TÉCNICO LISBOA

Harnessing the oxidative power of the yeast *Starmerella bombicola* for innovative biochemical production

Maria Rita Esteves Marçal

Thesis to obtain the Master of Science Degree in

Biological Engineering

Supervisors: Prof. Dr. Nuno Gonçalo Pereira Mira
Prof. Dr. Wim Soetaert

Examination Committee

Chairperson: Prof. Dr. Miguel Nobre Parreira Cacho Teixeira
Supervisor: Prof. Dr. Nuno Gonçalo Pereira Mira
Member of the Committee: Dr. Nuno Ricardo Torres Faria

November 2019

Preface

The work presented in this thesis was performed at Centre for Industrial Biotechnology and Biocatalysis (InBio.be) of Ghent University during the period February-July 2019, under the supervision of Prof. Dr. Wim Soetaert, and within the frame of the Erasmus program. The thesis was co-supervised at Instituto Superior Técnico by Prof. Dr. Nuno Mira.

All the information deemed as confidential by Ghent University was removed from the thesis or replaced by code names.

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

First of all, I would like to express my gratitude to my family for all their support throughout this five year journey. I specially have to thank my parents for providing me with everything I needed to succeed and for giving me the opportunity to embark on my Erasmus experience.

Thank you also to all my friends at Instituto Superior Técnico for helping me go through the more stressful times at university and for turning my time there into a more fun and pleasant journey. Outside the university, I also want to thank my best (and oldest) friends that have always been there for me throughout every important moment of my life.

I also have to thank Instituto Superior Técnico, for not being the monster that everyone painted out to be, and for providing me with some memorable moments. I had some amazing professors there, but a special thanks should be addressed to Prof. Nuno Mira for accepting to be my supervisor.

I also want to express my gratitude to Prof. Wim Soetaert and Dr. Sophie Roelants for giving me the opportunity of developing my master's thesis at InBio.be, and to thank everyone at this group for always being ready to help and for being so patient with the only non dutch speaker of the group.

At InBio.be, my deepest gratitude has to go to my tutor Jelle Remmery, without whom this project would not have happened. Thank you for all the laboratory techniques and concepts you taught me, but specially thank you for always being in a good mood, even when nothing seemed to work. Your motivation was the key for never giving up!

Finally, thank you to all my Erasmus friends that where with me this semester, specially to the people of Home Groningen, second floor. I will never forget the time spent cooking in the kitchen, the weekend trips and the night-outs. You turn this into an incredible experience!

To each and every one of you – Thank you.

Abstract

Biosurfactants have been gaining interest as environmentally friendlier alternatives to the classical petroleum-based surfactants. *Starmerella bombicola* is an industrial relevant yeast due to its capacity of producing large amounts of glycolipid-type biosurfactants known as sophorolipids. However, the full exploitation of its potential is dependent on overcoming the difficulty in obtaining a standardised product, since this yeast produces a mixture of structurally related molecules whose composition varies according to fermentation conditions and medium composition. This lack of uniformity can be attributed to the versatile oxidative power of *S. bombicola*. In fact, this yeast possesses several enzymes which can take part of the hydroxylation system of hydrocarbon-assimilating yeasts such as *S. bombicola*. In this work, engineered strains, with knock-outs in different genes encoding for oxidising enzymes, were tested on the type of sophorolipids produced. A lower number of compounds was detected for strains with a knock-out in genes 1 or 3, while the deletion of gene 2 did not produce a difference comparing to the wild-type. However, the double deletion of genes 1 and 2 led to a lower variety of products comparing to the respective single knock-outs. Finally, the selective production of product A was intended to be used as a proof of concept that gene 2 encodes for the enzyme responsible for hydroxylation of a specific substrate in *S. bombicola*, but the results indicated that another enzyme must be capable of catalysing this step.

Keywords

Starmerella bombicola; sophorolipids; biosurfactants; oxidative power.

Resumo

Os biosurfactantes têm vindo a ganhar interesse como alternativas verdes aos surfactantes derivados do petróleo. A levedura *Starmerella bombicola* apresenta potencial industrial devido à sua capacidade para produzir grandes quantidades de biosurfactantes da classe dos glicolipídeos - os soforolípidos. No entanto, a completa exploração deste potencial está dependente de ultrapassar a dificuldade em obter produtos *standard*, dado que esta levedura produz uma mistura de moléculas estruturalmente relacionadas, cuja composição varia de acordo com as condições de fermentação e a composição do meio. Esta falta de uniformidade pode ser atribuída ao poder oxidativo versátil de *S. bombicola*. De facto, esta levedura possui várias enzimas pertencentes ao sistema de hidroxilação de leveduras assimiladores de hidrocarbonetos, como é o caso de *S. bombicola*. Neste trabalho, estirpes geneticamente manipuladas, com um ou mais genes que codificam para enzimas oxidativas eliminados, foram testadas quanto ao tipo de soforolípidos que produziam. Um menor número de compostos foi detetado para estirpes em que os genes 1 ou 3 se encontravam eliminados, enquanto a eliminação do gene 2 não produziu diferenças comparativamente à estirpe selvagem. Porém, a dupla eliminação dos genes 1 e 3 levou à produção de uma menor variedade de produtos, comparativamente às estirpes com os respetivos genes eliminados individualmente. Finalmente, pretendia-se usar a produção seletiva do produto A como prova de conceito de que a enzima codificada pelo gene 2 é responsável pela hidroxilação de um substrato específico em *S. bombicola*, mas os resultados obtidos mostraram que existe outra enzima capaz de catalisar esta reação.

Palavras Chave

Starmerella bombicola; soforolípidos; biosurfactantes; poder oxidativo.

Contents

List of Figures

List of Tables

Acronyms

ACN	acetonitrile
AO	amine oxide
AE	alkyl ethoxylate
AES	alkyl ethoxy sulphate
APE	alkylphenol ethoxylate
APGs	alkyl polyglucosides
AS	alkyl sulphate
BB	back-bone
CBL	cellobioselipid
CFU	Colony Forming Units
CMC	Critical Micelle Concentration
CPEC	Circular Polymerase Extension Cloning
CPR	NADPH-cytochrome P450 oxidoreductase
CSM-A	Complete Supplement Mixture without adenine
CSM-U	Complete Supplement Mixture without uracil
DPGs	diolpolyglucosides
DTT	dithiothreitol
ELS	Evaporative Light Scattering
ER	endoplasmic reticulum
gDNA	genomic DNA
GM	glycerol-mannitol
HESI	Heated-Electrospray Ionization
HR	homologous region
KO	knock-out
LAS	linear alkylbenzene sulphonic acid
LB	Lysogeny Broth

LC-MS Liquid Chromatography – Mass Spectrometry

miRNA microRNA

NEB New England Biolabs Inc.

NHEJ Non-Homologous End Joining

OD optical density

P450s Cytochrome P450 monooxygenases

PCI Phenol:Chloroform:Isoamyl alcohol

PCR Polymerase Chain Reaction

PHA polyhydroxyalkanoate

QAC quaternary ammonium compound

RNAi RNA interference

SB *Starmerella bombicola*

SD+CSM-A Synthetic Dextrose with Complete Supplement Mixture without adenine

SD+CSM-U Synthetic Dextrose with Complete Supplement Mixture without uracil

SDS sodium dodecyl sulfate

siRNA small interfering RNA

SOB Super Optimal Broth

TEA triethylamine

TLC Thin Layer Chromatography

UPLC Ultra Performance Liquid Chromatography

WT wild type

YPD Yeast extract Peptone Dextrose

YNB Yeast Nitrogen Based

1 Introduction

The global concern about the environmental impact caused by the elevated use of surfactants leads to the demand of environmentally friendlier alternatives [1]. Contrary to these petroleum-based compounds, biosurfactants are produced by microorganisms and present several advantages, such as less toxicity and increased biodegradability, which permits their effective removal from the environment [2]. Amongst biosurfactants, sophorolipids attract particular commercial interest and already have applications in several companies, for example, in detergent and cosmetic formulations [3].

Sophorolipids are glycolipid-type biosurfactants consisting of a hydrophobic fatty acid tail of 16 or 18 carbon atoms and a hydrophilic carbohydrate head – the sophorose [4]. The yeast *Starmerella bombicola* has the ability to produce large amounts of sophorolipids from several substrates (oils, fatty acids, alkanes, etc.), including renewable sources and waste streams [5]. These characteristics, together with its non-pathogenic nature, open the door to the opportunity of transforming this yeast into a platform organism for the production of tailor-made biomolecules [6].

The creation of engineered *S. bombicola* strains has made it possible to direct the production towards glucolipids, bola-sophorolipids, cellobioselipids, among others, instead of the naturally produced acidic and lactonic sophorolipids [7]. However, the lack of uniformity of the product mixture obtained using this yeast, remains as a major drawback for their production at an industrial scale. One strategy that can be used to achieve a selective production of a certain compound is to knock-out or overexpress genes which take part in the sophorolipid biosynthesis [7]. However, this strategy is dependent on fully understanding the versatile oxidative power of *S. bombicola*.

The goal of this master's thesis was the preliminary investigation of the oxidative power of *Starmerella bombicola* by studying the influence of enzymes involved in oxygenation reactions on the type of sophorolipids produced.

2 Literature Review

2.1 Surfactants

2.1.1 Properties, Structure and Applications

The name surfactant is a contraction of the term surface-active agent. Therefore, surfactants are compounds with the capacity to adsorb onto the interfaces of a system composed of two immiscible phases, altering their interfacial free energies and, consequently, the interfacial tension between the two phases. The interfacial free energy is the amount of work required to create the interface, which divided per unit area translates into the interfacial (or surface) tension. [8] Surface tension results from the cohesive forces between molecules residing at or close to the interface which causes the surface to contract into the smallest possible area [9]. Thus, this parameter corresponds to the minimum amount of work required to expand the interface by unit of area. This measurement is also a reflection of the difference in nature of the two phases meeting at the interface, since the greater their dissimilarity, the higher will be the surface tension. [8]

Surfactants have a chemical structure composed of two different functional groups which have differences in affinity towards the same molecule: one group has very little attraction towards the solvent, whereas the other has a strong attraction. The first group is denominated hydrophobic, if in an aqueous system, or lipophobic, if in a lipid system. In the same way, the second group is called hydrophilic or lyophilic. [10]

Surfactants usually act to decrease the surface tension between the two phases. In the case of a surfactant dissolved in an aqueous medium, the hydrophobic groups orient themselves in order to minimise the contact with water molecules, which means that they will be oriented predominantly towards the air. [8] Therefore, the hydrophobic group distorts the structure of the water by weakening the cohesive forces between the water molecules (hydrogen bonds). Since intermolecular forces between surfactant and water molecule are much lower than between two water molecules, the interfacial tension will decrease. [11] This decrease in the interfacial tension is also accompanied by a decrease in the dissimilarity between the two phases at the interface, since air molecules, as well as the hydrophobic groups, are essentially non-polar in nature. [8] On the other hand, once the interface is saturated, the addition of more surfactants will not decrease the interfacial tension any further. Instead, the molecules will organise themselves in a structure which consist of several clustered surfactant molecules that shield their non-polar chains from the surrounding aqueous phase with their polar head groups. These structures are denominated micelles and the surfactant concentration at and above which micelles are formed is denominated Critical Micelle Concentration (CMC). [12]

Regarding the chemical groups composing the surfactant, the hydrophobic or lipophobic tail can be a hydrocarbon (branched, linear or aromatic), fluorocarbon or siloxane. The head, that is the hydrophilic or lyophilic group, is an ionic or highly polar group, which classifies the nature of the surfactant (figure 2.1). If the head group has no charge, the surfactant is called non-ionic, whereas if it has a negative or positive charge, it is called anionic or cationic, respectively. If it contains both positive and negative groups, then the surfactant is called zwitterionic. [11]

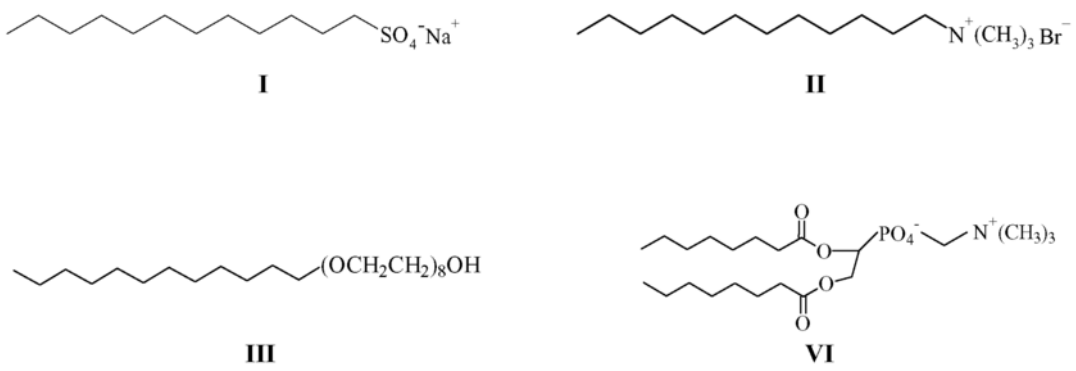

Figure 2.1: Examples of the different types of surfactants: **I** - anionic (Sodium dodecyl sulfate, SDS); **II** - cationic (dodecyltrimethylammonium bromide, CTAB); **III** - non-ionic (n-dodecyl tetra, $C_{12}E_4$) and **VI** - zwitterionic (dioctanoyl phosphatidylcholine, C_8-lecithin). [13]

The property of adsorbing onto interfaces, altering their surface tension, makes of surfactants the most versatile products of the chemical industry. They are used in many applications, such as cleaning, wetting, dispersing, emulsifying, foaming and anti-foaming agents. Consequently, they are present in products such as detergents, paints, paper coatings, inks, shampoos, toothpastes, firefighting foams, insecticides, amongst others. [14]

2.1.2 Environmental Impact

The widespread use of surfactants all over the world and the increasing environmental public awareness leads to the debate of the potential contamination associated with release of surfactants into the environment. [1]

After use, residual surfactants are discharged into sewage systems or directly into waste water, causing their accumulation in great quantities in wastewater treatment plants. [15] For this reason, several ecotoxicity studies evaluating surfactant breakdown and biological effect in the environment have already been developed ([1], [15], [16]). Many commonly used surfactants, such as the anionic linear alkylbenzene sulphonic acid (LAS), alkyl ethoxy sulphate (AES) and alkyl sulphate (AS) and the non-ionic alkyl ethoxylate (AE) [17], do not seem to pose a threat to the environment. However, there is still a lack of information concerning the toxic effects of cationic surfactants such as quaternary ammonium compound (QAC) and amine oxide (AO) types of surfactants. QAC's are widely used in detergents, fabric softeners, and hair conditioners, whereas AO's are used in textile industry, in rubber industry and in deodorant bars. Another source of concern is the alkylphenol ethoxylate (APE) type of surfactants, since their degradation products are octylphenols and nonylphenols, which can disrupt normal functioning of the fish endocrine system and accumulate in aquatic organisms. [15] In fact, the discovery of the toxicity of APE breakdown products, in 1984, led to its banishment or restriction of use in Europe. [18]

In summary, even though surfactants are often regarded as harmless due to their speculate low concentrations in the environment, statistical analysis of surfactant concentrations worldwide reveals that they are usually above their predicted "no effect" concentrations. Efforts should be made to limit the use of surfactants, from household level to large scale industries, since their accumulation in waste water can cause serious environmental problems. For example, in developing countries the excessive usage of phosphate-based detergents causes excessive growth of algae and, consequently, the eutrophication of ponds. On the other hand, surfactants also cause adverse effects in aquatic organisms and in bacteria and, even at low concentrations, they seem to significantly alter soil biochemistry and physical characteristics. [19]

Finally, most surfactants are produced from petroleum feedstocks. The global concern about oil reservations and their environmental impact leads to the quest for more sustainable alternatives, especially for those who use renewable sources as raw materials for production. [6]

2.2 Biosurfactants

2.2.1 Advantages over chemical surfactants

Biosurfactants are, as surfactants, surface-active molecules. The major difference between the two is that biosurfactants are produced by microorganisms and not petroleum based. Comparing the two, biosurfactants present several advantages over their chemically synthesised equivalents. [2]

The first one is their biodegradability, since microbial derived compounds can be easily degraded. [2] This is an important feature when evaluating environmental risk associated with surfactant use, since biological degradation is the most important mechanism for irreversible removal of these substances from aquatic and terrestrial environments. [20] In [21], the biodegradability of five different bacterial surfactants and one synthetic surfactant - sodium dodecyl sulfate (SDS) - was evaluated. In a liquid medium and in soil microcosms, the CO_2 evolution was measured during incubation with bacteria strains capable of degrading different petroleum hydrocarbons. It was observed that the microorganisms used the biosurfactants as carbon sources in a minimal medium, while SDS was also mineralized, but at a lower rate. In the same way, CO_2 emitted by a mixed bacterial culture in soil microcosms with biosurfactants was higher than in the microcosm containing SDS. These results show that biosurfactants are environmentally friendlier, which is complemented by the fact that they can be produced from industrial wastes and from by-products. Thus, the availability and price of the raw materials is another improvement. [22]

On the other hand, biosurfactants can also be active under extreme conditions when they are produced by extremophile organisms. For example, a biosurfactant produced by *Arthrobacter protophormiae* was found to be resistant to temperatures in a range between 30°C and 100°C and pH between 2 and 12 [23]. Since most industrial processes involve exposure to extreme temperatures, pH and pressure, the isolation of microbial products capable of functioning under these conditions is certainly beneficial. [24]

Regarding the surface and interface activity, generally speaking, biosurfactants are more effective and efficient and their CMC can be several times lower than petroleum-based surfactants [25], which means that for maximal decrease of surface tension, less surfactant is necessary. Finally, most biosurfactants also present low toxicity and are biocompatible and digestible making them suitable for pharmaceutical, cosmetic and food uses. [26]

2.2.2 Physiological role of biosurfactants

Biosurfactants are produced by a variety of microorganisms predominantly during growth on water immiscible substrates and secreted either extracellularly or attached to a part of the cell. [2] Since biosurfactants have different structures and surface properties and are produce by a wide variety of microorganisms, it is difficult to generalise their function. However, some of their roles have already been suggested. [27]

Firstly, biosurfactants allow microorganisms to grow on hydrophobic water-insoluble substrates by making them more readily available for uptake and metabolism. Emulsification has been suggested has a possible mechanism to increase the surface area between water and the substrate. [27] In [28], the authors observed that the production of biosurfactant by a *Pseudomonas* strain was accompanied by an increase in the aqueous concentration of the polycyclic aromatic hydrocarbon used as substrate which, consequently, increased its availability for biodegradation.

Biosurfactants can also play a role of antimicrobial agent towards various microorganisms. [27] Several biosurfactants have already been described as antibiotics, including the cyclic lipopeptide of *Bacillus subtilis* (surfactin) [29] and the extracellular peptide produced by *Streptomyces tendae* (streptofactin) [30]. Moreover, biosurfactants can be involved in cell adherence which imparts stability under hostile environmental conditions and virulence. [2] In order to attach to a surface, microorganisms can use their biosurfactants to regulate their cell surface properties - for example, the presence of cell-bound rhamnolipid increses the hydrophobicity of *P. aeruginosa* cell surface [31]. Furthermore, this biosurfactant is produced under the cell density-dependent control system (quorum sensing) and regulated in correlation with other virulence factors. [32]

Finally, biosurfactants can be involved in the removal of metal toxicity. In [33], the authors demonstrated the ability of the biosurfactant rhamnolipid to remove cadium, lead and zinc from soil.

In conclusion, biosurfactants can have multiple roles according to the needs of the producing organisms. Biosurfactant-producing organisms have been isolated from a wide diversity of environments including soil, sea water, marine sediments, oil fields and even extreme environments. [25] Some examples of these microorganisms will be given in the following section.

2.2.3 Classification and microbial origin

Contrary to chemically synthetized surfactants, biosurfactants are classified according to their chemical composition and microbial origin. The major classes of biosurfactants include glycolipids, lipopeptides and lipoproteins, phospholipids and fatty acids, polymeric surfactants, and particulate surfactants. [25]

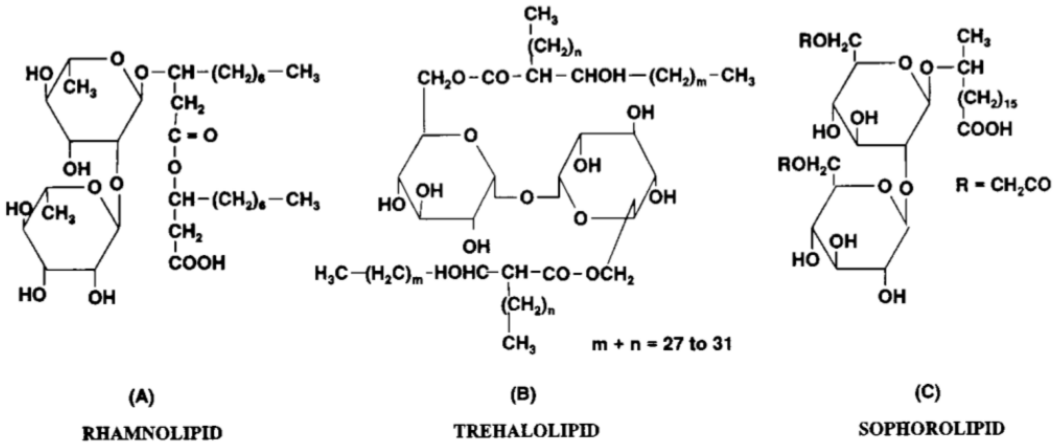

Figure 2.2: Structure of some common glycolipid biosurfactants: **A** - Rhamnolipid from *Pseudomonas aeruginosa* in which two rhamnose subunits are linked to two β-hydroxydecanoic acids in a side chain; **B** - Trehalose dimycolate from *Rhodococcus erythropolis* in which disaccharide trehalose is linked to two long-chain α-branched-β-hydroxy fatty acids; **C** - Sophorolipid from *Starmerella bombicola* in which dimeric sophorose is linked to a long-chain (C_{18}) hydroxy fatty acid. [25]

Glycolipids are carbohydrates in combination with long-chain aliphatic acids or hydroxyaliphatic acids. The best known biosurfactants included in this class are rhamnolipids, trehalolipids, and sophorolipids. Rhamnolipids have one or two molecules of rhamnose linked to one or two molecules of β-hydroxydecanoic acid and are produced by *Pseudomonas aeuginosa* [34] (Fig. 2.2.A) and *Pseudomonas* spp. Trehalolipids can have several structures, but a disaccharide trehalose linked at the 6 and 6' positions to mycolic acids is associated with most species of *Mycobacterium* [35], *Nocardia* [36], and

Corynebacterium [37]. Mycolic acids are long-chain α-branched-β-hydroxy fatty acids, which differ in size and structure depending on the producing organism - in Fig.2.2.B is presented trehalose dimycolate produced by *Rhodococcus erythropolis* [38]. Sophorolipids are produced mainly by yeasts such as *Starmerella bombicola* [39] and consist of a dimeric carbohydrate sophorose linked to a long-chain hydroxy fatty acid by glycosidic linkage (Fig. 2.2.C). [25]

Regarding lipopeptides and lipoproteins, these consist of a lipid attached to a polypeptide chain. [24] These include a large number of cyclic lipopeptides with antimicrobial action, such as decapeptide antibiotics (gramicidins) and lipopeptide antibiotics (polymyxins), produced by *Bacilus brevis* [40] and *B. polymyxa* [41], respectively. The cyclic lipopeptide surfactin (Fig. 2.3), produced by *B. subtillis*, is one of the most powerful biosurfactants known since is capable of reducing the surface tension of distilled water from 72 mN/m to 27.9 nM/m at a concentration of 0.005% [42]. Finally, lichenysin, produced by *B. licheniformis* exhibit excellent stability under extreme temperature, pH and salt conditions [43].

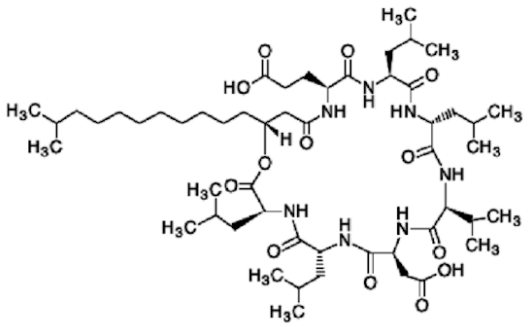

Figure 2.3: Structure of surfactin produced by *Bacillus subtillis*. [44]

Several bacteria and yeasts produce large quantities of fatty acid and phospholipid surfactants during growth on n-alkanes. [24] For example, an *Acinetobacter* strain produces rich vesicles that form microemulsions of alkanes in water [45].

Polymeric biosurfactants consist of polysaccharide-protein complexes such as emulsan, liposan, alasan, mannoprotein. [24] Emulsan, produced by an *Acinetobacter calcoaceticus* strain is an effective emulsifying agent for hydrocarbons in water, even at a concentration as low as 0.001-0.01% [46]. Liposan, an extracellular water-soluble emulsifier synthesized by *Candida lipolytica*, is composed of 83% carbohydrate and 17% protein and can be used as an emulsifier in food and cosmetic industries [47]. Mannoprotein, which is a protein produced by *Sacharomyces cerevisiae*, shows excellent emulsifier activity toward several oils, alkanes, and organic solvents [48].

Particulate biosurfactants consist of extracellular membrane vesicles which play an important role in alkane uptake by microbial cells by forming a microemulsion with hydrocarbons. [24] For example, *Sphingomonas* sp. presents a cell surface covered with extracellular vesicles when grown on polyaromatic hydrocarbons, whereas the surface is smooth when the cells are grown on a hydrophilic substrate [49].

2.2.4 Applications

2.2.4.A Oil Industry

One industrial application for biosurfactants is in the field of oil recovery and processing, in which they are involved in a process called microbial enhanced oil recovery. Two of the causes for poor oil recovery (around 40-45%) are the low permeability of the rocks forming the reservoir or the high viscosity of the

crude oil. This means that the ability of biosurfactants to reduce the oil/water interfacial tension and to form stable emulsions can improve the process efficiency. Moreover, the de-emulsifying properties of some biosurfactants may be used to break emulsions formed at various steps in oil extraction and processing, thus allowing a better recovery of the product. [3]

2.2.4.B Food Industry

Biosurfactants are also used in food processing industry to promote the formation and stabilisation of emulsions, which improves the texture and creaminess of dairy products. They are also used to delay staling, solubilise flavour oils and improve organoleptic properties in bakery and ice cream formulations and as fat stabilisers during cooking of fats. The biosurfactants used for this application are usually the ones obtained from yeast or *Lactobacilli*. [50]

2.2.4.C Cosmetic Industry

In cosmetic industry, due to their low toxicity and high biodegradability, biosurfactants are starting to replace chemically synthesised surfactants for their water binding capacity and their emulsifying, foaming, spreading and wetting properties, which have an effect on the product viscosity and consistency. [2, 3]

2.2.4.D Detergent formulations

In detergent formulations, low-foaming sophorolipids from *S. bombicola* appear suitable due to their low cytotoxicity and high biodegradability and general environmentally acceptable properties. [3] Worldwide many companies are developing and patenting cleaning products containing biosurfactants, such as Saraya [51], Naturell Clean [52] and Ecover [53].

2.2.4.E Environmental Biotechnology

Besides their industrial applications, biosurfactants demonstrate potential for environmental biotechnology. For example, biosurfactants are a promising method to improve the effectiveness of bioremediation of hydrocarbon contaminated environments. This environmentally friendly technique to remediate soil permits to preserve soil structure, requires little energy input, and involves the complete destruction or immobilisation of the contaminants. However, hydrocarbons exhibit low solubility in groundwater and tend to partition to the soil matrix, so they are poorly accessible to bacteria. Biosurfactants help to increase the mobility and bioavailability of hydrocarbons since they can increase their solubility in water or allow hydrophobic substrates to associate more easily with bacterial cells. This last one is achieve by interacting with the cell surface or by allowing the contact with pseudosolubilized or emulsified oil. [3, 54] Even without bioremediation, biosurfactants can still be use to release contaminants, characterised by low solubility, from the soil. This technique is called soil washing and its efficiency is based on the chemico-physical properties of the biosurfactant. [54]

Finally, biosurfactants can also be used for the removal of heavy metals from contaminated soil due to their ability to form complexes with metals. This way, they can be applied to a small part of contaminated soil inside a mixer so that anionic biosurfactants create complexes with metals by ionic bonds. This bond is so strong that flushing water through soil removes the biosurfactant-metal complex from the soil matrix to the soil solution due to the lowering of the interfacial tension. In the end, soil is deposited back, and the biosurfactant-metal complex treated to separate the two. [54]

2.2.4.F Biomedical applications

Biosurfactants show the ability to disturb the integrity and permeability of cell membranes, as well as to alter surface properties affecting microorganisms adhesion to them. These properties make them valuable for biomedical applications. [55]

Despite their potential, biomedical and pharmaceutical applications of biosurfactants still remain quite limited. However, lipopeptides, which have the capacity of disrupting lipid membranes, are already being used as substitutes of conventional antibiotics, which no longer work against resistant bacteria and fungi. For example, daptomycin (Cubist Pharmaceuticals [56]), a branched cyclic lipopeptide, was approved in 2003 for the treatment of skin infections caused by Gram-positive pathogens, including the methicillin-resistant *Staphylococcus aureus*. [55]

Biosurfactants could also find applications in preventing biofilm formation in medical devices. [55] Patents in this field are already being published, such as the use of the lipopeptide biosurfactant produced by the strain *B. subtilis* V9T14 in combination with antibiotics [57] and the use of *Lactobacillus* biosurfactants [58].

2.3 Sophorolipids

2.3.1 Characteristics and applications

Sophorolipids (Fig. 2.4) are glycolipid-type biosurfactants consisting of a ω/ω-1 hydroxy fatty acid tail, which can have one or more unsaturated bonds, and a hydrophilic carbohydrate head, the sophorose. Sophorose is a glucose disaccharide, with a β-1,2 bond between the two units, which can be acetylated on the 6′- and/or 6′′- positions. The fatty acid is β-glycosidically linked to the sophorose molecule and hydroxylated terminal or subterminal. Its carboxylic end is either free (acidic or open form) or internally esterified at the 4′′ or in some rare cases at the 6′- or 6′′- position (lactonic form). [4]

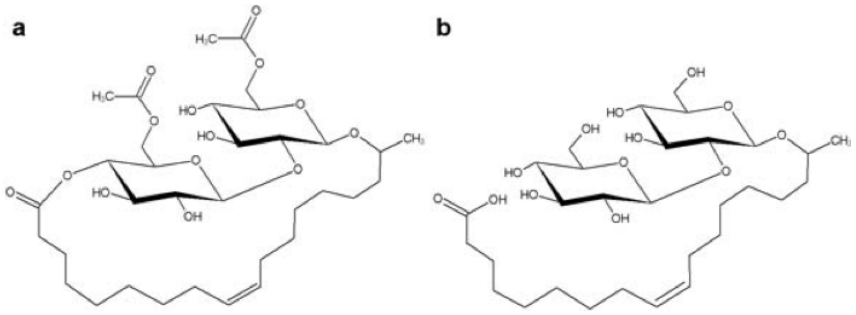

Figure 2.4: Sophorolipids produced by *S. bombicola*. **a)** diacetylated lactonic sophorolipid; **b)** non-acetylated open-chain (acidic) sophorolipid. [59]

Regarding its properties, when sophorolipids are solved in water, they lower the surface tension from 72.8 mN/m down to 35-36 mN/m, with a critical micelle concentration that varies between 35 and 200 mg/L depending on the substrate used [60]. Furthermore, their physicochemical properties vary depending if the sophorolipid is lactonized or in its acidic form. Generally, lactonic sophorolipids perform better at lowering surface tension and have an enhanced antimicrobial activity, whereas the acidic ones display better foam production and solubility. [4] The presence of acetyl groups also has an effect on the properties of the sophorolipid by enhancing their water solubility and their antiviral and cytokine stimulating effects [61].

There is a widespread commercial interested in using sophorolids as an environmental friendlier alternative to chemical surfactants. In fact, 24% of the worldwide patents on biosurfactants act on sophorolipids and commercialisation has already started. [62] For example, the Japanese company Saraya commercializes a dish washer containing sophorolipids as a cleaning agent (sophoron) and the French company Soliance produces sophorolipid based cosmetics, since they are claimed to trigger beneficial events regarding protection of hair and skin. Their emulsifying properties can also be exploited

in the petroleum industry and in the food industry. Furthermore, they can be used in environmental applications to decontaminate porous media (soils and groundwater tables) polluted by hydrocarbons or for the removal of heavy metals from sediments. [4]

However, the interest in sophorolipids in not limited to the production of surfactants, since their unique chemical structure can serve as the basis for synthesizing certain hydroxy fatty acids [63] and other compounds, such as fatty amines [64]. They also display antimicrobial activity against certain yeasts, plant pathogenic fungi and bacteria, so they can be applied in germicidal mixtures suitable for cleaning fruits and vegetables [65]. Finally, in the future sophorolipids could also be used for medical applications since they exhibit cytotoxic effects on several human cancer cell lines [66]. [4]

2.3.2 Biosynthetic pathway

Sophorolipids are produced by several closely related yeasts such as *Starmerella bombicola*, *Candida batistae*, *C. riodocensis*, *C. apicola* and *C. stellata*. Among these species, *S. bombicola* has been the most studied due to its non-pathogenic nature and ability to produced large amounts of biosurfactant (with yields of up to 400 g/L in optimised fermentations). [4, 62]

The complete pathway of sophorolipids biosynthesis by *S. bombicola* is already described, compromising six genes. The responsible genes are grouped in one large subtelomeric gene cluster [59], with the exception of the gene responsible for the lactonization of sophorolipids, which was found 2.5 Mbp upstream the cluster [67].

The first step in sophorolipids biosynthesis (Fig. 2.5) consists of terminal (ω) or subterminal (ω-1) hydroxylation of a fatty acid performed by the cytochrome P450 monooxygenase CYP52M1. After, two glucosylstransferases, which act in two distinct steps, are responsible for a subsequent glycosylation of the hydroxylated fatty acid: UGTA1 catalyses the transfer of a first glucose molecule from UDP-glucose, generating a glucolipid, which then receives a second glucose molecule transferred by UGTB1 from UDP-glucose. This results in the formation of an acidic, nonacetylated sophorolipid molecule. However, the majority these molecules are subsequently acetylated resulting in mono or di-acetylated sophorolipids. The acetylation at the 6′- and/or 6′′-position is carried out by an acetyl-coenzyme A (CoA) dependent acetyltransferase. Finally, the molecule is secreted by MDR proteins, which are membrane integrated transporters, to the extracellular space, where lactonic sophorolipids are formed. This final step is performed by a lactone esterase which catalyses the esterification of the carboxyl group of the hydroxy fatty acid with a hydroxyl group of sophorose at the 4”-position. [59]

2.3.3 Production conditions

Sophorolipid are considered to be synthesised by yeast cells as secondary metabolites, since they are only produced in the stationary phase and do not seem to play an essential role for cell viability. The regulation of sophorolipid biosynthesis is considerably complex, but it has already been established that it takes place under conditions of nitrogen limitation. Therefore, it is hypothesised that sophorolipids are used by cells as extracellular carbon storage material to safeguard the consumption by other organisms, which is complemented by their antimicrobial effect. [59] Furthermore, as *S. bombicola* and *C. apicola* inhabit environments with high osmotic strength, sophorolipid production may be a way of dealing with the high sugar concentrations encountered in their natural habitats (nectar-feeding insects and flowers) by converting and storing them. [4]

Even though sophorolipid production is possible when only one type of carbon source is provided [68], it is considerably more efficient when yeast cells are provided with both hydrophilic (glycidic) and a hydrophobic (lipidic) types of carbon sources. In most cases, glucose is used as the hydrophilic carbon source, whereas several substrates can act as the hydrophobic carbon source: oils, fatty acids, and their corresponding esters, alkanes, etc. These two types of carbon sources will provide the building blocks

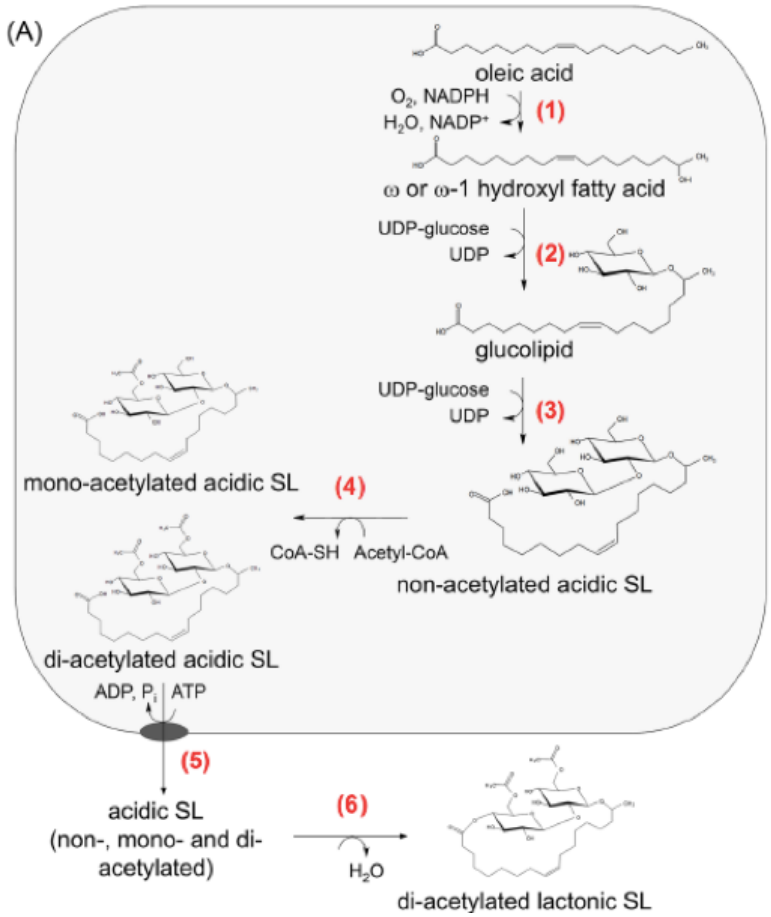

Figure 2.5: Sophorolipid biosynthetic pathway. **(1)** Cytochrome P450 monooxygenase CYP52M1, **(2)** UDP-glucosyltransferase UGTA1, **(3)** UDP-glucosyltransferase UGTB1, **(4)** acetyltransferase AT, **(5)** sophorolipid transporter MDR, **(6)** lactone esterase SBLE. [5].

for conventional sophorolipid synthesis: glucose and a fatty acid. [4] The fatty acid can be supplied as a triglyceride or a fatty acid methyl- or ethyl ester, since it will undergo extracellular hydrolysis by a lipase, being the resulting fatty acids uptaken by the cell. On the other hand, since sophorolipid-producing yeast strains, such as *S. bombicola* and *C. apicola*, are capable of growing on alkanes, they have the enzymes required for the terminal oxidation of alkanes, thereby generating fatty acids for further β-oxidation. Finally, if no hydrophobic substrate is present in the medium, fatty acids will also be formed *de novo* starting from acetyl-CoA derived from glycolysis. However, if the glucose concentration is low, part of the fatty acids will be conducted toward the β-oxidation for cell maintenance instead of sophorolipid synthesis. [69]

Although several molecules can be used as the hydrophobic substrate, it must be noted that the yield for sophorolipid production is not the same among them. [69] Regarding the use of alkane substrates, it has been observed that hexadecane and octadecane (C_{16} and C_{18} alkanes, respectively) achieve the best production yields, since native sophorolipids possess a fatty acid tail counting 16 or 18 carbon atoms. This means that divergent fatty acids will either be metabolised via β-oxidation or, in the case of shorter fatty acids, elongated to C_{16} or C_{18}. However, direct incorporation of smaller alkanes is also possible in a minor extent. On the other hand, fatty acid methyl- or ethyl esters derived from vegetable oils are preferred over their corresponding oils, because free fatty acids can disturb the electron balance

of the cells, and both of them perform better than alkanes. [70] The most common vegetable oils, namely the ones which contain oleic acid, are comprised of saturated or unsaturated fatty acids with the optimal chain length, making them an ideal substrate for direct incorporation, which consequently leads to a high sophorolipid production and yield. [69] This is definitely an advantage since vegetable oils are renewable resources making an environmentally friendly production possible for this biosurfactant. [4] It has been observed that rapeseed oil is one of the substrates of choice since it allows the best production yields when compared with other commonly available vegetable oils due to its abundance in oleic acid (optimal chain length) [71].

Regarding the diversity in sophorolipid composition, the fatty acid part of the molecule depends on the type of hydrophobic substrate used and varies in terms of chain length (16 or 18 carbon atoms), number of unsaturations (between 0 and 2) and position of the hydroxyl group (terminal or sub-terminal) [72]. Furthermore, the sophorose unit can be mono- or diacetylated and the fatty acid tail can either be free (acid form) or esterified to the sophorose head (lactonic form) [4]. Besides, the sophorose head can even be present on both ends of the molecule, forming the so-called bola-sophorolipids [73]. The great variety in sophorolipid composition is illustrated in [72], in which the authors analysed the sophorolipid mixture produced by *S. bombicola* in rapeseed oil and were able to identified more than 20 individual compounds. The ratio of homologs formed depends not only on the substrate used, but also on the medium composition and fermentation conditions. For example, when C_{16} or C_{18} alkanes are used, 85% or more of the sophorolipids are diacetylated lactones. Also, sophorolipids produced from oils always exhibited a higher level of diacetylated lactones than the ones produced from the corresponding esters [70]. The presence of citrate on the medium also has an influence on which type of sophorolipids are produced: in [74], the authors show that *S. bombicola* produces 50% versus 95% acidic sophorolipids in the presence or absence of citrate, respectively [5]. Finally, the use of substrates rich in polyunsaturated fatty acids seems to increase the level of acidic sophorolipids [70]. This lactonic/acidic balance also changes during the course of fermentation: in the beginning, acidic forms are predominant, whereas later on, conversion to the lactonic and acetylated forms is observed [75].

Lastly, for sophorolipid production the optimal temperature is set at 21°C [76], but most fermentations are run at 25°C or 30°C, with no significant differences between the two of them. [69] *S. bombicola* cultivation is associated with a strong pH drop, so addition of NaOH is necessary for maintaining it at the value of 3.5 [76]. Also, oxygen supply is very important throughout the whole fermentation process: it is essential for the exponential growth, but also for the sophorolipid biosynthesis, since at least one the enzymes involved, cytochrome P450 monooxygenase, requires molecular oxygen. [4] The optimal aeration rate for high sophorolipid yield, expressed in terms of oxygen transfer rate, lays between 50 and 80 mM O_2/L h^{-1} [77].

2.4 *Starmerella bombicola*

S. bombicola is an industrial relevant yeast, due to its non-pathogenic nature and ability to produced large amounts of sophorolipids. This yeast was first isolated from the honey of *Bombus* sp. (bumble-bee) by Spencer *et al.* (1970) [39]. It was first named *Torulopsis bombicola*, but, in 1978, it was reclassified as *Candida bombicola* by Yarrow *et al.* [78], since the inability of *Torulopsis* species to form pseudohyphae was not enough to discriminate from the *Candida* genus. The yeast was first thought to be a non-sporulating species, but after conjugation was observed, it was proposed a new genus – *Starmerella* – to accommodate the sexual stage of *C. bombicola* [79]. This genus was named after William T. Starmer, in recognition of his contributions to the ecology and evolution of yeasts associated with plants and insects. [5]

The biochemistry of *S. bombicola* it is still fairly unexplored, so little is known besides the substrate it requires, the sophorolipids and some other molecules it produces, and some aspects of the regulation of sophorolipid biosynthesis. [5]

Unlike other n-alkane utilising yeasts, sophorolipids stimulate the growth of *S. bombicola* on n-alkanes [80], which suggest that this biosurfactant plays an important role in alkane uptake. Besides alkanes, this yeast can grow on other hydrophobic carbon sources, such as fatty acids, fatty alcohols, fatty acid methyl or ethyl esters and vegetable/animal oils [70]. However, short fatty acids have a toxic effect on *S. bombicola*. In fact, the shorter the fatty acid, the more toxic for the yeast, which is probably due to the interference of fatty acids with the membrane [81].

Sophorolipid production takes place when nitrogen, the growth-limiting nutrient for *S. bombicola*, is depleted from the medium, since this allows cells to enter stationary phase [82]. However, sophorolipid biosynthesis has in fact a complex regulatory system: for example, limitation of phosphorus was also described to increase the production of sophorolipids [83] and a heme binding damage resistance protein 1 (DAP1) was identified as a possible regulator of the cytochrome P450 proteins [84].

Besides sophorolipids, *S. bombicola* also produces range of extracellular proteins, commonly found in other fungi, and a high quantity of intracellular lipids. [5] The yeast is also capable of producing alkyl polyglucosides (APGs) in a single step starting from secondary alcohols. However, if grown on primary alcohols, which are less costly and commercially available, the strong activity of a fatty acid oxidase (FAO1) will convert them into fatty acids. It is possible to engineer the yeast into producing APGs from primary alcohols by deleting the fatty alcohol oxidase gene, as it will be describe in more detail in section 2.4.2. [85]

2.4.1 Sophorolipid biosynthetic gene cluster

In [59], through the examination of genome sequencing data, the full sophorolipid biosynthetic cluster was described (Fig. 2.6).

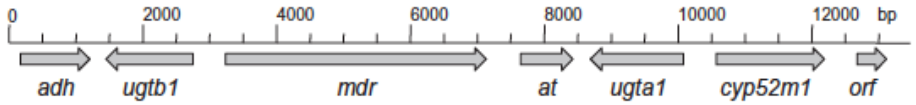

Figure 2.6: Genetic organisation of the sophorolipid gene cluster of *S. bombicola*. *adh*: putative alcohol dehydrogenase; *ugtB1*: second glucosyltransferase; *mdr*: transporter; *at*: acetyltransferase; *ugtA1*: first glucosyltransferase; *cyp52M1*: cytochrome P450 monooxygenase; *orf*: open reading frame with unknown function [59].

The cytochrome P450 monooxygenase CYP52M1 is the first enzyme of the sophorolipid biosynthetic pathway, being responsible for the synthesis of hydroxylated fatty acids. This is an essential step in sophorolipid synthesis, which explains why when this gene is knocked-out no sophorolipids could be detected in a HPLC (high-performance liquid chromatography) - ELSD (Evaporative Light Scattering Detection) analysis, while for the wild-type a clear production of lactonic sophorolipids was observed. [59]

The glucosyltransferase genes *ugtA1* and *ugtB1* were described in [86] and [87], respectively. These enzymes are responsible for the subsequent glycosylation steps of the hydroxylated fatty acid. These two different glucosyltransferases act in an independent manner, which means that disrupting the *ugtA1* gene has no influence on the second glucosylation step, which will occur provided there is a glucolipid to act as the acceptor. [86] On the other hand, the creation of Δ*ugtB1* deletions mutants leads to the production of glucolipids since the first glucosylation is still being performed. In fact, these mutants open the door for the creation of a time-efficient *in vivo* production process of these interesting glycolipid intermediates. [87]

The acetyltransferase gene *at*, responsible for acetylation of *de novo* synthesized sophorolipids in

S. bombicola, was identified in [88]. The same authors were able to create a Δat deletion mutant, which allowed the creation of a yeast strain producing unacetylated sophorolipids.

The putative multidrug resistance gene (*mdr*) encodes for the transporter responsible for sophorolipid secretion. Knocking-out this gene still leads to limited production of sophorolipid, since the transport occurs at the end of sophorolipid biosynthesis, so molecules still get a chance to accumulate. Furthermore, the detection of external sophorolipds with this strain could be attributed to less specific alternative secretion routes, which are less efficient than MDR proteins. [59]

The five genes above described are flanked by a putative alcohol dehydrogenase and a gene of unknown function at the 5′ and 3′ site respectively. Single knock-outs of these genes indicated that they do not seem to be essential for sophorolipid biosynthesis and do not affect cell viability. [59]

2.4.2 Industrial relevance

At the present, biosurfactants are still not able to compete with their chemically synthetized equivalents in terms of production costs. [25] However, *S. bombicola*'s capacity of producing large amounts of sophorolipids, together with the fact that it is capable of producing them from renewable sources [4] and even waste streams [89], opens the door to the opportunity of transforming this yeast into a platform organism for the production of (new-to-nature) biochemicals. [6]

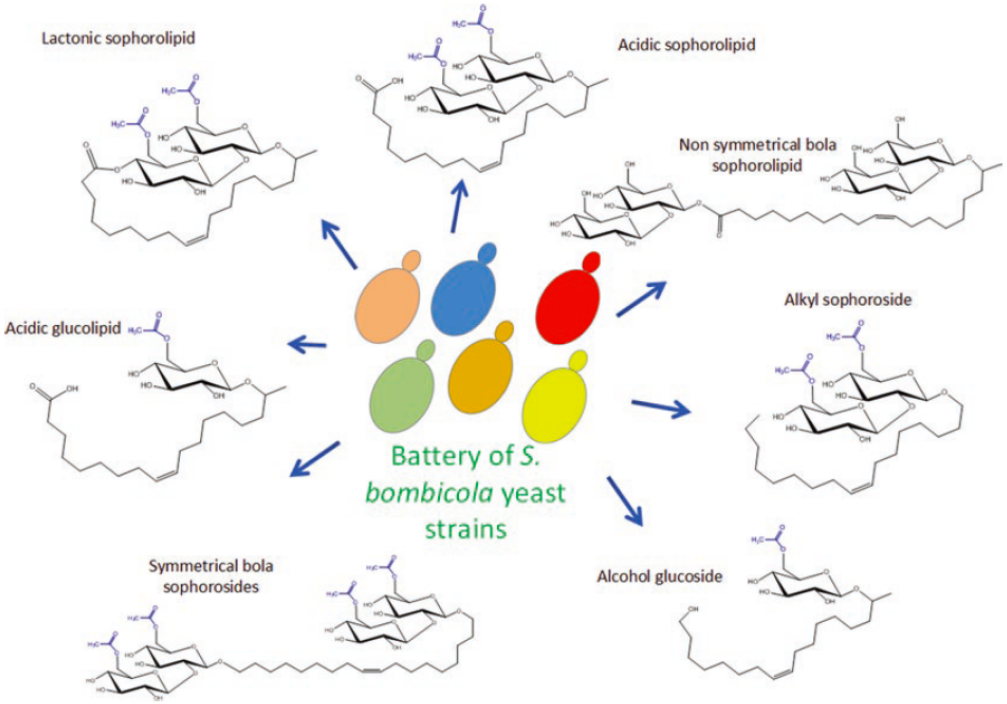

Figure 2.7: Overview of the molecular variety InBio.be has created by transforming *S. bombicola* into a platform organism for "new-to-nature" glycolipid production. [7]

To achieve the goal of transforming *S. bombicola* into a platform organism, the development of genetic manipulation tools for this yeast is essential. The InBio.be group has already started this pioneering work which allowed the creation of several new strains via knock-in's and knock-out's in the wild type *S. bombicola* strain. This enabled the generation of a battery of yeast strains whose production is directed towards glucolipids, bola-sophorolipids, non-acetylated sophorolipids, among others (Fig. 2.7).

Moreover, this strategy is also being used to achieve a selective production of either lactonic or acidic sophorolipids. [7]

The creation of the glucolipid-producing yeast strain was described in [87] and was achieved by deletion of the *ugtb1* gene in *S. bombicola*. Glucolipids constitute interesting intermediates for several kinds of biocatalytic or chemical conversions. For example, they can be used as substrates for different glycosyltransferases, leading to the incorporation of unique carbohydrate heads, which has an influence on their biological activity. [90] In the same way, the deletion of the acetyltransferase gene in [88] led to the creation of a strain producing unacetylated sophorolipids. This is a interesting variant of sophorolipids due to their decreased cytotoxicity and increased water solubility, which are important characteristics for applications in the pharmaceutical and detergent fields, respectively. Moreover, unacetylated sophorolipids have been used as starting molecules for synthesis of dispersible nanoparticles and glycolipid derivatives, in the production of glucolipids and fatty acids used for synthesis of polymers and as precursors for plastics and flavors. [88]

In [73], synthesis of bola-sophorolipids was achieved by disabling the sophorolipid acetyltransferase and the lactone esterase. The expected outcome of this genetic manipulation was a strain producing non-acetylated acidic sophorolipids. However, bolaform structures were also produced in high amounts due to the promiscuous activity of both UDP-glucosyltransferases from the core sophorolipid pathway. Even though their natural action is directed toward hydroxylated fatty acids and glucolipids, respectively, they apparently also display activity toward nonacetylated sophorolipid intermediates. Bola-sophorolipids have a unique structure which consists of a fatty acid molecule with a carbohydrate head on both sides. Thus, they are expected to behave completely different compared to the classic sophorolipids, which opens the door to new biotechnological applications. [73]

In [85], using a strain of *S. bombicola* with the fatty alcohol oxidase gene (*fao1*) knocked-out, the authors were able to increase the production of alkyl polyglucosides (APGs) and to produce novel diolpolyglucosides (diolpolyglucosides (DPGs)) from primary alcohols. APGs are non ionic surfactants which consist of polymeric acetals of glucose and fatty alcohols (Fig. 2.8). [91] The interest in APGs has been increasing in the past years due to their compatibility and synergistic effects when combine with other classes of surfactants, their excellent emulsifying properties and the fact that they are produced from natural, renewable resources and readily biodegradable. [92] In fact, when compared with classical sophorose lipids, these compounds display a better water solubility and a more effective reduction of the surface tension of water. APGs are naturally produced by *S. bombicola* when using secondary alcohol as a substrate. However, secondary alcohols, contrary to primary alcohols, are costly and few suppliers are available. Production of APGs from primary alcohols using *S. bombicola* is extremely difficult due to the strong activity of a fatty acid oxidase (FAO1). By disrupting the *fao1* gene in *S. bombicola* and growing the strain in medium containing 1-tetradecanol [85], the authors in were able to improved the yields of tetradecanol-based sophorosides. However, an unexpected high production of tetradecanediol-based sophorosides (Fig. 2.9) was also observed, which indicates that *S. bombicola* has ω and ω-1 activity toward alcohols. These DPGs are not produced by the parental strain since alcohols and diols are easily oxidised to their corresponding fatty acids by FAO1. [85]

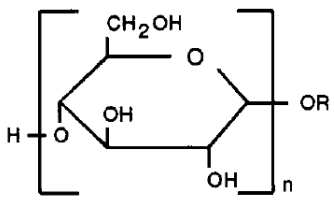

Figure 2.8: Alkyl polyglucoside: R= $C_8 - C_{14}$; n = 1-2. [91]

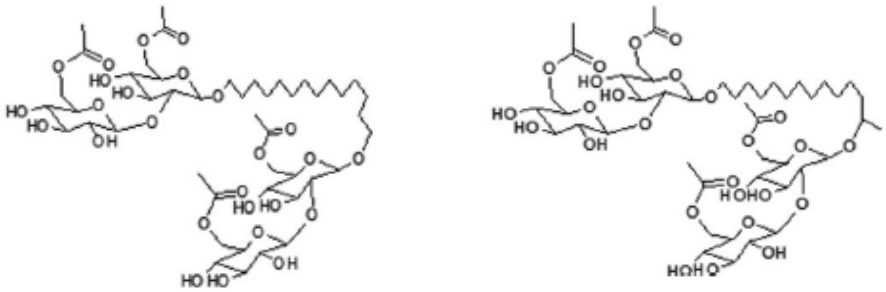

Figure 2.9: Structure of the DPGs produced by the Δ*fao1 S. bombicola* strain grown on 1-tetradecanol: 1,14-tetradecanediol (left) and 1,13-tetradecanediol (right) based sophorosides. [85]

In [6], the authors demonstrated that *S. bombicola* could be use as a platform organism for the production of tailor-made (glycolipid) biomolecules. To obtain this proof of concept, they aimed for the production of two totally different molecules: the bioplastic polyhydroxyalkanoate (PHA) and a new-to-nature cellobioselipid-biosurfactant. PHA's are biopolymers of β-hydroxy fatty acids naturally produced by a large number of organisms, mainly prokaryotes, which can be used as alternatives for synthetic plastics and/or polymers. The structure of both target molecules can be observed in Fig. 2.10. The structural variant of cellobioselipid (CBL) chosen is naturally produced by epiphytic fungi, of which *Ustilago maydis* is the most important example, and presents antifungal action.

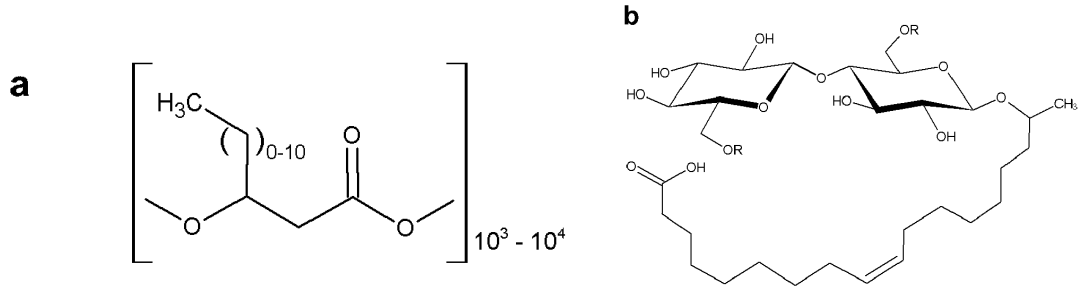

Figure 2.10: (a) Structure of PHA; (b) CBL's as produced by the engineered *S. bombicola*, R=H or $COCH_3$. [6]

The production of both PHA and CBL was achieved by using codon optimised expression cassettes obtained from PHAC1 synthase of *P. resinovorans* (PHAC1co) and UGT1 from *U. maydis* (UGT1co), respectively. These cassettes were integrated into the *S. bombicola* genome using a knock-out/knock-in strategy at the *locus* of one of the genes of the sophorolipid biosynthetic pathway - the *cyp52M1 locus* for the PHA strategy and the *ugtb1 locus* for the CBL strategy. This results in the production of the desire products whilst disabling naturally produced sophorolipids synthesis. Thus, expression of PHAC1co is under control of the 5' and 3' regulatory sequences of the *cyp52M1* gene and would thus be expected to be optimal under conditions for sophorolipids production. In the same way, the codon optimised UGT1co gene is flanked by the wild type *ugtb1* promotor and terminator sequences. This strategy of integrating new functionalities in *S. bombicola* by introducing genes at *loci* necessary for sophorolipid production safeguards the new biomolecules from sophorolipid contamination, while maintaining the advantages of a clear separation between growth and production phase. [6]

In conclusion, the successful production of both PHA and CBL in *S. bombicola* proofs that this yeast has the potential to become a flexible and robust production platform for the synthesis of tailor-made compounds. [6]

As a future perspective, work still has to be made to further expand the battery of yeast strains producing new and promising molecules with high efficiencies, since some strains still present low production levels. This is a consequence of the complex regulatory mechanisms of sophorolipid biosynthesis, which are still poorly understood, specially in the presence of modified substrates as it happens in engineered strains [93]. This is the key to decrease production costs, which is essential for the valorization of this technology. [7] Moreover, there is still a major drawback for the industrial scale production of the molecules produced by *S. bombicola* related with the variety of products obtained with this yeast. For commercialisation it is necessary to obtain a standardised product, since the physicochemical properties of the final product cannot vary according to the composition of the produced mixture. [67] Genetic engineering of *S. bombicola* can be used has a strategy to selective produce certain molecules. For example, in [67] the authors were able to obtain a strain producing 100% acidic sophorolipids (with mixed acetylation patters) by disabling the lactone esterase gene (*sble*), and a strain producing 99% of lactonic sophorolipids by using a *sble* overexpression strain.

2.5 Cytochrome P450 proteins

Cytochrome P450 monooxygenases (P450s), named for the absorption band at 450 nm of their carbon-monoxide bound form, are ubiquitously distributed in nature and constitute one of the largest superfamilies of enzyme proteins. [94, 95] They were first detected in a microsomal fraction of rodent liver, in 1958 [96], and later on they were identified in many microbial species, as well as various tissues of vertebrates, invertebrates and plants. The widespread distribution of P450s in nature demonstrates their important metabolic role. [97] In fact, these heme-thiolate proteins are involved, among others, in three main processes: biotransformation of drugs, xenobiotics, alkanes, terpenes, and aromatic compounds; biosynthesis of steroids, fatty acids, and bile acids; and degradation of herbicides and insecticides. [95] P450 enzymes catalyze regiospecific and stereospecific oxidative attacks on non-activated hydrocarbons. The reactions catalyse by them are extremely diverse since they can perform hydroxylations, N-, O- and S-dealkylations, sulphoxidations, epoxidations, deaminations, desulphurations, dehalogenations, peroxidations, and N-oxide reductions, amongst others. [98] These kind of reactions would require extremely high temperatures, but the involvement of these enzymes allow them to be performed at physiological temperatures. [94]

Regarding its structure, P450s are type-*b* hemoproteins, which means they have a protoporphyrin IX ring structure composed of four pyrrole subunits connected by methine bridges (Fig. 2.11). This structure produces four inward facing nitrogen atoms which coordinate a central, reduced iron atom (Fe_2^+). This atom makes four coordination bounds with the surrounding nitrogen atoms and one with the axial thiol ligand - a cysteine residue. [99]

The catalytic mechanism of P450s is not yet fully understood due to its complexity and involvement of a range of transient intermediates. However, the generally accepted catalytic mechanism is represented in Fig. 2.12. Initially, the substrate binds to the active site in the heme domain of the P450 protein, which prompts the release of a water molecule that covers the active site of the enzyme in the resting state. This leads to a conformational change bringing the flavins closer together to facilitate the transport of electrons down the chain. Consequently, there is an increase in the spin state and redox potential of the heme protein system, which allows one-electron reduction of the ferric complex (Fe_3^+) to a ferrous state (Fe_2^+). After that, molecular oxygen binds to the ferrous heme, which gives rise to a superoxide radical, which is very reactive and can disrupt the catalytic cycle. Then, a second reduction step occurs generating a negatively charged peroxo group which is very nucleophilic and, consequently, very short-lived. The complex is rapidly protonated twice due to hydrogen binding from surrounding water and amino acid side chains. This results in the release of a water molecule and the formation of the cytochrome P450 compound I intermediate (Cpd-I). This intermediate is widely accepted as being the final intermediate stage in the P450 catalytic cycle, and is thought to be the main oxidising agent. Thus, in the final step Cpd-I oxidises the bound substrate which is subsequently released as a hydroxylated

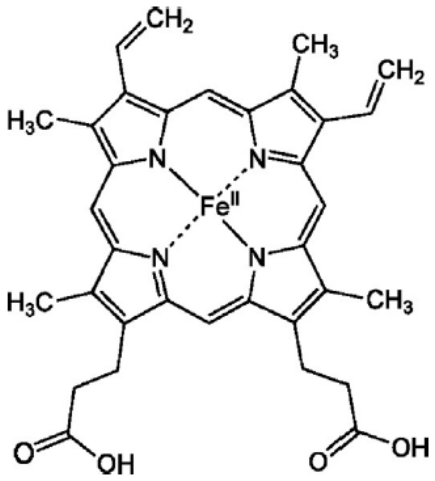

Figure 2.11: Type-b heme group of P450s which contains a protoporphyrin ring with a centrally bound iron atom. [99]

product. After that, water binds again to active site of the enzyme, which returns to its resting state. [99]

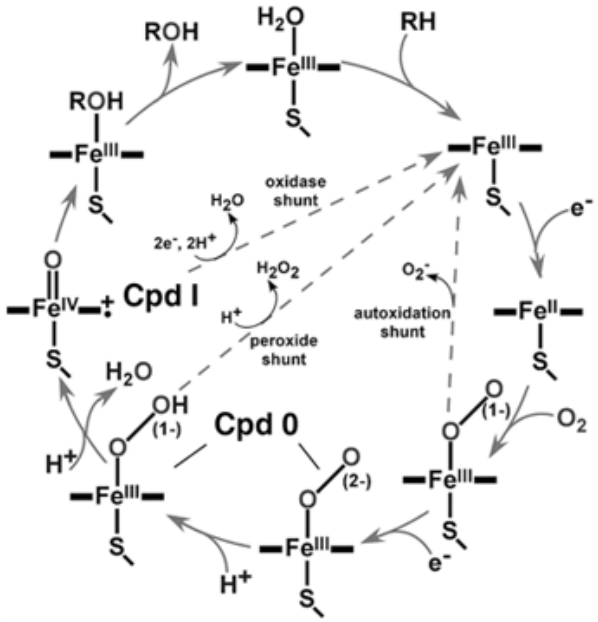

Figure 2.12: Graphical representation of the catalytic cycle of the cytochromes P450. [100]

As it was mentioned, dissociation can occur once molecular oxygen binds to the heme iron, which results in the cessation of catalysis. This can occur in one of the three ways (Fig. 2.12): (i) via the auto-oxidation shunt pathway, where the anion superoxide radical is formed after degradation of the peroxo-ferrous intermediate, (ii) via the peroxidase shunt pathway, where protonation of the hydroxyferric intermediate leads to dissociation and the release of hydrogen peroxide, or (iii) via the oxidase shunt pathway, where a water molecule is released after the double reduction and diprotonation of Cpd-I. These shunt pathways are related with the coupling efficiency of a P450, which is a measure of the number of electrons donated from NADPH which are successfully used for the catalysis of a bound organic substrate. The higher the coupling efficiency of P450 enzymes, the more favourable will be the

production of Cpd-I and, consequently, the production of the oxidised product. [99]

Most P450s use NAD(P)H-driven redox partner systems for delivery of the two electrons required for reduction of the ferric heme iron. The majority of mammalian P450s are attached to the endoplasmic reticulum (ER) (microsomal P450s) and are reduced by NADPH-cytochrome P450 oxidoreductase (CPR) - a FAD- or FMN-binding enzyme that is also attached to the ER membrane (class II P450 redox system). On the contrary, the P450s involved in steroid biosynthesis are linked to the inner mitochondrial membrane and interact with a soluble iron-sulfur protein (adrenodoxin), which sources electrons from the membrane-associated FAD-binding enzyme NADPH-adrenodoxin reductase (a class I system). On the other hand, prokaryotic P450s are soluble enzymes and most of them use a class I redox partner system. These class I/class II paradigm has been expanded in the last years after the discover of "unusual" electron transfer chains for different P450s. In fact, P450s can be classified into one of ten defined classes depending on the topology of the protein components involved in the electron transfer to the P450 enzyme. Of these 10 classes, there are two (class IX and X) whose P450s function without requirement for redox partners, while all the others require flavin-containing proteins, which are capable of being both oxidised and reduced, to facilitate electron transfer during catalysis. [98, 101]

In bacteria and yeasts, P450s play roles in the assimilation of carbon sources. More concretely, in hydrocarbon-assimilating yeasts, no hydroxylation system other than the cytochrome P450-dependent monooxygenase has been reported. [97] Thus, P450s were isolated in many n-alkane assimilating yeasts, such as *Yarrowia lipolytica*, *Candida tropicalis*, *C. maltosa* and more, recently, *S. bombicola*. They consist of n-alkane-inducible forms of P450 (P450alk), which, coupled with a CPR, catalyse the terminal or sub-terminal hydroxylation of alkanes. [102] This means they are used by the yeasts to convert alkanes into alcohols, which are then further converted into fatty acids enabling them to enter β-oxidation. Furthermore, as it was already mentioned, *S. bombicola* also produces sophorolipids which requires the synthesis of a hydroxy fatty acid via hydroxylation of a common saturated or unsaturated fatty acid of 16–18 carbon atoms. [95] This step is mediated by a P450 monooxygenase which belongs to this group, so these enzymes are also involved in hydroxylation of fatty acids, as it had already been observed for *C. tropicalis* [103].

2.5.1 Applications of P450s

The vast range of substrates used by P450s, together with their ability to perform regioselective and often stereoselective oxidation of those substrates, makes them attractive catalysts for a number of commercial applications. Namely, these enzymes find applications in synthetic biology and in oxychemical production, where they are used as a more economic alternative to synthetic chemistry. [101]

Recently, efforts have been made to extend even further the substrate repertoire of P450s by using techniques of protein engineering. These techniques are based on the rational redesign the P450 active site, or in random mutagenesis or directed evolution approaches, or even in a combination of both. *Bacillus megaterium* P450 BM3 (BM3, CYP102A1) consists of a natural fusion of the P450 heme group with a reductase domain (BMR), which means that this enzyme is catalytic self-sufficient. This, together with its high catalytic rate, makes of P450 BM3 a popular model system. Hence, variants of P450 BM3 with novel substrate selectivity have been produced using both rational (structure-guided) and random/directed evolution mutagenesis procedures. This has allowed the production of BM3 mutants with novel substrate selectivity which find applications in production of novel oxychemicals, such as human drug metabolites, which are used for drug safety testing. [101] However, natural CYP102A1 requires expensive NADPH as the electron donor, making it difficult to use CYP102A1 variants as the biocatalysts for producing bulk chemicals. Recently, variants of the P450 domain of CYP102A1 that can act with hydrogen peroxide have been created by directed evolution, but elevation of the affinity to hydrogen peroxide is still necessary to prevent destruction of the heme group by high concentrations of hydrogen peroxide. [104]

In order to establish a bioconversion process using P450s, the first steps necessary are screening for suitable P450s from the genome or cDNA libraries, and their expression in appropriate hosts. The next step is the modification of selected P450s by protein engineering to improve their catalytic properties, as it was mentioned above [104]. Several P450s and engineered variants are being used in synthetic biology approaches for valuable chemical production by fermentation in microbial cells. This system has the advantage of not requiring expensive NAD(P)H [104]. An example of the application of these systems is the production of the antimalarial artemisinin (Fig. 2.13.A) [105]. This compound is generated chemically from artemisinic acid, which in turn is originated from amorphadiene by three successive oxidation reactions performed in an engineered pathway in *Saccharomyces cerevisiae* using the P450 CYP71AV1 from *Artemisia annua*. Another example is a fermentation process using *Penicillium chrysogenum* which allowed the production of the cholesterol-lowering drug pravastatin (Fig. 2.13.B) [106]. This was done by introducing into the strain genes encoding compactin production and also an engineered form of the *Amycolatopsis orientalis* P450 CYP105AS1, to convert compactin to pravastin. This form of the enzyme containing five mutations (P450Prava), catalyses the stereo-selective hydroxylation of compactin to form the active pravastatin rather than its inactive epimer. [101]

In conclusion, the P450s superfamily shows a tremendous potential for practical applications, specially considering synthetic biology is now striving to replace synthetic chemistry for production of high-value chemicals. Since regio- and enantioselective hydroxylation makes chemical synthesis difficult, a bioconversion process using P450s proves to be quite attractive. [101, 104]

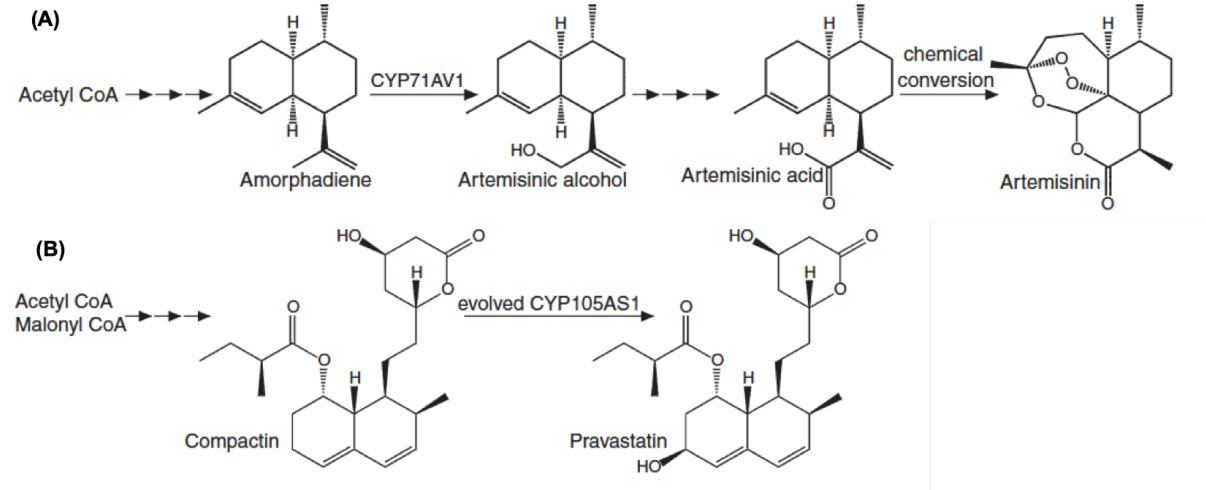

Figure 2.13: Examples of synthetic biology applications of P450 enzymes. **(A)** Production of antimalarial artemisinin from artemisinic acid produced in a reaction catalysed by CYP71AV1 in *S. cerevisiae*. **(B)** Production of pravastin in *Penicillium chrysogenum* using an engineered CYP105AS1. [101]

3 Material and Methods

3.1 Strains, media and culture conditions

3.1.1 Strains

Starmerella bombicola (SB) ATCC 22214 was used as the wild type (WT) strain. For transformation, the SB strains used are described in Table 3.1. For the growth trials, the following SB strains available in the laboratory were used: wild type; gene 1 knock-out (KO) strain (Δgene1); gene 2 KO strain (Δgene2); gene 3 KO strain (Δgene3); gene 1 and gene 2 KO strain (Δgene1Δgene2) and gene 1 and gene 6 KO strain (Δgene1Δgene6).

Table 3.1: *Starmerella bombicola* strains used for transformation.

Designation	Gene(s) knocked-out
PT36	*ura3*
Δgene6	gene 6
Δgene2Δgene6	gene 2 and gene 6
Δgene3	gene 3
Δgene3Δgene6	gene 3 and gene 6
Δgene1Δgene2	gene 1 and gene 2

Two markers could be present in *S. bombicola* strains. The *ura3* gene, encodes for an enzyme (Orotidine 5'-phosphate decarboxylase) that catalyses a reaction in the synthesis of pyrimidine ribonucleotides, so cells without this gene cannot grow unless uracil or uridine is added to the media. The hygromycin resistance gene encodes for hygromycine B phosphotransferase which inactivates this antibiotic. These two markers were essential for the selection of transformed cells using the selection media described in section 3.1.3.

The *Escherichia coli* Top10 strain was used for the *E. coli* transformations. *E. coli* Top10 strains previously transformed were also used to extract the respective plasmids. A *S. cerevisiae* strain previously transformed with a plasmid, which would be necessary for the *in vitro* tests, was also used for plasmid extraction.

Note: All strains are stored in 35% glycerol in cryovials at -80°C.

3.1.2 Non selective media

For the growth of *S. bombicola*, Yeast extract Peptone Dextrose (YPD) medium was used, while *E. coli* strains were grown on Lysogeny Broth (LB). The composition of both media is displayed in Tables 3.2 and 3.3, respectively.

Table 3.2: Composition of YPD medium. Note: Agar is only added when making solid medium.

Component	Concentration (g/L)
Glucose.H_2O (Cargill)	20
Yeast extract (Brenntach)	10
Bacto peptone (Becton & Dickinson)	20
Agar (Biokar diagnostics)	20

Table 3.3: Composition of LB medium. Note: Agar is only added when making solid medium.

Component	Concentration (g/L)
Tryptone (Biokar diagnostics)	10
Yeast extract	5
NaCl (Brenntach)	5
Agar	15

During growth trials, a 3C-agar solid medium was used to count Colony Forming Units (CFU). Its composition is displayed in table 3.4.

Table 3.4: Composition of the 3C-agar medium.

Component	Concentration (g/L)
Glucose.H_2O	110
Yeast extract	10
Urea (Sigma-Aldrich)	1
Agar	20

3.1.3 Selective media

Selective medium was used to favour the growth of transformed cells. For *S. bombicola*, the selective media used depends on the marker present on the knock-out cassette used for transformation. When the *ura3* marker was present, strains were selected on solid Synthetic Dextrose with Complete Supplement Mixture without uracil (SD+CSM-U) medium, which as the composition displayed in Table 3.5. When the hygromycin resistance gene was used as the marker, *S. bombicola* cells are grown on solid YPD medium supplemented with 1000 mg/L hygromycin B from *Streptococcus hygroscopus* (Sigma-Aldrich).

Table 3.5: Composition of solid SD+CSM-U medium.

Component	Concentration (g/L)
Yeast Nitrogen Based (YNB) without amino acids (Sigma-Aldrich)	6.7
Agar	20
Glucose.H_2O	20
Complete Supplement Mixture without uracil (CSM-U) (MP Biomedicals Europe)	0.77

Plasmid transformed *E. coli* strains were selected on LB medium supplemented with 50 mg/mL ampicillin (Sigma-Aldrich).

The *S. cerevisiae* strain used for plasmid extraction was grown in Synthetic Dextrose with Complete

Supplement Mixture without adenine (SD+CSM-A) medium (Table 3.6), since it was previously transformed with a plasmid containing the *ade2* marker.

Table 3.6: Composition of liquid SD+CSM-A medium.

Component	Concentration (g/L)
YNB without amino acids	6,7
Glucose	22
Complete Supplement Mixture without adenine (CSM-A) (MP Biomedicals Europe)	0.78

3.1.4 Production media

Lang medium, described by Lang *et al.* (2000) [107], was used for sophorolipid production during growth trials. After two days, the specific substrate was added in a concentration of 37 g/L. The composition of Lang medium is described in Table 3.7.

Table 3.7: Composition of Lang medium. [107]

Component	Concentration (g/L)
Glucose.H_2O	132
Yeast extract	4
Sodium citrate.2H_2O (Sigma-Aldrich)	5
NH_4Cl (Sigma-Aldrich)	1.5
KH_2PO_4 (Brenntach)	1
K_2HPO_4 (Brenntach)	0.16
$MgSO_4 \cdot 7\,H_2O$ (Sigma-Aldrich)	0.7
NaCl	0.5
$CaCl_2 \cdot 2\,H_2O$ (Sigma-Aldrich)	0.27

3.1.5 Culture conditions

S. bombicola was grown at 30°C and *E. coli* at 37°C. When liquid media was used, both cells were cultured at 200 rpm.

In case of liquid cultures, a pre-culture of 5 mL of medium was made before inoculating larger volumes. These pre-cultures were checked under the microscope for contamination before being inoculated at 1% in Erlenmeyers, with a ratio of liquid volume to total volume of 1/5, which were then covered with sterilised cotton plugs.

3.2 Molecular methods

3.2.1 Plasmid extraction

Plasmid were extracted from *E. coli* Top10 strains conserved at -80°C in 35% glycerol. Since all plasmids extracted contained the ampicillin resistance gene, the respective strains were inoculated in 5 mL of LB medium supplemented with 50 mg/mL of ampicillin and grown overnight at 37°C and 200 rpm. Plasmid extraction was performed using the Analytik Jena miniprep plasmid kit. In each extraction a volume of 4 mL of cell culture was used to elute the plasmid in 50 μL of elution buffer from the kit.

A plasmid extraction from *S. cerevisiae* was also performed. The respective strain was cultured in 5 mL of SD+CSM-A medium for two days at 30°C and 200 rpm. The plasmid was extracted using QIAprep Spin Miniprep Kit (Qiagen).

3.2.2 Genomic DNA extraction

The starting point to proceed to the genomic DNA (gDNA) extraction from *S. bombicola* was a stationary culture of yeast cells, which had to be inoculated 1 or 2 days before the extraction. The cells were centrifuged and washed with *Milli-Q*® water before proceeding to the disruption of the cell wall and nuclear envelope. This step required the addition of 200 μL of breaking buffer, whose composition is presented in Table 3.8, and 200 μL of glass beads (Sigma-Aldrich). A volume of 400 μL of Phenol:Chloroform:Isoamyl alcohol (PCI) saturated with TE buffer (Sigma-Aldrich) was also added to the mixture to allow the precipitation and removal of proteins. After vortexing for 10 minutes, the mixture was centrifuged for 10 min at 13K rpm, which allows the formation of an aqueous top layer that contains the DNA. The aqueous layer was recovered and 1 mL of absolute ethanol (Fisher) was added to precipitate DNA, which was then recovered by centrifugation (13K rpm at 4°C for 10 min) and dissolved in 200 μL of TE buffer - 10mM Tris, 1mM EDTA (Sigma-Aldrich). After performing the step of precipitation with ethanol and centrifugation one more time, two washing steps using 1 mL of 75% ethanol were still executed in order to remove all traces of organic solvents. Finally, the pellet was left to dry by contacted with air and after all ethanol had evaporated, the DNA was dissolved in 50-100 μL of TE buffer.

Table 3.8: Composition of the breaking buffer used to perform gDNA extraction.

Component	Concentration
Tris	50 mM
NaCl	100 mM
SDS (Sigma-Aldrich)	1%
Triton X-100 (Sigma-Aldrich)	2%
EDTA	1 mM

3.2.3 Polymerase Chain Reaction

DNA was amplified with Polymerase Chain Reaction (PCR) using different DNA polymerases. The choice of which polymerase should be used in each case depends on the purpose of the final PCR product. High fidelity polymerases are more costly, but their proofreading activity decreases the probability of having mismatched nucleotides in the generated PCR product. For this reason, high fidelity polymerases, such as PrimeSTAR® HS DNA Polymerase (Takara, Clontech), PrimeSTAR® GXL DNA Polymerase (Takara, Clontech) and Q5® High-Fidelity DNA Polymerase (New England Biolabs Inc. (NEB)), were used when the aim was to amplified DNA to be used in transformation and plasmid assembly. On the contrary, when the goal of the PCR was to detect a certain fragment of DNA in order to identify correctly transformed cells (colony PCR), proofreading activity was not necessary. In this cases, the polymerases used were Taq DNA polymerase (NEB) and OneTaq® Quick-Load 2X Master Mix with Standard Buffer (NEB). Tables 3.9 to 3.13 show the reaction mixtures used for each polymerase and their respective thermocycling conditions.

Besides the components mentioned in Tables 3.12 and 3.13, to perform colony PCR, a colony also had to be added to the reaction mixture. For that, the intended colony was picked-up from the selective plate using a sterile pipette-tip and suspended in the *Milli-Q*® water previously added to the PCR reaction tube. For Yeast colony PCR, before adding the other components of the reaction mixture, it was necessary to break the cells by a freezing and thawing treatment. The same was not necessary for *E. coli* colony PCR. The freezing and thawing treatment consisted in repeating three times the following steps: keep the mixture of cells in *Milli-Q*® water in the -80°C freezer until frozen and after microwaved it for 2 min.

Table 3.9: Reaction mixture composition (left) and thermocycling conditions to perform PCR using PrimeSTAR® HS DNA Polymerase (Takara, Clontech). *Note: The quantity of template in the mixture should be around 50 ng. **T_a = annealing temperature.

Component	**Volume** (μL)
PrimeSTAR® HS (Premix)	20
Primer forward (10 μM)	2
Primer reverse (10 μM)	2
Template	0.2-2*
Milli-Q® water	up to 50 μL
Total	**50**

Cycles	**Time**	**Temperature** (°C)
1x	1 min	98
30x	10 s	98
	5-15 s	T_a**
	1min/kb	72
1x	10 min	72
	∞	16

Table 3.10: Reaction mixture composition (left) and thermocycling conditions to perform PCR using PrimeSTAR® GXL DNA Polymerase (Takara, Clontech). *Note: The quantity of template in the mixture should be around 50 ng. **T_a = annealing temperature.

Component	**Volume** (μL)
5X PrimeSTAR® GXL Buffer (Takara)	10
dNTP Mixture (2.5 mM each) (Thermo Scientific)	4
Primer forward (10 μM)	2.5
Primer reversed (10 μM)	2.5
Template	0.2-2*
PrimeSTAR® GXL DNA Polymerase	1
Milli-Q® water	up to 50 μL
Total	**50**

Cycles	**Time**	**Temperature** (°C)
1x	30 s	98
29x	10 s	98
	5-15 s	T_a**
	1 min/kb	72
1x	10 min	72
1x	∞	16

The appropriate primers (Integrated DNA Technologies) for each PCR reaction were chosen using the *Clone Manager* Sci-Ed software, which also allows to determine the optimal annealing temperature between the primers and the template. In case the necessary primers were not designed yet, this was done using the same program, which allows to check the following parameters: GC content (%GC), melting temperature, dimers, stability, GC clamps, runs and hairpin. This way, the primer length was manipulated in order to achieve a GC content of approximately 50% and a stability parameter of 1.2 or above, which meant that not always primer secondary structures and GC clamps could be avoided.

Table 3.11: Reaction mixture composition (left) and thermocycling conditions to perform PCR using Q5® High-Fidelity DNA Polymerase (NEB). *Note: The quantity of template in the mixture should be around 50 ng. **T_a = annealing temperature.

Component	**Volume** (μL)
5X Q5® Reaction Buffer (NEB)	10
dNTP Mixture (2.5 mM each)	1
Primer forward (10 μM)	2.5
Primer reversed (10 μM)	2.5
Template	0.2-2*
Q5® High-Fidelity DNA Polymerase	0,5
5X Q5® High GC Enhancer (optional) (NEB)	10
Milli-Q® water	up to 50 μL
Total	**50**

Cycles	**Time**	**Temperature** (°C)
1x	30 s	98
30 x	10 s	98
	20 s	T_a**
	20-30s/kb	72
1x	2 min	72
1x	∞	16

Table 3.12: Reaction mixture composition (left) and thermocycling conditions to perform PCR using Taq DNA polymerase (NEB).

Component	**Volume** (μL)
Milli-Q® water	11.325
Standard Taq reaction buffer (NEB)	1.5
dNTPs (2mM)	1.5
Primer forward (10 μM)	0.3
Primer reverse (10 μM)	0.3
Taq polymerase	0.075
Total	**15**

Cycles	**Time**	**Temperature** (°C)
1x	7 min	95
25x	30 s	98
	30 s	45-60
	1min/kb	68
1x	10 min	68
	∞	16

Table 3.13: Reaction mixture composition (left) and thermocycling conditions to perform PCR using OneTaq® DNA polymerase (NEB).

Component	**Volume** (μL)
Milli-Q® water	6.9
Primer forward (10 μM)	0.3
Primer reversed (10 μM)	0.3
OneTaq® Quick Load 2X Master Mix with Standard Buffer	7.5
Total	**15**

Cycles	**Time**	**Temperature** (°C)
1x	7 min	95
25x	30 s	98
	30 s	45-60
	1 min/kb	68
1x	10 min	68
	∞	16

3.2.4 Gel eletrophoresis

Gel eletrophoresis was performed after every PCR reaction in order to check if target DNA region was amplified. The gel used was composed of 1% agarose (Sigma-Aldrich) in 1x TAE buffer - 40 mM Tris, 20 mM acetic acid (Fisher Scientific) and 1 mM EDTA. The eletrophoresis was run in 1x TAE buffer for 25 minutes under a voltage of 120 V.

For PCR products which were intended to purify after, a volume of 2 μL of template was mixed with 2 μL Gel Loading Dye Purple (6X) (NEB). In the case of colony PCR, the entire volume of the PCR reaction mixture was loaded into the gel and, if Taq DNA polymerase was used, mixed with 2 μL of loading dye, since OneTaq® already contains it. A volume of 2 μL of 2-Log DNA ladder (0.1–10.0 kb) (NEB) was also loaded into the gel.

After running, the gel was put in a buffer containing ethidium bromide for a minimum of 20 minutes. After this time, the gel was transferred to a Bio Rad imaging system to visualise it under UV radiation and a picture was taken using Quantity One software.

3.2.5 PCR product purification

In case eletrophoresis confirmed the amplification of the correct DNA fragment, the PCR product was purified using innuPREP PCRpure Kit from Analytik Jena. The first part of the protocol consisted on the binding of the PCR products which was done by adding 500 μL of the Binding Buffer to the PCR reaction mixture and transferring it to a Spin Filter coupled to a Receiver Tube. After a centrifugation step, the Receiver Tube was discarded and the PCR fragments were eluted to an Elution Tube by adding 20 μL of the Elution Buffer into the centre of the Spin Filter and centrifuging again.

3.2.6 Circular polymerase extension cloning

Circular Polymerase Extension Cloning (CPEC) was used in order to create circular plasmids from linear DNA fragments (insert piece and back-bone). This method requires the existence of overlapping regions between the two pieces. In order to fulfil this requirement, the primers used in the PCR, in which the DNA fragments are amplified, contained not only the region which overlaps with the template (20 bps), but also a region at the 5' end which overlaps with the other fragment (20 bps). The CPEC cloning mechanism is represent in Fig. 3.1.

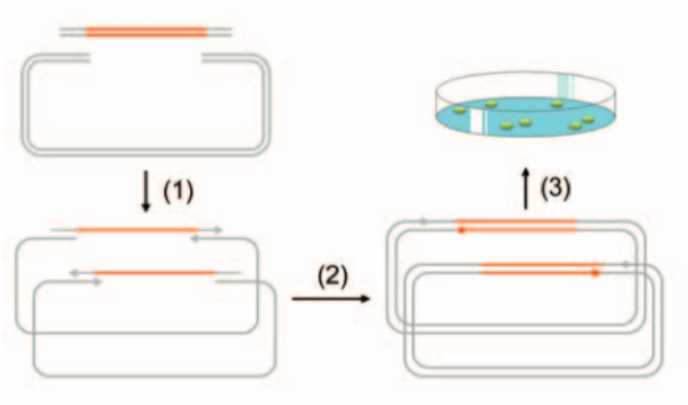

Figure 3.1: CPEC mechanism for cloning a single insert. **(1)** Denaturation and annealing of the overlapping ends of the back-bone and the insert. **(2)** Extension of the the hybridised insert and back-bone using each other as a template until they complete a full circle which has two nicks (marked by an arrow head); **(3)** Transformation of *E. coli* with the assembled plasmid. [108]

The composition of the reaction mixture used to perform CPEC is presented in Table 3.14 and the thermocycling conditions are the same used for a PCR reaction with Q5® High-Fidelity DNA Polymerase (Table 3.11), which is the polymerase used.

Table 3.14: Reaction mixture composition to perform CPEC using Q5® High-Fidelity DNA Polymerase.

Component	Volume (μL)
5X Q5 Reaction Buffer	5
dNTP Mixture (2.5 mM each)	5
DMSO	0.75
DNA fragments	2
Q5® High-Fidelity DNA Polymerase	0.5
Milli-Q® water	up to 25
Total	25

3.2.7 Sequencing

DNA Sanger sequencing of constructed plasmids was done by Macrogen inc. DNA solutions with a total volume of 10 μL containing 500 ng of DNA template and 2,5 μL of primer in *Milli-Q*® water were send to the company and the results were analyse using the *Clone Manager* Sci-Ed software.

3.3 Transformation techniques

3.3.1 *E. coli*

E. coli was used for storing and replicating plasmids. Two different techniques were used for transformation of *E. coli*: heat shock and eletroporation. Both of these techniques required the preparation of competent cells.

3.3.1.A Eletroporation

To prepare eletrocompetent cells, a pre-culture of *E. coli* Top10 cells was inoculated at 2% in a 250 mL Erlenmeyer containing Super Optimal Broth (SOB) (composition in Table 3.15) and grown until the optical density (OD) was past 0.6. After that, the culture was transferred to 50 mL falcon tubes which, after a period of 30 min on ice, were centrifuged (15 min, 4000 rpm, 4°C) to remove the supernatant. The pellet was ressuspended in 20 mL of cold *Milli-Q*® water and 10 mL of a glycerol-mannitol (GM) solution (composition in Table 3.16) was added bellow the cells, without disturbing the two layers formed. After centrifugation (15 min, 4000 rpm, 4°C), a supernatant with two layers was obtained, which were removed individually, starting from the top one. The pellet was ressuspended in cold GM solution and distributed into eppendorfs (50 μL in each), which are stored in the -80°C freezer.

Table 3.15: Composition of SOB medium. The pH of the medium must be adjusted to 7.0 using NaOH.

Component	Concentration
Tryptone	2% (w/v)
Yeast extract	0.5% (w/v)
NaCl	8.56 mM
KCl (Sigma-Aldrich)	250 mM
$MgCl_2$ (Sigma-Aldrich)	10 mM

Table 3.16: Composition of the GM solution.

Component	Concentration (% w/v)
Glycerol (Chem-Lab Analytical)	20
Mannitol (VWR)	1.5

To proceed to the transformation of the cells, 2 μL of the CPEC product was added into one of the previously prepared eppendorfs, which contain the eletrocompetent cells. This mixture was then transferred to a cold 0.2 cm sterile biorad® gene pulser cuvette. Eletroporation was performed in biorad gene pulser with a resistance of 200 Ω and a voltage of 2.5 kV. Immediately after, 1 mL LB medium was added to the cuvette and the mixture was then transferred to a falcon tube. The tube was incubated for 1 hour at 37°C and 200 rpm. After that period, 50μL of the cell culture were spread into the plate containing the selective medium (LB+ampicillin) and incubated at 37°C overnight.

3.3.1.B Heat shock

Contrary to eletroporation, the preparation of the cells for heat shock was done in the same day as the transformation. In order to make the cells competent a pre-culture of *E. coli* Top10 cells was inoculated at 10% in a 50 mL falcon tube, containing a minimum of 5 mL of LB medium, and incubated at 37°C and 200 rpm during approximately 3 hours. After leaving the falcon tube on ice for 1 hour, 1 mL of the culture was transferred to a sterile eppendorf and centrifuged for 30 seconds at 8000 rpm to remove the supernatant. A volume of 100 μL of cold sterile $CaCl_2$ (Sigma-Aldrich) 100 mM was added the eppendorf without pipeting up and down, since this could cause cell rupture. To proceed to

the transformation, 3μL of the CPEC product was added to the eppendorf and stirred. After a period of 20-30 minutes on ice, the eppendorf was put in a bath at 42°C for 45 seconds and then on ice again for 2 minutes. LB medium was added and the mixture transferred to a falcon tube and incubated at 37°C and 200 rpm for 1 hour. After that period, 50μL of the cell culture were spread into the plate containing the selective medium (LB+ampicillin) and incubated at 37°C overnight.

3.3.2 *S. bombicola*

In order to make knock-out mutants, *S. bombicola* was transformed by eletroporation with linearised DNA . Before transformation, cells had to be made competent, which was performed in a sterile environment and using sterilised solutions and *Milli-Q*® water. The solutions used were ice-cold and between every step the cells were kept in ice.

The first step was to prepare a culture of 50 mL of the correct strain in YPD medium with an OD between 0.6 and 1. The culture was prepared in an Erlenmeyer and transferred to a falcon tube to be centrifuged at 4000 rpm at 4°C for 5 minutes. The pellet obtained was washed with *Milli-Q*® sterile water two times and then dissolved in 2 mL of 1 M sorbitol (Dummy leverancier PMGE) and centrifuged at 4000 rpm at 4°C for 5 minutes. After, the pellet was incubated at room temperature for 15 min in a solution of 0.1 M Li-acetate (Sigma-Aldrich) and 2.5 mM of dithiothreitol (DTT) (Sigma-Aldrich) and centrifuged in the same conditions as before. The pellet obtained was washed with 2 mL of 1 M sorbitol and after centrifugation the pellet was suspended in 250 μL of 1 M sorbitol and divided into volumes of 50 μL which were put in sterile eppendorfs together with 50 μL of glycerol 70% and stored in the -80°C freezer.

To transform the previously prepared competent cells, 1 mL of cold sterile 1 M sorbitol was added to the eppendorf and centrifuged at 8000 rpm for 30 seconds. The supernatant was removed and the pellet was ressuspended in 50 μL of 1 M sorbitol and mixed with 500-1000 ng of linearised DNA. The mixture was transferred to a 0.2 cm sterile biorad® gene pulser cuvette and eletroporation was performed in biorad gene pulser with a resistance of 200 Ω and a voltage of 1.7 kV. Immediately after, 1 mL of YPD was added and the mixture was then transferred to a falcon tube. The tube was incubated for 1 hour at 30°C and 200 rpm. After that period, 50μL of the cell culture were spread into the plate containing the selective medium and incubated at 30°C until colonies appeared.

3.4 Analytical techniques

3.4.1 Nanodrop

After purification of PCR products or gDNA extraction, the spectrophotometer Nanodrop 2000 from Thermo scientic® was used to quantify and assess purity of DNA. The concentration of DNA was determined by measuring the absorbance at 260 nm. To analyse purity, absorbance at 280 nm and 230 nm was also measured. If the 260/280 ratio was lower than 1.8 it was an indication of protein pollution, whereas a 260/230 ratio lower than 2 indicated solvent pollution.

3.4.2 pH measurement

pH was measured with a Five easy F20 Mettler Toledo pH/mV meter.

3.4.3 Optical density

OD was measured with a Jasco V-630 Bio Spectrophotometer at a wavelength of 600 nm. Samples were diluted, if necessary, in order to obtain an OD value inferior to 1. Each sample was measured three times for statistical relevance.

3.4.4 Thin layer chromatography

Thin Layer Chromatography (TLC) was performed in order to detect sophorolipids produced during a growth trial. Samples were spotted in a TLC silica plate approximately at 1 cm from the bottom. The plate was put in a glass chamber filled with the eluent, composed of chloroform (Dummy leverancier PMGE), methanol (Sigma-Aldrich) and water (65:15:2, V:V:V), at a lower level than the sample spots. The eluent migrated on the plate and, when it reached almost the top, the plate was removed from the chamber and dried to remove the solvents. Lastly, a detection step was performed by immersing the plate in a 10% H_2SO_4 (Sigma-Aldrich) solution and heating it with a paint dryer until spots appeared on the plate.

3.4.5 Ultra performance liquid chromatography

Ultra Performance Liquid Chromatography (UPLC) was performed to determine glucose consumption and to analyse the type of sophorolipids produced during growth trials. The analysis was performed on a ACQUITY UPLC® system (Waters) and Evaporative Light Scattering (ELS) was used as the detection method. To perform the analysis, each sample, corresponding to a different strain and period of time, was placed on a well of a 96-well Sample Collection Plate.

3.4.5.A Glucose analysis

The UPLC analysis for glucose quantification was performed with a Waters Acquity UPLC BEH Amide 2.1x100 mm column (Hydrophilic Interaction Chromatography). An eluent, composed of 0.25% triethylamine (TEA) (Sigma-Aldrich), 75% acetonitrile (ACN) (Fisher Scientific) and 24.75% *Milli-Q*® water, was used in isocratic mode with a flow rate of 0.6 mL/min during a period of 4 min for each sample. The injection volume of each sample was 1 μL. After running all the samples in the column, there was still an injection of *Milli-Q*® water before the storage of the column in ACN.

3.4.5.B Sophorolipid analysis

The UPLC analysis for sophorolipid detection was performed with an Acquity UPLC CSHTM C18 1.7 μm, 2.1x50 mm column (Reversed Phase Chromatography) . A gradient elution with two solvents, 0.5% acetic acid (Fisher Scientific) in *Milli-Q*® water and 100% ACN, was used to separate the components. For each sample, the gradient started at 5% ACN and linearly increased to 95% in 10 min with a flow rate of 0.6 mL/min. The injection volume of each sample was 2 μL. After running all the samples in the column, there was an injection of ethanol before the storage of the column in ACN.

3.4.6 Liquid chromatography – mass spectrometry

Liquid Chromatography – Mass Spectrometry (LC-MS) was used in order to identify the type of sophorolipids present in a product sample. This analysis was performed in a Exactive™ Plus Orbitrap Mass Spectrometer (Thermo Scientific). The column and elution mode for the chromatography were the same as describe in section 3.4.5.B. The mass spectrometry had an m/z range of 215-1800 and Heated-Electrospray Ionization (HESI) in negative mode was used for ion source. The spectrometer was calibrated with a Pierce™ LTQ ESI Positive Ion Calibration Solution and Pierce™ Negative Ion Calibration Solution.

3.5 Shake flask growth trials

Growth trials with *S. bombicola* strains were performed in 250 mL Erlenmeyers which were incubated during 10 days at 30°C and 200 rpm. Each strain was cultured in triplicate in Lang Medium, and the substrate was added after 48h at a concentration of 37 g/L.

During the period of incubation, samples were taken to measure OD, pH, CFU, glucose concentration and sophorolipid production. During the first five days, these samples were taken everyday (except for CFU which was measure every two days and sophorolipid production whose samples were only taken after the addition of the substrate) and, after this period, every one or two days. To measure the pH and glucose, a sample of 500 μL of culture was centrifuged for 5 min at 13K rpm and the supernatant divided into two tubes: 50 μL to measure glucose, which was diluted 20 times in *Milli-Q*® water, and the rest to measure pH. To prepare the samples to detect sophorolipid production, 1 mL of absolute ethanol was added to 500 μL of culture. After shaking for 5 minutes, the mixture was centrifuged for 5 min at 13K rpm and the supernatant recovered. In order to count CFU, 90 μL of saline solution (0.9% NaCl) and 10 μL of culture were added into a well of a 96-well microplate and successive 10 time dilutions were applied until a dilution of 10^{-8} was obtained. A plate containing 3C medium was divided into 6 equal parts and three dilutions (10^{-5}, 10^{-6} and 10^{-7} or 10^{-6}, 10^{-7} and 10^{-8}) were plated in duplicate, using a volume of 10 μL, into each corresponding part. The plate was incubated at 30°C until individual colonies were formed.

4 Results and Discussion

4.1 Objectives

The goal of this master's thesis emerges from the problem already mentioned in section 2.4.2: the versatile oxidative power of *S. bombicola* leads to the production of a mixture of different products, which impedes its commercialisation. The natural mixture produced by SB can vary in terms of acidic, lactonic or bolaform sophorolipids, with different degrees of acetylation, number of unsaturations and position of the hydroxyl group. The composition of the produced mixture is influenced by the culturing conditions (pH, presence of citrate, aeration, etc.), but the complex regulation of sophorolipid biosynthesis is not yet understood. In this master's thesis, the influence of the enzymes responsible for oxygenation reactions in *S. bombicola* was studied. Several of these enzymes have already been identified in *S. bombicola*, however their role is not yet clear and only genomic information is available for most of them.

The general objective of this master's thesis was to start the construction of a database of observed oxidation reactions using engineered SB strains. A more particular objective was using the production of product A with SB engineered strains as a proof of concept to find the enzyme responsible for hydroxylation of a specific substrate in SB. Therefore, the first part of the experimental work consisted in genetic engineering SB strains by knocking-out genes expected to influence sophorolipid production. The second part of the experimental work was testing these strains, or strains previously produced in the laboratory, by performing shake flask growth trials, in order to determine the type of sophorolipids produced by them.

4.2 Creation of a plasmid containing the gene 4 KO cassette

The mechanism used to knock-out genes in *S. bombicola* strains was homologous recombination. In a successful transformation, the intended gene is replaced by the marker that permits to identify the correct transformants (*ura3* or hygromycin resistance gene). In order to achieve this result, SB has to be transformed with a cassette which contains the marker sequence flanked at the 5' and 3' side by 1000 bps of the sequence upstream and downstream the target gene, respectively. This regions will be designated, respectively, by homologous region (HR) 1 and 2.

In the case of the gene 4, the selection marker is the *ura3* gene, which has to include a promoter upstream (pura3) and a terminator downstream (tTK). The plasmid also contains a bacterial marker, the ampicillin resistance gene (AmpR), which permits the selection of plasmid-containing bacteria and also ensures its retention by selective pressure. This way, the plasmid is assembled in *E. coli*, which allows for it to be isolated and sequenced afterwards, since PCR is an error-prone technique. Furthermore, even though it is not annotated, the plasmid also contains an origin of replication, which allows its replication inside *E. coli*.

The region of the plasmid which contains the homologous regions of the gene 4 and the *ura3* marker constitutes the KO cassette, whereas the other part of the plasmid is called back-bone (BB).

In order to construct the plasmid, the pieces that constitute it have to be amplified by PCR. The region of the selection marker, as well as the BB of the plasmid, can be amplified using as template another available plasmid with the *ura3* marker. The homologous regions of the gene 4 have to be amplified using gDNA of *S. bombicola* as template.

The plasmid can be constructed using two different strategies. The first one consists in amplifying separately the four regions of the plasmid (HR 1, HR 2, *ura3* marker and BB) and performing a CPEC reaction to assemble the plasmid. This is the more simple strategy, but it involves performing the CPEC reaction with four fragments, which is not always effective. Since this strategy did not work in this case, another procedure was followed, which consists in performing two CPEC reactions. In the first one, a fragment containing the homologous regions together with the gene 4 (HR 1+ N2 +HR 2) was assembled together with the back-bone. After, using this plasmid as a template, a PCR reaction was performed to obtain a linear fragment containing the HR 1, the BB and the HR 2. Finally, this fragment was assembled together with the *ura3* marker in the second CPEC reaction. The details of each one of these steps are explicit in the following sections.

4.2.1 First CPEC reaction

The fragments amplified contain approximately 20 bps at each end of overlapping regions with each other, which allows for them to be assembled in a CPEC reaction. Electrocompetent *E. coli* cells were then transformed with the resulting CPEC mixture. The correct transformants were selected by performing an *E. coli* colony PCR using the OneTaq® mix and 52°C as the annealing temperature. The primers used were P1131 and P725 (Table A.1), which anneal with the back-bone before the HR 1 (forward primer) and after the HR 2 (reverse primer), respectively. Gel eletrophoresis with the product of the *E. coli* colony PCR allowed to confirm the presence of three colonies whose PCR product had the expected size considering the primers chosen (3.9 kb). One of this colonies was cultured in LB medium containing ampicillin, which allowed to extract the plasmid with a concentration of 355 ng/μL.

4.2.2 Second CPEC reaction

For the second CPEC reaction, the previously constructed plasmid was used as the template to amplify the HR 2+ BB +HR 1 fragment. The other CPEC piece, the *ura3* marker, was amplified using as template another available plasmid with the same marker.

E. coli cells were transformed by heat shock with the product resulting from the CPEC reaction. Twenty of the obtained colonies were tested with *E. coli* colony PCR using two sets of primers. Primers P1526 and P295 bind, respectively, to the back-bone before the HR 1 (forward primer) and to the *ura3* marker (reverse primer), so they are meant to test the presence of the HR 1 between the back-bone and the marker. On the other hand, primers P2220 and P725 bind, respectively, to the *ura3* marker (forward primer) and to the back-bone after the HR 2 (reverse primer), so they serve to test the presence of the HR 2. The colony PCR was performed using OneTaq® mix and 52°C as the annealing temperature. After the gel eletrophoresis, it was confirmed the presence of bands with the correct size (1466 bps for HR 1 and 1251 bps for HR 2) for all colonies tested. One of this colonies was chosen and, after being culture overnight in LB+ampicillin, it was stored in triplicate in 35% glycerol in cryovials at -80°C.

4.3 Knock-out of the gene 1, gene 4 and gene 5 genes in SB strains

Oxidising enzymes can be found in several hydrocarbon-assimilating yeasts, since they are involved in terminal or sub-terminal hydroxylation of alkanes which allows for them to be metabolised. Furthermore, this type of oxidising enzymes have also been identified as responsible for hydroxylation of other substrates. The strategy used to understand the influence of several oxidising enzymes on sophorolipid production was analysing the sophorolipid mixture produced by engineered strains, which have a gene

that encodes for an oxidising enzyme deleted. On the other hand, since the role of the enzyme encoded by gene 1 has already been identified, it was intended to also test the combination of deleting this gene together with other genes.

In this section it is described the attempt to KO three genes which encode for oxidising enzymes. Since genes 4 and 5 have never been deleted before, the strains in which this was attempted were the PT36 and Δgene6 strains. On the other hand, the KO of the gene 1 gene was attempted on the Δgene3, Δgene3Δgene6 and Δgene2Δgene6 strains. The strains which contain a deletion in gene 6 were intended for the particular objective of this master's thesis of producing product A from a specific substrate. For this to be possible, there cannot be activity of enzyme 6, since this enzyme oxidises this substrate. However, the isolated production of product A has not yet been possible, since there is also an enzyme with hydroxyation activity towards this substrate, which has not yet been identified. The combination of deletions in the gene 6 together with other oxidising genes was the method used to discover which enzyme is responsible for this step.

4.3.1 Plasmids

The plasmids who contain the knock-out cassettes of the genes 1 and 5 have the same structure as the plasmid containing the gene 4 KO cassette. However, the gene 1 plasmid contains an hygromycin resistance marker instead of the *ura3* marker. The presence of a different marker permits the KO of the gene 1 gene in strains which already contain the *ura3* marker as a result of previous transformations.

For the genes 1 and 4, the respective plasmids were extracted from the *E. coli* strains which contained them with the following concentrations, respectively: 260 ng/μL and 350.9 ng/μL. For the gene 5, the plasmid was already available, with a concentration of 304 ng/μL, as a result of a previous failed attempt to transform SB strains. For this reason, this plasmid was send for sequencing in order to re-evaluate if it was correctly constructed. Five primers were chosen to sequence the part of the plasmid which contains the KO cassette. However, the sequencing results for the *ura3* marker region were not conclusive, due to the weak quality of the sequencing on the regions where the primers bind. For this reason, three more primers were chosen to sequence the *ura3* marker in the forward direction and another three in the reverse direction in order to have overlapping sequencing results. This allowed to conclude that there were no bases mismatches in this region.

4.3.2 Linearisation of the knock-out cassettes

The gene 1 KO cassette (3936 bps) was amplified with PrimeSTAR® HS DNA Polymerase and 55°C as the annealing temperature. The 4 and 5 genes KO cassettes (3712 bps and 3788 bps, respectively) were amplified with PrimeSTAR® GXL DNA Polymerase and 50°C as the annealing temperature. The sets of primers used for the 1, 4 and 5 genes cassettes were, respectively, P1946 and P64; P10193 and P10194; P10191 and P10192.

After gel eletrophoresis confirmed the correct length of the PCR products, they were purified with a concentration of 336.7 ng/μL, 286.7 ng/μL, 322.9 ng/μL, for the 1, 4 and 5 genes KO cassettes, respectively.

4.3.3 Transformation of *S. bombicola* strains

A volume 3μL of linear DNA (approximately 1000 ng) was used to transform each of the strains mentioned in the beginning, which were then plated in the appropriate selective medium: SD+CSM-U for the strains transformed with the 4 and 5 genes KO cassettes and YPD supplemented with hygromycin for the strains transformed with the gene 1 KO cassette. For this last one, the strains were previously diluted 100 and 1000 times before plating since previous attempts to knock-out the gene 1 gene had

led to an overgrow of cells on the plate, which did not allow to isolate single colonies in order to perform Yeast Colony PCR. More details explaining this occurrence will be given in section 4.3.4.

After 5 days of incubation at 30°C multiple colonies appeared on the plates. Ten colonies from each transformed strain were picked-up from the plates and Yeast Colony PCR was performed using OneTaq® mix and 45°C as the annealing temperature. A positive control, which amplifies the alcohol dehydrogenase gene using primers P70 and P71, was also included in the mixture. The sets of primers used to test the insertion of the 1, 4 and 5 genes KO cassettes were, respectively, P954 and P755; P10904 and P621; and P1834 and P621. One of the primers binds to the genome of *S. bombicola* before the region where the cassette begins (forward primer) and the other one binds to the marker present in the cassette (reverse primer).

Despite multiple attempts, it was not possible to identify a correct transformant, since the eletrophoresis gel either showed incorrect size bands or no bands at all.

4.3.4 Discussion of the transformation results

In the case of the strains transformed with 4 and 5 genes KO cassettes, even before performing the Yeast Colony PCR, it was already foreseen that the transformation would not be successful, since a lot of colonies (more than 100) were obtained. This an indication that random integration could have occurred in the genome, instead of the indented homologous recombination. The KO method used relies on *S. bombicola* using homologous recombination as the repair mechanism of DNA double-strand breaks. However, a low transformation efficiency in *S. bombicola* could be explained by the fact that this yeast prefers in fact Non-Homologous End Joining (NHEJ), which would lead to non-homologous integration events. Even though there is no literature clearly confirming this hypothesis for *S. bombicola*, the fact that the knock-out of genes only occurs when using large homologous regions [7] is an indication that this is in fact the case. Furthermore, there were already identified other yeasts which prefer NHEJ to homologous recombination, such as *Yarrowia lipolytica* [109] and *Kluyveromyces lactis* [110].

Another reason behind the failed transformations attempts could be that the 4 and 5 genes are essential for cell viability. This hypothesis could be tested by analysing the gene expression levels of these genes during the exponential growth of SB in YPD medium, for example by quantitative PCR. In case these genes are involved in sophorolipid biosynthesis, which only occurs when cells enter stationary phase, it would not be expected for them to be induced during exponential growth. On the contrary, if they were involved in the primary metabolism of cells, they would be upregulated during the exponential growth, which could mean that they might be essential for the cells. If that was the case, they could not be directly knocked-out and another strategy had to be applied, for example, gene silencing using RNA interference (RNAi). This strategy reduces gene expression at the mRNA level (knockdown) by introducing into the cell specifically designed small interfering RNA (siRNA) or microRNA (miRNA), which, after being cleaved into smaller RNA fragments, will bind to the target complementary RNA. After this association, the target mRNA will be cleaved by a protein from the RNA-induced silencing complex (RISC), inhibiting the expression of the protein that it encodes [111]. However, this strategy is outside the scope of this master's thesis and was just presented has an example of future efforts that can be made to discover the function of these genes in sophorolipid biosynthesis.

On the other hand, the inability to knock-out the gene 1 gene cannot be due to the loss of cell viability, since this gene was already deleted successfully in the PT36 strain in the past. The fact that without previous dilution of the plates, an overgrow of cells was observed was a strange occurrence that could indicate that the selection system was not working. This was corroborate when the parental strain was plated on the same selective medium and the same overgrow was observed. This could mean that, contrary to what was expected, the strains being used for transformation already contained the hygromycin resistance gene. However, the same result was observed when plating the wild type strain, which did not undergo any genetic manipulation. A more plausible hypothesis to explain the lack

of selection is that the hygromycin antibiotic used to prepare the selective media was not working as expected, which can be explained by the recent change in supplier.

In conclusion, the inability to knock-out the 4 and 5 genes can imply the development of a new strategy. On the contrary, the knock-out of the gene 1 gene can have a simpler resolution, which passes by determining the optimum concentration of the hygromycin antibiotic supplemented to the YPD medium or by the acquisition of a new antibiotic bottle. Unfortunately, the resolution of this problem was not possible during the experimental period of this master's thesis. However, a different strategy was applied which allowed the creation of the Δgene1Δgene2Δgene6 SB strain.

4.4 Shake flask growth trials

4.4.1 Growth trial with WT, Δgene1, Δgene2 and Δgene3 SB strains

In this growth trial, strains with single knock-outs in genes which encode for oxidising enzymes were tested in order to evaluate the type of sophorolipids produced in comparison with the wild type. The goal of this experiment was to understand which oxidising enzymes are involved in sophorolipid biosynthesis and what is their influence on the type of products obtained.

The experiment was conducted in triplicate, so a total of 12 shake flasks were used. The strains were grown in approximately 50 mL of production medium described by Lang *et al.* (2000) [107] and the specific substrate was added after 48h, in order to begin the sophorolipid production. The parameters OD, pH, CFU and glucose concentration were measured and its values were displayed over time. Regarding the sophorolipids, samples of days 4, 5, 7, 8 and 9 of the growth trial were analysed in the UPLC in order to detect the type of sophorolipids produced.

4.4.1.A Optical Density

The variation of OD during the first three days of the growth trial is represented in Fig. 4.1. The OD was not represented over time for the rest of the days since the addition of the substrate, and posterior product formation, affects these values, which stop being representative of the amount of biomass present in the medium.

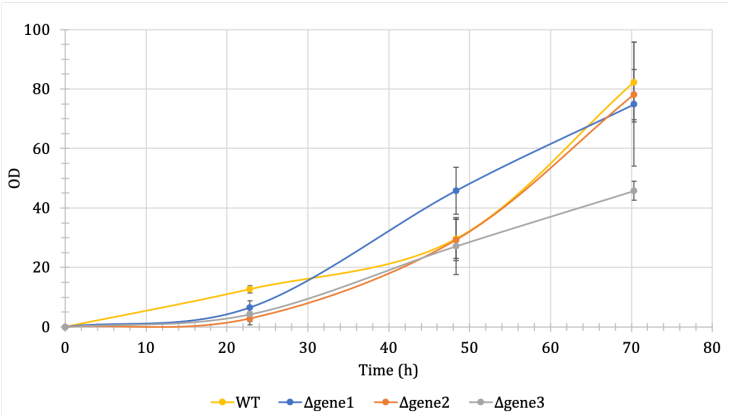

Figure 4.1: Variation of OD measured at 600 nm for the WT, Δgene1, Δgene2 and Δgene3 *S. bombicola* strains during the shake flask growth trial. The vertical error bars represent SD-values.

The strains Δgene1 and Δgene2 show a similar trend in terms of OD values when compared with the WT. On the other hand, the Δgene3 strain seems to have a sightly slower increase in OD. To

understand if there is a statistically significant difference between the Δgene3 strain and the WT, a t-test (Two Samples assuming unequal variance) between the OD values of the two was performed. The p-value obtained was 0.596, which is higher than 0.05, meaning that the OD values observed for the two strains are not statistically different.

4.4.1.B pH values

In Fig. 4.2, the variation of the pH during time can be observed. A drop in the pH over time was expected due to the utilisation of the nitrogen source and excretion of organic acids, such as citrate and isocitrate [112]. Once again, the Δgene3 strain seems to show a slightly slower decrease comparing with the WT, but the t-test confirmed that the difference is not statistically significant since the p-value obtained was of 0.179.

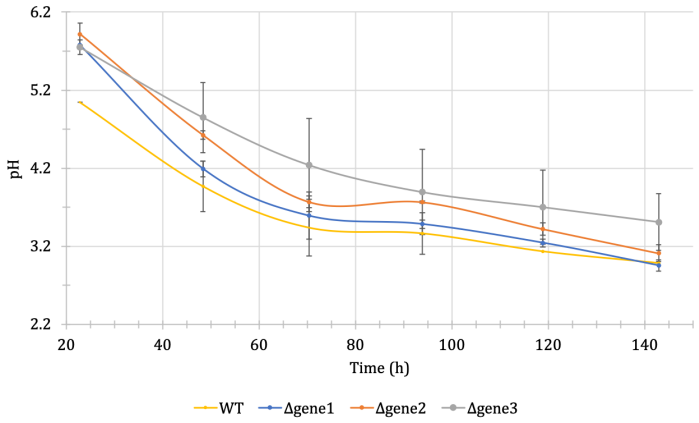

Figure 4.2: Variation of pH for the WT, Δgene1, Δgene2 and Δgene3 *S. bombicola* strains during the shake flask growth trial. The vertical error bars represent SD-values.

4.4.1.C Colony Forming Units

The value of CFU/μL was determined in days 3, 5, 7 and 9 of the growth trial. The representation of CFU/μL over time is represented in Fig. 4.3.

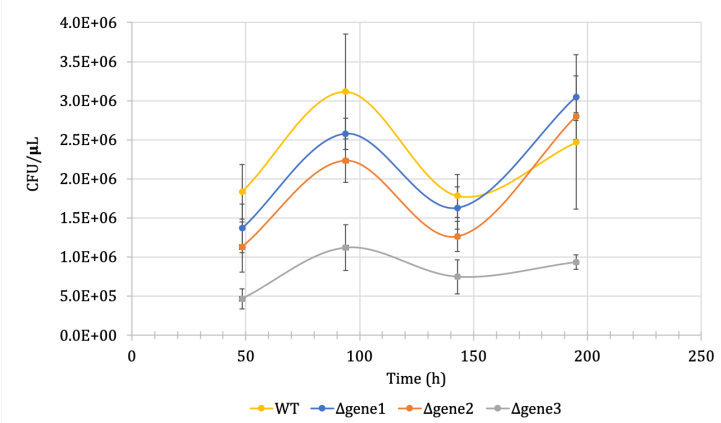

Figure 4.3: Variation of CFU/μL for the WT, Δgene1, Δgene2 and Δgene3 *S. bombicola* strains during the shake flask growth trial. The vertical error bars represent SD-values.

The values for day 7 are significantly bellow the expected for all strains, which suggest there might have been a problem during the dilution steps. In fact, it was expected that the CFU/μL value was approximately constant after day 3 of the growth trial, since cells already entered stationary phase. The decrease in CFU/μL on the day 9 for the WT can be due to counting errors, since the high number of colonies causes them to overlap with each other.

Once again, the growth of the Δgene3 strain seems to be slower comparing to the wild type. In this case, there is in fact a significant difference between the two since the p-value obtained was 0.012. The Δgene1 and Δgene2 strains also present lower values of CFU/μL compared to the WT (expect for the last day), but the difference is not as significant as for the Δgene3 strain.

4.4.1.D Glucose uptake

The variation of glucose concentration in the medium over time can be seen in Fig. 4.4.

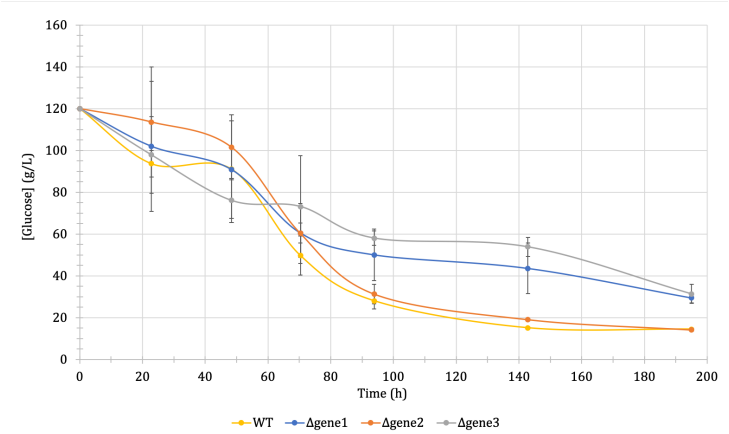

Figure 4.4: Variation of the glucose concentration in the medium for the WT, Δgene1, Δgene2 and Δgene3 *S. bombicola* strains during the shake flask growth trial. The vertical error bars represent SD-values.

The WT and Δgene2 strain show a similar trend in terms of glucose uptake, which appears to be slightly slower for the Δgene1 and Δgene3 strains. However, a t-test using the glucose values shows that there is no statistically significant difference for neither of them, since the p-values were of 0.569 and 0.488 in the tests using the glucose values of the Δgene1 and Δgene3 strains, respectively, together with the values of the WT.

In the end of the growth trial, the glucose was not completed depleted from the medium for any of the strains.

4.4.1.E Production of sophorolipids

The chromatograms resulting from the UPLC sophorolipid analysis on the fifth day of the growth trial are shown in Fig. 4.5.

Since the UPLC is performed using reverse phase chromatography, the more hydrophobic solutes will elute last. This means that the last peak (elution time at 7.30 min) corresponds to the specific substrate used, which is the most hydrophobic compound. For the Δgene1 and Δgene3 strains this is practically the only peak visible in the chromatogram, which means that the level of production is affected in these strains. On the other hand, both WT and Δgene2 strain present a high peak at 5.60 minutes, which corresponds to product B. The peaks between 3.70 and 4.40 minutes, which can also be

observed for both strains, correspond to product C with different degrees of acetylation. Finally, between minutes 2 and 3 there are some peaks less visible which might correspond to product F.

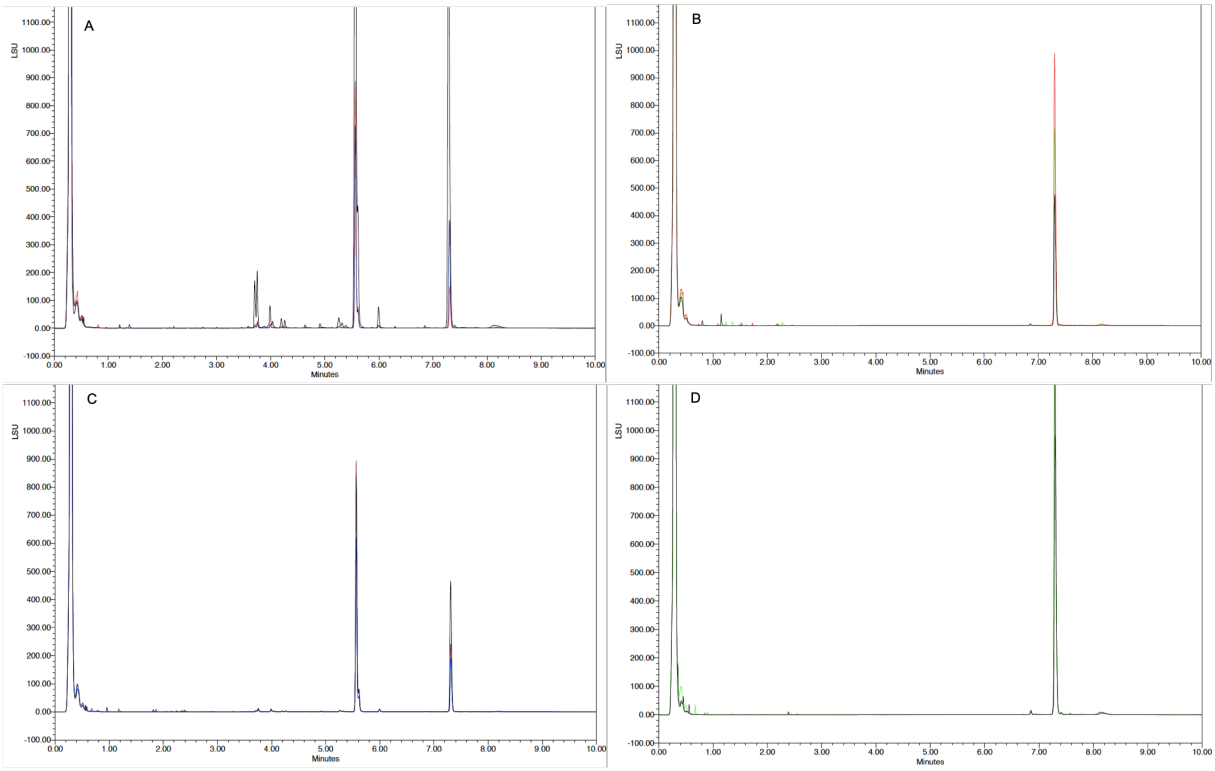

Figure 4.5: UPLC chromatograms of the WT (A), Δgene1 (B), Δgene2 (C) and Δgene3 SB strains from the fifth day of the growth trial.

In order to have more information about the different types of sophorolipids being produced by the strains, a LC-MS analysis was performed with the samples collected on the last day of the growth trial.

The strains in which the lowest number of peaks could be identified were the Δgene1 and Δgene3 strains. For both of them, the highest peak occurs at 7.30 minutes. This peak corresponds to the specific substrate used, considering its mass and the fact that it is the most hydrophobic compound and, consequently, the last to elute.

On the other hand, for both the WT and Δgene2 strain, the peak with the highest relative abundance occurs at 5.5 minutes and the mass spectrum indicates the presence of compounds with three different m/z values. The two m/z values with the highest relative abundance do not correspond to any compound already identified in the laboratory. A possible way to identify these compounds would be to perform LC-MS/MS, since this technique gives more structural information about the molecule, based on the fragmentation pattern. Regarding the third m/z value, this corresponds to product B, which also appears in the chromatograms of the Δgene1 and Δgene3 strains.

Besides product B, between 4.5 and 5 minutes, it is possible to detected for all strains peaks corresponding to product C. The Δgene1 strain is also producing product E, which was not detected for any other strain. For the WT and Δgene2 strain there is also the production of product D and product F.

Regarding the peaks that could not be identified, at 2.17 min there is a peak present in all chromatograms that might correspond to a component present in the medium, which is the same for all

strains. For the WT and Δgene2 strain, there is also a wide peak with three ends between minutes 5.20 and 5.55, which corresponds to several m/z values, none of which already identified. One possible explanation is that several compounds eluted at the same time, which led to the wrong mass being detected, which would also explained why this peak is so broad.

In conclusion, the production of sophorolipids appears to be affected by the knock-out in the genes 1 and 3 in terms of the variety of compounds being produced. For the Δgene2 strain, the analysis of the UPLC chromatogram might lead to the conclusion that the level of production was affected by this knock-out, since the peak corresponding to product B is practically the only one visible. Nevertheless, the LC-MS analysis allows to identify the same number of compounds produced by the WT and the Δgene2 strain. In Table 4.1 it can be seen the summary of the compounds identified.

Table 4.1: Summary of the compounds identified on the LC-MS chromatograms of the WT, Δgene1, Δgene2 and Δgene3 SB strains from the tenth day of the growth trial.

Compounds	Strains
Product F	WT, Δgene2
Product C	All
Product E	Δgene1
Product B	All
Product D	WT, Δgene2

It is interesting to notice that the WT, the Δgene2 and Δgene1 are the only ones producing compounds E and D. The absence of these products in the Δgene3 strain could mean that this oxidising enzyme is the one which has a preference for accepting a substrate with a different degree of saturation. However, the kinetic parameters might also be important to understand the fate of the substrate leading to the formation of products E and D, considering that *S. bombicola* also possesses an enzyme [77] that will be competing with the other oxidising enzymes for this same substrate. The fact that these products shift to the form of product D sooner in the WT and Δgene2 strain, comparing with Δgene1 strain, might mean that the saturated substrate is being used more quickly for these strains. This could indicate that both enzymes encoded by genes 1 and 3 are involved in the conversion of the substrate with a different degree of saturation. Nevertheless, is possible that the enzyme encoded by gene 1 is more selective towards the specific substrate used than the enzyme encoded by gene 3.

On the other hand, product F could only be identified in the WT and Δgene2 strain. The lack of this compounds in the Δgene1 and Δgene3 strains could be caused by changes in sophorolipid regulation, as a result of these gene deletions, that make the activity of the enzymes involved in the sophorolipid biosynthesis more specific.

Finally, it most be noted that no conclusion can be drawn in terms of quantities of sophorolipids being produced using this method. This is because the product sample extracted only corresponds to a part of the content of the shake flask, which is not homogeneous since lactonic sophorolipids have a tendency to precipitate. If quantification was the goal, it would be necessary to proceed to a product extraction at the end of the growth trial using the entire content of the shake flask, which was not the case. This way, even though the peaks in the Δgene3 strain are very plain compared to the other strains, there cannot be a direct conclusion that this strain is producing less quantity of sophorolipids. However, this can indeed be a possibility, since this KO seems to have affected the number of viable cells, which, consequently, would affect the quantity of biosurfactant being produced. On the other hand, it must also be considered that the lack of homogeneity of the medium, as well as batch to batch variations, might also affect the type of sophorolipids that can be detected for each strain.

4.4.2 Growth trial with Δgene1, Δgene2 and Δgene1Δgene2 SB strains

In the previous growth trial was demonstrated that there must be one or more oxidising enzymes capable of catalysing the hydroxylation of the substrate. To find these enzymes is essential to combine the KO in the gene 1 with a KO in another oxidising gene, which has been done to create the Δgene1Δgene2 SB strain (already available in the laboratory). In fact, even though the enzyme encoded by gene 2 does not seem do play a role in sophorolipid biosynthesis, based on the results of the previous growth trial, the effect of a double KO might lead to different conclusions. The activity of this enzyme might only become evident in the absence of the hydroxylation activity of the enzyme encoded by gene 1, since gene deletions lead to alterations in the transcriptome of cells. Furthermore, there is also the possibility of the enzymes forming complexes with each other, whose activity is only lost with the elimination of more than one enzyme composing this structure.

In this growth trial it will be determined if there is a difference on the type of sophorolipids produced by the Δgene1Δgene2 strain comparing with the respective single KO's.

4.4.2.A Optical Density

The variation of OD in the first four days of the growth trial is represented in Fig. 4.6. The strain Δgene1Δgene2 seems to present a slower increment in OD comparing with the other two strains, but this difference is not statistically significant - a p-value of 0.333 was obtained on a t-test using the values of the Δgene1 and Δgene1Δgene2 strains.

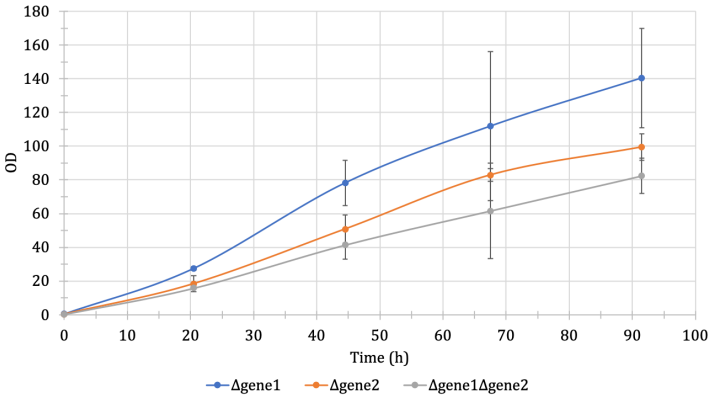

Figure 4.6: Variation of OD measured at 600 nm for the Δgene1, Δgene2 and Δgene1Δgene2 *S. bombicola* strains during the shake flask growth trial. The vertical error bars represent SD-values.

4.4.2.B pH values

All the strains show a similar trend in terms of pH drop, as can be observed in Fig. 4.7.

4.4.2.C Colony Forming Units

The values of CFU/μL for the days 3, 5, 7 and 9 of the growth trial are represented in Fig. 4.8.

For the Δgene2 and Δgene1Δgene2 strains, the values of CFU/μL are approximately constant over time, which corresponds to the expected, since cells have already entered stationary phase. For the Δgene1 strain there is an unexpected decrease in the CFU/μL value on the fifth day of the growth trial. This could be due to a dilution mistake when preparing the samples for plating. With the exception of this day, the values of CFU/μL for the Δgene1 strain are higher and there is a slight increase in the values over time. However, the values maintain the same order of magnitude over time and when compared

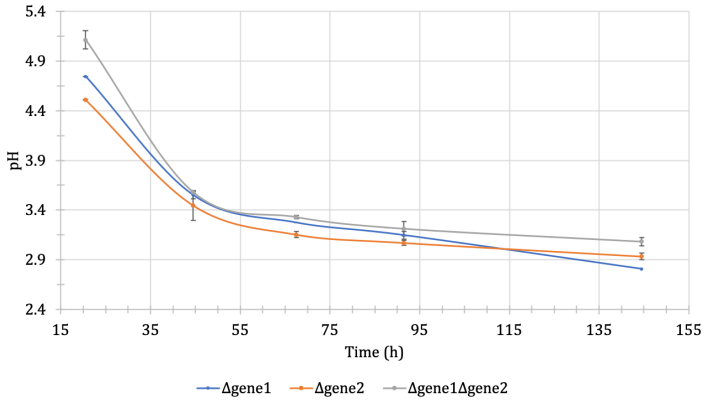

Figure 4.7: Variation of pH for the Δgene1, Δgene2 and Δgene1Δgene2 *S. bombicola* strains during the shake flask growth trial. The vertical error bars represent SD-values.

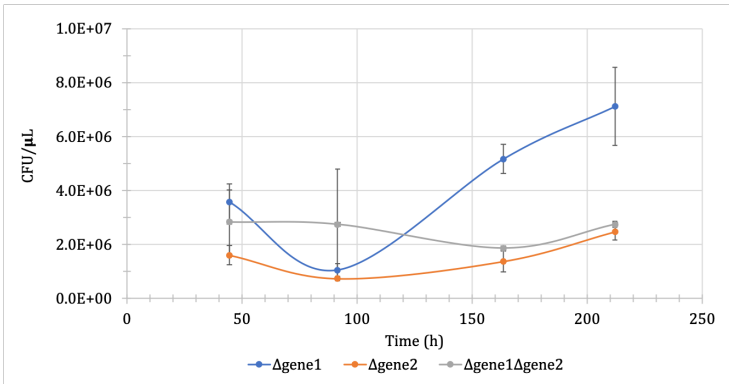

Figure 4.8: Variation of CFU/μL for the Δgene1, Δgene2 and Δgene1Δgene2 *S. bombicola* strains during the shake flask growth trial. The vertical error bars represent SD-values.

with the other two strains. The differences are, in fact, not statistically significant, as confirmed with the t-test performed - p-values of 0.138 and 0.288 for the Δgene2 and Δgene1Δgene2 strains, respectively, when compared with the Δgene1 values.

4.4.2.D Glucose uptake

The glucose uptake is approximately the same for the three strains, as it can be seen in the graphic of Fig. 4.9, which also shows that the glucose is not depleted from the medium for any of the strains.

4.4.2.E Production of sophorolipids

The chromatograms resulting from the UPLC sophorolipid analysis on the sixth day of the growth trial are shown in Fig. 4.10. Once again, for the strains which have the gene 1 knocked-out, it seems that there are no sophorolipids being produced, since the peak corresponding to the substrate is practically the only one visible. However, it can be observed that there are some solutes being eluted between the minutes 2 and 3 in all chromatograms. Furthermore, in the Δgene1 and Δgene1Δgene2 chromatograms, a peak around 4.80 minutes can also be observed, which might correspond to product C. For the Δgene2 strain the highest peak occurs around the 5.80 minutes and corresponds to a product B. Around minute 4, the peaks observed in the Δgene2 strain chromatogram correspond to product C.

It is interesting to notice that the peak corresponding to the substrate is almost imperceptible for

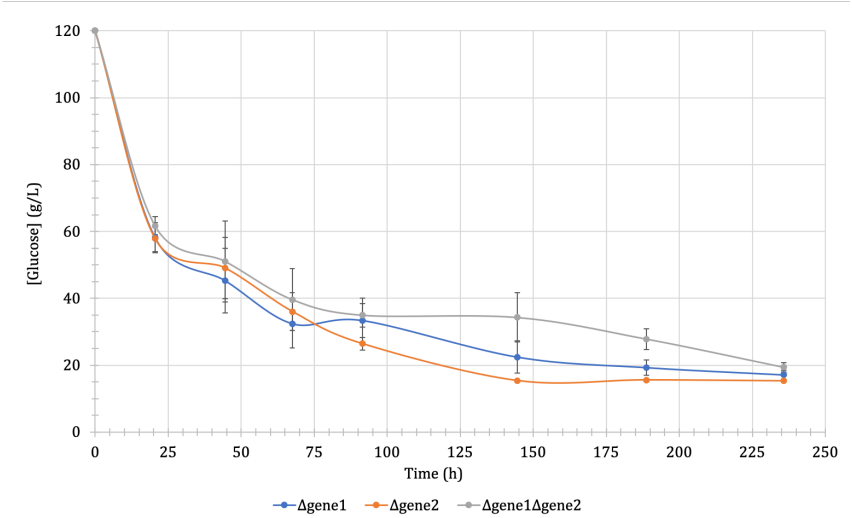

Figure 4.9: Variation of the glucose concentration in the medium for the Δgene1, Δgene2 and Δgene1Δgene2 *S. bombicola* strains during the shake flask growth trial. The vertical error bars represent SD-values.

the Δgene1Δgene2 strain, contrary to what is observed for the Δgene1 strain. Normally, the substrate peak is less visible for strains that are using it for production. However, for the Δgene1Δgene2 strain, a decrease on this peak is not accompanied by an increase in peaks corresponding to products. A possible explanation is that the Δgene1Δgene2 strain is directing this substrate more rapidly towards metabolism than the Δgene1 strain.

The UPLC is not sensitive enough to detect sophorolipids in strains which have a KO in gene 1, but they are in fact being produced by both strains, as confirmed by the TLC analysis (Fig. A.1). Therefore, a LC-MS analysis was performed with the samples collected on the last day of the growth trial. This analysis helps to understand if there is a significant difference between the sophorolipids produced by the Δgene1 and the Δgene1Δgene2 strains.

Analysing the chromatograms, the first thing that can be noticed is that the highest peak is different for each one of the strains. For the Δgene1Δgene2 strain, the highest peak occurs at 7.30 minutes and corresponds to the specific substrate used. For the Δgene1 strain, the highest peak corresponds to product B. The Δgene2 strain is also producing this compound, as well as the Δgene1Δgene2 strain, which in fact corresponds to the second highest peak in its chromatogram. Comparing the Δgene1 and the Δgene1Δgene2 strains there might be a tendency to conclude that this last one strain is producing less quantity sophorolipids than the Δgene1 strain. Although this might be true, it is not correct to infer about quantities using this method, as it was already explained. The only conclusion that can be drawn is that in fact the abundance of remaining substrate in comparison with the sophorolipid products is higher for the Δgene1Δgene2 strain.

For the Δgene2 strain, the highest peak corresponds to a m/z value already encountered for the WT and Δgene2 strain in the previous growth trial. Unfortunately, a correct identification of this compound was not possible. At the same elution time, the mass spectrum shows a compound with an m/z that corresponds to a product B. This compound is also present in the chromatograms of the Δgene1 and Δgene1Δgene2 strains. For the Δgene2 and Δgene1 strains, there is also the elution of products D and E, respectively. Finally, the Δgene2 strain is the only one producing product F.

Regarding the peaks non-identified, in the Δgene2 strain chromatogram it appears the same wide

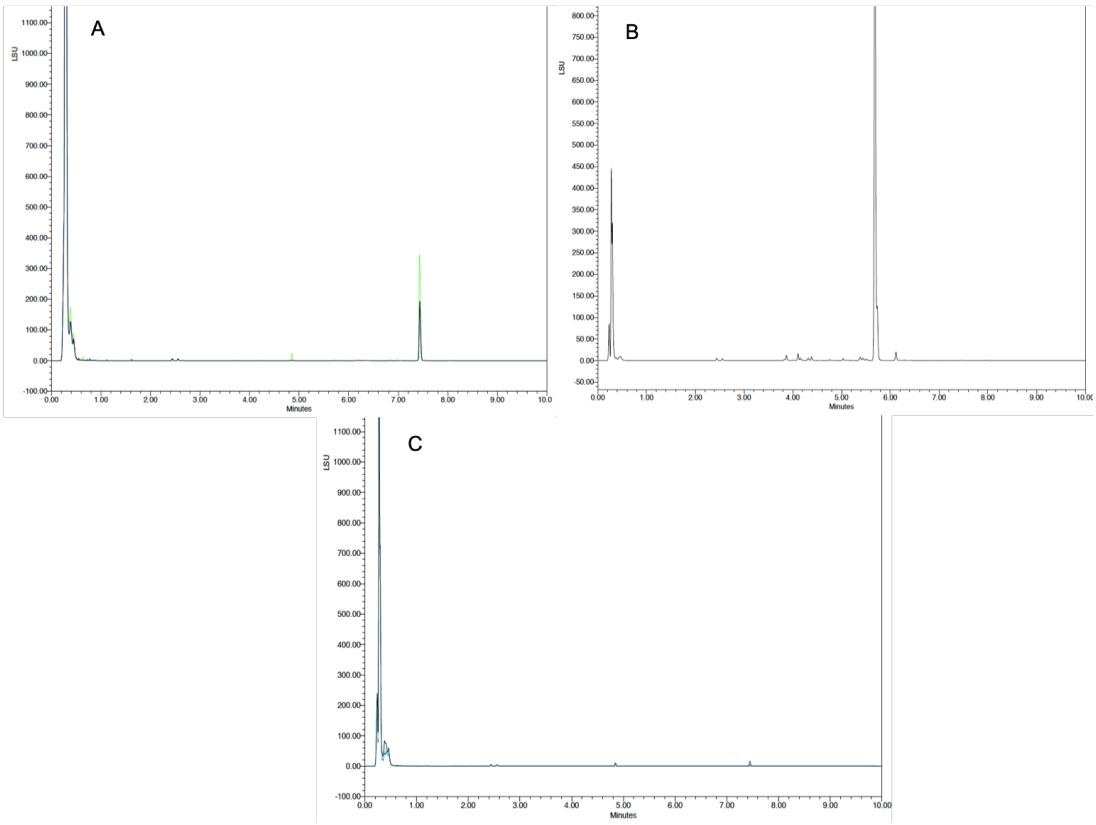

Figure 4.10: UPLC chromatograms of the Δgene1 (A), Δgene2 (B) and Δgene1Δgene2 (C) SB strains from the sixth day of the growth trial.

peak with three ends between 5.20 and 5.55 minutes that had already appeared for the same strain and for WT in the previous growth trial. As it was already discussed, this peak is probably the result of the co-elution of different compounds. In the Δgene1 and Δgene1Δgene2 chromatograms there is also the appearance of a peak, which could not be identified. Finally, in all chromatograms, the peaks that appear at 3.17 min and 6.70 min might correspond to components present in the media.

The summary of the compounds identified in this growth trial can be found in Table 4.2. Only the compounds which are being produced by all triplicates of each strain were considered in order to attenuate the effect of batch to batch variations.

Table 4.2: Summary of the compounds identified on the LC-MS chromatograms of the Δgene1, Δgene2 and Δgene1Δgene2 SB strains from the tenth day of the growth trial.

Compounds	Strains
Product F	Δgene2
Product C	All
Product E	Δgene1
Product B	All
Product D	WT, Δgene2

Similar to the previous growth trial, only the Δgene2 strain is producing product F, which, as it was explained, could be due to an unknown regulatory mechanism. Once again, there is also the production

of compounds in form E for the Δgene1 and form D for the Δgene2. In fact, the only compounds which could be identified for Δgene1Δgene2 strain were products C and D, so there might be a tendency towards a selective production of these compounds when the sophorolipid biosynthetic pathway is under stress due to the absence of two oxidising enzymes.

This experiment allowed to conclude that the Δgene1Δgene2 strain is producing a sophorolipid mixture with less variety than the Δgene1 strain. Therefore, gene 2 might play a role in the production, even though the previous growth trial did not show a difference between the WT and Δgene2 strain in terms of types of compounds identified. However, one fact that remains is that the production still occurs with both of these genes knocked-out, which means that one or more enzymes can secure the production. One hypothesis is the enzyme encoded by gene 3, due to the results obtained with the Δgene3 strain in the previous growth trial. However, it is not possible to conclude this with absolute certainty, since KO's have an effect on the transcriptome of cells. Therefore, to understand which enzymes are involved in sophorolipid production by the Δgene1Δgene2 strain, a RNA seq analysis should be performed. In case this analysis showed an upregulation of the gene 3 during stationary phase, then it would mean that this enzyme is probably involved in sophorolipid biosynthesis. Thereafter, it would also be interesting to test the effect of a double knock-out on genes 1 and 3 or even the triple KO of the 1, 2 and 3 genes.

4.4.3 Growth trial with Δgene1Δgene2, Δgene1Δgene6 and Δgene1Δgene2 Δgene6 SB strains

The Δgene1Δgene2Δgene6 SB strain was created in order to investigate if oxidising enzyme encoded by gene 2 is responsible for the hydroxylation of the specific substrate tested. The selective production of product A using this strain was intended to be used as a proof of concept of this hypothesis.

The results of OD, pH, CFU and glucose uptake are not shown since there is no significant difference between the strains. Therefore, only the results about the type of sophorolipids produced for each strain will be discussed.

4.4.3.A Production results

The chromatograms resulting from the UPLC sophorolipid analysis on the last day of the growth trial are shown in Fig. 4.11.

As it can be observed, the experiment did not produce the desired results since there is no significant difference between the chromatograms of the Δgene1Δgene6 and Δgene1Δgene2Δgene6 strains. In these two strains chromatograms, the peaks between 4 and 4.60 minutes corresponds to product H. The only peak identified as the desired product is the one that appears at minute 6.0, which corresponds to product A. On the other hand, the compounds eluted at 5.80 minutes and around 6.50 minutes could not be identified. To get more information about the compounds produced by the Δgene1Δgene6 and Δgene1Δgene2Δgene6 strains, a LC-MS analysis should be performed, but this was not possible during the period of experimental work of this master's thesis.

Regarding the Δgene1Δgene2 strain, since gene 6 is still present on this strain, the substrate will be oxidated to a compound that might correspond to the small peak at 7.50 min. No other product is visible in this chromatogram, which is not surprising considering the genes that are knocked-out - the same result had already been obtained with the Δgene1Δgene2 strain in the previous growth trial. Again, the absence of peaks in UPLC chromatogram does not mean that no products are being produced, since the UPLC is not sensitive enough to detect them when the gene 1 gene is knocked-out. Moreover, as had already been observed in the previous growth trial, the peak corresponding to the substrate is almost not visible, which can be explained by an accelerated fatty acid metabolism for this strain.

In conclusion, even though a LC-MS analysis would allow to get more information about the products formed by the strains, the UPLC chromatograms are sufficient to conclude that the enzyme responsible for hydroxylation of the specific substrate is not yet identified.

The next step should be performing a RNA seq analysis to analyse the differences in the transcription of the Δgene1Δgene6 and Δgene1Δgene2Δgene6 strains during growth on this specific substrate. Even though the enzyme encoded by gene 2 was identified as a good candidate in the past, KO's change the transcriptome of cells, so a new analysis might indicate which enzyme is securing the hydroxylation step on this new strain.

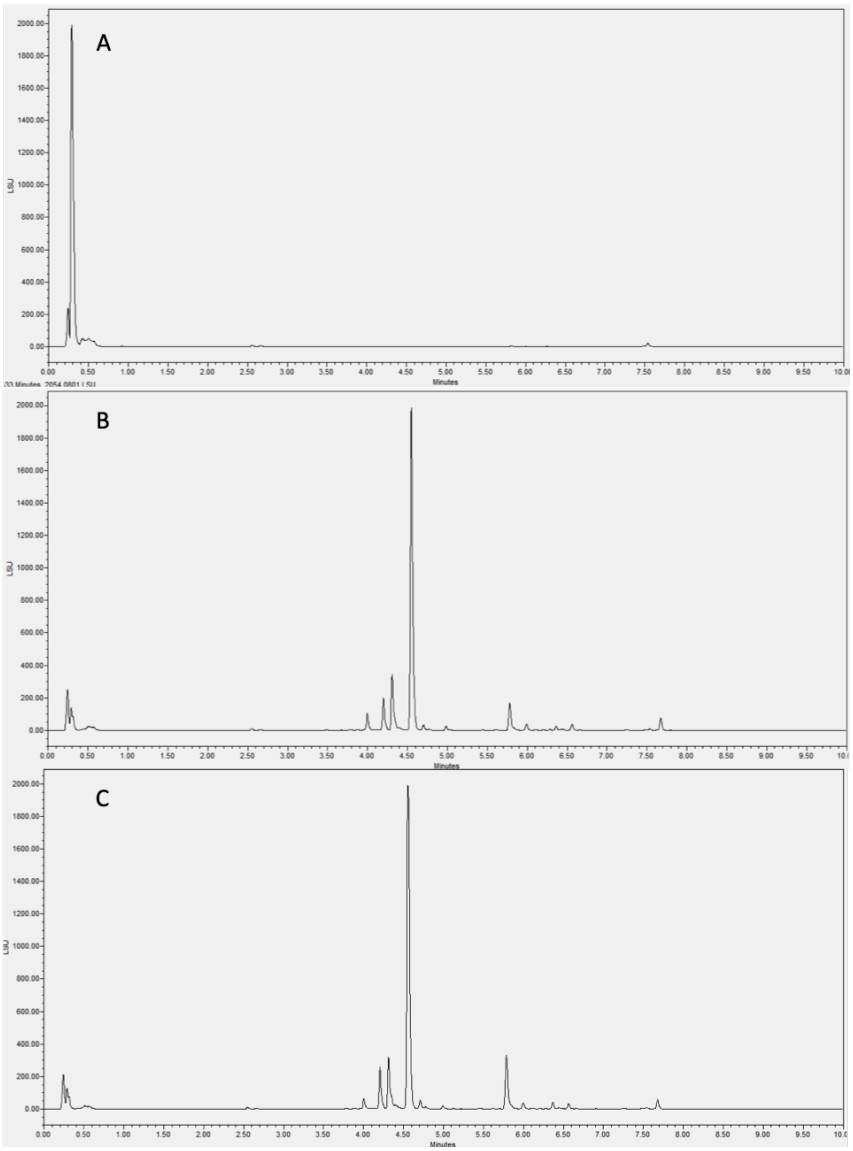

Figure 4.11: UPLC chromatograms of the Δgene1Δgene2 (A), Δgene1Δgene6 (B) and Δgene1Δgene2Δgene6 (C) SB strains from the tenth day of the growth trial.

5 Conclusion and Future Perspectives

The first part of this master's thesis consisted in performing molecular work in order to engineer *S. bombicola* strains. During the course of this work, it was possible to assemble two plasmids: one containing the KO cassette of the gene 4 and another one to delete the *ura3* gene. This last plasmid allowed the creation of a *ura3* negative Δgene1Δgene2 strain, which was then transformed with a gene 6 KO cassette to create the Δgene1Δgene2Δgene6 SB strain. Unfortunately, no more KO strains were created, since the transformations with the genes 1, 4 and 5 KO cassettes were not successful.

The major drawbacks during the attempts to engineer SB strains were the lack of efficiency of the hygromycin antibiotic and tendency to fail of the Yeast Colony PCR. The first one, did not allow the KO of the gene 1, which could have allowed the creation of three new strains. Regarding the Yeast Colony PCR, in several occasions no bands were detected, not even for the positive control, which complicated the task of finding positive transformants. Therefore, one possible improvement to facilitate the creation of new SB strains would be testing a different protocol to perform Yeast Colony PCR.

The second part of the experimental work was performing shake flask growth trials in order to test SB strains with different genes deleted. The single KO growth trial, in which the WT, Δgene1, Δgene2 and Δgene3 strains were tested, allowed to conclude that when the genes 1 and 3 are knocked-out, there is a lower variety in the mixture produced by *S. bombicola*, while no significant difference was encountered for the WT and Δgene2 strain. However, when the KO on the gene 2 was combined with a deletion in gene 1, there was a lower number of compounds being produced in comparison with the Δgene1 strain. The compounds which were common to all strains in these growth trials were products B and C, which were the only compounds detected for the Δgene3 and Δgene1Δgene2 strains. On the other hand, product F was only detected for the WT and Δgene2 strain. These strains, together with the Δgene1 strain, are also capable of producing products D and E. Both enzymes encoded by genes 1 and 3 seem to be involved in the production of these compounds, but the enzyme encoded by gene 1 is thought to only accept substrate with a different saturation degree when the majority of the specific substrate has been used. On the other hand, the compound corresponding to the highest peak in the LC-MS analysis of the WT and Δgene2 strain could not be identified. Hence, future efforts to identify this compound can consist in preforming a LC-MS/MS analysis.

Finally, the main objective of this master's thesis was to isolate the production of product A. Starting from the RNA sequencing results which indicate that enzyme encoded by gene 2 might be responsible for the hydroxylation of the specific substrate used to produce product A, the Δgene1Δgene2Δgene6 SB strain was used in a proof of concept to test this theory. However, the results obtained were not the expected ones, since, similarly to the Δgene1Δgene6, this new strain is still producing product H. Consequently, future efforts have to be made in order to construct a new strain which combines deletions in the genes 1 and 6 with deletions in other oxidising genes. Another candidate is the gene 5, but the attempts to knock-out this gene were not yet successful. Furthermore, a new candidate can emerge by performing a transcriptomic analysis on the Δgene1Δgene2Δgene6 strain.

The final attempt to discover the activity of each oxidising enzyme towards a series of different substrates would be performing *in vitro* enzymatic tests with engineered catalytic self-sufficient oxidising

enzymes. Unfortunately, it was not possible to perform these assays due to the impossibility in obtaining the back-bone of the plasmid which would be used for heterologous expression in *S. cerevisiae*. Therefore, one solution to move forward in this experimental procedure could be ordering the plasmid itself, since the plasmid supposedly already available in the laboratory, did not seem to be reliable.

In conclusion, this experimental work allowed for a preliminary investigation of the compounds produced using different engineered SB strains. Future efforts to understand the versatile oxidative power of *S. bombicola* include engineering more strains. Furthermore, it could also be included transcriptomic analysis of relevant strains, quantification information about the formed products and the analysis of the effect of fermentation conditions on the type of sophorolipids being produced by the engineered strains.

Bibliography

[1] L. N. Britton, "Surfactants and the environment," *Journal of Surfactants and Detergents*, vol. 1, no. 1, pp. 109–117, 1998.

[2] F. Md, "Biosurfactant: Production and application," *Journal of Petroleum & Environmental Biotechnology*, vol. 3, no. 4, 2012.

[3] I. M. Banat, A. Franzetti, I. Gandolfi, G. Bestetti, M. G. Martinotti, L. Fracchia, T. J. Smyth, and R. Marchant, "Microbial biosurfactants production, applications and future potential," *Applied Microbiology Biotechnology*, vol. 87, pp. 477–444, 2010.

[4] I. N. A. V. Bogaert, K. Saerens, C. D. Muynck, D. Develter, W. Soetaert, and E. J. Vandamme, "Microbial production and application of sophorolipids," *Applied Microbiology Biotechnology*, vol. 76, pp. 23–34, 2007.

[5] M. D. Graeve, S. D. Maeseneire, S. Roelants, and W. Soetaert, "*Starmerella bombicola*, an industrially relevant, yet fundamentally underexplored yeast," *FEMS Yeast Research*, vol. 18, no. 7, 2018.

[6] S. L. Roelants, K. M. Saerens, T. Derycke, Y.-C. L. Bing Li, Y. V. de Peer, S. L. D. Maeseneire, I. N. V. Bogaert, and W. Soetaert, "*Candida bombicola* as a platform organism for the production of tailor-made biomolecules," *Biotechnology and Bioengineering*, vol. 110, p. 9, 2013.

[7] S. Lodens, M. D. Graeve, S. L. K. W. Roelants, S. L. D. Maeseneire, , and W. Soetaert, "Transformation of an exotic yeast species into a platform organism: A case study for engineering glycolipid production in the yeast *Starmerella bombicola*," *Synthetic Biology: Methods and Protocols. Methods in Molecular Biology*, vol. 1772, pp. 95–123, 2018.

[8] M. J. Rosen and J. T. Kunjappu, *Surfactants and Interfacial Phenomena*. Wiley, 2012.

[9] OpenStaxCollege. Cohesion and adhesion in liquids: Surface tension and capillary action. Accessed 18/06/2019. [Online]. Available: https://opentextbc.ca/physicstestbook2/chapter/cohesion-and-adhesion-in-liquids-surface-tension-and-capillary-action/

[10] K. Sakamoto, R. Y. Lochhead, H. I. Maibach, and Y. Yamashita, *Cosmetic Science and Technology: Theoretical Principles and Applications*. Elsevier, 2017.

[11] S. Laurén. What are surfactants and how do they work? Accessed 21/02/2019. [Online]. Available: https://blog.biolinscientific.com/what-are-surfactants-and-how-do-they-work

[12] D. I. GmbH. Surfactants & critical micelle concentration (CMC). Accessed 21/02/2019. [Online]. Available: https://www.dataphysics-instruments.com/knowledge/understanding-interfaces/surfactants-cmc/

[13] C. O. Rangel-Yagui, A. Pessoa-Jr, and L. C. Tavares, "Micellar solubilization of drugs," *Journal of Pharmacy and Pharmaceutical Sciences*, vol. 8, pp. 147–65, 2005.

[14] M. Hirsch. Surface active agents (surfactants). Accessed 21/02/2019. [Online]. Available: https://knowledge.ulprospector.com/3106/pc-surface-active-agents-surfactants/

[15] T. Ivanković and J. Hrenović, "Surfactants in the environment," *Arhiv Za Higijenu Rada i Toksikologiju*, vol. 61, pp. 95–110, 2010.

[16] M. J. Scott and M. N. Jones, "The biodegradation of surfactants in the environment," *Biochimica et Biophysica Acta*, vol. 1, pp. 235–51, 2000.

[17] G.-G. Ying, "Fate, behavior and effects of surfactants and their degradation products in the environment," *Environment International*, vol. 32, p. 417 – 431, 2006.

[18] R. Renner, "European bans on surfactant trigger transatlantic debate," *Environmental Science and Technology*, vol. 31, pp. 316A–20A, 1997.

[19] S. Rebello, A. K. Asok, S. Mundayoor, and M. S. Jisha, "Surfactants: toxicity, remediation and green surfactants," *Environmental Chemistry Letters*, vol. 12, no. 2, p. 275–287, 2014.

[20] T. M. S. Lima, L. C. Procópio, F. D. Brandão, A. M. X. Carvalho, M. R. Tótola, and A. C. Borges, "Biodegradability of bacterial surfactants," *Biodegradation*, vol. 22, no. 3, p. 585–592, 2011.

[21] P. K. Mohan, G. Nakhla, and E. K. Yanful, "Biokinetics of biodegradation of surfactants under aerobic, anoxic and anaerobic conditions," *Water Research*, vol. 40, no. 3, p. 533–540, 2006.

[22] S. Satpute, G. Płaza, and A. Banpurkar, "Biosurfactants' production from renewable natural resources: Example of innovativeand smart technology in circular bioeconomy," *Management Systems in Production Engineering*, vol. 25, no. 1, pp. 46–54, 2017.

[23] V. Pruthi and S. S. Cameotra, "Production and properties of a biosurfactant synthesized by *Arthrobacter protophormiae* — an antarctic strain," *World Journal of Microbiology and Biotechnology*, vol. 13, p. 137–139, 1997.

[24] S. Vijayakumar and V. Saravanan, "Biosurfactants - types, sources and applications," *Research Journal of Microbiology*, vol. 10, no. 5, pp. 181–192, 2015.

[25] J. D. Desai and I. M. Banat, "Microbial production of surfactants and their commercial potential," *Microbiology and Molecular Biology Reviews*, vol. 61, pp. 47–642, 1997.

[26] N. Kosaric, "Biosurfactants in industry," *Pure and Applied Chemistry*, vol. 64, no. 11, pp. 1731–1737, 1992.

[27] E. Z. Ron and E. Rosenberg, "Natural roles of biosurfactants," *Environmental Microbiology*, vol. 3, no. 4, pp. 229–236, 2001.

[28] E. Deziel, G. Paquette, R. Villemur, F. Lepine, and J. Bisaillon, "Biosurfactant production by a soil *Pseudomonas* strain growing on polycyclic aromatic hydrocarbons," *Applied and Environmental Microbiology*, vol. 62, no. 6, p. 1908–1912, 1996.

[29] F. PeypouxJ., M. Bonmatin, and J. Wallach, "Recent trends in the biochemistry of surfactin," *Applied Microbiology and Biotechnology*, vol. 51, no. 5, p. 553–563, 1999.

[30] M. Richter, J. M. Willey, R. Süßmuth, G. Jung, and H.-P. Fiedler, "Streptofactin, a novel biosurfactant with aerial mycelium inducing activity from *Streptomyces tendae* Tü 901/8c," *FEMS Microbiology Letters*, vol. 163, no. 2, p. 165–171, 1998.

[31] Y. Zhang and R. M. Miller, "Effect of a *Pseudomonas* rhamnolipid biosurfactant on cell hydrophobicity and biodegradation of octadecane," *Applied and Environmental Microbiology*, vol. 60, no. 2, pp. 2101–2106, 1994.

[32] C. Olvera, J. B. Goldberg, R. Sánchez, and G. Soberón-Chávez, "The *Pseudomonas aeruginosa algC* gene product participates in rhamnolipid biosynthesis," *FEMS Microbiology Letters*, vol. 179, no. 1, pp. 85–90, 1999.

[33] D. C. Herman, J. F. Artiola, and R. M. Miller, "Removal of cadmium, lead, and zinc from soil by a rhamnolipid biosurfactants," *Environmental Science & Technology*, vol. 29, no. 9, pp. 2280–2285, 1995.

[34] F. G. Jarvis and M. J. Johnson, "A glycolipide produced by *Pseudomonas aeruginosa*," *Journal of the American Chemical Society*, vol. 71, no. 12, pp. 4124–4126, 1949.

[35] R. Anderson and M. Newman, "The chemistry of the lipids of *Tubercle bacilli*: XXXIII. isolation of trehalose from the acetone-soluble fat of the human *Tubercle bacillus*," *Journal of Biological Chemistry*, vol. 101, p. 499–504, 1933.

[36] Y. F. Ikuya Yano and M. Kusunose, "Occurrence of acylated trehaloses in *Nocardia*," *The Journal of General and Applied Microbiology*, vol. 17, pp. 329–334, 1971.

[37] T. Suzuki, K. Tanaka, I. Matsubara, and S. Kinoshita, "Trehalose lipid and α-branched-β-hydroxy fatty acid formed by bacteria grown on n-alkanes," *Agricultural and Biological Chemistry*, vol. 33, no. 11, pp. 1619–1627, 1969.

[38] A. Kretschmer, H. Bock, and F. Wagner, "Chemical and physical characterization of interfacial-active lipids from *Rhodococcus erythropolis* grown on n-alkanes," *Applied and Environmental Microbiology*, vol. 44, no. 4, pp. 864–870, 1982.

[39] J. F. T. Spencer, P. A. J. Gorin, and A. P. Tulloch, "*Torulopsis bombicola* sp. n." *Antonie van Leeuwenhoek - Journal of Microbiology*, vol. 36, no. 1, p. 129–133, 1970.

[40] M. A. Marahiel, W. Danders, M. Krause, and H. Kleinkauf, "Biological role of gramicidin S in spore functions. studies on gramicidin-S-negative mutants of *Bacillus brevis* ATCC 9999," *European Journal of Biochemistry*, vol. 99, no. 1, pp. 49–55, 1979.

[41] K. F. K. T. Tomoji Suzuki, Kyozo Hayashi, "The chemical structure of polymyxin E: The identities of polymyxin E_1 with colistin A and of polymyxin E_2 with colistin B," *The Journal of Biochemistry*, vol. 57, no. 2, p. 226–227, 1965.

[42] K. Arima, A. Kakinuma, and G. Tamura, "Surfactin, a crystalline peptidelipid surfactant produced by *Bacillus subtilis*: Isolation, characterization and its inhibition of fibrin clot formation," *Biochemical and Biophysical Research Communications*, vol. 31, no. 3, pp. 488–494, 1968.

[43] M. M. Yakimov, K. N. Timmis, V. Wray, and H. L. Fredrickson, "Characterization of a new lipopeptide surfactant produced by thermotolerant and halotolerant subsurface *Bacillus licheniformis* BAS50," *Applied and Environmental Microbiology*, vol. 61, no. 5, p. 1706–1713, 1995.

[44] W. Mongkolthanaruk, "Classification of *Bacillus* beneficial substances related to plants, humans and animals," *Journal of Microbiology and Biotechnology*, vol. 22, no. 12, pp. 1597–1604, 2012.

[45] O. Kappeli and W. R. Finnerty, "Partition of alkane by an extracellular vesicle derived from hexadecane-grown *Acinetobacter*," *Journal of Bacteriology*, vol. 140, no. 2, pp. 707–712, 1979.

[46] Z. Zosim, D. Gutnick, and E. Rosenberg, "Properties of hydrocarbon-in-water emulsions stabilized by *Acinetobacter* RAG-1 emulsan," *Biotechnology and Bioengineering*, vol. 24, no. 2, pp. 281–292, 1982.

[47] M. C. Cirigliano and G. M. Carman, "Isolation of a bioemulsifier from *Candida lipolytica*," *Applied and Environmental Microbiology*, vol. 48, no. 4, pp. 747–750, 1984.

[48] D. R. Cameron, D. G. Cooper, and R. J. Neufeld, "The mannoprotein of *Saccharomyces cerevisiae* is an effective bioemulsifier," *Applied and Environmental Microbiology*, vol. 54, no. 6, pp. 1420–1425, 1988.

[49] M. Gilewicz, T. Ni'matuzahroh, H. Nadalig, H. Budzinski, P. Doumenq, V. Michotey, and J. Bertrand, "Isolation and characterization of a marine bacterium capable of utilizing 2-methylphenanthrene," *Applied Microbiology and Biotechnology*, vol. 48, pp. 528– 533, 1997.

[50] M. Nitschke and S. Costa, "Biosurfactants in food industry," vol. 18, no. 5, pp. 252–259, 2007.

[51] T. Furuta, K. Igarashi, and Y. Hirata, "Low-foaming detergent compositions," Patent US20 040 171 512A1, September, 2004.

[52] M. Nero, "Methods for cleaning materials," Patent US7 556 654B1, July, 2009.

[53] D. Develter, I. Jacobs, and M. Renkin, "Detergent compositions," Patent ES2 266 767T3, March, 2007.

[54] M. Pacwa-Płociniczak, G. A. Płaza, Z. Piotrowska-Seget, and S. S. Cameotra, "Environmental applications of biosurfactants: Recent advances," *International Journal of Molecular Sciences*, vol. 54, pp. 449–444, 2007.

[55] L. Fracchia, C. Ceresa, A. Franzetti, M. Cavallo, I. Gandolfi, J. V. Hamme, P. Gkorezis, R. Marchant, and I. M. Banat, "Industrial applications of biosurfactants," *Biosurfactants: Production and Utilization—Processes, Technologies, and Economics*, p. 245–260, 2014.

[56] F. B. Oleson and F. P. Tally, "Use of daptomycin," Patent EP1 115 417B1, April, 2006.

[57] H. Ceri, R. J. Turner, M. G. M. F. Rivardo, and G. Allegrone, "Biosurfactant composition produced by a new *Bacillus licheniforms* strain, uses and products thereof," Patent US8 377 455B2, November, 2011.

[58] G. Reid, A. W. Bruce, H. J. Busscher, and H. C. V. der Mei, "*Lactobacillus* therapies," Patent US6 051 552A, April, 2000.

[59] I. N. A. V. Bogaert, K. Holvoet, S. L. K. W. Roelants, B. Li, Y.-C. Lin, Y. V. de Peer, and W. Soetaert, "The biosynthetic gene cluster for sophorolipids: a biotechnological interesting biosurfactant produced by *Starmerella bombicola*," *Molecular Microbiology*, vol. 88, no. 3, p. 501–509, 2013.

[60] R. D. Ashby, D. K. Y. Solaiman, and T. A. Foglia, "Property control of sophorolipids: influence of fatty acid substrate and blending," *Biotechnology letters*, vol. 30, no. 6, p. 3449–3461, 2008.

[61] V. Shah, G. F. Doncel, T. Seyoum, K. M. Eaton, I. Zalenskaya, R. Hagver, A. Azim, and R. Gross, "Sophorolipids, microbial glycolipids with anti-human immunodeficiency virus and sperm-immobilizing activities," *Antimicrobial Agents and Chemotherapy*, vol. 49, no. 10, p. 4093–4100, 2005.

[62] S. L. K. Roelants, S. L. D. Maeseneire, K. Ciesielska, I. N. A. V. Bogaert, and W. Soetaert, "Biosurfactant gene clusters in eukaryotes: regulation and biotechnological potential," *Applied Microbiology Biotechnology*, vol. 98, p. 3449–3461, 2014.

[63] S. Inoue and N. Miyamoto, "Process for producing a hydroxyfatty acid ester," Patent US4 201 844A, May, 1980.

[64] J. A. Zerkowski and D. K. Y. Solaiman, "Synthesis of polyfunctional fatty amines from sophorolipid-derived 17-hydroxy oleic acid," *Journal of the American Oil Chemists' Society*, vol. 83, no. 7, p. 621–628, 2006.

[65] D. Pierce and T. J. Heilman, "Germicidal composition," Patent US6 262 038B1, July, 2001.

[66] J. Chen, X. Song, H. Zhang, and Y. Qu, "Production, structure elucidation and anticancer properties of sophorolipid from *Wickerhamiella domercqiae*," *Enzyme and Microbial Technology*, vol. 39, no. 3, pp. 501–506, 2006.

[67] K. Ciesielska, I. Bogaert, S. Chevineau, B. Li, S. Groeneboer, W. Soetaert, Y. V. de Peer, and B.Devreese, "Exoproteome analysis of *Starmerella bombicola* results in the discovery of an esterase required for lactonization of sophorolipids," *Journal of Proteomics*, vol. 98, pp. 159–174, 2014.

[68] R. Homme, L. Weber, A. Weiss, U. Himmelreich, O. Rilke, and H.-P.Kleber, "Production of sophorose lipid by *Candida (Torulopsis) apicola* grown on glucose," *Journal of Biotechnology*, vol. 33, no. 2, pp. 147–155, 1994.

[69] I. N. V. Bogaert, J. Zhang, and W. Soetaert, "Microbial synthesis of sophorolipids," *Process Biochemistry*, vol. 46, p. 821–833, 2011.

[70] A.-M. Davila, R. Marchal, and J.-P. Vandecasteele, "Sophorose lipid production from lipidic precursors: Predictive evaluation of industrial substrates," *Journal of Industrial Microbiology*, vol. 13, no. 4, p. 249–257, 1994.

[71] Y.-B. Kim, H. S. Yun, and E.-K. Kim, "Enhanced sophorolipid production by feeding-rate-controlled fed-batch culture," *Bioresource Technology*, vol. 100, no. 3, pp. 6028–6032, 2009.

[72] A. Davila, R. Marchal, N. Monin, and J. Vandecasteele, "Identification and determination of individual sophorolipids in fermentation products by gradient elution high-performance liquid chromatography with evaporative light-scattering detection," *Journal of Chromatography A*, vol. 648, no. 1, pp. 139–149, 1993.

[73] I. N. A. V. Bogaert, D. Buyst, J. C. Martins, S. L. K. W. Roelants, and W. K. Soetaer, "Synthesis of bolaform biosurfactants by an engineered *Starmerella bombicola* yeast," *Biotechnology and Bioengineering*, vol. 113, pp. 2644–2651, 2016.

[74] S. L. Roelants, K. Ciesielska, S. L. D. Maeseneire, H. Moens, B. Everaert, S. Verweire, Q. Denon, B. Vanlerberghe, I. N. V. Bogaert, P. V. der Meeren, B. Devreese, and W. Soetaert, "Towards the industrialization of new biosurfactants: Biotechnological opportunities for the lactone esterase gene from *Starmerella bombicola*," *Biotechnology and Bioengineering*, vol. 113, no. 3, pp. 550–559, 2016.

[75] A. Davila, R. Marchal, and J. Vandecasteele, "Kinetics and balance of a fermentation free from product inhibition: sophorose lipid production by *Candida bombicola*," *Applied Microbiology and Biotechnology*, vol. 38, no. 1, pp. 6–11, 1992.

[76] ——, "Sophorose lipid formation by resting cells of *Torulopsis bombicola*," *Biotechnology Letters*, vol. 6, no. 4, p. 225–230, 1984.

[77] V. Guilmanov, A. Ballistreri, G. Impallomeni, and R. A. Gross, "Oxygen transfer rate and sophorose lipid production by *Candida bombicola*," *Biotechnology Letters*, vol. 77, no. 5, pp. 489–494, 2002.

[78] D. Yarrow and S. A. Meyer, "Proposal for amendment of the diagnosis of the genus *Candida* berkhout nom. cons." *International Journal of Systematic and Evolutionary Microbiology*, vol. 28, no. 4, pp. 611–615, 1978.

[79] C. A. Rosa and M.-A. Lachance, "The yeast genus *Starmerella* gen. nov. and *Starmerella bombicola* sp. nov., the teleomorph of *Candida bombicola* (Spencer, Gorin & Tullock) Meyer & Yarrow," *International Journal of Systematic Bacteriology*, vol. 48, pp. 1413–1417, 1998.

[80] S. Ito, M. Kinta, and S. Inoue, "Growth of yeasts on n-alkanes: Inhibition by a lactonic sophorolipid produced by *Torulopsis bombicola*," *Agricultural and Biological Chemistry*, vol. 44, no. 9, pp. 2221–2223, 1980.

[81] I. N. A. V. Bogaert, S. Roelants, D. Develter, and W. Soetaert, "Sophorolipid production by *Candida bombicola* on oils with a special fatty acid composition and their consequences on cell viability," *Biotechnology Letters*, vol. 32, no. 10, p. 1509–1514, 2010.

[82] A. Alcon, V. E. Santos, J. A. Casas, and F. García-Ochoa, "Use of flow cytometry for growth structured kinetic model development application to *Candida bombicola* growth," *Enzyme and Microbial Technology*, vol. 34, no. 5, pp. 399–406, 2004.

[83] A. Albrecht, U. Rau, and F. Wagnert, "Initial steps of sophoroselipid biosynthesis by *Candida bombicola* atcc 22214 grown on glucose," *Applied Microbiology and Biotechnology*, vol. 46, no. 1, p. 67–73, 1996.

[84] K. Ciesielska, B. Li, S. Groeneboer, I. V. Bogaert, Y.-C. Lin, W. Soetaert, Y. V. de Peer, and B. Devreese, "Silac-based proteome analysis of starmerella bombicola sophorolipid production," *Journal of Proteome Research*, vol. 12, no. 10, pp. 4376–4392, 2013.

[85] F. Takahashi, K. Igarashi, and H. Hagihara, "Identification of the fatty alcohol oxidase FAO1 from *Starmerella bombicola* and improved novel glycolipids production in an FAO1 knockout mutant," *Applied Microbiology Biotechnology*, vol. 100, p. 9519–9528, 2016.

[86] K. M. Saerens, S. L. Roelants, I. N. V. Bogaert, and W. Soetaert, "Identification of the UDP-glucosyltransferase gene UGTA1, responsible for the first glucosylation step in the sophorolipid biosynthetic pathway of *Candida bombicola* ATCC 22214," *FEMS Yeast Research*, vol. 11, p. 123–132, 2011.

[87] K. M. J. Saerens, J. Zhang, L. Saey, I. N. A. V. Bogaert, and W. Soetaert, "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*, vol. 28, p. 279–292, 2011.

[88] K. M. Saerens, L. Saey, and W. Soetaert, "One-step production of unacetylated sophorolipids by an acetyltransferase negative *Candida bombicola*," *Biotechnology and Bioengineering*, vol. 108, pp. 2923–2931, 2011.

[89] A. Daverey, K. Pakshirajan, and S. Sumalatha, "Sophorolipids production by *Candida bombicola* using dairy industry wastewater," *Clean Technologies and Environmental Policy*, vol. 13, no. 3, p. 481–488, 2011.

[90] K. Saerens, I. van Bogaert, W. Soetaert, and E. Vandamme, "Production of glucolipids and specialty fatty acids from sophorolipids by penicillium decumbens naringinase : optimization and kinetics," *Biotechnology Journal*, vol. 4, no. 4, pp. 517–524, 2009.

[91] D. Balzer, "Cloud point phenomena in the phase behavior of alkyl polyglucosides in water," *Langmuir*, vol. 9, pp. 3375–3384, 1993.

[92] D. Balzer and H. Luders, *Nonionic Surfactants: Alkyl Polyglucosides*. Marcel Dekker, Inc., 2000.

[93] K. M. Saerens, I. N. V. Bogaert, and W. Soetaert, "Characterization of sophorolipid biosynthetic enzymes from *Starmerella bombicola*," *FEMS Yeast Research*, vol. 15, no. 7, pp. 1–9, 2015.

[94] D. Werck-Reichhart and R. Feyereisen, "Cytochromes P450: a success story," *Genome Biology*, vol. 1, 2000.

[95] F.-C. Huang, A. Peter, and W. Schwab, "Expression and characterization of CYP52 genes involved in the biosynthesis of sophorolipid and alkane metabolism from *Starmerella bombicola*," *Applied and Environmental Microbiology*, vol. 80, p. 766–776, 2014.

[96] M. Klingenberg, "Pigments of rat liver microsomes," *Archives of Biochemistry and Biophysics*, vol. 75, no. 2, pp. 376–386, 1958.

[97] O. Kappeli, "Cytochromes P450 of yeasts," *Microbiology Reviews*, vol. 50, p. 244–258, 1986.

[98] F. Hannemann, A. Bichet, K. M. Ewen, and R. Bernhardt, "Cytochrome P450 systems—biological variations of electron transport chains," *Biochimica et Biophysica Acta*, vol. 1770, pp. 330–344, 2007.

[99] D. Cook, J. Finnigan, K. Cook, G. Black, and S. Charnock, "Cytochromes P450: History, classes, catalytic mechanism, and industrial application," *Advances in Protein Chemistry and Structural Biology*, vol. 105, pp. 105–126, 2016.

[100] S. L. U. of Illinois at Urbana-Champaign. Molecular mechanisms of cytochrome P450 catalysis. Accessed 6/04/2019. [Online]. Available: http://sligarlab.life.uiuc.edu/research/research-p450.html

[101] H. M. Girvan and A. W. Munro, "Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology," *Current Opinion in Chemical Biology*, vol. 31, p. 136–145, 2016.

[102] I. N. V. Bogaert, M. Demey, D. Develter, W. Soetaert, and E. J. Vandamme, "Importance of the cytochrome P450 monooxygenase CYP52 family for the sophorolipid-producing yeast *Candida bombicola*," *FEMS Yeast Research*, vol. 9, pp. 87–94, 2009.

[103] L. D. Eirich, D. L. Craft, L. Steinberg, A. Asif, W. H. Eschenfeldt, L. Stols, M. I. Donnelly, and C. R. Wilson, "Cloning and characterization of three fatty alcohol oxidase genes from *Candida tropicalis* strain atcc 20336," *Applied and Environmental Microbiology*, vol. 70, no. 8, pp. 4872–4879, 2004.

[104] T. Sakaki, "Practical application of cytochrome P450," *Biol. Pharm. Bull.*, vol. 35, p. 844–849, 2012.

[105] C. Paddon, P. Westfall, D. Pitera, K. Benjamin, and K. F. et al., "High-level semi-synthetic production of the potent antimalarial artemisinin," *Nature*, vol. 496, no. 7446, pp. 528–532, 2013.

[106] K. J. McLean, M. Hans, B. Meijrink, W. B. van Scheppingen, A. Vollebregt, K. L. Tee, J.-M. van der Laan, D. Leys, A. W. Munro, and M. A. van den Berg, "Single-step fermentative production of the cholesterol-lowering drug pravastatin via reprogramming of penicillium chrysogenum," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 9, pp. 2847–2852, 2015.

[107] S. Lang, A. Brakameier, R. Heckmann, S. Spockner, and U. Rau, "Production of native and modified sophorose lipids," *Chimica Oggi – Chemistry Today*, vol. 18, p. 76–79, 2000.

[108] J. Quan and J. Tian, "Circular polymerase extension cloning of complex gene libraries and pathways," *PLoS ONE*, vol. 4, no. 7, 2009.

[109] I.-S. Jang, B. J. Yu, J. Y. Jang, J. Jegal, and J. Y. Lee, "Improving the efficiency of homologous recombination by chemical and biological approaches in *Yarrowia lipolytica*," *PLoS ONE*, vol. 13, no. 3, 2018.

[110] R. Kooistra, P. J. J. Hooykaas, and H. Y. Steensma, "Efficient gene targeting in *Kluyveromyces lactis*," *Yeast Functional Analysis Report*, vol. 21, no. 9, p. 781–792, 2004.

[111] C. Yang, L. Qiu, , and Z. Xu, "Specific gene silencing using rnai in cell culture," *Methods in Molecular Biology*, vol. 793, p. 457–477, 2011.

[112] O. Stüwer, R. Homme, D. Haferburg, and H. Kleber, "Production of crystalline surface-active glycolipids by a strain of *torulopsis apicola*," *Journal of Biotechnology*, vol. 6, no. 4, pp. 259–269, 1987.

A Appendix

Table A.1: Primers used during the course of the experimental work.

Primers	Sequence
P64	GGTGTCGACTCGCCAAATTC
P70	ACGCAAGAGCTGTCTAACTG
P71	TTAAAGGGACTTGGCGTCTAGCAC
P202	CAGCCTCTTCTAGTCCGCTCACAATTCC
P295	GGGCGTTCCTCCTCTGATGTTC
P621	AGAGGTAGTTCGCACATCCAAGCT
P725	CGCCAAGCTATTTAGGTGAC
P755	GCGATTTGTGTACGCCCGACAGT
P954	TCATAGCGAGTTTCTTTGCATGTG
P1020	TGCCAAGTCGTTCAACACAG
P1021	CTGAGACAGCAGCTTGTCAC
P1131	AACGACGGCCAGTGAATTGT
P1347	GCCAGTGCAACAAGTATGAG
P1526	GCCATTCAGGCTGCGCAACTG
P1834	TCTTACTGTCGGGAATTCGCGTTGGCCGATCAATTGAGGAGTGGCAGCATATAG
P1946	GGGCTCCTGTCGTTGTCAAGTCC
P2211	CGACGTTGTAAAACGACGGCCAGTG
P2220	GGGAGATGGGGGAGGCTAACTGAAAC
P9958	CTGCCACCGCGAGTTACTTGGAGGACAATA
P9959	TGCGCCGATCTCACGCCGTGCTCATGCTTAG
P10025	GGCATCATCATGACCACAAGAAGGAGGAGAAC
P10026	GAAAGTCGGGCGAGTCCGTTTCTTCTCAGC
P10027	GTGGAATTGTGAGCGGATAAC
P10097	CAGTTTGTAAACTCGGCTCCGAACAGGCCGCCTGCAGGTCGACCATATGGGAGAG
P10103	CTGTTGCCCTTCGAGCTGGAAATCCCGCGGCCATGGCGGCCGGGAGCATG
P10104	CCGGCCGCCATGGCCGCGGGATTTCCAGCTCGAAGGGCAACAGGGTAG
P10105	TTGATCAGAATTCGAACACTACTTCATCCTTTCCCAACCCATGAAATCC

Table A.2: Primers used during the course of the experimental work (cont.).

Primers	Sequence
P10106	GGGTTGGGAAAGGATGAAGTAGTGTTCGAATTCTGATCAATGAAG
P10107	TCTTCCCATTAGATCTTATTTCACAAACAAACGACCCAACACCCGTGCGTTTTATTC
P10108	ACGCACGGGTGTTGGGTCGTTTGTTTGTGAAATAAGATCTAATGGGAAGAG
P10109	TATGGTCGACCTGCAGGCGGCCTGTTCGGAGCCGAGTTTACAAACTG
P10191	TCAAGCTAGGGAAAGCACTATC
P10192	ACGTATATAAAACCTCCGATAAG
P10193	TCCAGCTCGAAGGGCAACAG
P10194	TGTTCGGAGCCGAGTTTAC
P10771	AAGGGCTCATTGAATATTCGTAGGG
P10772	GATGCTCCCTTTGTTCTCAG
P10773	ACTCGCTGAGAGCAGCTTTTAGTC
P10774	CTTGCGCGTTCTGTTATTC
P10775	CAGCCTGCATGAGAAACTCGCTAA
P10776	GAACTGCAGCCATTCTTGAG
P10806	TGTATGCTCGAAGCACAGAAGAGGCCGTATTGAACAGTTTGGAC
P10807	CAAACTGTTCAATACGGCCTCTTCTGTGCTTCGAGCATACATATCATGAG
P10904	GTCTTGGTTAGAACGTGAGCAGATA
P10837	TACACACACTAAATTAATGATTCTTTATGCTGTGCTGGGCGCATT
P10838	GTGAAGGAATACCGCCAAGCGGAATTTTACTCAGTTTCACTCGGACGCCCGACCCTCCG
P10839	ACAAATACACACACTAAATTAATGATTTTTTATGCTGTGCTTGGCGCTGTG
P10840	AGGTGAAGGAATACCGCCAAGCGGAATTTTACTCAGTTTTACTCGGACGCCCGACCCTC
P10841	CAAATACACACACTAAATTAATGATTTTTTATGCTGTGCTTGGCACTGTG
P10842	AGGTGAAGGAATACCGCCAAGCGGAATTTTACTCAGTTTTACTCGGACGCCCGACCCTC
P10843	ACAAATACACACACTAAATTAATGAACATTAATTTCTCTGACGTGCTCGTGCTAG
P10844	AGGTGAAGGAATACCGCCAAGCGGAATTTTGCGAATGAACTTCGCTATGAC
P10845	CACAAATACACACACTAAATTAATGTTTATTGGACTCTCAGACGCCCTTG
P10846	AGGTGAAGGAATACCGCCAAGCGGAATTTTGCGGATGAACTTTGCCATTAC
P10847	CACGCAAACACAAATACACACACTAAATTAATGTTTATTGGACTCTC
P10848	AGGTGAAGGAATACCGCCAAGCGGAATTTTGCGGATGAACCTTGCCATTAC
P10849	CACGCAAACACAAATACACACACTAAATTAATGATTATTGATCTTTC
P10850	AGGTGAAGGAATACCGCCAAGCGGAATTTTACGAATGAATCTCGCCATCAC
P10851	AAAATTCCGCTTGGCGGTATTC
P10852	TAATTTAGTGTGTGTATTTGTGTTTGCGTGTCT
P10867	TAATTTAGTGTGTGTATTTGTGTTTGCGTGTCTATAG
P10868	AAAATTCCGCTTGGCGGTATTCCTTCACCTAGCACTG

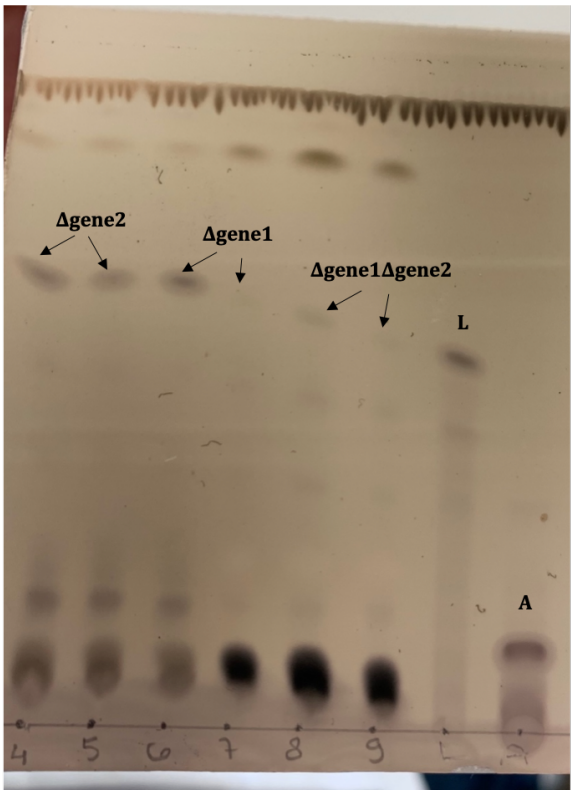

Figure A.1: TLC analysis of product samples from the Δgene2, Δgene1 and Δgene2Δgene1 SB strains of the fifth day of the growth trial. The last two coluns correspond to lactonic (L) and acidic (A) di-acetylated sophorolipids (standards).