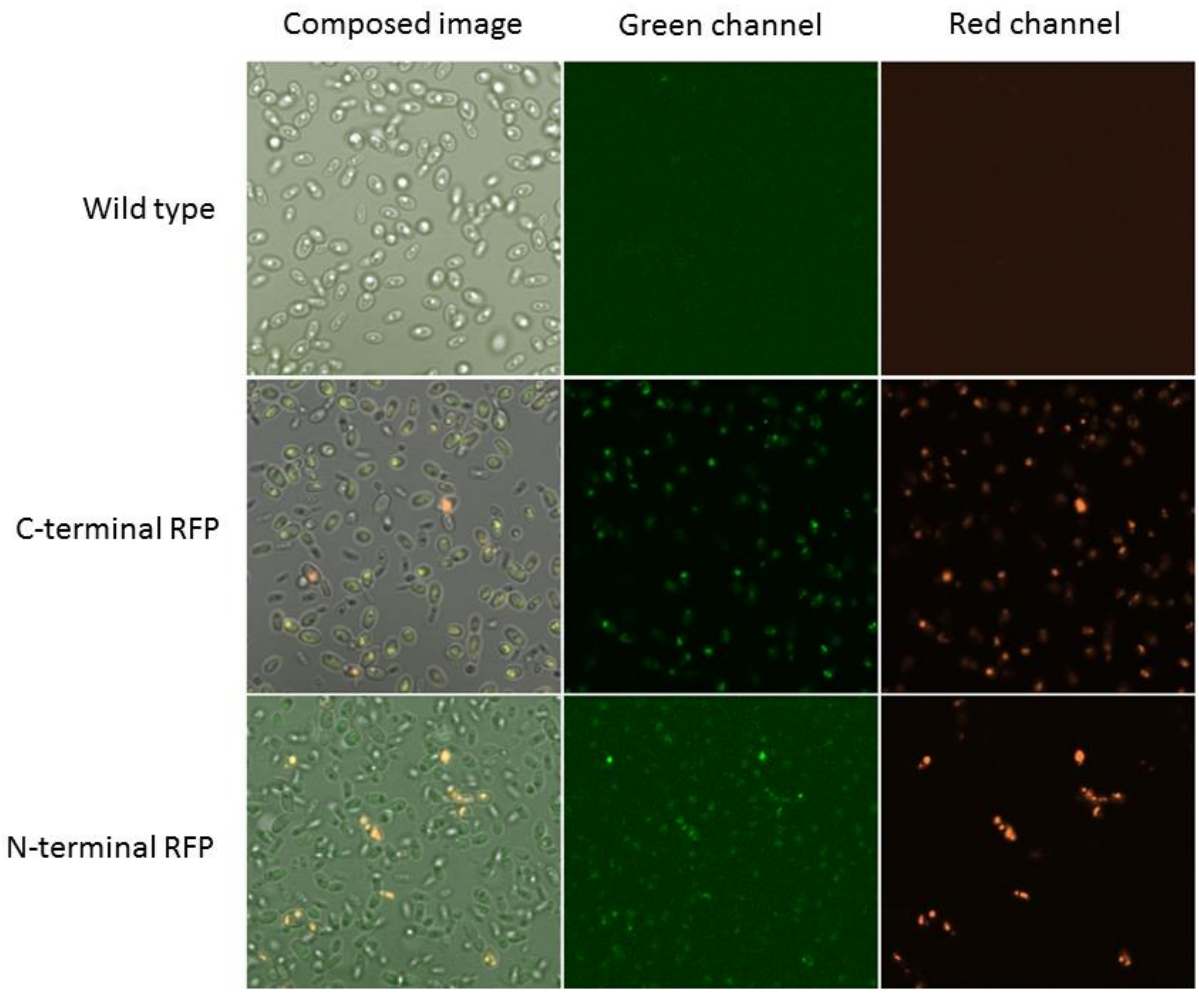


Figure 2.16: Fluorescence images of the different strains tested. Column one is the brigthfield image with overlay of the green (CYP52M1-GFP) and red (UGTA1-RFP) channels. Both colour channels are also given. The strains tested by confocal microscopy were the wild type S*. bombicola* and both UGTA1-RFP strains with the co-expressed CYP52M1-GFP.

The Pearson's correlation coefficient (PCC) calculates the covariance in pixel intensity of both channels on a pixel-by-pixel basis. Since it uses a mean pixel value calculated for both channels, which is subtracted from the recorded pixel value, this method is independent of background interference as well as dynamic range of the recorded values[384]. A value of 1 means perfect colocalisation, 0 means no colocalisation and -1 anti-localization. The Manders overlap coefficient (MOC) is similar to the Pearson's coefficient, but does not subtract a mean pixel value, this way preventing a negative value to be calculated troubling interpretation of the data. This results in values between 0 and 1, corresponding to no overlap and 100 % colocalisation. However, the MOC is highly sensitive towards co-occurrence of high value pixels while low value results or combinations of high and low values are of less importance. Also offsets in the data can heavily influence the calculated values[385]. A third coefficient is Li's colocalisation coefficient. It is based on the PCC, but instead of calculating covariance it counts the occurrence of pixels with both values above or below the mean image value. This has the downside that small differences above the mean have an equal weight in the calculation compared to strong signals, favouring pixel pairs with intensities near the mean intensity of each fluorophore used. The correlation scale runs from -0.5 for negative correlations to +0.5 for positive correlations.

Table 2.8: Calculated colocalisation values for the C-terminally fused UGTA1-RFP samples.

Parameter	Sample 1	Sample 2	Sample 3
Pearson's coefficient	0.838	0.908	0.907
Manders coefficient 1 Tr	0.768	0.895	0.899
Manders coefficient 2 Tr	0.8	0.815	0.781
Li's coefficient	0.428	0.451	0.446

When looking at the specific values of the correlation coefficients, it becomes clear that there is high correlation between the location of the CYP52M1-GFP and UGTA1-RFP. Comparing this with Figure 2.16 B, one can conclude with certainty that the P450 and glucosyltransferase are located in the same region of the yeast cell. This is presumably the membrane of the endoplasmatic reticulum since this is where the P450 is anchored to. Saerens *et al*.[175] reported that UGTA1 shows a high homology with enzymes of the GT1 family of glycosyltransferases. Most transferases of that family are bound to the membrane structures like the endoplasmatic reticulum or Golgi apparatus[386]. However, for UGTA1 no membrane anchors could be detected and cell lysate experiments without the insoluble membrane fraction seemed to prove their presence in the cell lysate[381]. Still, if the proteins would be completely soluble, one would expect to see fluorescence in the entire cell and not strictly associated with CYP52M1. Moreover, if some kind of hydrophobic region would be present to let the transferase associate with the membranes one still would expect to see fluorescence near the plasma membrane and other membrane structures.

A possible explanation for the fact that the enzymes are colocalised despite the fact that UGTA1 is not targeted to the ER (no signal present) might be P450-transferase interaction. In the case of class II P450s, interaction with the CPR enzyme is guided by electrostatical interactions on the protein surface of both enzymes. Something similar might be the case for the interaction between CYP52M1 and UGTA1. The technique used in this research does not allow to see how close the two enzymes are located, but techniques like Förster resonance energy transfer (FRET) could help to see if they are closely associated or not. Protein modelling might help as well to identify the responsible interacting regions and amino acid residues.

II.4. Conclusion

Relying on homologues integration for engineering can be a bottleneck when considering state-of-the-art engineering techniques. A plasmid would hence be a welcome addition to the molecular toolbox of *S. bombicola*. Unfortunately, no stable plasmid could be obtained using both *in silico* and *in vivo* strategies. Screening for potential ARS sequences in the genome of *S. bombicola* was unsuccessful and using heterologous ARS sequences from *S. cerevisiae* and *K. lactis* also did not result in a stable plasmid system. Interestingly, colonies were detected, but these all resulted from integration of the *ura3* marker somewhere in the genome of *S. bombicola*.

During the experiments, it became clear that the efficient recombination of the marker in the genome of *S. bombicola* is a major problem. Two causes arise when looking at the sequence of the ARS screening plasmid. One possibility is the *ura3* promotor. The promotor might be used as a guide for integration near or in the empty *ura3* locus in the genome of *S. bombicola* PT36 since only the coding sequence is missing. This problem can be eliminated by using non-homologues sequences like the *ura3* promotor off *S. cerevisiae*. Since in the experiments with the panARS the *ura3* promotor and marker of *S. bombicola* were active in *S. cerevisiae*, the other way around might be possible as well. This could greatly reduce the risk of targeting and integration in the genome of the marker in screening experiments. The remaining risk can be partially attributed to the genomic fragments inserted in the screening vector. These sequences can also be used for targeting the vector towards certain regions and facilitate recombination in or near non-essential sequences. Developing systems that don't use *S. bombicola* genomic sequences like an orthologous system involving mitochondrial DNA can cure this last problem. Though these strategies don't elevate the chances of finding a working *S. bombicola* ARS sequence, they will reduce the risk of genome integration.

When stable expression over longer periods of time is wanted, genomic integration is still the best strategy. Integration of complex cassettes in *S. bombicola* was often constrained due to uncertainties about the maximum length these cassettes can have. As could be seen from the results presented, very large constructs can be made and successfully integrated. In the case of the double cluster strain, five genes and a marker were introduced successfully and relatively easy. This opens the doors to more intensive engineering as large pathways can be introduced in one single transformation step. Indeed, there is no limit concerning the length of the final construct up to 14.000 base pairs.

The first real example of using a large expression cassette was the creation of the double cluster strain. A significant difference in sophorolipid production was found during the early stages of production when compared to the wild type *S. bombicola,* but this initial difference did not result in higher yields or productivity. Furthermore, follow-up qPCR and MRM experiments showed an inconsistency between the results from both techniques, which needs to be further investigated. An interesting observation in the workflow of the qPCR are the housekeeping genes that are taken along for normalisation between the different biological and technical replicates. In the case of the performed MRM experiments, no equivalent normalisation factor is used. Instead, BSA spiking is applied for differences in trypsin digest, but variability already present before this step is not taken into account. For qPCR, a set of housekeeping genes has been tested by Roelants *et al*.[368] A similar list could be composed for the MRM experiments[387,388] to enable more robust strain comparisons in future research.

Besides validation of both techniques, development of a strain lacking the sophorolipid gene cluster at its original location would be interesting for comparison. This way, introduction of the cluster at the *ura3* locus can be evaluated without the influence of the original one. A cluster negative strain can also be used for engineering the production of other glycolipids in *S. bombicola* since the sophorolipid biosynthetic enzymes can then not interfere in the synthesis anymore.

Finally, it was shown that coupling enzymes from the sophorolipid gene cluster to fluorescent proteins does not influence their biosynthetic activity. Though no “functional” domains where coupled to the P450 or transferase, they remained their original function in the production of sophorolipids. Furthermore, two new fluorescent proteins are available for *S. bombicola* during this research, offering the possibility of studying bidirectional promotors, protein (co)localisation or other cellular processes. The *in vivo* confirmation of the targeting of the P450 to the ER membrane and the finding that the UGTA1 and P450 are co-localized are the first results obtained by using multiple fluorescent proteins in one *S. bombicola* strain. Coupling to other enzymes like the second glucosyltransferase, the acetyltransferase or even the lactone esterase will enable to investigate their localisation and their role in the production of sophorolipids. This knowledge is crucial in the further development of *S. bombicola* as a platform production host for tailor-made glycolipids or other interesting compounds. Furthermore, the fluorescent chimera allow the rapid screening of promotors coupled to real production.

Another interesting possibility of these proteins could be identification and characterisation of localisation signals. Peroxisomal targeting is described by Roelants *et al*.[318] using the signal peptide present at the C-terminus of the peroxisomal multifunctional enzyme type 2[314]. Other signal peptides like nuclear, membrane, Golgi or mitochondrial localisation signals are still unknown or untested. For certain applications these sequences might be necessary, for example when integrating the 2µ plasmid maintenance proteins that need to be located inside the nucleus.

Chapter III

Engineering the P450 pool of S. bombicola

Chapter III - Engineering the P450 pool of *S. bombicola*

III.1. Introduction

Starmerella bombicola is a yeast known for its efficient production of sophorolipids, a kind of glycolipid biosurfactant[389]. The first step in the sophorolipid biosynthesis involves hydroxylation of a fatty acid by a cytochrome P450[39]. Cytochrome P450 enzymes are hemeproteins known for their capacity to use molecular oxygen to be incorporated in a wide variety of substrates. Besides oxygenation reactions, other less common reactions like nitration occur as well[390].

In the sophorolipid biosynthesis, hydroxylation of the fatty acids is carried out by CYP52M1 and it largely determines the length of the fatty acids introduced[315]. The enzyme has a high specificity towards fatty acids like stearic acid (C18:0) and oleic acid (C18:1), less towards shorter and longer fatty acids like palmitic (16:0) or myristic (14:0) acid. In combination with the specificity of the first glucosyltransferase UGTA1, this results in mainly C16:0, C18:0 and C18:1 fatty acid tails being incorporated. Therefore, most production processes are conducted with hydrophobic substrates like vegetable oils or longer fatty acids. Though the incorporation of these longer fatty acids is very efficient, it is also a burden when further molecular diversification of the sophorolipid molecules is desired. It is feasible to incorporate shorter alkyl chains by chemical or biocatalytical ways[363], *in vivo* production is still limited.

Growth trials with shorter fatty acids resulted mainly in long-chained sophorolipids and some smaller amounts of medium-chained ones. This is due to the elongation of the shorter fatty acids by elongase enzymes of the endoplasmatic reticulum. Shorter fatty acids can also be degraded in the β-oxidation pathway and those building blocks will serve as a substrate for *de novo* synthesis of fatty acids[391]. However when using specialized substrates like alcohols or ketones, short- and medium-chained sophorolipids can be obtained[392,393]. The main problems are still the high cost of for example ketones and the relatively low amounts of novel sophorolipids being produced. Engineering of *S. bombicola* by knocking-out the β-oxidation[314] or enzymes involved in long-chain alcohol oxidation[394] already improve the production of these novel new-to-nature molecules using cheaper substrates, but the traditional sophorolipids still remain a part of the mixture since *de novo* fatty acid synthesis is still occurring with subsequent incorporation of these fatty acids in the sophorolipids.

All in all, the major factors hindering the production of new-to-nature molecules are *de novo* sophorolipid synthesis, high substrates costs and less than optimal production. To circumvent this problem of elongation, degradation and *de novo* incorporation, one may try to express P450s with interesting substrate specificities, for example CYP52A4[249], CYP102A7[293], $P450_{foxy}$[295] or CYP1[395]. CYP52A4, CYP102A7 and $P450_{foxy}$ have a high activity towards myristic acid (C14:0) and CYP1 presumably towards palmitic acid (C16:0). All these P450s perform hydroxylations near the terminal or the terminal carbon atom of the fatty acid tail.

As already mentioned, natural P450s exhibit a wide range of potential reactions, substrates used and products formed. Still, research has focused on engineering specific P450 enzymes for studying for example drug metabolism. One such enzyme that has been studied extensively is CYP102A1 from *Bacillus megaterium*[275] also known as BM3. To date it is the most active P450 described in nature mostly due to the fact it does not depend on a separate reaction partner like most other P450 enzymes.

Structural analysis as well as mutagenic research on the enzyme proved that it is possible to shift substrate specificity and product formation when specific amino acid residues are altered[396–398]. Another technique used to widen the reaction horizon of CYP102A1 is so-called domain shuffling between the different CYP102A family members or mutants of CYP102A1 itself. The resulting synthetic chimeric P450s often have novel properties not directly linked to the parental enzymes[399,400]. To explore the full potential of other P450s, synthetic chimeric constructs have been constructed[401].

In this chapter, the creation of *S. bombicola* strains capable of producing sophorolipid mixtures not yet described by heterologous expression of wild type and engineered P450 enzymes and moreover enhancing this production by creation of chimeric P450s is described.

III.2. Material and Methods

III.2.1. Strains, plasmids and culture conditions

Starmerella bombicola ATCC22214 is the wild type strain used during the experiments. The derived PT36 strain, a *ura3* auxotrophic mutant, was used to engineer all the mutants described in this research. Transformation and growth trials were performed as described in II.2.1. As a hydrophobic substrate, rape seed oil was used at 37.5 g/L or another suitable hydrocarbon like ethyl palmitate, palmitic acid, myristic acid, ethyl myristate, lauric acid or ethyl myristate at concentrations of 5 g/L in Erlenmeyer experiments unless otherwise stated. All the different substrates were acquired from Sigma-Aldrich (USA)

Experiments in bioreactors were routinely carried out using the Biostat® B 3 L vessels (Sartorius BBI-systems; Germany) with 1.5 liter working volume. At inoculation, 1 liter of Lang-medium[370] was inoculated with 200 mL of a late exponential culture from an Erlenmeyer flask. The medium has an initial pH of 5.8 and is allowed to drop spontaneously to pH 3.5 and is kept at this pH by using a 5 M NaOH solution. Due to the acidification of the culture broth by the yeast itself, no acid has to be added. The reaction vessel is kept at 30 °C and stirred at 600 rpm. Aeration is kept at 1 L/min and closely monitored. Daily sampling of glucose concentrations and optical density is performed as stated in II.2.4.1. When the glucose concentration drops below 30 g/L, a precalculated volume from a 60 % glucose stock solution is added to the fermentor until a final concentration of 100 g/L. Larger fermentations were carried out in the 30 L B. Braun Biotech International C20-3 (Germany). Aeration and stirrer rates can be found in Table 3.8 Recuperation of the sophorolipids from the fermentation broth was done by using two techniques depending on the strain used. Isolation from highly productive strains was done by placing the culture broth overnight at 60 °C. This results in a precipitation of the sophorolipids produced. Recovery of sophorolipids from low productive strains was done by extraction with ethyl acetate.

III.2.2. Molecular Techniques

III.2.2.1. General techniques

High fidelity PCRs are performed with PrimeSTAR HS DNA Polymerase (Takara) when fragments are shorter than 3000 base pairs. Longer fragments are amplified with PrimeSTAR GXL DNA Polymerase (Takara). For colony PCR, TAQ-polymerase (New England Biolabs) is used. Genomic DNA was isolated using the Sigma Genomic DNA isolation kit.

Primers, plasmids and final strains were designed using the Clone Manager Basic Version 9 software suite from Sci-ED Software. Primers were ordered at Integrated DNA Technologies and diluted upon arrival to a 10 mM stock solution. Working solutions are derived from these by a 100x dilution to 10 µmol. Sequencing of plasmids and intermediate constructs was performed by LGC Genomics (Germany).

Transformation of *S. bombicola* was done following the protocol of Van Bogaert *et al*.[310]. Unless otherwise stated, cells were incubated on the previously described SD-medium. Transformation of *E. coli* was performed as described by Sambrook and Russel[371].

III.2.2.2. Vector construction

All the different plasmids constructed in this chapter are targeted towards the *cyp52m1* locus (Figure 3.). The P450s of interest are ligated between the *cyp52m1* promotor and terminator. Selection is performed by the *ura3* marker, which is situated downstream of the P450 coding sequence. In Table 3.1, an overview is given, which P450s were tested during this research.

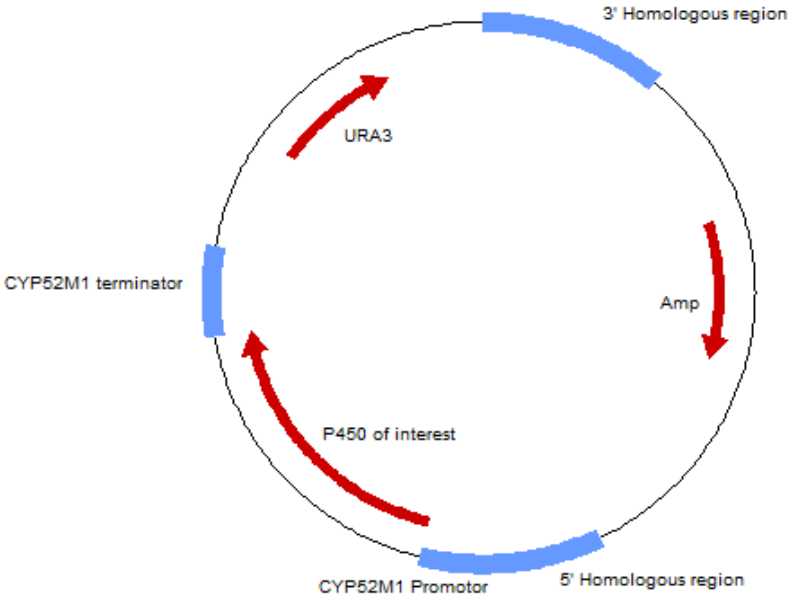


Figure 3.1: Overview of the plasmid used for knock-in strategies at the *cyp52m1* locus.

Table 3.1: P450 enzymes tested at the *cyp52m1* locus for the production of medium-chained sophorolipids.

P450 name	Organism	Coding sequence length	Preferred substrate(s)
CYP52A4	*Candida maltosa*	1617 bp	Lauric acid
CYP52M1	*Starmerella bombicola*	1617 bp	Oleic acid
CYP1	*Ustilago maydis*	2007 bp	Presumed palmitic acid
CYP102A1	*Bacillus megaterium*	3150 bp	Myristic acid
CYP102A7	*Bacillus licheniformis*	3225 bp	Myristic acid
$P450_{Foxy}$	*Fusarium oxysporum*	3201 bp	Lauric acid

Wild type CYP102A1, CYP102A7 and $P450_{Foxy}$ enzymes

Plasmid BS_EC_0360 was used for constructing the novel knock-in cassettes. The plasmid was constructed by using a pGEMT backbone for cloning in *E. coli* and contains homologues regions for incorporation in the genome of *S. bombicola* at the *cyp52m1* locus. The 5' homologous region encompasses 1000 base pairs upstream of the start codon of *cyp52m1*. Downstream of the locus are the *ura3* marker gene with its promotor and terminator followed by the 3' homologous region. The plasmids were constructed by linearizing the previous mentioned *cyp52m1* locus plasmid by high fidelity PCR (8228 bp). This results in a scarless vector backbone suitable for assembly cloning of the coding sequences of the to be tested P450s.

The coding sequence of CYP102A1 (3150 bp) and CYP102A7 (3225 bp) were amplified from genomic material from *Bacillus megaterium* ATCC14581 and *Bacillus licheniformis* LMG12363 respectively with the primers provided in Table 3.2. A similar strategy was used for creation of

the $P450_{Foxy}$ from *F. oxysporum* JCM11502. The coding sequence of $P450_{Foxy}$ was obtained by RNA isolation from a stationary cell culture using the QIAGEN RNeasy kit to avoid introns. After reverse transcription, amplification of the cDNA resulted in the expected PCR product (3201 bp). Ligation in the vector backbone was done by Gibson assembly for all sequences. Screening of the correct *E. coli* mutants by colony PCR is done with primers P843_A21knockhygroCasFor and P79_REV_tCYP_extBamHI. The resulting plasmids are called BS_EC_0358, BS_EC_0360 and BS_EC_0362 for respectively CYP102A1, CYP102A7 and $P450_{Foxy}$.After sequencing, the correct expression cassettes were amplified with primers P847_Cm2ExCasLongFor and P947_Cm2ExCasRev and transformed in *S. bombicola* PT36. Screening of the correct mutant yeasts happens with primer P954_Cm2CheckFor combined with P1231_BM3F87Arev, P243_CYP102A7checkrev, and P242_FoxyP450checkrev for respectively CYP102A1 (1335 bp), CYP102A7 (1545 bp) and $P450_{Foxy}$ (1642 bp) for the 5'-region. For the 3' region, only one primer pair is needed for screening all mutant strains; P953_URA3outendFor and P952_A21totRev (1172 bp).

Mutant CYP102A1F87A, CYP102A7F88A and $P450_{Foxy}$F88A enzymes

A previously acquired CYP102A1F87A plasmid was used as a template for amplification of the CYP102A1F87A insert. The plasmids described in III.2.2.2 for CYP102A7 and $P450_{Foxy}$ expression were used for mutagenesis of the phenylalanine at position 87-88 by Sanchis site saturation mutagenesis with primers P2086_CYP102A7mutF88ASanFor and P2089_FoxyP450F88AsanFor. As reverse primer their respective Gibson reverse primers were used for generation of the megaprimer. The resulting plasmids are BS_EC_0359, BS_EC_0361 and BS_EC_0363 for CYP102A1F87A, CYP102A7F88A and P450FoxyF88A. The same primers described in the previous section were used for screening correct yeast mutants.

Table 3.2: Primers used for the construction of wild type and mutated P450 expression plasmids. When overhanging primers were used, a dash represents the start of the binding region of the primer. For Sanchis primers the altered codon is marked in bold.

Primer name	Primer sequence
P52_REV_upCYP	ATATGTACTTTTCAATATGATAAACGGAGAAATAACG
P69_FOR_RegulUP	GTTTCTTAGCCTCCCATGGAAG
P79_REV_tCYP_extBamHI	CAGGATCCGTCGGTTAAACGCACTCCTTC
P395_UmCYP1InfFor	TGAAAAGTACATATGGTACC-ATGTCGCTCAAAGTGCAG
P396_UmCYP1InfRev	AACGCTAGCTTGGCG-TTATCTAGTGCCTTCCTTGCAAC
P843_A21knockhygroCasFor	GAGTCGGGCGTTATTTCTCC
P847_Cm2ExCasLongFor	CGTTGTCAAGTCCTAAGGTAT
P887_Cm2ExCasRev	AAGCGTGAAGCTCCTCTGACAATC
P954_Cm2CheckFor	TCATAGCGAGTTTCTTTGCATGTG
P679_UmCYP1checkRev	TGAGCTAAGTCCACCCTGAG
P953_URA3outendFor	TAAAGAAACGAAGGGCCCAGCAGTC
P952_A21totRev	GCTCTTGTTCGGTACTCTTATT
P1231_BM3F87Arev	GCGTCCAGCTTGTAGCTAACCCGTCTCCTGC
P2086_CYP102A7mutF88ASanFor	CGAGTTCAGCGGAGACGGGCTC**GCA**ACAAGCTGGACGGAAGAACCCAATTGG
P2089_FoxyP450F88AsanFor	GAGGGTGTTCACGATGGCCT**CGC**AACAGCATTCGAGGATGAGCCCAATTG
P2090_FoxyP450GibFor	GTTATTTCTCCGTTTATCATATTGAAAAGTACATAT-ATGGCTGAATCTGTCCCTA
P2091_FoxyP450GibRev	GAACGTTTCTTCCATGGGAGGCTAAGAAAC-CTAATCGAAAACATCAGTAGCA
P2092_CYP102A7GibFor	GTTATTTCTCCGTTTATCATATTGAAAAGTACATAT-ATGAACAAGTTAGATGGAATTCCAATC
P2093_CYP102A7GibRev	GAACGTTTCTTCCATGGGAGGCTAAGAAAC-TTACGTTCTTGCAATGCAGTC
P2094_CYP102A1GibFor	GTTATTTCTCCGTTTATCATATTGAAAAGTACATAT-ATGACAATTAAAGAAATGCCTCAG
P2095_CYP102A1GibRev	GAACGTTTCTTCCATGGGAGGCTAAGAAAC-TTACCCAGCCCACACGTC

Ustilago maydis CYP1

The wild type coding sequence of *cyp1* from *U. maydis* 521[402] was ordered from GeneArt. No codon optimisation was performed since the codon usage was similar to the codon usage of *S. bombicola* during the stationary phase. The sequence was amplified with primers P395_UmCYP1InfFor and P396_UmCYP1InfRev (2041 bp). The previous mentioned pGEMT_CYP102A7 plasmid was used as a vector backbone (7514 bp). Ligation was performed by the In-Fusion HD Cloning kit. This resulted in plasmid BS_EC_0058. After sequencing, the expression cassette was amplified from the plasmid with primers P847_Cm2ExCasLongFor and P947_Cm2ExCasRev (6512 bp). After transformation in *S. bombicola* PT36 and selection on SD-medium, the transformants were screened for correct integration of the cassette by yeast colony PCR using primers P954_Cm2CheckFor and P679_UmCYP1checkRev for the 5'-region and P953_URA3outendFor and P952_A21totRev for the 3' region. The 5' region fragment has a length of 1116 bp and the 3' region fragment of 1172 bp.

CYP1, CYP52M1 and CYP52A4 chimeric constructs

The earlier constructed plasmid BS_EC_0058 was used as a vector backbone to insert the *Bacillus megaterium* BM3 reductase. The vector backbone was amplified with primers P788_UmCYP1scarlessBMRInfRev and P69_FOR_RegulUP (8558 bp) (Table 3.3). This resulted in plasmid BS_EC_0091. The other plasmids are derived from the latter by amplification of the vector backbone (including the BM3 linker) with primers P1439_BMRlinkerRev and P69_FOR_RegulUP (8618 bp). The reductase of *B. megaterium* ATCC14581 was amplified from genomic material with primers P722_BM3CPRLinkerInfFor and P723_BM3CPRLinkerInfRev (1794 bp) and ligated using the In-Fusion HD Cloning Kit. The *S. bombicola* CPR was amplified from genomic DNA with primers P1440_*Sbombicola*CPRGibCYP52M1For and P1441_*Sbombicola*CPRGibCYP52M1Rev (1992 bp) and was assembled in the vector backbone using Gibson assembly. The expression cassettes were amplified with primers P847_Cm2ExCasLongFor and P887_Cm2ExCasRev (respectively 7310 bp, 7583 bp and 7691 bp).

Chimeric enzymes were also constructed for CYP52M1 and CYP52A4. The coding sequence of *cyp52m1* was amplified from genomic *S. bombicola* DNA with primers P834_Cyp52M1GibFor and P829_Cyp52M1BMRGibRev (1674 bp). The template of CYP52A4 was a plasmid kindly shared by Zimmer *et al*[249] named Cm2_correct_pl4. Primers P789_Cyp52A4scralessInfFor and P790_Cyp52A4scarlessBMRInfRev (1644 bp) were used for amplification of the sequence. CYP52M1 was combined with the reductase part of BM3 (BMR) or the *S. bombicola* CPR while CYP52A4 was only coupled to BMR.

Validation of CYP1BMR chimeric construct

Validation of the function of the reductase domain of *B. megaterium* ATCC14581 was done by introducing several mutations in the domains of the reductase. Part of the reductase was ordered as a gBlock from IDT (Leuven, Belgium) since PCR-based site-directed mutagenesis was impossible to achieve. The sequence can be found in Appendix C. For the mutation in the FMN domain, primers P1645_BMRmutgBlockMidGibFor and P1642_BMRmutbbGibgBlockRev were used for the vector backbone (9657 bp) with plasmid BS_EC_0091 as a template (Table 3.3). The mutated sequence was amplified from the gBlock with primers P1644_BMRmutblockGibgBlockFor and P1646_BMRmutgBlockMidGibRev (761 bp). A similar strategy was followed for the mutation in the NADPH binding domain with primers P1642_BMRmutbbGibgBlockRev and P1645_BMRmutgBlockMidGibFor for vector the backbone with BS_EC_0091 (9766 bp) as a template and primers P1643_BMRmutblockGibgBlockRev and P1645_BMRmutgBlockMidGibFor for the mutated sequence (645 bp). Finally, the double mutation construct was created with primers P1641_BMRmutbbGibgBlockFor and P1642_BMRmutbbGibgBlockRev for linearization of the vector BS_EC_0091 (9043 bp) and primer pair P1643_BMRmutblockGibgBlockRev and P1644_BMRmutblockGibgBlockFor for the sequence carrying the double mutation (1372 bp). All vectors were assembled by Gibson assembly. Transformation in *S. bombicola* PT36 and screening of the mutants is identical as described in II.2.2.

Table 3.3: Primers used for construction of the various chimeric constructs and for the validation cassettes of the CYP1BMR construct. Binding and non-binding parts of the primers are separated by a dash.

Primer name	Primer sequence
P52_REV_upCYP	ATATGTACTTTTCAATATGATAAACGGAGAAATAACG
P69_FOR_RegulUP	GTTTCTTAGCCTCCCATGGAAG
P722_BM3CPRLinkerInfFor	AAAATTCCGCTTGGCGGTATTCCTTCACC
P723_BM3CPRLinkerInfRev	GGGAGGCTAAGAAAC-TTACCCAGCCCACACGTCT
P788_UmCYP1scarlessBMRInfRev	GCCAAGCGGAATTTTTC-TAGTGCCTTCCTTGCAACATCGAAGGG
P789_Cyp52A4scarlessInfFor	TTGAAAAGTACATAT-ATGTCGGTTTCCTTTGTTCAC
P790_Cyp52A4scarlessBMRInfRev	GCCAAGCGGAATTTT-GTACATTTGGATATTGGCACCATC
P829_Cyp52M1BMRGibRev	AGGTGAAGGAATACCGCCAAGCGGAATTTT-AGAAAACCGCACAACCACACCGGGAGGAG
P834_Cyp52M1GibFor	TCTCCGTTTATCATATTGAAAAGTACATAT-ATGTTAATCAAAGACATTATTCTAACTCCAATGAG
P847_Cm2ExCasLongFor	CGTTGTCAAGTCCTAAGGTAT
P887_Cm2ExCasRev	AAGCGTGAAGCTCCTCTGACAATC
P1439_BMRlinkerRev	GTTTTCTGCCTTTTTGCGTAC
P1440_*Sbombicola*CPRGibCYP52M1For	GCTAAAAAAGTACGCAAAAAGGCAGAAAAC-GCCGATATTAATTTTATCGCT
P1441_*Sbombicola*CPRGibCYP52M1Rev	GAACGTTTCTTCCATGGGAGGCTAAGAAAC-CTACCAAACGTCCTCTTG
P1641_BMRmutbbGibgBlockFor	CGGAGACGGAAGCCAAATGGCACCTGCCGTTGAAGCAACG
P1642_BMRmutbbGibgBlockRev	CAAATACGGAGTAGCGAACGCCTTTTACTTCATCAGCA
P1643_BMRmutblockGibgBlockRev	CGTTGCTTCAACGGCAGGTGCCATTTGGCTTCCGTCTCCG
P1644_BMRmutblockGibgBlockFor	TGCTGATGAAGTAAAAGGCGTTCGCTACTCCGTATTTG
P1645_BMRmutgBlockMidGibFor	GCAAAACGTTTAACAATGCTTGAACTGCTTGAAAAATACC
P1646_BMRmutgBlockMidGibRev	CCGGGTATTTTTCAAGCAGTTCAAGCATTGTTAAACGTTTTGCC

Substrate specificity screening of CYP1BMR

For testing the substrate specificity of the chimeric CYP1BMR, the V60 vector system for *S. cerevisiae* W(R) was used for overexpression. The three parts of the V60 vector are obtained using the following primers (Table 3.4); P613_pYeDP60_Gibson_Piece1_Rev and P1032_V60Piece1For (4752 bp) for V60Piece1 and P614_pYeDP60_Gibson_Piece2_For and P1033_V60Piece2Rev (4498 bp) for V60Piece2. The primers used to obtain the coding sequence of the CYP1BMR are P2005_CYP1BMR_V60GibFor and P2006_CYP1BMR_V60GibRev (3857 bp). Final assembly of the plasmid was performed *in vivo* in *Saccharomyces cerevisiae* WT 630. For identification of the positive colonies in *S. cerevisiae* W(R) by colony PCR, the primers P274_V60insertRev and P273_V60insertFor were used (4094 bp).

Table 3.4: Primers used for construction of the V60 substrate screening plasmid. Binding and non-binding regions are separated by a dash.

Primer name	Primer sequence
P273_V60InsertFor	TGCATGTGCTCTGTATGTA
p274_V60InsertRev	TTTAGCGTAAAGGATGGG
P613_pYeDP60_Gibson_Piece1_Rev	TTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAAC
P614_pYeDP60_Gibson_Piece2_For	TTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAAC
P1032_V60Piece1For	GATCTCCCATGTCTCTACTGGTGGTGGTGCTTCTTT
P1033_V60Piece2Rev	TAATTTAGTGTGTGTATTTGTGTTTGCGTGTCTA
P2005_CYP1BMR_V60GibFor	CACGCAAACACAAATACACACACTAAATTA-ATGTCGCTCAAAGTGCAG
P2006_CYP1BMR_V60GibRev	GCACCACCACCAGTAGAGACATGGGAGATC-TTACCCAGCCCACACGTC

III.2.3. Isolation of yeast microsomes

The specificity with respect to fatty acids of the membrane-bound cytochrome P450 enzymes can be assayed by adding the substrate of interest to isolated microsomes. The *S. cerevisiae* W(R) strain is used in order to overexpress the cytochrome P450 enzyme of interest. *S. cerevisiae* does not contain endogenous hydroxylation activity to fatty acids. Furthermore, the *S. cerevisiae* cytochrome P450 reductase is overexpressed after induction with galactose. The following method is used for a culture containing a total of $5x10^9$ cells/mL. All components should be scaled up corresponding to culture size. For isolation of the microsomes, the protocol of Pompon *et al*.[403] was followed. To determine the protein concentration of obtained samples the Pierce™ BCA Protein Assay Kit of Thermo Scientific™ (product number #23225) is used. Protein concentrations are determined and reported with reference to standards of the common protein bovine serum albumin (BSA).

III.2.4. Activity determination of microsomes

III.2.4.1. Checking for the presence of microsomes

When the microsomes are isolated, one needs to check if the microsomes are present in the isolated protein mixture. This is done by a protocol where cytochrome c is reduced by the present cytochrome P450 reductase enzymes. This is accompanied by a colour change that can be observed at 550 nm. The stock solutions are made using distilled water. The composition of the reaction mixture can be found in Table 3.5. $NADP^+$ is added last to the

reaction mixture, since this component allows for the start of the reaction. The regeneration of NADPH is based on an enzymatic reaction. The glucose-6-phosphate dehydrogenase, present in the reaction mixture, will convert the oxidized $NADP^+$ back into NADPH in the presence of glucose-6-phosphate. NADPH then serves as an electron donor for the CPR. All components are acquired from Sigma-Aldrich (USA) unless otherwise stated.

Table 3.5: Composition of the reaction mixture to check for the presence of microsomes

Components	Volume (µL)
Protein (10 mg/mL)	100
K_2HPO_4 (2M; pH 7.4)	125
$MgCl_2$ (1 M)	3.3
Glucose-6-phosphate (60 mM)	55
Glucose-6-phosphate dehydrogenase (100 U/mL)	4
$NADP^+$ (100 mM)	13
Cytochrome c (9 mg/mL)	100
mQ	599.7
Total volume	**1000**

III.2.4.2. Cytochrome P450 activity protocol

To analyse the activity of the cytochrome P450, added to different substrates, the following reaction mixture is used. The stock solutions are prepared with distilled water, except for the fatty acids. For these, stock solutions are made in 100 % dimethylsulfoxide (DMSO). The composition of the reaction mixture for 1 substrate is given in Table 3.6. In order to start the reaction $NADP^+$ is added last. Everything is incubated at 30 °C and 120 rpm for 3 hours. After incubation the reaction is stopped by adding 0.5 mL 0.8 % (vol/vol) H_2SO_4. The formed products are extracted with ethyl acetate. This was done by adding 440 µL ethyl acetate and 10 µL acetic acid to the reaction mixture with subsequent vigorous mixing. After 5 minutes, the mixture is centrifuged at full speed for 5 minutes. After centrifugation, the top layer is collected for further analysis.

Table 3.6: Composition of the reaction mixture to screen for cytochrome P450 activity.

Components	Volume (µL)
Protein (10 mg/mL)	100
K_2HPO_4 (2M; pH 7.4)	125
$MgCl_2$ (1 M)	6.6
Glucose-6-phosphate (60 mM)	110
Glucose-6-phosphate dehydrogenase (100 U/mL)	8
$NADP^+$ (100 mM)	26
Substrate (100 mM)	2
mQ	622.4
Total volume	**1000**

III.2.5. *S. bombicola* self-sufficient P450 activity tests

Cell lysates of *S. bombicola* were generated as described by Saerens *et al*.[176] After determination of the protein concentration with the Pierce™ BCA Protein Assay Kit (ThermoFischer), the same reaction mixture described in Table 3.6 was used for activity measurements of the self-sufficient enzymes. Extraction of the products is performed as in 0.

III.2.6. Sampling and analysis

Determination of glucose, optical density and sophorolipid production were performed as described in II.2.4. The products of the CYP1BMR and *S. bombicola* self-sufficient P450 enzyme tests were analysed with a scan range from 100 to 400 m/z by LC-MS.

III.3. Results and discussion

As mentioned before, the physico-chemical and biological properties of surfactants are partially influenced by their hydrophilic-lipophilic balance (HLB). In the case of sophorolipids, the HLB value can be modified by changing the fatty acid tail to either shorter or longer congeners. In the past, sophorolipids differing from the wild type mixture have been made using atypical substrates like alcohols, ketones or related substrates[392,393,404]. Still, wild type sophorolipids often contaminate the product mixture leading to lowered uniformity. As can be seen in Figure 3.1, a simple substrate such as ethyl palmitate will be elongated to oleic and stearic acid or degraded in the β-oxidation and used for *de novo* fatty acid synthesis. Another interesting observation is the double signal for the lactonic C16 sophorolipids. Oleic and stearic acid are preferably hydroxylated at the ω-1 position, while palmitic acid is more equally hydroxylated at the ω and ω-1 positions. Since palmitic acid is shorter than both oleic and stearic acid, the terminal carbon atom might be better positioned for hydroxylation resulting in a mixture of ω and ω-1 products.

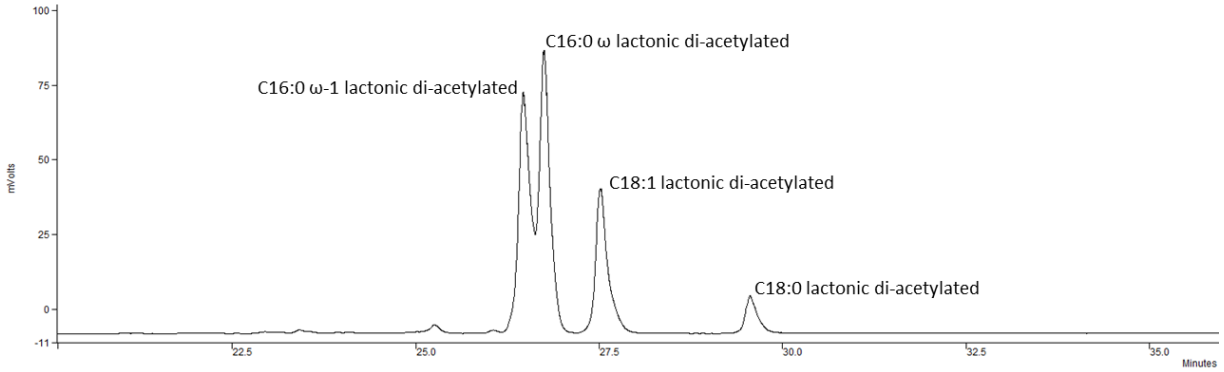


Figure 3.1: HPLC-ELSD chromatogram of a growth trial of wild type *S. bombicola* with ethyl palmitate as a hydrophobic substrate.

Still, *S. bombicola* is capable of utilizing shorter fatty acids with differing hydroxylation patterns as a substrate for glucosylation by UGTA1. However, a big drawback is the low uniformity caused by the promiscuity of the enzymes involved. As Figure 3.1 clearly depicts, a mixture of different C16 and C18 hydroxy fatty acids are incorporated in the sophorolipids rather than just one single congener. An interesting way to increase sophorolipid fatty acid uniformity would be to use P450s that are rather limited in their substrate while still converting enough to ensure adequate production of these molecules. In the following research, an attempt has been done to create new-to-nature sophorolipids with shorter fatty acid tails.

III.3.1. Self-sufficient P450 engineering

III.3.1.1. Expression of natural self-sufficient P450s in *S. bombicola*

Expression of self-sufficient P450s is an interesting strategy to alter the substrates available for the glucosyltransferases without risking incompatibility issues between the P450 and the cytochrome P450 reductase (CPR) of *S. bombicola*. In total three class VIII enzymes were expressed successfully in *S. bombicola* by integration at the *cyp52m1* locus and using the cyp52m1 promotor for expression (see III.2.2.2). These were CYP102A1 from *Bacillus megaterium*, CYP102A7 from *Bacillus licheniformis* and CYP505A1 (also known as $P450_{Foxy}$).

Though the substrate specificities differ slightly between them, it is fair to say that fatty acids ranging from lauric acid (C12:0) up to docosahexaenoic acid (C22:6) can be hydroxylated or oxidized.

However, sophorolipid synthesis tests with strains carrying CYP102A7 or $P450_{Foxy}$ using myristic acid as a hydrophobic substrate (5 g/L added 2 days after inoculation) resulted in no novel sophorolipid molecules being produced in detectable amounts. No adverse effects of the P450s could be detected in comparison with the wild type ATCC 22214 strain for parameters like growth rate or glucose consumption.

In order to verify enzyme activity in the *S. bombicola* cells, test with cell lysates from 4 days old cultures were conducted. Tested substrates were myristic acid (C14:0), oleic acid (C18:1), eicosapentaenoic acid (C20:5) and docosahexaenoic acid (C22:6) for CYP102A7 and lauric (C12:0), myristic acid (C14:0) and oleic acid (C18:1) for CYP505A1. Product formation was evaluated by HPLC and LC-MS (Figure 3.2 and Figure 3.3).

All evaluated substrates were clearly converted to hydroxylated compounds; proving the correct and active expression of the self-sufficient CYP1072A7 and $P450_{Foxy}$ genes in *S. bombicola*. In the case of CYP102A7, C14 and C18:1 are converted to hydroxylated products, which elute at 22.5 and 26.8 minutes respectively. For the polyunsaturated substrates more complex products combining several hydroxylations and epoxidations of the double bond are observed. The signal at 37.5 minutes originates from cell-own free fatty acids. For $P450_{Foxy}$, similar results are obtained for lauric and myristic acid. Both are converted to their respective hydroxy fatty acid and elute at 18.5 and 22.7 minutes for respectively hydroxy lauric acid and hydroxy myristic acid. No activity could be detected for oleic acid, which was unsuspected since according to Nakayama *et al*.[295] the relative activity of $P450_{Foxy}$ towards oleic acid is still 50 % compared to the activity towards lauric acid. It has to be noted that characterisation of $P450_{Foxy}$ by Nakayama *et al*.[295] was done *in vitro* on purified enzyme while the results discussed here are for raw cell lysates.

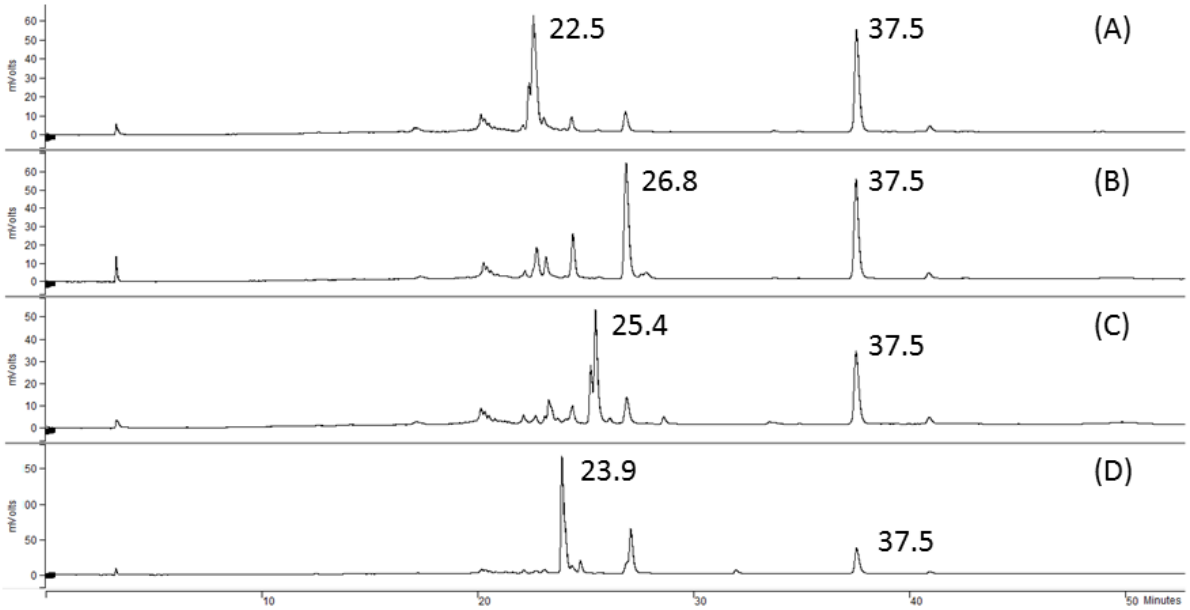


Figure 3.2: HPLC chromatograms of the CYP102A7 activity towards several fatty acids (A): myristic acid, (B): oleic acid, (C): eicosapentaenoic acid and (D): docosahexaenoic acid). The major hydroxylation or

epoxidation product is annotated for each substrate tested. The signal at 37.5 minutes represents the cell own oleic acid molecules from *de novo* fatty acid biosynthesis.

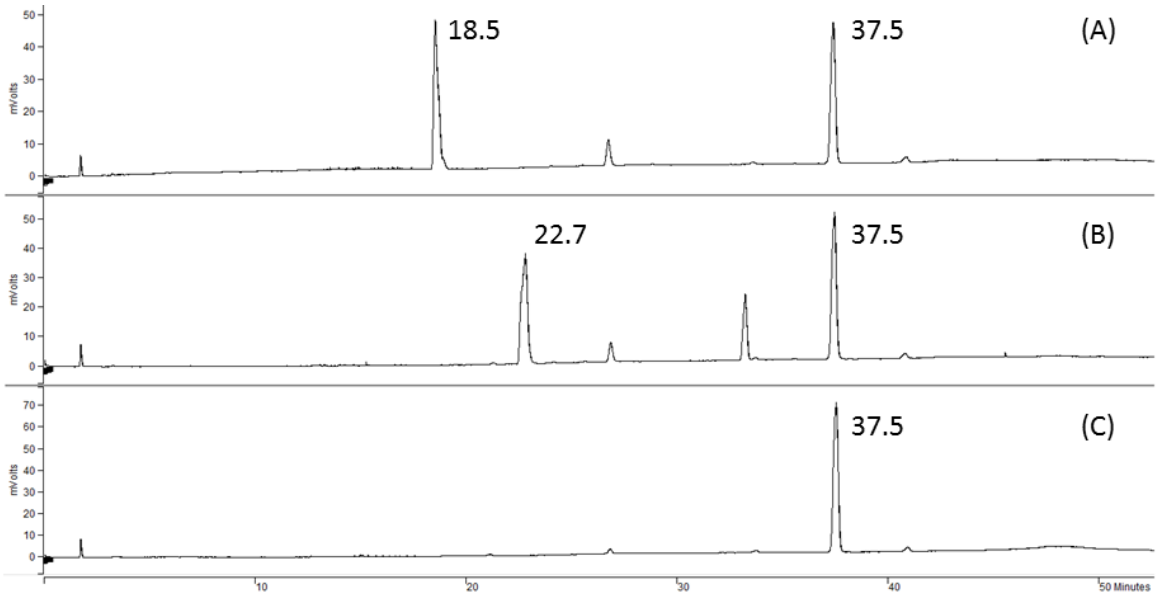


Figure 3.3: HPLC chromatograms of the enzyme assays using the $P450_{Foxy}$ mutant strains. (A): lauric acid, (B): myristic acid and (C): oleic acid. The major product formed is annotated. The molecule eluting after 37.5 minutes is oleic acid.

To conclude, both P450s were successfully expressed and able to hydroxylate the fatty acids *in vivo*. Yet, no incorporation of the hydroxylated compounds in sophorolipids was observed. This could on the one hand be attributed to the non-compatible region- and/or stereo-specificity of the hydroxylation. It is for instance not clear if the first glucosyltransferase can handle hydroxylated fatty acids different from the ω or ω-1 hydroxylated ones; CYP102A7 has a preference towards the ω-2 carbon atom, but side reactions towards the adjacent atoms are possible. The ratio of hydroxylation between the ω-1, ω-2 and ω-3 positions differs between different fatty acids, but in general the majority of hydroxylations occurs on the ω-2 carbon atom. This is consistent with other members of the CYP102A family. For myristic acid, both CYP102A2 and CYP102A3 also have higher activity towards the ω-2 or even ω-3 position and less for ω-1[291,405]. CYP102A5 also prefers ω-2 when incubated with palmitic and linoleic acid[292]. This stands in contrast with CYP102A1, which has a higher preference for the ω-1 carbon atom when the same substrates are compared. In the case of $P450_{Foxy}$ no ratios have been determined. In general no values are given for terminal hydroxylation, suggesting that this reaction is very uncommon for these enzymes. Still, ω-1 incorporation is indeed possible for C18 fatty acids since they occur in wild type sophorolipids, but it is not clear if this is also feasible for C12 and C14 fatty acids.

Furthermore, the unclear R/S configuration of both P450s might also be of importance; for the native sophorolipids, only the S-configured hydroxylated fatty acid is glucosylated by the first glucosyltransferase[377]. High sequence identity on the protein level with CYP102A1, a known R-hydroxylating P450, points towards a similar product being formed by both P450s. When looking at the ω-1 and ω-2 carbon atoms from several hydroxy fatty acids, the percentage of R-configured hydroxy groups ranges between 93 and 98%. At the ω-3 position this ratio drops to values between 74 and 87 %[406]. Though some of the hydroxy fatty acids have the correct stereospecificity, no production of sophorolipids on myristic acid could be observed.

III.3.1.2. Screening of alanine mutants of CYP102A1, CYP102A7 and CYP505A1

Shifting the P450s hydroxylating at the terminal position might be an interesting strategy to avoid possible problems with wrong regio- and stereospecificity. The wild type CYP102A1 is not only capable of accepting fatty acids, but also alcohols, alkenes and more bulky molecules like aromates[407]. Engineering of CYP102A1 has been done in several examples resulting in a broader substrate acceptability[408–410]. This is mostly done by rational design or directed evolution, but in general it comes down to a widening of the substrate channel inside the enzyme. Applications in drug metabolism research where big molecules need to access the enzyme for further oxidization are examples of this[398].

One amino acid residue stands out in all the mutagenesis experiments and this is the phenylalanine at position 87 of the wild type enzyme. According to several studies[285,411–413], this residue interacts with the hydrophobic tail of the fatty acids, resulting in a slight bend in the carbon atom chain of the substrate. Consequently, it is believed that the ω-1, ω-2 and ω-3 carbon atoms are more favourable positioned towards the heme iron and therefore are more easily oxidized than the terminal carbon atom. Phenylalanine 87 also plays a potential role in the stereoselectivity by blocking certain C-H bonds on the chain that might be used for S-hydroxylation. Mutating this residue to an alanine results in a shift of the hydroxylation position towards the terminal carbon atom for C12 and C14 fatty acids[411]. Molecular dynamics (MD) simulations and quantum-mechanical/molecular mechanics (QM/MM) calculations seem to support these findings[413]. It has to be noted that several reports don't support these findings concerning the fatty acid hydroxylation products[414,415] of these alanine mutants. As mentioned before, the wild type CYP102A1 has a high preference for the ω-1 position for myristic, lauric and capric acid. *In vitro* tests with both the entire enzyme and only the heme domain for the F87A mutants showed a shift towards the ω-4, ω-5 and ω-6 positions. Substitution with other residues like isoleucine have been reported as well to influence the regiospecificity of CYP102A1[416].

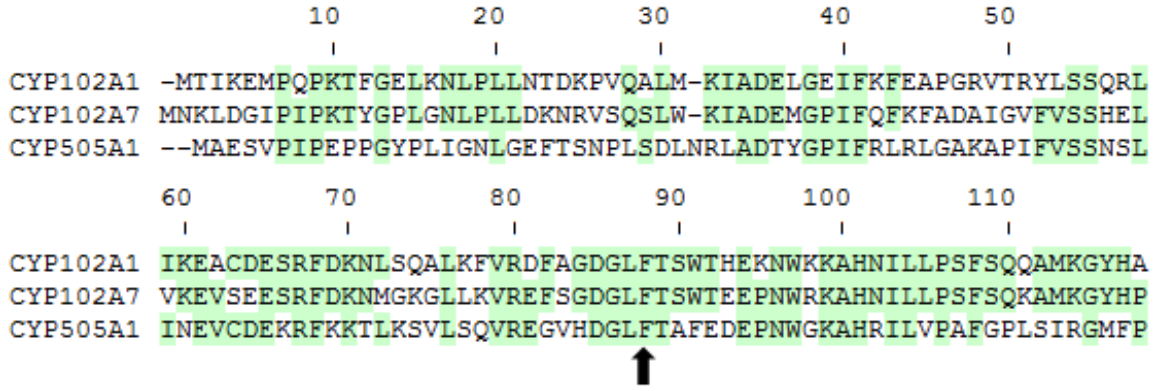


Figure 3.4: Alignment of the first amino acid residues of CYP102A1, CYP102A7 and CYP505A1 ($P450_{Foxy}$). The phenylalanine at position 87 is indicated by the black arrow.

In this work a total of three alanine mutants at position 87 were made. Sequence alignment of CYP102A7 and $P450_{Foxy}$ with CYP102A1 revealed that the same phenylalanine residue is present at position 88 for $P450_{Foxy}$ and 89 for CYP102A7. It has to be noted that historically the threonine at position 2 for CYP102A1 has been chosen as the starting residue of the protein sequence resulting in a slightly different numbering for the phenylalanine. In Figure 3.8, the first 118-119 residues of the three enzymes are aligned. As can be seen clearly, the phenylalanine is situated in a conserved DGLFT motif. This holds true for other members of the CYP102A family as well.[406] For CYP102A7 and $P450_{Foxy}$, this residue will be given the number 88 according to the alignment.

Creation of the alanine mutants was done starting from the wild type enzymes using site-directed mutagenesis[417]. Both wild type and mutated constructs were transformed in *S. bombicola* PT36 and correct transformants were selected. A first growth trial was done using ethyl myristate as a substrate. This substrate was chosen since it is liquid at room temperature, facilitating its admission, and because all three enzymes have a fairly high preference towards myristic acid, but the latter one is solid even at 30°C resulting in low substrate availability in the medium. Usage of methyl and ethyl esters of fatty acids has been proven successfully for production of sophorolipids[418] and generally resulting in high product titers.

At the end of the trail, LC-MS analysis was performed to detect the medium chain sophorolipids. Surprisingly, only longer congeners corresponding to C16 and C18 sophorolipids were found only in low amounts (Figure 3.5). Several explanations can be found for this phenomenon. First of all, the mutation does not have an effect on the kind of hydroxy fatty acid being produced. The mutant P450 shows the same activity as the wild type enzyme without the F87/88A mutation, C14:0 hydroxy fatty acids are still being produced in subterminal *R*-enantiomeric form, but can't be glucosylated by UGTA1. Secondly, the correct (S)-13-OH-C14:0 and 14-OH-C14:0 products are being formed, but UGTA1 is not or less active towards the C14 length. Though unlikely, the preference of UGTA1 drops significantly when shorter chains are presented in *in vitro* studies[175]. Thirdly, using the ethyl ester of myristic acid might be a problem. When using ethyl palmitate, this substrate is easily converted to palmitic acid in the culture medium. This is possibly done by a lipase. However, when using ethyl myristate, no hydrolysis of the ester bond could be seen.

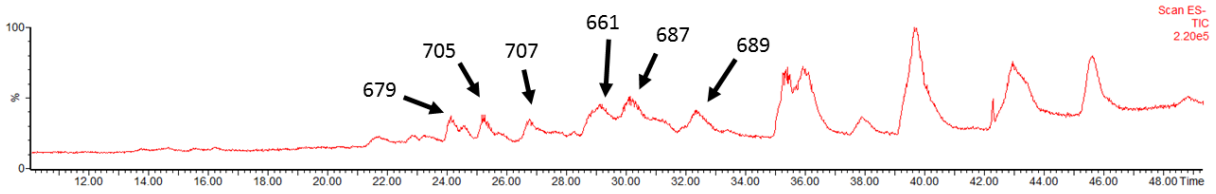


Figure 3.5: Sophorolipids produced by CYP102A1 on ethyl palmitate. M/z values of 679, 705 and 707 correspond to C16:0, C18:1 and C18:0 di-acetylated acidic sophorolipids. M/z values of 661, 687 and 689 correspond to C16:0, C18:1 and C18:0 di-acetylated lactonic sophorolipids

Therefore, a new trial was conducted using myristic acid. Again, no clear production could be seen for the shorter chain sophorolipids. A third growth trial was performed, this time using ethyl palmitate. This trail showed clear production of sophorolipids. The LC-MS spectra show the presence of C16:0 mono-acetylated, di-acetylated, lactonic di-acetylated, C16:1 lactonic di-acetylated and minor amounts of C18:1 di-acetylated acidic and lactonic sophorolipids (Figure 3.6). UPLC-ELSD analysis revealed that production was very low, but the wild type enzymes produced more than their corresponding F88A mutant. It has to be noted that acetate

adducts of sophorolipids are rare during LC-MS analysis, but they do occur occasionally. Though the reason for their occurrence is unknown, they can be resolved by applying higher voltages for the MS or by running the samples again independently from the first run. The conclusion why these are lactonic di-acetylated sophorolipids is based on the several arguments[419]. First of all, the elution times correspond with the lactonic sophorolipids. Secondly, when the molecular mass of acetate is subtracted from the measured mass they correspond to lactonic sophorolipids. Thirdly, adducts with other molecules might occur as well, but seem unlikely. For example, acetonitrile adducts are only measured in positive mode. In the case they do appear in negative mode, a molecular mass of 720 Dalton would be expected for di-acetylated C16 sophorolipids (680 + 41 – 1 = 720).

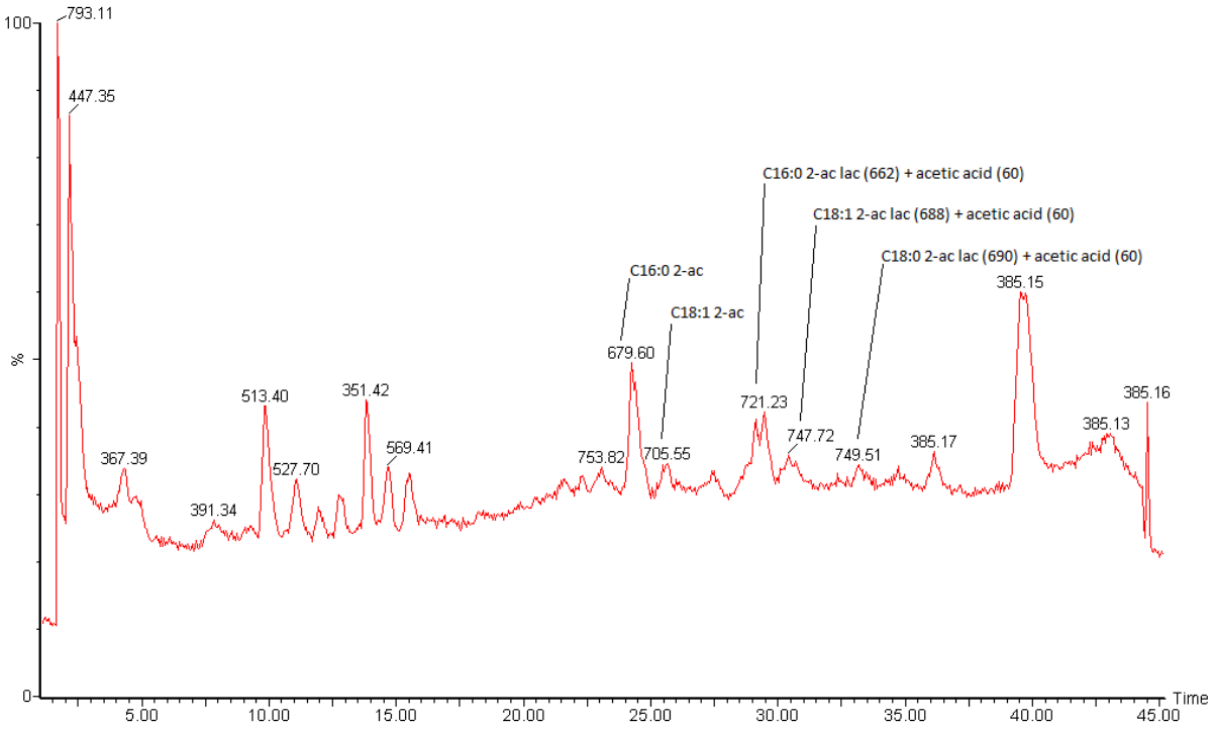


Figure 3.6: LC-MS spectrum of the broth of the *S. bombicola* CYP102A7 strain after 217 h with ethyl palmitate as the hydrophobic substrate

III.3.1.3. Enzyme tests for the F87A mutants

Though *in vivo* experiments showed lower productivity of the F87/88A mutants, it might still be that the mutants produce other kinds of products compared to the wild type enzymes, but that these molecules can't be incorporated by glucosyltransferases UGTA1 and UGTB1 into sophorolipids. Since the aim is to produce medium-chain sophorolipids, this lowered productivity of glycolipids might mask efficient production of other molecules. As explained in III.3.1.1, tests on crude cell lysate already showed high activity of the self-sufficient P450s towards several fatty acids.

All the F87A mutants were grown until three days in the stationary phase. As substrates, the fatty acids lauric, myristic and palmitic acid were chosen since all the P450s tested show high activity towards these molecules. The different cell lysates and molecules were incubated at 30 °C and sampled after three hours. LC-MS analysis of the different lysates clearly shows the lower activity of the F87A mutants compared to their respective wild type counterparts. In Figure 3.7, a comparison can be seen for CYP102A7 and its mutant variant incubated with

myristic acid. The wild type enzyme clearly produced more hydroxylated myristic acid (m/z 243) than the alanine mutant. Other products originating from the cell own fatty acids are visible as well. For example, hydroxylated oleic acid elutes at 30.5 minutes (m/z 297). Again the cell lysates originating from the CYP102A7 strain were the most potent for converting fatty acids to their corresponding hydroxylated congeners.

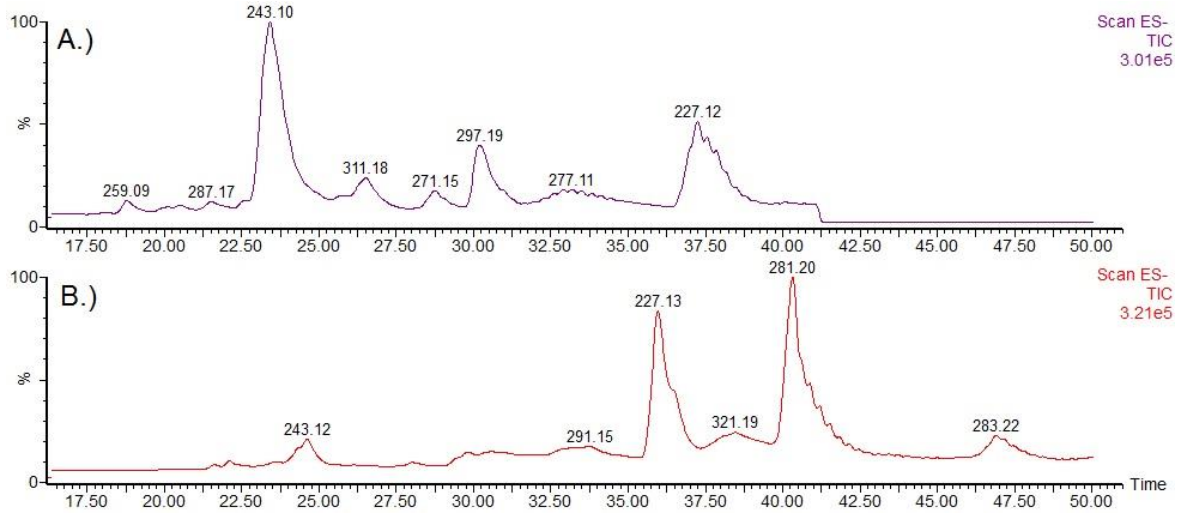


Figure 3.7: LC-MS analysis of an essay with cell lysate from *S. bombicola* A.) CYP102A7 and B.) CYP102A7 F88A. The substrate used was myristic acid (m/z 227). Hydroxylated myristic acid is visible at 24 minutes (m/z 243).

III.3.1.4. Hydroxylation of quercetin and resveratrol

The self-sufficient P450 CYP102A1 and its derived mutants is an oxidizer of many different molecules, among them quercetin[420] and resveratrol[290]. Several of the mutants mentioned are mutated at the F87 position as well. Though only tested for CYP102A1 in literature, the other two mutant P450s created in this work were tested as well together with their wild type enzymes to assess the possibility of *in vivo* hydroxylation of quercetin and resveratrol in parallel.

The wild type *S. bombicola* and the strains containing the wild type and mutated self-sufficient P450s were inoculated in production medium with both molecules added at the beginning of the stationary phase in 1 g/L quantities. After several hours, the cultures with quercetin started producing a very distinctive smell. Though the origin and composition is unknown, it is speculated that it might be a degradation product of quercetin. After several days, it became clear that the wild type *S. bombicola* was producing sophorolipids for both substrates tested. TLC and HPLC analysis of the different mutant strains showed that no detectable amounts of modified of resveratrol or quercetin occurred compared to the wild type *S. bombicola*.

Possibly both substrate accessibility/solubility and transport are the reason for this. Both resveratrol and quercetin have low solubility in water and tend to aggregate in small particles, further lowering the availability of these molecules.

III.3.2. Creation of chimeric P450 enzymes

III.3.2.1. Introduction

Most P450 enzymes need an extra reaction partner for delivering the electrons necessary for catalysing their specific reaction (see I.3.3 for a more detailed overview). For class II enzymes, this is done by cytochrome P450 reductase (CPR), a membrane bound protein capable of accepting electrons from NAD(P)H and transporting them towards the P450. In mammalian liver tissue, the ratio between the P450 population and CPR roughly is 20:1[421]. When looking at *S. bombicola*, no clear answer is available on the ratio P450:CPR. However, using RNA sequencing data available at the research group, an estimation can be made. During the stationary phase with oil addition, normalized reads per million (RPKM) values of all P450s amounts to 4153.96 while that of the CPR is 326.10 resulting in a ratio of 12.74:1 (Table 3.7). This value is only an estimation since final protein expression levels depend on RNA stability, RNA processing and protein stability.

Table 3.7: Table representing all P450s and the CPR in *S. bombicola* and their RNA expression levels. The RPKM values are used to calculate an approximation of the P450:CPR ratio.

Gene name	Exponential Phase	Stationary Phase with oil	Stationary Phase without oil
Cabom01g00060 (CYP52E4)	95.90	41.67	57.24
Cabom01g04780 (CYP52N2)	42.87	43.83	38.64
Cabom01g04790 (CYP52N3)	46.60	35.12	36.80
Cabom01g10350 (CYP61A1)	936.28	845,62	727.69
Cabom02g13890 (CYP52M1)	258.01	2391.91	3759.45
Cabom03g00070 (CYP52E5)	153.26	47.90	61.88
Cabom03g00110 (CYP52E6)	567.91	208.62	204.84
Cabom03g01690 (CYP52E3)	99.67	93.82	163.07
Cabom03g11440 (CYP52F1)	78.35	105.24	135.80
Cabom03g15730 (CYP52N1)	763.27	340.23	261.92
Total counts P450s	**3042.12**	**4153.96**	**5447.34**
Cabom01g04810 (CPR)	304.66	326.10	387.95
Ratio P450:CPR	**9.99**	**12.74**	**14.04**

Chimeragenesis of a P450 and a CPR has been applied successfully. By making the P450 self-sufficient, eukaryotic P450s can be expressed in prokaryotic expression and production systems. Furthermore, its catalytic properties like substrate affinity, product formation and electron transport can be altered or elevated[401]. Two hypothesis are tested concerning chimeric P450-CPR systems in *S. bombicola*. The first one is if an increased productivity for wild type sophorolipids is possible by the relatively high ratio of P450 (in this case CYP52M1) and CPR, a 1:1 ratio of CYP52M1 with its own CPR might result in higher electron transfer rates and subsequent P450 activity. As a result, this might push the sophorolipid pathway ending in higher productivity. The second postulate is researching if heterologously expressed class II P450s are active in *S. bombicola* and if their activity can be increased by coupling a reductase to their C-terminus. To test both assumptions, two different reductases were selected. The first CPR is from *S. bombicola*, this to maximize compatibility with CYP52M1 and to investigate and validate the lack of production in certain experiments. The second one is BMR, the reductase and linker region of CYP102A1 from *B. megaterium*. This reductase is chosen since it has been investigated thoroughly the last few decades. Based on sequence

alignment and amino acid conservation, the start of the linker of CYP102A1 was deduced to start at amino acid 452 for CYP102A1.

Figure 3.8: Sequence alignment with conserved regions highlighted of the CYP102A family. Amino acid conservation drops between positions 451 and start again at 482, suggesting the end of the P450 and the start of the reductase respectively[406].

III.3.2.2. CYP52M1 from *S. bombicola*

Production of sophorolipids is already very efficient in *S. bombicola*. Productivity values of up to 3.5 g/L.h are reported in literature for the wild type mixtures[422]. Engineered strains are also capable of exceeding 0.5 g/L.h[174]. Still, efforts in increasing these values are interesting since they result in shorter fermentation times or higher final yields. These advantages could be translated in lower costs of the final product. Furthermore, understanding genetic optimization of the production of wild type molecules can be translated to new types of molecules. In order to investigate the effect of chimeric constructs on the sophorolipid production, CYP52M1 was coupled to the reductase of CYP102A1 and its own *S. bombicola* CPR (SbCPR). For the enzyme with the SbCPR as reductase domain, the transmembrane helix was removed to ensure the SBCPR is not limited in its movability by this membrane anchor. The chimeric sequences were introduced at the *cyp52m1* locus and the resulting strains were tested on oleic acid and sampled regularly to assess any differences in production efficiency.

Unfortunately, no elevated production could be detected for the CYP52M1SbCPR and CYP52M1BMR strains (Figure 3.9). In the case of the BMR chimeric construct, a significant lower sophorolipid concentration was found at the end of the growth trial, but for the SbCPR chimer this drop is less clear. It is not known if both reductase domains are still contributing electrons to the P450 and if they have an influence on the activity of the uncoupled *S. bombicola* CPR or any interaction with other proteins it has to make. Coupling two enzymes can hinder their interaction if the linker hasn't got the proper length or form. As such it may be that electron transfer in the chimeric enzymes is hampered or even that the P450 domain still receives its electrons from the membrane bound CPR, but that this is less efficient due to the coupled reductase blocking the way. Furthermore, the membrane anchor of SbCPR was removed, which might affect its structure and/or activity.

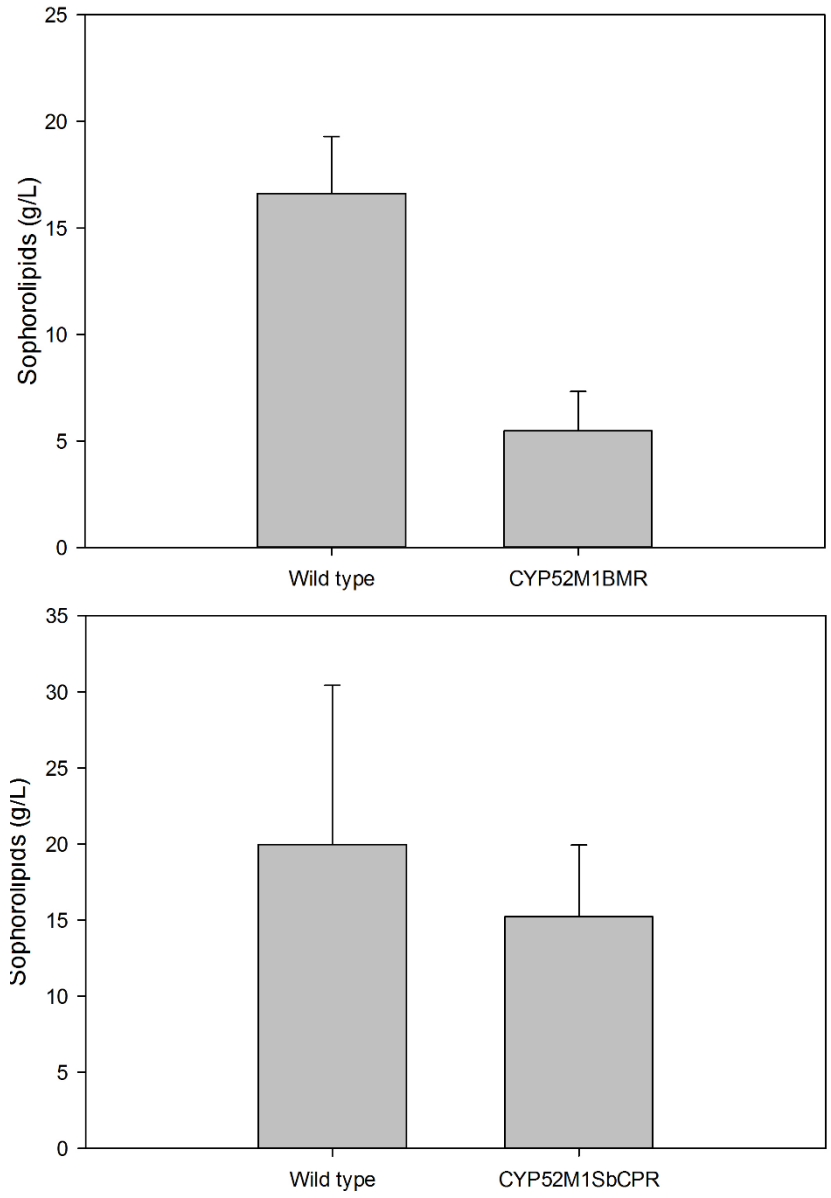


Figure 3.9: Sophorolipid concentrations for the growth trials with the wild type *S. bombicola* and the two different chimeric P450s strains. Error bars depict standard deviation. Sample size for each strain was 3.

Since no clear difference could be detected for the wild type molecules with the CYP52M1SbCPR chimer, another strain was chosen whose productivity is lower. The recently developed bola-sophorolipid producing strain[423] was a suitable candidate since production levels on fermentor scale are already at 0.5 g/L.h, but further improvement is necessary. If the chimeric CYP52M1SbCPR enzyme has a beneficial effect on the pathway by delivering more precursors, productivity might rise. When the CYP52M1SbCPR was introduced in a *ura3* negative bola-sophorolipid strain, production of these molecules was severally hampered. Though again the real reason is unknown, it might be that the earlier modification of the sophorolipid gene cluster (knock-out of the acetyltransferase) and the additional introduction of a less than optimal functioning chimeric P450 resulted in lower expression values of the cluster genes due to local disruption of the gene cluster and overall lowered production.

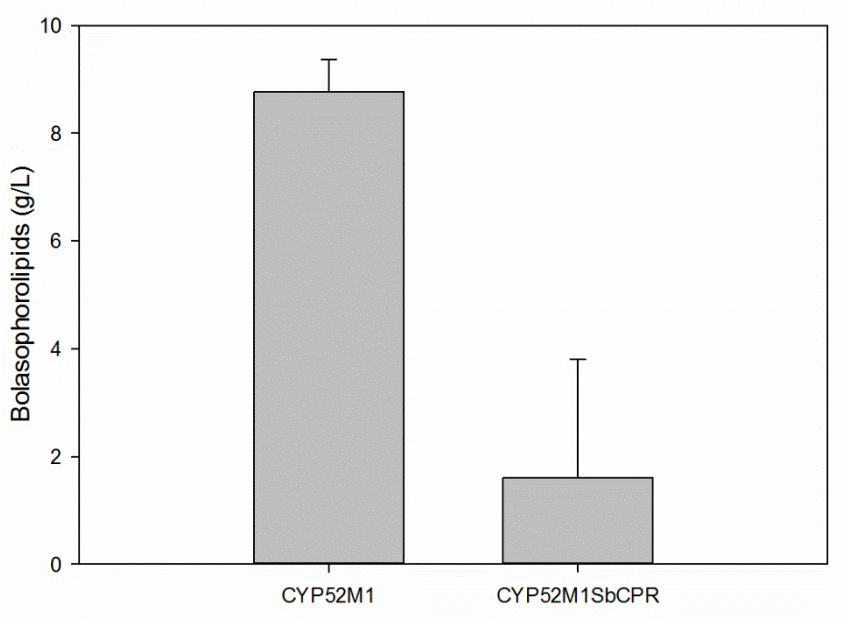


Figure 3.10: Production of bola-sophorolipids by the original strain and a strain carrying the chimeric CYP52M1SbCPR. Error bars depict standard deviation. Sample size for each strain was 3.

III.3.2.3. CYP52A4 from *Candida maltosa*

Heterologous expression of the wild type CYP52A4 in *S. bombicola* resulted in no detectable activity of this P450 both for sophorolipid synthesis and in fatty acid hydroxylations[424]. However, enzyme tests in *Saccharomyces cerevisiae* W(R) were successful. Potentially, an incompatibility of the P450 and the *S. bombicola* CPR might be the reason for the lack of activity. Therefore, a chimeric construct of CYP52A4 with the reductase of CYP102A1 was constructed to see if by introducing an additional electron donor, transport can be established.

A growth trial was performed using ethyl esters of lauric and myristic acid. Analysis of the sophorolipid samples was done by LC-MS. In Figure 3.11 and Figure 3.12, the resulting chromatograms are depicted. For the ethyl myristate growth trial, sophorolipids with m/z values of 623 m/z, 651 m/z, 679 m/z and 705 m/z are easily detected. They correspond to di-acetylated acidic sophorolipids with fatty acid tails of 12, 14, 16 and 18 carbon atoms. The ethyl laureate growth trial resulted in C12 sophorolipids, but no other congeners were detected. However, a clear signal for dicarboxylic acids appeared around 18 minutes.

One interesting observation is the wide range of fatty acids being incorporated in the sophorolipids while according to the research of Zimmer *et al*[249] CYP52A4 (also known as Cm2) shows a clear preference towards chain lengths of 12 and 14 carbon atoms *in vitro*. Besides the potential different behaviour of the P450 *in vivo* compared to *in vitro*, a possible explanation is the interplay between the specificities of the glucosyltransferases of *S. bombicola* and the new chimeric CYP52A4 enzyme. Furthermore, it is known that coupling of a reductase can change the substrate preference and products formed by the P450 domain. If this is the case, CYP52A4BMR might be producing hydroxylated C16 and C18 fatty acids as well. Then, the specificities of the chimer and the transferase can be called opposites of each other, resulting in a relatively uniform product spectrum from C12 until C18 sophorolipids.

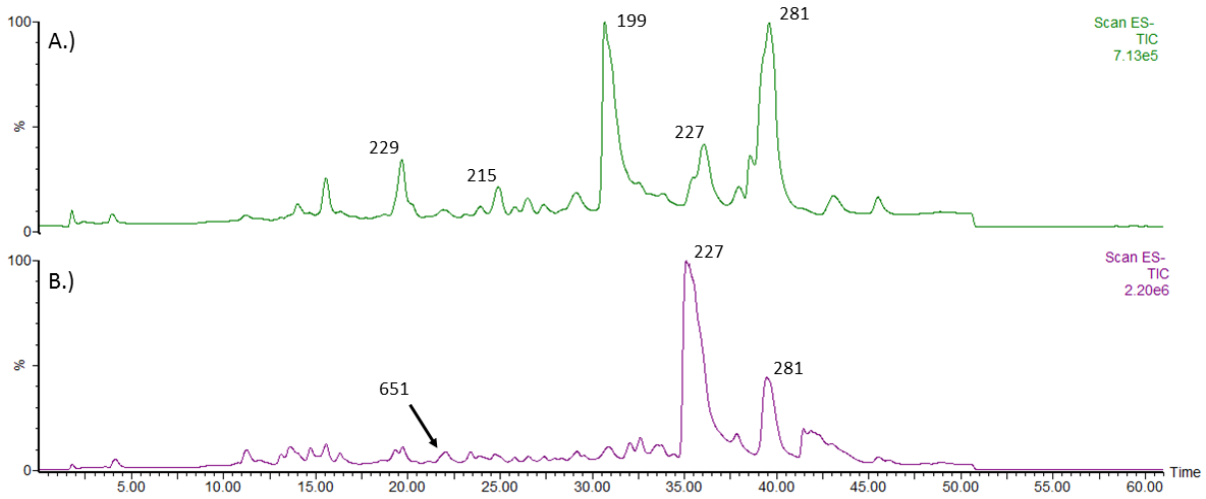


Figure 3.11: Total ion scans for the CYP52A4BMR growth trial on A.) ethyl laureate and B.) ethyl myristate. Prominent molecules are oleic acid (m/z = 281), myristic acid (m/z = 227), lauric acid (m/z = 199). Hydroxy lauric acid (m/z = 215) is visible at 25 minutes in chromatogram A.) while the dicarboxylic derivative elutes after 19 minutes (m/z = 229). Di-acetylated C14 sophorolipids (m/z = 651) are visible at 22 minutes in B.).

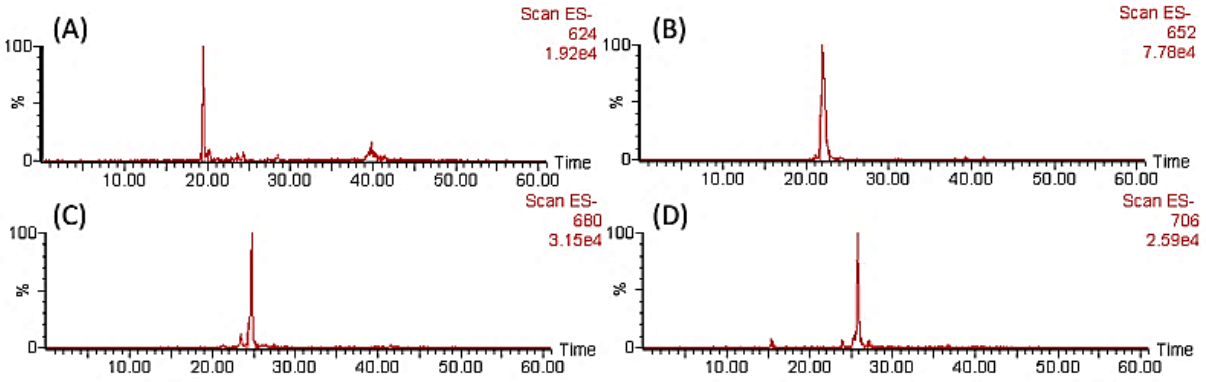


Figure 3.12: Mass scans of the ethyl myristate growth trial with the CYP52A4BMR strain. (A): C12 acidic sophorolipids; (B): C14 acidic sophorolipids; (C): C16 acidic sophorolipids; (D): C18:1 acidic sophorolipids

The first glucosyltransferase UGTA1 has a clear preference towards ω-1 hydroxylated oleic acid[176]. Preference for hydroxylated palmitic acid is only 29% of that for hydroxy oleic acid. For shorter chains like hydroxy lauric acid, no activity was detected. However, fermentation experiments using alcohols[392] and ketones[393] resulted in novel kinds of sophorolipids with yields up to 15 g/L after 10 days of fermentation. One may conclude that UGTA1 can glucosylate shorter chain lengths, but with significant lower activity. A remark has to be made that this still might depend on the substrates used. Alcohols and ketones differ from hydroxy fatty acids by the big carboxyl group at the other side of the chain (besides the differences between the hydroxy or keto group on those molecules). This dissimilarity can have an effect on the preference of the glucosyltransferase.

When looking at UGTB1, the activity towards the glucolipid 17-O-glucopyranosyloctadecenoic acid compared with decylglucoside is very similar with a drop of only 17% for decylglucoside. This is interesting since decylglucoside lacks a carboxyl group and is significantly shorter than the C18 glucolipid. Potentially the carbohydrate group is necessary for catalytic activity instead of any other functional group or chain length. Both observations relating to the glucosyltransferases result in UGTA1 having a big influence on the fatty acids being incorporated.

The lack of other kinds of sophorolipids with ethyl laureate compared to ethyl myristate is another interesting observation. Besides sophorolipids, dicarboxylic acids were detected as well for both substrates with the highest amount in the culture with ethyl laureate. Relating back to the specificity of the glucosyltransferases, it might be that the pool of hydroxy laureate is more easily converted to the corresponding carboxylic acids than incorporated in sophorolipids. Though mostly untested, one of the other endogenous P450s of *S. bombicola* might be responsible for this conversion. In other organisms like *Candida maltosa*[249,250] or *Candida tropicalis*[425], P450s belonging to the CYP52 family can convert alkanes or terminal hydroxy fatty acids to the corresponding α,ω diacids.

III.3.2.4. CYP1 from *Ustilago maydis*

Wild type CYP1 expressed in S. bombicola

CYP1 from *Ustilago maydis* takes part in the cellobiose lipid synthesis (see I.4.1) by terminal hydroxylating palmitic acid (C16:0). Hence, it is one of the few identified P450s hydroxylating fatty acids and being involved in glycolipid synthesis. This features render it a promising candidate for creating sophorolipid with an altered fatty acid chain in *S. bombicola*.

An initial growth trial was performed on palmitic acid (5 g/L) to investigate the activity of the unaltered CYP1 P450. A strain not capable of synthesizing sophorolipids by knocking out the P450 was used as a control. This way, even low amount of production can be detected by comparison against any background activity. Identification of the products was done by LC-MS analysis. The spectrum showed a signal for the production of the sophorolipid with mass 679 m/z (di-acetylated acidic sophorolipid with a C16 fatty acid chain). Peaks for the masses 661 m/z (di-acetylated lactonized sophorolipid with a C16 fatty acid chain) and 663 m/z (mono-acetylated acid sophorolipid with a mono-unsaturated C18 fatty acid chain) were detected as well (data not shown). The signal with a molecular weight of 679 m/z was clearly higher in the strain carrying the heterologously expressed CYP1 (Figure 3.13 and Figure 3.14) compared to the control. This shows that replacement of CYP52M1 by another unaltered P450 can result in altered sophorolipid production. It has to be noted that the amount of sophorolipids produced remained extremely low. After 8 days of production, signal was barely above the detection limit of the LC-MS.

Several explanations can be found for this. First of all, *U. maydis* produces relatively low amounts of cellobiose lipids. Spoeckner *et al*.[21] reported yields up to 30 g/L of a mannosylerythritol lipids and cellobiose lipids mixture of which 10% are the latter molecules. Similar values have been achieved by Saerens *et al*.[381]. This might be due to low activity of CYP1 in this pathway. When copying an enzyme with low overall activity, it shouldn't be a surprise that this is also observed in a heterologous system. Another problem might be non-optimal expression. Though the codon usage of CYP1 was compared to that of *S. bombicola* and deemed compatible, small problems might still exist upon translation of the RNA or folding of the amino acid sequence, resulting in suboptimal expression. Thirdly, one has to take into account that the P450 needs to interact with the cytochrome P450 reductase (CPR) for obtaining the required electrons. This is based on electrostatic interactions on the surface of both the P450 and CPR. When these surfaces are not compatible, inefficient electron transport can result from this and in the end, the catalytic turnover of the P450 will suffer.

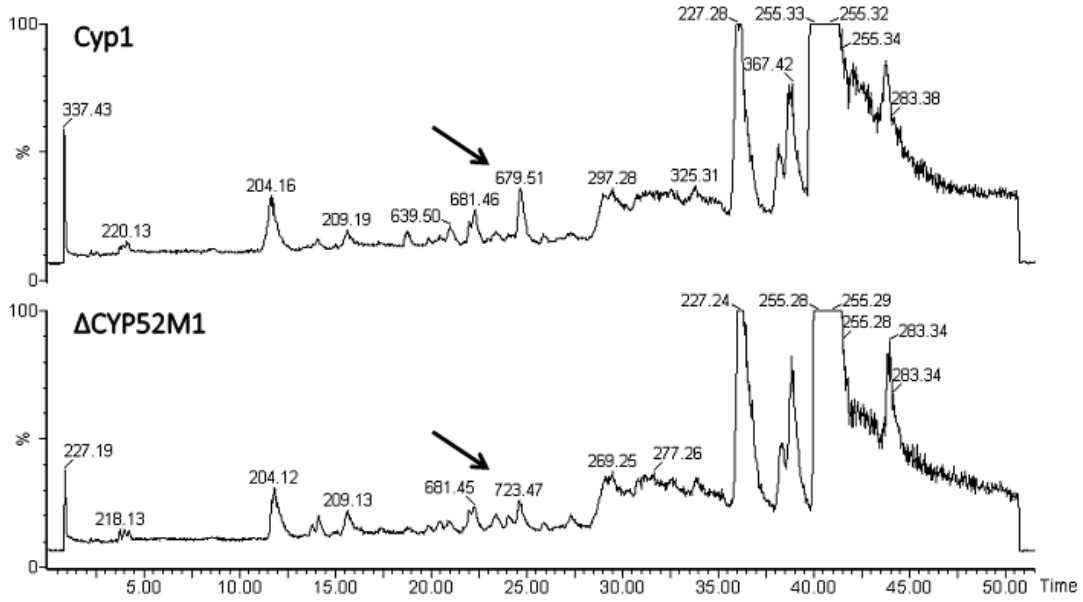


Figure 3.13: LC-MS chromatogram of the CYP1 knock-in strain and the CYP52M1 knock-out strain. The strain expressing the heterologous P450 CYP1 shows a novel mass of 679 m/z (680 – 1) at 24.5 minutes. In the knock-out strain, this signal is masked by another mass of 723 m/z.

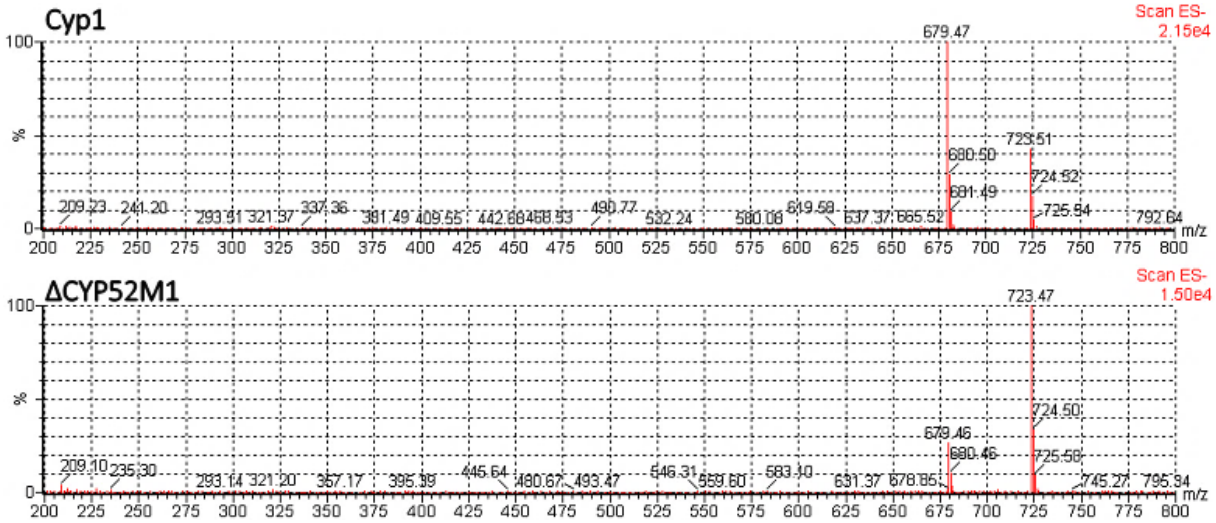


Figure 3.14: A mass spectrum scan at 24.5 minutes shows the shift in ratios between the mass of 679 m/z and 723 m/z.

Initial production analysis of the CYP1 chimeric P450

CYP1 is a typical class II P450 (I.3.3.3) requiring assistance of a CPR for delivery of electrons. Introduction of a chimeric CYP1BMR in *S. bombicola* in order to improve the activity and create higher yields of sophorolipids with a modified fatty acid tail was performed. The sophorolipid production on palmitic acid was investigated and increased amounts of C16 sophorolipids were detected for the strain with the CYP1BMR fusion compared to the CYP1 strain, but still much lower than regular wild type sophorolipid production. LC-MS analysis revealed that other congeners were present as well, mostly non- and mono-acetylated variant of the C16 sophorolipids. Small amounts of sophorolipids with an oleic acid (C18:1) tail were detected as well. They comprise roughly 5% of the total sophorolipid mixture produced. This small fraction can be explained by the fact that C18:1 is shorter than C18:0, more closely mimicking the natural substrate C16:0.

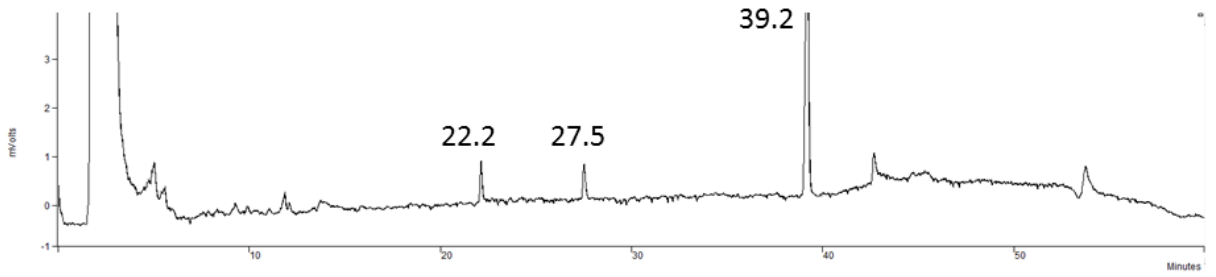


Figure 3.15: HPLC-ELSD chromatogram of the sophorolipid production of the CYP1BMR strain grown on palmitic acid after 8 days in the stationary phase. Products detected are di-acetylated acidic sophorolipids (22.2 minutes), di-acetylated lactonic sophorolipids (27.5 minutes) and the substrate palmitic acid (39.2 minutes). Both types of sophorolipids contain a C16 fatty acid tail.

Several interesting conclusions can be drawn from this growth experiment. First of all, it is possible to use the reductase of a self-sufficient P450 and couple it to a P450 of interest to produce more uniform kinds of molecules. In the case of the CYP1BMR fusion protein, only two different kinds of sophorolipids are observed in relatively high quantities, namely the acidic and lactonic di-acetylated C16 sophorolipids. Still, the total amount of sophorolipids in the culture broth remained below 100 mg/L.

Substrate and fermentation optimization of CYP1BMR

Further optimization of the process was carried out by testing different substrates like fatty acid ethyl esters. Substrate accessibility can be a bottleneck during production for most fatty acids as they are solid at 30 °C. Adding an emulsifier like Tween-20 can improve accessibility, but poses different problems during downstream processing. Ethyl esters of fatty acids are liquid at room temperature and have been used in the past for efficient production of sophorolipids[426]. Therefore, a comparison between palmitic acid and ethyl palmitate was conducted.

It was found that ethyl palmitate further enhanced productivity at least fiftyfold when comparing the amounts of the different sophorolipids produced (Figure 3.16). After 11 days in the stationary phase, a final concentration of 1.56 g/L was measured. The higher productivity can be explained by the fact that ethyl palmitate is liquid at 30°C while palmitic acid is not. The ethyl ester will easily disperse in the broth to form small droplets ensuring a high contact surface. The yeast will be able to use this higher contact area, and resulting higher substrate availability, for elevated production. Using ethyl palmitate did not have an effect on cell viability, both during the exponential as the stationary phase. During the optimization, several substrate concentrations were used between 5 and 25 g/L. Increasing the amount of ethyl ester did not further improve production of C16 sophorolipids. Therefore, in future Erlenmeyer experiments a concentration of 5 g/L was used.

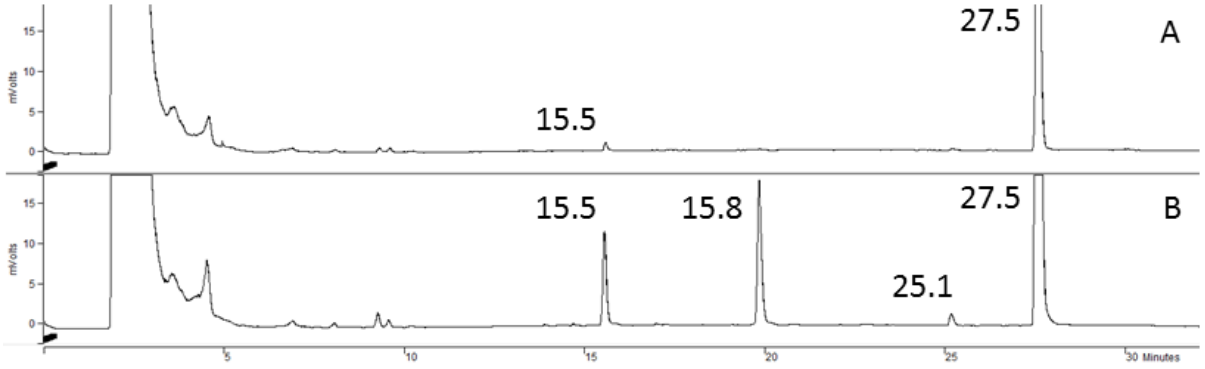


Figure 3.16: Productivity of C16 sophorolipids on palmitic acid (A) and ethyl palmitate (B). The acidic di-acetylated C16 sophorolipids elute at 15.5 minutes, the lactonic at 15.8 minutes. The substrate ethyl palmitate can be seen at 25.1 minutes while palmitic acid eluates at 27.5. It has to be noted that these elution times differ from the ones used elsewhere in this manuscript due to the usage of an experimental shorter HPLC-ELSD analysis protocol.

Validation of the CYP1BMR strains

During the previous experiments it was proven that by coupling a suitable electron donor to CYP1 it is feasible to have higher activity of this enzyme. Still, in order to fully confirm these conclusions, several validations need to occur. First of all, coupling CYP1 to the *S. bombicola* CPR should result in almost no production since it was proven that simple heterologous expression of CYP1 didn't result in efficient production. Secondly, mutating the BMR at certain essential residues rendering it less efficient in transporting the electrons to the P450 should again result in lower production levels, proving that the BMR domain is responsible for electron donation.

Essential residues in the FMN domain are the glycine at position 570, the tryptophan at 574 and the tyrosine at 536 in CYP102A1. Mutating these residues separately to aspartate resulted in a BM3 reductase domain with 30% of the original activity towards cytochrome c[427]. In the validation strategy, the glycine at position 570 will be altered. In the chimeric protein CYP1BMR this corresponds to the glycine at position 785. Similar to the FMN domain, several residues exist in the NADPH binding domain that have an effect on electron transfer and NADPH/$NADP^+$ binding and stability. One of these is the cysteine at position 999 (position 1216 in the chimeric enzyme). Mutating it to an alanine results in lowered performance due to higher affinity for $NADP^+$, thereby blocking the site for exchange with a new NADPH[404,405]. A 5-fold decrease for the reduction of cytochrome c was measured with the C999A mutant. The mutations will be tested on their own as well in a double mutant. It is expected that almost no sophorolipid production will be observed with this double mutation strain if the fused BMR is responsible for the electron delivery to the CYP1.

The growth trial was conducted using 5 g/L ethyl palmitate as a substrate with both single mutants, the double mutant and the non-mutated CYP1BMR strain and lasted until the glucose levels dropped below 10 g/L. Ethanol extraction of the sophorolipids and subsequent quantification resulted in the data presented in Figure 3.17. As can be seen, the mutation in the NADPH domain did not result in a drop in sophorolipid production. The mutation in the FMN domain resulted in a 5-fold drop while the double mutant has a 16-fold reduction compared to the non-mutated chimeric CYP1BMR. Though it was expected that both single mutations would result in a reduction of production, the results clearly show this is not the case.

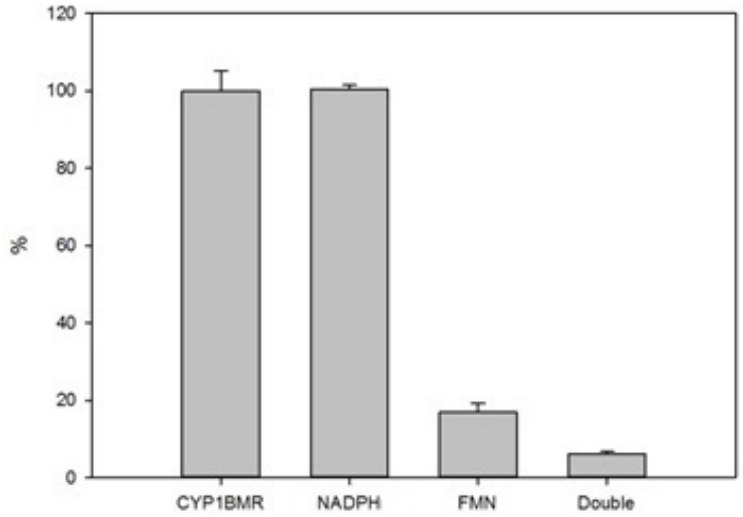


Figure 3.17: C16 sophorolipid production by the different CYP1BMR mutants and the non-mutated chimeric protein. Error bars depict standard deviation. Sample size for each strain was 3.

A potential explanation for this might be the difference between *in vitro* assays performed in the original papers and the *in vivo* experiments conducted in this research. Optimisation of an enzyme essay can result in reaction conditions differing significantly from *in vivo* conditions resulting in influenced enzyme performance. Nevertheless, the double mutation showed an additional effect of the mutated NADPH domain together with the FMN mutation. Another explanation might be that the mutation in the NADPH binding motif does not have an effect on its own, but combined with the one in the FMN domain, the BMR might be sufficiently disturbed to see an effect on activity.

This data, combined with the enhanced productivity of CYP1BMR compared with the wild type non-chimeric P450, clearly show that the BMR domain of CYP102A1 is responsible for the donation of electrons to the P450.

Scale up of C16 sophorolipid production by the CYP1BMR chimeric strain

Scale up of production of C16 sophorolipids was further done in a 3 L and 30 L bioreactor. In the 3 L bioreactor, 2 mL ethyl palmitate was added manually every day while aeration was kept at 1 vvm and stirring rate at 600 rpm. During this fermentation, heavy foaming occurred which was problematic. It was observed that this foaming was reduced each time ethyl palmitate was added to the fermentor. Therefore, the addition of hydrophobic substrate was split up in two portions of 1 mL a day. In Figure 3.18, glucose consumption and average optical density can be seen.

Depending on the growth phase, a glucose consumption rate of 12.5 g/L.day and 8.9 g/L.day was calculated for the exponential and stationary phase, respectively. This is lower than the glucose consumption of a wild type *S. bombicola,* which can easily reach 20 g/L.day during production. A likely explanation for this is the significant lower productivity of this new strain since biomass is on par with the wild type *S. bombicola*. The entire fermentation lasted 411.4 hours with a total yield of 9.9 ± 0.1 g of C16 sophorolipids. This results in a productivity of 24 ± 0.2 mg/L.h, roughly 40 fold lower than reported by Roelants et al.[174] for other engineered strains. Analysis by HPLC and LC-MS showed that the major products being produced are the lactonic di-acetylated C16 sophorolipids (Figure 3.19).

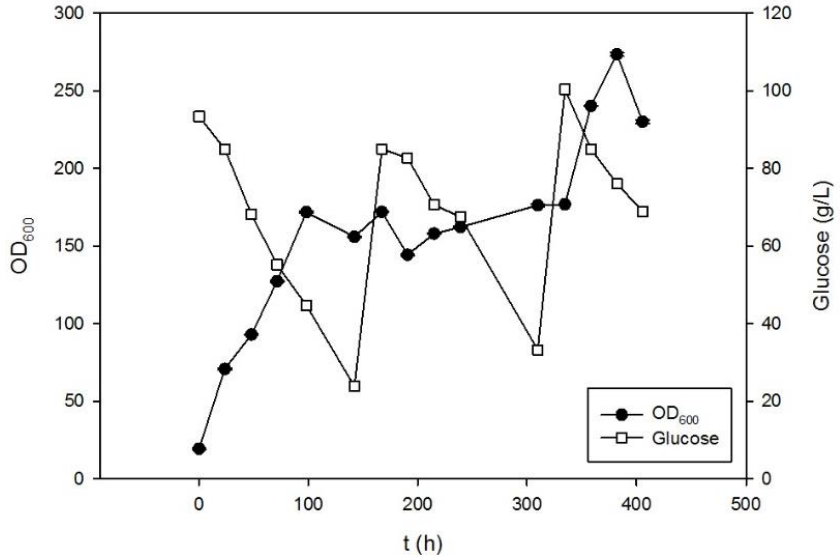


Figure 3.18: Glucose consumption and optical density for the 3 L fermentation with *S. bombicola* CYP1BMR

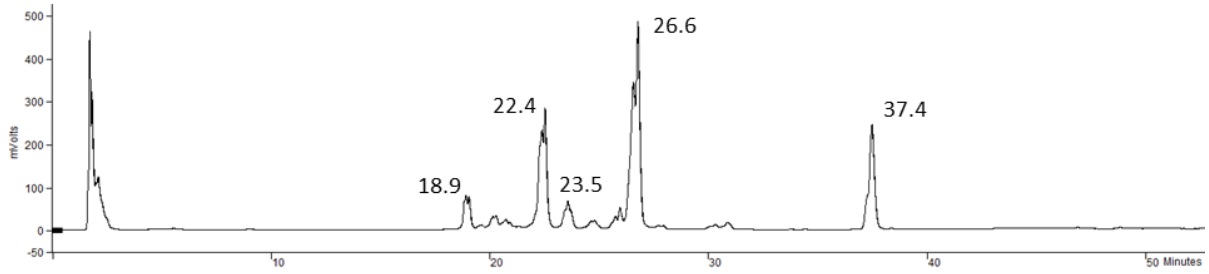


Figure 3.19: Sophorolipids produced in a 3 L fermentor using ethyl palmitate. At 18.9 minutes, non-acetylated C16:0 molecules can be seen. Di-acetylated congeners are present around 22.4 and 26.6 minutes, corresponding to acidic and lactonic sophorolipids respectively. At 37.4 palmitic acid can be seen while a small amount of C18:1 di-acetylated acidic sophorolipids are visible.

Foam formation became even more troublesome on a 30 L scale compared to 3 L. A first 30 L fermentation (fermentation A) was run using the same feeding and stirring strategy as the 3 L fermentor. However due to the excessive foam formation, aeration was reduced to approximately 0.1 vvm. A second 30 L fermentation (fermentation B) used a different feeding strategy.

Conventional anti-foam methods like overpressure or higher stirring speeds seemed inappropriate to combat the foaming during previous fermentations and usage of a silicon based antifoam was not opted due to the separation problems during downstream processing of these molecules from the sophorolipids. An interesting observation is that when a small amount of ethyl palmitate is added to the reactor, foam will disappear for 1 hour after which it will slowly start to accumulate. By using the substrate as an antifoam by adding it every 20 minutes in short bursts of 2 seconds (equalling 200 µL of ethyl palmitate), a relatively foam free fermentation process was possible. Achieving this required an automatic feeding system.

However, care had to be taken as ethyl palmitate has a melting point of 24-26 °C. Also, not every kind of tubing is resistant to ethyl palmitate or similar hydrophobic substrates. The regularly used Norprene® tubing (formulation A-60-G, Saint-Gobain) leaked this compound and even disintegrated. Eventually, Tygon® tubing (formulation F-4040-A, Saint-Gobain) was used. This type of tubing, however, is tougher and was not compatible with the peristaltic pump of the bioreactor control unit, so a very small diameter section had to be used for the part going through the pump. The ethyl palmitate feed bottle was kept on a heating plate with magnetic stirring at a temperature of 60 °C. This allowed in an overall higher aeration rate in fermentation B compared to A. The parameters can be found in Table 3.8. As can be seen, due to the new feeding and anti-foam strategy, aeration could be increased up to 0.3 vvm in fermentation B. Fermentation B resulted in a higher total amount of sophorolipids in a shorter time frame than fermentation A.

Table 3.8: Parameters for 30 L fermentations A and B. The average optical density was calculated for the stationary phase.

	Fermentation A (manual feed)	Fermentation B (automatic feed)
Aeration	± 1.2 L/ min (approx. 0.1 vvm)	± 5 L/min (approx. 0.3 vvm)
Stirring rate	600 rpm	600 rpm
Total glucose consumption	3492 g	2577 g
Total ethyl palmitate	54.85 g	162 g
Duration	448 h	359 h
Average optical density	131.9 ±16.7	120.9 ±13.6
Total amount of sophorolipids	9.0 ± 0.2 g	12.1 ± 0.2 g

At the end of both 30 L fermentations, the broth was harvested and stored at 4 °C for further downstream processing. After 48 hours, a brown precipitation could be observed. Analysis on HPLC revealed that these were C16 sophorolipids. When comparing the different fermentations, a striking difference can be seen. The ratio of the acidic and lactonic C16 sophorolipids (Figure 3.20) differs drastically between both fermentations. Fermentation A has a lower amount of lactonic sophorolipids compared to fermentation B. In literature, an effect of low aeration has been described[430–433] resulting in changes in saturation degree of the fatty acids incorporated and different levels of lactonic sophorolipids being produced.

Interestingly, Fermentation B already has a very low aeration compared to the 3 L, but still overall the product mixture remains the same. It has been reported that only at very low aeration rates the composition of the sophorolipids can change, but that from a certain threshold value their composition does not change that much[431–433]. The lower product titer for fermentation A might be explained by this as well, lower expression levels of the enzymes caused by lower oxygen levels and its inhibitory effect on cell growth and maintenance.

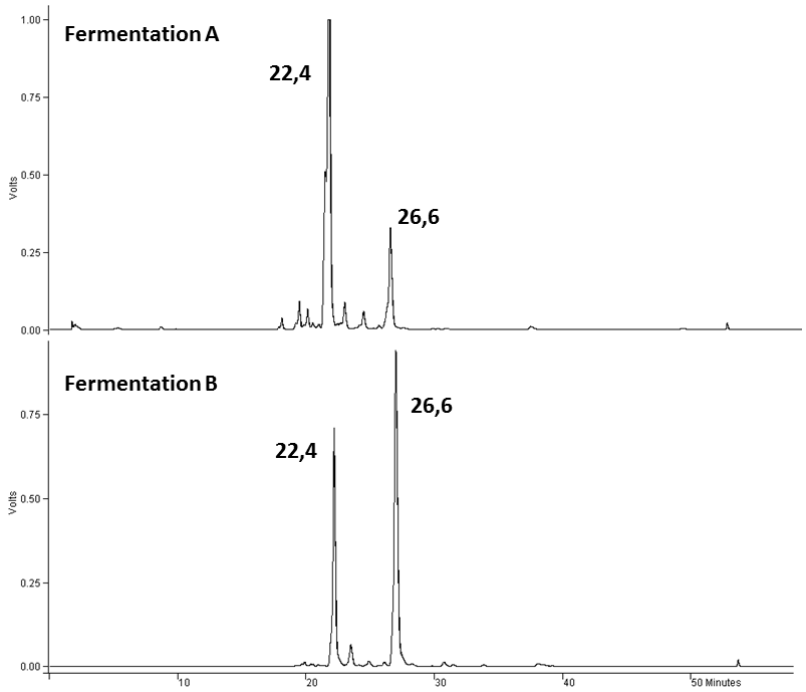


Figure 3.20: Difference between acidic (22.4 minutes) and lactonic (26.6 minutes) di-acetylated C16 sophorolipids.

Substrate specificity of CYP1BMR

From the chromatograms is it clear that CYP1BMR has a high preference towards palmitic acid, yet the true substrate specificity of CYP1 has never been experimentally proven. There are some strong indications that palmitic acid is the main substrate, for example due to the incorporation of this molecule in both cellobiose lipids and the C16 sophorolipids. In *S. bombicola* CYP1BMR some small amounts of stearic and oleic acid sophorolipids are detected as well, suggesting that the substrate specificity is not as strict as initially thought. In *U. maydis*, no cellobiose lipids with other chain lengths than 16 carbon atoms are detected. Chimeragenesis can alter substrate specificity, but when examining the products formed it is clear that the amounts of non-C16 sophorolipids are low. Glucosyltransferase UGTA1 has a clear preference towards longer chain lengths compared with highest activity for 17-S-hydroxy oleic acid. The final sophorolipid products of *S. bombicola* will be a result of the specificities of the chimeric CYP1BMR and both glucosyltransferases. For the *U. maydis* UGT1, substrate specificity has never been researched (IV.3.4.1).

To investigate the CYP1 and CYP1BMR specificity, an overexpression system in *S. cerevisiae* W(R) was used. To ensure that the enzymes would be isolated, both cytosolic and microsomal fractions are tested. The reason for this is the N-terminal region of CYP1 itself. All class II P450 are anchored to the ER membrane, but when CYP1 is analysed for this, no clear anchor could be found in the beginning of the protein sequence. A potential transmembrane helix is situated between residues 148 and 167, much further when compared to the CYP52M1 anchor (Figure 3.21). This might give problems to anchor the P450 to the ER membrane in *S. cerevisiae*. Since CYP1BMR has its own reductase domain, it can be active in the cytosol and convert fatty acids when available.

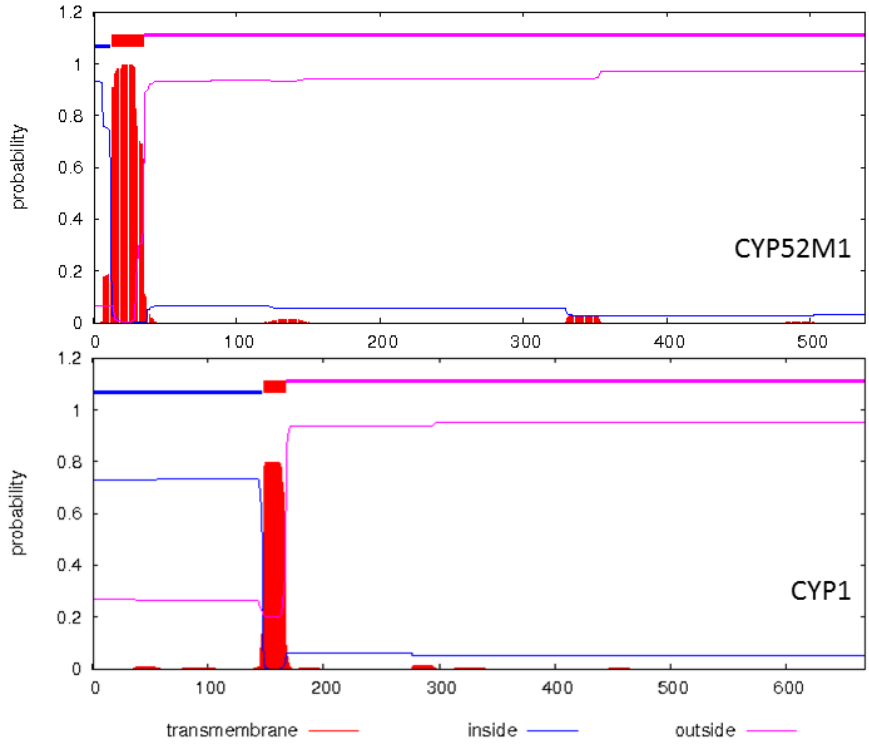


Figure 3.21: Prediction of the transmembrane helices of CYP52M1 and CYP1. Blue represents the part of the enzyme inside the ER, red the transmembrane helix and pink the cytosolic part. Prediction was made using the TMHMM Server v2.0 from the Centre for Biological Sequence Analysis situated at the Technical University of Denmark.

Another possibility lies in the quality of the sequence itself. The coding sequence used in this research was taken from the sequenced genome from the Broad institute database. This is a technical annotated genome of *U. maydis* 521[402]. Since no further data is available, gene prediction software was used. Potentially, the predicted coding sequence of CYP1 might have been faulty. When looking at the sequence, two introns were predicted in the genomic sequence, one of 193 nucleotides and one of 41 nucleotides. Both of them are positioned in the beginning of the genomic sequence. When taking the start of the transmembrane helix (position 148) and moving back towards the start of the protein sequence, a methionine is found at position 105, 43 amino acids before the anchor (Figure 3.22). If the methionine at position 105 would be the correct starting point of the P450, its overall size would be 564 amino acids, a bit longer than the CYP1 enzyme belonging to *Pseudozyma flocculosa*[434], which has a length of 534 amino acids.

Figure 3.22: Structure of the genomic sequence of *U. maydis* 521 CYP1. Red depicts the exons, blue the introns. M is the first methionine before the transmembrane helix (TMH).

In Figure 3.23, an LC-MS chromatogram can be seen for palmitic acid. From all the tested substrates (fatty acids with chain lengths of 12 to 20 carbon atoms), only palmitic acid resulted in detectable product formation for only the microsome fraction of the chimeric CYP1BMR. Though during the growth trials oleic acids were incorporated as well, this was only in very low amounts. No products could be seen for all the other different fatty acids tested, for the cytosolic fractions and for the wild type CYP1. Again very low activity of CYP1BMR is measured. While these results are obtained using *S. cerevisiae* W(R) instead of *S. bombicola* and relied on *in vitro* experiments, it is clear that the P450 chimera is not very active. Even though only palmitic acid was converted to the corresponding hydroxy fatty acid, other substrates might be accepted as well, but at much lower rates. As already mentioned before, the low productivity of sophorolipids in the CYP1BMR strain could easily be explained by low activity of the chimeric enzyme. When taking into account that UGTA1 still has 29% activity towards palmitic acid compared to oleic acid, it is most likely that CYP1BMR is the main cause for the low productivity of the C16 sophorolipids.

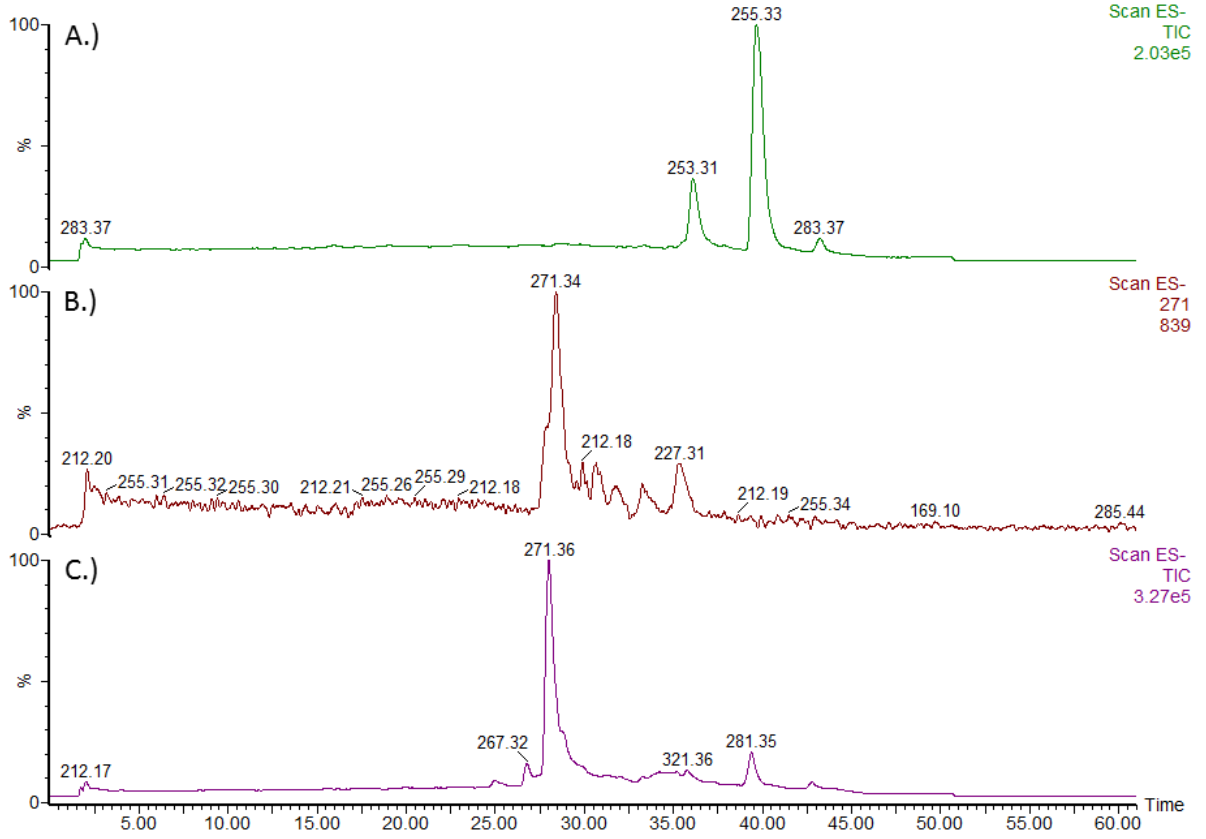


Figure 3.23: LC-MS chromatogram of the CYP1BMR microsome test. The value of 271 m/z corresponds to juniperic acid, the terminally hydroxylated variant of palmitic acid. Palmitic acid has a m/z value of 255, oleic acid corresponds to 281 m/z. A.) A total chromatogram for the microsome enzyme test. B.) Mass scan for m/z 271 C.) Standard of juniperic (16-hydroxy palmitic acid; m/z = 271).

III.4. Conclusion

Introducing a novel P450 in *S. bombicola* to alter the fatty acid tail of the molecules produced is a valuable tool in creating tailor-made glycolipids. Depending on the P450, a shift towards shorter chains of either 16 or 14 carbon atoms can be attained. In the first part of Chapter III, three different class VIII enzymes were investigated in their capability of contributing to novel kinds of sophorolipids. Since these P450s are self-sufficient, they can operate without the need of additional reaction partners. All three have activity towards shorter fatty acid chains. The conducted growth trials did show some minor activity, but the incorporated fatty acids all corresponded to the longer variants with C16 and C18 fatty acid tails. Further enzyme tests proved their activity, even towards shorter fatty acid chains.

It is clear that these enzymes are active in *S. bombicola,* but do not produce the right kind of hydroxy fatty acids. According to literature, replacing the phenylalanine at position 87 in CYP102A1 and related enzymes would enable the production of terminal hydroxy fatty acids. Still, other reports indicate a shift towards more subterminal positions as well. The different alanine mutants of the self-sufficient P450s were tested as well. Both *in vivo* and *in vitro* cell lysate experiments showed no evidence for the claims that these mutants would be able to produce terminal hydroxylated fatty acids and the enzyme tests even showed a reduction of activity for the alanine mutants. Their overall lower activity compared to the wild type enzymes is a major drawback for their use as biocatalysts. Still, modelling and engineering of P450s is a valuable tool for understanding the catalytic activity and steering product formation towards more desirable molecules.

In the second part of Chapter III, heterologous expression of class II P450s was attempted. In the case of CYP1 from *U. maydis*, low amounts of activity could be detected. To make sure sufficient electron transfer happened from the CPR to the P450, a strategy with the generation of chimeric P450s was attempted. The P450s CYP52A4 from *C. maltosa* and CYP1 from *U. maydis* were coupled to the reductase domain of CYP102A1. At the end, activity was detected for both chimeric proteins. Further validation that the BMR domain was responsible for enzyme activity was performed on the CYP1BMR. This further proved that coupling P450s to a reductase is an interesting strategy to increase activity of heterologous enzymes. In the case of CYP1BMR, further optimisation and scale-up of the production process resulted in a final amount of 30 grams of C16 sophorolipids.

Increasing wild type sophorolipid production was attempted as well. By coupling CYP52M1 to the reductase of CYP102A1 and its own *S. bombicola* CPR, it was suspected that these new strains might have a higher productivity. Though it can only be verified on RNA level, more P450 transcripts are present compared to the CPR. By creating a fusion protein, each P450 domain would have its own electron donor. However, this strategy proved to be incapable of increasing productivity. Though the exact reason is unknown, interference of the reductase domain with for example UGTA1 interactions might be a possible explanation. As it was proven in Chapter II, CYP52M1 and UGTA1 are colocalised. Even though the coupled reductase might enhance electron transfer rates, if the next step in the biocatalytic pathway is hindered by it overall lower productivity might be observed.

Chapter IV

Enhancing glycolipid production in S. bombicola

Chapter IV - Enhancing glycolipid production in *S. bombicola*

IV.1. Introduction

Glycolipids are compounds composed of a carbohydrate hydrophilic group coupled to a hydrophobic fatty acid or lipid derived structure. In nature, these compounds have very specific roles in cell signalling and recognition, membrane structure, bacterial motility and so on. Recent years, due to growing concerns about the production and environmental effects of petrochemical derived surfactants, interest in glycolipids produced by means of biotechnology has grown significantly. These molecules find applications in for example medicine[435] and cosmetics[436,437].

Especially bacterial and fungal molecules are being investigated intensely. Well studied examples are the rhamnolipids produced by *Pseudomonas* and *Burkholderia* species and the sophorolipids produced by *Starmerella bombicola*.

Both kinds of surfactants are already commercially available (see Chapter I Table 1.3). However, research to elevate production titers and enhance product uniformity is still continuing. In the case of *S. bombicola*, the sophorolipids are still a mixture of congeners differing in acetylation degree, integrated fatty acid tail and potential lactonisation of the molecules. Recent attempts to engineer this yeast to enhance product uniformity resulted in strains capable of producing either only lactonic or acidic sophorolipids[174]. Control over the fatty acid tail can be done by feeding special substrates[392,393,404,438]. As *S. bombicola* prefers to incorporate oleic acid using a substrate rich in these molecules can increase product uniformity. These strategies result in industrially interesting titers of the target molecules.

Engineering *S. bombicola* for the production of other kinds of glycolipids than sophorolipids has been attempted as well. Glucolipids, intermediates in the sophorolipid pathway, are being produced as a final product in strains lacking the second glucosyltransferase UGTB1. However, compared to the wild type *S. bombicola*, production is severely impaired with yields of about 4 g/L after 14 days of production[317].

Another engineering example are the cellobiose lipids, naturally produced by fungi like *U. maydis* and *P. flocculosa*. The hydrophilic head group of these molecules only differs from sophorose by how the two glucose molecules are coupled. In the case of cellobiose both glucose molecules are connected β-1,4 instead of β-1,2 for sophorose. Introduction of the glucosyltransferase of *U. maydis* resulted in a strain producing novel kinds of glycolipids in *S. bombicola*. However, again the titers remained below the 5 g/L mark and only 10% of the glycolipids produced were presumed to be cellobiose lipids[318].

In the following chapter, optimization of the productivity of the glucolipid and cellobiose lipid producing strains is described. First of all, the genetic background of the existing strains was evaluated. Starting from these results, new strains were developed to increase production titers, and finally the best performing strain was scaled-up to a 3L bioreactor.

IV.2. Material and Methods

IV.2.1. Strains and culture conditions

Culture conditions can be found in II.2.1. Strains used for glucolipid and cellobiose lipid production are the ones created by Saerens *et al*.[317] and Roelants *et al*.[318]. More details on their genetic background and genetic modifications are given in the Results and Discussion section. New strains were derived from the *ura3* auxotrophic PT strain[311]. Protein expression for protein characterization was done in *E. coli* BL21 (Bioké; The Netherlands).

IV.2.2. Molecular Techniques

IV.2.2.1. General techniques

The general molecular techniques used are identical to the ones described in II.2.2.

IV.2.2.2. Construction of integrative cassettes

Cassettes targeting the ugtb1 locus

A first strain (B11new) was constructed with the original knock-out cassette described by Saerens *et al*.[317], but this time using *S. bombicola* PT36 as a background instead of the G9 auxotrophic strain used in previous research[310,317,318] (Figure 4.1 A). The other production cassettes are derived from an earlier constructed plasmid (BS_EC_0269_ADH_UGTB1_loc) for integration at the *ugtb1* locus. The 5' and 3' homologous regions of this new plasmid correspond to parts of the alcohol dehydrogenase gene (*adh*) upstream of the sophorolipid cluster and of the sophorolipid transporter (*mdr*) situated downstream of the *ugtb1* locus. The gene of interest can be integrated between the *ugtb1* promotor (pUGTB1) and terminator (tUGTB1). As a marker the *ura3* marker is used. It is inserted between the 5' homologous region and the terminator of *ugtb1*. It is controlled by its endogenous promotor pURA3, but terminated by the bidirectional terminator t14190 (terminator of genes *cabom01g14190* and *cabom01g14200*). This way, both the *ura3* and *adh* transcripts will be terminated correctly. Removal of the *ura3* marker from engineered *S. bombicola* strains is possible by amplification of a recovery cassette and subsequent transformation and selection on 5-fluoroorotic acid (5-FOA). The cassette can be created from genomic wild type *S. bombicola* DNA with primers P337_3'ura3delUGT1Rev and P653b_FOR_vectorUgtUMKI_ura3_UGTB1locus_tUGTB1.

A second strain created with this new plasmid has its *ugtb1* coding sequence replaced with that of *ugta1* (Figure 4.1 B) called BS_EC_0321_UGTA1_KI_UGTB1_KO. A third one was engineered by coupling the promotor and terminator of *ugtb1* without a coding sequence between (Figure 4.1 C) with the name BS_EC_0380_pGEMT_PUGTB1_TUGTB1_Ura_CORRECT. The same plasmid for the *ugbt1* locus was used to express the *U. maydis ugt1* at the *ugtb1* locus as well (Figure 4.1 D) named BS_EC_0305_UGT1_Kol1. In Table 4.1, an overview is given of the primers used to create these different constructs. In Figure 4.2, a schematic representation of the sophorolipid gene cluster can be seen. The region where the different cassettes are integrated can be seen inside the red square.

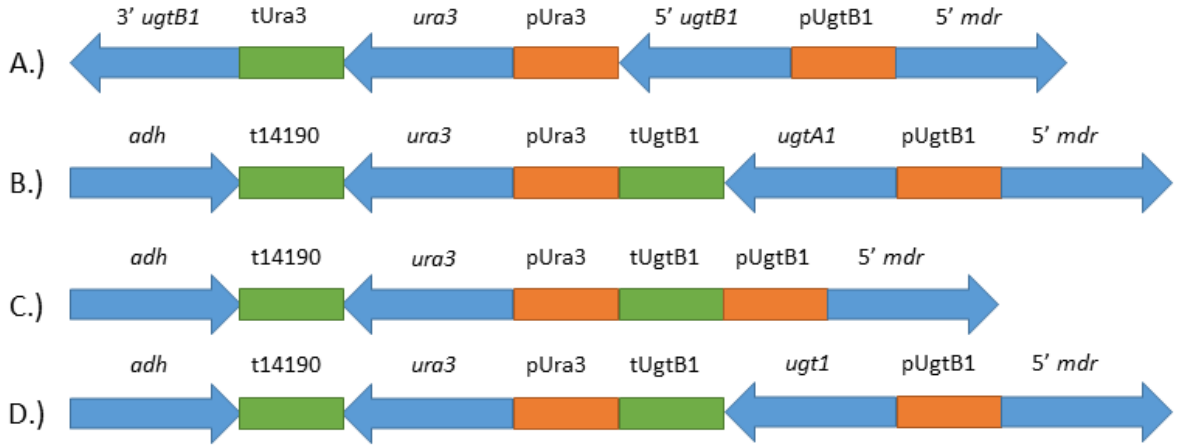


Figure 4.1: Different knock-in and -out cassettes for the development of glucolipid (A, B and C) and a cellobiose lipid (D) producing strain. A.) Original knock-out cassette developed by Saerens *et al.*[317]. B.) New developed cassette for introduction of the *ugta1* coding sequence at the *ugtb1* locus. C.) Knock-out cassette for the *ugtb1* locus derived from the new plasmid for recombination at the *ugtb1* locus. D.) New *ugt1* knock-in cassette.

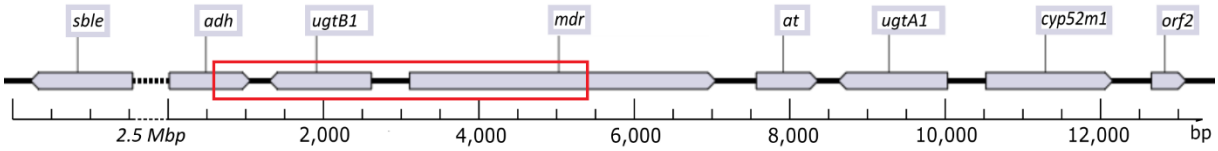


Figure 4.2: Schematic overview of the sophorolipid gene cluster. The location where the different cassettes are integrated is depicted within the red square.

Table 4.1: Primers used for the creation of the different *ugtb1* knock-out and the *ugta1* knock-in cassettes. Non-binding regions are separated by a dash.

Primer name	Primer sequence
P337_3'ura3delUGT1Rev	ACCCAGTTCCTGGCACTATTG
P350_CepBURA3F	TCTGGCGAATAACAGCACTC
P653b_FOR_vectorUgtUMKI_ura3_UGTB1locus_tUGTB1	CCTGGCTCTTTTTCTAGATATGTCTG
P662_pJetMakeFor	ATCTTTCTAGAAGATCTCCTACAATATTC
P663_pJetMakeRev	ATCTTGCTGAAAAACTCGAGCCATC
P983_ADHMDRinfpJETFor	CTTCCGGATGGCTCGAGTTTTTCAGCAAGAT-ACCCAGTTCCTGGCACTAT
P984_ADHMDRinfpJETRev	GAGAATATTGTAGGAGATCTTCTAGAAA-GATCCTCGCCACCACCTAGTT
P987_tUGTB1Rev	TTCTGCTCTCAACACCGAGTGTAG
P988_tCabomInfADHFor	GCCAAGTCCCTTTAA-ATTTGCTAGGCTGCAGTGCTTATA
P989_pUraInftUGTB1Rev	TTGAGAGCAGAAGTT-GAACCATGATGGCAGTGTTC
P1426_PromGTIIandMDR_Rev	AAGCCAAAATCAGAGAGTGGGACC
P1596_ACP_check_out_fw_2	TATGCTGGTGCATCCAGTTG
P1822_UGT1InfFor	AGAAAAAGAGCCAGG-TCAAAAGAGGCGGACTTCTGC
P1823_UGT1InfRev	AATAAAGTCTCCTAT-ATGGCGACCGAACATATTCTTTTG
P1861_UGTB1URAKO_lig_fw	ATAGGAGACTTTATTGTTTTATGC
P1862_UGTB1URAKO_lig_rv	CCTGGCTCTTTTTCTAGATATGTC
P1863_UGTB1URAKO_cass_fw	AAAGCAGCATAGACACGAGCTGAC
P1864_UGTB1URAKO_cass_rv	TGCTTCCGGCTCGTATAATG
P1894_UGTB1_KO_Cass_rv	CCTCGCCACCACCTAGTTTGTC
P2246_UGTB1KOCass_Karen_fw	GAGAGTGGGACCTGATTC
P2247_UGTB1KOCass_Karen_rv	CTGCTCTCAACACCGAGTGTAG

Cassettes targeting the ura3 locus

To circumvent potential negative regulation, the coding sequence of *ugt1* was introduced at the *ura3* locus under the control of the phosphoglycerate kinase promotor (pGKI)[312]. This way no need for extra engineering inside the cluster is necessary. Introduction of this construct was done in *S. bombicola* PT36 Δ*ugta1*. The expression cassette contains the nourseothricin acetyltransferase gene (NAT), necessary for nourseothricin resistance. Expression of the resistance gene is achieved by the glyceraldehyde 3'-phosphate dehydrogenase promotor (pGAPD). Both promotors show high expression during the exponential phase and no downregulation during the stationary phase[368]. The primers used for the construction of this expression cassette (BS_EC_0395_UmUGT1_OE_Uraloc_kol4) can be found in Table 4.2. The vector backbone was amplified from plasmid BS_EC_0291_pGAPD1555_nat1.

Construction of the plasmid for UGT1 characterisation in E. coli BL21

Expression of the *ugt1* gene in *E. coli* BL21 was done using the commercial pET21a plasmid. Expression is achieved by induction with isopropyl β-D-1-thiogalactopyranoside (IPTG). The coding sequence of *ugt1* was equipped with a 6xhis tag at its C-terminus. The primers needed for amplifying the vector backbone and *ugt1* insert can be found in Table 4.2. The final expression plasmid is BS_EC_0392_pET21a_UmUGT1_kol1.

Table 4.2: Primers used for the creation of the *ura3* locus *ugt1* overexpression cassette and the pET21a plasmid for expression in *E. coli*. Non-binding regions are separated by a dash.

Primer name	Primer sequence
P1_FOR_URA3v	AGAACAAGGCCGAGTATGTC
P32_REV_cassette	GTCAGATTAGCCTCCGACATAG
P33_FOR_checkcassIN	CCATACTCAAGCGCGAACAC
P34_REV_checkcassIN_URA3	GATGTCGAATAGCCGGGCTGCTAC
P35_REV_checkcassIN_DOWN	GAGCTCAAGACGCGTTTACTCAATGC
P306_trypcastestrev	ACTAGCCGAGATCGTTGC
P1600_GAPD_insert_check_rv	GTCCCGATACTGGGATGTTG
P2365_nat1_sequencing_1	ATCGAGGCACTGGATGGGTCCTTCAC
P2483_UmUGT1GibpPKIFor	GTAAACTACCCAAGGTCCTCCAAACCAGAAAAA-ATGGCGACCGAACATATT
P2484_UmUGT1GibtTKRev	CCGTGTTTCAGTTAGCCTCCCCCATCTCCC-CTAAAAGAGGCGGACTTC
P2482_pGKIRev	TTTTTCTGGTTTGGAGGACCTT
P2475_pET21promrev	CCCGTATATCTCCTTCTTAAAG
P2476_pET21termfor	GATCCGGCTGCTAACAAAGC
P2477_UmUGT1GibpETpromFor	CTTTAAGAAGGAGATATACGGG-ATGGCGACCGAACATATTCTTTTGGTCTGCT
P2479_UmUGT1hisGebpETtermRev	CCTTTCGGGCTTTGTTAGCAGCCGGATCTCA-GTGATGGTGATGGTGATGAAAGAGGCGGACTTCTGC
P2388_tTK_For	GGGAGATGGGGGAGGCTAAC

IV.2.3. UGT1 enzyme assays

IV.2.3.1. *E. coli* BL21 *ugt1* expression

Protein expression

One colony of the desired *E. coli* BL21 strain was inoculated in 5 mL of LB medium with 100 µg/mL ampicillin and incubated overnight. This preculture was used for inoculation of 250 mL LB medium containing 100 µg/L ampicillin. The Erlenmeyer flasks were incubated at 37 °C at 200 rpm. When the optical density of the culture reached 0.6, 250 µL of an 1M IPTG stock was added and further incubated at 16 °C at 200 rpm overnight. The cells are harvested by centrifugation at 9000 rpm for 30 minutes at 4 °C. The cell pellets were stored for at least 1 hour at -20 °C before lysis and purification.

Lysis of E. coli BL21

The pellet originating from 250 mL broth were suspended in 8 mL lysing buffer (300 mM NaCl, 50 mM $NaPO_4$, pH 8, 1 mg/mL Lysozyme) and placed on ice for 30 minutes. After incubation on ice, the cell suspension was sonicated twice with a duty cycle of 50 % and an output control of 3. All sonication steps were performed on ice. Afterwards, the cell debris was removed by centrifugation at 10.000 rpm at 4 °C during 30 minutes. The supernatant was filter sterilized through a sterile 0.22 µm sterile filter.

6xHis tag protein purification

The His-tag purification columns were packed before the purifications. This was done by suspending the Ni-NTA agarose in its original buffer and container by agitation and subsequently pouring 1.5 mL of the agarose resin suspension into a 10 mL purification column and allowing the resin to settle by gravity. The supernatant remaining on top of the resin was removed and the resin was rinsed with 6 mL of sterile distilled water. The resin was settled by gravity. Afterwards, the supernatant was removed and the resin was rinsed a second time, with 6 mL of equilibrium buffer (10 mM imidazole in 50 mM $NaPO_4$, 300 mM NaCl; pH 8). After settling and removing the supernatant, this rinsing step was repeated twice.

A volume of 8 mL of the cell lysate was loaded on the prepared Ni-NTA column. The column was incubated for 60 minutes at 4 °C and gentle agitation was applied to keep the resin suspended in the lysate. After the hour, the resin was settled by gravity, the supernatant was recovered and stored at 4° for SDS-PAGE analysis. The resin was washed with 8 mL wash buffer (30 mM imidazole in 50 mM $NaPO_4$, 300 mM NaCl; pH 8) and again allowed to settle by gravity. The supernatant was removed and stored for further analysis by SDS-PAGE. This step was repeated three more times. Each time a sample of the flow-through was kept for SDS-PAGE analysis afterwards. Finally, the His-labeled proteins were eluted with 2x 4 mL elution buffer (250 mM ididazole in in 50 mM $NaPO_4$, 300 mM NaCl; pH 8).

SDS-PAGE

An SDS-PAGE was performed on all the collected fractions to assess any losses of the His-tag purified enzyme. The composition of both the stacking and resolving gels can be found in Table 4.3. First the resolving gel was made and allowed to settle for 30 minutes. On top of the mixture a layer of isopropanol was added to straighten to top layer of the gel. Next the stacking gel was poured.

Table 4.3: Composition of the stacking and resolving gels used for SDS-PAGE.

Components	Stacking gel (5%)	Resolving gel (10%)
Distilled water (mL)	2.85	4.1
30% Acrylamide/Bis (mL)	0.85	3.3
1.5 M Tris-HCl pH 8.8 (mL)	/	2.5
0.5 M Tris-HCl pH 8.8 (mL)	0.25	/
10 % SDS (µL)	50	100
10 % Ammonium Persulfate (µL)	50	50
TEMED (µL)	5	5
Total (mL)	**5**	**10**

A volume of 10 µL of sample was added to 20 µL of sample buffer (Table 4.4). The purified samples have a standard concentration of 1 mg/mL. For crude cell extracts and the initial flow-through of the extract through the purification column, only 5 µL of sample was taken diluted in sample buffer. The insoluble fraction after cell lysis was diluted in 50 mL of distilled water and 10 µL was added to the sample buffer. The samples were incubated at 95 °C for 5 minutes. 10 µL of the prepared sample was loaded on the gel and the gel was run at 200 V for 60 minutes.

After the gel was run, excess SDS was rinsed from the gel with distilled water and the gel was fixed with fixation buffer (40% Ethanol and 10% acetic acid in mQ) for 15 minutes. The fixed gel was washed with distilled water twice for 2 minutes. Staining of the gel was done with staining solution (34% (v/v) MeOH, 3% (w/v) H_3PO_4, 17% (w/v) $(NH_4)_2SO_4$ and 0.2% (w/v) CCB-G-250) overnight at room temperature. An orbital shaker was used for efficient mixing of the staining solution and assuring full coverage of the gel. After incubation, the gel was destained using distilled water and incubated again using an orbital shaker.

Table 4.4: Composition of the sample buffer required for loading the samples on the SDS-Page gel.

Components	Volumes
Distilled water (mL)	3.55
0.5 M Tris-HCl pH 6.8 (mL)	1.25
Glycerol (mL)	2.5
10 % (w/v) SDS (mL)	2.0
0.5 % (w/v) Bromophenol blue (µL)	200
Total (mL)	**10**

Buffer exchange

A buffer exchange was needed before enzyme tests were performed. To exchange the buffer, Millipore Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-30 membrane were used. A volume of 4 mL of the enzyme solution was applied on the filter and centrifuged at 4 °C and 4500 rpm for 9 minutes or until a volume of 0.5 mL remained on the filter unit. The filter units were washed with the washing buffer four times. To prevent clogging of the filter, the remaining liquid was disturbed by pipetting with the washing buffer before each washing step. Store the enzyme at -20 °C.

Enzyme tests

The enzyme tests were performed as described by Saerens *et al*.[176] As glycosyl donor UDP-glucose was tested. Glycosyl acceptors were juniperic acid and hydroxy fatty acids derived from sophorolipids by acidic hydrolysis[16].

The enzyme mixtures were sampled after 3 hours of incubation at 30 °C. Extraction of the glycolipids was done by adding 440 µL ethyl acetate and 10 µL acetic acid to the reaction mixture. After vigorous mixing, the samples were centrifuged at maximum speed for 5 minutes. The supernatant was then analysed on LC-MS for glycolipids with a scan range from 400 to 800 m/z as described in II.2.4.2.

IV.2.3.2. *S. bombicola* enzyme tests

Cell lysates of the *ugt1* overexpression strain were created and subsequently tested as described in Saerens *et al*.[176] Sampling and analysis were equal as for the *E. coli* BL21 enzyme tests described in IV.2.3.1.

IV.2.4. Follow-up of growth and glucose consumption

Sampling and analysis of glucose, optical density and glycolipids is performed as described in II.2.4.1

IV.3. Results and discussion

IV.3.1. Introduction

Production of novel glycolipids in *S. bombicola* has been achieved the recent years by genetic engineering of a *ura3* auxotrophic strain. In this strain, either the second glucosyltransferase *ugtb1* was knocked-out for glucolipid (Figure 4.3 A) production or replaced by *ugt1* from *U. maydis* for cellobiose lipid (Figure 4.3 B) production. Still, the amounts produced by both strains were not at the same level as those of the wild type sophorolipids[318]. Several hypotheses can be made on the cause of this drop in product titers. First of all, altering the structure of the sophorolipid gene cluster might have an effect on the regulation of its genes. This effect can be due to deletion of essential regulatory sequences embedded in the promotors, terminators or even the coding sequences themselves. Deletion of these regulatory regions can influence the association of transcription factors or other DNA binding proteins necessary for gene expression. Another possibility related to regulation is chromatin structure. In fungi, regulation of many secondary metabolite gene clusters is due to epigenetic marking of the histones involved in DNA packaging and thereby accessibility of the DNA for other proteins[439,440].

A third reason for the lowered production can be transporter specificity. Though the MDR transporter of *S. bombicola* is classified as a multi-drug resistant ATP-binding cassette transporter[39], its specificity has never been tested. If this transporter is specific towards the sophorose head group, different carbohydrate structures might impair the transport of non-sophorolipid glycolipids across the plasma membrane.

A fourth possibility is the strain used for engineering. In the past, two *ura3* auxotrophic strains were developed. *S. bombicola* G9[310] was the first *ura3* negative strain developed for further engineering. It was selected as a spontaneous mutant when the wild type yeast was grown on medium containing 0.1% 5-fluoroorotic acid (5-FOA). Though this strain is effectively *ura3* negative, using a spontaneous mutant can have an influence on other factors as well since mutations in other regions or even altered epigenetic regulation can occur. Therefore, a second *ura3* negative strain was developed by deleting the *ura3* coding sequence through targeted engineering methods, only leaving the promotor and terminator at the original locus. This strain is called PT36.

In the following experiments, these two strains will be evaluated for the production of different glycolipids. Thereby, a better understanding of their production can be achieved.

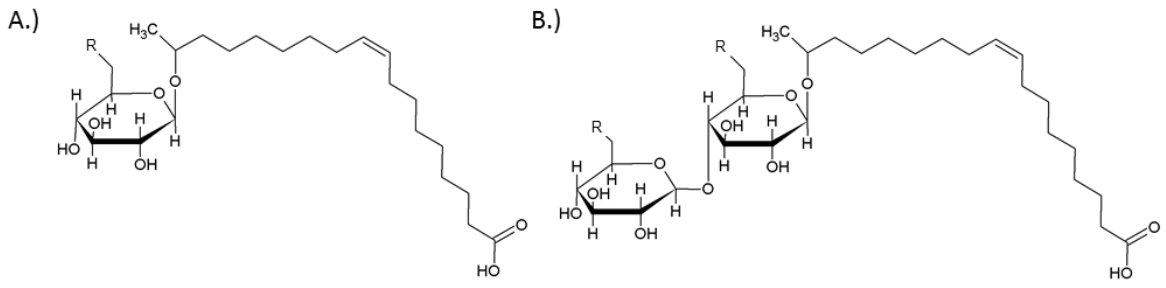


Figure 4.3: Molecular structures of A.) glucolipids and B.) cellobiose lipids produced by *S. bombicola*. The R-group on the 6' carbon atom of each glucose molecule can be either an hydroxy or acetyl group.

IV.3.2. *S. bombicola* G9 versus PT36

As a difference in production between the two *ura3* auxotrophic strains, the G9 strain and the PT36 strain, can explain some details about the fitness of both strains, a production experiment was conducted with both strains on production medium described by Lang et al.[370] complemented with uracil (0.3 g/L) and uridine (0.5 g/L). At the start of the secondary phase, 37.5 g/L rapeseed oil was added. In Figure 4.4, optical density and sophorolipid production are plotted. A striking difference between the G9 and PT36 can be seen for both parameters evaluated.

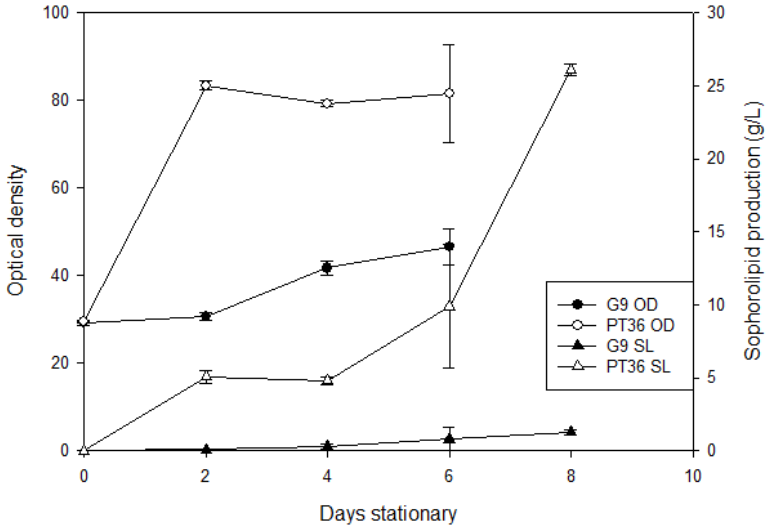


Figure 4.4: Optical density and sophorolipid production for the *ura3* negative *S. bombicola* strains G9 and PT36. G9 is a spontaneous *ura3* negative strain while the PT36 is an engineered deletion strain. Error bars depict standard deviation. Sample size for each strain was 3.

One easy can see that for biomass, the G9 only reaches a value of 40 while the wild type easily reaches 80, a two-fold difference. For the sophorolipids, titers of 25 g/L can be achieved for the PT36, rivalling the wild type production levels. The G9 hardly rises above 1 g/L. This clearly shows that the engineered *ura3* negative PT36 is superior compared to the G9 strain, both in fitness and sophorolipid production. The true nature of this difference is not known, but possibly additional mutations are present in the G9 genome influencing the fitness and production capacity of this strain.

IV.3.3. Engineering *S. bombicola* for glucolipid production

IV.3.3.1. Characterization of glucolipid producing strains

Based on the striking superiority of the PT36 strain over the G9 strain it was decided to develop new glucolipid producing strains (see also Figure 4.1). A first one was made using the same expression cassette as used for the already existing G9 glucolipid strain (B11)[317], but this time in a PT36 background. This strain is referred to as B11new. Other cassettes were designed and introduced as well, in an attempt to circumvent potential negative influence of the old construct on production since parts of the *ugtb1* coding sequence are still present in the original recombination cassette. They might have a negative influence on for example cluster regulation, preventing the efficient expression of the remaining sophorolipid genes. One new

strain is again a knock-out of the *ugtb1* gene, but this knock-out was made by fusing the promotor and terminator (strain UGTB1_PT). Finally, the *ugtb1* coding sequence was replaced by the one of *ugta1*, effectively doubling the copy number of the gene producing the glucolipids (strain A²).

Characterisation of the novel strains was done in a production experiment using rapeseed oil. As a control, the original glucolipid producing strain B11 was taken along. Besides glucolipid production, consumption of glucose and optical density were analysed. During the production trial, a clear visual difference was noticed between the new strain A², carrying two copies of *ugta1*, and the G9 derived B11 (Figure 4.5). The original glucolipid strain B11 shows a stronger yellow colour compared to the new strains. A similar observation was made for the B11new and UGTB1_PT strain. Besides the colour, a brown sediment could be detected in the Erlenmeyer flasks of the new glucolipid strains, strongly resembling non-soluble glycolipids.

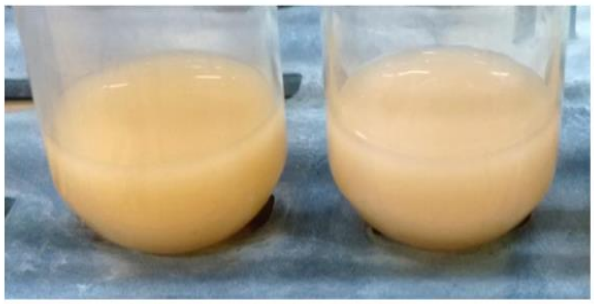


Figure 4.5: Colour difference between the original B11 glucolipid strains and the strain carrying the first glucosyltransferase twice (A²-strain).

When looking at the optical densities and glucose consumption of the different strains, it is clear that the wild type *S. bombicola* reaches a high optical density faster than the glucolipid strains (Figure 4.6). Though the difference becomes negligible at the end of the experiment for the A², B11new and UGTB1_PT strains, for the original B11 strain the discrepancy increases during the production trial. For glucose consumption, a similar observation can be made. The wild type, A², B11new and UGTB1_PT strains consume the glucose within 8 days after inoculation while in the culture of B11 still 15 g/L remains at this time point. This comes down to an average daily consumption of 9.3 g/L.day for the B11 compared to 15 g/L.day for the other strains.

Analysis of the glucolipids production by the different strains was done by TLC during the growth trial and by UPLC and LC-MS afterwards (see II.2.4.2). To be able to compare the different strains, only ethanol samples were used (II.2.4.2). UPLC analysis shows that the B11 strain is capable of producing glucolipids, but only in low amounts compared to the other glucolipid producing strains (Figure 4.7). Analysis of the molecules on LC-MS proved that the produced glucolipid mixture of all strains is similar in profile. Molecular masses of 504, 502 and 500 are easily detected and correspond to acetylated glucolipids with respectively stearic, oleic and linoleic acid incorporated. Other masses corresponding to non-acetylated glucolipids were detected as well with a mass of 460 Dalton (Figure 4.8).

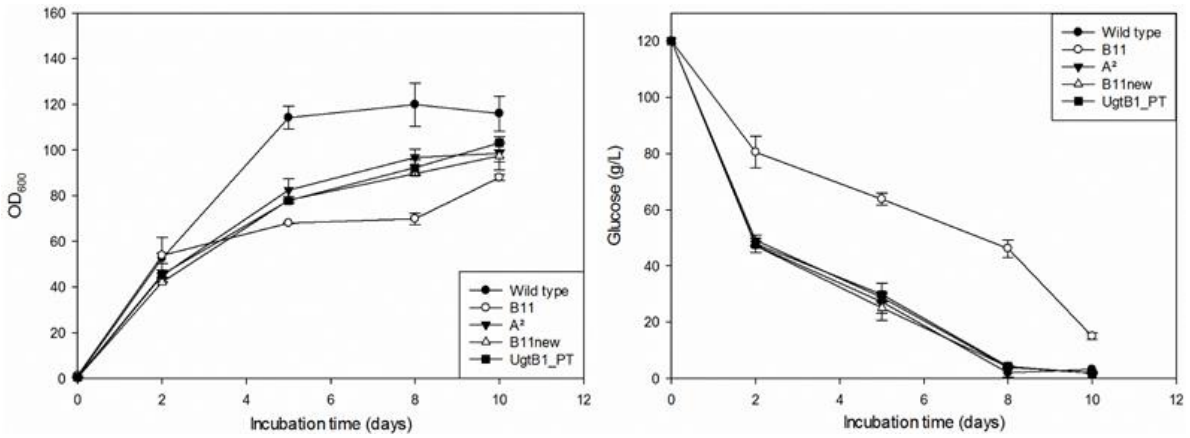


Figure 4.6: Optical density and glucose consumption for the wild type *S. bombicola* and the B11, B11new, UGTB1_PT and A² strains. Error bars depict standard deviation. Sample size for each strain was 3.

Lab-scale purification of the glycolipids was performed by solvent extraction using methyl tert-butyl ether (MTBE), hexane and water. It appeared necessary to wash the glycolipids first with hexane to remove any fatty acids and oils left from the growth trial. This step is essential to be able to extract the glucolipids. Total production in shake flask for the A² strain amounted to 31.52 g/L, 27.17 g/L for the B11new strain and 31.03 g/L for the UGTB1_PT. These concentrations are similar to sophorolipid production values found in literature for other engineered strains[174]. A significant difference is the glucolipid concentration obtained with the original B11 strain, which amounts to 3.55 g/L. No significant difference could be found between the new strains.

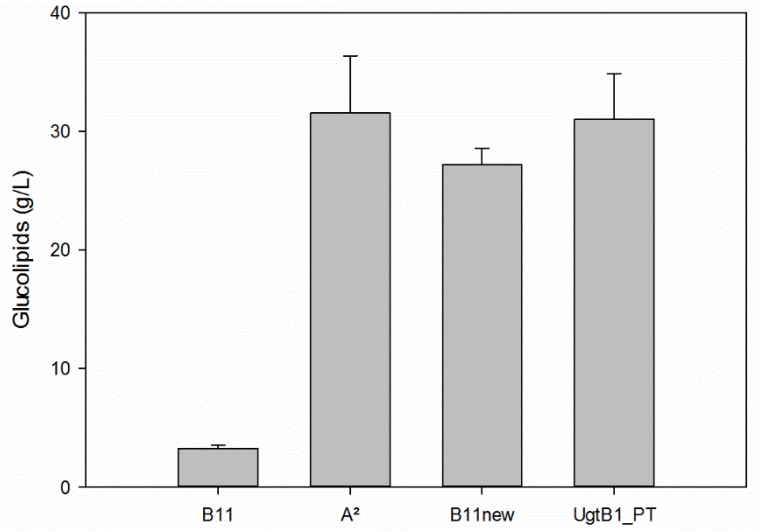


Figure 4.7: Concentration of glucolipids 10 days after inoculation produced by the old B11 strain and the new developed A², B11new and UGTB1_PT strains. Error bars depict standard deviation. Sample size for each strain was 3.

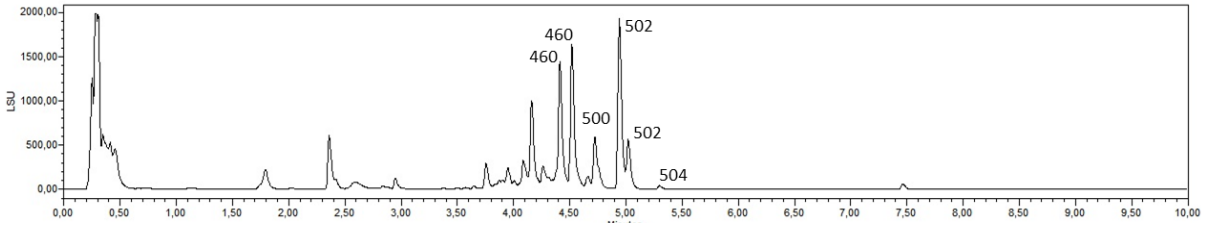


Figure 4.8: UPLC chromatogram of the glycolipids produced by the *S. bombicola strain* carrying two copies of the *ugta1* gene. The main glucolipids present are C18:1 non- and mono-acetylated ones (mw 460 and 502 Dalton).

Curiously, several masses corresponding to molecules with two glucose moieties coupled to the same fatty acids were present among the glucolipids (Figure 4.9). These are for example 705 m/z corresponding to a di-acetylated molecule with an oleic acid tail. Other masses like 661 m/z and 663 m/z are also significantly present, representing molecules with one acetyl-group and respectively a stearic or oleic acid tail. Interestingly, no masses corresponding to lactonized molecules could be detected, suggesting that glucolipids as well as the higher mass molecules are not substrates of the lactone esterase. If lactonized molecules would be present, they would elute from the column between 27 and 33 minutes, but no such signals could be detected.

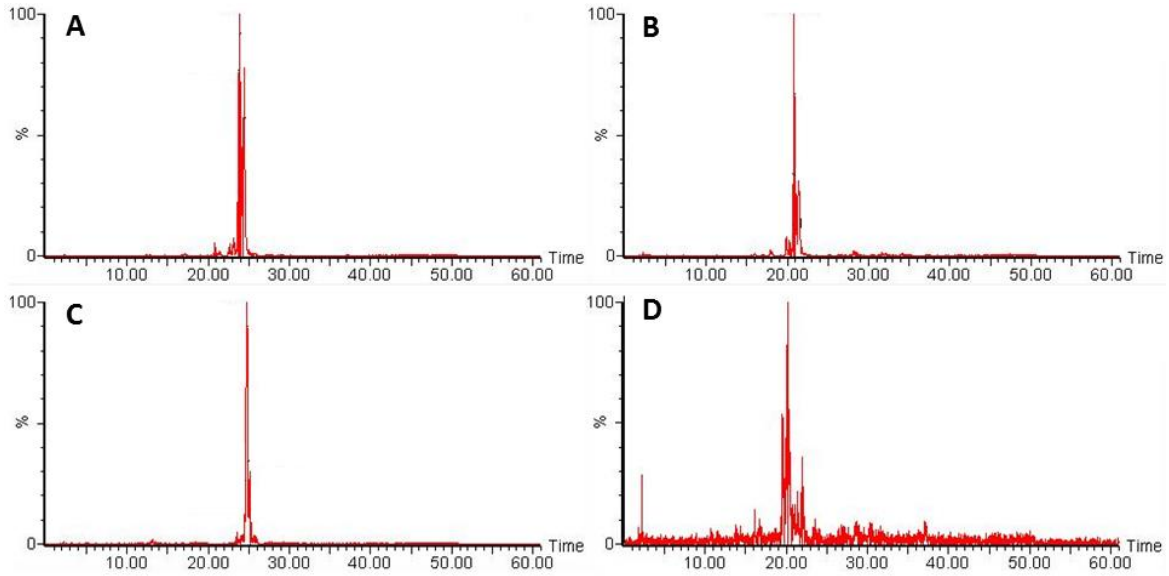


Figure 4.9: Mass scan for glycolipids produced by the A² glucolipid strain containing an oleic acid tail. A = 459 m/z (non-acetylated C18:1 glucolipid); B = 621 m/z (non-acetylated C18:1 bola-glucolipid); C = m/z 705 (di-acetylated C18:1 bola-glucolipid); D = 687 m/z (lactonized di-acetylated C18:1 glycolipid). Lactonized molecules elute at 30 minutes. As can be clearly seen, no such signal is present.

The promiscuity of UGTA1 might clarify the presence of several molecules with higher masses. These molecules do not correspond to disaccharides coupled to fatty acids like sophorolipids or cellobiose lipids, but might be a novel class of glucolipids with one glucose unit coupled to both the hydroxyl and carboxyl group. A visual representation of the proposed structure of these molecules can be found in Figure 4.10. The readiness of UGTA1 to glucosylate the carboxyl group of fatty acids is observed as well in *S. bombicola* strains producing bola-type sophorolipids[423]. Though bola-amphiphilic sophorolipids are only produced at high amounts when the acetyltransferase and lactone esterase both are knocked out, bola glucolipids readily appear in the new strains without these modifications.

One explanation might be that the molecules reside longer in the cytosol, giving UGTA1 more time to glucosylate the carboxylic group. This longer presence can be triggered due to slower transport of the molecules across the membrane. When looking at the new glucolipid strains, production levels are on par with the wild type as well, making the hypothesis about slower transport a difficult one to maintain. Still, the exact reason for this "bolafication" of glycolipids a mystery.

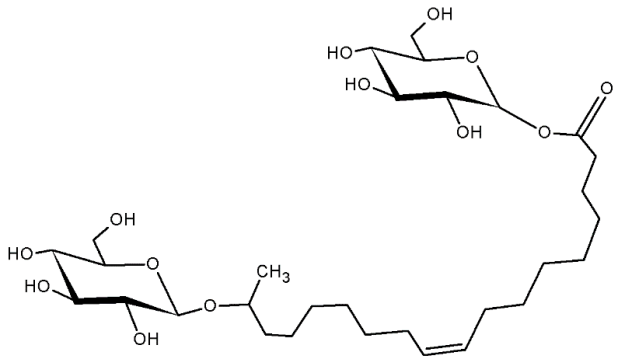


Figure 4.10: Proposed structure of a non-acetylated bola-glucolipid with an oleic acid linker.

IV.3.3.2. 3 L Bioreactor for A²-strain

Upscaling in a 3L bioreactor was performed for the A²-strain. In the beginning, a start volume of 1.2 L was present. A late exponential culture of 200 mL was used as inoculum. Oleic acid was added as a substrate every day in shots of 18 mL. When glucose levels dropped below 30 g/L, a 60 % glucose solution was added until a final concentration of 100 g/L was obtained. The fermentation time between each glucose addition will be referred to as a phase (not to be confused with a cyclic fermentation). Acidity of the broth was kept at pH 3.5 by adding aliquots of 5 M NaOH. This level was reached after 23.8 hours, another indication besides production that the cells were transitioning from exponential growth to stationary production. At the end of the fermentation, a final volume of 1.55 L was attained due to the different additions of glucose and oleic acid.

Consumption of glucose appeared normal during the fermentation compared to other strains tested in this PhD and literature[174] (Figure 4.11). During the first phase including the non-productive exponential phase, 25.5 g/L.day was consumed. When looking only at the stationary phase, this value rises to 30.3 g/L.day. The second phase showed a lower consumption of glucose with only 16.9 g/L.day while the third uncompleted phase this value dropped to 8 g/L.day. First signs of production were detected on TLC after 23.8 hours.

Microscopic analysis of the broth revealed large globules devoid of any cells. After centrifugation, a distinct layer could be observed resting above the cells. When heated to 60 °C, this layer rapidly melted into a brown oily substance. Purification of the molecules was performed by the Bio Base Europe - Pilot Plant because of the large volume of both broth and molecules produced.

HPLC analysis of the different glycolipid samples from the fermentation shows a total production of 206.6 grams of glycolipids. At the first refeeding of glucose, the concentration of glycolipids amounted for 60.35 g/L. When looking at the productivity this results in 0.61 g/L.h during the entire phase. When excluding the non-productive exponential phase, this value rises to 0.81 g/L.h. For the second phase productivity seems similar to the first phase, but no exact figure could be calculated due to a missing data point before the second glucose refeeding. For the part of the third phase sampled, a value of 0.67 g/L.h was obtained, again not differing much from the first one.

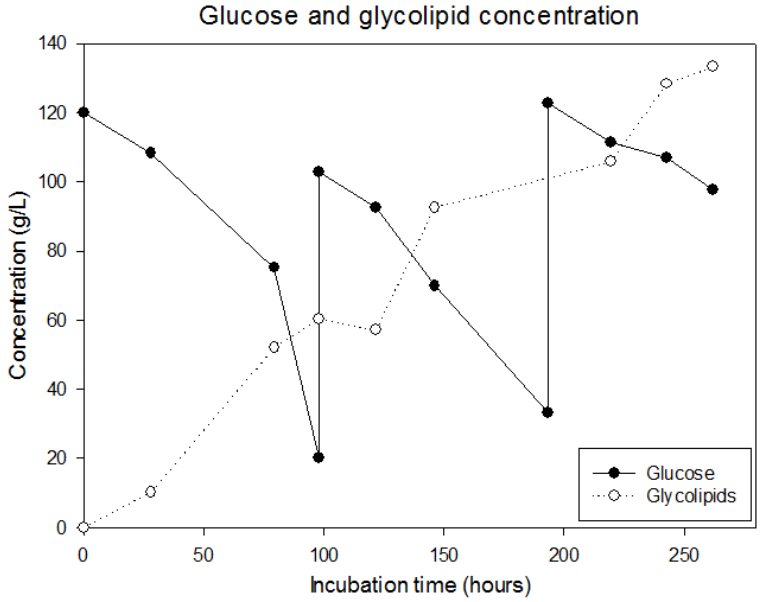


Figure 4.11: Evolution of glucose and glycolipid concentration during the fermentation of the novel A² strain.

IV.3.4. Optimization of cellobiose lipid production

During previous research it has been shown that modification of the carbohydrate moiety from a sophorose to either a glucose or a cellobiose (4-O-β-D-glucopyranosyl-β-D-glucopyranose) is possible[318]. However, production remained limited for both molecules. Moreover, the cellobiose lipid producing, G9 derived *S. bombicola Δugtb1::ugt1* described by Saerens *et al.*[318] produces a glycolipid mixture consisting of maximum 25% cellobiose lipids, the remaining fraction being glucolipids. The large remainder of glucolipids in the produced glycolipid mixture makes it difficult to isolate the cellobiose lipids. In order to enhance product uniformity, the *Δugtb1::ugt1* cassette described by Saerens *et al.* carrying the *ugt1* gene from the cellobiose lipid producer *U. maydis* was introduced in strain A113, an earlier created, G9 derived *S. bombicola* strain lacking the first glucosyltransferase *ugta1*[175]. Indeed, in *U. maydis* and *P. flocculosa* only one glucosyltransferase is present in the cellobiose lipid gene cluster, suggesting a dual catalytic function for this transferase, both glucosylation reactions of the hydroxy fatty acid and subsequent of the glucolipid intermediate. The characterisation of the both strains is described in IV.3.4.1. An overview of their engineered cluster is given in Figure 4.12.

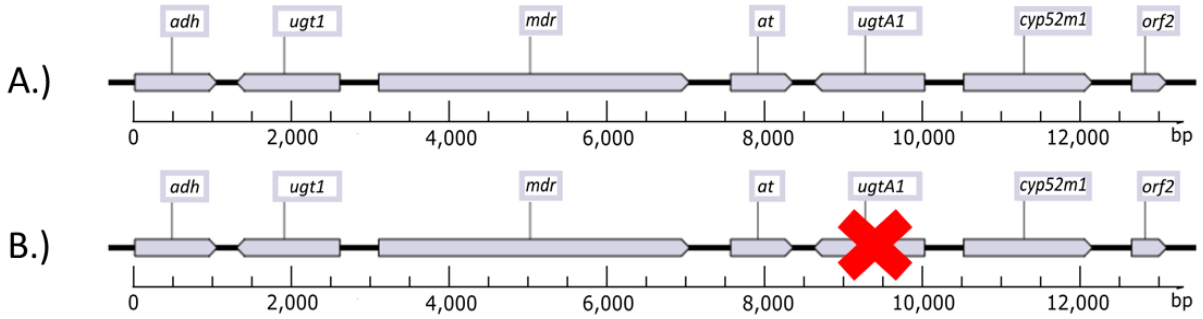


Figure 4.12: Differences between the original cellobiose strain (A.) and the novel strain wherein the *ugta1* coding sequence was deleted (B.). Both strains are derived from the G9 *ura3* auxotrophic *S. bombicola*.

In a next step it was envisaged to increase production. In IV.3.3 it has been shown that it is possible to enhance glucolipid productivity by using a different strain background (PT36 strain obtained by targeted engineering compared to *S. bombicola* G9, a spontaneous *ura3* auxotrophic strain). The improvement was obtained combining several new knock-in and -out cassettes in the defined genetic background. For the production of cellobiose lipids, a similar strategy was followed: a new designed cassette for the *ugtb1* locus was constructed and introduced in *S. bombicola* PT36, thereby resulting in a strain similar to the original cellobiose lipid strain described by Saerens *et al.* (Figure 4. D.).

IV.3.4.1. Characterisation of G9 cellobiose lipids strains

In a first growth trial, the G9 background strains were incubated in Lang medium with rapeseed oil as a hydrophobic substrate. After seven days, the cultures were analysed for their cellobiose lipid production. In Figure 4.13, chromatograms of *S. bombicola Δugtb1::ugt1* (Figure 4.12 A.) and *S. bombicola Δugtb1::ugt1 Δugta1* (Figure 4.12 B.) can be seen. A clear difference is visible between the original cellobiose lipid strain and the strain lacking both *ugtA1* and *ugtB1, i.e.* the strain only expressing the *U. maydis ugt1*. *S. bombicola Δugtb1::ugt1* clearly produces more glycolipids than *S. bombicola Δugtb1*::*ugt1* Δ*ugta1*.

In the chromatogram of the original *S. bombicola Δugtb1*::*ugt1*, the signals at 21.5 and 22.9 correspond to non-acetylated glucolipids with respectively oleic and stearic fatty acid tails. Acetylated variants are also detected after 25 minutes. Interestingly, also molecules with two glucose units are present, which elute at 23.9 and 24.4 minutes. Possibly these are di-acetylated cellobiose lipid molecules. When analysing the products of *S. bombicola Δugtb1*::*ugt1* Δ*ugta1,* it becomes clear that the strain produces less glycolipids in general. The signal at 22.3 minutes corresponds to a molecule with a mass of 724. Though no glycolipids are known with this molecular weight, analyses of the mass spectrum of this peak reveals a molecule with a molecular weight of 664 Dalton. This corresponds to a mono-acetylated disaccharide, presumably cellobiose, with an oleic acid tail. It is known that the acetate in the eluents can form adducts with glycolipids. Another possibility is the modification of the fatty acid tail. Since the mass of the unknown compound only differs by 16 Dalton from the known di-acetylated C18:0 cellobiose lipid, two potential explanations can be found.

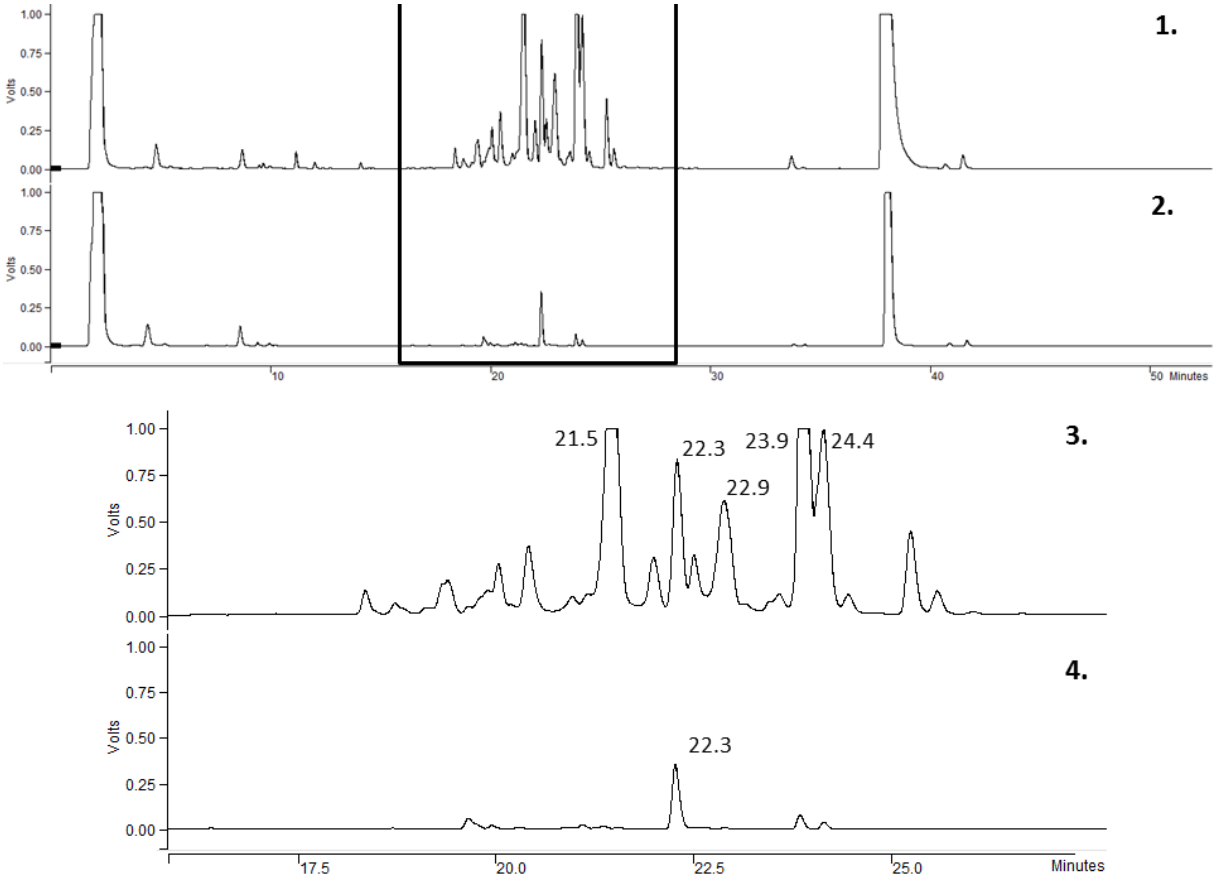


Figure 4.13: Comparison of *S. bombicola Δugtb1::ugt1* (1 and 3) and *S. bombicola Δugtb1::ugt1 Δugta1* (2 and 4). Cross-analysis of the samples on LC-MS resulted in identification of the different signals. Non-acetylated glucolipids with an oleic acid tail elute at 21.5 while acetylated ones elute at 24.4. Glucolipids with a stearic acid tail elute at 22.9. The signal at 22.3 minutes has a molecular mass of 724.

A first one might be the activity of an enzyme such as a P450 on the stearic acid tail adding a second oxygen atom somewhere in the fatty acid tail. Several other P450s are present in the genome of *S. bombicola,* but only few have been characterised[254] , hence such a reaction could be facilitated (see III.4). Another explanation could the activity of an enzyme such as a hydratase on the double bond of oleic acid. Several enzymes have been described that effectively hydrate the double bond[441,442]. This way, a product with a molecular mass of 724 is also reached. The higher hydrophilicity of the molecule by the addition of an extra oxygen also explains the earlier elution of the compound. Still, further structural analysis of the molecules needs to be done before any further conclusions on its structure can be made.

The difference in the amount of glycolipids being produced can be attributed to several bottlenecks. It is possible that the glucolipids produced by UGTA1 in the original cellobiose lipid strain are a substrate for UGT1. Another possibility might be lower production of hydroxy fatty acids in the *ugta1* knock-out strain. Since the *ugta1* coding sequence is positioned upstream of the bidirectional promotor of both the *ugta1* and *cyp52m1*, removal or alterations in the sequence might have an impact on *cyp52m1* gene expression. Yet another possibility is that the masses detected in the original cellobiose lipid strain are not all cellobiose lipids, but the earlier described bola glucolipids. To test the latter hypothesis, alkaline hydrolysis of the glycolipid mixture was performed. Hereby, the glucose molecules attached to the carboxylic group will be removed, resulting in a disappearance or reduction of the presumably produced cellobiose lipids. If all the molecules however are cellobiose lipids, only deacetylation will occur.

In Figure 4.14 analysis by LC-MS of the glycolipids can be seen after hydrolysis for the *S. bombicola Δugtb1*::*ugt1*. Clearly visible is the cellobiose lipid signal. Further analysis on UPLC-ELSD for both alkaline hydrolysis proves that the amount of cellobiose lipid molecules is much lower than previously reported. It is estimated that the total amount of cellobiose lipids in the sample is only 14,31% of the total glycolipid pool. For the strain only carrying the *ugt1* gene, deacetylated molecules can be seen with a mass of 622 Dalton (621 m/z), corresponding to a disaccharide with an oleic acid tail.

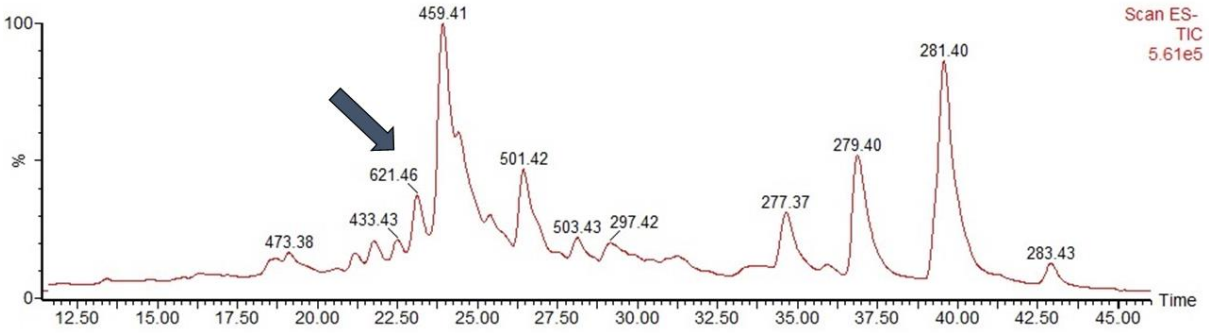


Figure 4.14: LC-MS analysis of the glycolipid mixture produced by the original cellobiose lipid strain after alkaline hydrolysis. The majority of the molecules correspond to glucolipids with either an oleic (m/z 459 and 501) or stearic (m/z 503) fatty acid tail. The molecule with m/z 621 corresponds to a disaccharide coupled to an oleic acid.

Though the reason for the low production of cellobiose lipids is unknown, several possibilities exist. First of all, expression of the transferase might be impaired due to problems arising during transcription or translation. Secondly, activity of the enzyme itself might be impaired due to incompatible or even absent post-translational modifications. Thirdly, the enzyme might be not that active compared to the *S. bombicola* UGTA1 and UGTB1 to start with.

A second growth trial was performed comparing the original G9 *S. bombicola Δugtb1*::*ugt1* strain and a PT36 derived *Δugtb1*::*ugt1*. As can be seen in Figure 4.15, in the original G9 strain a small remainder of the *ugtb1* coding sequence is still present as part of the homologous regions. In the new cassette transformed in PT36 (see Figure 4.1 D), the entire coding sequence is replaced by the one from *ugt1*. However, after the growth trial only the G9 derived strain showed production of glycolipids.

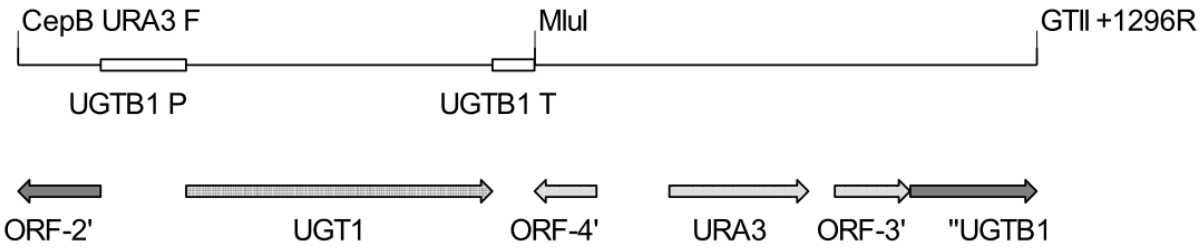


Figure 4.15: Expression cassette developed by Saerens *et al*.[381] for the production of cellobiose lipids *in S. bombicola* G9. ORF-2' is part of the sophorolipid transporter. UGTB1 P and T are the original promotor and terminator of the *ugtb1* gene.

IV.3.4.2. Overexpression of *ugt1* at *ura3* locus

To counter potential low production due to expression problems caused by altering or downregulation of the sophorolipid gene cluster, a *ugt1* overexpression strain with *ugt1* integrated at the *ura3* locus was created in a *ugta1* negative strain of *S. bombicola* with PT36 background. The reason for the *ura3* locus is to evade potential negative influence of or on the *ugt1* coding sequence if it would be integrated in the sophorolipid gene cluster. The choice for a *ugta1* negative strain is because UGTA1 still produced high amounts of glucolipids, a molecule that might not at all be a substrate for UGT1. This way, all the hydroxy fatty acids produced by *S. bombicola* will be available for UGT1 to potentially convert them to cellobiose lipids. The choice for PT36 background was made since it became clear that strains developed from the spontaneous *ura3* auxotrophic G9 are heavily impaired in their glycolipid production. To ensure high product uniformity, oleic acid was used as a substrate instead of rapeseed oil.

As can be seen in Figure 4.16, a mixture of different cellobiose lipid molecules is being produced. The most prevalent molecules are variants carrying a C18:1 fatty acid tail. The differing masses correspond to varying acetylation patterns such as non-acetylated (m/z 621), mono-acetylated (m/z 663) and di-acetylated (m/z 705). Interestingly, no lactonic molecules with m/z values such as 645 and 687 for respectively mono- and di-acetylated lactonic molecules can be detected for oleic acid derived molecules. This suggests that the lactone esterase, which is responsible for sophorolipid ring-closure, is not capable of performing the same reaction on the slightly different cellobiose head group. A similar observation was made when looking at the glucolipids produced in the different strains described in IV.3.3. Though the exact mechanism of lactonisation is unknown, the presence of a sophorose group and a sufficiently long fatty acid are apparently both necessary.

Two other signals are present between the cellobiose lipids (21 to 26 minutes) and the fatty acids linoleic and oleic acid (36 and 41 minutes). These signals correspond to masses belonging to hydroxylated variants of oleic (m/297) and stearic acid (m/z 299). The presence of such molecules is likely due to CYP52M1 producing more hydroxy fatty acids than UGT1 can convert in cellobiose lipids or that can be degraded by the cell own fatty acid metabolism.

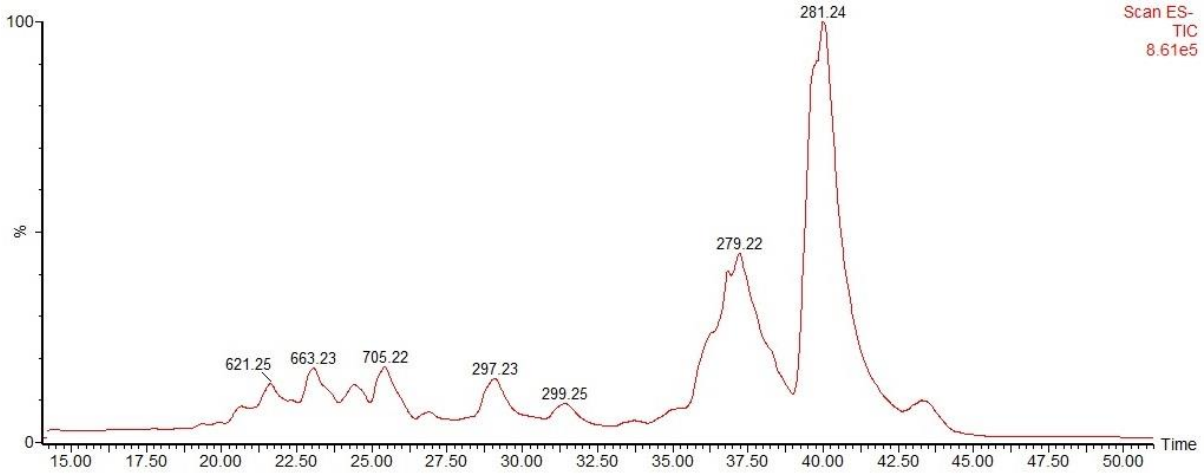


Figure 4.16: Detailed view of the cellobiose lipids produced on oleic acid by the *ugt1* overexpression strain. Non-acetylated cellobiose lipid = m/z 621; mono-acetylated cellobiose lipid = m/z 663 and di-acetylated cellobiose lipid = m/z 705. The substrate oleic acid can be seen with an m/z value of 281. Cell own linoleic acid = m/z 279. Hydroxy fatty acids derived from oleic and stearic acid have m/z values of respectively 297 and 299.

The overall production of the strains remained below 100 mg/L, a level far lower than the titers achieved with *U. maydis*[21] and the reported cellobiose lipid strain already available at the lab[318]. The relatively low levels of production can be attributed to several reasons. First of all, as described above, activity of the enzyme itself might be impaired due to incompatible or even absent post-translational modifications or the enzyme might be not that active compared to the *S. bombicola* UGTA1 and UGTB1 to start with. Second, it might be that the transferase is not that active towards longer hydroxy fatty acid chains. In *U. maydis*, the only fatty acid being incorporated is palmitic acid. Though oleic acid is shorter than stearic acid, it still might be too long for efficient incorporation. Also in *Cryptococcus humicola*, only a small portion of the cellobiose lipids have a C18 fatty acid tail[443].

Another problem arises from the hydroxylation profile of CYP52M1. Most of the products formed are ω-1 hydroxy fatty acids. In *U. maydis*, the same position is hydroxylated as well, but the glucosylation reaction always occurs on the terminal group. Knocking out CYP1, hydroxylating the terminal position, in *U. maydis* results in abolishment of the cellobiose lipid production while knocking out CYP2, hydroxylating the ω-1 position, only has an effect on the ω-1 hydroxylation[327]. This suggests that UGT1 has very low or even no activity towards ω-1 hydroxy fatty acids as a substrate. It has to be noted that chirality also plays a role since in *S. bombicola*, only S-configured ω-1 hydroxy fatty acids are formed by CYP52M1. CYP2 reaction products have never been subjected to detailed analysis on the absolute configuration of the hydroxy group incorporated. Hence it is unknown if UGT1 can accept either R- or S-hydroxy fatty acids. So, it might be that the ω-1 hydroxy palmitic acids produced in *U. maydis* by this enzyme are unsuitable as a glucose acceptor for UGT1 due to the stereochemistry of the hydroxy group.

A second production trial was performed using different substrates. This time oleic acid, ethyl palmitate and juniperic acid (16-OH palmitic acid) were fed to the *ugt1* overexpression strain. No palmitic acid was used since this substrate tends to be less available for the yeast to use as a substrate since it is not liquid at 30 °C. In Figure 4.17, a detailed LC-MS chromatogram of the final ethyl acetate extractions can be seen. Using the different substrates, no clear difference can be seen regarding the mixture of produced cellobiose lipid molecules. All cellobiose lipids have either a stearic, oleic or palmitic acid tail, most likely due to *de novo* synthesis of C18 fatty acids. No molecules with masses for the lactonized variants are detected, again pointing towards the cellobiose lipids being an unsuitable substrate for the lactone esterase. Analysis of the total extracted cellobiose lipid content of the broth revealed that ethyl palmitate and juniperic acid both raised the production of cellobiose lipids at least 2-fold compared to oleic acid. This is most likely due to the higher preference of UGT1 towards hydroxylated palmitic acid. Furthermore, the products and titers don't differ when adding juniperic acid or ethyl palmitate. The low activity of UGT1 might result in CYP52M1 being able to provide enough hydroxylated palmitic acid from ethyl palmitate to mask the bypassing of this step when feeding juniperic acid. As can be concluded, the use of a C16 substrate results in more efficient production of cellobiose lipids.

An interesting observation is again the molecule with an m/z value of 723 observed in the production trial with oleic acid. Though also present in the molecules produced on the other two substrates, it is masked by the di-acetylated C16 cellobiose lipids (m/z 679). If the signal corresponds to a C18 cellobiose lipid with a modified fatty acid tail, other acetylation variants should be present as well. A non-acetylated variant would have an m/z value of 639. Such a value is indeed visible at 20.5 minutes. The mono-acetylated variant is not visible in the total

mass scan, but when performing a single ion scan, a signal can be detected (Figure 4.18). However, further molecular characterisation such as NMR is needed before final conclusions can be made on the new detected masses.

Table 4.5: Overview of the different cellobiose lipids produced by the *S. bombicola ugt1* overexpression strain. In Appendix F, an overview is given of several types of glycolipids, fatty acids and derivatives and their corresponding mass.

Fatty acid tail	Non-acetylated (m/z)	Mono-acetylated (m/z)	Di-acetylated (m/z)
Palmitic acid (C16:0)	595	637	679
Stearic acid (C18:0)	621	663	705
Oleic acid (C18:1)	623	665	707

A direct comparison between the different cellobiose lipid strains tested in the production trails is biased due to the different genetic backgrounds of the auxotrophic strains used. Still, overexpression of UGT1 in a *Δugta1 S. bombicola* strain is an interesting strategy for the production of cellobiose lipids. Total production remained below 100 mg/L, but the product uniformity is clearly better than the G9 derived *Δugtb1*::*ugt1*. The presence of hydroxy fatty acids in the supernatant points towards sufficient supply of substrate for the glucosyltransferase. Still, as already mentioned, several uncertainties remain about the reasons for the low production levels. Therefore, a characterisation of the transferase is attempted.

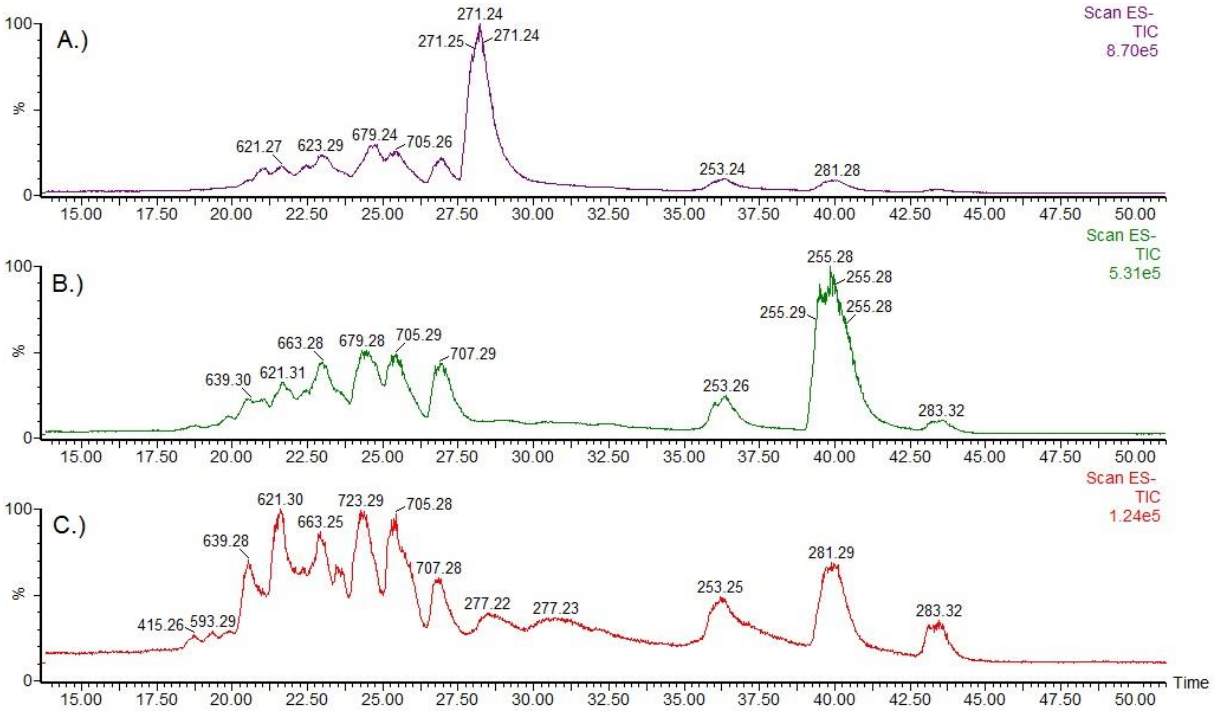


Figure 4.17: Detailed view of the molecules produced on A.) juniperic acid (m/z 271); B.) ethyl palmitate (m/z 255 for palmitic acid) and C.) oleic acid (m/z 281) by the *S. bombicola* PT36 *Δugta1 ugt1* overexpression strain.

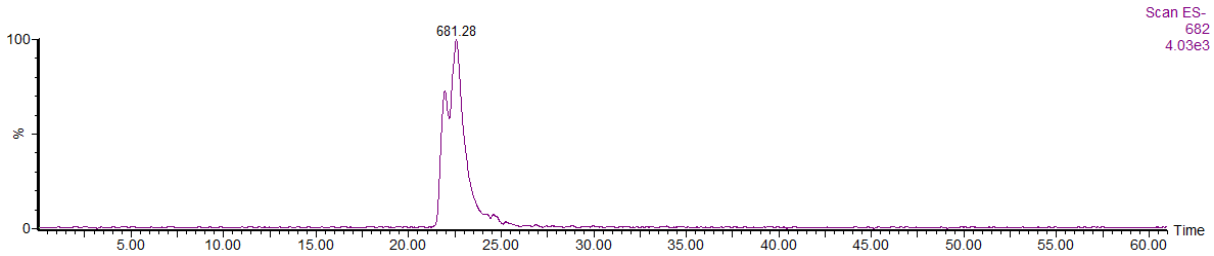


Figure 4.18: Single ion scan of the potential mono-acetylated C18 cellobiose lipid with an extra oxygen atom incorporated.

IV.3.4.1. *UGT1* enzyme characterization

In order to fully understand the glucosylation capacity of UGT1, an *in vitro* enzyme characterization was performed. Initially, his-tag purification of the UGT1 enzyme was attempted to eliminate any background activity from other transferases present in *E. coli* BL21. However, after several attempts it became clear that his-tag purification is not feasible for this enzyme: SDS-PAGE analysis provided proof that the enzyme is not retained on the columns. Though no models are existing for this enzyme, it might be that the C-terminal 6xhis-tag is not freely available, thus hindering interaction of the tag with the Ni-NTA resin. Further tests were done using cell lysate from the *ugt1* expressing *E. coli* strain as well as a BL21 strain carrying an empty expression plasmid as a control for background activity. Juniperic acid (16-OH palmitic acid) and a mixture of C18 ω-1 hydroxy fatty acids (mostly C18:1), derived from sophorolipids by Bio Base Europe Pilot Plant, were incubated for three hours at 30°C.

As can be seen in Figure 4.19, there was activity on the terminal hydroxylated palmitic acid (m/z 595). No activity could be detected on the longer C18 variants (Appendix E Figure A.1). The latter can be due to the fairly low amounts of ω-hydroxy fatty acids in the substrate combined with lower activity towards longer chain lengths. As discussed above (IV.3.4.2), it seems that UGT1 is not active towards subterminal hydroxy groups, though R/S chirality incompatibility cannot be excluded. Interestingly, glucolipids were also produced with juniperic acid as a substrate (m/z 433). This indicates that UGT1 can perform the sequential addition of glucose molecules, but that some loss of the glucolipid intermediate occurs. The reason for this is unknown, but potentially this might be due to the enzyme being expressed in *E. coli*, resulting in different (suboptimal) conditions compared to expression in *S. bombicola* or *U. maydis*. These might for example result in a less stable binding of the glucolipid intermediate.

As it became apparent that the enzyme was active, characterisation of the *S. bombicola* overexpression strain was performed as well. When the cells were in the stationary phase for three days, they were harvested and lysed. The cell lysate was again incubated with hydroxylated C16 and C18 fatty acids as well as 12-OH lauric acid and primary alcohols ranging from 10 to 20 carbon atoms in length. This was done to see if similar activity could be detected for these molecules as for the fatty acids. That way they could be used as a proxy for activity towards other hydroxy fatty acids that are not commercially available like 14-OH myristic acid. The substrates were incubated for 24 hours and analysed on LC-MS. In Figure 4.20, an ethyl acetate extraction can be seen with juniperic acid as a substrate. Though the strain used was overexpressing the *ugt1* gene product, only marginal activity was detected Non-acetylated cellobiose lipids were detected in low amounts (m/z 595).

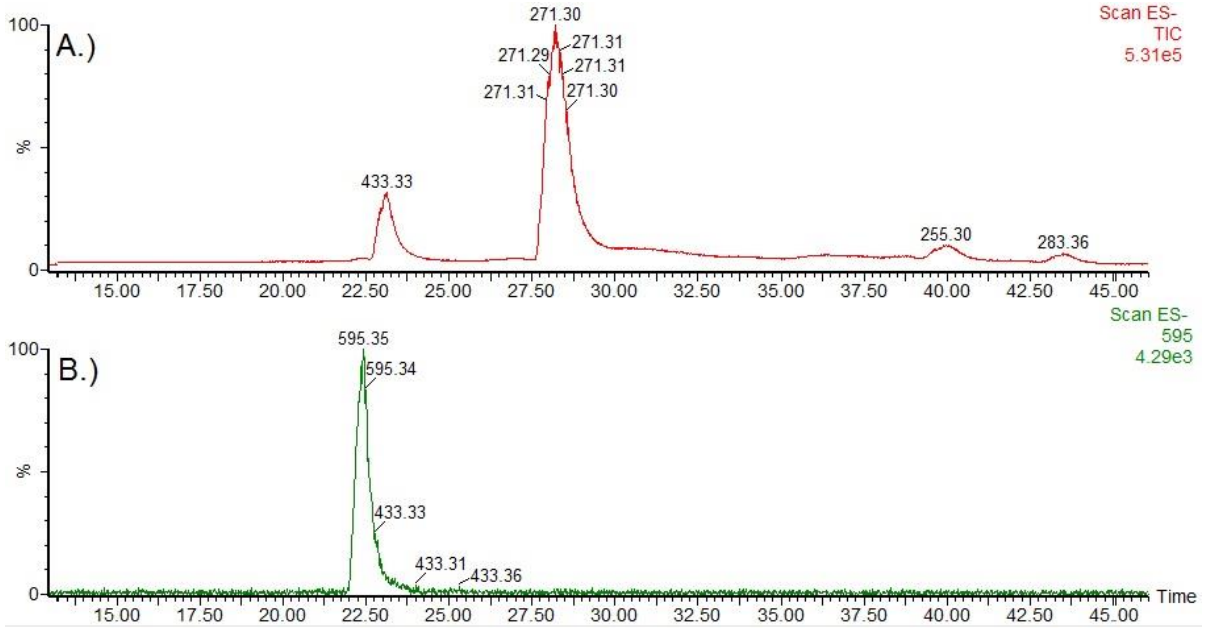


Figure 4.19: Detailed LC-MS chromatogram from enzyme tests on the cell lysate from *E. coli* BL21 expressing the *ugt1* gene. The chromatograms showed are the tests with juniperic acid as a substrate (m/z = 271) with A.) total ion scan with glucolipids visible at 23 minutes. The m/z value of 433 corresponds to non-acetylated C16 glucolipids and B.) mass scan for m/z 595 corresponding with a palmitic acid coupled to a presumed cellobiose head group.

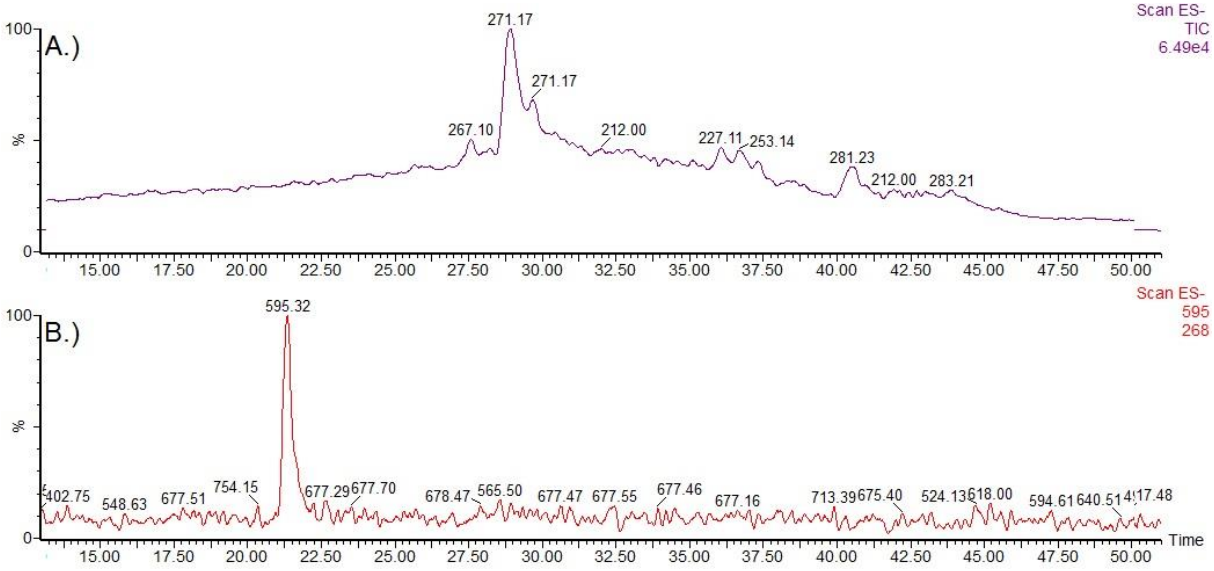


Figure 4.20: Detailed LC-MS chromatogram of enzyme tests on raw cell lysate of the mutant *S. bombicola* overexpressing the UGT1 enzyme. A.) Total ion scan showing the substrate juniperic acid (m/z = 271) and; B.) Mass scan for m/z 595, a non-acetylated cellobiose lipid.

IV.4. Conclusion

Production of glycolipids with a carbohydrate head group different from sophorose is possible when several precautions are taken into account. Firstly, it is necessary to start from strains thoroughly tested for their fitness. Mutant strains that are severely impaired in growth or other biological functions compared to the wild type *S. bombicola* will result in bad producing strains, as illustrated by the glucolipids production with the G9 background strains. Secondly, it is of great importance to be able to differentiate the newly produced molecules from naturally produced ones and to eliminate substrate loss in side reactions performed by native enzymes, while simultaneously improving product uniformity. It is therefore necessary to specifically engineer *S. bombicola* in this way that unwanted side reactions are eliminated. This was clearly essential in the case of the cellobiose lipid production, where the new molecules co-occurred with other glycolipids if the native UGTA1 was still active.

In the case of the glucolipids, several new strains were created that are capable of producing significantly higher amounts compared to the original strain. Concentrations above 25 g/L in Erlenmeyer flasks and up to 130 g/L in a bioreactor were easily achieved. Productivity of the A² strain was on par with other engineered strains of *S. bombicola*[174]. Still, the molecules produced remain a mixture. Different fatty acid tails and acetylation patterns are observed. A knock-out of the acetyltransferase or an overexpression of it can already help to achieve a more uniform product[177]. Purification of the molecules was achieved using solvents during the initial tests, but solvent-free methods could be developed as well. For sophorolipids, purification of acidic and lactonic molecules has been described using neutral polymeric sorbents[444]. Using a similar approach for the glucolipids, a more uniform pure product could be achieved in a more cost-effective way.

For the cellobiose lipids, the low production titers are probably not solely caused by an excessive burden of previous manipulations of *S. bombicola* or by substrate loss in side reactions. Major causes of the lower productivity might be related to suboptimal transport of the cellobiose lipids out of the cell, to non-optimal/non-fine-tuned activity of the UGT1 enzyme in the constructed strains or to problems with the localisation of UGT1 (since UGTA1 is localised near the CYP52M1 enzyme). The *in vitro* experiments indicate towards a main contributor to the low productivity being the UGT1 enzyme itself. The results of the enzyme tests performed with *E. coli* cell lysates support the hypothesis that UGT1 is solely responsible for the glucosylation of hydroxy fatty acids to cellobiose lipids and that these glucosylations are performed sequentially. Indeed, strong indications were present due to it being the only transferase in the *U. maydis* cellobiose lipid gene cluster[327], but no characterisation has been done before.

Besides optimisation of the resulting production strain, for example through optimising the expression and activity of the UGT1 enzyme, and besides fermentation optimisation, detailed structural analysis of the produced molecules is essential. Ideally nuclear magnetic resonance should be performed, but given the tendency of molecules being produced as a mixture instead of a uniform product this is not always feasible. Still, other possibilities exist to investigate the molecular structure even without state-of-the-art methods. For example, hydrolysis of the molecules, either alkaline or acidic, will show the composition for the hydrophobic part as well as the location of the carbohydrates on the fatty acid, either on a hydroxy group or the carboxyl group.

Chapter V

Production of rhamnolipids in S. bombicola

Chapter V – Production of rhamnolipids in S. bombicola

V.1. Introduction

Glycolipid biosurfactants are produced by several microorganisms like *Pseudomonas aeruginosa*, *Starmerella bombicola*, *Ustilago maydis* and many others. They find applications in medicine, pharmacy and nanoindustry. Rhamnolipids, a well-studied glycolipid, is produced by several species from the *Pseudomonas* and *Burkholderia* genera. The top producer is *P. aeruginosa* PAO1[19], an opportunistic pathogen, with titers up to 120 g/L. Other species are less productive. Rhamnolipid synthesis required precursor input from two pathways. Firstly, activated dTDP-L-rhamnose is generated from glucose-1-phosphate. Thymidylation and subsequent (de)hydrogenase and epimerase activity results in the rhamnose precursor for the hydrophilic moiety. Secondly, the hydrophobic part originates from the bacterial fatty acid synthesis. Since this synthesis uses a dissociated enzyme system[445], the intermediates produced are easily accessible for other enzymes to divert them towards production molecules different from fatty acids. One group of these intermediates are the β-hydroxy fatty acids. The first rhamnosyltransferase RhlA is able to couple two of these molecules into a dimer called a 3-(3-hydroxyalkanoyloxy) alkanoic acids (HAA). In *P. aeruginosa*, this enzyme has a high preference towards chain lengths varying from 8 to 12 carbon atoms with the highest activity for 10 carbon atoms. In *Burkholdera* species, longer congeners with 14 carbon atoms are produced. Though a dimer is the most common variant, monomers and trimers occur as well[44]. Finally, transferases RhlB and RhlC will utilize dTDP-L-rhamnose as a donor to rhamnosylate the HAA group.

Sophorolipids, another well-studied kind of glycolipid, is produced by the yeast *S. bombicola*[446]. It is a secondary metabolite composed out of a hydroxy fatty acid and two sugar moieties. The glucosylation of the hydroxy fatty acid is achieved by a glucosyltransferase named UGTA1[175] while the second transferase UGTB1 couples a second glucose to the first one by a β-1,2 glycosidic bond. High production titers have been achieved in optimized production setups[370] with productivities up to 3.5 g/L.h[422]. To achieve these high levels, an efficient flux of metabolites has to exist inside the cell. UDP-glucose is generated from glucose-1-phosphate and UTP. Two sources of glucose-1-phosphate are present in the central metabolism of many organisms. A first one is gluconeogenesis. In this pathway, glucose-1-phosphate is generated in a mechanism similar to reversed glycolysis. It is proven in *C. apicola* that some percentage of glucose-1-phosphate is indeed created this way[447]. A second route is catalysed by phosphoglucomutase (PGM). It interconverts glucose-6-phosphate to glucose-1-phosphate[448]. Though it is not known for *S. bombicola*, it has been proven that the route via PGM to UDP-glucose can be engineered for higher fluxes through the Leloir pathway[449,450].

Harnessing the capability of *S. bombicola* to glucosylate fatty acids in an efficient manner opens up possibilities for production of other kinds of glycolipids. As was demonstrated in Chapter IV, changes to the carbohydrate head group are possible. During this research, heterologous production of rhamnolipids was attempted. The precursor pathways for both activated rhamnose and HAA molecules were introduced and tested for their activity.

V.2. Material and Methods

V.2.1. Strains and culture conditions

General cultivation conditions for *S. bombicola* and *E. coli* can be found in II.2.1. For the quercetin glycosylation experiments with *S. bombicola*, 1 g/L of quercetin dissolved in DMSO was added to the cultures. *Pseudomonas aeruginosa* PAO1 LMG24986 was obtained from the Belgian Co-Ordinated Collections of Micro-organisms and cultured in the same medium as E. coli Dh5α.

V.2.2. Molecular Techniques

V.2.2.1. General techniques

General molecular techniques are identical as described in II.2.2.

V.2.2.2. Plasmid development

Creation of a UDP-rhamnose strain

The plasmid developed in II.2.3.2 was used for efficient recombination at the *cyp52m1* locus. The vector backbone was amplified with primers P52_REV_upCYP and P69_FOR_RegulUP (Table 5.2) (6539bp). The coding sequence of Rhm2/Mum4 (GenBank: AEE32949.1) was codon optimized by backtranslation from the protein sequence and ordered as a gBlock at Integrated DNA technologies (Leuven) (Appendix D). The gBlock was used as a template for amplification with primers P1052_Rhm2GibCypFor and P1781_Rhm2GibCypRev (2073bp). Assembly was performed with Gibson assembly and correct plasmids are screened with primers P945_Rhm2SeqFor and P79_REV_tCYP_extBamHI (1870bp). After sequencing, the expression cassette is amplified with primers P847_Cm2ExCasLongFor and P887_Cm2ExCasRev (5516bp) from BS_EC_0285_Cyplocus_Rhm2 and transformed in *S. bombicola* PT36. The mutants are finally screened with primers P946_Cm2checkRev and P79_REV_tCYPextBamHI for the 5'-region and P38_For_qPCRURA3 combined with P952_A21TotRev for the 3'-region. All the necessary primers can be found in Table 5.2.

Development of a dTDP-L-rhamnose strain

Creation of the dTDP-L-rhamnose strain involved the introduction of three genes at the three final loci of the sophorolipid biosynthetic cluster. The original bacterial pathway was shortened by one enzyme by using the *Arabidopsis thaliana* AtNRS/ER to perform the final two steps. The coding sequence was ordered as a gBlock (GenBank: AEE34035.1) at IDT (Appendix D) and subsequently amplified and assembled into the final expression vector. The vector backbone used was amplified from the earlier developed *cyp52m1* locus plasmid developed in II.2.3.2. All other sequences were amplified from genomic material from *P. aeruginosa* PAO1 and *S. bombicola* (Table 5.2). If necessary, intermediate fragments were generated by fusion PCR with Primestar and subsequently stored in a pJET vector. The final assembly (BS_EC_0102_pGEMTRhamnose 1) was performed by Gibson assembly and is shown in Figure 5.1.

The final expression cassette was created by primers P508_MDRendFor and P887_CM2ExCasREV (8117bp). Mutants were screened with P264_MDRseq4 and P940_ugtA1_start (2886bp) and P953_URA3outendFor combined with P952_A21totRev (1172bp).

Table 5.1: Sequences needed for the construction of a dTDP-L-rhamnose *S. bombicola* strain.

Sequence name	Length (base pairs)	Homology 5'-region	Homology 3'-region
5' homologues region	951	pJET vector backbone	*Atnrs/er*
Atnrs/er	906	*Atnrs/er*	Terminator UGTA1
Terminator UGTA1	262	Terminator UGTA1	*rmlB*
rmlB	1059	Terminator UGTA1	Promotor UGTA1
Promotor UGTA1	550	*rmlB*	*rmlA*
rmlA	882	Promotor UGTA1	3' homologues region
3' homologues region	3499	*rmlA*	pJET vector backbone

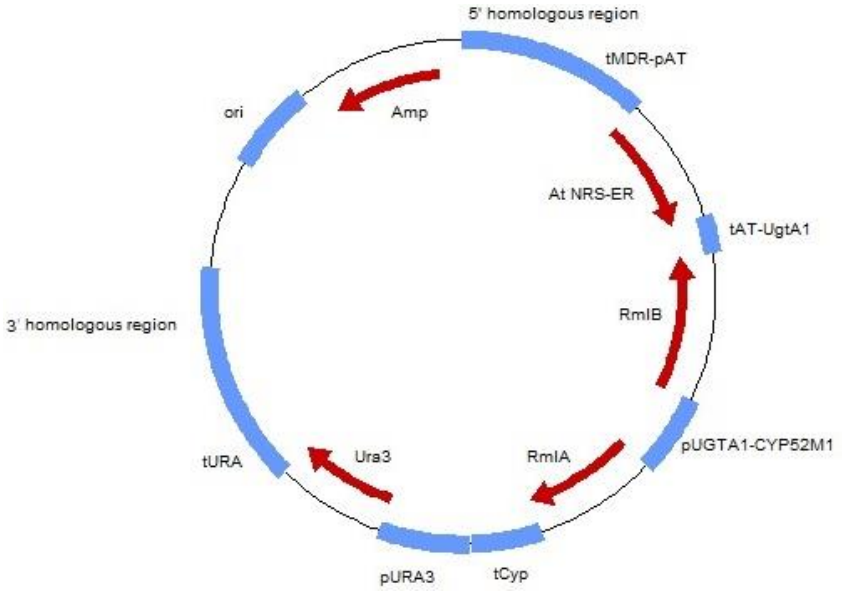


Figure 5.1: Visual representation of the plasmid carrying the expression cassette for the production of dTDP-L-rhamnose in *S. bombicola* (BS_EC_0102_pGEMTRhamnose 1).

Recovery of the NDP-rhamnose strains.

The strains that are able of synthesizing NDP-rhamnose all have the *ura3* marker in their genome. To perform additional modifications, this marker needs to be removed. This can be done by using a recovery cassette for transformation and plating out the cells on selective medium containing 5-FOA. The recovery cassette can be amplified from genomic DNA of wild type *S. bombicola* ATTC22214 with primers P69_FOR_RegulUP and P887_Cm2ExCasRev (930bp). After several days, the emerging colonies are screened with P69_FOR_RegulUP and P206_PfCYP1Rev (Table 5.2). Correct mutants produce a PCR product of 1060bp.

NDP-rhamnose validation plasmid

Validation of the abovementioned strains is done by glycosylation of quercetin by the *A. thaliana* AT78D1[451] glycosyltransferase. A previous constructed plasmid for the *ura3* locus (BS_EC_0395_UmUGT1_OE_Uraloc_kol4) was reused for construction of the expression cassette BS_EC_0393_AT78D1_OE_Uraloc kol 2. The coding sequence of the transferase was ordered as a gBlock (GenBank: AEE31240.1) at IDT (Appendix D) and ligated by Gibson

assembly in the linearized vector. The cassette was transformed in strains containing the previously constructed UDP- and dTDP-L-rhamnose pathways as well in a *S. bombicola* strain lacking the *ugta1* gene. Selection was performed on nourseothricin. The gene *at78d1* is controlled by the strong constitutive promotor from phosphoglycerate kinase. Amplification of the expression cassette is performed from this final plasmid. A second validation cassette containing the second rhamnosyltransferase *rhlB* was designed as well. The RhlB enzyme will also be expressed *ura3* locus using nourseothricin as marker (pCARBO425 pGEMT_pGKIRhlB_Nourseo). All the primers used can be found in Table 5.2.

Table 5.2: Primers used for the construction of the different expression cassettes for nucleotide sugars and glycosylated quercetin. Binding and non-binding domains of the primers are separated by ligature.

Primer name	Primer sequence
P1_FOR_URA3v	AGAACAAGGCCGAGTATGTC
P32_REV_cassette	GTCAGATTAGCCTCCGACATAG
P33_FOR_checkcassIN	CCATACTCAAGCGCGAACAC
P35_REV_checkcassIN_DOWN	GAGCTCAAGACGCGTTTACTCAATGC
P38_FOR_qPCRURA3	AGGTGGAGTGTTTAAGA
P52_REV_upCYP	ATATGTACTTTTCAATATGATAAACGGAGAAATAACG
P69_FOR_RegulUP	GTTTCTTAGCCTCCCATGGAAG
P79_REV_tCYP_extBamHI	CAGGATCC-GTCGGTTAAACGCACTCCTTC
P377_MegaHit8For	CTCCCTCCGTTGGCAATCAAACTG
P554_MDRpAT5homGibFOR	TTCCGGATGGCTCGAGTTTTTCAGCAAGAT-TGGTCTGGCCCTGAGTCTGAAG
P555_MDRpAT5homoverREV	TGCGTCTGCAACCAT-TGAATTCTAGAATGTGAGGTGGAATGAGGCAAGG
P556_NRSERoverFOR	ACATTCTAGAATTCA-ATGGTTGCAGACGCAAACG
P557_NRSERoverREV	ATGAATTTAGTTTAC-TCAAGCTTTAACTTCAGTCTTCTTGT
P558_tATroverFor	GAAGTTAAAGCTTGA-GTAAACTAAATTCATGACAGCCTTT
P559_tATgibRev	CGCGAGTGGGTGGGTAAGCAGTACGCATGA-AATCGTACGATCAAATCAGAT
P560_RmlBcasGibFor	TCTTCCCTGATCTGATTTGATCGTACGATT-TCATGCGTACTGCTTACCCACCCACT
P561_RmlBcasoverrev	CTTTCAGGCCATATA-ATGACGATTCTCGTGACCGGCAGCGCCGGCTT
P562_pUGTAoverfor	CACGAGAATCGTCAT-TATATGGCCTGAAAGAGG
P563_pUGTAoverrev	GCCCTTGCGTTTCAT-ATATGTACTTTTCAATATGATAAAC
P564_RmlAcasoverfor	TTGAAAAGTACATAT-ATGAAACGCAAGGGCATCATCC
P565_RmlAcasGibrev	GAACGTTTCTTCCATGGGAGGCTAAGAAAC-TCAGTACACGGTCTCGGTCAG
P566_DOWNCYPFOR	AAGCGCCTGCTGACCGAGACCGTGTACTGAGTTTCTTAGCCTCCCATGGAAG
P567_DOWNCYPREV	AGAATATTGTAGGAGATCTTCTAGAAAGATAAGCGTGAAGCTCCTCTGACAATC
P568_pJETrhamnGibfor	AGTTTTGATTGTCAGAGGAGCTTCACGCTT-ATCTTTCTAGAAGATCTCCTACAAT
P569_pJETrhamnGibrev	CTGCGTTACTTCAGACTCAGGGCCAGACCA-ATCTTGCTGAAAAACTCGAGCCATC
P840_MDRSeq3	TGATTGGCGGCATAGTTACTGG
P841_MDRSeq4	TGGTCTGGCCCTGAGTCTGAAG
P847_Cm2ExCasLongFor	CGTTGTCAAGTCCTAAGGTAT
P887_Cm2ExCasRev	AAGCGTGAAGCTCCTCTGACAATC
P940_ugtA1_start	CACAATGAGCCCTTCATCACAC
P945_Rhm2SeqFor	TGGCACTCATGTGCTCTTGG
P946_Cm2checkRev	CAACAGTACGAACATTTCCGTTAGAG
P952_A21TotRev	GCTCTTGTTCGGTACTCTTATT
P953_ura3outendfor	TAAAGAAACGAAGGGCCCAGCAGTC
P1052_Rhm2GibCypFor	CGTTATTTCTCCGTTTATCATATTGAAAAGTACATAT-ATGGATGATACAACATACAA
P1344_REV_MDR	CGAGAAACGCACTGCCGAGCATAC
P1580_AOX2_sequencing_3	CGCTGAGACTCACTCACTTG
P1596_ACP_check_out_fw_2	TATGCTGGTGCATCCAGTTG
P1600_GAPD_insert_check_rv	GTCCCGATACTGGGATGTTG
P1772_nat1_check_rv	TCATGTAGAGCGCCTGCTCG
P1781_Rhm2GibCypRev	GGGAACGTTTCTTCCATGGGAGGCTAAGAAAC-TTAGGTTCGCTTATTGGGTTCGAAAAC
P1814_tATrev	AATCGTACGATCAAATCAGATC
P2312_tGALFor	ATATACACTAGATGCTATCGATAATATCG
P2391_At78D1GibpGKIFor	CTCCAAACCAGAAAAA-ATGACCAAATTCTCCGAGCCAA
P2392_At78D1GibtTKRev	CCGTGTTTCAGTTAGCCTCCCCCATCTCCCCTA-AACTTTCACAATTTCGTCCAACAA
P2482_pGKIRev	TTTTTCTGGTTTGGAGGACCTT

ACP and PPTase expression cassette at ura3 locus

Construction of the first expression cassette carrying the acyl carrier protein (ACP) and the phosphopantheteine transferase (PPTase) was done using the pJET vector backbone for cloning in *E. coli.* All the different parts of the expression cassette can be found in Table 5.3 in their order of assembly. The pJET vector was chosen as a vector backbone to construct the expression cassette. The sequences of the ACP and PPTase were amplified from isolated genomic DNA from *P. aeruginosa* PAO1. The primer used for amplification of the ACP contained a 6xHis tag. All other sequences were picked up from *S. bombicola* ATCC22244 genomic DNA except for the vector backbone. Primers were ordered at Integrated DNA Technology (IDT, Leuven, Belgium). The final expression plasmid is BS_EC_0130_pJETACPPPTase 1 (Figure 5.2). The expression cassette for integration in the genome of *S. bombicola* PT36 *Δcyp52m1* is created by using primers P49_FOR_URA3upextNheI and P1013_GAPDAT_FusionF (Table 5.4). Screening of the mutants after incubation on SD medium was performed by colony-PCR with primers P1_FOR_URA3v and P690_ACPhisGibpCypFor (2330bp) for the 5' region and P843_A21knockhygroCasFor combined with P35_REV_checkcassIN_DOWN (1897bp) for the 3' region.

Table 5.3: Fragments used for the ACP-PPTase expression cassette

Sequence name	Length (base pairs)	Homology 5'-region	Homology 3'-region
5' homologues region	1849	pJET vector backbone	Terminator AT
Terminator AT	330	5' homologues region	*acp*
acp	229	Terminator AT	Promotor UGTA1
Promotor UGTA1	590	*acp*	*pptase*
pptase	774	Promotor UGTA1	3' homologues region
3' homologues region	1062	*pptase*	pJET vector backbone

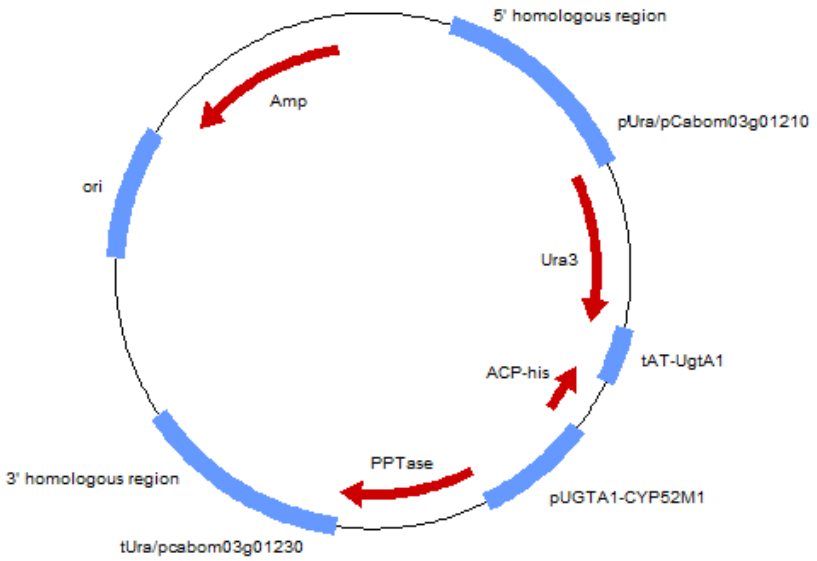


Figure 5.2: Overview of the ACP-PPTase expression cassette (BS_EC_0130_pJETACPPPTase 1).

Table 5.4: Primers used for the ACP-PPTase expression cassette.

Primer name	Primer sequence
P1_FOR_URA3v	AGAACAAGGCCGAGTATGTC
P35_REV_checkcassIN_DOWN	GAGCTCAAGACGCGTTTACTCAATGC
P49_FOR_URA3upextNheI	CGCTAGCCATGCTGAAGACAGCACCGCTGCTATC
P685_5-regionURA3GibpJETFor	CCGCCAGATCTTCCGGATGGCTCGAGTTTTTCAGCAAGAT-AACAAGGCCGAGTATGTC
P686_5-regionURA3GibtAtRev	CTGTCATGAATTTAGTTTAC-TCATCTTGACTGAACTTTTCT
P687_tATGib5-regionFor	GAAAAGTTCAGTCAAGATGA-GTAAACTAAATTCATGACAGCCTTTTC
P688_tATGibACPRev	CATCATCACCATCACCACTAA-AATCGTACGATCAAATCAGATCA
P689_ACPhisGibtAtRev	GATCTGATTTGATCGTACGATT-TTAGTGGTGATGGTGATGATGTTGCTGGTGAGCAACGATGTAG
P690_ACPhisGibpCypFor	CTTGCCTCTTTCAGGCCATATA-ATGAGCACCATCGAAGAACG
P691_pUGTAGibACPFor	CGTTCTTCGATGGTGCTCATTAT-ATGGCCTGAAAGAGGCAAG
P692_pUGTAGibPPTaseRev	CGGTCGTTCATGGCGCGCAT-ATATGTACTTTTCAATATGATAAACGGAGAAAT
P693_PPTaseGibpUGTAFor	GTTTATCATATTGAAAAGTACATAT-ATGCGCGCCATGAACGACCGTCTCCC
P694_PPTaseGibp3-regionRev	GAAGGAACTGTTTGAGAAAA-TCAGGCGCCGACCGCCACCAGGCT
P695_3-regionGibPPTaseFor	CTGGTGGCGGTCGGCGCCTGA-TTTTCTCAAACAGTTCCTTCAATGCAACTT
P696_3-regionGibpJetRev	CATGGCAGCTGAGAATATTGTAGGAGATCTTCTAGAAAGAT-GAAGCAGGCGAGTCGGAGCA
P843_A21knockhygroCasFor	GAGTCGGGCGTTATTTCTCC
P1013_GAPDAT FusionF	GCCTCGTCAACCATCTTATC

ACP, PPTase, FabD and FabY expression cassette at ura3 locus

Introduction of the fabD and fabY genes in plasmid BS_EC_0130_pJETACPPPTase was done by linearising the vector with primers P697_5-regionGibtCabom03g16530Rev and P706_3-regionGibFabYFor (Table 5.6) and ligating the different fragments in the order found in Table 5.5 by Gibson assembly. The final plasmid is BS_EC_0403_ACP_FabDY_final (Figure 5.3).

Table 5.5: Fragments used for the ACP-PPTase-FabD-FabY expression cassette.

Sequence name	Length (base pairs)	Homology 5'-region	Homology 3'-region
tCabom03g16530	202	Vector backbone	FabD
fabD	982	tCabom03g16530	Promotor UGTB1
Promotor UGTB1	527	FabD	FabY
fabY	1970	Promotor UGTB1	Vector backbone

Table 5.6: Primers used for the ACP-PPTase-FabD-FabY expression cassette. Non-binding regions are separated by a dash.

Primer name	Primer sequence
P697_5-regionGibtCabom03g16530Rev	GAAATAAATTATTATAAGCACTGCAGCCTAGCAAAT-TCAGGCGCCGACCGCCACCAGGC
P698_tCabom03g16530Gib5-regionFor	CTGGTGGCGGTCGGCGCCTGA-ATTTGCTAGGCTGCAGTGCTTATAATAATTTATTTC
P699_tCabom03g16530GibFabDRev	CGCGCGCCGCCCTGGCCTGA-TTTGGTTTCGCAGTGTCGAATTG
P700_FabDGibUGTB1For	CCACTCTCTGATTTTGGCTT-ATGTCTGCATCCCTCGCATTCGTCTTCCCTGG
P701_FabDGibtCabom03g16530Rev	CAATTCGACACTGCGAAACCAAA-TCAGGCCAGGGCGGCGCGCGT
P702_pUGTB1GibFabDFor	GAATGCGAGGGATGCAGACAT-AAGCCAAAATCAGAGAGTGG
P703_pUGTB1GibFabYRev	CAATGACCGGTAGTCGAGACATA-TAGGAGACTTTATTGTTTTATGC
P704_FabYGibpUGTBFor	CATAAAACAATAAAGTCTCCTAT-ATGTCTCGACTACCGGTCATTGTTG
P705_FabYGibp3-regionRev	GTATTCATGTGCAAGTTGCATTGAAGGAACTGTTTGAGAAAA-TCAGTCGAGCATGTCGCTGAAA
P706_3-regionGibFabYFor	TTCAGCGACATGCTCGACTGA-TTTTCTCAAACAGTTCCTTCAATGCAACTT
P1057_FabYSeqFor	CTGGTCAAGGCCGAAGGCGGCCGCTA

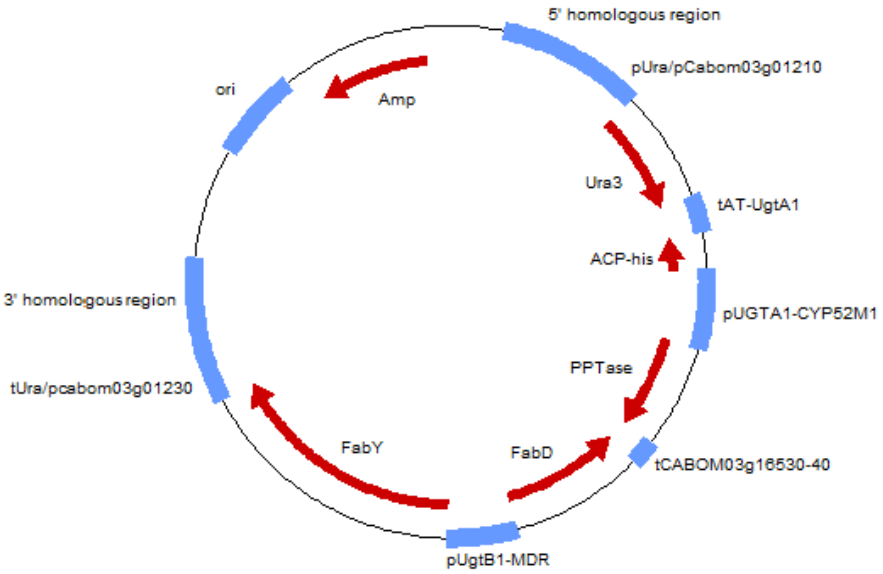


Figure 5.3: Overview of the ACP-PPTase-FabD-FabY expression cassette (BS_EC_0403_ACP_FabDY_final).

FabV, FabZ, FabF1 and FabG expression cassette for the lactone esterase locus

The second expression cassette for the introduction of the bacterial fatty acid synthesis was designed for the lactone esterase locus. A similar strategy was followed as for the previous mentioned expression cassettes for the *ura3* locus. In Table 5.7 the different fragments and their properties are mentioned. The final expression cassette can be amplified with primers P1666_lactcasrev and P297_kolip2flankrev (Table 5.8) from the plasmid BS_EC_0345_FabVZ1 (9043bp). Screening of the mutants is done by PCR with primers P940_ugtA1start, P1664_lactcheckcasfor, P551_ura3endrev and P1665_lactcheckcasrev resulting of fragments of 2321bp and 2408bp for the 5' and 3'-region. Alternatively, a second cassette with hygromycine selection was created as well. This was to encounter potential *ura3* recovery problems. The final plasmid with hygromycine resistance is BS_EC_0404_FabGFZV_hygromycine kol 1.

Table 5.7: Fragments used for the FabF1-FabV-FabG-FabZ expression cassette

Sequence name	Length (base pairs)	Homology 5'-region	Homology 3'-region
5' region	1000	Vector backbone	FabF1
fabF1	1245	5' region	Promotor UGTA1
Promotor UGTA1	550	FabF1	FabV
fabV	1197	Promotor UGTA1	Terminator At
Terminator At	262	FabV	FabG
fFabG	744	Terminator At	Promotor UGTB1
Promotor UGTB1	484	FabG	FabZ
fabZ	441	Promotor UGTB1	Terminator 01g4190
Terminator 01g4190	816	FabZ	URA3-marker
ura3 marker	1412	Terminator 01g4190	3' region
hygromycine marker	1814	Terminator 01g4190	3' region
3'region	1000	URA3-marker	Vector backbone

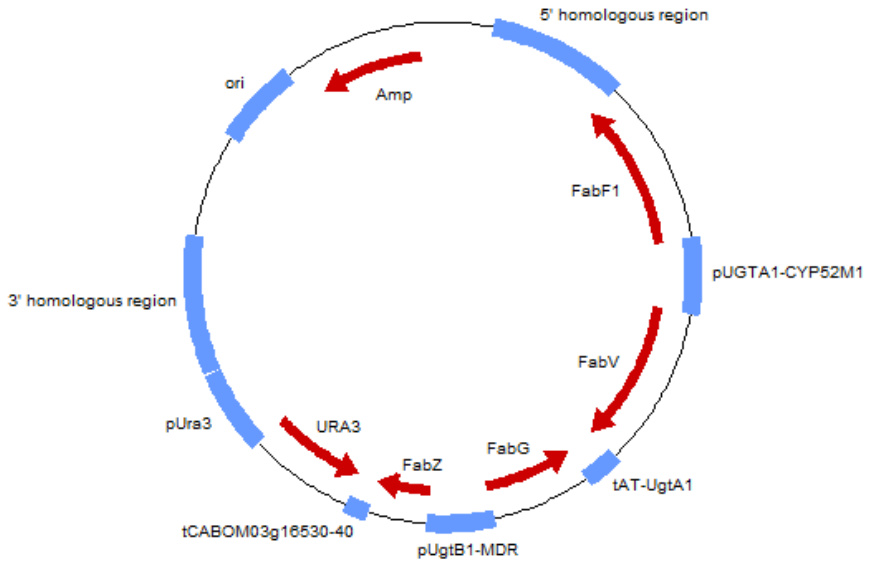


Figure 5.4: Overview of the FabF1-FabV-FabG-FabZ expression cassette with *ura3* selection (BS_EC_0345_FabVZ1)

Table 5.8: Primers used for the FabV-FabZ-FabF1-FabG expression cassette

Primer name	Primer sequence
P297_kolip2flankrev	TACTGCTCTGCCGATCGTTG
P551_ura3endrev	TCATCTTGACTGAACTTTTCTCAGATA
P940_ugtA1start	CACAATGAGCCCTTCATCACAC
P955_lacton5GibpjetFor	CAGATCTTCCGGATGGCTCGAGTTTTTCAGCAAGAT-AACCAAGGGGTGATCTCTCG
P956_Lacton5GibfabF1Rev	CGCAGGTTCGCCGACTGA-TTTGAGACCATTTGAATCCC
P957_fabF1Gib5regionRev	GATTCAAATGGTCTCAAA-TCAGTCGGCGAACCTGCGGA
P958_fabF1GibpCypFor	CCTCTTTCAGGCCATATA-ATGTCGCGTAGACGCGTCGTCA
P959_pCYPGibfabF1For	CGACGCGTCTACGCGACAT-TATATGGCCTGAAAGAGGCA
P960_pCypGibfabVRev	GCGCGGTTTGATGATCAT-ATATGTACTTTTCAATATGATAAACGGAGAA
P961_fabVGibpCYPFor	CATATTGAAAAGTACATAT-ATGATCATCAAACCGCGCGTG
P962_fabVGibtATRev	GTCATGAATTTAGTTTAC-TCAGGCCTGGATCAGGTTGG
P963_tATGibfabVFor	CAACCTGATCCAGGCCTGA-GTAAACTAAATTCATGACAGCCTTT
P964_tATGibfabGRev	GGGATGTACATGAGCTGA-AATCGTACGATCAAATCAGA
P965_fabGGibtAtRev	GATTTGATCGTACGATT-TCAGCTCATGTACATCCC
P966_fabGGibpMDRFor	CACTCTCTGATTTTGGCTT-ATGAGTCTGCAAGGTAAGG
P967_pMDRGibfabGFor	CTTACCTTGCAGACTCAT-AAGCCAAAATCAGAGAGTG
P968_pMDRGibfabZRev	CTCGTTGATGTCCATCAT-ATAGGAGACTTTATTGTTTTATGC
P969_fabZGibpMDRFor	CATAAAACAATAAAGTCTCCTAT-ATGATGGACATCAACGAGATTC
P970_fabZGibtcabomRev	CACTGCAGCCTAGCAAAT-TCATAGTTTGCGTTCCGC
P974_pUraGibLacton3For	CATGTAGTCGCACCCGCCCC-GTTGAACCATGATGGCAGTGTT
P975_lacton3GibUraFor	CTGCCATCATGGTTCAAC-GGGGCGGGTGCGACTACATG
P976_lacton3GibpJetRev	CAGCTGAGAATATTGTAGGAGATCTTCTAGAAAGAT-GATCGCAGTAAGGCCCCTCGT
P1660_FabZGibTermRev	TCATAGTTTGCGTTCCGCACAGATGATTTCAGCC
P1661_Term41904200GibFabZFor	CATCTGTGCGGAACGCAAACTATGA-GCACACACGAAGAGTGCC
P1662_Term41904200GibURA3Rev	CATATCTGAGAAAAGTTCAGTCAAGATGA-GTCTCATTCTATCAAAAGCCAGTACTA
P1664_lactcheckcasfor	ACCAAGGGGTGATCTCTCGAGATGG
P1665_lactcheckcasrev	GCAGTAAGGCCCCTCGTTACCC
P1666_lactcasrev	TGCTCGTCTGAGACGGCTTGGG
P2293_HygroGibUGTB1locFor	ACTAGTACTGGCTTTTGATAGAATGAGAC-TCAGTTAGCCTCCCCCATCTCCCGATCC
P2294_HygroGibUGTB1locRev	CAAGAACTACACTCGGTGTTGAGAGCAGAACCAATGGCAGTGGCTTAC
P2295_LactonasebbFor	GGGGCGGGTGCGACTACATGAACATC
P2296_HygroGibLaclocRev	GAGCGATGTTCATGTAGTCGCACCCGCCCC-CCAATGGCAGTGGCTTAC

HAA production cassette

Expression of the *P. aeruginosa rhlA* was obtained by creating an expression vector designed for the *cyp52m1* locus (see II.2.3.2). The coding sequence of *rhlA* was amplified from genomic material DNA from *P. aeruginosa* PAO1 isolated using the GenElute Bacterial Genomic DNA Kit (SigmaAldrich; USA). The vector was linearized and both fragments were assembled with the In-Fusion® HD Cloning Kit (Takara; Japan). After transformation in *S. bombicola ΔPT36*, the mutants were screened with primers P954_Cm2checkFor and 938_ura3outbeginrev for the 5' region (Table 5.9). The 3' region was screened with P946_Cm2checkRev and P952_A21TotRev. Again both *ura3* and nourseothricin (respectively BS_EC_0403 and BS_EC_0404) selection constructs were designed. The expression cassette was also introduced in in a *S. bombicola* strain carrying the cassettes.

Table 5.9: Primers used for the creation of the *rhlA* expression cassette designed for the *cyp52m1* locus of *S. bombicola*.

Primer name	Primer sequence
P52_REV_upCYP	ATATGTACTTTTCAATATGATAAACGGAGAAATAACG
P69_FOR_RegulUP	GTTTCTTAGCCTCCCATGGAAG
P163_RhlainfFor	ATTGAAAAGTACATAATGCGGCGCGAAAGTCTGTTGG
P164_RhlainfRev	AACGCTAGCTTGGCGTCAGGCGTAGCCGATGGCCATCTC
P938_ura3outbeginrev	ACTGCCATCATGGTTCAACCTCAC
P946_Cm2checkRev	CAACAGTACGAACATTTCCGTTAGAG
P952_A21TotRev	GCTCTTGTTCGGTACTCTTATT
P954_Cm2checkFor	TCATAGCGAGTTTCTTTGCATGTG

V.2.3. Protein extraction from *S. bombicola*

Extraction of proteins from *S. bombicola* strains is described in III.2.5.

V.2.4. Native PAGE and Western blotting

V.2.4.1. Native PAGE

For the native separation of basic and neutral proteins (pI < 7.0), a basic gel can be applied for separation. The composition of both the stacking and running gel can be found in Table 5.10. The gels are both let to polymerize for 1 hour. The total amount of protein to be loaded depends on the application, but may vary from 0.5 to 5 µg. The samples are loaded with a 5 x concentrated sample buffer (50 % glycerol 50 %, 27 % mQ, 22.9 % Tris-HCl pH 6.8, 0.1 % w/v Bromophenol Blue). The gels run at 250 V in a cooled buffer (0.05 M Tris, 0.38 M glycine in distilled water).

Table 5.10: Composition of a basic native polyacrylamide gel.

Components	Stacking Gel	Resolving Gel 15%
Tris-HCl 1.5M pH 8.9 (mL)	-	1.3
Tris-HCl 0.5M pH 6.8 (mL)	1.25	-
30% Acryl-Bis	500 µL	2.63 mL
Distilled water (mL)	3.2	1.17
TEMED (µL)	5	11.5
10% APS (µL)	50	20
Total	5.005	5.1315

V.2.4.2. Native Western Blotting

The proteins are transferred from the native PAGE to a nitrocellulose membrane. The membrane, Watmann-paper, spunch and native gel are put in CAPS buffer (0.1 M CAPS, pH 11) and stacked according to the order seen in Figure 5.5. The transfer is at 100 V for 60 minutes.

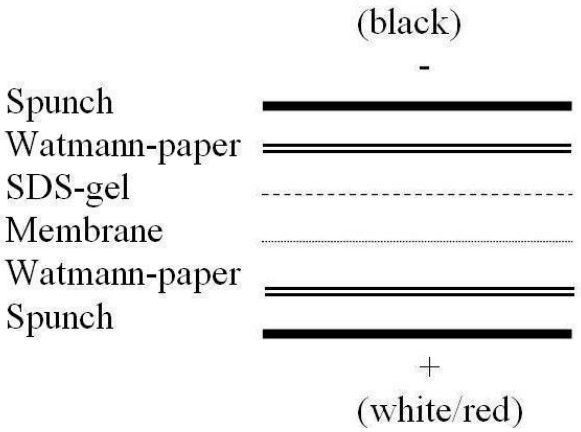


Figure 5.5: Composition and order of the transfer stack for native Western blotting.

After transfer, the membrane is blocked for 1 hour in phosphate buffer solution (PBS) (0.1 M NaCl, 0.083 M Na_2HPO_4 and 0.017 M NaH_2PO_4 at pH 7.2) containing 1% casein and incubated at room temperature on an orbital shaker overnight. Afterwards, the membrane is washed thrice with PBS containing 0.2 % Triton X100. Blotting of the membrane is done by applying the first antibody (anti-His from mouse, Sigma) in a 1/2700 diluted solution in 20ml PBS containing 1 % casein and 0.2 % Triton X100. Incubate for 2h on an orbital shaker. Wash the membrane thrice with PBS containing 0.2 % Triton X100. Blot with the second antibody (anti-mouse from goat with alkaline phosphatase conjugate, Sigma) in a 1/6700 diluted solution in 20ml PBS containing 1 % casein and 0.2 % Triton and incubate for 1 hour on an orbital shaker. Wash the membrane again three times with PBS. Develop the blotted membrane by phosphatase conjugation. Add 50µl NBT/BCIP stock solution (Sigma) to 10ml 10mM Tris pH 9.5 containing 100mM NaCl and 50mM MgCl2. Incubate the membrane during 30min at 37°C in the dark.

V.2.5. Enzyme assays for flavonoid glycosyltransferase activity

5'-Uridinediphospho glucose disodium salt hydrate (UDP-glucose) and 2'-Deoxy-Thymidine diphosphate rhamnose (dTDP-rhamnose) were obtained from Carbosynth. Quercetin and 2'-Deoxy-Thymidine diphosphate glucose (dTDP-glucose) was ordered at SigmaAldrich. Substrate solutions were prepared freshly in 50 mM KH_2PO_4, pH 7.7 except for quercetin, which was dissolved in DMSO. Enzyme assays contained 2 mM of all the substrates used and 200 µl fresh protein solution (III.2.5) in a total volume of 250 µl. For the blank reactions phosphate buffer replaced either the donor, the acceptor or the protein solution. All enzyme reactions were incubated at 30°C for 3, 6 and 24h. Reactions were stopped by addition of 200 µl HCl (2N) and analysed by LC-MS and TLC.

V.2.6. Sampling and analysis

V.2.6.1. Analysis of NDP-sugars and monosaccharides

Nucleotide activated sugars were analyzed on HPLC and LC-MS with a method described by Räbinä *et al*.[452]. The column used was a Grace™ Alltech™ Prevail™ C18 250mm column. The HPLC system was a Varian Prostar HPLC system coupled to a Varian dual wavelength detector. The LC-MS consisted of an Intertek ASG (Manchester, UK) with a Micromass Quattro Ultima LIMS 1107 (Waters). Two types of detection were used. When negative ion mode was applied, the scanning range was set between 550 and 700 m/z. With single ion mode, the masses below (Table 5.11) were used for detection. The same column and LC conditions as for the HPLC analysis were used.

Table 5.11: Masses of the different nucleotide sugars used in single ion mode.

Nucleotide sugar	Mass (Dalton; m/Z + 1)
UDP-Galacturonic acid / Glucuronic acid	580
UDP-D-Galactose / Glucose	566
UDP-N-Acetylgalactosamine / glucosamine	607
UDP-L-Rhamnose	550
GDP-D-Mannose	605
dTDP-L-Rhamnose	548
dTDP-L-Glucose	564
dTDP-4-dehydro-6-deoxy -L-mannose/D-glucose	546

Monosaccharides were analysed on the LC-MS using the X-Bridge Amide column made by Waters. Both single ions (163, 179) as a range (100-350 m/z) were used for detection. Elution of the sugar molecules from the column was done isocratic with 75% acetonitrile and 25% mQ and 0.15% trimethylamine. The flow rate was 1 mL/minute while the column was kept at 35 °C. Standard series of rhamnose were used to determine the concentration in the samples.

V.2.6.2. Analysis of quercetin glycosides

Quercetin glycosides from culture broth or enzymatic reactions were analyzed by TLC on Silica gel 60 F254 (Merck, Germany). Separation of quercetin and the different quercetin derivatives was done by using ethyl acetate:acetic acid:formic acid:water (100:11:11:27) as a mobile phase. Compounds were visualized either by staining with 10 % H_2SO_4 and subsequent charring or by UV-detection at 254 nm.

LC-MS analyses was done using the method described in Pandey *et al*[453].

V.2.6.3. Sampling of glycolipids and HAA molecules

Sampling of glycolipids and HAA molecules was done as described in II.2.4.2. LC-MS analysis was conducted to identify the different congeners possible. In Table 5.12 the different masses of the most likely HAA molecules can be found.

Table 5.12: Table of the different HAA molecules that can be produced. Though shorter and longer chains can be incorporated, only the most likely molecules are searched for.

HAA molecule	Molecular weight
C_8-C_8	302
C_8-C_{10} / C_{10}-C_8	330
C_8-C_{12} / C_{12}-C_8	356
C_{10}-C_{10}	358
C_{10}-C_{12} / C_{12}-C_{10}	386
C_{12}-C_{12}	414

V.3. Results and discussion

V.3.1. Production of activated rhamnose precursors

Engineering of *S. bombicola* for the production of activated rhamnose is undertaken by following two strategies. The first one tries to import the bacterial dTDP-L-rhamnose pathway using glucose-1-phosphate as a substrate. In nature, this pathway uses four distinct steps to generate the activated rhamnose. However, in *Arabidopsis thaliana*, a 3,5-epimerase/4-keto reductase called AtNRS/ER (AT1G63000) was discovered that shows activity towards dTDP-4-keto-6-deoxy-glucose[454] and produces dTDP-L-rhamnose. A second strategy is based on the plant based system for the generation of UDP-rhamnose from UDP-glucose. This is catalysed by one single enzyme called RHM2/MUM4 (AT1G53500)[455]. One benefit of this strategy is that the enzyme can directly use the already available pool of UDP-glucose, which is efficiently generated in *S. bombicola*. Both kinds of activated rhamnose pathways will be introduced in *S. bombicola* and their products will be measured.

Isolation of nucleotide sugars and their subsequent detection is possible, but one needs to keep in mind that they are intracellular metabolites. Breaking open the cells and isolate them can result in relatively large variations so that real quantification can be tricky. A more robust way of analysis is using the available pool for further modifications like rhamnosylations. The presence of NDP-sugars can be proven and furthermore demonstrated to be accessible and usable for other enzymes. As a proof of concept, this pool will be used as a substrate for rhamnosylation of quercetin. For this, another gene from *A. thaliana* will be expressed in *S. bombicola*. The enzyme ATUGT78D1 (AT1G30530) is a flavonol-3-*O*-rhamnosyltransferase capable of using UDP-D-glucose, UDP-L-rhamnose and dTDP-L-rhamnose for *in vivo* glycosylations[451,456]. An overview of all the different steps in the creation of a flavonoid glycosylation platform can be seen in Figure 5.6. The different chemical structures of quercetin and its glycosides can be seen in Figure 5.7.

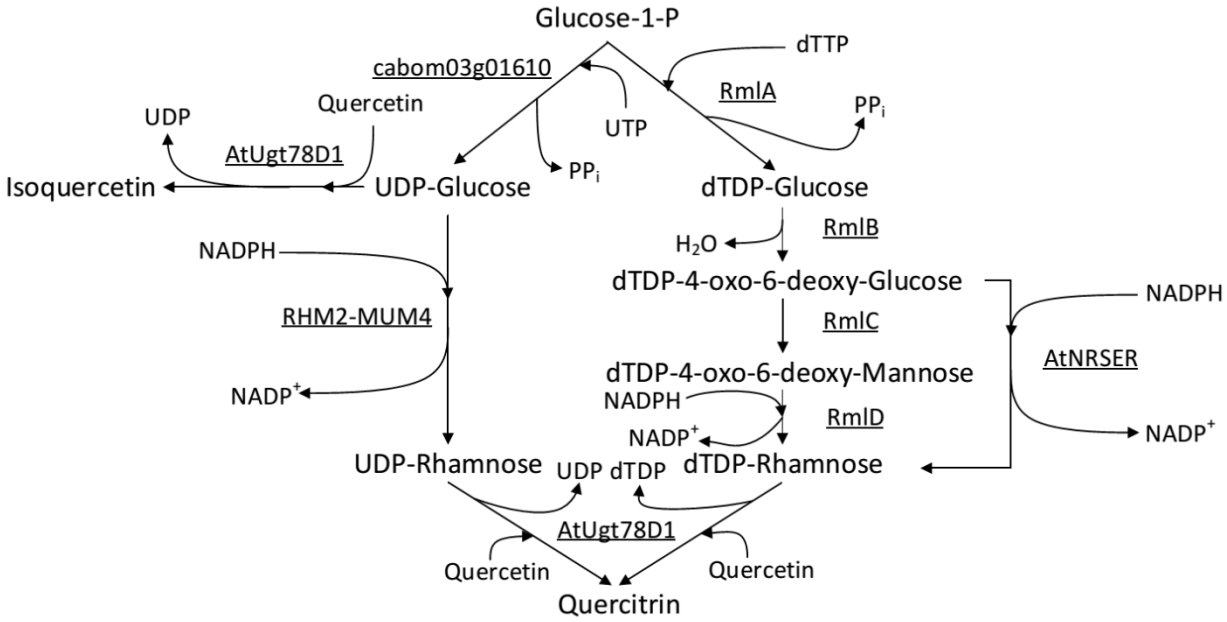


Figure 5.6: Overview of the steps required to produce UDP-L-rhamnose and dTDP-L-rhamnose from glucose-1-phosphate. Utilizing these pools for the glycosylation of quercetin is done by ATUGT78D1.

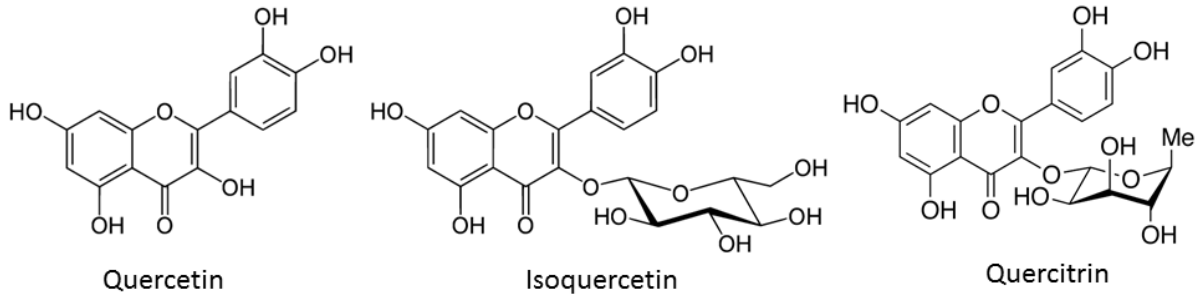


Figure 5.7: Chemical structures of quercetin and the glycosides that are produced by ATUGT78D1. Quercetin glucoside can be seen in the middle named isoquercetin while the rhamnoside can be seen as quercitrin.

V.3.1.1. Engineering and evaluation of NDP-L-rhamnose strains

A first production trial was conducted with the dTDP-L-rhamnose and UDP-L-rhamnose strains in comparison with the wild type and *cyp52m1* knock-out *S. bombicola* strains (for strain construction see Chapter V.2.2.2). Both NDP-L-rhamnose strains contain the genes necessary for production of activated rhamnose. The coding sequences are knocked-in at the *cyp52m1* locus for the UDP-L-rhamnose strain and at the loci of the acetyltransferase, *ugta1* and *cyp52m1* all situated inside the sophorolipid gene cluster. Regulatory sequences such as promotors and terminators were recycled for their respective locus. The *cyp52m1* knock-out was taken along since it does not produce sophorolipids, a product that might interfere with both the availability of UDP-glucose and further analysis. Both the supernatant and the cells were collected and processed. Nucleotide sugars ware purified and concentrated as described in materials and methods. Initially, the methods described by Räbinä *et al*.[452] was tested both for purification as analysis. Separation of the different nucleotide sugars was feasible using a Grace Alltech Prevail C18 HPLC Column as can be seen in Figure 5.8. Purification of cell lysate remained troublesome. The method developed by Räbinä *et al*.[452] appeared not suitable. A different method designed by Behmüller *et al*.[457] was tested as well, but the results remained inconclusive. Though the purification of a stock solution of UDP-glucose was possible, using the same technique for purification of nucleotide sugars from yeast cells remained inconclusive.

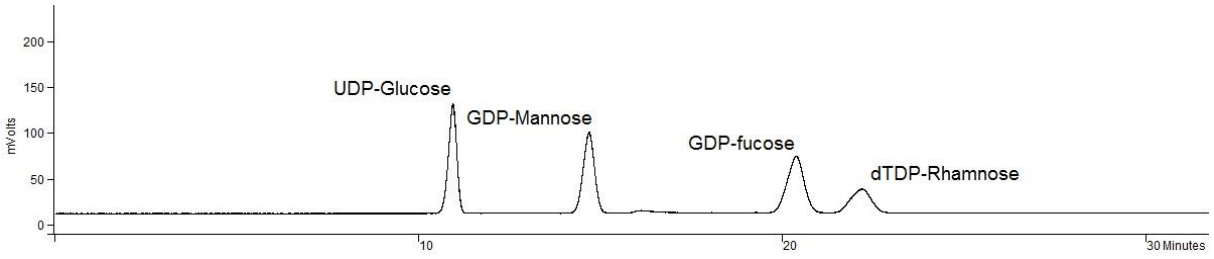


Figure 5.8: Standard of different nucleotide sugars used to optimise separation on a Grace Alltech Prevail C18 HPLC Column

Though nucleotide sugars can be isolated efficiently from cells according to several papers, no success was achieved with the various different methods tested. However, these molecules tend to be instable during longer periods of time by hydrolysis in the corresponding monosaccharide by enzymes belonging to the Nudix hydrolase superfamily. It is known that *S. bombicola* is not capable of utilizing rhamnose[458] for growth so any hydrolysis will result in a small accumulation of rhamnose over time. Therefore, detection of the monosaccharide rhamnose would in parallel prove the existence of the activated rhamnose inside and potentially outside the cell.

The same strains as mentioned above were grown on standard production medium without the addition of any hydrophobic substrate. After 8 days, the cells were collected by centrifugation and both the remaining broth and the cells were analysed for rhamnose production. In Figure 5.9, a direct comparison can be found between the different strains tested. Rhamnose has an elution time of approximately 14.5 minutes. When looking at the wild type and the *cyp52m1* knock-out strains, no signal can be detected at this time point. When comparing the NDP-rhamnose strains, one can see in comparison a new signal at 14.5 minutes. Though this is indicative of a novel molecule being formed in the rhamnose strains, it still shows a slightly different position compared to the standard. This might be explained by the more complex matrix in which the samples are dissolved. Standard addition of rhamnose was performed as well and confirmed the position of the new signal to be likely rhamnose. Though it is not a direct measurement of the nucleotide sugar, the origin of the rhamnose can only be derived from this activated molecule since it cannot be found in the control strains.

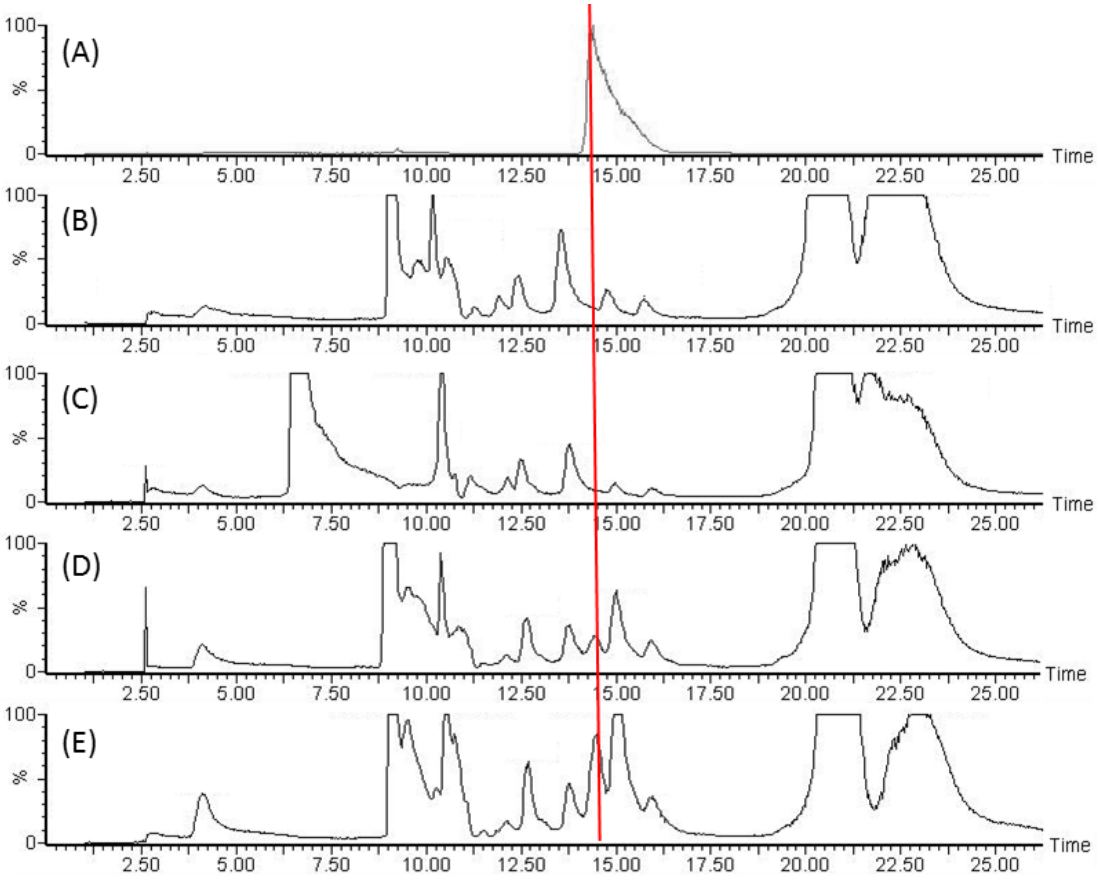


Figure 5.9: LC-MS chromatograms of the analysed supernatant for the different strains tested. (A) rhamnose standard, (B) wild type *S. bombicola*, (C) Δ*cyp52m1 S. bombicola*, (D) dTDP-L-rhamnose strain, (E) UDP-L-rhamnose strain. The elution time of rhamnose is 14.5 minutes and is marked by the yellow line.

V.3.1.2. Glycosylation of interesting biomolecules

Growth trial with quercetin

To assess the availability of the different NDP-rhamnose pools in the engineered *S. bombicola* strains, the *A. thaliana ugt78d1* gene was introduced in the NDP-L-rhamnose producing strains developed in V.2.2.2 as well in a *Δugta1* strain of *S. bombicola*. The latter one was chosen as a control since it doesn't produce any activated rhamnose so the *A. thaliana* transferase would only be able to produce quercetin glucosides. The NDP-L-rhamnose strain would be capable of producing both the rhamnoside and the glucoside. The three different mutant strains were inoculated and grown until the stationary phase at 48 hours. At this time point, quercetin dissolved in DMSO was added in a concentration of 1 g/L. After 8 days the cultures were harvested and both the supernatant and the intracellular matrix were analysed by TLC and LC-MS.

TLC already showed a clear difference between the dTDP-L-rhamnose strains and the two other strains. In Figure 5.10, a TLC is visible from an ethyl acetate extraction from the culture broth of all the strains tested. Though the amount of quercetin rhamnoside produced remains low, it is clearly visible for sample 2, the extraction corresponding to the dTDP-L-rhamnose strain. Interestingly, no quercetin glucosides are detected for all the strains tested. This is unexpected since the AT78D1 transferase is capable of utilizing UDP-glucose for the glucosylation of flavonoids.

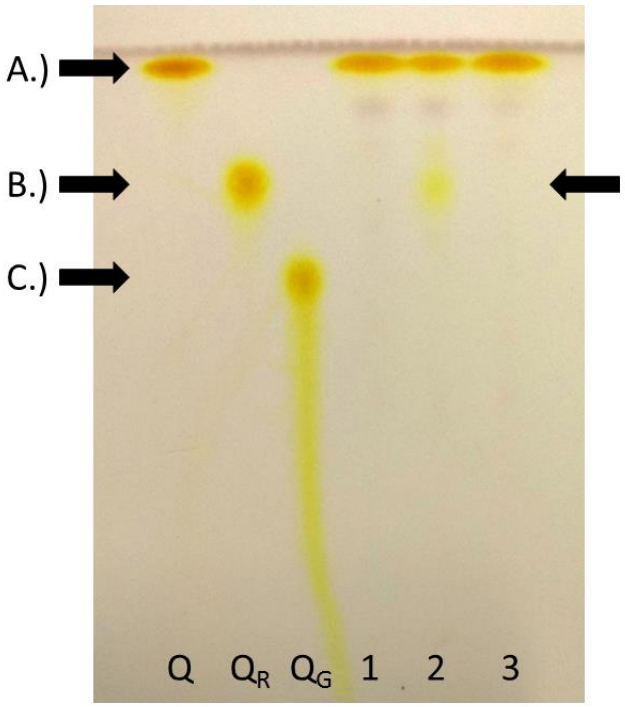


Figure 5.10: TLC of the ethyl acetate extraction from the growth trial with the different at78d1 carrying strains. Standards are visible in the first 3 lanes. Q = Quercetin, Q_R = Quercitrin, Q_G = Isoquercetin. Lane 1 corresponds to the strain only carrying the *at78d1* gene while lanes 2 and 3 are respectively the dTDP- and UDP-rhamnose strains carrying the *at78d1* gene.

Enzyme tests on crude cell lysate

Since the production levels were low, enzyme assays were conducted in parallel to test the production of glycosylated quercetin molecules using the activated rhamnose molecules. The 3 strains carrying the *at78d1* gene were inoculated in production medium and protein isolation was performed after 3 days in the stationary phase. For the UDP-rhamnose strain, only UDP-glucose and quercetin were added as substrates since biosynthesis of UDP-rhamnose starts from UDP-glucose. For the dTDP-rhamnose strain, dTDP-rhamnose and dTDP-glucose were tested. As a control, the enzyme test was also performed on the Δ*ugta1* cell lysate carrying the same *at78d1* gene. The enzyme assay was sampled at 3, 6 and 24 hours and analysed on both TLC and LC-MS. Interestingly, no activity could be detected for these experiments. This stands in contrast to the samples from the growth experiment were the quercetin rhamnoside was detected for the dTDP-rhamnose strain.

Though the exact nature for the lack of activity is unknown, several explanations can be found. First of all, quercetin is almost insoluble in water. This can lead to low substrate availability resulting in no or very low amounts of product formation. Secondly, the enzyme assays were conducted following the protocol also used for assaying the glucosyltransferases involved in the sophorolipid production[175]. It might be that the assay is not suitable for AT78D1, again resulting in no detectable activity.

V.3.2. Integration of a bacterial fatty acid synthase in *S. bombicola*

As in all living organisms, fatty acids can be synthesised *de novo* by the fatty acid synthase complex. In bacteria, this system is constituted out of different enzymes that are not associated with each other (see I.2.2). The growing acyl chain is transported between the different catalytic steps by the so-called acyl carrier protein (ACP), which buries the hydrophobic acyl chain inside the protein body shielding it from the watery cytosolic environment. In eukaryotes, the same types of catalytic steps are utilized for production of fatty acids. However, one big difference is the arrangement of the synthase complex. Instead of utilizing different enzymes, only two big proteins called α and β carry all the necessary domains[459]. These two enzymes form a multizyme consisting out of several α and β proteins and resembles a cage with the ACP protein (or better domain) trapped inside. This has as an advantage that locally high concentrations of fatty acid intermediates are present, pushing the equilibrium of the reactions towards the next step. However, this has as a consequence that the growing acyl chain is not available for the synthesis of other molecules.

To ensure that the precursors needed for rhamnolipid production are present, it was opted to introduce all the genes necessary for cytosolic, dissociated fatty acid syntheses. To accomplish this, eight different genes need to be introduced in *S. bombicola.* Though more enzymes can catalyse a specific step in the pathway, a selection was made to limit the amount of genes required to be knocked-in and to ensure an as high as possible productivity and product uniformity or that might be interesting enzyme engineering targets for further optimization. An overview is given in Figure 5.11.

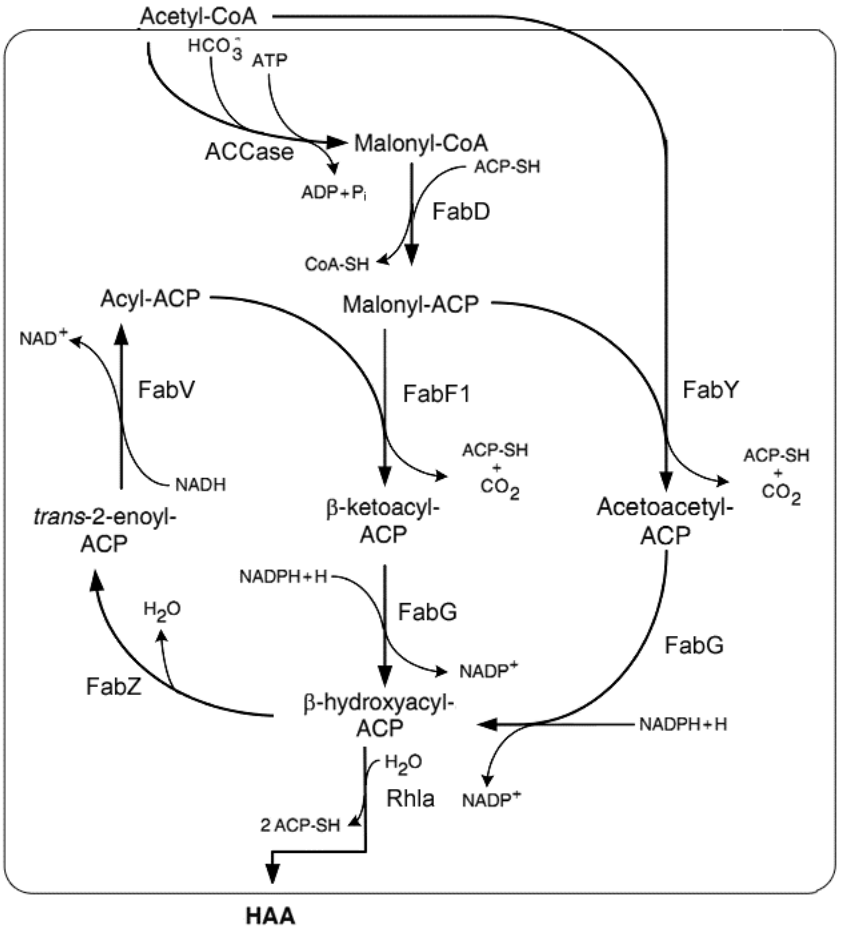


Figure 5.11: Proposed pathway for the production of HAA molecules in *S. bombicola*. The general precursor is Acetyl-CoA, originating from either glycolysis or β-oxidation of fatty acids.

As a proof of concept, HAA molecules will be produced. One limiting factor for the engineering of this bacterial fatty acid synthase complex in *S. bombicola* is the limited amount of defined promotors and terminators. Therefore, the promotors and terminators of the sophorolipid biosynthetic cluster will be used. These promotors are all active during the stationary phase. Another benefit is the bidirectional nature for the *ugtb1* and *ugta1* promotors, saving space in the expression cassettes. Though it is unknown if essential regulatory regions are present in the coding sequences of the sophorolipid biosynthetic genes upstream of these promotors, activity has been measured for the *cyp52m1* promotor[368].

V.3.2.1. Expression of rhlA in *S. bombicola Δcyp52m1*.

As already mentioned previously, RhlA required β-hydroxy fatty acids coupled to an ACP. Though highly unlikely, an initial experiment is performed by expressing RhlA in *S. bombicola* without the entire bacterial fatty acid synthase complex. The only β-hydroxy fatty acids coupled to an ACP are present in the fungal synthase complex and not freely accessible. Still it can't be ruled out 100% that RhlA cannot access them. To ensure no interference can occur from the native sophorolipid production, a *cyp52m1* knock-out strain[39] was used as a parent strain.

The analysis of the final samples was done using LC-MS and HPLC-ELSD. The masses of the expected congeners are described in V.2.6.3. As can be seen in Figure 5.12, no difference can be seen between the strain expressing the *rhlA* gene and the *cyp52m1* knock-out strain. Analysis on LC-MS did not reveal any masses corresponding with the HAA molecules. Therefore, the strategy of the implementation of a bacterial fatty acid synthase in an eukaryotic host was chosen as the next step.

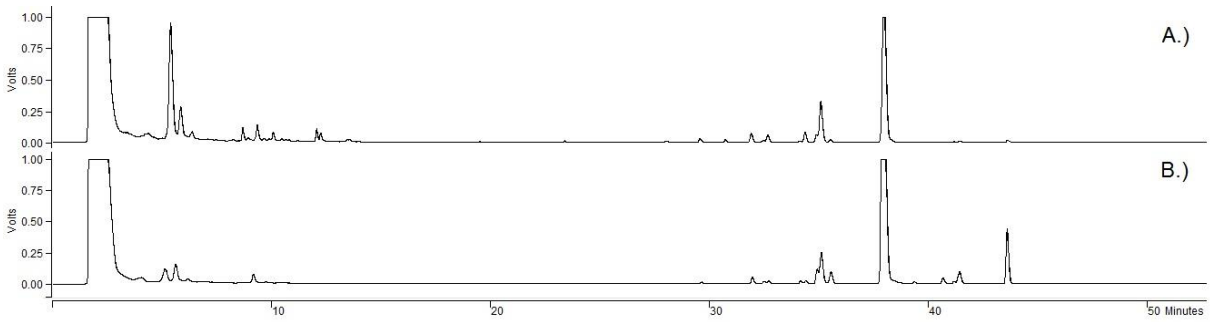


Figure 5.12: HPLC-ELSD chromatogram of A.) the *rhla* strain and B.) *cyp52m1* knock-out strain. The expected molecules for the RhlA strain would elute from the column between 25 and 35 minutes.

V.3.2.2. Introduction of a bacterial fatty acid synthase in *S. bombicola*

The different cassettes carrying the genes required for the FAS II pathway were created as described in V.2.2.2. Initially, after every transformation, the *ura3* marker would be removed by a recovery cassette. Though due to difficulties with this process, antibiotic resistance genes were used as an alternative. The coding sequence of the *acp* gene was equipped with a 6xHis tag for purification and detection by Western blotting. Due to the changing confirmation of the ACP during the elongation of the coupled acyl chain, a native polyacrylamide gel electrophorese allows the separation of different intermediate acyl-ACP congeners[81]. Both purification and detection of the ACP appeared difficult since no consistent results could be obtained.

The final strain carrying all the genes necessary was compared in a growth trial to the Δ*cyp52m1 S. bombicola* as the constructed HAA strain is not capable of producing sophorolipids. When analysing the chromatograms, it became clear that there was no detection of HAA molecules. The most likely explanation is a problem in the synthesis of these molecules, as the pathway relies on the correct expression of 9 genes and their correct interaction, it cannot be ruled out that the metabolism of *S. bombicola* interferes with production. HAA molecules are two or more β-hydroxy fatty acids coupled together. Growth trials performed by Roelants *et al*.[368] show that metabolism of rhamnolipids is possible, but only after prolonged incubation times and when glucose has been depleted. This was confirmed by the presence of rhamnose and monorhamnolipids in the culture broth, molecules not present in the substrate.

Analysis of the samples for this specific product was performed on LC-MS (Figure 5.13). When scanning for the masses corresponding to those formulated in V.2.6.3, no product could be identified. It has to be concluded that either one or several steps for the production of HAA molecules is not functioning or that the produced amounts are too low to compete with degradation or to be detected by the standard analytical methods at the lab.

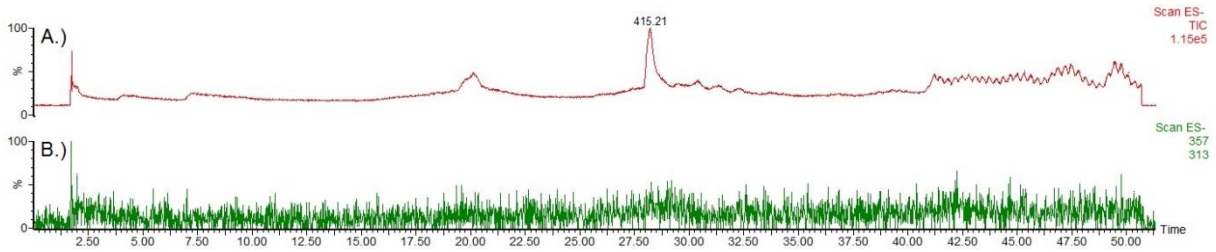


Figure 5.13: LC-MS chromatogram for the growth trial with the strain carrying the FAS II synthase complex and the first rhamnosyltransferase RhlA. A.) Total chromatogram; B.) Mass scan of the most common HAA congener consisting out of two β-hydroxy decanoic acid molecules (m/z 357).

V.3.3. Production of rhamnolipids by *S. bombicola*

Since no strain was created that could produce rhamnolipids *in vivo,* but only the necessary precursor strains, it was opted to test the *in vivo* and *in vitro* capabilities of the strains. Both NDP-rhamnose strains were equipped with the *rhlB* coding sequence, necessary for rhamnosylation of HAA molecules. During an initial production trial, both NDP-rhamnose strains were co-cultured with the HAA strain. After eight days, the broth was sampled and analysed on LC-MS. No molecules with masses similar to rhamnolipids could be detected.

During a second experiment, crude cell lysates were generated from both NDP-rhamnose strains as well as from the HAA strain. Both crude lysates were combined and incubated at 30 °C. Samples were taken after 24 hours. The lysate was extracted with an equal volume of ethyl acetate. This fraction was dried and the remaining pellet was dissolved. Analysis of the extract on LC-MS clearly showed that no new molecules are present when compared to the blanc. As can be concluded, the cell lysates of the precursor strains were not able to generate either HAA molecules or rhamnolipid like molecules.

V.4. Conclusion

Production of molecules structurally different from sophorolipids is possible with *S. bombicola*[318]. Still, often these molecules are derived from precursors that are easily accessible inside the cell. In the case of polyhydroxyalkanoate (PHA) production, the PHA synthase was targeted towards the peroxisome where the acyl-CoA precursors are readily available thanks to the β-oxidation. In the case of rhamnolipids production, the necessary precursors are not available or accessible for the different rhamnosyltransferases. Therefore, introduction of the bacterial derived pathways was needed. For the production of monorhamnolipids, the variant with only one rhamnose molecule attached, 11 genes need to be introduced when using UDP-rhamnose as the carbohydrate donor.

To assess the availability of the activated rhamnose, rhamnosyltransferase ATUGT78D1 from *A. thaliana* was introduced in the new strains. After an initial growth trial, quercitrin was detected on TLC for the dTDP-l-rhamnose strain. Though the acquired levels of quercitrin are low, together with the detection of rhamnose in the culture broth it proves the existence of an activated rhamnose molecule.

No limitations are present when looking at the regions where these genes can be introduced in *S. bombicola,* but the limited knowledge and availability regarding promotors, terminators and their activity strongly hamper the engineering. The different expression cassettes created in this chapter, all contain promotors belonged to the sophorolipid synthesis cluster. This choice was made in the beginning to save space in the expression cassettes, as two out of three of them are bidirectional, and they are active during the stationary phase. This way, a lower burden would be introduced on the yeast cells compared to expression during the stationary phase. Furthermore, due to a lack of a promotor library, it seemed the only possible strategy to follow.

During the engineering and testing of the novel strains, several problems turned up. First of all, in the case of the activated rhamnose, low product titers are present, even after prolonged incubation of the different production strains. This could be due to the presence of heterologous sequences inside the sophorolipid gene cluster. In Chapter II, several problems were encountered as well with different configurations of UGTA1-RFP constructs. Potentially, as stated previously, knocking in genes inside the sophorolipid gene cluster might have an influence on the activity of the promotors and subsequently the genes regulated by them.

For the ACP-coupled β-hydroxy fatty acids and derived HAA-molecules, no activity could be detected. Production of these molecules required introducing 9 genes at 3 different loci. Problematic expression of just one of those genes can already result in no products being formed. One reason might be that the genes introduced originate from a Gram-negative bacterium with a differing codon usage. In general, *S. bombicola* has an AT ratio of approximately 50 % while *P. aeruginosa* can reach values of 34 %. For the genes introduced, similar ratios were found. Though there might be no regulatory effect on the promotors, the differing codon usage can still influence the expression of all those genes. To test whether expression is the reason for no production, qPCR or MRM experiments could be conducted. They don't provide information of the confirmation of the enzymes being produced, it might learn something more about the bottlenecks in the research.

Chapter VI

Conclusion and perspectives

Chapter VI - Conclusion and perspectives

Traditional (bio)surfactants have many applications in for example personal care, oil recovery, cosmetics and agriculture (Figure 6.1). They promote solubilisation, emulsification and dispersion of other molecules. Globally, the market for surfactants is one of the largest when looking at the different industrial chemicals available. Their market share encompasses over €34 billion in 2016 and projected to grow each year with 5.6%. This translates in production volumes exceeding 15 million MT. Due to rising ecological awareness, a shift towards biobased surfactants can be seen. While the traditional surfactants are based on petrochemical resources, these biosurfactants are produced by renewable means. The first generation of biosurfactants are produced from oleochemicals like coconut and palm oil. It has to be noted that this is still a chemical process with often only the hydrophobic substrate being renewable. Second generation biosurfactants involve microbial produced molecules like sophorolipids (SLs), rhamnolipids (RLs) and mannosylerythritol lipids (MELs). Several multinationals show particular interest in these biosurfactants. MELs are mainly produced by Toyobo (Japan) while RLs mostly by Jeneil (USA) and Urumqi Unite Bio-Technology Co., Ltd. (China). SLs have a more international profile with producers, distributers and applicants such as Soliance (France), Ecover (Belgium), Evonik (Germany), Wheatoleo (France), Saraya (Japan), MG Intobio (Korea), SyntheZyme (USA) and multinationals such as Henkel. As already mentioned, these biosurfactants have their use in cosmetics, food, feed and drug formulation. When surveying the IP-landscape, multiple patents can be found for these applications. Still, a limited structural variety and high production costs hamper market penetration. Engineering of wild type or platform organisms can solve this problem. Since all commercial biosurfactants belong to the glycolipid class, expanding the diversity of the glyco- and lipo- part seem logical approaches to tackle the structural diversification. Lowering production cost can be done by optimization of the biosynthetic pathways and producer strains and state-of-the-art downstream processing.

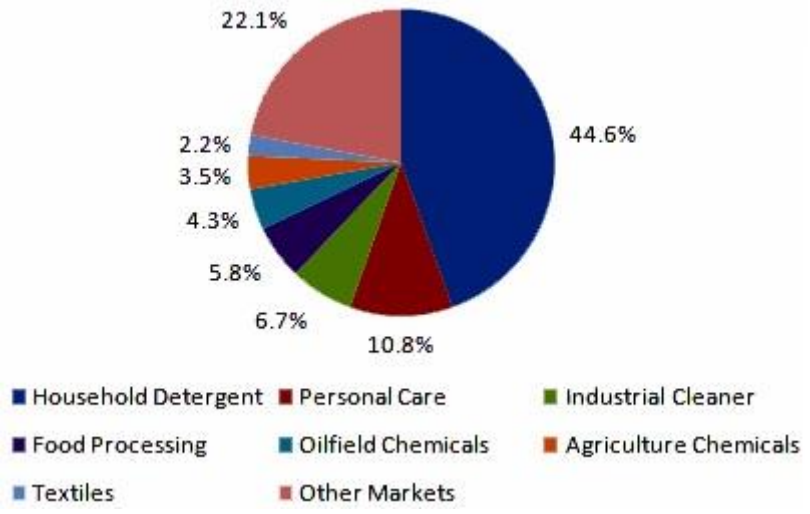


Figure 6.1: Biosurfactant market volume share by application in 2013[460]

In essence, molecular diversification can be achieved relatively easy when analysing the diversity already available in nature. When only glucose is considered, numerous disaccharides can be produced depending on the glycosidic bond between both monomers. However, certain carbohydrate moieties will be preferred over others because of their ease to be synthesised or other properties such as biological activity. From a practical point of view, the expansion of carbohydrate diversity can be reduced by searching for molecules with added biological benefits and by varying the degree of polymerisation. Depending on the hydrophobic group used, new physico-chemical properties will become available.

When looking at the commercialised glycolipids, several different carbohydrate and lipidic groups are available. The most studied glycolipids are composed of for example sophorose, rhamnose, mannosylerythritol and trehalose as carbohydrate group. Variation of the lipidic group includes different kinds of hydroxy fatty acids, either as is or modified and/or polymerised. To further expand this variety, several strategies can be utilized. Genetic engineering has been proven to be a valid way to either increase product uniformity or to obtain novel kinds of molecules. Cellobiose lipids without the subterminal hydroxyl group on the palmitic acid tail are obtained by deleting the gene coding for the second P450 involved in the biosynthesis. In the case of sophorolipids, either knocking-out or overexpressing certain enzymes such as the second glycosyltransferase UGTB1, the acetyltransferase or the lactone esterase gives rise to either glucolipids, sophorolipids enriched in for example acidic or lactonic variants and even bola-sophorolipids. Though these example all involve direct interference into the biosynthetic pathway, other metabolic processes can have an impact as well. A deletion strain of *S. bombicola* lacking the MFE-2 enzyme essential to the β-oxidation was able to produce higher amounts of novel kinds of sophorolipids than the wild type strain with no adverse effects on cell viability[314]. Process engineering can also help to further expand and control the structural variability. In essence, the test with the *mfe-2* knock-out is a good example of both kinds of engineering for enhanced sophorolipid production.

S. bombicola has the potential to be a platform organism for the production of glycolipids or other molecules. Still, decent strain development is only possible when a solid foundation of molecular tools is available. Turning *S. bombicola* in the organism of choice will require further expanding the fundamental knowledge and ways to steer the metabolism. In Chapter II, both expanding the molecular tools and understanding sophorolipid synthesis are described. Transformation and selection has been developed by Van Bogaert *et al*.[310] and further expanded to see the upper limits of the recombination system in *S. bombicola*. As became apparent, no clear limit can be detected up to 13.8 kb. During the research in Chapter II, it became clear that the stepwise lengthening and subsequent integration of cassettes for the *cyp52m1* locus were not hampered in integration efficiency up to 11 kb. This opened the door for a proof-of-concept integration cassette with the potential of having a beneficial effect on the production of sophorolipids. Therefore, an additional copy of the sophorolipid biosynthetic cluster was integrated at the *ura3* locus. The final integration cassette encompassed 14 kb in length and harboured five genes and a selection marker. The general idea was that an additional copy of each gene, except the lactone esterase, would help to enhance the flux throughout the pathway. Furthermore, it would enable future research on the regulation of the sophorolipid gene cluster. Unfortunately, no big gains in productivity could be measured and qPCR and MRM follow-up experiments provided opposite results and RNA and protein levels inside the cell. Still, the different experiments on integration length prove that it is possible to introduce entire pathways in one single transformation event at one specific location in the

genome. For the moment, only the *ura3* locus has been tested for cassettes of 14 kb in length. However, there is no reason to believe this is a feature unique to this locus since throughout the research, several different loci were targeted for engineering and often integration cassettes larger than 8 kb were introduced. The lactone esterase locus was engineered in Chapter V by an integration cassette of 9 kb carrying four genes and a marker.

One has to keep in mind that there're several limiting factors for this technique. First of all, the recombination cassettes are dependent on large homologues regions both at the 5' and 3' part. Smaller sequences can be used, but this will have an impact on efficiency. During the research, no severe problems were encountered with low transformation efficiencies. Second drawback is the lack of suitable markers to use. Though the *ura3* gene is already extensively used for engineering and counter selection is possible by 5-FOA, it slows down the speed at which new strains can be designed. Throughout this thesis, two antibiotic markers were used as well to circumvent problems with *ura3* removal. Still, expanding the number of auxotrophic markers is necessary for complex metabolic engineering of *S. bombicola*. In the case of *S. cerevisiae*, up to 12 auxotrophic markers are available[461]. The lack of markers combined with relative efficient recombination events became problematic in Chapter II where during the screening for a functional ARS sequence in *S. bombicola*, (il)legitimate recombination of the screening vector resulted in many mutants being created of which none had a functional plasmid. Targeting of the plasmid towards either the *ura3* locus (due to the promotor of the *ura3* gene), the *cabom03g11120* locus (source of the terminator for the *ura3* marker in the screening vector) or even other locations in the genome (due to the screened DNA sequences) made it possible for S. bombicola to integrate the marker. Interestingly, when the same screening vectors were transformed in *S. cerevisiae*, it became apparent that the *S. bombicola ura3* promotor was active in this yeast, allowing growth without integration of the plasmid. A potential solution to the integration problems might be using the *ura3* promotor of for example *S. cerevisiae* in *S. bombicola* for expression of the *ura3* marker. This way, homology is reduced.

Essential to homologues recombination is the double-stranded break (DSB) of the DNA at the location where recombination is wanted. Several strategies are available to enhance the efficiency of homologues recombination by deliberately introducing these DSB. A powerful technique is using nucleases. Nucleases such as zinc finger nucleases (ZFN)[462] or transcription activator-like effector nucleases (TALEN)[463] can be targeted towards specific location in the genome and introduce double stranded DNA breaks. Both techniques rely on combining DNA interacting domains with the non-specific nuclease *Fok*I. Engineering of the DNA interacting domain results in identification of specific sequences unique for the targeted locus or genomic region. Still, a major drawback for both techniques is the nuclease used. *Fok*I needs to be a dimer before genome cutting can be performed. Therefore, an upstream and downstream ZFN or TALEN needs to be designed. This not only complicates the design of both classes, it also leads to the risk of homodimers being formed resulting in potential cleavage of genomic regions not targeted for engineering. Another strategy is using meganucleases[464,465], DNA cleaving enzymes that recognize DNA sequences of 15 to 40 bp. Engineering of these enzymes is more complicated since the recognition and cleavage regions of the enzyme are grouped into a single domain. Still, several examples of successful engineering in plants, animals[466] or even human cells[467] can be found for all the techniques mentioned.

One more drawback for the nuclease derived techniques is their locus dependency. The DNA interacting domains are all specific for one locus. If extensive engineering is required at multiple

locations several ZFNs, TALENs or meganucleases will have to be designed and introduced in the cells. A locus independent technique such as clustered regularly interspaced short palindromic repeat RNA-guided Cas9 nucleases (better known as CRISPR/Cas9) has already been used for yeast engineering. In *S. cerevisiae*, the technique has been used successfully for multigene knock-outs with efficiencies up to 87 %[468]. Using CRISPR/Cas9 in *S. bombicola* would require some engineering before the tool can be used. In other organisms, the coding sequences for the proteins can be delivered by a transient plasmid, but this is not possible for *S. bombicola*. The genes required would have to introduced in the genome first, preferably under the control of an inducible promotor to have full control over the activity of the system.

Another essential part of the genetic engineering is the availability of standardised sequences. In the case of *S. bombicola*, only a couple of promotors have been characterised throughout the years. Reusing promotors with known properties was attempted in Chapter V to circumvent this problem, but ideally a promotor library is available for optimal balancing of the engineered pathways. This balancing is crucial for achieving optimum strain fitness and production. An imbalance in gene expression or enzyme activity can lead to stress responses due to excessive protein build-up, precursor depletion and accumulation of intermediates, which can inhibit enzymes or even can become cytotoxic. An example for this is the production of dTDP-L-rhamnose. When intracellular levels of this molecule become too high, it effectively inhibits RmlA, the first step of its own biocatalytic pathway. Ideally these molecules are further used such as rhamnosylation of biomolecules as attempted in Chapter V. Creation of such a library can be done using the fluorescent proteins of Chapter II.

Bidirectional promotors are of special interest since they can express two genes while often using the same amount of base pairs as a monodirectional promotor. This way, more genes can be introduced into an expression cassette with limited length. This was proven to be a valid strategy in the case of activated rhamnose production in Chapter V. For the production of dTDP-L-rhamnose, three genes were introduced at the loci of the acetyltransferase *at*, glucosyltransferase *ugta1* and P450 *cyp52m1* in the sophorolipid biosynthetic cluster. The detected production of rhamnose proved the activity of the (bidirectional) promotor pUGTA1-pCYP52M1.

Still, some caution is necessary as it became apparent in Chapters II and IV where suboptimal expression and production was achieved with certain configurations of genes and promotors. In the case of the RFP-coupled UGTA1, the N-terminal coupled RFP resulted in almost no production and fluorescence. Potentially one has to take into account the local DNA structure or regulatory sequences encoded in the coding sequences of the originally upstream of a promotor. Though no information is available on the regulation of sophorolipid synthesis, in other organisms the genes encoding secondary metabolites are often clustered in subtelomeric regions. Several ways of gene regulation are possible. In the case of cellobiose lipids, *rua1* encodes a transcription factor necessary for the expression of the entire biosynthetic cluster. For *S. bombicola*, no such element can be found either inside the cluster or outside. The data obtained from RNA sequencing also did not yield potential candidates since no transcription factor or transcription related genes seem to behave the same or opposite of the sophorolipid biosynthetic genes. Another way of gene regulation is by global regulators that integrate environmental changes. Sophorolipid synthesis is initiated at high carbon/nitrogen ratios. Potentially, such a global regulator might be involved. A third kind of regulation is epigenetic. Certain modifications of histones can result in either dense chromatin called heterochromatin and the more open euchromatin. The latter one is more accessible for

enzymes acting on the DNA such as polymerases. Local disruptions of the DNA can have an influence on the state of the chromatin and subsequently the expression of the genes in that area. An interesting experiment to investigate the potential regulation can be done using the double cluster strain developed in Chapter II. Though the results obtained do not support any kind of conclusion, the technique itself could be used. Development of a strain lacking the entire cluster at its original location, but having a copy at the *ura3* locus might result in a loss of regulation due to the different genomic location.

Another interesting strategy to circumvent the potential negative influences of regulation on the production of novel molecules would be using characterised promotors and terminators. However, multi-step pathways require each gene to be suited with its own promotor and terminator, quickly resulting in large expression cassettes. Until recently, polycistronic expression of pathways was a quality exclusively given to bacteria. Several techniques are developed to allow multiple proteins to be translated from a single mRNA molecule in eukaryotes. A first example is the introduction of Internal ribosome entry sites (IRES)[469] (Figure 6.2 A). These IRES allow the ribosomes to initiate protein synthesis at any place on the mRNA. Eukaryotic mRNA translation starts by binding of the cap-binding complex eukaryotic initiation factors (eIF) 4F to the mRNA 5'-m_7G cap-structure. This complex acts as a scaffold for several initiation factors as well as ATP-dependent RNA helicase eIF4A, which dissolves the secondary structure of the 5'-untranslated region (5'-UTR). Furthermore, the 40S ribosomal unit will be recruited and positioned with several other enzymes at the first encountered start codon. Final recruitment of the 60S ribosomal unit forms the elongation-competent 80S ribosome[470]. When no cap structure is present, initiation of translation can still occur when certain secondary and tertiary mRNA structures are present in de 5-UTR. Several types of IRES sequences have been identified and have been grouped depending on the amount of canonical eIFs and IRES trans-activating factors are needed[471]. The higher the group number, the more the IRES initiation complex starts to resemble the cap-dependent system.

However, problems with expression levels of proteins encoded in front or behind the IRES[472] required the development of different systems. Several viruses employ so-called 2A peptides to create polycistronic expression[473]. The coding sequences of these peptides are situated between the open reading frames (ORFs) of the genes that need to be expressed. At the end of the 2A amino acid sequence, several highly conserved residues introduce a ribosomal skip, successfully preventing the subsequent condensation of the next amino acid to the already formed peptide chain and thereby releasing smaller individual proteins instead of one big protein (Figure 6.2 B). This technique has been used in several eukaryotes like *Pichia pastoris*[474], *Arabidopsis thaliana*[475] and even mammalian cells originating from human[476], zebrafisch[476] or mice[477] tissue. All 2A peptides have a conserved GDVE(E/S)NPGP motif at the end of their amino acid sequence. In front of this motif, some sequence conservation can be seen for several residues, but their exact role remains unclear.

Caution is necessary as several drawbacks exist for this technique. The major one is that this has never been attempted in *S. bombicola* so it is uncertain that it will work. Secondly, in the case of the 2A peptides, there exists a high degree of homology between the different types. Bicistronic expression is not bothered by this, but polycistronic expression might have to deal with illegitimate recombination and thereby a loss of function. The 2A tool can also help in the development of an episomal vector for *S. bombicola*. When looking at for example the 2µ plasmid, four genes are involved. These are *rep1*, *rep2*, *rep3* and *flp*. Introduction of these genes in the *S. bombicola* genome might result in a strain able to use the 2µ plasmid.

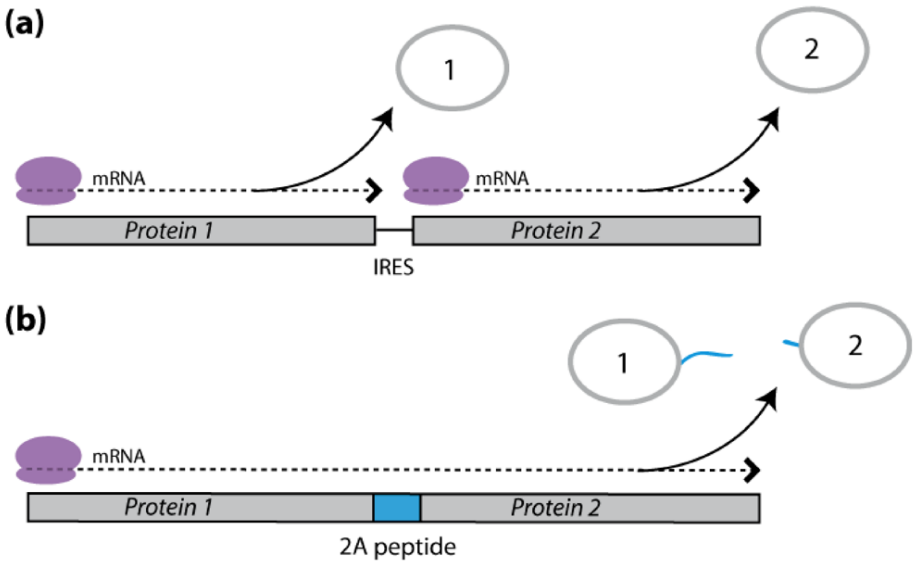


Figure 6.2: The mechanism behind IRES (a) and 2A peptides (b).

Still, it is possible to locally change the DNA sequence and not influence expression of neighbouring genes. This is clear from several experiments conducted in chapters II, III and IV. Substituting or altering coding sequences still resulted in good activity for several mutant strains. This clearly points towards the possibility of generating new kinds of precursors for glycolipid production and further on utilizing either the sophorolipid biosynthetic genes or heterologous ones.

In Chapter III, P450 engineering is proven as a way to provide shorter hydroxy fatty acids to the glucosyltransferases. Several P450s with interesting properties were either introduced without any modifications, as a part of a fusion protein or mutated at an essential amino acid for their regiospecificity. In the case of unmodified P450s, CYP1 from *U. maydis*, CYP52A4 from *C. maltosa*, CYP102A1 from *B. megaterium*, CYP102A7 from *B. licheniformis* and $P450_{Foxy}$ from *F. oxysporum* were used. Though activity was detected for all except CYP52A4, production of specific molecules remained difficult. For the self-sufficient CYP102A1, CYP102A7 and $P450_{Foxy}$, mixtures of sophorolipids with palmitic, oleic and stearic acid were detected. For CYP1, the produced molecules were more uniform with mostly palmitic acid being incorporated. Still, only marginal activity was detected for CYP1. The reason for this is unknown, but the most likely is CPR incompatibility. P450s of class II require electrons to be donated from a second reaction partner (see I.3.3). The interaction between P450 and CPR is electrostatically and small changes on the surface of either enzyme can have a big impact on electron delivery[478]. Therefore, the creation of fusion proteins, so called chimeric P450s was attempted. By coupling a reductase domain to the P450, a more efficient electron transfer could be obtained. In the case of CYP1 and CYP52A4, activity could be detected when the P450 was coupled to the reductase domain of CYP102A1 called BMR. Even though the productivity was low, it was a successful proof of concept for P450 engineering. For CYP1, further optimisation of the production process resulted in the production of highly uniform C16 sophorolipids using a strain carrying the CYP1BMR chimeric P450. This proves that introducing novel P450s in *S. bombicola* is a valid strategy to alter the fatty acid tail of the glycolipid produced.

Still, several strategies remain to improve the production of novel kinds of sophorolipids. For the already created chimeric strains, further engineering attempts could be made. A first one

is the overexpressing of these P450 chimeras. Though more protein not necessarily translates to higher production values, it is a first simple strategy to investigate. A second one focusses on the coding sequence and structure of the P450 itself. As already mentioned in III.3.2.4, several remarks can be made on the quality of the coding sequence used for CYP1. If the starting methionine is indeed the one from position 105, a new chimeric CYP1 can be made with a shorter region of the protein inside the ER. A third solution lays in the resistance of *S. bombicola* towards shorter fatty acid chains. If the stress caused by shorter chains impairs productivity, adaptive evolution of these strains towards higher resistance might help to reach higher product titers of short- and medium-chain sophorolipids.

A more fundamental and extensive solution requires exploring the linker between the reductase and P450. For the moment the natural CYP102A1 linker is used for all the different chimeric P450s, but numerous literature reports have proven the importance of the linker in P450 chimeragenesis[479–482] since the mobility of both domains relative to each other plays an important role for efficient electron transport. CYP102A1 might function as a dimer *in vivo* with the BMR domain donating electrons to the P450 domain of the second enzyme[481]. If the BMR is less-than-optional positioned, this can severely interfere with electron transport. In the case of chimeragenesis of P450s, the rigidity of the linker will have a serious influence on the degree of freedom of the reductase domain towards its own or other P450s. On the other hand, if the electron transfer happens within the chimeric P450, linker length will have a role as well. If any improvement could be made at all for the P450 system in *S. bombicola* through chimeragenesis, different linkers should be investigated. Molecular tools are available to construct libraries of P450 chimeras with different linker lengths and structures, for example P-Link[483] and DuaLinX[484]. Related to the linker is the reductase used for creating the chimeric enzyme. As could be seen for CYP52M1, production was impaired by the coupling of the reductase while it enabled production for CYP52A4 and CYP1. Combinatorial design of the chimera using different reductases and linkers can further enhance the knowledge on creating these proteins.

Though all these strategies tested are based on introducing heterologous P450s with interesting characteristics, *S. bombicola* has some other potential candidates as well. A first candidate is *cyp52m1*. A structural study of might result in identification of the essential amino acid residues responsible for either substrate, regio- and stereospecificity and potentially altering them. Besides *cyp52m1*, the coding sequences of 7 other P450s belonging to the CYP52 family are present in *S. bombicola*. In Table 6.1, an overview is given and their sequence identities and similarities. Recently, two were tested by *in vitro* experiments. Both CYP52N1 and CYP52E3 showed activity towards shorter fatty acids like myristic acid[254]. High amino acid sequence similarities exist between the members of the same subfamily, but they still might be sufficiently different to allow different products to be formed. Characterisation of these enzymes might lead to the discovery of P450s active towards longer or shorter fatty acids. Still, modification of the lipidic part is no guarantee for the efficient production of novel sophorolipids due to the specificity of the glucosyltransferases. UGTA1 is known to have already lower activity towards hydroxy palmitic acid and almost no activity towards hydroxy lauric acid. Identification of the residues responsible for this restricted specificity can lead to engineered mutant enzymes with a tailor-made substrate and product profile. Another benefit of utilizing *S. bombicola* P450 is the compatibility of the P450s with the CPR of this yeast. In the same light P450s from closely related species like *C. apicola* or *C. batistae* could be used as well.

Table 6.1: Overview of all the CYP52 family P450s present in *S. bombicola*. Sequence identities are given above the diagonal while sequence similarities are given below the diagonal. All percentages are calculated using the ORCAE blast tool.

P450\P450	CYP52E3	CYP52E4	CYP52E5	CYP52E6	CYP52M1	CYP52N1	CYP52N2	CYP52N3
CYP52E3		85%	85%	84%	47%	42%	43%	43%
CYP52E4	92%		99%	85%	46%	42%	42%	42%
CYP52E5	82%	100%		85%	46%	42%	41%	42%
CYP52E6	93%	92%	92%		45%	42%	42%	43%
CYP52M1	62%	62%	62%	62%		46%	45%	45%
CYP52N1	59%	59%	59%	60%	64%		87%	86%
CYP52N2	58%	58%	57%	58%	63%	95%		95%
CYP52N3	58%	58%	58%	59%	63%	94%	97%	

Modification of the carbohydrate part is possible as well. Though this has been done in the past by Saerens et al.[317] and Roelants et al.[318] for the production of glucolipids and cellobiose lipids, productivity remained low compared to other engineered strains described[174,423]. Interestingly, all the strains altered in their carbohydrate biosynthesis are derived from *S. bombicola* G9, a spontaneous *ura3* mutant. Other, better producing strains are either derived from the wild type *S. bombicola* ATCC22214 or the derived PT36 strain. This PT36 strain is a specifically engineered *S. bombicola* strain where the coding sequence of the *ura3* marker has been removed by an engineered recombination cassette. A comparing production trial between both auxotrophic strains clearly showed the difference in production capacity. While the PT36 easily reaches titers up to 25 g/L, comparable to wild type production levels in shake flask experiments, the G9 hardly achieved 1 g/L. Though these results do not explain the low productivity of the glucolipid and cellobiose lipid strains, it is a good indication that the spontaneous mutant is not a suitable starting strain for engineering.

Introduction of several knock-out and knock-in cassettes in the PT36 strain resulted in production of glucolipids comparable to wild type sophorolipid levels for all the strategies tested. For the original glucolipid producing strain, the most likely explanation according to Saerens et al.[317] was transporter incompatibility. Though the *S. bombicola* sophorolipid has never been characterised, selectivity towards certain molecules can occur. This is for example observed for MELs where 2' and 3' acylations are necessary for export[328]. The growth trial and subsequent upscaling however do not support this hypothesis since production can be improved dramatically by only changing the starting strain.

Further carbohydrate modifications were obtained by introduction of the *U. maydis ugt1* gene involved in cellobiose lipid production. The original G9 derived cellobiose lipid strain produced a mixture of glucolipids and cellobiose lipids in low amounts. Higher product uniformity was obtained by utilizing a *S. bombicola* strain deficient in the *ugta1* gene. This way, no glucolipids could be produced. Further enzyme characterisation of UGT1 proved that this enzyme is indeed the sole transferase responsible for the sequential double glucosylation of 16-OH palmitic acid. The experiments on the cell lysate of an *S. bombicola* UGT1 overexpression strain further indicated that low amounts of activity are present. The real cause of this low activity is unknown, but several explanations can be made. It might be that UGT1 is not that active resulting in low productivity in an *in vivo* system. Another explanation is the potential localisation of the enzyme. As was proven in Chapter II, UGTA1 and CYP52M1 are colocalised inside the cell. The reason for this colocalization is unknown, but potentially electrostatic interactions might play a role. Since *ugt1* comes from a non-related fungus, interaction between UGT1 and CYP52M1 might be non-existent resulting in the enzyme not efficiently

receiving the hydroxy fatty acids required for cellobiose lipid production. Furthermore, as already mentioned in Chapter 4, the hydroxy fatty acids produced by CYP52M1 might not have the correct regio- and stereospecificity. In *U. maydis*, only terminal hydroxy fatty acids are incorporated in the cellobiose lipids. It is still unknown whether UGT1 can accept subterminal hydroxylated fatty acids and if so, which stereochemistry they must have. Still, by introducing heterologous transferases carbohydrate modifications can be made possible. Though only cellobiose lipids are being produced for the moment, other types of glycolipids are possible as well. Finding suitable enzymes by screening transferases for their capability of using hydroxy fatty acids or intermediate glycolipids can be done relatively fast *in vitro* or *in silico*. However, this does not ensure optimal production of the desired molecules when using them in *S. bombicola*, it can give an idea on the possibilities for the *in vivo* glycorandomasation possibilities. Large data sets are available for carbohydrate-processing enzymes thanks to glycoarrays.

The substrate pool used for glucolipids and cellobiose lipids is UDP-glucose (Chapter IV), also efficiently used for the sophorolipid biosynthesis. Other kinds of glycolipids, for example rhamnolipids or xylolipids, require non-existing or suboptimal precursor pools. However, as can be seen in Figure 6.3, several interesting nucleotide sugars can be produced starting from existing pools. An interesting strategy is using the already available UDP-glucose pool as a start and by introducing suitable NDP-sugar modifying enzymes, new precursor pools can be created. It is for example possible to convert UDP-glucose to UDP-rhamnose[455] and UDP-galactose[485] using only one catalytic step. Creating these carbohydrate donors and combining them with capable transferases can lead to new-to-nature glycolipids with altered properties originating from the altered hydrophilic moiety. In Chapter V it is shown that is indeed possible to utilize the UDP-glucose and glucose-1-phosphate pool for the creation of activated rhamnose.

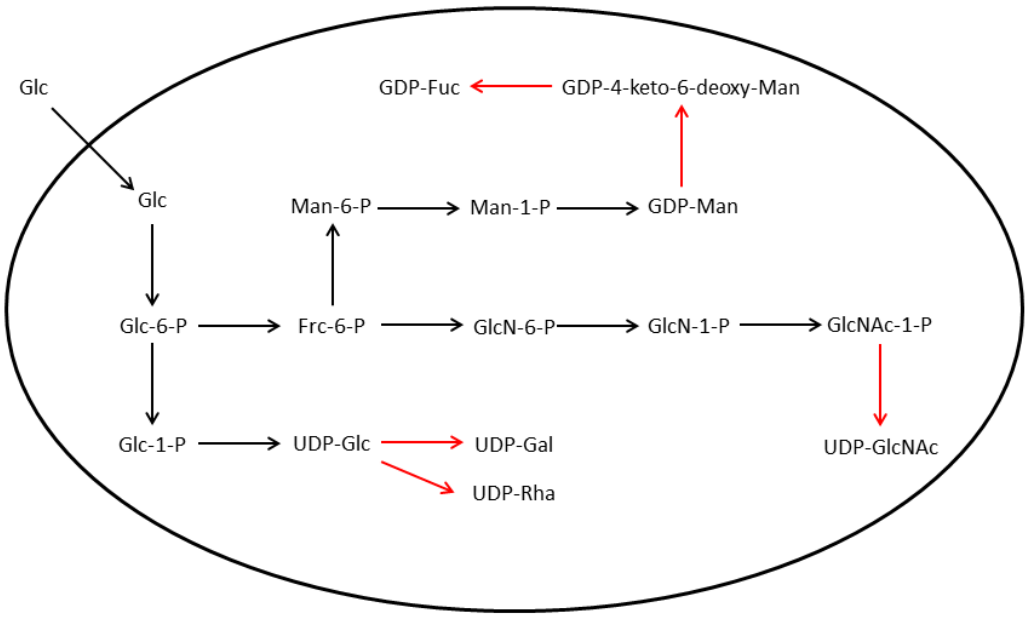


Figure 6.3: Potential production of several nucleotide sugars starting from readily available pools inside *S. bombicola*. Glc = glucose; Frc = fructose; Fuc = fucose; Man = mannose; Gal= galactose; GlcN = glucosamine; GlcNAc = N-acetyl-glucosamine; Rha = rhamnose.

As can be seen throughout this work, *S. bombicola* can be turned into a platform organism for not only different kinds of sophorolipids, but also other kinds of glycolipids. Several obstacles still need to be tackled. A first hurdle to be taken is the limited fundamental knowledge on promotors, the position effect of the targeted locus and the regulation of sophorolipid synthesis. Though the latter one can be evaded by integrating pathways at other locations than the sophorolipid gene cluster, production of glycolipids related to sophorolipids might benefit from a better understanding on how the sophorolipid gene cluster is actually regulated. Expanding the molecular toolkit with for example a stable plasmid or the above mentioned polycistronic expression methods can further enhance the speed of engineering resulting in both more knowledge and the generation of new production strains.

References

References

1. Plinius Secundus Senior, C. *Naturalis Historia Volume VIII Book 28*.
2. Ceresana. *Surfactants – Study: Market, Analysis, Trends*. (2015).
3. Ivanković, T. & Hrenović, J. Surfactants in the environment. *Arch. Ind. Hyg. Toxicol.* **61,** 95–110 (2010).
4. Zeng, G. *et al.* Co-degradation with glucose of four surfactants, CTAB, Triton X-100, SDS and Rhamnolipid, in liquid culture media and compost matrix. *Biodegradation* **18,** 303–310 (2007).
5. Lima, T. M. S. *et al.* Evaluation of bacterial surfactant toxicity towards petroleum degrading microorganisms. *Bioresour. Technol.* **102,** 2957–2964 (2011).
6. Baccile, N. *et al.* Development of a Cradle-to-Grave Approach for Acetylated Acidic Sophorolipid Biosurfactants. *ACS Sustain. Chem. Eng.* **5,** 1186–1198 (2017).
7. Fischer, E. Ueber die Glucoside der Alkohole. *Berichte der Dtsch. Chem. Gesellschaft* **26,** 2400–2412 (1893).
8. von Rybinski, W. & Hill, K. Alkyl Polyglycosides—Properties and Applications of a new Class of Surfactants. *Angew. Chemie Int. Ed.* **37,** 1328–1345 (1998).
9. Rosevear, P., Van Aken, T., Baxter, J. & Ferguson-Miller, S. Alkyl glycoside detergents: a simpler synthesis and their effects on kinetic and physical properties of cytochrome c oxidase. *Biochemistry* **19,** 4108–4115 (1980).
10. Koenigs, W. & Knorr, E. Ueber einige Derivate des Traubenzuckers und der Galactose. *Berichte der Dtsch. Chem. Gesellschaft* **34,** 957–981 (1901).
11. Franzetti, A. *et al.* Environmental fate, toxicity, characteristics and potential applications of novel bioemulsifiers produced by *Variovorax paradoxus* 7bCT5. *Bioresour. Technol.* **108,** 245–251 (2012).
12. Daverey, A. & Pakshirajan, K. Sophorolipids from *Candida bombicola* using mixed hydrophilic substrates: Production, purification and characterization. *Colloids Surfaces B Biointerfaces* **79,** 246–253 (2010).
13. Khopade, A. *et al.* Production and stability studies of the biosurfactant isolated from marine Nocardiopsis sp. B4. *Desalination* **285,** 198–204 (2012).
14. Reis, R. S., Pacheco, G. J., Pereira, A. G. & Freire, D. M. G. in *Biodegradation - Life of Science* (eds. Chamy, R. & Rosenkranz, F.) 31–61 (InTech, 2013).
15. Desai, J. D. & Banat, I. M. Microbial production of surfactants and their commercial potential. *Microbiol. Mol. Biol. Rev.* **61,** 47–64 (1997).
16. Rau, U., Hammen, S., Heckmann, R., Wray, V. & Lang, S. Sophorolipids: A source for novel compounds. *Ind. Crops Prod.* **13,** 85–92 (2001).
17. Zhang, J., Saerens, K. M. J., Van Bogaert, I. N. A. & Soetaert, W. Vegetable oil enhances sophorolipid production by *Rhodotorula bogoriensis*. *Biotechnol. Lett.* **33,** 2417–2423 (2011).
18. Konishi, M., Fukuoka, T., Morita, T., Imura, T. & Kitamoto, D. Production of new types of sophorolipids by *Candida batistae*. *J. Oleo Sci.* **57,** 359–369 (2008).
19. Lang, S. & Wullbrandt, D. Rhamnose lipids--biosynthesis, microbial production and application potential. *Appl. Microbiol. Biotechnol.* **51,** 22–32 (1999).
20. Dubeau, D., Déziel, E., Woods, D. E. & Lépine, F. *Burkholderia thailandensis* harbors two identical *rhl* gene clusters responsible for the biosynthesis of rhamnolipids. *BMC Microbiol.* **9,** 263–275 (2009).
21. Spoeckner, S., Wray, V., Nimtz, M. & Lang, S. Glycolipids of the smut fungus *Ustilago maydis* from cultivation on renewable resources. *Appl. Microbiol. Biotechnol.* **51,** 33–39 (1999).
22. Rau, U., Nguyen, L. A., Roeper, H., Koch, H. & Lang, S. Fed-batch bioreactor production of mannosylerythritol lipids secreted by *Pseudozyma aphidis*. *Appl. Microbiol. Biotechnol.* **68,** 607–613 (2005).
23. Uchida, Y., Tsuchiya, R., Chino, M., Hirano, J. & Tabuchi, T. Extracellular accumulation of mono- and di-succinoyl trehalose lipids by a strain of *Rhodococcus erythropolis* grown on n-alkanes. *Agric. Biol. Chem.* **53,** 757–763 (1989).
24. Hammami, W., Labbé, C., Chain, F., Mimee, B. & Bélanger, R. R. Nutritional regulation and kinetics of flocculosin synthesis by *Pseudozyma flocculosa*. *Appl. Microbiol. Biotechnol.* **80,** 307–315 (2008).

25. Matsuyama, T., Kaneda, K., Ishizuka, I., Toida, T. & Yano, I. Surface-active novel glycolipid and linked 3-hydroxy fatty acids produced by *Serratia rubidaea*. *J. Bacteriol.* **172,** 3015–3022 (1990).
26. Joshi-Navare, K., Singh, P. K. & Prabhune, A. A. New yeast isolate *Pichia caribbica* synthesizes xylolipid biosurfactant with enhanced functionality. *Eur. J. Lipid Sci. Technol.* **116,** 1070–1079 (2014).
27. Yeh, M.-S., Wei, Y.-H. & Chang, J.-S. Bioreactor design for enhanced carrier-assisted surfactin production with *Bacillus subtilis*. *Process Biochem.* **41,** 1799–1805 (2006).
28. Li, H., Tanikawa, T., Sato, Y., Nakagawa, Y. & Matsuyama, T. *Serratia marcescens* gene required for surfactant serrawettin W1 production encodes putative aminolipid synthetase belonging to nonribosomal peptide synthetase family. *Microbiol. Immunol.* **49,** 303–310 (2005).
29. Lin, S.-C., Minton, M. A., Sharma, M. M. & Georgiou, G. Structural and immunological characterization of a biosurfactant produced by *Bacillus licheniformis* JF-2. *Appl. Environ. Microbiol.* **60,** 31–38 (1994).
30. Ng, I.-S., Ye, C., Zhang, Z., Lu, Y. & Jing, K. Daptomycin antibiotic production processes in fed-batch fermentation by *Streptomyces roseosporus* NRRL11379 with precursor effect and medium optimization. *Bioprocess Biosyst. Eng.* **37,** 415–423 (2014).
31. Vandamme, E. J., Leyman, D., De Visschera, P., De Buyser, D. & Vansteenkiste, G. Effect of Aeration and pH on Gramicidin S Production by *Bacillus brevis*. *Culture* **31,** 247–257 (1981).
32. Lépine, F., Déziel, E., Milot, S. & Villemur, R. Liquid chromatographic/mass spectrometric detection of the 3-(3-hydroxyalkanoyloxy) alkanoic acid precursors of rhamnolipids in *Pseudomonas aeruginosa* cultures. *J. Mass Spectrom.* **37,** 41–46 (2002).
33. Tabuchi, T., Nakamura, I., Higashi, E. & Hideyuki, K. Factors Affecting the Production of the Open-ring Acid of Spiculisporic Acid by *Penicillium spiculisporum*. *J. Ferment. Technol.* **55,** 43–49 (1977).
34. Rubinovitz, C., Gutnick, D. L. & Rosenberg, E. Emulsan production by *Acinetobacter calcoaceticus* in the presence of chloramphenicol. *J. Bacteriol.* **152,** 126–132 (1982).
35. Sarubbo, L. A., Farias, C. B. B. & Campos-Takaki, G. M. Co-utilization of canola oil and glucose on the production of a surfactant by *Candida lipolytica*. *Curr. Microbiol.* **54,** 68–73 (2007).
36. Navon-Venezia, S. *et al.* Alasan, a new bioemulsifier from *Acinetobacter radioresistens*. *Appl. Environ. Microbiol.* **61,** 3240–3244 (1995).
37. Shabtai, Y. Production of exopolysaccharides by *Acinetobacter* strains in a controlled fed-batch fermentation process using soap stock oil (SSO) as carbon source. *Int. J. Biol. Macromol.* **12,** 145–152 (1990).
38. Ron, E. Z. & Rosenberg, E. Natural roles of biosurfactants. *Environ. Microbiol.* **3,** 229–236 (2001).
39. Van Bogaert, I. N. A. *et al.* The biosynthetic gene cluster for sophorolipids: A biotechnological interesting biosurfactant produced by *Starmerella bombicola*. *Mol. Microbiol.* **88,** 501–509 (2013).
40. Zulianello, L. *et al.* Rhamnolipids are virulence factors that promote early infiltration of primary human airway epithelia by *Pseudomonas aeruginosa*. *Infect. Immun.* **74,** 3134–3147 (2006).
41. Randhawa, K. K. S. & Rahman, P. K. S. M. Rhamnolipid biosurfactants – Past , Present and future scenario of global market. *Front. Microbiol.* **5,** 1–7 (2014).
42. George, S. & Jayachandran, K. Production and characterization of rhamnolipid biosurfactant from waste frying coconut oil using a novel *Pseudomonas aeruginosa* D. *J. Appl. Microbiol.* **114,** 373–383 (2013).
43. Coutte, F. *et al.* New integrated bioprocess for the continuous production, extraction and purification of lipopeptides produced by *Bacillus subtilis* in membrane bioreactor. *Process Biochem.* **48,** 25–32 (2013).
44. Abdel-Mawgoud, A. M., Lépine, F. & Déziel, E. Rhamnolipids: Diversity of structures, microbial origins and roles. *Appl. Microbiol. Biotechnol.* **86,** 1323–1336 (2010).
45. Giani, C., Meiwes, J., Rothert, R. & Wullbrandt, D. *Pseudomonas aeruginosa* and its use in a process for the biotechnological preparation of L-rhamnose. 6 (1995).
46. Hoermann, B., Müller, M. M., Syldatk, C. & Hausmann, R. Rhamnolipid Production by *Burkholderia plantarii* DSM 9509T. *Chemie Ing. Tech.* **82,** 1593–1594 (2010).

47. Nitschke, M., Costa, S. G. V. A. O. & Contiero, J. Rhamnolipid surfactants: An update on the general aspects of these remarkable biomolecules. *Biotechnol. Prog.* **21,** 1593–1600 (2005).
48. Mysels, K. J. Surface tension of solutions of pure sodium dodecyl sulfate. *Langmuir* **2,** 423–428 (1986).
49. Adamczyk, Z., Para, G. & Warszyński, P. Influence of ionic strength on surface tension of cetyltrimethylammonium bromide. *Langmuir* **15,** 8383–8387 (1999).
50. Ochoa-Loza, F. J., Artiola, J. F. & Maier, R. M. Stability constants for the complexation of various metals with a rhamnolipid biosurfactant. *J. Environ. Qual.* **30,** 479–485 (1992).
51. Aşçı, Y. B., Nurbaş, M. & Açıkel, Y. S. Sorption of Cd(II) onto kaolin as a soil component and desorption of Cd(II) from kaolin using rhamnolipid biosurfactant. *J. Hazard. Mater.* **139,** 50–56 (2007).
52. Whang, L. M., Liu, P. W. G., Ma, C. C. & Cheng, S. S. Application of biosurfactants, rhamnolipid, and surfactin, for enhanced biodegradation of diesel-contaminated water and soil. *J. Hazard. Mater.* **151,** 155–163 (2008).
53. Chrzanowski, Ł. *et al.* Interactions between rhamnolipid biosurfactants and toxic chlorinated phenols enhance biodegradation of a model hydrocarbon-rich effluent. *Int. Biodeterior. Biodegrad.* **65,** 605–611 (2011).
54. Shin, K. H., Kim, K. W. & Ahn, Y. Use of biosurfactant to remediate phenanthrene-contaminated soil by the combined solubilization-biodegradation process. *J. Hazard. Mater.* **137,** 1831–1837 (2006).
55. Zhao, F. *et al.* Production of rhamnolipids by *Pseudomonas aeruginosa* is inhibited by H2S but resumes in a co-culture with *P. stutzeri*: applications for microbial enhanced oil recovery. *Biotechnol. Lett.* **37,** 1803–1808 (2015).
56. Leibundgut, M., Maier, T., Jenni, S. & Ban, N. The multienzyme architecture of eukaryotic fatty acid synthases. *Curr. Opin. Struct. Biol.* **18,** 714–725 (2008).
57. Kutchma, A. J., Hoang, T. T. & Schweizer, H. P. Characterization of a *Pseudomonas aeruginosa* fatty acid biosynthetic gene cluster: Purification of acyl carrier protein (ACP) and malonyl-coenzyme A:ACP transacylase (FabD). *J. Bacteriol.* **181,** 5498–5504 (1999).
58. Rawlings, M. & Cronan, J. E. J. Escherichia coli Acyl Carrier Protein Lies within a Cluster of Fatty Acid Biosynthetic Genes. *J. Biol. Chem.* **267,** 5751–5754 (1992).
59. Shen, Z. & Byers, D. M. Isolation of *Vibrio harveyi* acyl carrier protein and the *fabG, acpP,* and *fabF* genes involved in fatty acid biosynthesis. *J. Bacteriol.* **178,** 571–573 (1996).
60. Morbidoni, H. R., DeMendoza, D. & Cronan, J. E. J. *Bacillus subtilis* acyl carrier protein is encoded in a cluster of lipid biosynthesis genes. *J. Bacteriol.* **178,** 4794–4800 (1996).
61. Raychaudhuri, A., Jerga, A. & Tipton, P. A. Chemical mechanism and substrate specificity of RhlI, an acylhomoserine lactone synthase from *Pseudomonas aeruginosa*. *Biochemistry* **44,** 2974–2981 (2005).
62. Finking, R. *et al.* Characterization of a new type of phosphopantetheinyl transferase for fatty acid and siderophore synthesis in *Pseudomonas aeruginosa*. *J. Biol. Chem.* **277,** 50293–50302 (2002).
63. Beld, J., Cang, H. & Burkart, M. D. Visualizing the chain-flipping mechanism in fatty-acid biosynthesis. *Angew. Chemie - Int. Ed.* **53,** 14456–14461 (2014).
64. Rock, C. O. & Cronan, J. E. J. Re-evaluation of the solution structure of acyl carrier protein. *J. Biol. Chem.* **254,** 9778–9785 (1979).
65. Cronan, J. E. J. Molecular properties of short chain acyl thioesters of acyl carrier protein. *J. Biol. Chem.* **257,** 5013–5017 (1982).
66. Cronan, J. E. J. The chain-flipping mechanism of ACP (acyl carrier protein)-dependent enzymes appears universal. *Biochem. J.* **460,** 157–163 (2014).
67. Masoudi, A., Raetz, C. R. H., Zhou, P. & Pemble, C. W. Chasing acyl carrier protein through a catalytic cycle of lipid A production. *Nature* **505,** 422–426 (2014).
68. Zhang, Y.-M. *et al.* Identification and Analysis of the Acyl Carrier Protein (ACP) Docking Site on Beta-Ketoacyl-ACP Synthase III. *J. Biol. Chem.* **276,** 8231–8238 (2001).
69. Zhang, Y.-M., Wu, B., Zheng, J. & Rock, C. O. Key Residues Responsible for Acyl Carrier Protein and β-Ketoacyl-Acyl Carrier Protein Reductase (FabG) Interaction. *J. Biol. Chem.* **278,** 52935–52943 (2003).
70. Davis, M. S., Solbiati, J. & Cronan, J. E. J. Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in *Escherichia coli*. *J. Biol. Chem.* **275,** 28593–28598 (2000).

71. Li, S.-J. & Cronan, J. E. J. The genes encoding the two carboxyltransferase subunits of *Escherichia coli* Acetyl-CoA carboxylase. *J. Biol. Chem.* **267,** 16841–16847 (1992).
72. Cronan, J. E. J. & Li, S.-J. The Gene Encoding the Biotin Caroxylase Subunit of *Escherichia coli* Acetyl-CoA Carboxylase. *J. Biol. Chem.* **267,** 855–863 (1992).
73. Guchhait, R. B. *et al.* Acetyl Coenzyme A Carboxylase System of *Escherichia coli*. *J. Biol. Chem.* **249,** 6633–6645 (1974).
74. Polakis, S. E., Guchhait, R. B., Zwergel, E. E., Lane, M. D. & Cooper, T. G. Acetyl Coenzyme A Carboxylase system of *Escherichia coli*. Studies on the mechanisms of the biotin caboxylase- and carboxyltransferase-catalyzed reactions. *J. Biol. Chem.* **249,** 6657–6667 (1974).
75. Oefner, C., Schulz, H., D'Arcy, A. & Dale, G. E. Mapping the active site of *Escherichia coli* malonyl-CoA-acyl carrier protein transacylase (FabD) by protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62,** 613–618 (2006).
76. Brenner, S. The molecular evolution of genes and proteins: a tale of two serines. *Nature* **334,** 528–530 (1988).
77. Ruch, F. E. & Vagelos, P. R. Characterization of a malonyl enzyme intermediate and identification of the malonyl binding site in malonyl coenzyme A acyl carrier protein transacylase of *Escherichia coli*. *J. Biol. Chem.* **248,** 8095–8106 (1973).
78. Tsay, J. T., Oh, W., Larson, T. J., Jackowski, S. & Rock, C. O. Isolation and characterization of the beta-ketoacyl-acyl carrier protein synthase III gene (*fabH*) from *Escherichia coli* K-12. *J. Biol. Chem.* **267,** 6807–6814 (1992).
79. Davies, C., Heath, R. J., White, S. W. & Rock, C. O. The 1.8 Angström crystal structure and active-site architecture of β-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli*. *Structure* **8,** 185–195 (2000).
80. Heath, R. J. & Rock, C. O. Inhibition of β-ketoacyl-acyl carrier protein synthase III (FabH) by acyl-acyl carrier protein in *Escherichia coli*. *J. Biol. Chem.* **271,** 10996–11000 (1996).
81. Heath, R. J. & Rock, C. O. Regulation of fatty acid elongation and initiation by acyl-acyl carrier protein in *Escherichia coli*. *J. Biol. Chem.* **271,** 1833–1836 (1996).
82. Rock, C. O. & Jackowski, S. Regulation of phospholipid synthesis in *Escherichia coli*. Composition of the acyl-acyl carrier protein pool *in vivo*. *J. Biol. Chem.* **257,** 10759–10765 (1982).
83. Cho, H. & Cronan, J. E. J. Defective export of a periplasmic enzyme disrupts regulation of fatty acid synthesis. *J. Biol. Chem.* **270,** 4216–4219 (1995).
84. Jiang, P. & Cronan, J. E. J. Inhibition of fatty acid synthesis in *Escherichia col* in the absence of phospholipid synthesis and release of inhibition by thioesterase action. *J. Bacteriol.* **176,** 2814–2821 (1994).
85. Choi, K.-H., Heath, R. J. & Rock, C. O. β-ketoacyl-acyl carrier protein synthase III (FabH) is a determining factor in branched-chain fatty acid biosynthesis. *J. Bacteriol.* **182,** 365–370 (2000).
86. Zhang, Y.-M. & Rock, C. O. Will the initiator of fatty acid synthesis in *Pseudomonas aeruginosa* please stand up? *J. Bacteriol.* **194,** 5159–5161 (2012).
87. Yuan, Y., Leeds, J. A. & Meredith, T. C. *Pseudomonas aeruginosa* directly shunts beta-oxidation degradation intermediates into de novo fatty acid biosynthesis. *J. Bacteriol.* **194,** 5185–5196 (2012).
88. Yuan, Y., Sachdeva, M., Leeds, J. A. & Meredith, T. C. Fatty acid biosynthesis in *Pseudomonas aeruginosa* is initiated by the FabY class of β-ketoacyl acyl carrier protein synthases. *J. Bacteriol.* **194,** 5171–5184 (2012).
89. Zhang, Y. & Cronan, J. E. J. Transcriptional analysis of essential genes of the *Escherichia coli* fatty acid biosynthesis gene cluster by functional replacement with the analogous *Salmonella typhimurium* gene cluster. *J. Bacteriol.* **180,** 3295–3303 (1998).
90. Campos-Garcia, J. *et al.* The Pseudomonas aeruginosa *rhlG* gene encodes an NADPH-dependent β-ketoacyl reductase which is specifically involved in rhamnolipid synthesis. *J. Bacteriol.* **180,** 4442–4451 (1998).
91. Miller, D. J., Zhang, Y.-M., Rock, C. O. & White, S. W. Structure of RhlG, an essential β-ketoacyl reductase in the rhamnolipid biosynthetic pathway of *Pseudomonas aeruginosa*. *J. Biol. Chem.* **281,** 18025–18032 (2006).
92. Kass, L. R. & Bloch, K. On the enzymatic synthesis of unsaturated fatty acids in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **58,** 1168–1173 (1967).

93. Heath, R. J. & Rock, C. O. Roles of the FabA and FabZ β-hydroxyacyl-acyl carrier protein dehydratases in *Escherichia coli* fatty acid biosynthesis. *J. Biol. Chem.* **271,** 27795–27801 (1996).
94. Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M. & Smith, J. L. Structure of a dehydratase–isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure* **4,** 253–264 (1996).
95. Guerra, D. J. & Browse, J. A. *Escherichia coli* β-hydroxydecanoyl thioester dehydrase reacts with native C10 acyl-acyl-carrier proteins of plant and bacterial origin. *Arch. Biochem. Biophys.* **280,** 336–345 (1990).
96. Heath, R. J. & Rock, C. O. Enoyl-acyl carrier protein reductase (*fabI*) plays a determinant role in completing cycles of fatty acid elongation in *Escherichia coli*. *J. Biol. Chem.* **270,** 26538–26542 (1995).
97. Schwab, J. M., Klassen, J. B. & Lin, D. C. T. β-Hydroxydecanoylthioester dehydrase: A rapid, convenient, and accurate product distribution assay. *Anal. Biochem.* **150,** 121–124 (1985).
98. Hoang, T. T. & Schweizer, H. P. Fatty acid biosynthesis in *Pseudomonas aeruginosa*: Cloning and characterization of the *fabAB* operon encoding β-hydroxyacyl-acyl carrier protein dehydratase (FabA) and β-ketoacyl-acyl carrier protein synthase I (FabB). *J. Bacteriol.* **179,** 5326–5332 (1997).
99. Clark, D. P., DeMendoza, D., Polacco, M. L. & Cronan, J. E. J. Beta-hydroxydecanoyl thio ester dehydrase does not catalyze a rate-limiting step in *Escherichia coli* unsaturated fatty acid synthesis. *Biochemistry* **22,** 5897–5902 (1983).
100. Sharma, S. K. *et al.* Identification, Characterization, and Inhibition of *Plasmodium falciparum* β-Hydroxyacyl-Acyl Carrier Protein Dehydratase (FabZ). *J. Biol. Chem.* **278,** 45661–45671 (2003).
101. Liu, W. *et al.* A new β-hydroxyacyl-acyl carrier protein dehydratase (FabZ) from *Helicobacter pylori*: Molecular cloning, enzymatic characterization, and structural modeling. *Biochem. Biophys. Res. Commun.* **333,** 1078–1086 (2005).
102. Kimber, M. S. *et al.* The structure of (3R)-hydroxyacyl-acyl carrier protein dehydratase (FabZ) from *Pseudomonas aeruginosa*. *J. Biol. Chem.* **279,** 52593–52602 (2004).
103. Waag, H. & Cronan, J. E. J. Functional replacement of the FabA and FabB proteins of *Escherichia coli* fatty acid synthesis by *Enterococcus faecalis* FabZ and FabF homologues. *J. Biol. Chem.* **279,** 34489–34495 (2004).
104. Marrakchi, H., Choi, K.-H. & Rock, C. O. A New Mechanism for Anaerobic Unsaturated Fatty Acid Formation in *Streptococcus pneumoniae*. *J. Biol. Chem.* **277,** 44809–44816 (2002).
105. Turnowsky, F., Fuchs, K., Jeschek, C. & Hogenauer, G. *envM* genes of *Salmonella typhimurium* and *Escherichia coli*. *J. Bacteriol.* **171,** 6555–6565 (1989).
106. Bergler, H., Fuchsbichler, S., Högenauer, G. & Turnowsky, F. The enoyl-[acyl-carrier-protein] reductase (FabI) of *Escherichia coli*, which catalyzes a key regulatory step in fatty acid biosynthesis, accepts NADH and NADPH as cofactors and is inhibited by palmitoyl-CoA. *Eur. J. Biochem.* **242,** 689–694 (1996).
107. Heath, R. J. & Rock, C. O. A triclosan-resistant bacterial enzyme. *Nature* **406,** 145–146 (2000).
108. Zhu, L., Lin, J., Ma, J., Cronan, J. E. J. & Wang, H. Triclosan resistance of *Pseudomonas aeruginosa* PAO1 is due to FabV, a triclosan-resistant enoyl-acyl carrier protein reductase. *Antimicrob. Agents Chemother.* **54,** 689–698 (2010).
109. Heath, R. J., Su, N., Murphy, C. K. & Rock, C. O. The enoyl-[acyl-carrier-protein] reductases FabI and FabL from *Bacillus subtilis*. *J. Biol. Chem.* **275,** 40128–40133 (2000).
110. Massengo-Tiassé, R. P. & Cronan, J. E. J. Diversity in enoyl-acyl carrier protein reductases. *Cell. Mol. Life Sci.* **66,** 1507–1517 (2009).
111. Marrakchi, H. *et al.* Characterization of *Streptococcus pneumoniae* enoyl-(acyl-carrier protein) reductase (FabK). *Biochem. J.* **370,** 1055–1062 (2003).
112. Massengo-Tiassé, R. P. & Cronan, J. E. J. *Vibrio cholerae* FabV defines a new class of enoyl-acyl carrier protein reductase. *J. Biol. Chem.* **283,** 1308–1316 (2008).
113. Zheng, C. J. *et al.* Cephalochromin, a FabI-directed antibacterial of microbial origin. *Biochem. Biophys. Res. Commun.* **362,** 1107–1112 (2007).
114. Zheng, C.-J., Sohn, M.-J. & Kim, W.-G. Atromentin and leucomelone, the first inhibitors specific to enoyl-ACP reductase (FabK) of *Streptococcus pneumoniae*. *The Journal of Antibiotics* **59,** 808–812 (2006).
115. Rosenfeld, I. S., D'Agnolo, G. & Vagelos, P. R. Synthesis of unsaturated fatty acids and the lesion in fab B mutants. *J. Biol. Chem.* **248,** 2452–2460 (1973).

116. Garwin, J. L., Klages, A. L. & Cronan, J. E. J. Structural, enzymatic, and genetic studies of β-ketoacyl-acyl carrier protein synthases I and II of *Escherichia coli*. *J. Biol. Chem.* **255,** 11949–11956 (1980).
117. Jackowski, S., Murphy, C. M., Cronan, J. E. J. & Rock, C. O. Acetoacetyl-acyl carrier protein synthase. A target for the antibiotic thiolactomycin. *J. Biol. Chem.* **264,** 7624–7629 (1989).
118. Price, A. C., Zhang, Y.-M., Rock, C. O. & White, S. W. Structure of β-ketoacyl-[acyl carrier protein] reductase from *Escherichia coli*: Negative cooperativity and its structural basis. *Biochemistry* **40,** 12772–12781 (2001).
119. Marumo, K. *et al.* Enzymatic synthesis and isolation of thymidine diphosphate-6-deoxy-D-xylo-4-hexulose and thymidine diphosphate-L-rhamnose. Production using cloned gene products and separation by HPLC. *Eur. J. Biochem.* **204,** 539–545 (1992).
120. Melo, A. & Glaser, L. The mechanism of 6-deoxyhexose synthesis. II. Conversion of deoxythymidine diphosphate 4-keto-6-deoxy-D-glucose to deoxythymidine diphosphate L-rhamnose. *J. Biol. Chem.* **243,** 1475–1478 (1968).
121. Melo, A., Elliott, W. H. & Glaser, L. The mechanism of 6-deoxyhexose synthesis. I. Intramolecular hydrogen transfer catalyzed by deoxythymidine diphosphate D-glucose oxidoreductase. *J. Biol. Chem.* **243,** 1467–1474 (1968).
122. Marolda, C. L. & Valvano, M. A. Genetic analysis of the dTDP-rhamnose biosynthesis region of the *Escherichia coli* VW187 (O7:K1) *rfb* gene cluster: Identification of functional homologs of *rfbB* and *rfbA* in the *rff* cluster and correct location of the *rffE* gene. *J. Bacteriol.* **177,** 5539–5546 (1995).
123. Ma, Y. *et al.* Determination of the pathway for rhamnose biosynthesis in mycobacteria: Cloning, sequencing and expression of the *Mycobacterium tuberculosis* gene encoding α-D-glucose-1-phosphate thymidylyltransferase. *Microbiology* **143,** 937–945 (1997).
124. Lindqvist, L., Kaiser, R., Reeves, P. R. & Lindberg, A. A. Purification, characterization and HPLC assay of *Salmonella* glucose-1-phosphate thymidylyltransferase from the cloned *rfbA* gene. *Eur. J. Biochem.* **211,** 763–770 (1993).
125. Blankenfeldt, W., Asuncion, M., Lam, J. S. & Naismith, J. H. The structural basis of the catalytic mechanism and regulation of glucose-1-phosphate thymidylyltransferase (RmlA). *EMBO J.* **19,** 6652–6663 (2000).
126. Melo, A. & Glaser, L. The Nucleotide Specificity and Feedback Control of Thymidine. *J. Biol. Chem.* **240,** 398–405 (1965).
127. Allard, S. T. M. *et al.* The crystal structure of dTDP-D-Glucose 4,6-dehydratase (RmlB) from *Salmonella enterica* serovar *typhimurium*, the second enzyme in the dTDP-l-rhamnose pathway. *J. Mol. Biol.* **307,** 283–295 (2001).
128. Gross, J. W., Hegeman, A. D., Vestling, M. M. & Frey, P. A. Characterization of enzymatic processes by rapid mix-quench mass spectrometry: The case of dTDP-glucose 4,6-dehydratase. *Biochemistry* **39,** 13633–13640 (2000).
129. Giraud, M. F., Leonard, G. A., Field, R. A., Berlind, C. & Naismith, J. H. RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase. *Nat. Struct. Mol. Biol.* **7,** 398–402 (2000).
130. Graninger, M., Nidetzky, B., Heinrichs, D. E., Whitfield, C. & Messner, P. Characterization of dTDP-4-dehydrorhamnose 3,5-epimerase and dTDP-4- dehydrorhamnose reductase, required for dTDP-L-rhamnose biosynthesis in *Salmonella enterica* serovar *typhimurium* LT2. *J. Biol. Chem.* **274,** 25069–25077 (1999).
131. Ochsner, U. A., Fiechter, A. & Reiser, J. Isolation, characterization, and expression in *Escherichia coli* of the *Pseudomonas aeruginosa rhlAB* genes encoding a rhamnosyltransferase involved in rhamnolipid biosurfactant synthesis. *J. Biol. Chem.* **269,** 19787–19795 (1994).
132. Zhu, K. & Rock, C. O. RhlA converts Beta-hydroxyacyl-acyl carrier protein intermediates in fatty acid synthesis to the Beta-hydroxydecanoyl-Beta-hydroxydecanoate component of rhamnolipids in *Pseudomonas aeruginosa*. *J. Bacteriol.* **190,** 3147–3154 (2008).
133. Lotfabad, T. B. *et al.* Structural characterization of a rhamnolipid-type biosurfactant produced by *Pseudomonas aeruginosa* MR01: Enhancement of di-rhamnolipid proportion using gamma irradiation. *Colloids Surfaces B Biointerfaces* **81,** 397–405 (2010).
134. Costa, S. G. V. A. O., Déziel, E. & Lépine, F. Characterization of rhamnolipid production by *Burkholderia glumae*. *Lett. Appl. Microbiol.* **53,** 620–627 (2011).
135. Rahim, R. *et al.* Cloning and functional characterization of the *Pseudomonas aeruginosa rhlC* gene that encodes rhamnosyltransferase 2, an enzyme responsible for di-rhamnolipid biosynthesis. *Mol. Microbiol.* **40,** 708–718 (2001).

136. Glessner, A., Smith, R. S., Iglewski, B. H. & Robinson, J. B. Roles of *Pseudomonas aeruginosa las* and *rhl* quorum-sensing systems in control of twitching motility. *J. Bacteriol.* **181,** 1623–1629 (1999).
137. Pesci, E. C. *et al.* Regulation of las and rhl quorum sensing in *Pseudomonas aeruginosa*. *J. Bacteriol.* **179,** 3127–3132 (1997).
138. Reis, R. S., Pereira, A. G., Neves, B. C. & Freire, D. M. G. Gene regulation of rhamnolipid production in *Pseudomonas aeruginosa* - A review. *Bioresour. Technol.* **102,** 6377–6384 (2011).
139. Fuqua, W. C., Winans, S. C. & Greenberg, E. P. Quorum sensing in bacteria: The LuxR-LuxI family of cell density- responsive transcriptional regulators. *J. Bacteriol.* **176,** 269–275 (1994).
140. Medina, G., Juárez, K., Valderrama, B. & Soberón-Chávez, G. Mechanism of *Pseudomonas aeruginosa* RhlR Transcriptional Regulation of the *rhlAB* Promoter. *J. Bacteriol.* **185,** 5976–5983 (2003).
141. Dekimpe, V. & Déziel, E. Revisiting the quorum-sensing hierarchy in *Pseudomonas aeruginosa*: The transcriptional regulator RhlR regulates LasR-specific factors. *Microbiology* **155,** 712–723 (2009).
142. Lizewski, S. E., Lundberg, D. S. & Schurr, M. J. The transcriptional regulator AlgR is essential for *Pseudomonas aeruginosa* pathogenesis. *Infect. Immun.* **70,** 6083–6093 (2002).
143. Okkotsu, Y., Tieku, P., Fitzsimmons, L. F., Churchill, M. E. & Schurr, M. J. *Pseudomonas aeruginosa* AlgR phosphorylation modulates rhamnolipid production and motility. *J. Bacteriol.* **195,** 5499–5515 (2013).
144. Morici, L. A. *et al. Pseudomonas aeruginosa* AlgR represses the Rhl quorum-sensing system in a biofilm-specific manner. *J. Bacteriol.* **189,** 7752–7764 (2007).
145. Carty, N. L. *et al.* PtxR modulates the expression of QS-controlled virulence factors in the *Pseudomonas aeruginosa* strain PAO1. *Mol. Microbiol.* **61,** 782–794 (2006).
146. Rampioni, G., Schuster, M., Greenberg, E. P., Zennaro, E. & Leoni, L. Contribution of the RsaL global regulator to *Pseudomonas aeruginosa* virulence and biofilm formation. *FEMS Microbiol. Lett.* **301,** 210–217 (2009).
147. Kahraman, H. & Erenler, S. O. Rhamnolipid production by Pseudomonas aeruginosa engineered with the *Vitreoscilla* hemoglobin gene. *Appl. Biochem. Microbiol.* **48,** 188–193 (2012).
148. Wittgens, A. *et al.* Growth independent rhamnolipid production from glucose using the non-pathogenic *Pseudomonas putida* KT2440. *Microb. Cell Fact.* **10,** 80–96 (2011).
149. Ochsner, U. A., Reiser, J. & Fiechter, A. Production of *Pseudomonas aeruginosa* Rhamnolipid Biosurfactants in Heterologous Hosts. *Appl. Environ. Microbiol.* **61,** 3503–3506 (1995).
150. Cha, M., Lee, N., Kim, M., Kim, M. & Lee, S. Heterologous production of *Pseudomonas aeruginosa* EMS1 biosurfactant in *Pseudomonas putida*. *Bioresour. Technol.* **99,** 2192–2199 (2008).
151. Cabrera-Valladares, N. *et al.* Monorhamnolipids and 3-(3-hydroxyalkanoyloxy) alkanoic acids (HAAs) production using Escherichia coli as a heterologous host. *Appl. Microbiol. Biotechnol.* **73,** 187–194 (2006).
152. Wang, Q. *et al.* Engineering bacteria for production of rhamnolipid as an agent for enhanced oil recovery. *Biotechnol. Bioeng.* **98,** 842–853 (2007).
153. Tavares, L. F. D. *et al.* Characterization of rhamnolipids produced by wild-type and engineered *Burkholderia kururiensis*. *Appl. Microbiol. Biotechnol.* **97,** 1909–1921 (2013).
154. Daniel, H. J., Reuss, M. & Syldatk, C. Production of sophorolipids in high concentration from deproteinized whey and rapeseed oil in a two stage fed batch process using *Candida bombicola* ATCC 22214 and *Cryptococcus curvatus* ATCC 20509. *Biotechnol. Lett.* **20,** 1153–1156 (1998).
155. Chen, J., Song, X., Zhang, H. & Qu, Y. Production, structure elucidation and anticancer properties of sophorolipid from *Wickerhamiella domercqiae*. *Enzyme Microb. Technol.* **39,** 501–506 (2006).
156. Asmer, H. J., Lang, S., Wagner, F. & Wray, V. Microbial production, structure elucidation and bioconversion of sophorose lipids. *J. Am. Oil Chem. Soc.* **65,** 1460–1466 (1988).
157. Ciesielska, K. *et al.* Exoproteome analysis of *Starmerella bombicola* results in the discovery of an esterase required for lactonization of sophorolipids. *J. Proteomics* **98,** 159–174 (2014).
158. Ashby, R. D., Solaiman, D. K. Y. & Foglia, T. A. Property control of sophorolipids: Influence of fatty acid substrate and blending. *Biotechnol. Lett.* **30,** 1093–1100 (2008).

159. Schippers, C., Geßner, K., Müller, T. & Scheper, T. Microbial degradation of phenanthrene by addition of a sophorolipid mixture. *J. Biotechnol.* **83,** 189–198 (2000).
160. Kang, S. W., Kim, Y. B., Shin, J. D. & Kim, E.-K. Enhanced biodegradation of hydrocarbons in soil by microbial biosurfactant, sophorolipid. *Appl. Biochem. Biotechnol.* **160,** 780–790 (2010).
161. Hirata, Y. *et al.* Novel characteristics of sophorolipids, yeast glycolipid biosurfactants, as biodegradable low-foaming surfactants. *J. Biosci. Bioeng.* **108,** 142–146 (2009).
162. Imura, T. *et al.* Spontaneous vesicle formation from sodium salt of acidic sophorolipid and its application as a skin penetration enhancer. *J. Oleo Sci.* **63,** 141–147 (2014).
163. Valotteau, C. *et al.* Biocidal Properties of a Glycosylated Surface: Sophorolipids on Au(111). *ACS Appl. Mater. Interfaces* **7,** 18086–18095 (2015).
164. Fu, S. L. *et al.* Sophorolipids and Their Derivatives Are Lethal Against Human Pancreatic Cancer Cells. *J. Surg. Res.* **148,** 77–82 (2008).
165. Hagler, M. D. *et al.* Sophorolipids Decrease IgE Production in U266 Cells by Downregulation of BSAP (Pax5), TLR-2, STAT3 and IL-6. *J. Allergy Clin. Immunol.* **119,** S263 (2007).
166. Baccile, N., Babonneau, F., Jestin, J., Pehau-Arnaudet, G. & Van Bogaert, I. N. A. Unusual, pH-induced, self-assembly of sophorolipid biosurfactants. *ACS Nano* **6,** 4763–4776 (2012).
167. Zhou, S. *et al.* Supramolecular assemblies of a naturally derived sophorolipid. *Langmuir* **20,** 7926–7932 (2004).
168. Penfold, J. *et al.* Solution self-assembly of the sophorolipid biosurfactant and its mixture with anionic surfactant sodium dodecyl benzene sulfonate. *Langmuir* **27,** 8867–8877 (2011).
169. Singh, S. *et al.* A direct method for the preparation of glycolipid–metal nanoparticle conjugates: sophorolipids as reducing and capping agents for the synthesis of water re-dispersible silver nanoparticles and their antibacterial activity. *New J. Chem.* **33,** 646–652 (2009).
170. Al-Jasim, A. *et al.* Isolation of sophorose during sophorolipid production and studies of its stability in aqueous alkali: epimerisation of sophorose to 2-O-β-d-glucopyranosyl-d-mannose. *Carbohydr. Res.* **421,** 46–54 (2016).
171. Zerkowski, J. A., Solaiman, D. K. Y., Ashby, R. D. & Foglia, T. A. Head group-modified sophorolipids: Synthesis of new cationic, zwitterionic, and anionic surfactants. *J. Surfactants Deterg.* **9,** 57–62 (2006).
172. Ciesielska, K. *et al.* SILAC-based proteome analysis of *Starmerella bombicola* sophorolipid production. *J. Proteome Res.* **12,** 4376–4392 (2013).
173. Albrecht, A., Rau, U. & Wagner, F. Initial steps of sophoroselipid biosynthesis by *Candida bombicola* ATCC 22214 grown on glucose. *Appl. Microbiol. Biotechnol.* **46,** 67–73 (1996).
174. Roelants, S. L. K. W. *et al.* Towards the industrialization of new biosurfactants: Biotechnological opportunities for the lactone esterase gene from *Starmerella bombicola*. *Biotechnol. Bioeng.* **113,** 550–559 (2015).
175. Saerens, K. M. J., Roelants, S. L. K. W., Van Bogaert, I. N. A. & Soetaert, W. 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 (2011).
176. Saerens, K. M. J., Van Bogaert, I. N. A. & Soetaert, W. Characterization of sophorolipid biosynthetic enzymes from *Starmerella bombicola*. *FEMS Yeast Res.* **15,** fov075 (2015).
177. Saerens, K. M. J., Saey, L. & Soetaert, W. One-step production of unacetylated sophorolipids by an acetyltransferase negative *Candida bombicola*. *Biotechnol. Bioeng.* **108,** 2923–2931 (2011).
178. Kirk, O. & Christensen, M. W. Lipases from *Candida antarctica*: Unique biocatalysts from a unique origin. *Org. Process Res. Dev.* **6,** 446–451 (2002).
179. Price, N. P. J. J., Ray, K. J., Vermillion, K. E., Dunlap, C. A. & Kurtzman, C. P. Structural characterization of novel sophorolipid biosurfactants from a newly identified species of *Candida* yeast. *Carbohydr. Res.* **348,** 33–41 (2012).
180. Tulloch, A. P., Spencer, J. F. T. & Deinema, M. H. A new hydroxy fatty acid sophoroside from *Candida bogoriensis*. *Can. J. Chem.* **46,** 345–348 (1968).
181. Nuez, A., Ashby, R. D., Foglia, T. A. & Solaiman, D. K. Y. LC/MS analysis and lipase modification of the sophorolipids produced by *Rhodotorula bogoriensis*. *Biotechnol. Lett.* **26,** 1087–1093 (2004).
182. Guengerich, F. P. & Munro, A. W. Unusual cytochrome P450 enzymes and reactions. *J. Biol. Chem.* **288,** 17065–17073 (2013).

183. Omura, T. & Sato, R. Fractional solubilization of haemoproteins and partial purification of carbon monoxide-binding cytochrome from liver microsomes. *Biochim. Biophys. Acta* **71,** 224–226 (1963).
184. Omura, T. & Sato, R. The Carbon Monoxide-biding pigment of Liver Microsomes. *J. Biol. Chem.* **239,** 2370–2378 (1964).
185. Omura, T. & Sato, R. The Carbon Monoxide-Binding Pigment of Liver Microsomes. I. Evidence For Its Hemoprotein Nature. *J. Biol. Chem.* **239,** 2370–2378 (1964).
186. Nelson, D. R. The cytochrome p450 homepage. *Hum. Genomics* **4,** 59–65 (2009).
187. Nelson, D. R. *et al.* The P450 Superfamily: Update on New Sequences, Gene Mapping, Accession Numbers, Early Trivial Names of Enzymes, and Nomenclature. *DNA Cell Biol.* **12,** 1–51 (1993).
188. Nelson, D. R. *et al.* P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* **6,** 1–42 (1996).
189. Hasemann, C. A., Kurumbail, R. G., Boddupalli, S. S., Peterson, J. A. & Deisenhofer, J. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure* **3,** 41–62 (1995).
190. Mestres, J. Structure conservation in cytochromes P450. *Proteins Struct. Funct. Genet.* **58,** 596–609 (2005).
191. Presnell, S. R. & Cohen, F. E. Topological distribution of four-alpha-helix bundles. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 6592–6596 (1989).
192. Gotoh, O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J. Biol. Chem.* **267,** 83–90 (1992).
193. Ost, T. W. B. *et al.* Phenylalanine 393 exerts thermodynamic control over the heme of flavocytochrome P450 BM3. *Biochemistry* **40,** 13421–13429 (2001).
194. Ost, T. W. B. *et al.* Structural and spectroscopic analysis of the F393H mutant of flavocytochrome P450 BM3. *Biochemistry* **40,** 13430–13438 (2001).
195. Chen, Z., Ost, T. W. B. & Schelvis, J. P. M. Phe393 Mutants of Cytochrome P450 BM3 with Modified Heme Redox Potentials Have Altered Heme Vinyl and Propionate Conformations. *Biochemistry* **43,** 1798–1808 (2004).
196. Adak, S. *et al.* Tryptophan 409 controls the activity of neuronal nitric-oxide synthase by regulating nitric oxide feedback inhibition. *J. Biol. Chem.* **274,** 26907–26911 (1999).
197. Wilson, D. J. & Rafferty, S. P. A structural role for tryptophan 188 of inducible nitric oxide synthase. *Biochem. Biophys. Res. Commun.* **287,** 126–129 (2001).
198. Voegtle, H. L. *et al.* Spectroscopic characterization of five- and six-coordinate ferrous-NO heme complexes. Evidence for heme Fe-proximal cysteinate bond cleavage in the ferrous-NO adducts of the Trp-409Tyr/Phe proximal environment mutants of neuronal nitric oxide synthase. *Biochemistry* **42,** 2475–2484 (2003).
199. Martinis, S. A., Atkins, W. M., Stayton, P. S. & Sligar, S. G. A Conserved Residue of Cytochrome P-450 Is Involved in Heme-Oxygen Stability and Activation. *J. Am. Chem. Soc.* **111,** 9252–9253 (1989).
200. Kimata, Y., Shimada, H., Hirose, T. & Ishimura, Y. Role of Thr-252 in cytochrome P450cam: a study with unnatural amino acid mutagenesis. *Biochem. Biophys. Res. Commun.* **208,** 96–102 (1995).
201. Vidakovic, M., Sligar, S. G., Li, H. & Poulos, T. L. Understanding the role of the essential Asp251 in cytochrome P450cam using site-directed mutagenesis, crystallography, and kinetic solvent isotope effect. *Biochemistry* **37,** 9211–9219 (1998).
202. Peterson, J. A. & Graham, S. E. A close family resemblance: the importance of structure in understanding cytochromes P450. *Structure* **6,** 1079–1085 (1998).
203. Woggon, W. D., Wagenknecht, H. A. & Claude, C. Synthetic active site analogues of heme-thiolate proteins: Characterization and identification of intermediates of the catalytic cycles of cytochrome P450cam and chloroperoxidase. *J. Inorg. Biochem.* **83,** 289–300 (2001).
204. Sharrock, M. *et al.* Cytochrome P450cam and its complexes, Mössbauer parameters of the heme iron. *BBA - Protein Struct.* **420,** 8–26 (1976).
205. Momenteau, M. & Reed, C. a. Synthetic Heme-Dioxygen Complexes. *Chem. Rev.* **94,** 659–698 (1994).
206. Tani, F., Matsu-ura, M., Nakayama, S. & Naruta, Y. Synthetic models for the active site of cytochrome P450. *Coord. Chem. Rev.* **226,** 219–226 (2002).

207. Macdonald, I. D. G., Sligar, S. G., Christian, J. F., Unno, M. & Champion, P. M. Identification of the Fe-O-O bending mode in oxycytochrome P450cam by Resonance Raman spectroscopy. *J. Am. Chem. Soc.* **121,** 376–380 (1999).
208. Loew, G. H. & Harris, D. L. Role of the heme active site and protein environment in structure, spectra, and function of the cytochrome p450s. *Chem. Rev.* **100,** 407–419 (2000).
209. Denisov, I. G., Makris, T. M., Sligar, S. G. & Schlichting, I. Structure and chemistry of cytochrome P450. *Chem. Rev.* **105,** 2253–2277 (2005).
210. Karuzina, I. I. & Archakov, A. I. The oxidative inactivation of cytochrome P450 in monooxygenase reactions. *Free Radic. Biol. Med.* **16,** 73–97 (1994).
211. Karuzina, I. I., Zgoda, V. G., Kuznetsova, G. P., Samenkova, N. F. & Archakov, A. I. Heme and apoprotein modification of cytochrome P450 2B4 during its oxidative inactivation in monooxygenase reconstituted system. *Free Radic. Biol. Med.* **26,** 620–632 (1999).
212. Bernhardt, R. Cytochrome P450: Structure, function, and generation of reactive oxygen species. *Rev. Physiol. Biochem. Pharmacol.* **127,** 137–221 (1995).
213. Hannemann, F., Bichet, A., Ewen, K. M. & Bernhardt, R. Cytochrome P450 systems-biological variations of electron transport chains. *Biochim. Biophys. Acta - Gen. Subj.* **1770,** 330–344 (2007).
214. Tyson, C. A., Lipscomb, J. D. & Gunsalus, I. C. The Roles of Putidaredoxin and P450cam in Methylene Hydroxylation. *J. Biol. Chem.* **247,** 5777–5784 (1972).
215. Pochapsky, T. C., Ye, X. M., Ratnaswamy, G. & Lyons, T. A. An NMR-Derived Model for the Solution Structure of Oxidized Putidaredoxin, a 2-Fe, 2-S Ferredoxin from *Pseudomonas*. *Biochemistry* **33,** 6424–6432 (1994).
216. Shimada, H. *et al.* Putidaredoxin-cytochrome P450(cam) interaction: Spin state of the heme iron modulates putidaredoxin structure. *J. Biol. Chem.* **274,** 9363–9369 (1999).
217. Green, A. J. *et al.* Expression, purification and characterisation of a *Bacillus subtilis* ferredoxin: A potential electron transfer donor to cytochrome P450 BioI. *J. Inorg. Biochem.* **93,** 92–99 (2003).
218. Trower, M. K., Emptage, M. H. & Sariaslani, F. S. Purification and characterization of a 7Fe ferredoxin from *Streptomyces griseus*. *Biochim. Biophys. Acta (BBA)/Protein Struct. Mol.* **1037,** 281–289 (1990).
219. Seliskar, M. & Rozman, D. Mammalian cytochromes P450 - Importance of tissue specificity. *Biochim. Biophys. Acta - Gen. Subj.* **1770,** 458–466 (2007).
220. Bernhardt, R. & Gunsalus, I. C. Reconstitution of cytochrome P4502B4 (LM2) activity with camphor and linalool monooxygenase electron donors. *Biochem. Biophys. Res. Commun.* **187,** 310–317 (1992).
221. Watanabe, I., Nara, F. & Serizawa, N. Cloning, characterization and expression of the gene encoding cytochrome P-450sca-in2 from *Streptomyces carbophilus* involved in production of pravastatin, a specific HMG-CoA reductase inhibitor. *Gene* **163,** 81–85 (1995).
222. Hlavica, P. On the function of cytochrome b5 in the cytochrome P-450-dependent oxygenase system. *Arch. Biochem. Biophys.* **228,** 600–608 (1984).
223. Vergéres, G. & Waskell, L. Cytochrome b5, its functions, structure and membrane topology. *Biochimie* **77,** 604–620 (1995).
224. Porter, T. D. The roles of cytochrome b5 in cytochrome P450 reactions. *J. Biochem. Mol. Toxicol.* **16,** 311–316 (2002).
225. Elahian, F., Sepehrizadeh, Z., Moghimi, B. & Mirzaei, S. A. Human cytochrome b5 reductase: structure, function, and potential applications. *Crit. Rev. Biotechnol.* **34,** 134–143 (2014).
226. Henderson, C. J., McLaughlin, L. A. & Wolf, C. R. Evidence That Cytochrome b5 and Cytochrome b5 Reductase Can Act as Sole Electron Donors to the Hepatic Cytochrome P450 System. *Mol. Pharmacol.* **83,** (2013).
227. Yamazaki, H., Johnson, W. W., Ueng, Y.-F., Shimada, T. & Guengerich, F. P. Lack of Electron Transfer from Cytochrome b5 in Stimulation of Catalytic Activities of Cytochrome P450 3A4: CHARACTERIZATION OF A RECONSTITUTED CYTOCHROME P450 3A4/NADPH-CYTOCHROME P450 REDUCTASE SYSTEM AND STUDIES WITH APO-CYTOCHROME b5. *J. Biol. Chem.* **271,** 27438–27444 (1996).
228. Estrada, D. F., Laurence, J. S. & Scott, E. E. Substrate-modulated cytochrome P450 17A1 and cytochrome b5 interactions revealed by NMR. *J. Biol. Chem.* **288,** 17008–18 (2013).
229. Porter, T. D. & Kasper, C. B. NADPH-cytochrome P-450 oxidoreductase: flavin mononucleotide and flavin adenine dinucleotide domains evolved from different flavoproteins. *Biochemistry* **25,** 1682–1687 (1986).

230. Hubbard, P. A., Shen, A. L., Paschke, R., Kasper, C. B. & Kim, J.-J. P. NADPH-cytochrome P450 oxidoreductase. Structural basis for hydride and electron transfer. *J. Biol. Chem.* **276,** 29163–29170 (2001).
231. Vermilion, J. L. & Coon, M. J. Purified liver microsomal NADPH-cytochrome P-450 reductase. Spectral characterization of oxidation-reduction states. *J. Biol. Chem.* **253,** 2694–2704 (1978).
232. Li, H.-L., Shao, Z.-B. & He, Z.-L. Hydrocarbon generation characteristics and potential of bitumen in the Tarim Basin. *Pet. Geol. Exp.* **31,** 373–398 (2009).
233. Iyanagi, T. & Mason, H. S. Properties of hepatic reduced nicotinamide adenine dinucleotide phosphate-cytochrome c reductase. *Biochemistry* **12,** 2297–2308 (1973).
234. Iyanagi, T., Makino, N. & Mason, H. S. Redox properties of the reduced nicotinamide adenine dinucleotide phosphate-cytochrome P-450 and reduced nicotinamide adenine dinucleotide-cytochrome b5 reductases. *Biochemistry* **13,** 1701–1710 (1974).
235. Iyanagi, T., Anan, F. K., Imai, Y. & Mason, H. S. Studies on the microsomal mixed function oxidase system: redox properties of detergent-solubilized NADPH-cytochrome P-450 reductase. *Biochemistry* **17,** 2224–2230 (1978).
236. Gutierrez, A. *et al.* Interflavin electron transfer in human cytochrome P450 reductase is enhanced by coenzyme binding: Relaxation kinetic studies with coenzyme analogues. *Eur. J. Biochem.* **270,** 2612–2621 (2003).
237. Laursen, T., Jensen, K. & Müller, B. L. Conformational changes of the NADPH-dependent cytochrome P450 reductase in the course of electron transfer to cytochromes P450. *Biochim. Biophys. Acta - Proteins Proteomics* **1814,** 132–138 (2011).
238. Schwarz, D. *et al.* Arachidonic and eicosapentaenoic acid metabolism by human CYP1A1: Highly stereoselective formation of 17(R),18(S)-epoxyeicosatetraenoic acid. *Biochem. Pharmacol.* **67,** 1445–1457 (2004).
239. Lewis, D. F. V., Gillam, E. M. J., Everett, S. A. & Shimada, T. Molecular modelling of human CYP1B1 substrate interactions and investigation of allelic variant effects on metabolism. *Chem. Biol. Interact.* **145,** 281–295 (2003).
240. Choudhary, D., Jansson, I., Stoilov, I., Sarfarazi, M. & Schenkman, J. B. Metabolism of retinoids and arachidonic acid by human and mouse cytochrome P450 1B1. *Drug Metab. Dispos.* **32,** 840–847 (2004).
241. Amet, Y. *et al.* Validation of the (ω-1)-hydroxylation of lauric acid as an in vitro substrate probe for human liver CYP2E1. *Biochem. Pharmacol.* **50,** 1775–1782 (1995).
242. Adas, F. *et al.* Requirement for ω and (ω-1)-hydroxylations of fatty acids by human cytochromes P450 2E1 and 4A11. *J. Lipid Res.* **40,** 1990–1997 (1999).
243. Fukuda, T. *et al.* Different Mechanisms of Regioselection of Fatty Acid Hydroxylation by Laurate (ω-1)-Hydroxylating P450s, P450 2C2 and P450 2E1. *J. Biochem.* **115,** 338–344 (1994).
244. Seghezzi, W., Sanglard, D. & Fiechter, A. Characterization of a second alkane-inducible cytochrome P450-encoding gene, CYP52A2, from *Candida tropicalis*. *Gene* **106,** 51–60 (1991).
245. Seghezzi, W. *et al.* Identification and Characterization of Additional Members of the Cytochrome P450 Multigene Family CYP52 of *Candida tropicalis*. *DNA Cell Biol.* **11,** 767–780 (1992).
246. Craft, D. L., Madduri, K. M., Eshoo, M. & Wilson, C. R. Identification and Characterization of the CYP52 Family of *Candida tropicalis* ATCC 20336, Important for the Conversion of Fatty Acids and Alkanes to α,ω-Dicarboxylic Acids. *Appl. Environ. Microbiol.* **69,** 5983–5991 (2003).
247. Picataggio, S. *et al.* Metabolic engineering of Candida tropicalis for the production of long-chain dicarboxylic acids. *Nat. Biotechnol.* **10,** 894–898 (1992).
248. Ohkuma, M., Tanimoto, T., Yano, K. & Takagi, M. CYP52 (Cytochrome P450alk) Multigene Family in *Candida maltosa* : Molecular Cloning and Nucleotide Sequence of the Two Tandemly Arranged Genes. *DNA Cell Biol.* **10,** 271–282 (1991).
249. Zimmer, T., Ohkuma, M., Ohta, A., Takagi, M. & Schunck, W.-H. The CYP52 Multigene Family of *Candida maltosa* Encodes Functionally Diverse n-Alkane-Inducible Cytochromes P450. *Biochem. Biophys. Res. Commun.* **224,** 784–789 (1996).
250. Scheller, U., Zimmer, T., Kärgel, E. & Schunck, W.-H. Characterization of the n-alkane and fatty acid hydroxylating cytochrome P450 forms 52A3 and 52A4. *Arch. Biochem. Biophys.* **328,** 245–254 (1996).

251. Blasig, R. *et al.* Degradation of long-chain n-alkanes by the yeast *Candida maltosa*. *Appl. Microbiol. Biotechnol.* **31–31,** 571–576 (1989).
252. Lottermoser, K., Schunck, W.-H. & Asperger, O. Cytochromes P450 of the sophorose lipid-producing yeast *Candida apicola*: Heterogeneity and polymerase chain reaction-mediated cloning of two genes. *Yeast* **12,** 565–575 (1996).
253. Van Bogaert, I. N. A., De Mey, M., Develter, D., Soetaert, W. & Vandamme, E. J. Importance of the cytochrome P450 monooxygenase CYP52 family for the sophorolipid-producing yeast *Candida bombicola*. *FEMS Yeast Res.* **9,** 87–94 (2009).
254. Huang, F. C., Peter, A. & Schwab, W. Expression and Characterization of CYP52 Genes Involved in the Biosynthesis of Sophorolipid and Alkane Metabolism from *Starmerella bombicola*. *Appl. Environ. Microbiol.* **80,** 766–776 (2014).
255. Kurtzman, C. P., Price, N. P. J., Ray, K. J. & Kuo, T. M. Production of sophorolipid biosurfactants by multiple species of the *Starmerella* (*Candida*) *bombicola* yeast clade. *FEMS Microbiol. Lett.* **311,** 140–146 (2010).
256. Hawkes, D. B., Slessor, K. E., Bernhardt, P. V. & De Voss, J. J. Cloning, expression and purification of cindoxin, an unusual Fmn-containing cytochrome P450 redox partner. *ChemBioChem* **11,** 1107–1114 (2010).
257. Hawkes, D. B., Adams, G. W., Burlingame, A. L., Ortiz De Montellano, P. R. & De Voss, J. J. Cytochrome P450cin (CYP176A), isolation, expression, and characterization. *J. Biol. Chem.* **277,** 27725–27732 (2002).
258. McLean, M. A., Maves, S. A., Weiss, K. E., Krepich, S. & Sligar, S. G. Characterization of a Cytochrome P450 from the Acidothermophilic Archaea *Sulfolobus solfataricus*. *Biochem. Biophys. Res. Commun.* **252,** 166–172 (1998).
259. Park, S.-Y. *et al.* Thermophilic cytochrome P450 (CYP119) from *Sulfolobus solfataricus* : high resolution structure and functional properties. *J. Inorg. Biochem.* **91,** 491–501 (2002).
260. Puchkaev, A. V. & Ortiz De Montellano, P. R. The *Sulfolobus solfataricus* electron donor partners of thermophilic CYP119: An unusual non-NAD(P)H-dependent cytochrome P450 system. *Arch. Biochem. Biophys.* **434,** 169–177 (2005).
261. Jackson, C. J. *et al.* A novel sterol 14α-demethylase/ferredoxin fusion protein (MCCYP51FX) from *Methylococcus capsulatus* represents a new class of the cytochrome P450 superfamily. *J. Biol. Chem.* **277,** 46959–46965 (2002).
262. McLean, K. J. *et al.* Biophysical characterization of the sterol demethylase P450 from *Mycobacterium tuberculosis*, its cognate ferredoxin, and their interactions. *Biochemistry* **45,** 8427–8443 (2006).
263. Bellamine, A., Mangla, A. T., Nes, W. D. & Waterman, M. R. Characterization and catalytic properties of the sterol 14α-demethylase from Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 8937–8942 (1999).
264. Rylott, E. L. *et al.* An explosive-degrading cytochrome P450 activity and its targeted application for the phytoremediation of RDX. *Nat. Biotechnol.* **24,** 216–219 (2006).
265. Jayamani, I., Manzella, M. P. & Cupples, A. M. RDX degradation potential in soils previously unexposed to RDX and the identification of RDX-degrading species in one agricultural soil using stable isotope probing. *Water. Air. Soil Pollut.* **224,** 1745–1759 (2013).
266. Jackson, R. G., Rylott, E. L., Fournier, D., Hawari, J. & Bruce, N. C. Exploring the biochemical properties and remediation applications of the unusual explosive-degrading P450 system XplA/B. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 16822–16827 (2007).
267. Andeer, P. F., Stahl, D. A., Bruce, N. C. & Strand, S. E. Lateral Transfer of Genes for Hexahydro-1,3,5-Trinitro-1,3,5-Triazine (RDX) Degradation. *Appl. Environ. Microbiol.* **75,** 3258–3262 (2009).
268. Gassner, G. T., Ludwig, M. L., Gatti, D. L., Correll, C. C. & Ballou, D. P. Structure and mechanism of the iron-sulfur flavoprotein phthalate dioxygenase reductase. *FASEB J.* **9,** 1411–1418 (1995).
269. Gibson, D. T. & Parales, R. E. Aromatic hydrocarbon dioxygenases in environmental biotechnology. *Curr. Opin. Biotechnol.* **11,** 236–243 (2000).
270. Roberts, G., Grogan, G., Greter, A., Flitsch, S. L. & Turner, N. J. Identification of a New Class of Cytochrome P450 from a *Rhodococcus* sp. *J. Bacteriol.* **184,** 3898–3908 (2002).
271. Roberts, G. A. *et al.* A self-sufficient cytochrome p450 with a primary structural organization that includes a flavin domain and a [2Fe-2S] redox center. *J. Biol. Chem.* **278,** 48914–48920 (2003).

272. De Mot, R. & Parret, A. H. A. A novel class of self-sufficient cytochrome P450 monooxygenases in prokaryotes. *Trends Microbiol.* **10,** 502–508 (2002).
273. Nagy, I. *et al.* Degradation of the thiocarbamate herbicide EPTC (S-ethyl dipropylcarbamothioate) and biosafening by *Rhodococcus* sp. Strain NI86/21 involve an inducible cytochrome P-450 system and aldehyde dehydrogenase. *J. Bacteriol.* **177,** 676–687 (1995).
274. Warman, A. J. *et al.* Characterization of *Cupriavidus metallidurans* CYP116B1 - A thiocarbamate herbicide oxygenating P450-phthalate dioxygenase reductase fusion protein. *FEBS J.* **279,** 1675–1693 (2012).
275. Miura, Y. & Fulco, A. J. ω-1, ω-2 and ω-3 Hydroxylation of long-chain fatty acids, amides and alcohols by a soluble enzyme system from *Bacillus megaterium*. *Biochim. Biophys. Acta (BBA)/Lipids Lipid Metab.* **388,** 305–317 (1975).
276. Sevrioukova, I. F. & Peterson, J. A. NADPH-P-450 reductase: Structural and functional comparisons of the eukaryotic and prokaryotic isoforms. *Biochimie* **77,** 562–572 (1995).
277. Daff, S. N. *et al.* Redox control of the catalytic cycle of flavocytochrome P-450 BM3. *Biochemistry* **36,** 13816–13823 (1997).
278. Munro, A. W., Daff, S. N., Coggins, J. R., Lindsay, J. G. & Chapman, S. K. Probing electron transfer in flavocytochrome P-450 BM3 and its component domains. *Eur. J. Biochem.* **239,** 403–409 (1996).
279. Tsotsou, G. E., Cass, A. E. G. & Gilardi, G. High throughput assay for cytochrome P450 BM3 for screening libraries of substrates and combinatorial mutants. *Biosens. Bioelectron.* **17,** 119–131 (2002).
280. Di Nardo, G. *et al.* Wild-type CYP102A1 as a biocatalyst: Turnover of drugs usually metabolised by human liver enzymes. *J. Biol. Inorg. Chem.* **12,** 313–323 (2007).
281. Reinen, J. *et al.* Efficient screening of cytochrome P450 BM3 mutants for their metabolic activity and diversity toward a wide set of drug-like molecules in chemical space. *Drug Metab. Dispos.* **39,** 1568–1576 (2011).
282. Schneider, S., Wubbolts, M. G., Sanglard, D. & Witholt, B. Production of chiral hydroxy long chain fatty acids by whole cell biocatalysis of pentadecanoic acid with an *E. coli* recombinant containing cytochrome P450BM-3 monooxygenase. *Tetrahedron Asymmetry* **9,** 2832–2844 (1998).
283. Truant, G., Komandla, M. R., Falck, J. R. & Peterson, J. A. P450BM-3: absolute configuration of the primary metabolites of palmitic acid. *Arch. Biochem. Biophys.* **366,** 192–198 (1999).
284. Cryle, M. J. & De Voss, J. J. Facile determination of the absolute stereochemistry of hydroxy fatty acids by GC: application to the analysis of fatty acid oxidation by a P450BM3 mutant. *Tetrahedron Asymmetry* **18,** 547–551 (2007).
285. Graham-Lorence, S. *et al.* An active site substitution, F87V, converts cytochrome P450 BM-3 into a regio- and stereoselective (14S,15R)-arachidonic acid epoxygenase. *J. Biol. Chem.* **272,** 1127–1135 (1997).
286. Vottero, E. *et al.* Role of residue 87 in substrate selectivity and regioselectivity of drug-metabolizing cytochrome P450 CYP102A1 M11. *J. Biol. Inorg. Chem.* **16,** 899–912 (2011).
287. Wolter, F., Clausnitzer, A., Akoglu, B. & Stein, J. Piceatannol, a natural analog of resveratrol, inhibits progression through the S phase of the cell cycle in colorectal cancer cell lines. *J. Nutr.* **132,** 298–302 (2002).
288. Geahlen, R. L. & McLaughlin, J. L. Piceatannol (3,4,3',5'-tetrahydroxy-trans-stilbene) is a naturally occurring protein-tyrosine kinase inhibitor. *Biochem. Biophys. Res. Commun.* **165,** 241–245 (1989).
289. Potter, G. A. *et al.* The cancer preventative agent resveratrol is converted to the anticancer agent piceatannol by the cytochrome P450 enzyme CYP1B1. *Br. J. Cancer* **86,** 774–778 (2002).
290. Kim, D. H. D.-H., Ahn, T., Jung, H.-C. H. C., Pan, J.-G. J. G. & Yun, C. H. C.-H. Generation of the Human Metabolite Piceatannol from the Anticancer-Preventive Agent Resveratrol by Bacterial Cytochrome P450 BM3. *Drug Metab. Dispos.* **37,** 932–936 (2009).
291. Gustafsson, M. C. U. *et al.* Expression, Purification, and Characterization of *Bacillus subtilis* Cytochromes P450 CYP102A2 and CYP102A3: Flavocytochrome Homologues of P450 BM3 from Bacillus megaterium. *Biochemistry* **43,** 5474–5487 (2004).
292. Chowdhary, P. K., Alemseghed, M. & Haines, D. C. Cloning, expression and characterization of a fast self-sufficient P450: CYP102A5 from Bacillus cereus. *Arch. Biochem. Biophys.* **468,** 32–43 (2007).

293. Dietrich, M. *et al.* Cloning, expression and characterisation of CYP102A7, a self-sufficient P450 monooxygenase from *Bacillus licheniformis*. *Appl. Microbiol. Biotechnol.* **79,** 931–940 (2008).
294. Kitazume, T., Takaya, N., Nakayama, N. & Shoun, H. *Fusarium oxysporum* fatty-acid subterminal hydroxylase (CYP505) is a membrane-bound eukaryotic counterpart of *Bacillus megaterium* cytochrome P450BM3. *J. Biol. Chem.* **275,** 39734–39740 (2000).
295. Nakayama, N., Takemae, A. & Shoun, H. Cytochrome P450foxy, a catalytically self-sufficient fatty acid hydroxylase of the fungus *Fusarium oxysporum*. *J. Biochem.* **119,** 435–440 (1996).
296. Daiber, A., Shoun, H. & Ullrich, V. Nitric oxide reductase (P450) from *Fusarium oxysporum*. *J. Inorg. Biochem.* **99,** 185–193 (2005).
297. Shiro, Y. *et al.* Spectroscopic and kinetic studies on reaction of cytochrome P450nor with nitric oxide: Implication for its nitric oxide reduction mechanism. *J. Biol. Chem.* **270,** 1617–1623 (1995).
298. Takaya, N. *et al.* Cytochrome p450nor, a novel class of mitochondrial cytochrome P450 involved in nitrate respiration in the fungus *Fusarium oxysporum*. *Arch. Biochem. Biophys.* **372,** 340–346 (1999).
299. Oshima, R. *et al.* Structural evidence for direct hydride transfer from NADH to cytochrome P450nor. *J. Mol. Biol.* **342,** 207–217 (2004).
300. Kudo, T., Tomura, D., Liu, D. L., Dai, X. Q. & Shoun, H. Two isozymes of P450nor of Cylindrocarpon tonkinense: Molecular cloning of the cDNAs and genes, expressions in the yeast, and the putative NAD(P)H-binding site. *Biochimie* **78,** 792–799 (1996).
301. Kaya, M. *et al.* Cloning and Enhanced Expression of the Cytochrome P450nor Gene (*nicA* ; CYP55A5) Encoding Nitric Oxide Reductase from *Aspergillus oryzae*. *Biosci. Biotechnol. Biochem.* **68,** 2040–2049 (2004).
302. Zhang, L., Takaya, N., Kitazume, T., Kondo, T. & Shoun, H. Purification and cDNA cloning of nitric oxide reductase cytochrome P450nor (CYP55A4) from *Trichosporon cutaneum*. *Eur. J. Biochem.* **268,** 3198–3204 (2001).
303. Laudert, D., Pfannschmidt, U., Lottspeich, F., Holländer-Czytko, H. & Weiler, E. W. Cloning, molecular and functional characterization of *Arabidopsis thaliana* allene oxide synthase (CYP74), the first enzyme of the octadecanoid pathway to jasmonates. *Plant Mol. Biol.* **31,** 323–335 (1996).
304. Froehlich, J. E., Itoh, A. & Howe, G. A. Tomato allene oxide synthase and fatty acid hydroperoxide lyase, two cytochrome P450s involved in oxylipin metabolism, are targeted to different membranes of chloroplast envelope. *Plant Physiol.* **125,** 306–317 (2001).
305. Brodhun, F. & Feussner, I. Oxylipins in fungi. *FEBS J.* **278,** 1047–1063 (2011).
306. Toporkova, Y. Y. *et al.* Structure-function relationship in the CYP74 family: Conversion of divinyl ether synthases into allene oxide synthases by site-directed mutagenesis. *FEBS Lett.* **587,** 2552–2558 (2013).
307. Cho, K. Bin, Lai, W., Hamberg, M., Raman, C. S. & Shaik, S. The reaction mechanism of allene oxide synthase: Interplay of theoretical QM/MM calculations and experimental investigations. *Arch. Biochem. Biophys.* **507,** 14–25 (2011).
308. Turner, J. G., Ellis, C. & Devoto, A. The Jasmonate Signal Pathway. *Plant Cell* **Supplement,** 153–164 (2002).
309. Grechkin, A. N. Hydroperoxide lyase and divinyl ether synthase. *Prostaglandins Other Lipid Mediat.* **68–69,** 457–470 (2002).
310. Van Bogaert, I. N. A., De Maeseneire, S. L., Develter, D., Soetaert, W. & Vandamme, E. J. Development of a transformation and selection system for the glycolipid-producing yeast *Candida bombicola*. *Yeast* **25,** 273–278 (2008).
311. Van Bogaert, I. N. A. *et al.* Cloning, characterization and functionality of the orotidine-5'-phosphate decarboxylase gene (URA3) of the glycolipid-producing yeast *Candida bombicola*. *Yeast* **24,** 201–208 (2007).
312. Van Bogaert, I. N. A., De Maeseneire, S. L., Develter, D., Soetaert, W. & Vandamme, E. J. Cloning and characterisation of the glyceraldehyde 3-phosphate dehydrogenase gene of *Candida bombicola* and use of its promoter. *J. Ind. Microbiol. Biotechnol.* **35,** 1085–1092 (2008).
313. Soetaert, W., Van Bogaert, I. N. A. & Roelants, S. L. K. W. Methods to produce bolaamphiphilic glycolipids. 50 (2013).

314. Van Bogaert, I. N. A., Sabirova, J. S., Develter, D., Soetaert, W. & Vandamme, E. J. Knocking out the MFE-2 gene of *Candida bombicola* leads to improved medium-chain sophorolipid production. *FEMS Yeast Res.* **9,** 610–617 (2009).
315. Van Bogaert, I. N. A. *et al. Preparation of 20-HETE using multifunctional enzyme type 2 (MFE-2)-negative Starmerella bombicola*. *Journal of Lipid Research* **54,** 3215–3219 (American Society for Biochemistry and Molecular Biology, 2013).
316. Takahashi, F. Method for producing glycolipid. *Patent* 1–13 (2015).
317. Saerens, K. M. J., Zhang, J., Saey, L., Van Bogaert, I. N. A. & Soetaert, W. 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 (2011).
318. Roelants, S. L. K. W. *et al. Candida bombicola* as a platform organism for the production of tailor-made biomolecules. *Biotechnol. Bioeng.* **110,** 2494–2503 (2013).
319. De Oliveira, V. C., Maeda, I., Delessert, S. & Poirier, Y. Increasing the carbon flux toward synthesis of short-chain-length-medium-chain-length polyhydroxyalkanoate in the peroxisome of *Saccharomyces cerevisiae* through modification of the β-oxidation cycle. *Appl. Environ. Microbiol.* **70,** 5685–5687 (2004).
320. Vijayasankaran, N., Carlson, R. & Srienc, F. Synthesis of poly[(R)-3-hydroxybutyric acid) in the cytoplasm of *Pichia pastoris* under oxygen limitation. *Biomacromolecules* **6,** 604–611 (2005).
321. Sabirova, J. S. *et al.* The 'LipoYeasts' project: using the oleaginous yeast *Yarrowia lipolytica* in combination with specific bacterial genes for the bioconversion of lipids, fats and oils into high-value products. *Microb. Biotechnol.* **4,** 47–54 (2011).
322. Lemieux, R. U., Thorn, J. A., Brice, C. & Haskins, R. H. Biochemistry of the *Ustilaginales*: II. Isolation and Partial Characterization of Ustilagic Acid. *Can. J. Chem.* **29,** 409–414 (1951).
323. Lemieux, R. U. the Biochemistry of the *Ustilaginales*: III: the Degradation Products and Proof of the Chemical Heterogeneity of Ustilagic Acid. *Can. J. Chem.* **29,** 415–425 (1951).
324. Lemieux, R. U. & Charanduk, R. Biochemistry of the *Ustilaginales*: VI: the Acyl Groups of Ustilagic Acid. *Can. J. Chem.* **29,** 759–766 (1951).
325. Lemieux, R. U., Thorn, J. A. & Bauer, H. F. Biochemistry of the *Ustilaginales*: IX. the β-d-Cellobioside Units of the Ustilagic Acids. *Can. J. Chem.* **31,** 1054–1059 (1953).
326. Teichmann, B. *et al.* Identification of a biosynthesis gene cluster for flocculosin a cellobiose lipid produced by the biocontrol agent *Pseudozyma flocculosa*. *Mol. Microbiol.* **79,** 1483–1495 (2011).
327. Teichmann, B., Linne, U., Hewald, S., Marahiel, M. a. & Bölker, M. A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from *Ustilago maydis*. *Mol. Microbiol.* **66,** 525–533 (2007).
328. Hewald, S. *et al.* Identification of a gene cluster for biosynthesis of mannosylerythritol lipids in the basidiomycetous fungus *Ustilago maydis*. *Appl. Environ. Microbiol.* **72,** 5469–5477 (2006).
329. Teichmann, B. *et al.* Beta hydroxylation of glycolipids from Ustilago maydis and Pseudozyma flocculosa by an NADPH-dependent β-hydroxylase. *Appl. Environ. Microbiol.* **77,** 7823–7829 (2011).
330. Teichmann, B., Liu, L., Schink, K. O. & Bölker, M. Activation of the ustilagic acid biosynthesis gene cluster in *Ustilago maydis* by the C2H2 zinc finger transcription factor Rua1. *Appl. Environ. Microbiol.* **76,** 2633–2640 (2010).
331. Mimee, B., Labbé, C. & Bélanger, R. R. Catabolism of flocculosin, an antimicrobial metabolite produced by *Pseudozyma flocculosa*. *Glycobiology* **19,** 995–1001 (2009).
332. Arutchelvi, J. I., Bhaduri, S., Uppara, P. V. & Doble, M. Mannosylerythritol lipids: A review. *J. Ind. Microbiol. Biotechnol.* **35,** 1559–1570 (2008).
333. Fukuoka, T., Morita, T., Konishi, M., Imura, T. & Kitamoto, D. A basidiomycetous yeast, *Pseudozyma tsukubaensis*, efficiently produces a novel glycolipid biosurfactant. The identification of a new diastereomer of mannosylerythritol lipid-B. *Carbohydr. Res.* **343,** 555–560 (2008).
334. Morita, T. *et al.* Production of a novel glycolipid biosurfactant, mannosylmannitol lipid, by *Pseudozyma parantarctica* and its interfacial properties. *Appl. Microbiol. Biotechnol.* **83,** 1017–1025 (2009).
335. Morita, T. *et al.* Efficient production of di- and tri-acylated mannosylerythritol lipids as glycolipid biosurfactants by *Pseudozyma parantarctica* JCM 11752(T). *J. Oleo Sci.* **57,** 557–565 (2008).

336. Worakitkanchanakul, W. *et al.* Aqueous-phase behavior and vesicle formation of natural glycolipid biosurfactant, mannosylerythritol lipid-B. *Colloids Surfaces B Biointerfaces* **65,** 106–112 (2008).
337. Inoh, Y., Furuno, T., Hirashima, N., Kitamoto, D. & Nakanishi, M. Rapid delivery of small interfering RNA by biosurfactant MEL-A-containing liposomes. *Biochem. Biophys. Res. Commun.* **414,** 635–640 (2011).
338. Morita, T., Fukuoka, T., Imura, T. & Kitamoto, D. Production of mannosylerythritol lipids and their application in cosmetics. *Appl. Microbiol. Biotechnol.* **97,** 4691–4700 (2013).
339. Strieker, M. & Marahiel, M. A. The structural diversity of acidic lipopeptide antibiotics. *ChemBioChem* **10,** 607–616 (2009).
340. Baumgart, F., Kluge, B., Ullrich, C., Vater, J. & Ziessow, D. Identification of amino acid substitutions in the lipopeptide surfactin using 2D NMR spectroscopy. *Biochem. Biophys. Res. Commun.* **177,** 998–1005 (1991).
341. Huang, X. *et al.* Antiviral activity of antimicrobial lipopeptide from *Bacillus subtilis* fmbj against *Pseudorabies* Virus, *Porcine* Parvovirus, Newcastle Disease Virus and Infectious Bursal Disease Virus in vitro. *Int. J. Pept. Res. Ther.* **12,** 373–377 (2006).
342. Steenbergen, J. N., Alder, J., Thorne, G. M. & Tally, F. P. Daptomycin: A lipopeptide antibiotic for the treatment of serious Gram-positive infections. *J. Antimicrob. Chemother.* **55,** 283–288 (2005).
343. Vance, F. G. J. *et al.* Daptomycin versus standard therapy for bacteremia and endocarditis caused by *Staphylococcus aureus*. *N. Engl. J. Med.* **355,** 653–665 (2006).
344. Coëffet-Le Gal, M. F., Thurston, L., Rich, P., Miao, V. & Baltz, R. H. Complementation of daptomycin *dptA* and *dptD* deletion mutations in trans and production of hybrid lipopeptide antibiotics. *Microbiology* **152,** 2993–3001 (2006).
345. Doekel, S. *et al.* Non-ribosomal peptide synthetase module fusions to produce derivatives of daptomycin in *Streptomyces roseosporus*. *Microbiology* **154,** 2872–2880 (2008).
346. Chiocchini, C., Linne, U. & Stachelhaus, T. *In Vivo* Biocombinatorial Synthesis of Lipopeptides by COM Domain-Mediated Reprogramming of the Surfactin Biosynthetic Complex. *Chem. Biol.* **13,** 899–908 (2006).
347. Alexander, D. C. *et al.* Production of novel lipopeptide antibiotics related to A54145 by *Streptomyces fradiae* mutants blocked in biosynthesis of modified amino acids and assignment of *lptJ, lptK* and *lptL* gene functions. *J. Antibiot. (Tokyo).* **64,** 79–87 (2011).
348. Baltz, R. H. in *Antimicrobials: New and Old Molecules in the Fight Against Multi-Resistant Bacteria* (eds. Marinelli, F. & Genilloud, O.) **9783642399,** 364 (Springer Berlin Heidelberg, 2014).
349. Powell, A. *et al.* Engineered blosynthesis of nonribosomal lipopeptides with modified fatty acid side chains. *J. Am. Chem. Soc.* **129,** 15182–15191 (2007).
350. Snydman, D. R., Jacobus, N. V. & McDermott, L. A. Activity of a novel cyclic lipopeptide, CB-183,315, against resistant *Clostridium difficile* and other gram-positive aerobic and anaerobic intestinal pathogens. *Antimicrob. Agents Chemother.* **56,** 3448–3452 (2012).
351. Winn, M., Fyans, J. K., Zhuo, Y. & Micklefield, J. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* **33,** 317–347 (2016).
352. Lin, S., Lin, K., Lo, C. & Lin, Y. Enhanced biosurfactant production by a *Bacillus licheniformis* mutant. *Appl. Microbiol. Biotechnol.* **23,** 267–273 (1998).
353. Zhao, J. *et al.* Genome shuffling of *Bacillus amyloliquefaciens* for improving antimicrobial lipopeptide production and an analysis of relative gene expression using FQ RT-PCR. *J. Ind. Microbiol. Biotechnol.* **39,** 889–896 (2012).
354. Huang, D. *et al.* In silico aided metabolic engineering of *Streptomyces roseosporus* for daptomycin yield improvement. *Appl. Microbiol. Biotechnol.* **94,** 637–649 (2012).
355. Wang, L. *et al.* Improvement of A21978C production in *Streptomyces roseosporus* by reporter-guided rpsL mutation selection. *J. Appl. Microbiol.* **112,** 1095–1101 (2012).
356. Hu, H. & Ochi, K. Novel Approach for Improving the Productivity of Antibiotic-Producing Strains by Inducing Combined Resistant Mutations. *Appl. Environ. Microbiol.* **67,** 1885–1892 (2001).
357. Tanaka, Y. *et al.* Antibiotic overproduction by *rpsL* and *rsmG* mutants of various actinomycetes. *Appl. Environ. Microbiol.* **75,** 4919–4922 (2009).
358. Chng, C., Lum, A. M., Vroom, J. A. & Kao, C. M. A key developmental regulator controls the synthesis of the antibiotic erythromycin in *Saccharopolyspora erythraea*. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 11346–11351 (2008).

359. Laurent, N., Voglmeir, J. & Flitsch, S. L. Glycoarrays—tools for determining protein–carbohydrate interactions and glycoenzyme specificity. *Chem. Commun.* **1473,** 4400 (2008).
360. Delbeke, E. I. P. *et al.* Petroselinic acid purification and its use for the fermentation of new sophorolipids. *AMB Express* **6,** (2016).
361. Delbeke, E., Roman, B. I., Marin, G. B., Van Geem, K. M. & Stevens, C. V. A new class of antimicrobial biosurfactants: quaternary ammonium sophorolipids. *Green Chem.* **17,** 3373–3377 (2015).
362. Delbeke, E. I. P. *et al.* Sophorolipid Amine Oxide Production by a Combination of 2 Fermentation Scale-up and Chemical Modification. *Ind. Eng. Chem. Res.* **55,** 7273–7281 (2016).
363. Delbeke, E. I. P., Movsisyan, M., Van Geem, K. M. & Stevens, C. V. Chemical and enzymatic modification of sophorolipids. *Green Chem.* **18,** 76–104 (2016).
364. Trummler, K., Effenberger, F. & Syldatk, C. An integrated microbial/enzymatic process for production of rhamnolipids and L-(+)-rhamnose from rapeseed oil with *Pseudomonas* sp. DSM 2874. *Eur. J. Lipid Sci. Technol.* **105,** 563–571 (2003).
365. Recke, V. K. *et al.* Lipase-catalyzed acylation of microbial mannosylerythritol lipids (biosurfactants) and their characterization. *Carbohydr. Res.* **373,** 82–88 (2013).
366. Beggs, J. D. Transformation of yeast by a replicating hybrid plasmid. *Nature* **275,** 104–109 (1978).
367. Newlon, C. S. & Theis, J. F. The structure and function of yeast ARS elements. *Curr. Opin. Genet. Dev.* **3,** 752–758 (1993).
368. Roelants, S. *Starmerella bombicola* as a platform organism for the production of biobased compounds. (Ghent University. Faculty of Bioscience Engineering, 2013).
369. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. & Cormier, M. J. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* **111,** 229–233 (1992).
370. Lang, S., Brakemeier, A., Heckmann, R., Spöckner, S. & Rau, U. Production of native and modified sophorose lipids. *Chim. Oggi* **18,** 76–79 (2000).
371. Mamiatis, T., Fritsch, E. F., Sambrook, J. & Engel, J. *Molecular Cloning. A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, 1989).
372. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
373. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26,** 966–968 (2010).
374. Liachko, I. & Dunham, M. J. An autonomously replicating sequence for use in a wide range of budding yeasts. *FEMS Yeast Res.* **14,** 364–367 (2014).
375. Li, W.-C. *et al.* Sequence analysis of origins of replication in the *Saccharomyces cerevisiae* genomes. *Front. Microbiol.* **5,** 574 (2014).
376. Williams, J. S., Eckdahl, T. T. & Anderson, J. N. Bent DNA functions as a replication enhancer in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8,** 2763–9 (1988).
377. Van Bogaert, I. Microbial synthesis of sophorolipids by the yeast *Candida bombicola*. (Ghent University. Faculty of Bioscience Engineering, 2008).
378. Longtine, M. S. *et al.* Additional Modules for Versatile and Economical PCR-based Gene Deletion and Modification in *Saccharomyces cerevisiae*. *Yeast* **14,** 953–961 (1998).
379. Lütke-Eversloh, T. & Stephanopoulos, G. Combinatorial pathway analysis for improved L-tyrosine production in *Escherichia coli*: Identification of enzymatic bottlenecks by systematic gene overexpression. *Metab. Eng.* **10,** 69–77 (2008).
380. Palmer, J. M. & Keller, N. P. Secondary metabolism in fungi: does chromosomal location matter? *Curr. Opin. Microbiol.* **13,** 431–436 (2010).
381. Saerens, K. M. J. Synthesis of glycolipids by *Candida bombicola*. (Ghent University, 2012).
382. Zhou, J., Lin, J., Zhou, C., Deng, X. & Xia, B. *Cytotoxicity of red fluorescent protein DsRed is associated with the suppression of Bcl-xL translation*. *FEBS Letters* **585,** (2011).
383. Preuss, D. *et al.* Structure of the yeast endoplasmic reticulum: Localization of ER proteins using immunofluorescence and immunoelectron microscopy. *Yeast* **7,** 891–911 (1991).
384. Dunn, K. W., Kamocka, M. M. & McDonald, J. H. A practical guide to evaluating colocalization in biological microscopy. *Am. J. Physiol. Cell Physiol.* **300,** C723–C742 (2011).
385. Adler, J. & Parmryd, I. Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytom. Part A* **77,** 733–742 (2010).
386. Desmet, T. & Soetaert, W. Enzymatic glycosyl transfer: mechanisms and applications. *Biocatal. Biotransformation* **29,** 1–18 (2011).

387. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4,** 222 (2008).
388. Zhang, H. *et al.* Methods for peptide and protein quantitation by liquid chromatography-multiple reaction monitoring mass spectrometry. *Mol. Cell. Proteomics* **10,** M110.006593 (2011).
389. Davila, A.-M., Marchal, R. & Vandecasteele, J.-P. Sophorose lipid fermentation with differentiated substrate supply for growth and production phases. *Appl. Microbiol. Biotechnol.* **47,** 496–501 (1997).
390. Barry, S. M. *et al.* Cytochrome P450–catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. *Nat. Chem. Biol.* **8,** 814–816 (2012).
391. Tehlivets, O., Scheuringer, K. & Kohlwein, S. D. Fatty acid synthesis and elongation in yeast. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1771,** 255–270 (2007).
392. Van Bogaert, I. N. A., Fleurackers, S., Van Kerrebroeck, S., Develter, D. & Soetaert, W. Production of new-to-nature sophorolipids by cultivating the yeast *Candida bombicola* on unconventional hydrophobic substrates. *Biotechnol. Bioeng.* **108,** 734–741 (2011).
393. Brakemeier, A., Wullbrandt, D. & Lang, S. Microbial alkyl-sophorosides based on 1-dodecanol or 2-, 3- or 4-dodecanones. *Biotechnol. Lett.* **20,** 215–218 (1998).
394. Takahashi, F., Igarashi, K. & Hagihara, H. Identification of the fatty alcohol oxidase FAO1 from *Starmerella bombicola* and improved novel glycolipids production in an FAO1 knockout mutant. *Appl. Microbiol. Biotechnol.* 1–10 (2016). doi:10.1007/s00253-016-7702-6
395. Hewald, S., Josephs, K. & Bolker, M. Genetic analysis of biosurfactant production in *Ustilago maydis*. *Appl. Environ. Microbiol.* **71,** 3033–3040 (2005).
396. Damsten, M. C. *et al.* Application of drug metabolising mutants of cytochrome P450 BM3 (CYP102A1) as biocatalysts for the generation of reactive metabolites. *Chem. Biol. Interact.* **171,** 96–107 (2008).
397. Whitehouse, C. J. C. *et al.* A highly active single-mutation variant of P450BM3 (CYP102A1). *ChemBioChem* **10,** 1654–1656 (2009).
398. Kim, K. H. *et al.* Generation of human chiral metabolites of simvastatin and lovastatin by bacterial CYP102A1 mutants. *Drug Metab. Dispos.* **39,** 140–150 (2011).
399. Landwehr, M., Carbone, M., Otey, C. R., Li, Y. & Arnold, F. H. Diversification of Catalytic Function in a Synthetic Family of Chimeric Cytochrome P450s. *Chem. Biol.* **14,** 269–278 (2007).
400. Kang, J. Y. *et al.* Chimeric cytochromes P450 engineered by domain swapping and random mutagenesis for producing human metabolites of drugs. *Biotechnol. Bioeng.* **111,** 1313–1322 (2014).
401. Sadeghi, S. J. & Gilardi, G. Chimeric P450 enzymes: Activity of artificial redox fusions driven by different reductases for biotechnological applications. *Biotechnol. Appl. Biochem.* **60,** 102–110 (2013).
402. Kämper, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444,** 97–101 (2006).
403. Pompon, D., Louerat, B., Bronine, A. & Urban, P. Yeast expression of animal and plant P450s in optimized redox environments. *Methods Enzymol.* **272,** 51–64 (1996).
404. Brakemeier, A., Wullbrandt, D. & Lang, S. *Candida bombicola*: Production of novel alkyl glycosides based on glucose/2-dodecanol. *Appl. Microbiol. Biotechnol.* **50,** 161–166 (1998).
405. Lentz, O., Urlacher, V. & Schmid, R. D. Substrate specificity of native and mutated cytochrome P450 (CYP102A3) from *Bacillus subtilis*. *J. Biotechnol.* **108,** 41–49 (2004).
406. Whitehouse, C. J. C., Bell, S. G. & Wong, L.-L. P450 BM3 (CYP102A1): connecting the dots. *Chem. Soc. Rev.* **41,** 1218–1260 (2012).
407. Whitehouse, C. J. C., Bell, S. G. & Wong, L.-L. Desaturation of alkylbenzenes by cytochrome P450(BM3) (CYP102A1). *Chemistry (Easton).* **14,** 10905–8 (2008).
408. Whitehouse, C. J. C. *et al.* Evolved CYP102A1 (P450BM3) variants oxidise a range of non-natural substrates and offer new selectivity options. *Chem. Commun.* **1,** 966–8 (2008).
409. Appel, D., Lutz-Wahl, S., Fischer, P., Schwaneberg, U. & Schmid, R. D. A P450 BM-3 mutant hydroxylates alkanes, cycloalkanes, arenes and heteroarenes. *J. Biotechnol.* **88,** 167–171 (2001).
410. Seifert, A. *et al.* Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo- and chemoselectivity. *ChemBioChem* **10,** 853–861 (2009).
411. Oliver, C. F. *et al.* A single mutation in cytochrome P450 BM3 changes substrate orientation in a catalytic intermediate and the regiospecificity of hydroxylation. *Biochemistry* **36,** 1567–1572 (1997).

412. Chen, C.-K. J., Shokhireva, T. K., Berry, R. E., Zhang, H. & Walker, F. A. The effect of mutation of F87 on the properties of CYP102A1-CYP4C7 chimeras: altered regiospecificity and substrate selectivity. *J. Biol. Inorg. Chem.* **13,** 813–824 (2008).
413. Dubey, K. D., Wang, B. & Shaik, S. Molecular Dynamics and QM/MM Calculations Predict the Substrate-Induced Gating of Cytochrome P450 BM3 and the Regio- and Stereoselectivity of Fatty Acid Hydroxylation. *J. Am. Chem. Soc.* **138,** 837–845 (2016).
414. Cirino, P. C. C. & Arnold, F. H. Regioselectivity and Activity of Cytochrome P450 BM-3 and Mutant F87A in Reactions Driven by Hydrogen Peroxide. *Adv. Synth. Catal.* **344,** 932–937 (2002).
415. Dietrich, M., Do, T. A., Schmid, R. D., Pleiss, J. & Urlacher, V. B. Altering the regioselectivity of the subterminal fatty acid hydroxylase P450 BM-3 towards γ- and δ-positions. *J. Biotechnol.* **139,** 115–117 (2009).
416. Brühlmann, F. *et al.* Engineering cytochrome P450 BM3 of *Bacillus megaterium* for terminal oxidation of palmitic acid. *J. Biotechnol.* **184,** 17–26 (2014).
417. Reikofski, J. & Tao, B. Y. Polymerase chain reaction (PCR) techniques for site-directed mutagenesis. *Biotechnol. Adv.* **10,** 535–547 (1992).
418. Ashby, R. D., Solaiman, D. K. Y. & Foglia, T. A. The use of fatty acid esters to enhance free acid sophorolipid synthesis. *Biotechnol. Lett.* **28,** 253–260 (2006).
419. Tong, H., Bell, D., Tabei, K. & Siegel, M. M. Automated Data Massaging, Interpretation, and E-Mailing Modules for High Throughput Open Access Mass Spectrometry.
420. Kitamura, E. *et al.* Production of Hydroxlated Flavonoids with Cytochrome P450 BM3 Variant F87V and Their Antioxidative Activities. *Biosci. Biotechnol. Biochem.* **77,** 1340–1343 (2013).
421. Peterson, J. A., Ebel, R. E., O'Keeffe, D. H., Matsubara, T. & Estabrook, R. W. Temperature dependence of cytochrome P-450 reduction. A model for NADPH-cytochrome P-450 reductase:cytochrome P-450 interaction. *J. Biol. Chem.* **251,** 4010–4016 (1976).
422. Gao, R., Falkeborg, M., Xu, X. & Guo, Z. Production of sophorolipids with enhanced volumetric productivity by means of high cell density fermentation. *Appl. Microbiol. Biotechnol.* **97,** 1103–1111 (2013).
423. Van Bogaert, I. N. A. A., Buyst, D., Martins, J. C., Roelants, S. L. K. W. K. W. & Soetaert, W. K. Synthesis of bolaform biosurfactants by an engineered *Starmerella bombicola* yeast. *Biotechnol. Bioeng.* (2016). doi:10.1002/bit.26032
424. Geys, R. & Soetaert, W. Engineering van de P450 enzympopulatie van *Candida bombicola* voor de productie van new-to-nature surfactanten. (Ghent University, 2012).
425. Eschenfeldt, W. H. *et al.* Transformation of fatty acids catalyzed by cytochrome P450 monooxygenase enzymes of *Candida tropicalis*. *Appl. Environ. Microbiol.* **69,** 5992–9 (2003).
426. Davila, A.-M., Marchal, R. & Vandecasteele, J.-P. Kinetics and balance of a fermentation free from product inhibition: sophorose lipid production by *Candida bombicola*. *Appl. Microbiol. Biotechnol.* **38,** 6–11 (1992).
427. Kleine, M. L. & Fulco, A. J. Critical residues involved in FMN binding and catalytic activity in cytochrome P450BM-3. *J. Biol. Chem.* **268,** 7553–7561 (1993).
428. Joyce, M. G. *et al.* The crystal structure of the FAD/NADPH-binding domain of flavocytochrome P450 BM3. *FEBS J.* **279,** 1694–1706 (2012).
429. Roitel, O., Scrutton, N. S. & Munro, A. W. Electron transfer in flavocytochrome P450 BM3: kinetics of flavin reduction and oxidation, the role of cysteine 999, and relationships with mammalian cytochrome P450 reductase. *Biochemistry* **42,** 10809–10821 (2003).
430. Van Bogaert, I. N. A. *et al.* Microbial production and application of sophorolipids. *Appl. Microbiol. Biotechnol.* **76,** 23–34 (2007).
431. Oliveira, M. R., Camilios-neto, D., Baldo, C., Magri, A. & Celligoi, M. A. P. C. Biosynthesis and Production of Sophorolipids. *Int. J. Sci. Technol. Res.* **3,** 133–143 (2014).
432. Ratsep, P. & Shah, V. Identification and quantification of sophorolipid analogs using ultra-fast liquid chromatography – mass spectrometry. *J. Microbiol. Methods* **78,** 354–356 (2009).
433. Guilmanov, V. *et al.* Oxygen Transfer Rate and Sophorose Lipid Production by *Candida bombicola*. *Biotechnol. Bioeng.* **77,** 489–494 (2002).
434. Marchand, G. *et al.* Identification of Genes Potentially Involved in the Biocontrol Activity of *Pseudozyma flocculosa*. *Phytopathology* **99,** 1142–1149 (2009).
435. Rodrigues, L., Banat, I. M., Teixeira, J. & Oliveira, R. Biosurfactants: potential applications in medicine. *J. Antimicrob. Chemother.* **57,** 609–18 (2006).
436. Lourith, N. & Kanlayavattanakul, M. Natural surfactants used in cosmetics: glycolipids. *Int. J. Cosmet. Sci.* **31,** 255–261 (2009).

437. Morya, V. K., Ahn, C., Jeon, S. & Kim, E.-K. Medicinal and cosmetic potentials of sophorolipids. *Mini Rev. Med. Chem.* **13,** 1761–8 (2013).
438. Fleurackers, S., Van Bogaert, I. N. A. & Develter, D. On the production and identification of medium-chained sophorolipids. *Eur. J. Lipid Sci. Technol.* **112,** 655–662 (2010).
439. Reyes-Dominguez, Y. *et al.* Heterochromatic marks are associated with the repression of secondary metabolism clusters in *Aspergillus nidulans*. *Mol. Microbiol.* **76,** 1376–1386 (2010).
440. Shwab, E. K. & Keller, N. P. Regulation of secondary metabolite production in filamentous ascomycetes. *Mycol. Res.* **112,** 225–230 (2008).
441. Kim, B.-N., Joo, Y.-C., Kim, Y.-S., Kim, K.-R. & Oh, D.-K. Production of 10-hydroxystearic acid from oleic acid and olive oil hydrolyzate by an oleate hydratase from *Lysinibacillus fusiformis*. *Appl. Microbiol. Biotechnol.* **95,** 929–937 (2012).
442. Joo, Y.-C. *et al.* Production of 10-hydroxystearic acid from oleic acid by whole cells of recombinant *Escherichia coli* containing oleate hydratase from *Stenotrophomonas maltophilia*. *J. Biotechnol.* **158,** 17–23 (2012).
443. Puchkov, E. O. *et al.* The mycocidal, membrane-active complex of *Cryptococcus humicola* is a new type of cellobiose lipid with detergent features. *Biochim. Biophys. Acta - Biomembr.* **1558,** 161–170 (2002).
444. Ribeiro, I. A. C., Bronze, M. R., F. Castro, M. & Ribeiro, M. H. L. Selective recovery of acidic and lactonic sophorolipids from culture broths towards the improvement of their therapeutic potential. *Bioprocess Biosyst. Eng.* 1–13 (2016). doi:10.1007/s00449-016-1657-y
445. Rock, C. O. in *Biochemistry of Lipids, Lipoproteins and Membranes* 59–96 (2008).
446. Cooper, D. G. & Paddock, D. A. Production of a Biosurfactant from *Torulopsis bombicola*. *Appl. Environ. Microbiol.* **47,** 173–176 (1984).
447. Hommel, R. K. *et al.* Production of sophorose lipid by *Candida* (*Torulopsis*) *apicola* grown on glucose. *J. Biotechnol.* **33,** 147–155 (1994).
448. Najjar, V. A. The Isolation And Properties Of Phosphoglucomutase. *J. Biol. Chem.* **175,** 281–290 (1948).
449. Bro, C., Knudsen, S., Regenberg, B., Olsson, L. & Nielsen, J. Improvement of galactose uptake in *Saccharomyces cerevisiae* through overexpression of phosphoglucomutase: Example of transcript analysis as a tool in inverse metabolic engineering. *Appl. Environ. Microbiol.* **71,** 6465–6472 (2005).
450. Levander, F., Svensson, M. & Rådström, P. Enhanced exopolysaccharide production by metabolic engineering of *Streptococcus thermophilus*. *Appl. Environ. Microbiol.* **68,** 784–790 (2002).
451. Jones, P. *et al.* UGT73C6 and UGT78D1, Glycosyltransferases Involved in Flavonol Glycoside Biosynthesis in *Arabidopsis thaliana*. *J. Biol. Chem.* **278,** 43910–43918 (2003).
452. Räbinä, J. *et al.* Analysis of nucleotide sugars from cell lysates by ion-pair solid-phase extraction and reversed-phase high-performance liquid chromatography. *Glycoconj. J.* **18,** 799–805 (2001).
453. Pandey, R. P., Malla, S., Simkhada, D., Kim, B.-G. & Sohng, J. K. Production of 3-O-xylosyl quercetin in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **97,** 1889–1901 (2013).
454. Watt, G., Leoff, C., Harper, A. D. & Bar-Peled, M. A bifunctional 3,5-epimerase/4-keto reductase for nucleotide-rhamnose synthesis in *Arabidopsis*. *Plant Physiol.* **134,** 1337–1346 (2004).
455. Oka, T., Nemoto, T. & Jigami, Y. Functional analysis of *Arabidopsis thaliana* RHM2/MUM4, a multidomain protein involved in UDP-D-glucose to UDP-L-rhamnose Conversion. *J. Biol. Chem.* **282,** 5389–5403 (2007).
456. Ren, G. *et al.* Synthesis of flavonol 3-O-glycoside by UGT78D1. *Glycoconj. J.* **29,** 425–432 (2012).
457. Behmüller, R., Forstenlehner, I. C., Tenhaken, R. & Huber, C. G. Quantitative HPLC-MS analysis of nucleotide sugars in plant cells following off-line SPE sample preparation. *Anal. Bioanal. Chem.* **406,** 3229–37 (2014).
458. Spencer, J. F. T., Gorin, P. A. J. & Tulloch, A. P. *Torulopsis bombicola* sp. n. *Antonie Van Leeuwenhoek* **36,** 129–133 (1970).
459. Jenni, S., Leibundgut, M., Maier, T. & Ban, N. Architecture of a fungal fatty acid synthase at 5 A resolution. *Science (80-.).* **311,** 1263–1267 (2006).

460. *Global Biosurfactants Market Analysis And Segment Forecasts To 2020 - Biosurfactants Industry, Outlook, Size, Application, Product, Share, Growth Prospects, Key Opportunities, Dynamics, Trends, Analysis, Biosurfactants Report - Grand View Research Inc*. (2015)
461. Siewers, V. An overview on selection marker genes for transformation of *Saccharomyces cerevisiae*. *Methods Mol. Biol.* **1152,** 3–15 (2014).
462. Geurts, A. M. *et al.* Knockout Rats via Embryo Microinjection of Zinc-Finger Nucleases. *Science (80-.).* **325,** (2009).
463. Li, T. *et al.* TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and *Fok*I DNA-cleavage domain. *Nucleic Acids Res.* **39,** 359–72 (2011).
464. Smith, J. *et al.* A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.* **34,** e149 (2006).
465. Epinat, J.-C. *et al.* A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.* **31,** 2952–62 (2003).
466. Cade, L. *et al.* Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Res.* **40,** 8001–10 (2012).
467. Arnould, S. *et al.* Engineered I-CreI Derivatives Cleaving Sequences from the Human XPC Gene can Induce Highly Efficient Gene Correction in Mammalian Cells. *J. Mol. Biol.* **371,** 49–65 (2007).
468. Bao, Z. *et al.* Homology-Integrated CRISPR–Cas (HI-CRISPR) System for One-Step Multigene Disruption in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **4,** 585–594 (2015).
469. Vagner, S., Galy, B. & Pyronnet, S. Irresistible IRES. *EMBO Rep.* **2,** 1781–1790 (2001).
470. Komar, A. A. & Hatzoglou, M. Cellular IRES-mediated translation. *Cell Cycle* **10,** 229–240 (2011).
471. Kieft, J. S. Viral IRES RNA structures and ribosome interactions. *Trends Biochem. Sci.* **33,** 274–283 (2008).
472. Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E. & Hayakawa, T. IRES-Dependent Second Gene Expression Is Significantly Lower Than Cap-Dependent First Gene Expression in a Bicistronic Vector. *Mol. Ther.* **1,** 376–382 (2000).
473. Halpin, C., Cooke, S. E., Barakate, A., Amrani, A. El & Ryan, M. D. Self-processing 2A-polyproteins - a system for co-ordinate expression of multiple proteins in transgenic plants. *Plant J.* **17,** 453–459 (1999).
474. de Amorim Araújo, J. *et al.* Coexpression of cellulases in *Pichia pastoris* as a self-processing protein fusion. *AMB Express* **5,** 84–93 (2015).
475. Burén, S. *et al.* Use of the Foot-and-Mouth Disease Virus 2A Peptide Co-Expression System to Study Intracellular Protein Trafficking in *Arabidopsis*. *PLoS One* **7,** e51973 (2012).
476. Kim, J. H. *et al.* High Cleavage Efficiency of a 2A Peptide Derived from *Porcine* Teschovirus-1 in Human Cell Lines, Zebrafish and Mice. *PLoS One* **6,** e18556 (2011).
477. Trichas, G., Begbie, J. & Srinivas, S. Use of the viral 2A peptide for bicistronic expression in transgenic mice. *BMC Biol.* **6,** 40–52 (2008).
478. Kandel, S. E. & Lampe, J. N. Role of protein-protein interactions in cytochrome P450-mediated drug metabolism and toxicity. *Chem. Res. Toxicol.* **27,** 1474–86 (2014).
479. Govindaraj, S. & Poulos, T. L. Role of the linker region connecting the reductase and heme domains in cytochrome P450 BM3. *Biochemistry* **34,** 11221–11226 (1995).
480. Govindaraj, S. & Poulos, T. L. Probing the structure of the linker connecting the reductase and heme domains of cytochrome P450BM-3 using site-directed mutagenesis. *Protein Sci.* **5,** 1389–93 (1996).
481. Neeli, R. *et al.* The dimeric form of flavocytochrome P450 BM3 is catalytically functional as a fatty acid hydroxylase. *FEBS Lett.* **579,** 5582–5588 (2005).
482. Munro, A. W., Girvan, H. M. & McLean, K. J. Cytochrome P450-redox partner fusion enzymes. *Biochim. Biophys. Acta - Gen. Subj.* **1770,** 345–359 (2007).
483. Belsare, K. D. *et al.* P-LinK: A method for generating multicomponent cytochrome P450 fusions with variable linker length. *Biotechniques* **57,** 13–20 (2014).
484. Bakkes, P. J. *et al.* Design and improvement of artificial redox modules by molecular fusion of flavodoxin and flavodoxin reductase from Escherichia coli. *Nat. Publ. Gr.* 1–13 (2015).
485. Thoden, J. B., Frey, P. A. & Holden, H. M. Molecular Structure of the NADH/UDP-glucose Abortive Complex of UDP-galactose 4-Epimerase from *Escherichia coli*: Implications for the Catalytic Mechanism†,‡. *Biochemistry* **35,** 5137–5144 (1996).

Summary

Summary

Biosurfactants, surfactant molecules originating from microbial synthesis and/or renewable resources, have been proven to be a valuable alternative to the (petro)chemical based variants widely used today. Due to the structural variability of these molecules, a wide variety of (potential) applications ranging from household usage to pharmaceuticals have been identified. Thanks to their biological degradability and production using renewable resources combined with their (biological) activity even at low concentrations, the focus of surfactant production is slowly turning towards these molecules. Still, widespread usage of these fascinating molecules remains low mostly due to the higher production cost and limited structural variability of the commercialised compounds. When looking at glycolipids, carbohydrate or lipid variety is often limited to just a handful of different types of groups with minor modifications. In the case of sophorolipids, the fatty acid tail is most often limited to a C18 fatty acid, either with its carboxylic group free or condensed to the sophorose moiety. Longer or shorter congeners are possible, but only in low amounts or as a part of a more complex mixture of molecules.

When taking productivities into account, only a few microorganisms are capable of producing enough biosurfactant in a short amount of time to reduce the cost of making them. Unfortunately, one has to take into account certain drawbacks like potential pathogenicity of the producer strain. Using robust producer strains for the production of wild type, modified or new-to-nature molecules has been proven to be an interesting technique to enhance the production of these molecules. Both on a genetic and process level optimisation has been done with varying degrees of success.

During this thesis, *S. bombicola* was turned into a robust strain for the production of different kinds of molecules. Both genetic engineering and better understanding the production parameters allowed to gain deeper insights in how *S. bombicola* produces biosurfactants and how to steer the production towards novel kinds of molecules.

In the second chapter, the main focus was creating new molecular tools for easier modification and/or screening of novel strains. A first tool was the development of an episomal vector for *S. bombicola*. Though many yeast species have one or several plasmids, no such thing is known for *S. bombicola*. By using *in silico* characterisation techniques and using heterologous and homologous sequences in a suitable screenings vector, the aim was to identify ARS sequences. Unfortunately, no active sequence could be found. The second tool is the maximum size of the integration cassettes used for knock-in and -out strategies. As it became clear that no clear upper limit could be found up until 11600 base pairs, a more practical was followed to double the genetic cluster of sophorolipid production. Though the idea was that more expression would result in more production, the results were not straight forward pointing towards higher transcription, translation and final productivity. Still, it proved that large constructs can be designed and transformed efficiently in *S. bombicola*, opening the possibility of introducing complex multistep pathways in a few steps. The final tool can be seen as multiple smaller tools together. Firstly, new kinds of fluorescent proteins were tested successfully in *S. bombicola* besides the already available GFP. Secondly, it was shown that coupling multiple domains does not necessarily impair protein function. This opens the possibility of creating larger chimeric enzymes where the catalytic domains are coupled for enhanced turnover.

Thirdly, using the sophorolipid biosynthetic enzymes for the coupling resulted in their cellular localisation and provided a tool for comparing productivity with expression levels of the enzymes involved. When integrating this knowledge with qPCR data, better fine tuning of gene expression levels can be achieved.

Chapter three and four both focus on modifying either the carbohydrate or the lipidic group. In chapter three, is was proven that introducing novel P450s in *S. bombicola* can be an interesting strategy to alter the fatty acids being incorporated. Several self-sufficient P450 were introduced in *S. bombicola* for the production of new kinds of sophorolipids. Their activity could be measured, but no new molecules were produced. Engineering of the self-sufficient P450s by altering the critical phenylalanine at position 87 did not result in more favourable intermediates being produced. Still, they can be used as a starting point for other strategies. One of those was explored in depth by creating self-sufficient chimeric variants of CYP52A4 and CYP1. Though the production levels remained low, a high product uniformity was achieved for the CYP1BMR strain. Further optimisation including the substrate used resulted the production of C16 sophorolipids, both acidic and lactonic.

In chapter four, the main concept was modification of the carbohydrate group. The production levels of glucolipids and cellobiose lipids produced by previously designed strains remained low. Redesigning the original constructs as well as using optimized engineering strains resulted in the production of glucolipids up to 130 g/L. Interestingly, new molecules were observed as well in the produced mixture. It was already known that the UGTA1 transferase is capable of coupling a glucose moiety to both hydroxyl and carboxyl groups, the observation of bola glucolipids remained elusive until the first growth trials with the newly engineered glucolipid strains. For the cellobiose lipids, higher product uniformity was achieved by only expressing the *ugt1* gene from *U. maydis* in *S. bombicola*. This not only resulted in losing the contaminating glucolipid fraction from the older strains, it also showed that UGT1 is solely responsible for the sequential glucosylation of the hydroxy palmitic acid. Further optimisation of the production strains is necessary to produce these molecules in bigger volumes, but for the first time relatively uniform cellobiose lipids have been produced in a non-wild type producer.

Chapter five was about turning *S. bombicola* into a producer of molecules that are unknown to this yeast. Rhamnolipids are molecules that originate from a bacterial producer and are structurally not similar to sophorolipids or cellobiose lipids. Production of these molecules involved integrating several bacterial pathways for the production of the necessary precursors. To obtain the hydrophobic part, the genes necessary for the bacterial FAS II system were introduced at two different locations in the genome. As a proof-of-concept, the production of HAA molecules was attempted. After several growth trials, it became clear to no activity could be measured. For the production of activated rhamnose, two different pathways were designed resulting in dTDP-L-rhamnose or UDP-L-rhamnose. To assess the availability of these rhamnose donors, a proof-of-concept was designed around the rhamnosylation of quercetin by the transferase At78D1 from *Arabidopsis thaliana*. After a growth trial, it became clear that only the strain producing dTDP-L-rhamnose was capable of producing a rhamnosylated product. A final experiment was carried out by combining cell lysates of both the NDP-strains equipped with the second rhamnosyltransferase RhlB and the HAA-strain. No production of rhamnolipids or related molecules could be observed.

In conclusion, utilizing the robustness of *S. bombicola* for creating a platform for biosurfactant production is a valuable strategy. Though several obstacles remain concerning the fundamental knowledge of the organism, it is already possible to engineer strains that are capable of producing novel molecules on scales relevant for industrial production. The results in this work show that *S. bombicola* is a relevant player in the field of glycolipid production.

Samenvatting

Samenvatting

Biosurfactanten, surfactant moleculen geproduceerd door micro-organismen of vanuit hernieuwbare grondstoffen, zijn een evenwaardig alternatief voor de tegenwoordig veelgebruikte (petro)chemische surfactanten. Een veelvoud aan potentiele toepassingen zijn reeds bekend gaande van huishoudelijke tot farmaceutische applicaties zijn al reeds bekend. Andere aspecten zoals hun biodegradeerbaarheid, hernieuwbare productie en potentiele biologische activiteit hebben ervoor gezorgd dat de interesse naar deze moleculen gevoelig gestegen is. Ondanks deze voordelen zijn nog enkele grote nadelen verbonden aan deze verbindingen. Zo is de productiekost nog steeds significant hoger en is de structurele variatie van de reeds gecommercialiseerde moleculen beperkt.

Een interessante deelklasse van de biosurfactanten, namelijk de glycolipiden, werden reeds uitvoerig bestudeerd. Men dient echter in gedachten te houden dat deze klasse van moleculen slechts een handvol variaties kent voor zowel het koolhydraat als het vetgedeelte van de glycolipide. Een specifiek voorbeeld zijn de sophorolipiden. Alle moleculen van deze subklasse zijn samengesteld uit een sophorose en een C18 vetzuur. Eventuele verzadiging van de C18 staart of de locatie van de hydroxylatie kunnen variëteit veroorzaken maar ketens met langere of kortere varianten zijn quasi onbestaande. De enige uitzondering hiervoor zijn de moleculen van *R. bogoriensis* die standaard een C22 vetzuurketen hebben. Het is echter niet onmogelijk afwijkende vetzuren of analoge moleculen te incorporeren maar vaak leidt dit tot een meer complex mengel moleculen of tot lagere productieniveaus. Verder is het belangrijk de productiecapaciteit van bepaalde organismen niet te vergeten. Slechts enkele zijn in staat tot hoge en snelle productie van biosurfactanten zoals *P. aerugiosa* of *S. bombicola*. Nadelen zoals potentiele pathogeniciteit kunnen echter een grote rol spelen in de kostprijs en applicaties van bepaalde moleculen.

Het gebruik maken van robuuste micro-organismen voor de productie van eigen wild-type, gemodificeerde wild-type of nieuwe types van moleculen kan een interessante strategie zijn om enerzijds de kostprijs van biosurfactanten te laten dalen en anderzijds om de moleculaire diversiteit te vergroten. Dit is in het verleden als reeds enkele malen toegepast op kleine schaal.

Gedurende dit eindwerk werd getracht *S. bombicola* om te vormen tot een robuuste producent van verscheidene moleculen. Nieuwe inzichten op het vlak van genetische modificaties alsook op procesniveau zorgden voor een diepere kennis betreffende de productie van deze moleculen.

In hoofdstuk twee werden verscheidene nieuwe technieken getest voor het modificeren en opvolgen van *S. bombicola*. Een eerste techniek bestond uit het ontwikkelen van een plasmide voor *S. bombicola*. Hoewel verscheidene micro-organismen een of meerdere types van plasmide hebben is er tot op heden geen bekend voor *S. bombicola*. Door gebruik te maken van *in silico* screeningstechnieken alsook het uittesten van enkele homologe en heterologe sequenties werd getracht een gelijkaardig systeem te ontwikkelen maar tot op heden zonder succes. De tweede techniek betreft de grootte van de recombinatie cassettes. Hoewel slechts enkele groottes getest zijn werd snel duidelijk dat voor het gros van de modificaties geen reële limiet bestaat. Vanuit een meer praktisch oogpunt werd besloten om de gencluster van de sophorolipiden synthese te verdubbelen. Hoewel initieel verwacht werd dat het hogere kopij

aantal van de genen tot verhoogde transcriptie, translatie en productie zou leiden, werd snel duidelijk dat dit niet het geval was. Desalniettemin was het duidelijk dat grote cassettes gebruikt kunnen worden voor het efficiënt introduceren van grote en/of complexe metabolische netwerken in een beperkt aantal tussenstappen. De laatste tool bestond uit het introduceren van fluorescente eiwitten en kan beschouwd worden als meerdere kleine tools tezamen. Allereerst werden twee nieuwe fluorescente eiwitten correct tot expressie gebracht in *S. bombicola*. Dit opent mogelijkheden met meerdere promotoren of localisatiesignalen tegelijkertijd te testen. Ten tweede werd aangetoond dat het koppelen van extra domeinen aan bepaalde enzymen niet zorgt voor een daling in activiteit. Dit kan in de toekomst verder benut worden door het koppelen van andere katalytische elementen tot grote chimere eiwitten met een verhoogd omzettingsvermogen. Tot slot zorgde het koppelen van de fluorescente eiwitten aan deze enzymen van de sophorolipiden synthese voor enerzijds het lokaliseren van deze moleculen binnenin de cel en anderzijds voor een visuele tool om expressieniveaus te kunnen meten en tegelijkertijd te koppelen aan productiviteit. Wanneer deze data gekoppeld wordt met qPCR data kan een beter begrip van de eiwitexpressie bekomen worden en kan deze beter afgesteld worden om een zo min mogelijke impact te hebben op celvitaliteit.

Hoofdstukken drie en vier focusten beide op de modificatie van een bepaald deelaspect van de structuur van glycolipiden. Hoofdstuk drie draait om het veranderen van de vetzuurketen door middel van expressie van P450s enzymen. Het vervangen van het celeigen CYP52M1 door andere, eventueel zelfvoorzienende, P450s bleek een interessante doch niet succesvolle manier te zijn om nieuwe compatibele hydroxy vetzuren te produceren en incorporeren in glycolipiden. Hun activiteit was echter eenvoudig meetbaar wat bevestigde dat heterologe expressie van P450s functioneel kan zijn. Het muteren van enkele zelfvoorzienende P450s zorgde eveneens niet voor de gehoopte productie. De reden hiervoor kan echter gezocht worden in enerzijds een lagere activiteit van de mutant en anderzijds de productie van minder compatibele intermediairen. Het tweede deel van hoofdstuk drie was gericht op het creëren van zogenaamde chimere, zelfvoorzienende P450s. Enkele varianten zoals deze met CYP52A4 en CYP1 bleken interessante eigenschappen te hebben voor de productie van nieuwe types sophorolipiden. Voor het CYP1BMR eiwit bleek het mogelijk specifiek C16 sophorolipiden te produceren. Verdere optimalisatie resulteerde in hogere productiviteit van deze moleculen.

In hoofdstuk 4 lag de focus op het modificeren van de koolhydraatgroep. Productie van glucolipiden en cellobiose lipiden was al reeds mogelijk dankzij eerder geconstrueerde stammen maar de behaalde opbrengsten waren laag in vergelijking met de productie van sophorolipiden. Verder was de productuniformiteit niet optimaal. Het opnieuw ontwerpen van de originele recombinatiecassettes zorgde in het geval van de glucolipiden in een significant hogere productie. Hoewel deze stijging grotendeels te danken was aan het gebruiken van een specifiek ontworpen *ura3* auxotrofe stam van *S. bombicola* bleek er een groot verschil te zijn in de verschillende cassettes en hun effect op de productie. Een andere interessante waarneming waren de bolaglucolipiden. Het was al reeds bekend dat UGTA1 in staat is om zowel hydroxyl als carboxylgroepen te glucosyleren *in vivo*. Voor de productie van cellobiose lipiden werd een hogere product uniformiteit behaald door enkel het *ugt1* gen van *U. maydis* te expresseren in een *ugta1* knock-out strain van *S. bombicola*. Hierdoor werden er geen glucolipiden meer geproduceerd en kon definitief aangetoond worden dat enkel het UGT1 verantwoordelijk is voor de dubbele glucosylering van het hydroxy palmitinezuur. Verdere optimalisatie van de productiestammen is echter nodig voor het behalen van hogere titers.

Het laatste hoofdstuk gaat over het aanwenden van *S. bombicola* voor het produceren van rhamnolipiden. Rhamnolipiden zijn moleculen die structureel niet verwant zijn aan de sophorolipiden en hun synthese verloopt verschillend dan deze van de sophorolipiden. Om productie te bekomen was het noodzakelijk om bacteriële metabolische processen te integreren in *S. bombicola* aangezien de celeigen eukaryote netwerken niet flexibel genoeg of toereikend waren. De integratie van een type II vetzuursynthese werd ondernomen door de vereiste genen te integreren op twee verschillende locaties in het genoom. Als *proof-of-concept* molecule werd getracht om dimeren van β-hydroxy vetzuren te bekomen. Na enkele groeiproeven werd duidelijk dat er geen activiteit gemeten kan worden. Voor de geactiveerde rhamnose werden twee pathways in parallel getest, één specifiek voor UDP-L-rhamnose, de andere voor dTDP-L-rhamnose. Om te beschikbaarheid van het geactiveerde rhamnose te testen werd een poging ondernomen voor het rhamnosyleren van quercetine door het transferase At78D1 van *Arabidopsis thaliana*. Een groefproef wees uit dat de stam uitgerust met de pathway voor dTDP-L-rhamnose productie in staat was om een gerhamnosyleerd product te vormen. Een finaal experiment waarbij cellysaten van beide precursorstammen gecombineerd werden leidde uiteindelijk niet tot de productie van rhamnolipiden of andere verwante moleculen.

Als conclusie kan met stellen dat *S. bombicola* de nodige robuustheid bezit om te dienen als productieplatform voor verscheidenen types biosurfactanten. Hoewel enkele obstakels nog aanwezig zijn, zowel op een fundamenteel als een meer procesmatige niveau, is gebleken dat het reeds mogelijk is om industrieel relevante giststammen te ontwikkelen. De resultaten van dit werk vertellen duidelijk dat *S. bombicola* een relevant organisme is voor glycolipidenproductie.

Curriculum vitae

Curriculum vitae

Personal Information

Name: Robin Geys
Date of Birth: 8th May 1987
Place of Birth: Kapellen
Nationality: Belgian
Phone number: +32 494 90 88 90
Email: robingeys@gmail.com

Languages:
- Dutch (Native)
- English (Fluent)
- French (Fluent)
- German (Intermediate)

Education

Jan 2013 – Dec 2016	Phd, Ghent University, Faculty of Bioscience Engineering Centre for Industrial Biotechnology and Biocatalysis PhD thesis: "Engineering the metabolism of *Starmerella bombicola* for the production of tailor-made glycolipids." Promotor: Prof. Dr. Ir. Wim Soetaert Co-promotor: Prof. Dr. Ir. Inge Van Bogaert
Sep 2010 – Jul 2012	Master of Bioscience Engineering: Cell and Gene Biotechnology Ghent University, Faculty of Bioscience Engineering, Ghent, Belgium Master thesis: "Engineering van de P450 enzympopulatie van Candida bombicola voor de productie van new-to-nature surfactanten" Promotor: Prof. Dr. Ir. Wim Soetaert Tutor: Prof. Dr. Ir. Inge Van Bogaert
Feb 2008 – Jul 2010	Bachelor Bioscience engineering University of Antwerp, Antwerp, Belgium
Sep 2006 - Feb 2008	Bachelor Biomedical Sciences University of Antwerp, Antwerp, Belgium
Sep 1999 – Jul 2006	Sint-Michiels College Brasschaat, Modern Languages – Science

Courses attended in the framework of the doctoral school program (Bioscience Engineering)

- Ghent Biobased Economy Summer School (2013)
- 7th From PhD to Job Market (To Company) (2014)
- Advanced Academic English: Conference Skills – Academic Posters (2014)
- N2N Multidisciplinary Seminar Series on Bioinformatics (2015)
- qPCR (2016)

Personal interests

- Photography
- Traveling
- Astronomy
- Sports (climbing and cycling)

Publications

Published

Geys, R., Soetaert, W., Van Bogaert I., 2014. Biotechnological opportunities in biosurfactant production. Current Opinion in Biotechnology, Volume 30, p66-72

In Preparation

Research Papers

Roelants, S., Ciesielska, K., De Maeseneire, S., Lodens, S., Coussement, P., **Geys, R.**, Saerens, K., Saey, L., Verweire, S., Van Bogaert, I., Devreese, B., Soetaert, W., 2016. Development and application of a molecular toolkit for the industrial yeast *Starmerella bombicola*. submitted Metabolic Engineering

Attended Conferences and Workshops

Enabling Technologies for Eukaryotic Synthetic Biology (21st – 23th June 2015). Poster presentation: "Creation of a fungal production platform for biosurfactants by P450 engineering". **Geys, R.**, Lemmens, C., De Vos, G., Van Bogaert, I., Soetaert, S. Flash Presentation: "Creation of a fungal production platform for biosurfactants by P450 engineering"

Frontiers in Fungal Systems Biology (28th-30th September 2014). Poster presentation: "Creation of a fungal production platform for biosurfactants by P450 engineering". **Geys, R.**, De Smet, M., Van de Velde, I., Van Bogaert, I., Soetaert, S.

19th National Symposium on Applied Biological Sciences (7th Fabruary 2014). Poster presentation: "Promotor screening with fluorescent proteins for the yeast Starmerella bombicola". **Geys, R.**, Roelants, S., De Maeseneire, S., Soetaert, W.

Ghent Biobased Economy Summer School on Integration of Green & White Biotechnology: From Plant to Product (19th-22nd august 2013). Poster presentation: "Genetic engineering of Starmerella bombicola for the production of cellobiose lipids". **Geys, R.**, D'Hondt, Y., Van Bogaert, I., Soetaert, W.

Yeasterday 2013 (24th May 2013). Poster presentation: "Genetic engineering of Starmerella bombicola for the production of cellobiose lipids". **Geys, R.**, D'Hondt, Y., Van Bogaert, I., Soetaert, W.

Student guidance

Practical exercises: General Microbiology (2013-2016)

Tutor of bachelor and master theses

- Yaika D'Hondt, Bachelor Thesis (2013)
- Anouk van Nassauw, Bachelor Thesis (2013)
- Laurens De Ras, M. Sc. Thesis (2014)
- Belén Adiego Pérez, Bachelor Thesis (2014)
- Margaux De Smet, M. Sc. Thesis (2015)
- Sylwia Jezierska, M. after Master (2015)
- Julius Ramboer, Bachelor Thesis (2015) Did not graduate
- Gwen De Vos, Bachelor Thesis (2015)
- Christophe Lemmens, Bachelor Thesis (2015)
- Jeroen Van Malderen, M. Sc. Thesis (2016)
- Niels Mariën, M. Sc. Thesis (2016) Did not graduate
- Jelle Remmery, M. Sc. Thesis (2017)

Appendices

Appendices

Appendix A

Codon usage tables used for codon optimization of synthetic sequences.

10 Genes				Stationary Genes			
AAA	0.62%	GAA	0.71%	AAA	1.42%	GAA	1.71%
AAC	4.14%	GAC	3.19%	AAC	1.33%	GAC	1.17%
AAG	4.14%	GAG	4.04%	AAG	1.84%	GAG	2.14%
AAT	0.62%	GAT	1.57%	AAT	1.48%	GAT	1.67%
ACA	0.57%	GCA	0.38%	ACA	1.68%	GCA	1.60%
ACC	1.95%	GCC	1.81%	ACC	1.25%	GCC	1.61%
ACG	0.24%	GCG	0.24%	ACG	1.15%	GCG	1.28%
ACT	2.00%	GCT	2.28%	ACT	1.54%	GCT	2.03%
AGA	0.29%	GGA	0.71%	AGA	1.78%	GGA	1.53%
AGC	0.86%	GGC	2.05%	AGC	1.90%	GGC	1.61%
AGG	0.24%	GGG	0.10%	AGG	1.57%	GGG	1.06%
AGT	0.29%	GGT	1.90%	AGT	1.24%	GGT	1.33%
ATA	0.10%	GTA	0.24%	ATA	1.03%	GTA	0.73%
ATC	2.85%	GTC	2.47%	ATC	1.69%	GTC	1.23%
ATG	4.76%	GTG	0.43%	ATG	2.00%	GTG	1.59%
ATT	1.81%	GTT	1.62%	ATT	1.73%	GTT	1.59%
CAA	0.76%	TAA	4.19%	CAA	2.16%	TAA	0.67%
CAC	3.85%	TAC	4.04%	CAC	1.33%	TAC	1.16%
CAG	4.00%	TAG	0.62%	CAG	2.05%	TAG	0.61%
CAT	0.90%	TAT	0.71%	CAT	1.49%	TAT	1.27%
CCA	0.48%	TCA	0.29%	CCA	1.74%	TCA	2.38%
CCC	1.86%	TCC	1.47%	CCC	1.30%	TCC	1.26%
CCG	0.62%	TCG	0.52%	CCG	1.24%	TCG	1.59%
CCT	1.86%	TCT	1.28%	CCT	1.47%	TCT	2.01%
CGA	0.29%	TGA	0.00%	CGA	1.58%	TGA	2.02%
CGC	1.14%	TGC	3.47%	CGC	1.35%	TGC	1.90%
CGG	0.05%	TGG	4.76%	CGG	1.14%	TGG	2.36%
CGT	2.85%	TGT	1.28%	CGT	1.05%	TGT	1.43%
CTA	0.14%	TTA	0.10%	CTA	1.06%	TTA	1.10%
CTC	2.19%	TTC	3.95%	CTC	1.98%	TTC	2.14%
CTG	0.57%	TTG	0.38%	CTG	2.31%	TTG	2.42%
CTT	1.38%	TTT	0.81%	CTT	2.05%	TTT	1.88%

Appendix B

Optimized coding sequences of *gfpco10genes*, *bfpcostat* and *rfpco10stat.*

gfpco10genes

ATGTCGAAAGGAGAGGAGCTGTTCACTGGCGTCGTCCCAATCCTGGTCGAACTTGATGGCGACG
TTAACGGTCACAAGTTTAGCGTGAGCGGTGAAGGTGAGGGTGATGCTACCTACGGCAAGCTGAC
TCTTAAGTTCATCTGCACCACTGGCAAGCTTCCAGTTCCATGGCCCACACTCGTTACAACCTTCG
GCTATGGCGTTCAGTGCTTTGCCCGTTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCG
GCCATGCCTGAGGGCTATGTTCAGGAACGTACTATCTTCTTCAAGGACGACGGCAACTACAAAAC
TCGTGCTGAAGTCAAGTTTGAGGGCGATACTTTGGTAAACCGTATCGAGCTCAAGGGCATTGACT
TCAAGGAGGACGGTAACATTCTTGGTCATAAGCTTGAGTATAATTACAACTCTCACAACGTCTACA
TTATGGCTGACAAGCAGAAGAATGGAATTAAAGTCAACTTCAAGATTCGCCATAACATCGAGGAC
GGCAGCGTACAGCTCGCTGACCACTACCAACAGAACACTCCCATTGGCGACGGCCCTGTTCTTC
TCCCTGACAACCACTATCTATCTACCCAGTCCGCCTTGTCCAAAGACCCCAACGAGAAGCGCGAC
CACATGGTTCTTTTGGAGTTCGTTACTGCCGCTGGCATCACGCACGGAATGGACGAGCTGTACAA
ATAA

bfpcostat

ATGTCGAAAGGAGAGGAGCTGTTCACTGGCGTCGTCCCAATCCTGGTCGAACTTGATGGCGACG
TTAACGGTCACAAGTTTAGCGTGAGCGGTGAAGGTGAGGGTGATGCTACCTACGGCAAGCTGAC
TCTTAAGTTCATCTGCACCACTGGCAAGCTTCCAGTTCCATGGCCCACACTCGTTACAACCTTCA
GCCACGGCGTTCAGTGCTTTAGCCGTTACCCCGACCACATGAAGCGGCACGACTTCTTCAAGTC
GGCCATGCCTGAGGGCTATGTTCAGGAACGTACTATCTCCTTCAAGGACGACGGCAACTACAAAA
CTCGTGCTGAAGTCAAGTTTGAGGGCGATACTTTGGTAAACCGTATCGAGCTCAAGGGCATTGAC
TTCAAGGAGGACGGTAACATTCTTGGTCATAAGCTTGAGTATAATTACAACTCTCACAACGTCTAC
ATTACGGCTGACAAGCAGAAGAATGGAATTAAAGCAAACTTCAAGATTCGCCATAACATCGAGGA
CGGCAGCGTACAGCTCGCTGACCACTACCAACAGAACACTCCCATTGGCGACGGCCCTGTTCTT
CTCCCTGACAACCACTATCTATCTACCCAGTCCGCCTTGTCCAAAGACCCCAACGAGAAGCGCGA
CCACATGGTTCTTTTGGAGTTCGTTACTGCCGCTGGCATCACGCACGGAATGGACGAGCTGTACA
AATAA

rfpco10stat

ATGAGGTCTTCCAAGAACGTTATCAAGGAATTTATGCGCTTCAAGGTCCGCATGGAAGGTACTGT
TAACGGTCACGAGTTCGAGATTGAGGGAGAGGGTGAGGGTCGTCCCTACGAGGGCCACAACAC
TGTCAAGCTCAAGGTTACCAAGGGTGGTCCTCTTCCATTCGCATGGGACATTCTTTCCCCTCAGT
TCCAGTATGGCTCCAAGGTCTACGTCAAGCACCCCGCAGACATCCCCGATTACAAGAAACTTTCC
TTCCCTGAAGGCTTCAAGTGGGAGCGTGTGATGAACTTCGAGGACGGGGGCGTCGTTACAGTTA
CGCAGGATAGCTCTCTCCAAGACGGTTGCTTCATCTATAAGGTCAAGTTCATCGGTGTCAACTTC
CCCTCAGACGGTCCCGTCATGCAAAAGAAGACGATGGGTTGGGAAGCCAGCACAGAGCGCCTTT
ACCCCCGTGACGGTGTCCTTAAGGGTGAAATTCACAAGGCACTCAAGCTGAAGGATGGAGGCCA
CTACCTCGTCGAGTTCAAGTCGATTTATATGGCTAAGAAGCCTGTTCAGCTCCCTGGTTACTACTA
CGTCGACTCGAAGCTGGACATCACATCGCACAATGAGGATTACACGATCGTTGAGCAGTACGAA
CGTACCGAGGGCCGTCACCACCTCTTCCTCTAG

Appendix C

Ordered gBlock for *cyp1bmr* validation. Mutated codons are in bold font and underlined.

TGCTGATGAAGTAAAAGGCGTTCGCTACTCCGTATTT**GAT**TGCGGCGATAAAAACTGGG
CTACTACGTATCAAAAAGTGCCTGCTTTTATCGATGAAACGCTTGCCGCTAAAGGGGCA
GAAAACATCGCTGACCGCGGTGAAGCAGATGCAAGCGACGACTTTGAAGGCACATATG
AAGAATGGCGTGAACATATGTGGAGTGACGTAGCAGCCTACTTTAACCTCGACATTGAA
AACAGTGAAGATAATAAATCTACTCTTTCACTTCAATTTGTCGACAGCGCCGCGGATATG
CCGCTTGCGAAAATGCACGGTGCGTTTTCAACGAACGTCGTAGCAAGCAAAGAACTTCA
ACAGCCAGGCAGTGCACGAAGCACGCGACATCTTGAAATTGAACTTCCAAAAGAAGCTT
CTTATCAAGAAGGAGATCATTTAGGTGTTATTCCTCGCAACTATGAAGGAATAGTAAACC
GTGTAACAGCAAGGTTCGGCCTAGATGCATCACAGCAAATCCGTCTGGAAGCAGAAGA
AGAAAAATTAGCTCATTTGCCACTCGCTAAAACAGTATCCGTAGAAGAGCTTCTGCAATA
CGTGGAGCTTCAAGATCCTGTTACGCGCACGCAGCTTCGCGCAATGGCTGCTAAAACG
GTCTGCCCGCCGCATAAAGTAGAGCTTGAAGCCTTGCTTGAAAAGCAAGCCTACAAAGA
ACAAGTGCTGGCAAAACGTTTAACAATGCTTGAACTGCTTGAAAAATACCCGGCGTGTG
AAATGAAATTCAGCGAATTTATCGCCCTTCTGCCAAGCATACGCCCGCGCTATTACTCG
ATTTCTTCATCACCTCGTGTCGATGAAAAACAAGCAAGCATCACGGTCAGCGTTGTCTCA
GGAGAAGCGTGGAGCGGATATGGAGAATATAAAGGAATTGCGTCGAACTATCTTGCCG
AGCTGCAAGAAGGAGATACGATTACGTGCTTTATTTCCACACCGCAGTCAGAATTTACG
CTGCCAAAAGACCCTGAAACGCCGCTTATCATGGTCGGACCGGGAACAGGCGTCGCGC
CGTTTAGAGGCTTTGTGCAGGCGCGCAAACAGCTAAAAGAACAAGGACAGTCACTTGG
AGAAGCACATTTATACTTCGGCTGCCGTTCACCTCATGAAGACTATCTGTATCAAGAAGA
GCTTGAAAACGCCCAAAGCGAAGGCATCATTACGCTTCATACCGCTTTTTCTCGCATGC
CAAATCAGCCGAAAACATACGTTCAGCACGTAATGGAACAAGACGGCAAGAAATTGATT
GAACTTCTTGATCAAGGAGCGCACTTCTATATT**GCC**GGAGACGGAAGCCAAATGGCACC
TGCCGTTGAAGCAACG

Appendix D

Ordered gBlocks with coding sequence of the *A. thaliana atnrs/er.* Due to the big size of the coding sequence and the limit of IDT for a gBlock at the time, two separate gBlocks were ordered. At the 5' of gBlock 1 and 3' of gBlock 2 homologous regions are present for assembly in a pJET vector (underlined). Both gBlocks also carry homologues regions for each other to enable easy assembling (bold).

gBlock 1

GTTTTTCAGCAAGATATGGTTGCAGACGCAAACGGTTCATCATCAAGCTCATTTAACTTC
CTAATCTACGGTAAAACCGGATGGATCGGTGGTTTACTCGGTAAACTCTGCGAAGCTCA
AGGAATCACTTACACTTACGGCTCCGGTCGTCTTCAAGATCGTCAATCGATCGTCGCCG
ACATCGAATCCGTGAAACCTAGCCACGTGTTCAACGCTGCTGGAGTCACCGGTCGTCCT
AATGTTGATTGGTGCGAATCTCACAAAGTTGAGACGATTCGTACTAATGTCGCCGGAAC
CCTAACTCTCGCTGACATTTGCAGAGAGAAAGGACTTGTTCTGATCAATTACGCTACGG
GTTGTATATTTGAGTATGATTCGGGTCATCCTCTCGGGTCGGGTATTGGATTCAAGGAG
GAGGATACTCCTAATTTCACCGGATCTTTCTACTCTAAAAC**CAAAGCTATGGTGGA**

gBlock 2

CAAAGCTATGGTGGAGGAGCTGCTCAAGAACTATGAAAATGTATGCACGCTAAGAGTG
CGAATGCCGATTTCATCGGATCTAACAAACCCGAGAAACTTCATCACGAAGATTGCTCG
GTATGAGAAAGTTGTGGACATCCCAAACTCGATGACAATCCTCGATGAGCTTCTCCCGA
TATCAATCGAAATGGCGAAGAGGAACTTAACCGGGATCTACAATTTCACTAACCCGGGT
GTTGTGAGCCACAACGAGATCTTGGAGATGTACAGAGACTACATTGACCCGAGTTTTAC
TTGGAAGAACTTCACATTGGAGGAACAAGCTAAAGTGATTGTGGCGCCAAGGAGTAACA
ATGAGCTTGATGCAACTAAGTTGAAGACTGAGTTCCCTGAGTTGATGTCTATCAAAGAGT
CTCTGATCAAGTTCGTGTTTGAGCCCAACAAGAAGACTGAAGTTAAAGCTTGAATCTTTC
TAGAAGAT

Ordered gBlocks with coding sequence of the *A. thaliana rhm2/mum4*. Due to the big size of the coding sequence and the limit of IDT for a gBlock at the time, two separate gBlocks were ordered. At the 5' of gBlock 1 and 3' of gBlock 2 homologous regions are present for assembly in the expression vector for the *cyp51m1* locus (underlined). Both gBlocks also carry homologues regions for each other to enable easy assembling (bold).

gBlock 1

CGTTATTTCTCCGTTTATCATATTGAAAAGTACATATATGGATGATACTACGTATAAGCCA
AAGAACATTCTCATTACTGGAGCTGCTGGATTTATTGCTTCTCATGTTGCCAACAGATTA
ATCCGTAACTATCCTGATTACAAGATCGTTGTTCTTGACAAGCTTGATTACTGTTCAGAT
CTGAAGAATCTTGATCCTTCTTTTTCTTCACCAAATTTCAAGTTTGTCAAAGGAGATATCG
CGAGTGATGATCTCGTTAACTACCTTCTCATCACTGAAAACATTGATACGATAATGCATT
TTGCTGCTCAAACTCATGTTGATAACTCTTTTGGTAATAGCTTTGAGTTTACCAAGAACAA
TATTTATGGTACTCATGTTCTTTTGGAAGCCTGTAAAGTTACAGGACAGATCAGGAGGTT
TATCCATGTGAGTACCGATGAAGTCTATGGAGAAACCGATGAGGATGCTGCTGTAGGAA
ACCATGAAGCTTCTCAGCTGTTACCGACGAATCCTTACTCTGCAACTAAGGCTGGTGCT
GAGATGCTTGTGATGGCTTATGGTAGATCATATGGATTGCCTGTTATTACGACTCGCGG
GAACAATGTTTATGGGCCTAACCAGTTTCCTGAAAAAATGATTCCTAAGTTCATCTTGTT
GGCTATGAGTGGGAAGCCGCTTCCCATCCATGGAGATGGATCTAATGTCCGGAGTTACT
TGTACTGCGAAGACGTTGCTGAGGCTTTTGAGGTTGTTCTTCACAAAGGAGAAATCGGT
CATGTCTACAATGTCGGCACAAAAAGAGAAAGGAGAGTGATCGATGTGGCTAGAGACAT
CTGCAAACTTTTCGGGAAAGACCCTGAGTCAAGCATTCAGTTTGTGGAGAACCGGCCCT
TTAATGATCAAAGGTACTTCCTTGATGATCAGAAGCTGAAGAAATTGGGGTGGCAAGAG
CGAACAAATTGGGAAGATGGATTGAAGAAGACAATGGACTGGTACACTCAGAATCCTGA
GTGGT**GGGGTGATGTTTCTGGAGCTTTGCTTCCTCATCCGAG**

gBlock 2

GGGGTGATGTTTCTGGAGCTTTGCTTCCTCATCCGAGAATGCTTATGATGCCCGGTGG
AAGACTTTCTGATGGATCTAGTGAGAAGAAAGACGTTTCAAGCAACACGGTCCAGACAT
TTACGGTTGTAACACCTAAGAATGGTGATTCTGGTGACAAAGCTTCGTTGAAGTTTTTGA
TCTATGGTAAGACTGGTTGGCTTGGTGGTCTTCTAGGGAAACTATGTGAGAAGCAAGGG
ATTACATATGAGTATGGGAAAGGACGTCTGGAGGATAGAGCTTCTCTTGTGGCGGATAT
TCGTAGCATCAAACCTACTCATGTGTTTAATGCTGCTGGTTTAACTGGCAGACCCAACGT
TGACTGGTGTGAATCTCACAAACCAGAGACCATTCGTGTAAATGTCGCAGGTACTTTGA
CTCTAGCTGATGTTTGCAGAGAGAATGATCTCTTGATGATGAACTTCGCCACCGGTTGC
ATCTTTGAGTATGACGCTACACATCCTGAGGGTTCGGGTATAGGTTTCAAGGAAGAAGA
CAAGCCAAATTTCTTTGGTTCTTTCTACTCGAAAACCAAAGCCATGGTTGAGGAGCTCTT
GAGAGAATTTGACAATGTATGTACCTTGAGAGTCCGGATGCCAATCTCCTCAGACCTAA
ACAACCCGAGAAACTTCATCACGAAGATCTCGCGCTACAACAAAGTGGTGGACATCCCG
AACAGCATGACCGTACTAGACGAGCTTCTCCCAATCTCTATCGAGATGGCGAAGAGAAA
CCTAAGAGGCATATGGAATTTCACCAACCCAGGGGTGGTGAGCCACAACGAGATATTG
GAGATGTACAAGAATTACATCGAGCCAGGTTTTAAATGGTCCAACTTCACAGTGGAAGA
ACAAGCAAAGGTCATTGTTGCTGCTCGAAGCAACAACGAAATGGATGGATCTAAACTAA
GCAAGGAGTTCCCAGAGATGCTCTCCATCAAAGAGTCACTGCTCAAATACGTCTTTGAA
CCAAACAAGAGAACCTAAGTTTCTTAGCCTCCCATGGAAGAAACGTTCCCTCCTTA

Ordered gBlock with coding sequence of *A. thaliana atugt78d1*. At the 5' and 3' side of the gBlock homologous regions are present for assembly in the expression vector for the *ugtb1* locus (underlined).

AATAAAGTCTCCTATATGACCAAATTCTCCGAGCCAATCAGAGACTCCCACGTGGCAGT
TCTCGCGTTTTTCCCCGTTGGCGCTCATGCCGGTCCTCTCTTAGCCGTCACTCGCCGTC
TCGCCGCCGCTTCTCCCTCCACCATCTTTTCTTTCTTCAACACCGCAAGATCAAACGCGT
CGTTGTTCTCCTCTGATCATCCCGAGAACATCAAGGTCCACGACGTCTCTGACGGTGTT
CCGGAGGGAACCATGCTCGGGAATCCACTGGAGATGGTCGAGCTGTTTCTCGAAGCGG
CTCCACGTATTTTCCGGAGCGAAATCGCGGCGGCAGAGATAGAAGTTGGAAAGAAAGT
GACATGCATGCTAACAGATGCCTTCTTCTGGTTCGCAGCGGACATAGCGGCTGAGCTG
AACGCGACTTGGGTTGCCTTCTGGGCCGGCGGAGCAAACTCACTCTGTGCTCATCTCT
ACACTGATCTCATCAGAGAAACCATCGGTCTCAAAGATGTGAGTATGGAAGAGACATTA
GGGTTTATACCAGGAATGGAGAATTACAGAGTTAAAGATATACCAGAGGAAGTTGTATTT
GAAGATTTGGACTCTGTTTTCCCAAAGGCTTTATACCAAATGAGTCTTGCTTTACCTCGT
GCCTCTGCTGTTTTCATCAGTTCCTTTGAAGAGTTAGAACCTACATTGAACTATAACCTA
AGATCCAAACTTAAACGTTTCTTGAACATCGCCCCTCTCACGTTATTATCTTCTACATCG
GAGAAAGAGATGCGTGATCCTCATGGCTGCTTTGCTTGGATGGGGAAGAGATCAGCTG
CTTCTGTAGCGTACATTAGCTTCGGCACCGTCATGGAACCTCCTCCTGAAGAGCTTGTG
GCGATAGCACAAGGGTTGGAATCAAGCAAAGTGCCGTTTGTTTGGTCGCTGAAGGAGA
AGAACATGGTTCATCTACCAAAAGGGTTTTTGGATCGGACAAGAGAGCAAGGGATAGTG
GTTCCTTGGGCTCCACAAGTGGAACTGCTGAAACACGAGGCAATGGGTGTGAATGTGA
CACATTGTGGATGGAACTCAGTGTTGGAGAGTGTGTCGGCAGGTGTACCGATGATCGG
CAGACCGATTTTGGCGGATAATAGGCTCAACGGAAGAGCAGTGGAGGTTGTGTGGAAG
GTTGGAGTGATGATGGATAATGGAGTCTTCACGAAAGAAGGATTTGAGAAGTGTTTGAA
TGATGTTTTTGTTCATGATGATGGTAAGACGATGAAGGCTAATGCCAAGAAGCTTAAAGA
AAAACTCCAAGAAGATTTCTCCATGAAAGGAAGCTCTTTAGAGAACTTCAAGATCTTGTT
GGACGAAATTGTGAAAGTTTAGCCTGGCTCTTTTTCT

Appendix E

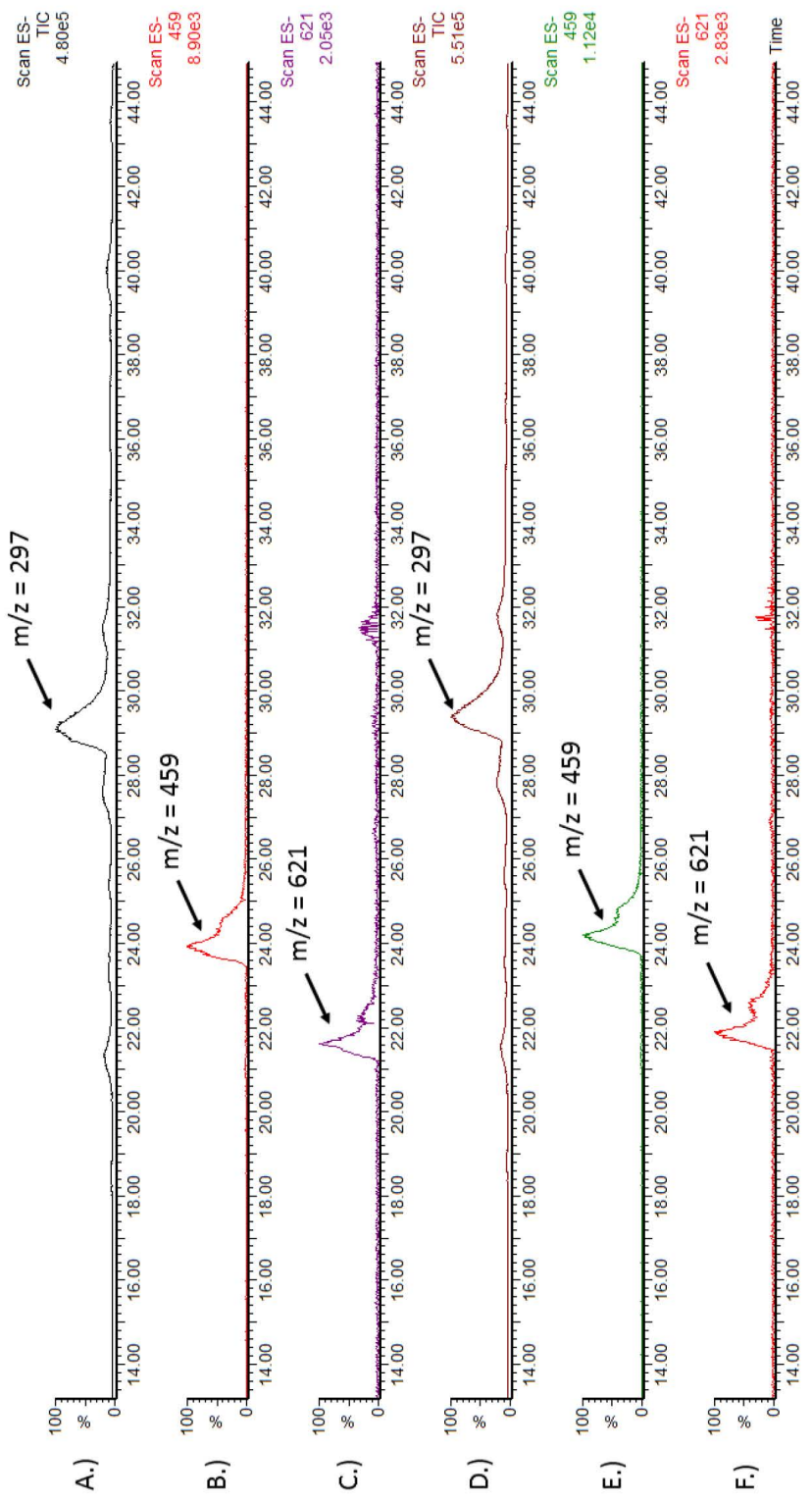


Figure A.1: LC-MS chromatograms of the *E. coli* BL21 *ugt1* enzyme test described in IV.3.4.1. A.) Total ion scan for the enzyme test extract. Hydroxy oleic acid is visible with a m/z = 297. B.) Mass scan for non-acetylated C18:1 glucolipids (m/z = 459) from the enzyme test extract. C.) Mass scan for non-acetylated C18:1 disaccharides (m/z = 621) from the enzyme test extract. D.) Total ion scan for the substrate used. Hydroxy oleic acid is visible with a m/z = 297. E.) Mass scan for non-acetylated C18:1 glucolipids (m/z = 459) from the substrate. F.) Mass scan for non-acetylated C18:1 glycolipids, most likely sophorolipids (m/z = 621) from the substrate.

Appendix F

Molecular masses of several common types of fatty acids, fatty acid derivatives and glycolipids. The molecular mass for the saccharide monomers used in the calculations was 180 Dalton. OH-FA = hydroxy fatty acid; gl = glycolipid with one saccharide; gl-gl = glycolipid with a disaccharide carbohydrate moiety; ac = acetylation; lac = lactonisation

Fatty acids	Mass	OH-FA	ω-al FA	Dicarboxylic acid	gl	ac-gl
C22:0	340	356	354	370	518	560
C20:0	312	328	326	342	490	532
C18:0	284	300	298	314	462	504
C18:1	282	298	296	312	460	502
C18:2	280	296	294	310	458	500
C18:3	278	294	292	308	456	498
C16:0	256	272	270	286	434	476
C14:0	228	244	242	258	406	448
C12:0	200	216	214	230	378	420
C10:0	172	188	186	202	350	392
C8:0	144	160	158	174	322	364

Fatty acids	Mass	gl-gl	ac-gl-gl	lac-ac-gl-gl	ac-ac-gl-gl	lac-ac-ac-gl-gl
C22:0	340	680	722	704	764	746
C20:0	312	652	694	676	736	718
C18:0	284	624	666	648	708	690
C18:1	282	622	664	646	706	688
C18:2	280	620	662	644	704	686
C18:3	278	618	660	642	702	684
C16:0	256	596	638	620	680	662
C14:0	228	568	610	592	652	634
C12:0	200	540	582	564	624	606
C10:0	172	512	554	536	596	578
C8:0	144	484	526	508	568	550