“The future depends on what you do today.”

— M. Gandhi

Examination committee: Prof. Dr.Bart Devreese (promotor)

Prof. Dr. Savvas Savvides (chairman)

Prof. Dr. Nico Callewaert

Prof. Dr. Ronnie Willaert

Prof. Dr. Tom Desmet

Prof. Dr. Chris de Koster

Dr. ir. Inge Van Bogaert

Dr. Sara Groeneboer

This work was supported by IWT-Vlaanderen “Biosurf” project and the European FP7 project “Biosurfing”.

Department of Biochemistry and Microbiology

Laboratory for Protein Biochemistry and Biomolecular Engineering

Unit for Mass Spectrometry and Proteomics

Proteomic study of the sophorolipid producer *Starmerella bombicola*

Katarzyna Ciesielska

2013

Dissertation presented to the Faculty of Sciences in partial fulfillment

of the requirements for the degree of Doctor of Science: Biochemistry

Promotor: Prof. Dr. Bart Devreese

Titel van het doctoraatproefschrift in het Nederlands:

Proteoom onderzoek van de sophorolipide producent *Starmerella bombicola*.

To refer to this thesis:

Ciesielska K. (2013) Proteomics study of the sophorolipid producer *Starmerella bombicola*.

PhD thesis, Faculty of Science, Gent University,Gent, Belgium.

Cover: Soap bubbles

http://graphics.ucsd.edu/~iman/SoapBubbles/Soap_bubbles_RGB.png

Acknowledgements

My adventure as a PhD student has reached its end. Looking back, I feel very happy that I took up this challenge and embarked upon a journey of scientific discovery. It was one on the best decisions I ever made and, over the past 5 years, my stay at the L-ProBE laboratory greatly influenced my development as a person and scientist.

Firstly, I would like to express my very great appreciation to my promotor Prof. Bart Devreese. I want to thank him for his support, critical comments and his always open office door. He continuously believed in me and encouraged me to explore many different disciplines of science. Secondly, I wish to thank Dr. Sara Groeneboer who helped me see things in a wider perspective. Thank you for pointing me in the right direction, and for your kindness and friendliness. Our brainstorming sessions were priceless. I am very lucky I could work with you.

I wish to thank all colleagues from L-ProBE for sharing their knowledge about mass spectrometry and proteomics. Thank you for a warm welcome, Prof. Jozef Van Beeumen, Isabel, Ingrid, Rita, Gonzalez, Isaak, Isabel, Griet, Laurence, Simon, Yurong, Andreia, Agata, Silke, Catherine, Debbie, Jan, Erwin, Sammy, Aleksandra, Ruben, Wouter, Kedar and Yehudi. I specially want to thank Prof. Savvas Savvides for replying to my Erasmus application in 2008, which meant the start of my L-ProBE experience, and to Dr. Bjorn Vergauven and Ann who instigated my interest in science.

The Biosurf/ing projects would not exist without Dr. Inge Van Bogaert. Inge, thank you for sharing your knowledge about *S. bombicola* and biosurfactants, and for your valuable comments to my research. The interest you showed to my work was strongly motivating and gave me energy to go further, deeper, faster. I also wish to thank everyone from the InBio team for a great collaboration: Prof. Wim Soetaert, Karen, Robin, Kwok en Stijn. Sophie, thank you for infecting me with you enthusiasm and energy during our stimulating conversations. Your drive to solve the mysteries of *S. bombicola* is inspiring.

I am also thankful for the collaboration with team of Prof. Yves Van de Peer, Yao and Bing. Thank you for supporting my research with the transcriptomics study and for initiating me in the world of bioinformatics.

I would like to offer my special thanks to the members of my examination committee: Prof. Tom Desmet, Prof. Nico Callewaert, Prof. Ronnie Willaert and Prof. Chris de Koster. I highly appreciate your guidance and helpful suggestions, which were precious to improve my thesis and future publications.

My Polish and Belgium friends also deserve "Dziękuję !!!" and "Dankuwel !!!!" for their support, motivation, entertainment, an escape from work and great times together.

Goedele, Patrick and Jens, my Belgium family, thank your for unconditional support and helping hand in every situation. My husband, Kenneth, for saying “you can make it !!!” and for his love and friendship.

I particularly wish thank my parents, Anna and Jozef, and my brother and his family, Jacek, Ewa and Maciek. Without you I would not be here. You are the best supporters I could ever dream of, and you always believe more in me than I do myself. Thanks for your everlasting support!

Thank you !!!

Dankuwel !!!

Dziękuję !!!

September 27th, 2013 Kasia

Table of contents

Abbreviations **1**

Chapter 1 Literature review 5

1 Introduction **7**

2 Biosurfactants **8**

2.1 Glycolipid biosurfactants 8

2.2 Sophorolipids 9

2.2.1 Supramolecular assemblies of sophorolipids 10

2.2.2 Sophorolipid producing organisms 12

2.2.3 Sophorolipid biosynthesis 13

2.2.4 Application of sophorolipids 16

3 Proteomics **17**

3.1 Quantitative proteomics 19

3.1.1 Chemical labelling 19

3.1.2 Metabolic labelling 20

3.1.3 Label-free proteomics 21

3.2 Yeast proteomics 22

3.2.1 Sacharomyces cerevisiae as a model organism 22

3.2.2 Other examples of yeasts proteomes 22

3.3 The yeast exoproteome 23

3.3.1 The secretion pathway in yeast 24

3.3.2 Cell wall bound protein 24

3.3.3 Yeast secretome studies 25

4 Lipase-catalysed intra-esterification **26**

4.1 Lipases 26

4.2 Intra-esterification of hydroxy fatty acids 27

5 Conclusion **31**

Aim of the study **33**

Chapter 2 SILAC-based proteome analysis of Starmerella bombicola sophorolipid production 37

Abstract **39**

1 Introduction **40**

2 Materials and methods **41**

2.1 Materials 41

2.2 Creation of lys1Δ mutant 42

2.3 Growth conditions 42

2.4 Sophorolipid extraction 43

2.5 Protein extraction and separation 43
2.6 Digestion with endoproteinase Lys-C 43
2.7 NanoLC-ESI-FT MS analysis 44
2.8 Protein sequence database generation 44
2.9 Analysis of SILAC datasets using MaxQuant 44
2.10 RNA-seq analysis 45
2.11 Data mining with Blast2GO 45
2.12 Glucose, ammonium and phosphate measurements 46

3 Results 46
3.1 Construction of a SILAC compatible S. bombicola lys1 Δ mutant 46
3.2 Culture conditions and determination of sampling time 46
3.3 General properties of the datasets: gene ontology analysis and ratio distribution 47
3.4 Protein quantification 49
3.5 mRNA profiling 57
3.6 Glucose, ammonium and phosphate measurements 58

4 Discussion 59
4.1 The sophorolipid biosynthesis pathway is induced upon leaving the exponential growth phase . 59
4.2 Stationary phase related changes in protein abundance 60
4.3 Proteins involved in metabolic changes accompanying the sophorolipid synthesis pathway 61
4.4 Stress response 62
4.5 Cytoskeleton/vesicle transport 63
4.6 What triggers sophorolipid production? 63

5 Conclusion 63

Supplementary tables 65

Supplementary figures 68

Supplementary data 70

Chapter 3 Exoproteome analysis of Starmerella bombicola results in the discovery of an esterase required for lactonization of sophorolipids 71

Abstract 73

1 Introduction 74

2 Materials and methods 74
2.1 Materials 74
2.2 Strain and culture conditions for exoproteome analysis and genetic engineering 75
2.3 Sophorolipid analysis 75
2.4 Extracellular proteins collection and concentration 76
2.5 PNGase F treatment of the exracellular proteins 76
2.6 SDS-PAGE separation 76
2.7 In-gel trypsin digestion 77
2.8 Nano LC-ESI-LTQ-FT ICR 77

2.9 Protein identification and sequence analysis 78
2.10 Invertase assay 78
2.11 Creation of the lactonase knock-out cassette 79

3 Results 80
3.1 Exoproteome identification 80
3.2 Gene ontology analysis of the exoproteome 80
3.3 Comparison between exponential and stationary phase extracellullar proteins 81
3.4 Semiquantitative protein abundance analysis 86
3.5 Exoproteome comparison among yeast species 86
3.6 Invertase 87
3.7 Discovery of missing lactone esterase 88

4 Discussion 89
4.1 Comparison with other exoproteomes 89
4.2 Characterization of the exoproteome 90
4.3 Phase specific proteins 91
4.4 GO 92
4.5 Lactone esterase 92

5 Conclusion 93

Supplementary tables 95

Chapter 4 Lactone esterase of Starmerella bombicola : expression, purification and enzymatic properties 101

Abstract 103

1 Introduction 104

2 Material and Methods 105
2.1 Materials 105
2.2 Sophorolipids preparation 105
2.3 DNA construct 105
2.4 Protein expression 106
2.5 Protein purification 107
2.6 SDS-PAGE and trypsin digestion 107
2.7 Mass spectrometric analysis of tryptic fragments 107
2.8 N-terminal protein sequence analysis 108
2.9 HPLC and LC-MS analysis of glycolipids 108
2.10 Analysis of substrate selectivity 108
2.11 Temperature and pH dependency of rSBLE 109
2.12 Preliminary kinetic analysis 109
2.13 Evaluation of lipase activity 109
2.14 Evaluation of esterase activity 109
2.15 Analysis of glycosylation 110
2.16 Site-directed mutagenesis and mutant analysis 110

3 Results 110

3.1 Expression in Pichia pastoris 110
3.2 Purification of rSBLE 111
3.3 Amino acid sequencing and glycosylation analysis of rSBLE 112
3.4 Homology with other lipases 113
3.5 Substrate (SL) specifity 114
3.6 pH and temperature dependency of rSBLE 115
3.7 Preliminary kinetic analysis 122
3.8 Esterase activity 125
3.9 Site-directed mutagenesis 126

4 Discussion 127

5 Conclusion 130

Supplementary tables 131

Chapter 5 General conclusions and future perspectives 133

Intracellular proteome 135

Extracellular proteome characterization 136

Lactone esterase 138

Chapter 6 Summary 141

Samenvatting 146

References 149

Curriculum vitae 163

Abbreviations

2D-PAGE	two-dimensional gel electrophoresis
ABC	ATP-binding cassette
acetyl-CoA	acetyl coenzyme A
ACN	acetonitrile
AOX	alcohol oxidase
AT	acetyltransferase
ATP	adenosine triphosphate
BCA	bicinchoninic acid
BMGY	buffered-glycerol complex medium
BMMY	buffered-methanol complex medium
bp	base pairs
C18	octadecyl carbon chain
CALA	*C. antarctica* lipase A
CALB	*C. antarctica* lipase B
cDNA	complementary DNA
CMC	critical micelle concentration
Cyp52M1	cytochrome p450 gene
Dap1	damage resistance protein1
diAc	diacetylated
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediamine tetra acetic acid
emPAI	exponentially Modified Protein Abundance Index
EndoH	endoglycosidase H
ER	endoplasmic reticulum
ESI	electrospray ionization
EtOH	ethanol
FA	formic acid
FDR	false discovery rate
FPKM	fragments per kilobase of exon per million fragments mapped
FT-IR	fourier transform infrared spectroscopy
FTICR	fourier transform ion cyclotron resonance mass spectrometry

FW	forward
G3P	glyceraldehyde 3-phosphate
GeLC	in-gel tryptic digestion followed by liquid chromatography
GO	gene ontology
GPI	glycophosphatidylinositol
GPP	(DL)-glycerol-3-phosphatase
H/L	heavy to light ratio
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOG	high osmolarity glycerol
HPLC	high-performance liquid chromatography
ICAT	isotope-coded affinity tags
ICDH	isocitrate dehydrogenase
ICPL	isotope-coded protein label
iTRAQ	isobaric tags for relative and absolute quantitation
(k)Da	(kilo)Dalton
KO	knock-out
LB	Luria-Bertani
LC	liquid chromatography
LTQ-FT	Linear Ion Trap Mass Spectrometer combined with a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer.
lys-C	endoproteinase Lys-C
Lys1	saccharopine dehydrogenase
m/m%	mass to mass concentration
m/v%	mass to volume concentration
m/z	mass-to-charge ratio
MALDI	matrix assisted laser desorption/ionisation
Mbp	megabase pairs
MDR	multidrug resistance gene
MEL	mannosylerythritol lipids
MeOH	methanol
MM	Michaelis-Menten
monAc	monoacetylated
MS	mass spectrometry
MS/MS	tandem mass spectrometry

MW	molecular mass
n	Hill coefficient (cooperativity coefficient)
NAD(+)	nicotinamide adenine dinucleotide
NADP(+)	nicotinamide adenine dinucleotide phosphate
nonAc	nonacetylated
OD600	optical density of a sample measured at a wavelength of 600 nm
ORCAE	online Resource for Community Annotation of Eukaryotes
pI	isoelectric point
PngaseF	peptide-N-glycosidase F
pnpa	paranitrophenyl acetate
pnpb	paranitrophenyl butyrate
pnpp	paranitrophenyl palmitate
PPL	porcine pancreatic lipase
REV	reverse
RNA	ribonucleic acid
rSBLE	recombinant *S. bombicola* lactone esterase
rSBLEm	rSBLE mutant
RT-PCR	real-time polymerase chain reaction
r	correlation coeficient
SBLE	*S. bombicola* lactone esterase
SD	synthetic dextrose
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	stable isotope labeling by amino acids in cell culture
SL	sophorolipids
TAG	triacylglyceride
THF	tetrahydrofuran
TLC	thin layer chromatography
TMT	tandem mass tag
UGTA1	UDP-glucosyltransferase A1
UGTB1	UDP-glucosyltransferase B1
URA3	orotidine 5’-phosphate decarboxylase
WGS	whole-genome shotgun
WT	wild type
wt%	mass percentage

v/v%	volume concentration
XIC	extracted-ion chromatogram
YNB	yeast nitrogen base
α	degree of ionization
µRPLC	microcolumn reversed phase liquid chromatography

Chapter 1

Literature review

1 Introduction

A 'surfactant' is described as a ***surf**ace **act**ing **a**ge**nt.*** Surfactants are typically organic compounds, composed of both a hydrophilic and a hydrophobic moiety that interact with phase boundaries in heterogeneous systems. They reduce the surface tension at the interfaces between liquids, solid or gas phases allowing them to mix and disperse in water. This ability makes them very useful in everyday life as chemicals or detergents, and they are applied in foods, cosmetics or household products [1]. Unfortunately, most surfactants are produced chemically, often as by-products from oil refinery processes. Moreover, they have a low biodegradability which leads to significant environmental issues. Some surfactants are reported as carcinogens or endocrine disrupters indicating the potential danger for both human health and the ecosystem.

Fortunately, recently environmental friendly substitutes produced by microorganisms became available, termed biosurfactants, which are non-toxic and biodegradable. Among the different types of biosurfactants, glycolipid biosurfactants caught the most attention of the industrial and academic world. Besides displaying the desired physicochemical properties, they can be produced in high yields using renewable resources, which is in line with the worldwide policy towards a more sustainable industry.

In this thesis, we will focus on a specific class of biosurfactants, sophorolipids which are composed of a disaccharide, sophorose, linked to C-16 or C-18 hydroxylated fatty acids. Sophorolipids are already included in commercially available cleaning products (e.g. Ecover) or cosmetics (e.g. Soliance). Their structures and applications will be described further in detail in this chapter. To have full control over their production in an industrial setting, the biochemistry behind sophorolipid biosynthesis needs to be completely understood. The fungus *Starmerella bombicola* is the best sophorolipid producer and was selected as a model organism to study the mechanism of sophorolipid biosynthesis. In this introduction, we will further summarize the current knowledge about sophorolipid biosynthesis in this organism, both at the genetic and biochemical level. We will also address some specific topics that introduce the reader to forthcoming chapters.

As a part of the experimental work performed in this doctoral research project we performed quantitative proteomic technologies on this yeast. Using a SILAC-based proteomic experiment, we compared the *S. bombicola* proteome of a sophorolipid producing and non-producing state. Therefore, in this introduction we will also outline the basics of this proteomic technology and highlight the importance of yeast proteomics in biotechnology and medicine. In this study, we also focussed on the extracellular proteome. Although this part of the proteome contains many interesting proteins that are important in food management and protection from the environment, it received less attention. We will provide a brief overview of the current knowledge about the main functionalities found in the fungal extracellular environment and the proteomic strategies used to address.

Finally, to provide the necessary background to chapter 4, the biochemistry of lactonization is outlined, with a focus on substrates similar to sophorolipids (*i.e.* ω-hydroxy fatty acids).

2 Biosurfactants

Naturally occurring biosurfactants are environment-friendly alternatives for chemical surfactants. They are secreted by microorganisms and their biological role can differ from species to species. They can be used to assimilate water-insoluble nutrients or to improve attachment of cells to a hydrophobic surface. They also can serve as external carbon source or they can be involved in defence mechanisms as a consequence of their antimicrobial properties [2]. Biosurfactants are preferred over their chemical counterparts because they can be produced under mild conditions, and they display a lower toxicity, higher biodegradability and environmental compatibility. They are already employed in environmental applications such as bioremediation and dispersion of oil spills [3]. Other potential applications of biosurfactants relate to food, cosmetics, health care and cleaning. According to Transparency Market Research, the global biosurfactants market has grown extremely over the last few years and has a great prognosis for the future [4]. An increasing interest for bio-based products and concerns among consumers about the environmental effects of available products are driving the growth of this market. In 2011 the global biosurfactants market was worth USD 1,735.5 million while it is expected to reach USD 2,210.5 million in 2018 [4]. It is predicted that the household detergents and personal care segment will contribute 56.8% of the global biosurfactants market in 2018.

Biosurfactants can be divided into low-molecular-weight molecules that lower surface and interfacial tensions efficiently, and high-molecular-weight polymers that bind tightly to surfaces [2]. Structurally, biosurfactants are composed of a hydrophilic moiety which can be in the form of an acid, peptide cations or anions, mono-, di- or polysaccharides linked to a hydrophobic moiety composed of unsaturated or saturated hydrocarbon chains or fatty acids. Biosurfactants can be classified according to their chemical composition: glycolipids, oligopeptides and lipopeptides, phospholipids, fatty acids and neutral lipids, and polymeric biosurfactants. In this work, we performed research on glycolipid biosurfactants and therefore focus our introduction to this family of compounds.

2.1 Glycolipid biosurfactants

Glycolipid biosurfactants caught the most attention from the industrial and academic world because their production yield is higher than for other types of biosurfactants. In addition, they can be produced from renewable resources like sugarcane molasses, biodiesel co-products, waste from dairy industry or frying oil waste [5]. Their structures are difficult to mimic by chemistry. Glycolipid biosurfactants are composed of a hydrophilic part, represented by a mono- or disaccharide, and a hydrophobic part containing fatty acids, hydroxylated fatty acids or hydroxylated alcohols. Glycolipid biosurfactants are classified according to the nature of the sugar head. For example, rhamnolipids, produced by *Pseudomonas aeruginosa,* are comprised of either one or two rhamnose molecules attached to one or two β-hydroxydecanoic acids [6]. Mannosylerythritol lipids (MEL) are synthesized by *Pseudozyma antarctica* and occur as a mixture of four components: the major compounds MEL-A and MEL-B and the minor compounds MEL-C and MEL-D. The backbone of these molecules is a mannose-erythritol disaccharide on which short (C2-8) or long (C10-18) fatty acid chains are esterified [7]. In **Table 1.1**, some other known glycolipids are listed together with their producing

organism and industrial applications. The sophorolipids are the main subject of this study and they will be further discussed in more detail.

Table 1.1. The most popular glycolipids and their producing organism according to [8].

Glycolipid biosurfactants	Microorganism	Example of industrial application
Sophorolipids	*Starmerella bombicola, Candida apicola, Rhodotorula bogoriensis, C. antartica, Torulopsis petrophilum, C. botistae, C. riodocensis, C.stellata, Candida sp. Y-27208, C. floricola, Wickerhamiella domercqiae, Pichia anomala*	Antimicrobial, antiviral, spermicidal properties used in detergents and cosmetics
Rhamnolipids	*Pseudomonas aeruginosa, Pseudomonas chlororaphis, Serratia rubidea*	Bioremediation, antimicrobial and biocontrol properties
Cellobioselipids	*Ustilago maydis*	Antifungal compounds
Mannosylerythritol lipids	*Ustilago maydis, Pseudozyma anatarctica, Kurzanomyces sp*	Cosmetics, personal care and medical industry
Trehaloselipids	*Rhodococcus erythropolis, Arthrobacter sp., Nocardia erythropolis, Corynebacterium sp., Mycobacterium sp.*	Dissolution of hydrocarbons

2.2 Sophorolipids

Sophorolipids (SL) are glycolipid biosurfactants consisting of a sophorose sugar head and a hydrophobic fatty acid tail. Sophorose is a glucose disaccharide with a β-1,2 bond and can be acetylated at the 6'- and/or 6''positions. To the sophorose, a single terminal or subterminal hydroxylated fatty acid (C16 or C18) is β-glycosidically linked. The carboxylic end of this fatty acid can be free (acidic/open form) or internally esterified at the 4" position of the sophorose head (lactone form) (**Fig. 1.1**). Rarely, the esterification can occur at the 6'- or 6"-position. The hydroxyl fatty acid can contain one or more unsaturated bonds [9, 10]. The main producers are yeast species belonging to the *Starmerella* clade. They synthetize sophorolipids as a mixture of molecules which differ in the fatty acid part (chain length, saturation, and position of hydroxylation), in the acetylation pattern as well in lactonization [11].

The structure of the SL has a big influence on their physicochemical properties. Lactonic sophorolipids have better antimicrobial properties and have a better capacity to lower the surface tension, whereas the acidic sophorolipids are more soluble and are better foam formers [12]. Despite the fact that di- or monoacetylated SL are less soluble, they have better antibacterial [13], antiviral and cytokine stimulating effects [14].

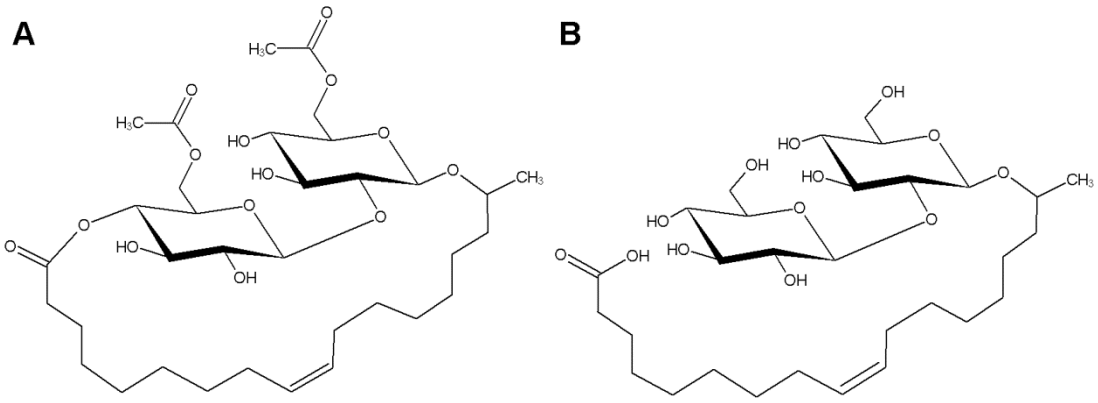


Figure 1.1. A. Lactonic diacetylated form of the sophorolipids. B. Acidic nonacetylated form of the sophorolipids.

2.2.1 Supramolecular assemblies of sophorolipids

Self-assembly is a process in which "single" components organize themself in a structure or pattern as a consequence of specific local interactions, without external forces. Self-assembly of sophorolipids is entropically driven and takes place above a certain concentration, termed the critical micelle concentration (CMC) [15]. Above CMC, surfactants organize themselves in water such that their polar head groups become oriented towards the water, whereas the hydrophobic tails cluster together. This orientation leads to the formation of various superstructures such as micelles, vesicles or multilayers [16].

In SL different types of assemblies are detected depending on SL concentration and pH. Especially acidic SL received a lot of attention because of their unique structural features including an asymmetrical polar head size (disaccharide vs. COOH) and a kinked hydrophobic core (cis-9-octadecenoic chain) [17]. Different techniques were employed to investigate the various structures of the self-assembly including light microscopy, small- and wide-angle X-ray scattering, FT-IR spectroscopy, and dynamic laser light scattering [17]. In acidic solutions (pH < 5.5), giant twisted and helical ribbons of 5−11 μm width and several hundreds of μm length were observed (**Fig. 1.2**). The ribbon formation decreases with increasing pH while at the same time the helicity and entanglements increases. The proposed model of for self-assembly was based on both strong hydrophobic associations between the fatty acid chains and strong disaccharide−disaccharide hydrogen bonding. At concentrations below 1.0 mg/mL, the size of self-assembled aggregates increased as the concentration increased. At concentrations above 1.0 mg/mL, micellar aggregates with a constant hydrodynamic radius of about 100 nm are formed [17].

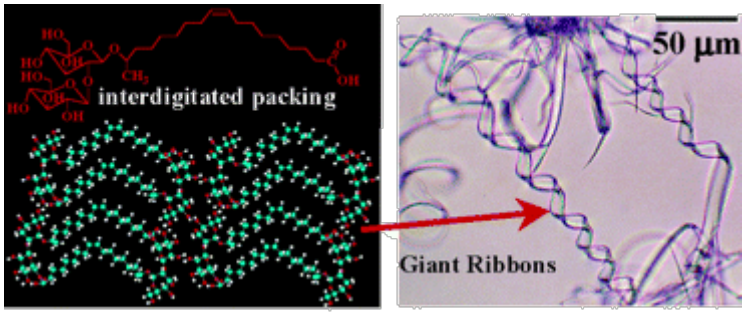


Figure 1.2. Proposed model of supramolecular assemblies of an acidic sophorolipid integrated in ribbons and a light microscopy image of ribbons formed in water at concentration below 1 mg/mL and at pH= 4.1 [17].

Other studies report that the degree of ionization, α, of the -COOH group, greatly influences the sophorolipid self-assembly [18]. When the degree of ionization increases, negative charges at the micellar surfaces are introduced, which initiate changes in shape, aggregation state and surface properties (**Fig. 1.3**). Micelles are formed at low (pH<5) and medium (5<pH<8) degrees of ionization. At high α (pH>8), large net-like aggregates are observed.

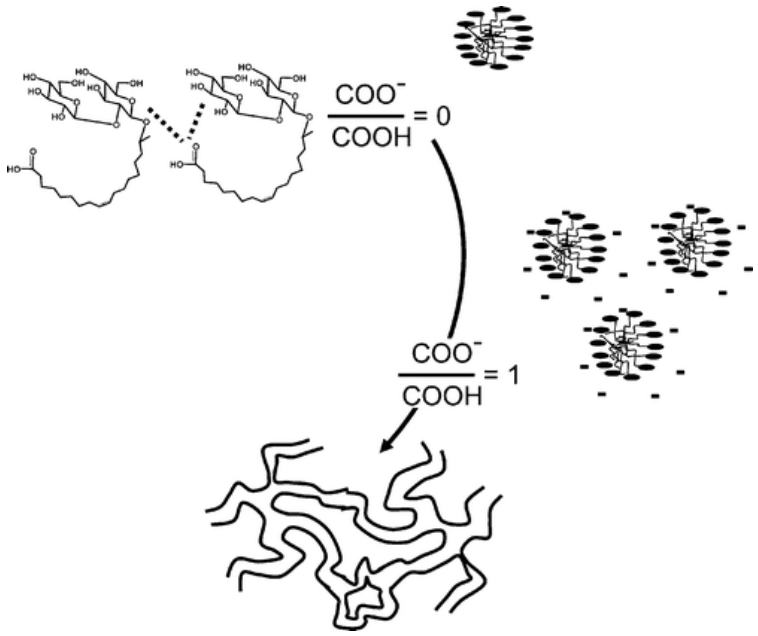


Figure 1.3. The self-assembly of acidic sophorolipids varies according to the degree of ionization [18].

In the same report, the morphology of the assemblies was evaluated in function of SL concentration. For sophorolipid concentrations <1 wt%, spherical micelles having an average radius of 3.0 nm existed. At a sophorolipid concentrations ≥ 1 wt%, micelles were no longer spherical and started to elongate from c > 0.5 wt%, and at c = 5 wt%, a cylindrical micelle shape was observed [18].

Similar self-assembly studies were performed on the acidic and lactonic forms of sophorolipids and their mixtures, using small-angle neutron scattering [19]. It was found that lactonic sophorolipids, which are more hydrophobic, form small unilamellar vesicles and at higher concentration a

disordered dilute phase of tubules. Acidic sophorolipids, in the concentration range of 0.5 to 30 mM, predominantly existed in the form of small globular micelles, along with low concentrations of larger, more planar aggregates (lamellar or vesicular). Their self-assembly properties were compared to weakly ionic or non-ionic surfactants with relatively large head groups. In mixtures of acidic and lactonic sophorolipids a micellar structure associated with acidic sophorolipids dominated.

2.2.2 Sophorolipid producing organisms

Candida apicola was the first species described to produce sophorolipids. The structure of the hydroxy fatty acid/sophoroside conjugate was elucidated as a partially acetylated sophorose unit β-glycosidically attached to 17-L-hydroxyoctadecanoic or 17-L-hydroxy-Δ9-octadecenoic acid [20]. However, for industrial applications, the most important producer of SL is *Candida* (syn. *Torulopsis*) *bombicola*, a non-pathogenic yeast isolated from the honey of *Bombus sp.* (the bumble-bee) (**Fig. 1.4)** by Spencer in 1970 [21]. This species can produce up to 400 g/L sophorolipids in fed-batch fermentation [22]. Later, the name *Starmerella bombicola* was proposed by taxonomists because they discovered a new clade *Starmerella* to which *Candida bombicola* was classified [23]. Strains from the *Starmerella* clade are fermentative and utilize a few carbon sources like glucose, galactose, raffinose and sucrose. They are osmotolerant, which indicates a specialization towards a microenvironment with a high osmotic pressure such as nectar. In 2010, Kurtzman discovered new members of the *Starmerella* genus which produce sophorolipids: *C. stellate, C. ricodensis and Candida sp. NRRL Y-27208* [24]. All these species produce predominantly acidic sophorolipids in contrast to *S. bombicola* and *C. apicola* which produce mostly the lacton form. Interestingly, in *Candida sp. NRRL Y-27208* a novel form of dimeric and trimeric sophorose containing sophorolipids were identified by MALDI-TOFMS (**Fig. 1.5**) [25]. Konishi *et al.* [26] discovered *C. batista,* which produces mainly diacetylated acidic SL containing mostly terminally hydroxylated octadecanoic acid as the lipid tail. Imura *et al.* [27] described *C. floricola* TM 1502 as a new SL producer, secreting mainly diacetylated acidic sophorolipids. Generally, in the last 3 years, the interest in the *Starmerella* clade of organisms increased. Other new strains were isolated from bees, flowers and fruits like *Starmerella caucasica sp.* nov. [28], *Starmerella jinningensis sp. nov* [29] and *Candida kuoi sp*. [30]. Only the last one was shown to produce acidic sophorolipids similar to *C. batiste, C. riodocensis* and *C. stellate.*

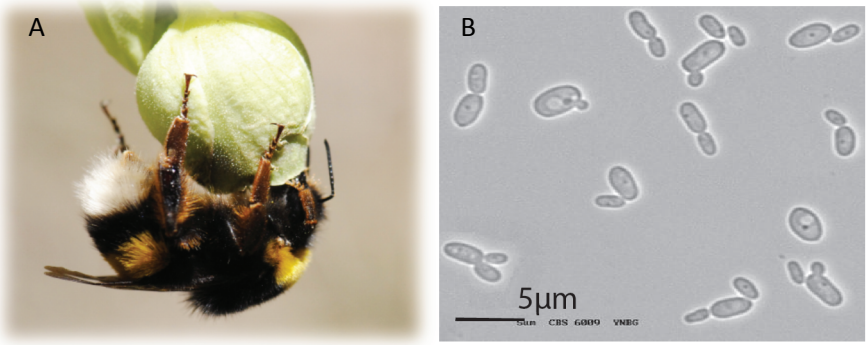


Figure 1.4. A. *Bombus terrestris* queen gathering probably the yeast-containing nectar [31]. B. Microscopic view of yeast *Starmerella (Candida) bombicola* (www.cbs.knaw.nl/yeast/BioloMICS.aspx).

Also other microorganisms, not belonging to the *Starmerella* clade were reported to produce sophorolipids. *Wickerhamiella domercqiae*, a strain isolated from oil waste was shown to secrete molecules that are almost identical to the major components of sophorolipids produced by *S. bombicola* and *C. apicola* [32]. Moreover, *W. domercqiae* is also able to produce lactonic diacetylated sophorolipids, with 17-hydroxyoctadecanoic acid as a lipid, in a high yield. Sophorolipid production was also reported in the thermotolerant yeast *Pichia anomala* [33]. However, the yield was very low and the structural properties of the products have not been described. Finally, Tulloch *et al.* [34] found a new form of sophorolipids produced by *Candida bogoriensis* (*Rhodotorula bogoriensis*). Its structure differs from the sophorolipids of *C. apicola* in the hydroxy fatty acid moiety, which is 13-hydroxyl dodecosanoic acid (C22).

This doctoral research focused on the biosynthesis of sophorolipids in *S. bombicola* ATCC 22214. This strain is the best producer of SL and is already used in the industry. In our project, we could build on a recent studies on *S. bombicola* sophorolipid production, conducted by Dr. IN. Van Bogaert and Prof. W. Soetaert at the Faculty of Bioengineering of Ghent University. They selected *S. bombicola* ATCC 22214 as a host for the production of new-to-nature biosurfactants.

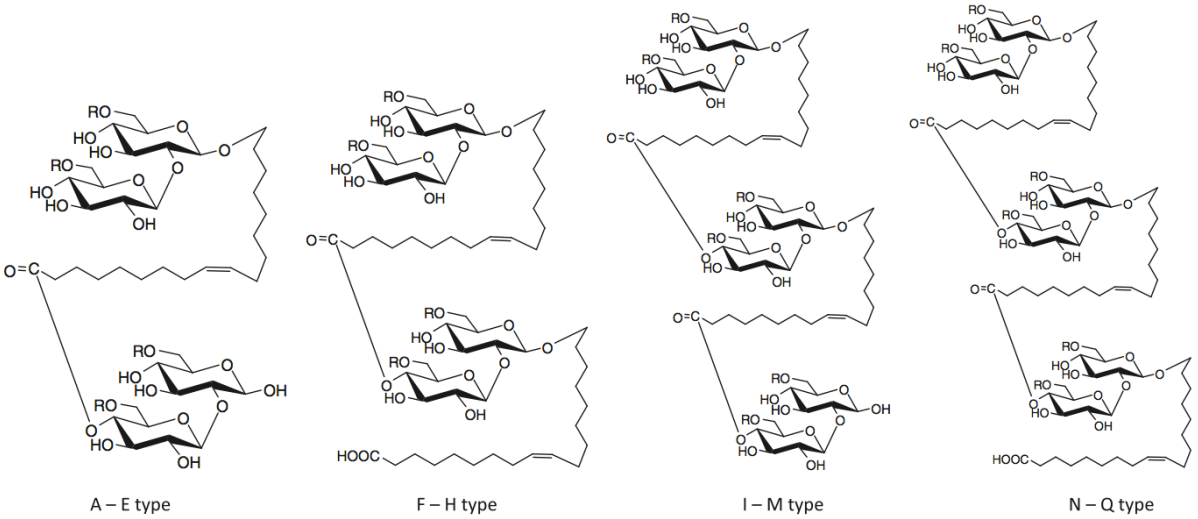


Figure 1.5. The structure of dimeric and trimeric sophorolipids produced by *Candida sp. NRRL Y-27208*. Seventeen polymeric sophorolipids (A-Q) were detected by MALDI-TOF MS. The structures shown are A-E, mono-acyl-disophorose; F-H, di-acyl-disophorose; I-M, di-acyl-trisophorose; N-Q, tri-acyl-trisophorose. -OR = O-acetyl groups [25].

2.2.3 Sophorolipid biosynthesis

Till now, most sophorolipid-related research concentrated on the optimization of the *S. bombicola* fermentation process, mainly to achieve a higher yield of SL's or to modify the SL structure. Different studies focused on the use of cheaper carbon sources to decrease the cost of SL production, aiming to make this process more attractive for the industry. According to Ashby *et al.* [35], the estimated price of the SL production is at $1.00—$3.00/kg while the production of synthetic surfactants is approximately $2.00/kg which make these bio-products already competitive.

For example, using molasses, rich in sugars and proteins, together with oleic oil, 53 g/L sophorolipids was produced [36]. A similar result was obtained with a biodiesel co-product stream, opulent in glycerol and fatty acids [35]. Interesting is also the utilization of the lactose rich dairy products [37]

and waste frying oil [38] as new carbohydrate sources for biosurfactant production. Moreover, modulating the fermentation technology can also influence yield. In fed-batch fermentation 400 g/L SL was generated compare to batch fermentation [22]. Generally, SL yields increase extremely when both hydrophilic and hydrophobic carbon sources are present in medium. This can be explained by the fact that when only a hydrophilic substrate is present, SL synthesis requires *de novo* fatty acid synthesis at cost of additional energy, and the efficiency of the process drops [39]. However, when a hydrophobic substrate is present, it is directly incorporated into SL [40]. Different hydrophobic carbon sources have been already tested, from alkanes, fatty acids, alcohols to esters. Interestingly, only those with a chain length similar to *de novo* produced sophorolipids (C16-C18) seem to be easily incorporated and result in high fermentation yields. Among the most favoured are rapeseed oil/esters which contain mainly C18:1 and C18:2 fatty acids [39]. In contrast, a rather poor integration of oils originating from coconut and meadow foam is observed, probably because they contain either mid or very long chain fatty acids [41].

This preference for C16 and C18-fatty acids can be explained by the specificity of cytochrome p450, which is proposed to be the first enzyme in the SL pathway and which hydroxylates mainly C16:0, C18:1 and C18:0 fatty acids [42]. Therefore, in case of a hydrophobic source with a different carbon chain length, the fungus needs to either shorten or lengthen the fatty acid chains, or, most likely, completely degrade them. In this case, *de novo* synthesis of the sophorolipids is required.

Biologically derived sophorolipids can be modified by chemo-enzymatic processes like amino acid conjugation [43], ester and amide formation [44]; acylation and lactonization [45]. However, all described methods are expensive and time consuming. The availability of genetically modified strains able to produce new–to–nature molecules with different properties would bypass the need for these expensive downstream processes. However, in order to create such strains, the knowledge about sophorolipid biosynthesis needs to be improved.

2.2.3.1 Sophorolipid biosynthetic enzymes are organized in a gene cluster

In 2007, Van Bogaert [11] proposed a group of enzymes that are involved in sophorolipid biosynthesis in *S. bombicola* (**Fig. 1.6**). In the first step of this pathway, the fatty acids are terminally or subterminally hydroxylated by a membrane-bound cytochrome P450 enzyme. Its expression during SL production was confirmed at the transcription level, by RT-PCR [46]. Next, a glucose molecule is bound to the hydroxylated fatty acid forming glucolipids. Then, a second glucose is attached creating an acidic sophorolipid. As described by Saerens *et al.* [47, 48], those two reactions are carried out by two glucosyltransferases UGTA1 and UGTB1, which use UDP-glucose as substrate. These molecules can be secreted as such or, alternatively, they first undergo an acetylation on the sophorose 6’ and 6’’hydroxyl groups by an acetyl-CoA dependent acetyltransferase before secretion [49]. Acetylated molecules then can undergo lactonization. At the start of this work, it was uncertain whether this was a spontaneous chemical reaction, or whether it is catalysed by a lactone esterase. It was also uncertain if this reaction occurs in the intra- or extracellular compartment.

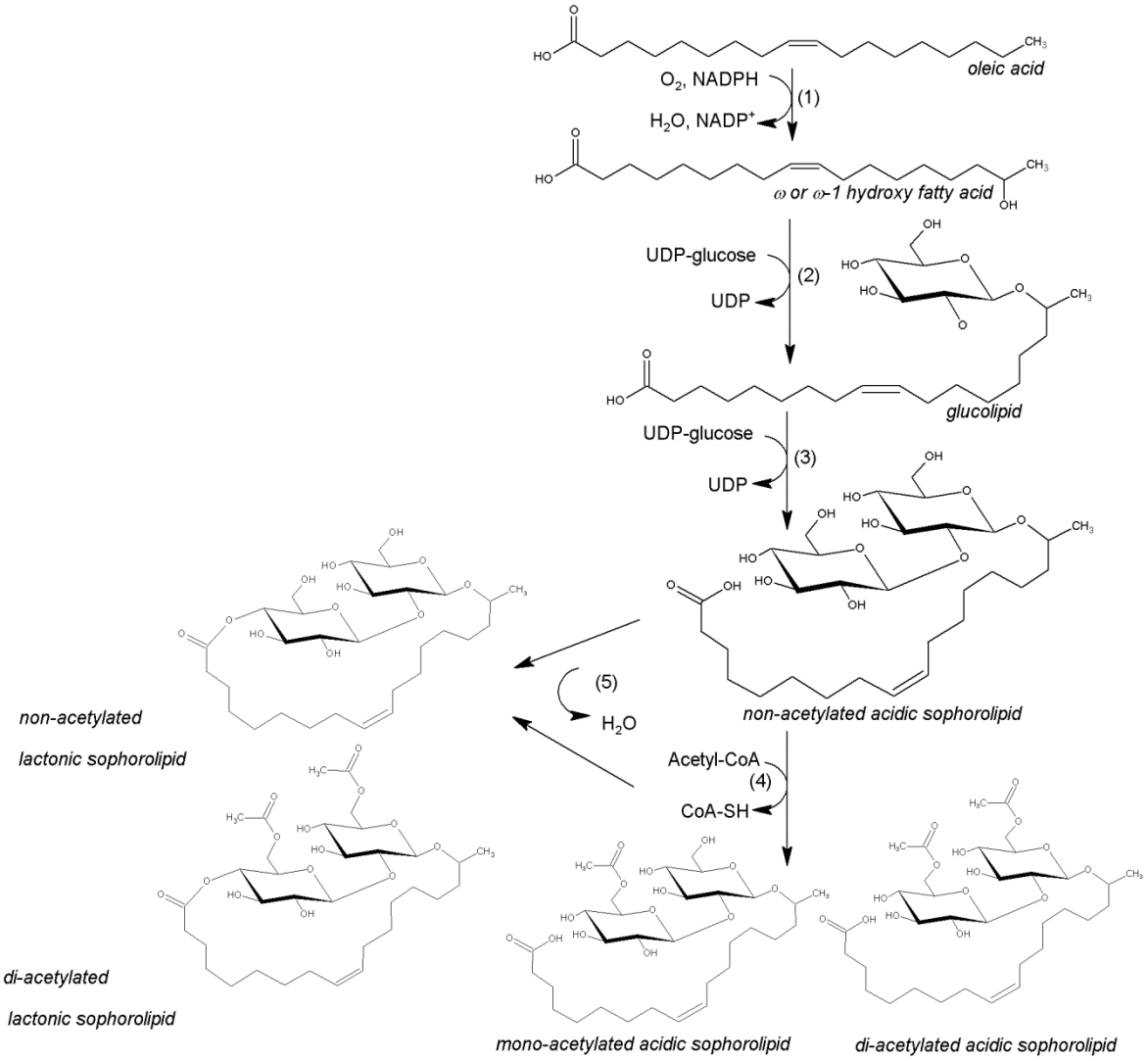


Figure 1.6. Sophorolipids biosynthesis according to [12]. (1) Cytochrome p450, (2) Glucosyltransferase 1, (3) Glucosyltransferase 2, (4) Acetyltransferase, (5) Probably lactone esterase.

Sophorolipids are mainly secreted during the stationary phase but their presence is not obligatory for cell viability. They are produced under starvation conditions and, therefore it is suggested that they act as an external carbon source and thus should be regarded as secondary metabolites. A large body of evidence in the scientific literature suggests that many pathways of secondary metabolites are maintained by enzymes encoded in a gene cluster [50]. Indeed, during the last years several reports describe how clustered genes are responsible for the biosynthesis of comparable biosurfactant molecules in other species, like *Ustiligo maydis* [51] and *Pseudozyma flocculosa* [52]. Very recently, a sophorolipid cluster in *S. bombicola* was identified [53], thanks to the genome sequencing in the framework of the Biosurf project supported by the IWT, in which framework also this work was initiated. The earlier described enzymes: cytochrome p450, the two glucosyltransferases as well as a putative acetyltransferase were detected in the genome in close location (**Fig. 1.7**).

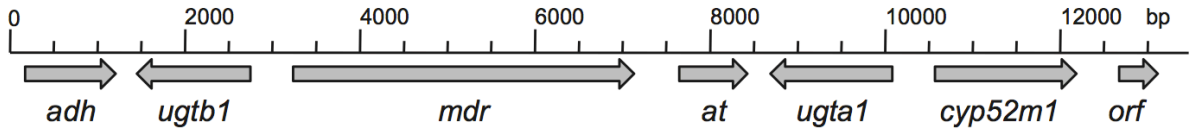


Figure 1.7. Cluster of the sophorolipids producing genes [53]. Genetic organization of the sophorolipid gene cluster of *S. bombicola*. *adh*: putative alcohol dehydrogenase; *ugtB1*: second glucosyltransferase; *mdr*: transporter; *at*: acetyltransferase; *ugtA1*: first glucosyltransferase; *cyp52m1*: cytochrome P450 monooxygenase; *orf*: open reading frame with unknown function.

Additionally, a gene encoding an ABC transporter (MDRSL) was discovered. A KO strain secreted a significant lower amount of sophorolipids. However, some sophorolipids were still present in the medium which may indicate passive transport by other, perhaps non-specific, transporters [54]. Knocking-out the open reading frames left and right from the first glycosyltransferase (ugta1) and P450 gene (cyp52m1), respectively, did not influence SL secretion. Curiously, no gene that could encode for a lactone esterase was discovered.

2.2.3.2 Regulation of sophorolipid biosynthesis

In both the cellobiose lipid and MEL gene cluster, gene expression is controlled by a transcription factor that is activated in response to a changing environment. For example, in *Ustilago maydis,* the production of biosurfactants occurs after transcription factor activation upon nitrogen limitation [55]. In the *S. bombicola* sequence cluster, no transcription factor was detected [53]. However, for *Candida apicola* it is suggested that the ammonium ion concentration is also important for sophorolipid production [56]. Also Davila *et al.* [57] described that sophorolipid production in *S. bombicola* is connected with nitrogen limitation. Later, Albrecht *et al.* [58] followed the nitrogen and phosphate concentration during *S. bombicola* growth and concluded that SL production occurred at total phosphate exhaustion and nitrogen limitation. It was also proposed that under these conditions NAD(+) and NADP(+)-dependent isocitrate dehydrogenase (NAD/P-ICDH) has a declined specific activity. This, together with a normal isocitrate synthase activity, causes the release of excessive citrate into the cytosol where ATP-dependent citrate lyase converts it into oxaloacetate and acetyl-CoA, the building block of fatty acids. Albrecht *et al.* followed the activity of NAD and NADP-ICDH during growth in a bioreactor and also tested the cofactor's influence on the activity of those enzymes. It was suggested that the decline in specific activity of NAD/P-ICDH is not regulated at the enzyme activity but at the enzyme synthesis level.

2.2.4 Application of sophorolipids

Over the last decade, the unique properties of the sophorolipids found many applications in different areas of daily life. Their ability to lower the surface tension, together with their biodegradability and low toxicity makes them attractive compounds for a new generation of environmental friendly cleaning products. As such, sophorolipids are the active agents of different household detergents produced by the Belgian company Ecover (http://www.ecover.com/). Also a Japanese company Saraya (http://worldwide.saraya.com/) includes them in their dishwasher product. Moreover, sophorolipids' emulsifying properties can be used in the petroleum industry, especially in secondary oil recovery and for removing the hydrocarbons from drill material [59, 60]. Sophorolipids are helpful in decontaminating soils and waters from hydrocarbons [61] and in removing heavy metals from sediments [62]. In the food industry, sophorolipids, as emulsifiers, can improve the quality of wheat flour products [63]. In cosmetics, the French company Soliance

(http://www.soliance.com/) uses sophorolipids because of their excellent emulsifying and antibacterial properties. These glycolipids act against acne, dandruff and body odours. Recently, Soliance introduced a new product on the market, Sophogreen, a plant-based solubilizer containing a high concentration of sophorolipids. Sophorolipids are also claimed to have protective effects on skin, hair and nails, and therefore they are interesting substrates for cosmetic and dermatological products [64]. Moreover, they stimulate dermal fibroblast metabolism, collagen neosynthesis [65] and participate in wound healing processes [66]. Additionally, sophorolipids were described to have some anticancer properties connected with their special interaction with the plasma membrane [67]. Recently, glycolipids found an application in nanotechnology. As capping molecules of nanoparticles they improve the hydrophilicity and solubility [68], and could therefore find applications in carrying various biomolecules [69]. Moreover, they can act as structure directing agents for the production of nanostructured silica thin films [70]. Finally, the hydrophobic part of sophorolipids, terminally or subterminally hydroxylated fatty acids, can be used as monomers for forming polymers or for creating macrocyclic esters used in perfume industry [71].

3 Proteomics

Proteomics is defined as the study of all proteins (proteome) produced at a certain time, under certain conditions by a cell [72]. Proteomics is considered as the next level in the analysis of entire biological systems after genomics and transcriptomics (**Fig. 1.8**). The genome can be considered as stable and, as such, it does not reflect the adaptations of an organism to withstand a certain environment. Transcriptomics provides information on the differential expression of the gene content and can be used to analyse which genes are necessary to survive under certain conditions. This type of information is valuable to better understand the mechanism of biological processes in cells. The transcriptome, on the other hand, does not provide any information about post-transcriptional events (alternative splicing, translational regulation) and post-translational modifications that may significantly affect the real functionality of genes. Therefore, analysis of the protein content, responsible for the actual cellular functions in the cells, is proposed to reflect cellular conditions more accurately.

Currently, mass spectrometry is the most widely used technique for proteomic studies [73]. Proteins are identified and quantified by the characterization of their tryptically digested peptides and, following HPLC-separation, by electrospray ionization (ESI) or matrix assisted laser desorption/ionisation (MALDI)–based mass spectrometric analysis methods. The mass spectrometer generates spectra of the eluting peptides (MS) and allows fragmenting the most abundant peptide ions in tandem mass spectrometry (MS/MS). The tandem mass spectra are then searched against a protein sequence database, resulting in the identification of peptides from which a protein list is compiled. Protein identification is usually based on the detection of at least two unique peptides. While quantitation used to be based on spot intensity in 2D-PAGE, nowadays this is based on the mass spectral signals. This will be outlined further.

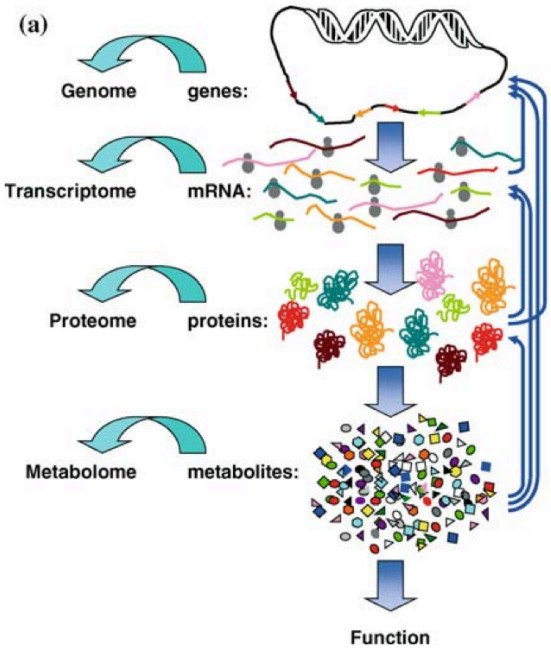


Figure 1.8. Scheme illustrating the 'omic' hierarchy: genomics, transcriptomics, proteomics, and metabolomics [74].

A major bottleneck in proteomic analysis is the high complexity and the large dynamic concentration range in the sample. To enhance the capability to also identify and quantify minor components in the sample, different methods of sample fractionation were created. Most of them are based on liquid chromatography. In our experiments, we use a two-dimensional separation strategy where SDS-PAGE is combined with nano-HPLC. In this so-called GELC approach, proteins are first separated according to their molecular weight on the SDS-PAGE gel. Individual bands are then cut, submitted to tryptic digestion and the resulting peptides are separated on a C18 column according to their hydrophobicity, by reverse phase chromatography. The latter is the most widely used method for peptide separation prior to MS as it can be directly coupled to the MS instrument thanks to the compatibility in terms of flow rate and solvents. In order to increase the sensitivity of LC/MSMS applications we used a nano LC system coupled to a high-resolution FTICR mass spectrometer [75].

This MS–based technology allows to create a platform to study the quantitative changes in the proteome upon a changing environment. This information is useful to understand the molecular function of each protein component and provide insights into the mechanism of the various biological processes.

3.1 Quantitative proteomics

In proteomics, one can focus on the identification of proteins in a cell or tissue. However, the technology can also be employed to determine the abundance of a certain protein in a sample and, in a comparative experiment, protein expression ratios can be calculated. Quantitative proteomics approaches fall into two groups, which either employ absolute or relative quantification methods. Using absolute methods, the exact amount of a protein in a biological sample can be measured. With relative quantification methods, the up- or down-regulation of protein production relative to a control sample is obtained, and results are expressed as a "fold" change increase or decrease [76]. In most relative quantification strategies, samples are labelled by adding a mass tag that allows comparing proteins from different samples with each other in a single experiment. Those tags are adding to the mass of the proteins or peptides without changing the biochemical properties. Typically, stable isotopic labels are used as mass tags. These can be introduced by a chemical, enzymatic or metabolic method. The choice of the labelling strategy depends on the biological problem that is addressed. In **figure 1.9** the most commonly used methods for differential labelling are displayed.

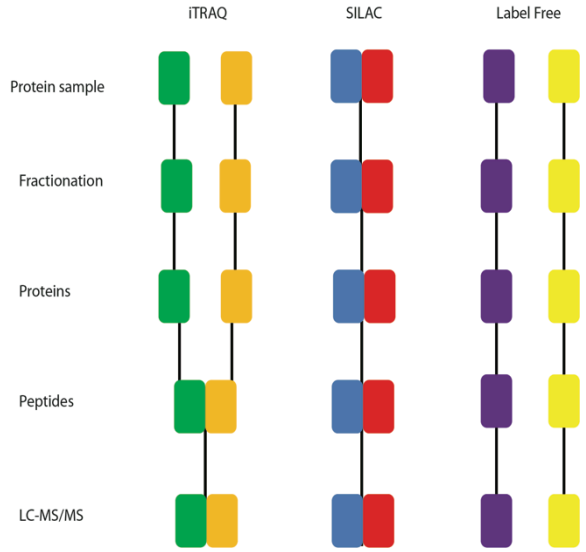


Figure 1.9. Experimental workflow for chemical (iTRAQ) and metabolic (SILAC) labelling and label-free quantitative proteomics. In iTRAQ the sample is labelled at the level of peptides while in SILAC labelling takes place earlier, during protein synthesis. In label free techniques samples are separately analysed while they are mixed in labelling methods.

3.1.1 Chemical labelling

Using chemical labelling strategies, isotopes are incorporated by attaching a chemical compound to the peptide or protein, mostly via the sulphydryl group of cysteines or to the primary amines of the N-terminus and lysine residues. In ICAT (isotope-coded affinity tags) a thiol-specific reactive group is attached to a biotin group via an isotope coded linker which contains 8 isotopes (H/D or $^{12}C/^{13}C$) for the light or heavy tags [77] (**Fig. 1.10A**). The main advantage of this method is that biotin-avidine affinity chromatography can be used to select cysteine containing peptides after tryptic digestion of the protein mixture, which helps to reduce sample complexity. However, the disadvantage of this method is the low sequence coverage, and the fact that not all proteins contain cysteine residues. A second type of the chemical labelling is iTRAQ (Isobaric Tag for Relative and Absolute Quantification)

(**Fig. 1.10B**) or related chemicals like ICPL and TMT [78]. It involves chemical labelling with isobaric reagents where stable isotopes are differently distributed between a so-called reporter group and a balance group. Up to 8 different of these reagents can be constructed, allowing to perform multiplex experiments by differently labelling 8 samples and consequently mixing them together. The isobaric labels allow the co-elution of the peptides originating from several samples and they are observed as a single peak in MS. The quantification is based on MS/MS analysis, as the isobaric labels each yield a different reporter fragment ion that allows relative quantification. The disadvantage of the method is that the labelling occurs at the very end of sample preparation, after tryptic digestion, which makes the technique prone to technical variations. However, the fact that 8 samples can be co-monitored, makes it an interesting tool for time-resolved experiments.

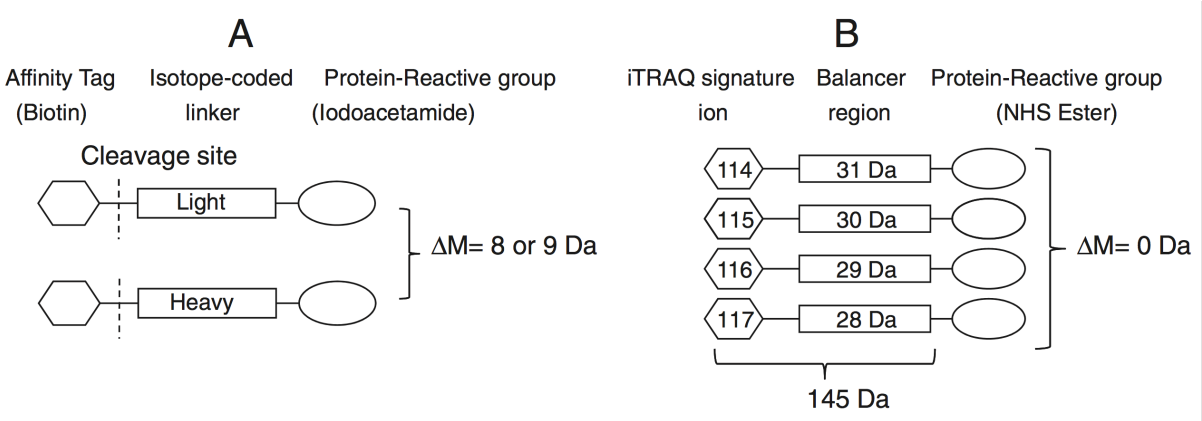


Fig 1.10 Chemical labelling methods: A, ICAT consists of a biotin affinity tag, isotopically labeled linker and peptide reactive group (PRG). It contains an acid cleavable site as depicted in the diagram resulting in a difference in mass of 9 Da between the labels. B, iTRAQ labels consist of asignature group, a balancer region and a PRG. All four labels are isobaric with the signature group and balancer adding up to 145 Da [79].

3.1.2 Metabolic labelling

In metabolic labelling, stable isotopes are introduced during protein biosynthesis. The first type of metabolic labelling applied in MS-based proteomics was ^{15}N labelling. Yeast were grown in ^{15}N-labelled salts as the sole nitrogen source in the medium [80]. This technique can be a good choice for autotrophic organisms, which can synthetize their own amino acids and incorporate the labelled nitrogen. However, nitrogen integration can vary from peptide to peptide according to their sequence, so the mass shift cannot be easily predicted [81]. The most widely used method for *in vivo* labelling is SILAC. SILAC is an acronym for 'stable isotope labeling of amino acids in cell culture' and has been introduced by Mathias Mann's group in 2002 [82]. In SILAC, samples to be compared are grown in two different media, one containing a 'light' form of an amino acid, the other containing a 'heavy' form of an amino acid, which actually carries stable isotopes such as ^{13}C or ^{15}N (**Fig. 1.11**). Lysine and arginine are the two most common labelled amino acids used in this technique because, after tryptic digestion, each peptide will contain typically a single isotopically labelled lysine or arginine at its C-terminus. After mixing of the samples, followed by tryptic digestion, each peptide will therefore appear as a doublet in the mass spectrum. In this case, the mass shift is predictable, even with the tolerance of one or two missed cleavage due to inefficient digestion. Relative quantification is then based on the intensity difference of the different isotopes in the mass spectrum. The SILAC technique is extremely sensitive and allowed to detect more than 2000 proteins

in a bakers' yeast proteomics study, as reported by de Godoy *et al* [75]. The early addition of the label into living cells reduces sample preparation bias, which is generally accepted as a major source of experimental variation in proteome analysis. Due to the availability of different labels multiple comparisons can be performed in parallel. Unfortunately, SILAC is not generally applicable for all organisms. Plants, for example, can synthesize their own amino acids from other nitrogen sources. Only organisms auxotroph for lysine or/and arginine can be used for this technique. In this work, for example, SILAC could be applied on *S. bombicola* only after genetic engineering.

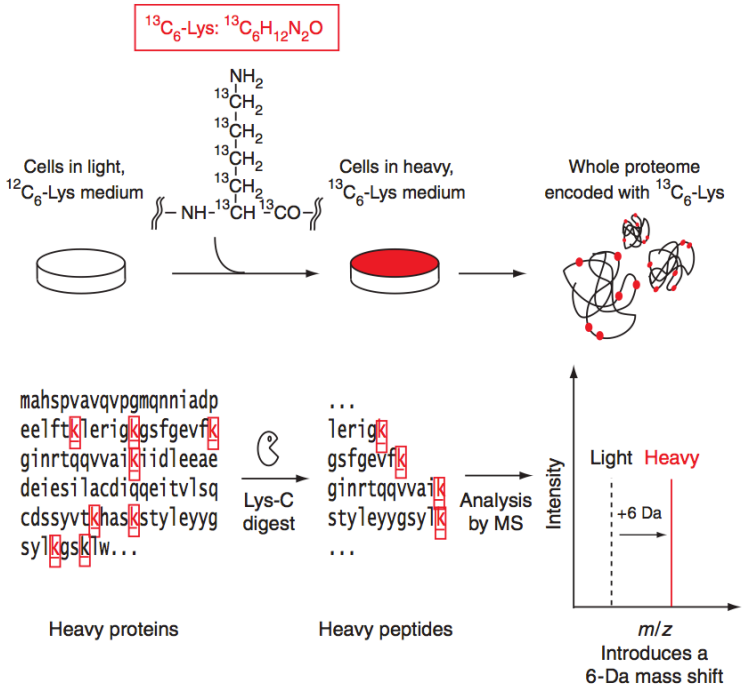


Figure 1.11. Outline of the SILAC strategy [83]. Cells growing in normal, light medium are subcultured in medium containing heavy, stable isotope-labeled amino acids like lysine with six $^{13}C_6$ (highlighted in red). During cell growth the labelled lysine is incorporated in every protein in the proteome. Digesting proteins with lys-C results in peptides bearing lysine-$^{13}C_6$. On MS, a mass shift of 6 Da of heavy from the light peptide is visible which allows to distinguish two differently labelled samples.

3.1.3 Label-free proteomics

Label-free quantification is a new trend in proteomics. In this method no label is added and the different samples are analysed by separate LC-MSMS runs. The quantification is then based on aligning the chromatograms and determining the differences in peak intensities. This requires very reproducible LC-chromatography, advanced LC-MS software tools, and a lot of MS measurements [84].

Some more simple methodologies are based on spectral counting. In a method referred to as Exponentially Modified Protein Abundance Index (emPAI), relative protein quantification is based on the number of identified peptides belonging to that particular protein. EmPAI correlates a protein's abundance to the number of detected unique peptides belonging to that protein divided by the theoretically number of predicted peptides of that protein [85].

3.2 Yeast proteomics

3.2.1 Sacharomyces cerevisiae as a model organism

Most applications of fungal proteomics evidently concern *S. cerevisiae* because it is the model organism for eukaryotes. It was selected as a model organism because it is easy to grow and culture, and it is rather straightforward to manipulate and transform this organism through homologous recombination. As a eukaryote, it shares a cellular structure with plants and animals. Despite the evolutionary distance, around 31% of its genes have human orthologs [86]. *S. cerevisiae* was the first eukaryotic organism for which the genome was completely sequenced (in 1996) [87]. The genome size is 12 Mb with 5800 predicted genes. Today, a *S. cerevisiae* deletion mutant library is available that covers 90% of the yeast genome [88]. The well-established genome information and the availability of molecular and cellular tools for this organism explain the vast amount of yeast proteomic studies. The number of new publications now reached a certain steady level indicating that proteomics has become a standard tool in yeast studies (**Fig. 1.12**).

S. cerevisiae was actually used as a model organism to develop and improve different proteomic technologies. Shevchenko [89] used 2D-PAGE and MS technology to link the yeast genome and proteome. De Godoy analysed the yeast proteome to study factors affecting the coverage of mass spectral analysis [75]. He managed to detect 2000 proteins in a single yeast proteome experiment and concluded that analysis of complex mixtures is not limited by sensitivity but by a combination of a high dynamic concentration range – where abundant peptides prevent sequencing of low abundance ones – and by the effective sequencing speed. Lastly, this year, Picotti presented an almost complete map of the *S. cerevisiae* proteome where 97% of the genome-predicted proteins were detected [90]. To achieve this, a strategy was developed by applying two types of mass-spectrometric mapping: one based on a discovery-driven (shotgun) strategy and a second supporting hypothesis-driven (targeted) proteomic approach. Together, the two versions of the map encompass almost the entire proteome and support many studies using modern proteomic technologies.

3.2.2 Other examples of yeasts proteomes

S. cerevisiae is not the only fungal species on which proteomic experiments have been performed. Indeed, several reports describe how proteome analysis can be useful in other fungi relevant to biotechnology and medicine. In case of pathogenic species like *Candida albicans* and *Aspergillus fumigatus*, proteome studies were used to determine their virulence mechanisms and to define new drug targets [91]. Further, a proteome study on wine [92] and brewing [93] yeast appeared to be a powerful tool to increase our knowledge about the biological processes in yeast during fermentation. A study of the *Pichia pastoris* proteome under conditions of osmotic stress [94] revealed that increased osmolarity of the growth medium does not have a beneficial effect on recombinant protein production, in contrast to other host organisms. These results might be used in future strain optimization efforts and bioprocess engineering of this biotechnologically relevant yeast.

Very interesting in the context of our work is the report on the oleaginous yeast *Lipomyces starkeyi* [95]. *L. starkeyi* has the capability to accumulate over 70% of its cell biomass as lipid. Analysis of *L. starkeyi* AS 2.1560 proteome samples from different culture stages during a typical lipid production process was performed using an online multidimensional µRPLC/MS/MS method. The results indicated a global response associated with lipid production upon nitrogen deficiency. Many of the

down-regulated proteins were related to glycolysis, whereas the majority of the up-regulated proteins were involved in proteolysis, carbohydrate metabolism and lipid metabolism. Such research provides essential information about cellular responses to nutrient availability and on the basic biochemistry of lipid accumulation, which can be further used in the design of engineered oleaginous yeasts.

Another interesting proteomic study was performed on the yeast–like fungus *Pseudozyma flocculosa* [96], which produces antifungal glycolipids, termed flocculosins. To understand the mechanism behind flocculosin production, a 2D-PAGE-MS experiment was performed to compare the proteome under conditions favouring fungal growth (control) with conditions conducive to flocculosin synthesis (stress). Up-regulated proteins or proteins specific for the flocculosin production-phase were associated with carbon and fatty acid metabolism, and also with the filamentous change of the fungus leading to flocculosin production. This report suggested that flocculosin synthesis is connected with specific stress limiting conditions. However, a biosynthetic pathway was not discovered.

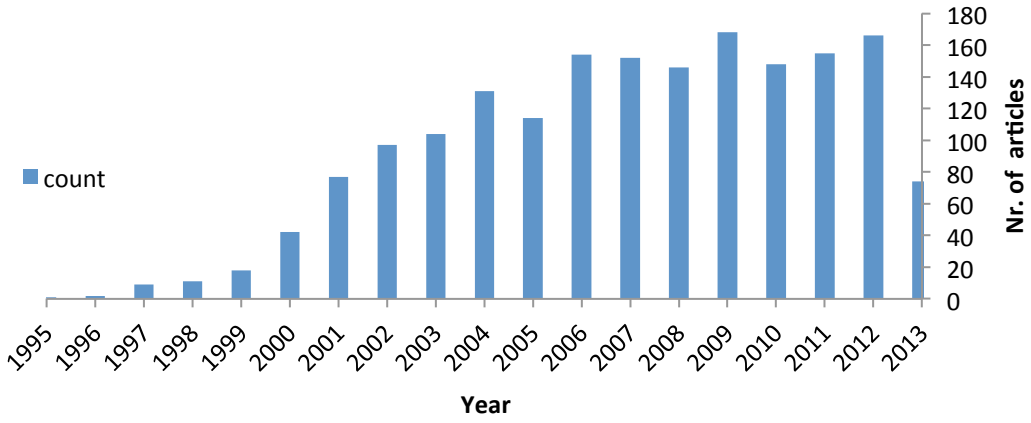


Figure 1.12. Diagram representing the distribution of publications connected with " yeast proteome" in PubMed database.

3.3 The yeast exoproteome

Since sophorolipids are extracellularly secreted, we also considered the extracellular protein landscape for proteomic analysis. Proteins secreted by yeast have multiple biological functions. They play a role in gathering nutrition, in maintenance and remodelling the cell wall, signalling and defence mechanisms. The term "secretome" is used to describe the pool of proteins found outside of the cell that are actively secreted by the classical secretion machinery. Mostly, these secreted proteins contain a signal peptide. As for the entire proteome, the content of the secretome is not stable and it can adapt to new conditions. Unfortunately, the secretome as such is difficult to determine. Indeed, the extracellular protein fraction also contains cytosolic proteins secreted via a non-classical pathway or resulting from cell lysis, and proteins released from the cell surface in the medium by mechanical shaking or resulting from cell wall remodelling. Nonetheless, by analysing the

extracellular protein content, which is termed exoproteomics, we may learn about the mechanism of cellular adaptation to the environment.

3.3.1 The secretion pathway in yeast

Proteins identified in the exoproteome can be divided into actively secreted - which possess a secretion signal (including cell wall bound proteins) - and intracellular proteins without secretion signal. The first group are proteins secreted by the classical secretory mechanism using the ER-Golgi exocytosis pathway [97] (**Fig. 1.13A**). A protein being generated by the ribosome enters the endoplasmic reticulum (ER) as a nascent protein chain through the interaction of the signal recognition particle with the secretion signal. The newly synthetized protein is typically glycosylated and then leaves the ER via specific exit sites from which vesicles are formed. These vesicles are used to transfer the proteins to the Golgi apparatus where they are modified, processed and sorted, and ultimately packed into vesicles directed to the cell surface from where proteins can leave the cell after merging of the vesicle with the cell membrane [98]. This pathway is also followed by proteins attached to the cell surface which form the so called “surfome”.

To the second group of secreted proteins are proteins without a secretion signal that originate from the cytosol or nucleus. To explain this, non-classical secretion pathways are proposed. Two of such non-classical ways were described in yeast, vesicular (**Fig. 1.13B**) or non-vesicular [99]. In non-classical vesicular secretion proteins are released through membrane bound structures that fuse with each other. An example of such a mechanism is the secretion of the stress-related Acb1 protein by autophagy [100]. Alternatively, the non-vesicular mode involves a direct passage of a protein across the plasma membrane. The best-known example for this mechanism in yeast is the mating pheromone α-factor [101]. This protein is secreted directly across the cell surface by a dedicated ABC transporter, Ste6.

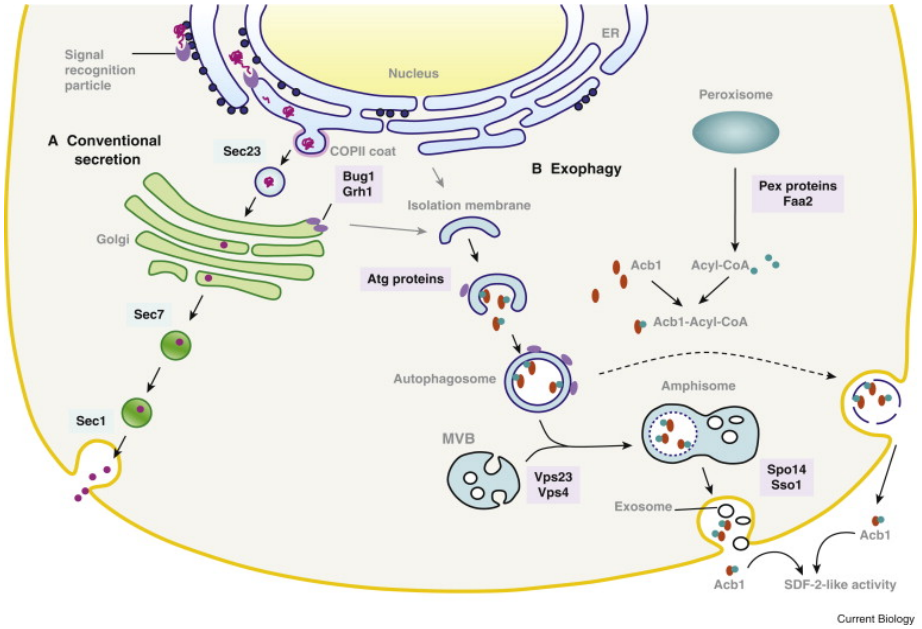


Figure 1.13. The conventional (A) and non-conventional vesicular (B) secretion pathway [102].

3.3.2 Cell wall bound protein

The yeast cell wall is a dynamic structure that participates in different cellular processes [103]. An important part of this structure consists of cell wall-bound proteins, also called the “surfome’’, which

assist in many yeast biological processes. They play a role in mating, colony morphology, biofilm formation or interaction with hosts [104]. The composition of the surfome depends on the growth conditions such as nutrition, temperature, oxygen level and pH.

Cell wall bound proteins can be attached to the cell wall via polysaccharides or an anchor molecule. Proteins bound to the cell wall polysaccharides can be either directly linked (through the γ-carboxyl group of glutamic acids) or indirectly link (through a β-1,6-glucan moiety) to the β-1,3-glucan in the cell wall [105]. Alternatively, proteins can be attached to the cell wall via disulphide-linkage with other cell wall proteins [106].

Proteins can also be hooked in the cell membrane via a glycosylphosphatidylinositol (GPI) anchor. Proteins undergo this modification in a post-translational process. The carboxyl-terminal signal peptide is removed from the protein and the new carboxyl-terminal is then combined with the GPI precursors. GPI is composed of a phosphatidylinositol group linked through a carbohydrate-containing linker (glucosamine and mannose glycosidically bound to the inositol residue) and, via an ethanolamine phosphate bridge, to the C-terminal amino acid residue of the mature protein. The two fatty acids within the hydrophobic phosphatidyl-inositol group anchor the protein to the cell membrane (**Fig. 1.14**). By adding the GPI anchor, yeast targets the protein to the cell surface and covalently immobilizes it there. GPI anchored proteins may have different functions.

To predict if a protein has a GPI anchor, its sequence is can be scanned for the presence of the N-terminal secretion signal and C-terminal GPI signal. This is implemented in software such as PredGPI [107].

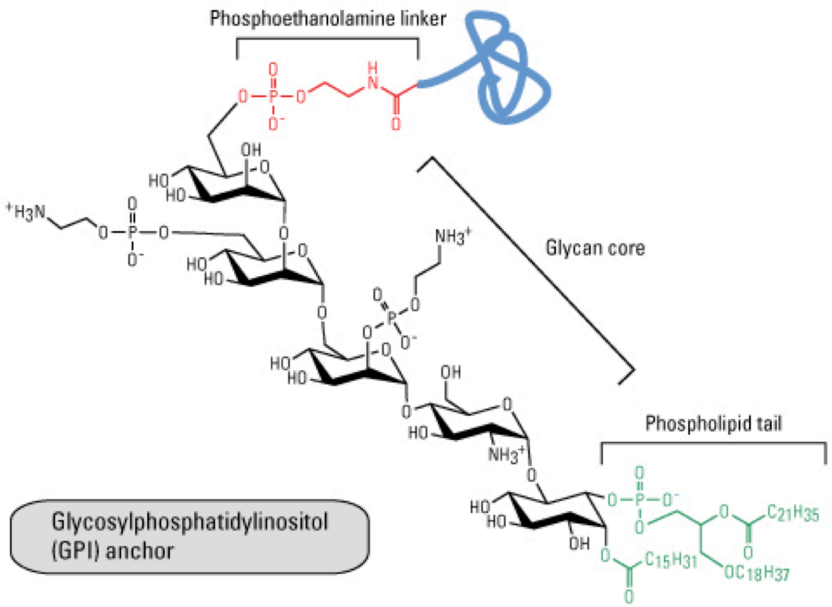


Figure 1.14. GPI anchor structure (http://www.piercenet.com)

3.3.3 Yeast secretome studies

The adaption of fungi to certain conditions is usually analysed by the characterization of the internal proteome of the cell, while, in general, the extracellular proteins are omitted. This is curious, since actually those proteins directly participate in cellular adjustments in response to a changing

environment. However, lately the interest in the yeast secretome increased. The secretomes of the biotechnologically important yeasts: *P. pastoris* [108], *K. lactis* [109, 110] and *C. utilis* [111] have been described. Those yeasts found biotechnological applications as hosts for recombinant protein secretion. In most of these studies, the content and size of the secretome was analysed to reveal possible factors negatively effecting heterologous protein production, as the presence of proteases. In other studies, the response of the secretome to different conditions was tested to collect knowledge for improvement of the heterologous protein production. Also the secretome of pathogenic yeasts like *C. albicans* [112] and *C. grablata* [113] was studied. These studies aimed to discover novel virulence factors that could be used to develop clinical markers or vaccins. In *C. albicans,* many secreted and cell wall-bound proteins are already known to mediate host cell adhesion during invasion, promote biofilm formation or protect against immune systems.

4 Lipase-catalysed intra-esterification

4.1 Lipases

Lipases are lipid-degrading enzymes widely distributed in microorganisms, plants and animals. In water, lipases hydrolyse the ester bond of compounds like triglycerides at the interface of water and the insoluble substrate. Lipases can also perform esterification of carboxylic acids with an alcohol (esterification) or with an alcohol residue of a second ester (inter-esterification) in organic solvent. Further, in a transesterification reaction an acyl residue is either transferred to an alcohol (alcoholysis) or to a carbonic acid (acidolysis). The product of an ammonolysis reaction is an amide and an alcohol (**Fig. 1.15**) [114].

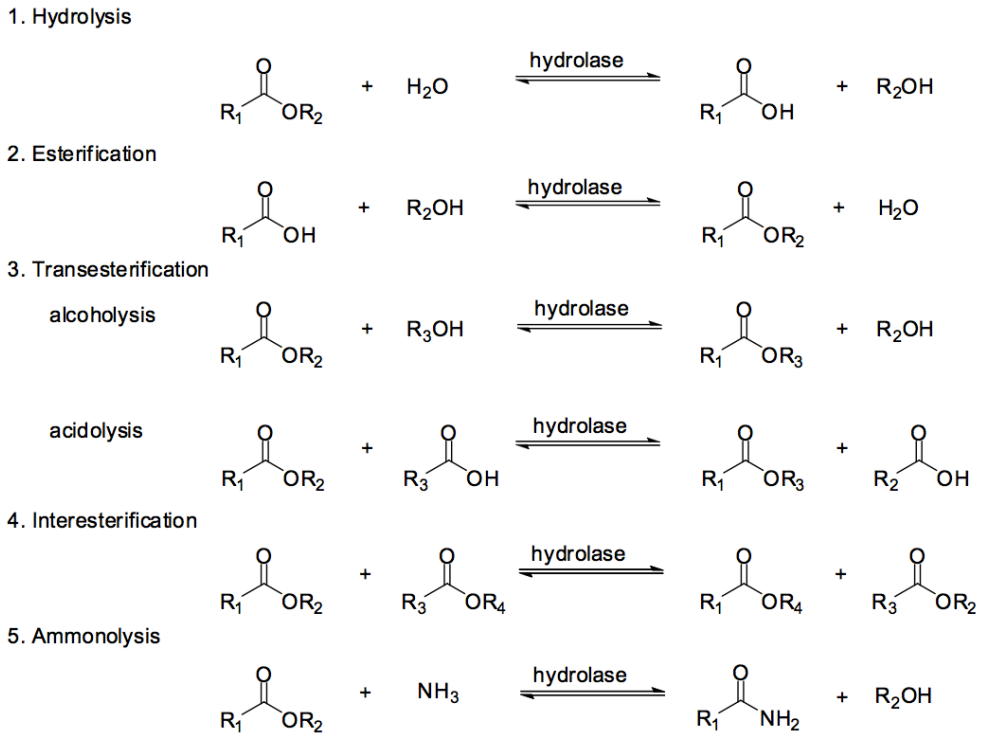


Figure 1.15: Possible reactions that are catalysed by lipases [115].

Most lipases contain a conserved GxSxG sequence motif that forms the substrate binding site [116]. Lipases belong to the serine hydrolases and the serine residue in the GxSxG motif belongs to the catalytic triad together with an aspartate and a histidine residue. The mechanism for ester bond hydrolysis or synthesis mediated by lipases is the same. First, the active-site serine reacts with the carbonyl-group of the substrate, yielding a tetrahedral intermediate that is stabilized by the catalytic histidine and aspartate residues. Then alcohol is released and an acyl-enzyme complex is formed. Attack of a nucleophile (water in hydrolysis, an alcohol in esterification) forms again a tetrahedral intermediate, followed by the release of the product (acid or ester) [117] .

Structurally, lipases are hallmarked by a α/β-hydrolase fold [118], which consists of eight-stranded β-sheet, flanked α-helices on both sides (**Fig. 1.16**).

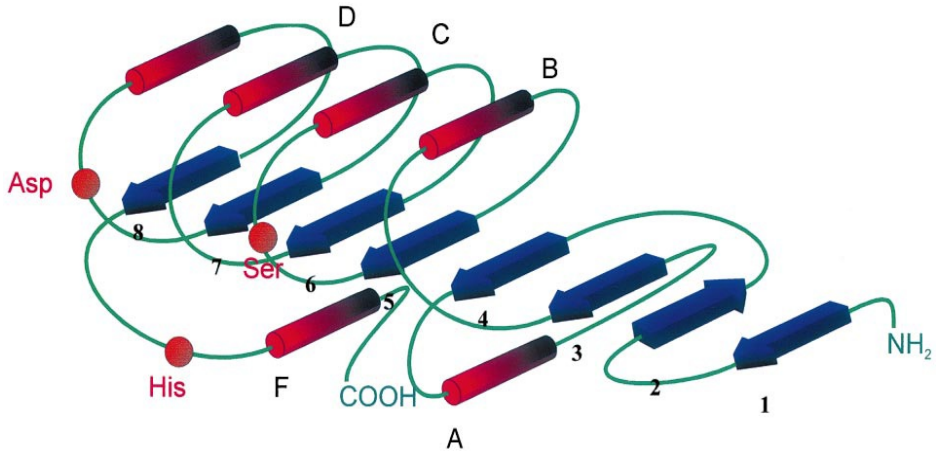


Figure 1.16. Schematic presentation of the α/β-hydrolase fold. β-sheets (1-8) are shown as blue arrows, α-helices (A-F) as red cylinders. The relative positions of the amino acids of the catalytic triad are indicated as red circles [118].

Many lipases undergo interfacial activation where the so-called lid-domain which covers the active-site pocket plays an important role. In water, when the substrate concentration is low, the lid is closed and the lipase is not active. As the substrate concentration increases and crosses the CMC, the substrate creates lipid micelles. By interacting with the oil-water interface, the lid then opens and the lipase is transformed into its active, open state. That is why the most favourite substrates for lipases are glycerides with low water-solubility [119].

Lipases are often used as catalysts in organic synthesis, especially for asymmetric synthesis. They are preferred above other enzymes because they are commercially available at a relatively affordable prize, and display a high stability and catalytic efficiency.

4.2 Intra-esterification of hydroxy fatty acids

In view of the production of sophorolipids, we will focus here on the use of lipases for *in vitro* intra-esterification reactions, with lactones as the resulting products.

Lactone rings composed of C-5 to C-6-containing organic acids can be formed spontaneously from the corresponding hydroxy fatty acids, while the formation of the macrolactones containing a larger fatty acid chain (C14-C18) is not straightforward. Indeed, the entropically disfavoured monomeric

macrolacton synthesis has to compete with the formation of linear oligomers and multimeric lactones like diolides (dimeric lactones). Macrolactones have applications in different areas. For instance a C-16 macrolactone, hexadecanolide, is a musk lactone (**Fig. 1.17A**), and is an important product for the fragrance industry. Another example is C- 14, 15 or 16 macrolactone, a key structure of the macrolide antibiotics produced by *Streptomyces erythreus* (**Fig. 1.17B**). Likewise, the lactonic form of the sophorolipids can be regarded as a macrolactone-like structure.

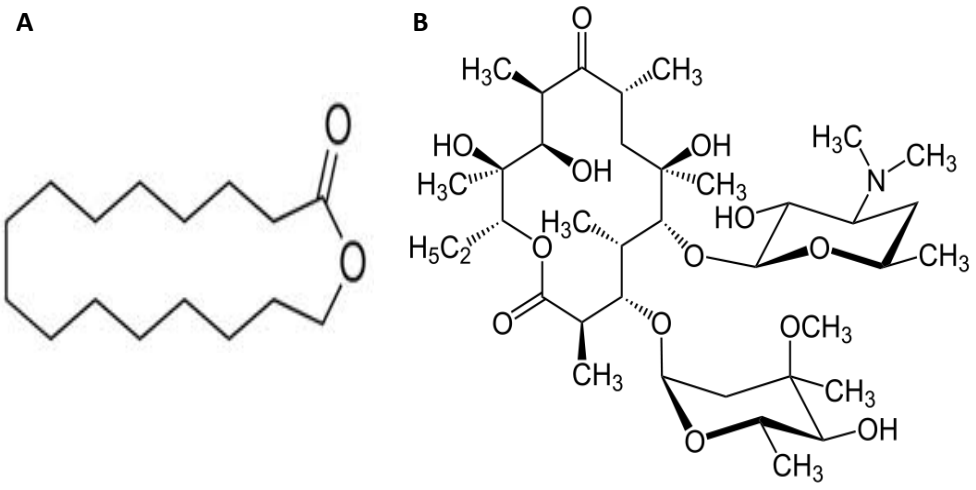


Figure 1.17. A. Hexadecanolide, a musk lactone (www.sigmaaldrich.com) B. Erythromycin, a macrolide antibiotic (http://www.epharmacognosy.com)

Several chemical methods were developed to produce macrolactones. They are mainly based on a weak activation of the acyl carbonyl group using specific reagents, such as 2,2'-dipyridyl disulphide [120], 2-chloro-1-methylpyridinium iodide [121], and carboxylic 2,4,6-trichlorobenzoic anhydride [122]. However, the chemical synthesis is still a complicated and expensive procedure that occurs under extreme conditions. This explains the great interest for biocatalysts for the industrial production of lactones. In the scientific literature, there are a few reports that describe the lipase-catalysed intra-esterification of industrial or medically relevant macrolactones.

The first group describing enzyme-catalysed lactonization consisted of Japanese scientists (in 1987) [123] who describe the lactonization reaction of the methyl ester of ω-hydroxy fatty acids in an organic solvent by a lipase. By screening several enzymes, they detected that only Lipase P from *Pseudomonas sp.* and porcine pancreatic lipase are able to catalyse this reaction with high efficiency. Therefore, they claimed that lactonization is not a common feature of lipases but it is rather a specific activity of some lipases. They speculated that special structural features are necessary to catalyse the synthesis of macrolactones. For this reaction, the best performing organic solvents appeared to be heptane and benzene. The lipase-catalysed lactonisation also depends on the chain length of the substrate. When the methyl ester of 16-hydroxyhexadecanoate or 15-hydroxypentadeconoate was used, about 80 m/m% of it was converted to lactone. With the methyl ester of 13-hydroxytrideconoate, the yield decreased to 38 m/m%, and a diolide (dilactone) was formed with a yield of 25 m/m% (**Fig. 1.19**). Dilactone formation decreased when the chain length was increased, and dropped to 3 m/m% with C-16 substrates. Consequently, it was suggested that

for efficient intramolecular transesterification, there need to be a defined distance between –COOH and a free -OH group in the substrate [123].

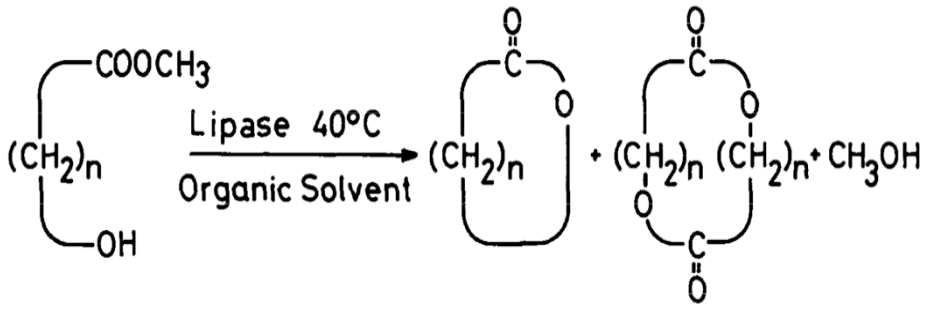


Figure 1.19. Lactone synthesis using 16-hydroxyhexadecanoate [123]

A year later, these results were improved substantially [124]. The authors showed that the lactonization product profile not only depends on the chain length of the substrate but also on the lipase itself and that the relation is more complex than originally proposed. Here, the methyl ester of 10-hydroxydecanoic acid was used as a substrate for lipase-catalysed lactonization in anhydrous isooctane. No monolactone was detected, but instead di-, tri-, tetra- and pentalactones were formed. Lipase from *Pseudomonas sp.* (AKand K-10), and porcine pancrease (PPL) created di- and trilactones as a major product, whereas lipases from *Candida cylindracea* and *Mucor meihei* give mostly tri- and tetralactones. The reactions catalysed by these enzymes appeared to be stereospecific, since for the tested lipases, mainly R-isomers were consumed.

More detailed work on the lactonizing lipase P from *Pseudomona sp*. was described in 1991 [125]. Different substrates for the lactonization and hydrolysis reactions were tested. For the lactonization, anhydrous conditions were required to produce intramolecular transesterification because the presence of water promotes hydrolysis. **Table 1.2** represents the product yield from the lactonization and hydrolysis reactions for various substrates with different chain lengths. For the lactonization reaction, it was confirmed that long chain methyl esters of ω-hydroxy acids are preferable for monolactone formation while shorter chains lead to dilactone formation. To evaluate hydrolysis activity, different p-nitrophenyl esters and triglycerides were tested. The results were surprising because lactonizing lipases appeared to have a preference for p-nitrophenyl esters and triglycerides from C-8 acyl or shorter acyl chains. This suggests that the catalytic site can host octanoyl or shorter acyl groups. To create a monolactone, a hydroxy group should be within short distance to the acyl carboxyl group. The authors claimed that for lactone formation, the chain length should be doubled compared to what is required for hydrolysis. They proposed a mechanism in which, after formation of the enzyme-acyl intermediate, the C-8 portion near the acyl moiety should stay in close contact with the enzyme while the rest of the molecule, containing the hydroxyl group, can move freely and fold back towards the acyl carbonyl function and attack it to form a lactone. The shorter the substrate, the more difficult it is for the –OH group to attack the acyl carbonyl group intramolecularly, and the probability that the –OH group attacks the carboxyl group from a neighbouring substrate increases, the latter resulting in the increase of dilactone formation [125].

Table 1.2. Results of the lactonization activity of lipase P towards various methyl ester of the ω-hydroxyacids in an organic solvent (hexane, 40°C, substrate conc. 1 mM) and hydrolysing activity of lipase towards p-nitrophenyl esters and triglycerides [125]. The highest product yields are indicated in bold. *Product yield was determinated by gas chromatography as described in [123].

Lactonizing activity			**Hydrolysing activity**		
Substrate HO-$(CH_2)_n$-$COOCH_3$	Product yield*		Substrate chain length for the hydrolysis reaction	Product yield	
	Monolactone m/m[%]	Diolide m/m[%]		p-Nitrophenol m/m[%]	Free fatty acids m/m [%]
n=10	0	**100**	n=0	3.4	49.2
n=11	ND	**80**	n=2	18.2	92.9
n=12	12	**30**	n=4	29.5	96
n=13	20	11	n=6	100	100
n=14	25	ND	n=8	24.3	46.4
n=15	25	4	n=10	21.7	12.7
n=16	**57**	0.1	n=12	13.6	4.3
n=17	**100**	0.5	n=14	12.1	7.4
n=18	**80**	0	n=16	1.1	10.2
n=20	38	0			

The second lipase which was proven to produce hexadecanolide, the musk lactone, from 16-hydroxyhexadecanoic acid (juniperic acid) was lipase B from *Pseudozyma (Candida) antarctica* [126]. *P. antarctica* lipase can catalyse highly enantioselective transformations of many chiral alcohols in esterification, transesterification and hydrolysis reactions [127]. With this enzyme no formation of di-, tri- or tertralactones were noticed when a high substrate concentration was used and the maximal yield was 80 m/m% monolactone. Moreover, product inhibition was observed only by monolactones whereas for other enzymes, inhibition was mainly due to oligomeric lactones [126]. Another study where *P. antarctica* lipase B (CALB) was used, resulted in macrolactone synthesis by lipase in a water-in-oil microemulsion [128]. It was suggested that lactonization may be promoted by improving the competition with the interesterification reaction by employing a solution of reverse micelles (w/o microemulsion droplets). This way, the substrate/droplet occupancy level is kept below unity. It is expected that the reaction takes place at the droplet interface where the lipase is concentrated and activated while the substrate is located in the continuous oil phase. According to the hypothesis of Jaeger and Ippoliti [129], the partitioning of the substrate will increase the yield of the reaction even when substrate concentration exceeds droplet concentration. However, this hypothesis could not be confirmed and a decrease in lactone yield was observed when the substrate concentration was higher than the droplet concentration. The authors suspected that the proposed hypothesis is too simple and does not take into account that lipase might “create” its own microemulsion droplet environment.

P. antarctica N-435 lipase activity was also described to create a new type of macrolactone [130]. This lipase catalysed the lactonization of (+)-coriolic acid to a 13 carbon-ring with a 5 carbon- tail (**Fig. 1.20**). The production of this macrolactone using this biocatalyst was more efficient than using a complicated chemical method. In this study, the enzyme performed well in many solvents but the best result was obtained using diisopropyl-ether and toluene. It was suggested that polar organic

solvents are unfavourable to maintain protein stability and probably act by distorting the essential water layer. With a small concentration of the substrate, less than 1 mM, a 50 m/m% higher product yield achieved. In contrast, at high substrate concentration (20 mM) a small amount of polyester was detected (with a MW of around 4000). So increasing the substrate concentration creates competition between the reactive groups of different substrate molecules, near the enzyme active site, which is unfavourable for an intramolecular reaction between the –OH and -COOH [130].

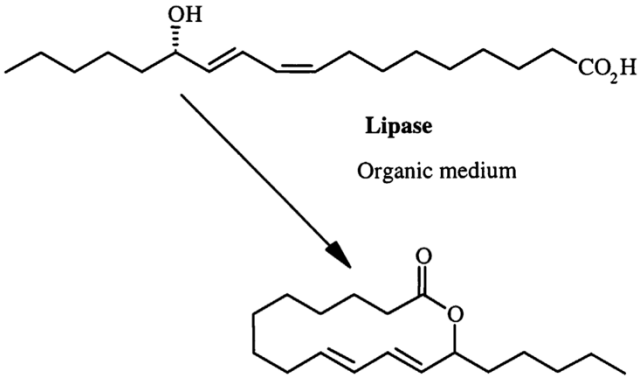


Figure 1.20. The enzymatic synthesis of the lactone form (+)-coriolic acid in an organic medium [130].

Lastly, the enzyme Novozyme 435 was used to lactonize methyl esters of acidic sophorolipids *in vitro* in organic THF solvents at 30°C (see **Fig. 1.4**). Surprisingly, the resulting product was an unnatural compound, since the ester bond was formed between the carboxyl group and the 6'' –OH group instead of the 4'' -OH of sophorose [45].

5 Conclusion

In recent years many studies addressing sophorolipid production, structure, modification, and industrial applications were performed. This indicates that biosurfactants are a popular and important subject of study, especially in the epoch of a growing awareness for the environment. However, none of these studies addressed the proteomic changes that coincide with the biosynthesis of those molecules. Nowadays, the increased accessibility of genome sequencing data and widely developed proteomic strategies allow detailed and rapid data generation. This type of information can be used to design biotechnological processes. In the case of *S. bombicola* they can be important for improving the production of natural sophorolipids, to have control over the SL structures, or to secrete a new-to-nature molecule. In the following chapters an investigation of the *S. bombicola* proteome and the exoproteome will be presented.

Aim of the study

This thesis was performed in the framework of two research projects, *i.e.* the IWT-SBO 'BIOSURF' project and the European FP7-project 'BIOSURFING'. These two projects aimed to create a platform for producing new-to-nature biosurfactants using the fungus *Starmerella bombicola* as host organism. This strain was selected because it naturally produces high amounts of a type of biosurfactants termed sophorolipids (SL). To have full control over SL production in an industrial setting, the biochemistry underlying SL-biosynthesis needs to be completely understood. Therefore, those projects applied a multidisciplinary approach, combining molecular biology, transcriptomics and proteomics. The proteomic experiments conducted in this framework form the subject of this doctoral thesis and in the following chapters the results of different experimental setups are described.

SL are synthesized from two fundamental building blocks: glucose and fatty acids which can be easily added to the growth medium [40]. However, in the absence of a hydrophobic substrate, SL production still occurs. Besides activating glucose molecules to UDP-glucose, *S. bombicola* creates *de novo* fatty acids to integrate them into SL. As we focused on *de novo* formation of the SL, we were interested in all molecular players connected to the proposed SL biosynthesis pathway, either upstream or downstream.

In the **second chapter** of this thesis we describe the analysis of the intracellular proteome of *S. bombicola*. Using a SILAC-based proteomic experiment, we compared the *S. bombicola* proteome under sophorolipid producing and non-producing conditions. These proteomic studies aimed to confirm co-expression of the genes of the SL biosynthesis cluster and possibly discover new participants in the SL-production process. We also hoped to provide more details about the regulation of the SL gene cluster. All that information could be useful to better understand the mechanism of SL biosynthesis and therefore guide the engineering of *S. bombicola* to create new-to-nature molecules. Our work was supported by quantitative transcriptomic data generated by the group of Prof. Van de Peer. Moreover, their effort to sequence and annotate the *S. bombicola* genome was essential to interpret our mass spectrometric data. In this chapter, we also determined the glucose, ammonium and phosphate concentrations during the transition from the exponential to the stationary growth phase in order to confirm previous reports claiming that SL production is associated with nutrient starvation [131].

Sophorolipids can be present in an open, acidic form or a closed, lactonic form. Until now, it was not clear if SL lactonization is an enzymatic reaction or a spontaneous process. However, it was proposed that a cell wall-bound or extracellular lipase is responsible for SL lactone formation in *Candida apicola* [40]. In addition, lactonization of acidic unacetylated SL was obtained *in vitro* by extracellular lipase B from *Pseudozyma antarctica* (former *Candida antarctica*) in an anhydrous tetrahydrofuran environment [45]. In **chapter 3**, we focused on the extracellular proteome analysis of *S. bombicola* in order to confirm the existence of a protein responsible for sophorolipid lactonization. We further validated our findings by creating a knock out mutant for the proposed enzyme, lactone esterase. With the help of the group of Dr. Van Bogaert and Prof. Soetaert, HPLC and LC-MS analyses on the SL mixtures were used to monitor the production of acidic or lactonic SL. We also investigated the influence of sophorolipids on the exoproteome content. Therefore, the analyzed *S. bombicola* exoproteome was compared with published data on the secretome of other

fungi. We also performed semiquantitative analyses of protein abundances during the transition from the exponential to stationary phase.

In **chapter 4** we further characterized the lactone esterase (SBLE) described in chapter 3. We created an enzymatic assay to study the substrate preference and temperature and pH dependence. Moreover, a preliminairy kinetic study was performed to estimate its K_m and V_{max}. In addition, we performed site-directed mutagenesis to validate the activity of SBLE.

Chapter 2

SILAC-based proteome analysis of *Starmerella bombicola* sophorolipid production

Ciesielska K.[a]; Li B.[b]; Groeneboer S.[a]; Van Bogaert, I.[c]; Lin Y.-C.[b]; Soetaert, W.[c]; Van de Peer Y.[b]; Devreese B.[a]

[a] Laboratory for Protein Biochemistry and Biomolecular Engineering, Gent University

[b] VIB Department of Plant Systems Biology and Department of Plant Biotechnology and Bioinformatics, Gent University

[c] Laboratory of Industrial Biotechnology and Biocatalysis , *Gent University*

This chapter is in accepted for publication in the 'Journal of proteome research'

Contribution: K. Ciesielska engineered the *S. bombicola* SILAC mutant and generated and analyzed proteomic data. B. Li created a *S. bombicola* proteomic database and generated transcriptomic data. The phosphate, ammonium and glucose concentration determination took place in the Laboratory of Industrial Biotechnology and Biocatalysis.

Abstract

Starmerella (Candida) bombicola is the biosurfactant-producing species, which caught the greatest deal of attention in the academic and industrial world due to its ability of producing large amounts of sophorolipids. Despite its high economic potential, the biochemistry behind the sophorolipid biosynthesis is still poorly understood. Here, we present the first proteomic characterization of *S. bombicola*. In order to investigate the processes behind the production of these biosurfactants, the proteome of sophorolipid producing (early stationary phase) and non-producing cells (exponential phase) was compared. The *lys1*Δ mutant was created which allowed the use of SILAC for quantitative analysis. We report simultaneous production of all known enzymes involved in sophorolipid biosynthesis including the predicted sophorolipid transporter. In addition, the heme binding protein Dap1, a possible regulator for Cyp52M1, was identified. Our results, supported by mRNAseq analysis, further indicate that ammonium and phosphate limitation are not the only factors inducing sophorolipid biosynthesis.

Keywords: *Starmerella bombicola*, sophorolipid biosynthesis, yeast, quantitative proteomics, SILAC, mRNAseq

1 Introduction

Driven by an augmented awareness about the environmental effects of synthetic detergents, the number of applications of biosurfactants has grown steadily in recent years. These natural detergents not only possess interesting physicochemical properties but above all they are biodegradable and display low toxicity. Several microorganisms that produce glycolipids with surfactant properties have been identified. Among the best described organisms are the rhamnolipid-producing bacterium *Pseudomonas aeruginosa* [132] and the cellobiose lipids producing fungi *Pseudozyma flocculosa* [52] and *Ustilago maydis*, the latter also produces mannosylerythritol lipids [51]. However, the biosurfactant-producing species that caught the greatest deal of attention in the academic and industrial world is *Starmerella bombicola* (previously named *Candida bombicola*) [23]. Using renewable resources [36, 133], this non-pathogenic yeast is able to produce large amounts of sophorolipids (SL) during its stationary growth phase. Sophorolipids produced by this species are glycolipids composed of a sophorose-head of which the reducing end is connected through a β-glycosidic bond to a terminally or subterminally hydroxylated C16 or C18 fatty acid. In addition they can be decorated by acetyl groups at the 6 and/or 6' position of the sophorose group (**Fig. 2.1**). Sophorolipids are mainly secreted as ring-closed lactonic molecules but can also appear in an open acidic form, which has different physicochemical properties. For example, lactonic forms are more potent in decreasing the surface tension and have stronger antimicrobial properties than their acidic counterparts, which on their turn are better water soluble [12]. There already exists a number of applications for sophorolipids in the industry, as they are included as active agents in a number of cleaning products, cosmetics and personal care products [11].

Despite their economic potential, the biochemistry behind the sophorolipid biosynthesis is still not completely understood. In 2007, Van Bogaert *et al.* [11] proposed a pathway for their production starting from fatty acids and UDP-glucose building blocks that depends on the subsequent actions of five different enzymes: a microsomal cytochrome p450 (Cyp52M1), two glucosyltransferases (UgtA1 and UgtB1), an acetyltransferase (AT) and, possibly, a lactone esterase (**Fig. 2.1**). During the past few years, molecular biology experiments allowed to identify these enzymes in *S. bombicola* ATCC 22214 except for the lactone esterase [46-49]. Recently, it has been shown that the CYP52M1, UGTA1, UGTB1 and AT genes are clustered with a multidrug resistance gene (MDR) that is required for extracellular transport [53]. This suggests that the genes are possibly co-regulated. No gene for a lactone esterase was discovered.

Here, we present a proteomic study aiming to confirm the coordinated production of the enzymes involved in sophorolipid biosynthesis and, more importantly, to identify additional proteins connected to its metabolism. Understanding cell metabolism during sophorolipid production is necessary to allow engineering of the biosynthetic pathway, e.g. in order to create new-to-nature molecules. Therefore, we set up a proteome analysis experiment based on stable isotope labelled amino acids (SILAC) using an engineered lysine biosynthesis deficient *S. bombicola* strain. We performed a quantitative comparison between cells grown to the exponential and the early stationary phase, respectively representing a sophorolipid non-producing and producing state. Since sophorolipid biosynthesis and metabolism is most likely membrane linked, we used extracts

enriched in hydrophobic proteins. SILAC data were verified using results from an mRNAseq transcriptome analysis. This work represents the first proteomic analysis on sophorolipid producing *S. bombicola* and offers a platform for further investigating on the biology of this organism of biotechnological interest.

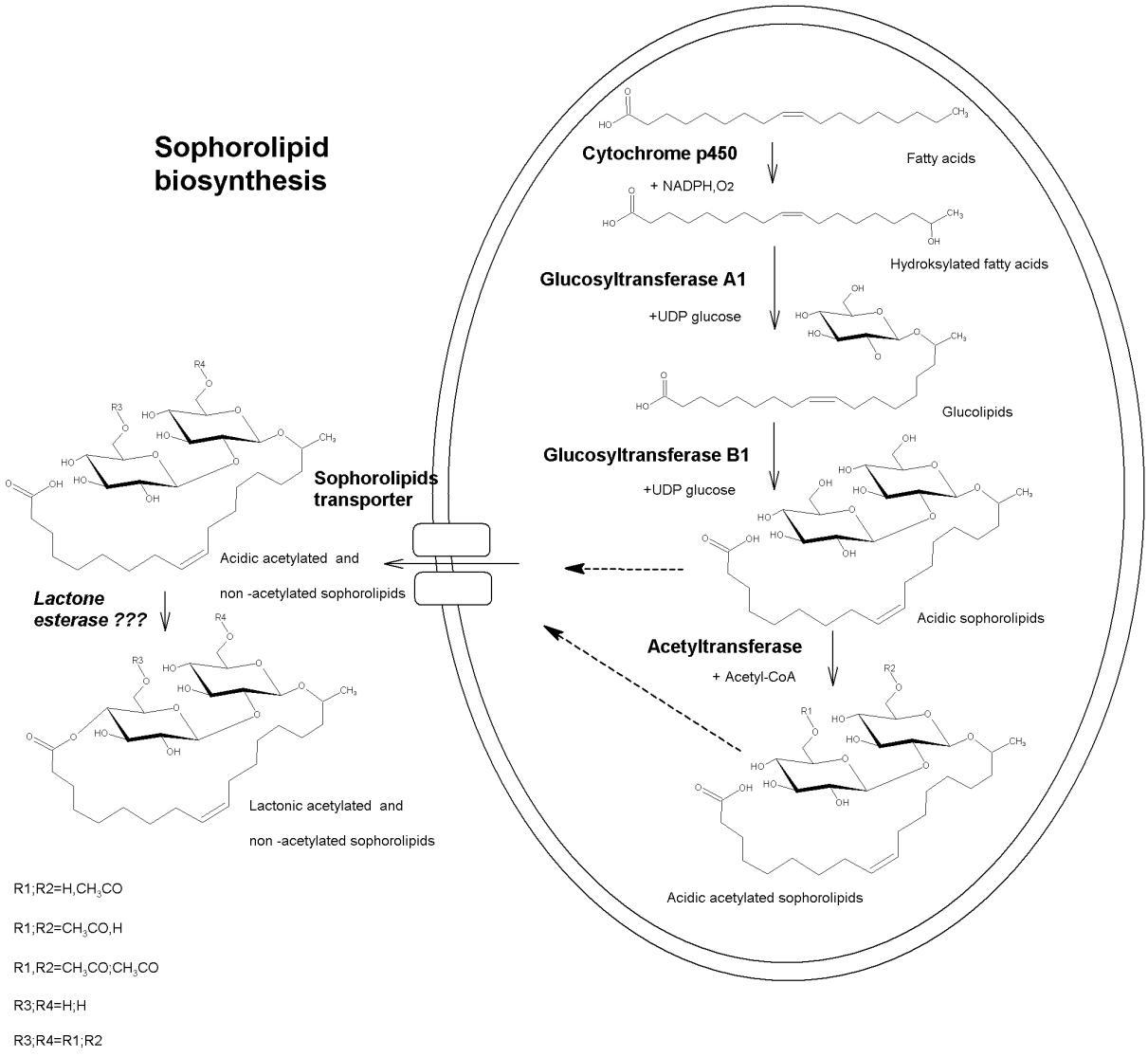


Figure 2.1. Sophorolipid biosynthesis in *S. bombicola* proposed by Van Bogaert [12]. From the predicted enzymes only lactone esterase is not yet identified.

2 Materials and methods

2.1 Materials

Materials for fungal growth, SILAC medium including $^{12}C_6$ L-Lysine- and $^{13}C_6$ L-Lysine(isotopical purity: 99 atom % $^{13}C_6$), for SDS-PAGE analysis and LysC digestion were purchased from Sigma-Aldrich (USA) except for glucose, trisodium citrate and sodium chlorate that were ordered from Merck Millipore (Germany). Yeast nitrogen base medium without amino acids was from ForMedium

(UK), acrylamide from National Diagnostics (USA) and enzyme endoproteinase LysC from Wako Chemicals USA, VA. LC-MS grade acetonitrile and formic acid were bought from Biosolve (The Netherlands).

2.2 Creation of *lys1*Δ mutant

To perform a SILAC analysis in *S. bombicola*, we needed to create a lysine auxotrophic strain. Therefore, we generated a knock-out strain defective for the final step in lysine biosynthesis, *i.e.* for saccharopine dehydrogenase (Lys*1*) [134]. The LYS1 gene replacement procedure was based on a one-step gene disruption with a linear fragment of DNA, containing the *URA3* marker flanked by a 5' and 3' homologous regions (1000 bp) of the target gene (**Fig. 2.2**). For this, the *LYS1* gene and its flanking region were amplified and ligated into the TOPO XL (Invitrogen) vector. After restriction digestion, cutting both in the *LYS1* coding region, the functional orotidine 5'-phosphate decarboxylase (*URA3*) gene was inserted in the *LYS1* region, resulting in a TOPO XL (Invitrogen) plasmid harbouring the knock-out cassette. Starting from this vector, a linear fragment was created with Lys1For (CACGGAGGCAGTTTCGTCAAGGAT) and Lys1Rev (AACGTGCGCGATGTGACGACAA). This cassette was purified and used to transform a *URA3* auxotrophic (*ura3*Δ) *S. bombicola* strain as described previously [135]. Transformants were selected on a modified SD plate containing $^{12}C_6$-Lys (**Suppl. Table 2.1**). On SD plates without lysine, no transformants were observed.

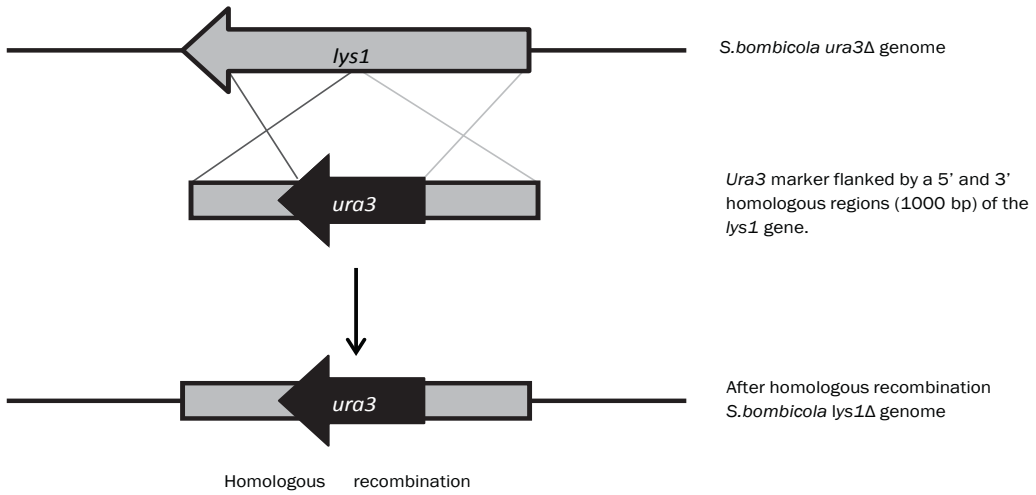


Figure 2.2. In order to develop a SILAC strategy in *S. bombicola* a lysine auxotrophic strain was created. The *lys1* gene was interupted by the homologous recombination with the cassette containing the *ura3* marker.

2.3 Growth conditions

Lang *et al*. described a medium that is optimized for sophorolipid production by *S. bombicola* [136]. For the SILAC experiment this medium was adapted by removing any source of lysine. Therefore, yeast extract was replaced by Yeast Nitrogen Base (YNB) without amino acids to which a mixture of 0.3 g/L Met, 0.15 g/L His, 0.3 g/L Trp and 0.2 g/L isotopically enriched Lys was added (**Suppl. Table 2.2**). To achieve complete integration of isotopically labelled Lys, the mutant was first grown 48 h on

an agar plate containing the modified SD medium and then a single colony was used to inoculate 5 ml modified Lang medium. Overnight grown precultures at OD_{600} = 0.8-0.9 were used to start cultures with the same amount of cells ($3x10^6$ cells/ml). Growth took place at 28°C in a rotating shaker at 130 rpm in 500 ml flasks containing 75 ml of culture. The SILAC experimental set up consisted of yeast grown either on "heavy" $^{13}C_6$-Lys or "light" $^{12}C_6$-Lys. Three biological replicas were performed. In two replicate experiments A and B, cells grown to the exponential phase (OD_{600}= 10, at 35h) were labelled with "heavy" Lys (H) containing medium while cells grown to the stationary phase (OD_{600}= 30, at 75 h) were labelled with "light" Lys medium (L). In the third replica, experiment Rev, the labelling was reverted.

2.4 Sophorolipid extraction

Thin layer chromatography (TLC) was used to test if sophorolipids were present in the samples. To extract sophorolipids, 1 ml of culture was mixed with 440 µL ethyl acetate and 11 µL acetic acid and shaken vigorously for 5 min. After centrifugation at 9 000 g for 5 min, the upper solvent layer was removed and transferred into a fresh eppendorf tube. 10 µl of sample was spotted on a silica plate (Kieselgel 60_{F254}) and separated using a mixture composed of chloroform/methanol/water (ratio 65/15/2) [9]. After 1h of solvent migration, the TLC plate was visualized by spraying a mixture of acetic acid, sulfuric acid and anisaldehyde (ratio 100/2/1), followed by heating.

2.5 Protein extraction and separation

Cell pellets from sophorolipid non-producing (exponential phase) and producing (early stationary phase) cultures were resuspended in water and mixed in a 1:1 ratio according to the OD (**Fig. 2.3**). Next, this mixture of cells was used to extract proteins using the Mem-PER® Eukaryotic Membrane Protein Extraction Reagent Kit (Thermo Scientific, Pierce Biotechnology, USA). The hydrophobic protein fraction from the mixed cells from both growth conditions was concentrated by acetone precipitation. Then, protein pellets were dissolved in 2 m/v% SDS and the protein concentration was measured using the Thermo Scientific Pierce BCA Protein Assay Kit. Hundred microgram of protein was loaded on a 12.5 m/v% SDS-PAGE gel and electrophoresis was performed at 110V. The gel was stained overnight with Coomassie Brilliant Blue G250 and consequently destained in 30 v/v% MeOH.

2.6 Digestion with endoproteinase Lys-C

The total SDS-PAGE gel lane was sliced into 17 fragments. The gel pieces were destained by three 20 min incubation cycles at 30°C with 150 µl of a 50 v/v% acetonitrile (ACN)/ 200mM NH_4HCO_3 mixture. The proteins were digested in-gel by adding 8 µl of 0.002 µg/µl LysC endoproteinase (Wako Chemicals USA, VA) in 50mM NH_4HCO_3 and incubating overnight at 37°C. The resulting peptides were collected from the supernatant. More peptides were extracted from the gel piece with 60 µl of 60 v/v% ACN/0.1 v/v% formic acid (FA) during 20 min incubation at 30°C. Next, the samples were vortexed for 3 min, shortly centrifuged and extraction was repeated with 20 µl of 60 v/v%ACN/0.1 v/v% FA. These extracts were pooled with the supernatant and the resulting peptide mixture was dried under vacuum centrifugation and dissolved in 15 µl of 2 v/v% ACN/0.1 v/v% FA.

2.7 NanoLC-ESI-FT MS analysis

Five µl of the extracted peptides were loaded on a Zorbax 300SB-C18 trap column 5mm x 0.3mm and separated by a Zorbax 300SB-C18 analytical column 150 mm x 75 µm (Agilent) connected to an Agilent 1200 chromatographic system (Agilent, Santa Clara, CA, USA). The LC system was coupled to an LTQ-FT Ultra mass spectrometer (Thermo Fisher Scientic, Waltham, MA) as previously described [137]. The separation was performed by reversed phase chromatography using a 50 min linear gradient ranging from 2 v/v% acetonitrile to 80 v/v% acetonitrile in 0.1 v/v% formic acid/water at a 300 nl/min flow rate. The LC eluent was directly introduced to a Triversa NanoMate ESI source (Advion, Ithaca, NY), working in the nanoLC mode and equipped with D-chips to which a 1.55 kV voltage was applied.

The FT-ICR mass analyzer acquired MS scans (3 s) at a resolution of 100,000 during the LC separation. The three most intense precursor peptides (minimal intensity : 500) from each MS scan were automatically selected and fragmented in the LTQ ion trap mass analyser which was also used to collect MS/MS spectra. Mass range for precursor ion scans was from m/Z 300 to 1400, charge state 1 was rejected. The isolation width was 2 m/Z. The normalized collision energy was set at 35 with an activation Q of 0.25 and an activation time of 30 ms. The dynamic exclusion parameters were as follows : repeat count 2, repeat duration 30s, exclusion duration 90s with an exclusion mass width of ±5 ppm.

2.8 Protein sequence database generation

The genome sequencing and assembly was outsourced to LGC Genomics (Berlin). It was performed on a 454 whole-genome shotgun (WGS) sequencing and assembler platform, resulting in 310 contigs. After the gap closure by primer walking on fosmid clones between these contigs, the 310 contigs were further assembled into 3 chromosomes, with a length of 2.91 Mbp, 3.08 Mbp and 3.50 Mbp respectively. From the 3 chromosomes, 4617 genes were predicted using the AUGUSTUS [138] gene prediction program (unpublished data). Functional annotation of predicted genes was performed by Blast [139] analysis against the Swiss Prot database (version 2012/06) [140]. The protein domain information and Gene Ontology terms was obtained through InterProScan (InterPro V.34) [141]. The functional annotation of *S. bombicola* genes were transferred from the best hit of the BLAST search result and the genome annotation is stored in an ORCAE database [142].

2.9 Analysis of SILAC datasets using MaxQuant

Proteomic data analysis was performed using the MaxQuant package [143] version 1.1.1.36. Protein identification was performed using the Andromeda search engine against our preliminary annotated *S. bombicola* genome database merged with a decoy (reverse) database and contaminants list. As variable modification, methionine oxidation was selected. Endoproteinase LysC was specified as the enzyme with up to two missed cleavages allowed. Quantification analysis for SILAC experiments was performed using MaxQuant (Quant). Peptides were filtered for the presence of maximum 3 Lys per peptide. Peptide and protein false discovery rates were set to 0.01. The maximum posterior error probability was set to 1 and 6 amino acids were required as the minimum peptide length. Unique and razor peptides were both considered for MS quantification and minimum 2 SILAC peptide ratios

were counted to calculate the SILAC protein ratio. An experimental design template was created in which the three SILAC replicate experiments were combined (A,B and Rev). More details about the parameters can be found in **Suppl. Table 2.3.** MaxQuant automatically determined SILAC peptide and protein ratios. Values for heavy to light ratios (H/L) at the protein level were calculated as the median of all measured peptide ratios for a given protein. Proteomic data from Maxquant were statistically analysed using the Perseus software (version 1.2.0.17.). Here, the data were first processed, removing contaminants and decoy hits. The ratios between stationary and exponential phase (L/H) from the reversed replicate experiment (Rev) were inverted to H/L. After calculating the log2 of normalized H/L ratios for all replicas, the proteins quantified in at least 2 of 3 replicas were retained for further analysis. A Student's t-test was performed on the log2 values of H/L ratios at the 95% confidential level ($p<0.05$) to indicate a significant protein abundance difference between the two conditions. In-depth analysis of the function of proteins was performed when the log2 fold change was equal or higher than 1.5.

2.10 RNA-seq analysis

S. bombicola WT cells were collected precisely at the same time points as for the proteomic experiment. From each phase 20 mL culture was centrifuged for 5 min at 4000 g. Next, pellets were suspended in 15 mL of RNase free water and lyophilized. The lyophilized powder was crushed in liquid nitrogen and RNA was extracted with the RNeasy Midi kit (Qiagen). The single-end, stranded shotgun sequencing of transcriptome (RNA-seq) was performed by FASTERIS SA (Switzerland) according to Illumina library preparation protocol and kit. Briefly, two micrograms of total RNA was treated with DNAseq to remove any DNA contamination and was enriched for poly-A tailed mRNA using magnetic beads. The strand specific 3' and 5' adapter were ligated with transcripts respectively. RNA fragments were converted to cDNA by the reverse transcriptase and amplified by random priming. The amplified forward strand cDNA with insert size ~200bp were used to prepare the Illumina sequencing library. High-throughput sequencing of the cDNA library (RNA-seq) was performed on an Illumina HiSeq 2000 platform.

The first 10 bases and low quality bases with PHRED score < 20 were removed by FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). RNA-sequence reads were mapped onto predicted genes using Tophat 1.3.1. [144]. The abundance of gene expression was reported in FPKM units (fragments per kilobase of exon per million fragments mapped) by Cufflinks 1.3.0. [145]. The gene expression from exponential to stationary phase (H/L) was compared and their relative expression values were normalized by log2 transformation. A Multiple Testing Correction was performed and significant genes were selected based on the q-value. The 615 genes, which were quantified in at least 2 of 3 replicas in proteomic experiment, were selected and used for further investigation.

2.11 Data mining with Blast2GO

The group of 615 proteins quantified in at least 2 of 3 replicas in the LC-MS analyses, was analysed with the Blast2GO 2.6.2 program [146] to examine the overrepresented Gene Ontology terms. As a reference set, a GO term annotated *S. bombicola* genome was used. It was generated using IntraProScan (V.34) as described in 2.8. The Fisher's exact test and Enrichment test with Multiple Testing Correction of FDR 0.05 (Benjamini and Hochberg) were used.

2.12 Glucose, ammonium and phosphate measurements

To measure nutrient concentration during transition from exponential to stationary phase the *S. bombicola lys1*Δ was cultured in modified Lang medium, using the same setup as for the proteomic experiment. One ml of culture was collected at different time points between 0-96h of growth. Next, after centrifugation of 1 mL of culture for 5 min at 9000 g, the supernatant was used to detect the level of 3 components and the pH. The ammonium level was measured with the Ammonia (Rapid) kit (Megazyme), while the phosphate content was analysed using the Gawronski method [147]. Additionally, the concentration of glucose in the supernatant was determinate using a YSI 2700 SELECT biochemical analyser. The measurements were performed for three biological replicates.

3 Results

3.1 Construction of a SILAC compatible *S. bombicola lys1* Δ mutant

We report here the creation of a SILAC-compatible mutant of *S. bombicola*. A lysine auxotrophic *S. bombicola* strain was created by replacing the *LYS1* gene with a linear fragment of DNA containing the *URA3* marker flanked by a 5' and 3' homologous regions (1000 bp) of the *LYS1* gene. The *S. bombicola lys1Δ* strain showed no growth on media lacking lysine. Furthermore, mass spectral analysis of peptides obtained from a protein extract submitted to endoprotease Lys-C showed complete incorporation of the heavy labelled $^{13}C_6$-lysine when cells were grown on SILAC medium (data not shown). This ensures that all peptides, digested with LysC, will possess a predicted mass shift during LC-MSMS measurement. Analysis of the auxotroph growth showed that there were no differences compared to the wild type under our experimental conditions. The sophorolipid pattern analysed by TLC was also identical to the wild type (**Suppl. Fig. 2.1.**). For quantitative proteomic analysis, two OD points were selected which differ in sophorolipids production and which are separated by a short timespan to avoid the proteolytic activity in the later stationary phase.

3.2 Culture conditions and determination of sampling time

In order to analyse the proteome of *S. bombicola,* with a focus onto the proteins involved in sophorolipid biosynthesis, we aimed a comparison between samples from cultures at producing and non-producing conditions. We performed a growth curve measurement together with TLC analysis of the cultures to monitor sophorolipid production. It was confirmed that sophorolipid biosynthesis takes place when the stationary phase is approached [10]. As shown in **Fig. 2.3**, there is no production in the beginning of the exponential phase at OD_{600}=10. At OD_{600}=30, in the beginning of stationary growth phase, sophorolipids are detected in the medium. Therefore, we decided to take samples for comparative proteomic analysis at these two time points.

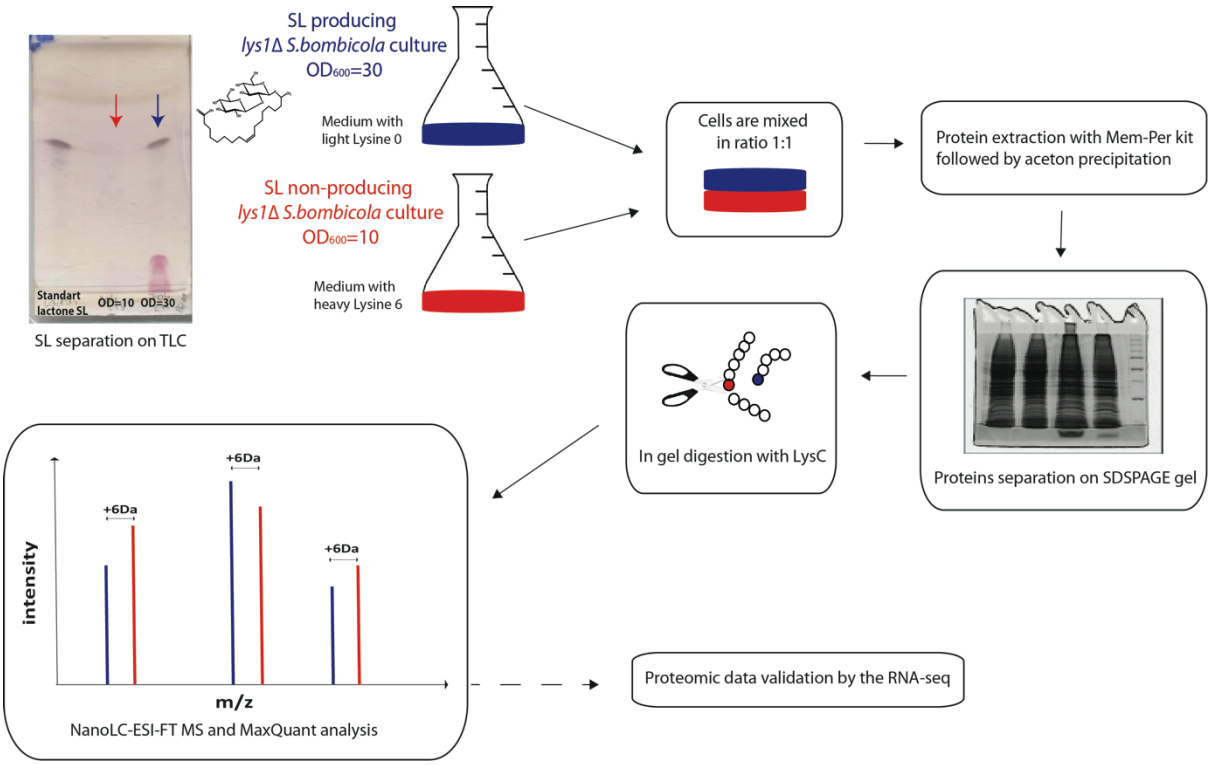


Figure 2.3. Workflow of the SILAC experiment.

3.3 General properties of the datasets: gene ontology analysis and ratio distribution

The difference between exponential and early stationary phase cells was analysed by quantitative proteomics using metabolic labelling and validated by transcript quantification using mRNA-sequencing. SILAC-data acquired from the three replicate GeLC-MS experiments were processed using MaxQuant and Perseus software. For further analysis, we only considered 615 proteins for which the SILAC ratio was obtained in at least 2 out of 3 replicate experiments. The correlation coefficient between these experiments is on average 0.7. The corresponding data are available in **Suppl. Fig. 2.2.**

We submitted the list of those 615 proteins for a Gene Ontology analysis using Blast2GO. The Fisher exact test was carried out to detect which GO terms are upgraded in the test group. Next, an enrichment test with FDR 0.05 (Benjamini and Hochberg) was performed to select the most specific GO terms (**Fig. 2.4**). The largest amount of proteins taken for quantitative analysis is involved in the energy metabolic processes and the translation machinery. Typically, fewer proteins were found in the Gene Ontology groups relating to transcription and other DNA related processes. This can be explained in part by the fact that we enriched our samples in more hydrophobic proteins, thereby removing polar DNA-binding proteins. No strong enrichment in membrane proteins is observed from the GO terms analysis. Only 97 out of 615 proteins have at least one transmembrane domain as predicted by the Phobius version 1.05 algorithm. This reflects the fact that the kit used in this study should be rather regarded as enriching in hydrophobic proteins than as a true membrane protein enrichment tool.

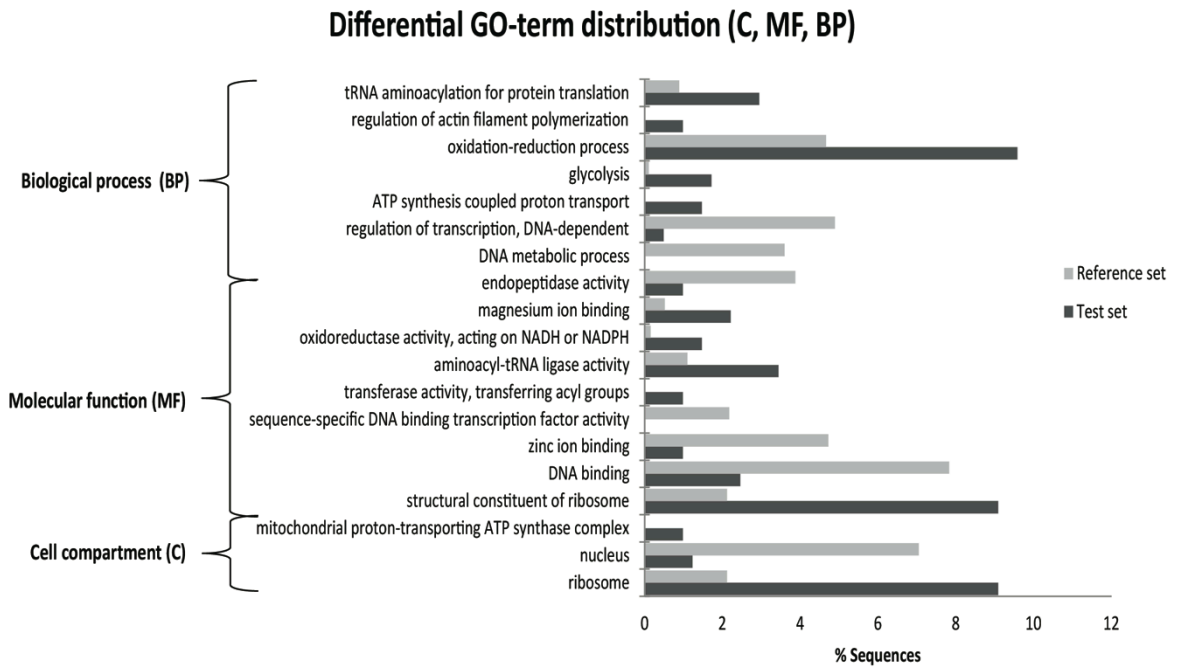


Figure 2.4. Graph represents the most specific GO protein classes enriched in proteins quantified in 2 of 3 replicas of SILAC experiment. From the graph we can read the level of enrichment expressed in % sequence in the test and the reference set.

In **Fig. 2.5** the distribution of the log2 SILAC protein ratios obtained from these 615 proteins and log2 FPKM ratios for the corresponding genes are presented. The distribution of mRNA and protein expression ratios is centred around a log2 value of 0, which reflect that expression of most genes does not change during transition to the stationary phase. However in the protein abundance ratio, the distribution is slightly shifted towards higher log2 ratios which reflect underrepresentation in the stationary phase. This can be consistent with the higher metabolic activity of exponential phase cells.

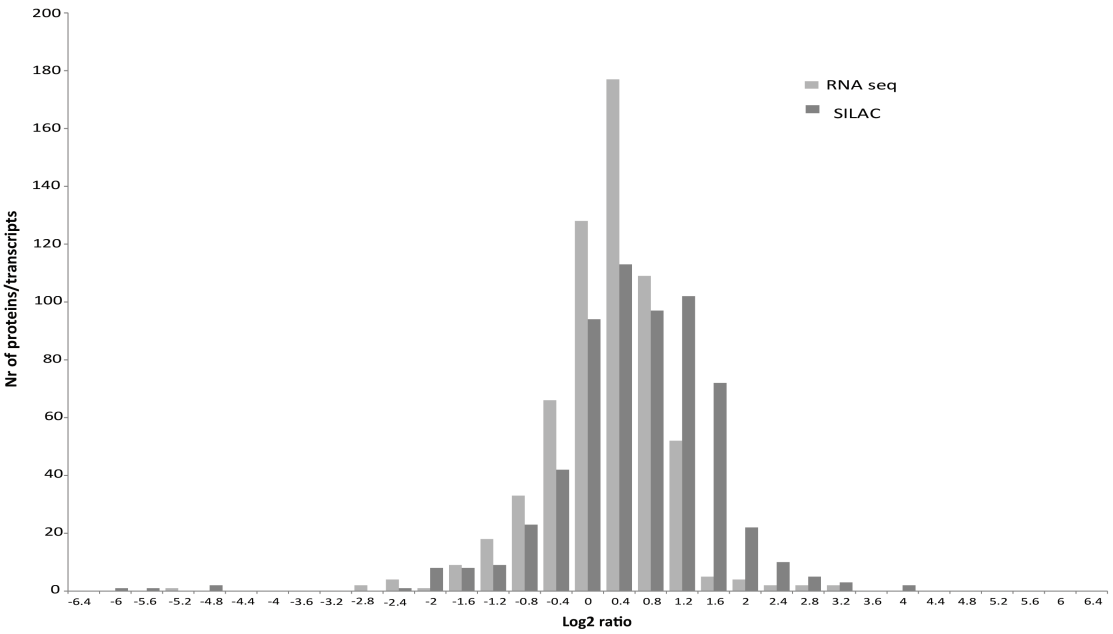


Figure 2.5. Log2 distribution of SILAC and FPKM ratios for proteins quantified in at least 2 of 3 replicas. The natural logarithm of SILAC and FPKM ratios was scaled into units of 0.4, and the total number of proteins and transcript in each bin was plotted against the ratio.

3.4 Protein quantification

The abundance ratios of the 615 proteins, for which a SILAC ratio was obtained in at least 2 out of the 3 replicate experiments, were statistically analysed using a t-test with p-value < 0.05. From the 147 proteins that passed the t-test (the details about the identification and quantification can be found in **Suppl. data**, Proteomics data), 89 displays at least a 2-fold change, the large majority (70) of those are more abundant in the exponential phase. 42 more proteins display at least a 1.5-fold change. These results are displayed in **Table 2.1** where significantly regulated proteins are grouped according to the process in which they are involved. Proteins that decreased in abundance in the early stationary phase are mainly involved in translation, amino acids metabolism, protein folding and stabilization and sterol biosynthesis. In contrast, the up-regulated proteins are involved in sophorolipid biosynthesis, tricarboxylic acid cycle and related processes, stress response and in cytoskeleton formation.

Table 2.1. **The table contains proteomic and transcriptomic results which are described in the discussion. Proteins presented here are significantly regulated with fold change >1.5 at p-value <0.05 while genes are significant if q-value is <0.05. Because proteins and genes log2 H/L rations were calculated by dividing their expression from exponential phase (heavy labelled in SILAC) by stationary phase condition (light labelled in SILAC), its "-" values should be understand as overrepresentation in stationary phase while "+" stands for underrepresentation in that phase .**

Protein ID	Protein description		log2 protein fold change (log2 H/L)	log2 p-value	significant	log2 FKMF fold change (log2 H/L)	p-value	q-value	significant
SL pathway									
cabom02g13870	AT	Acetyltransferase	-6.10	1.67	yes	-2.72	1.45E-07	2.13E-05	yes
cabom02g13850	UgtA1	UDP-glucosyltransferase A	-4.93	2.79	yes	-2.68	1.05E-06	0.00011 3108	yes
cabom02g13880	UgtB1	UDP-glucosyltransferase B	-5.09	2.60	yes	-3.19	2.73E-05	0.00174 502	yes
cabom02g13890	Cyp52M1	Cytochrome p450	N/A	N/A		-3.87	5.77E-10	1.34E-07	yes
cabom02g13860	Mdr	ABC transporter	-2.00	0.90	no	-2.92	2.05E-05	0.00143 728	yes
TCA support									
cabom03g05520	Mls1	Malate synthase, glyoxysomal	-2.47	2.89	yes	-1.42	0.00	0.03	yes
cabom02g02830	Icl1	Isocitrate lyase	-1.91	1.77	yes	-0.09	0.83	0.93	no
cabom02g01860	Gor1	Glyoxylate reductase 1	-2.13	1.90	yes	-2.55	8.56E-08	1.40E-05	yes
cabom02g04740	Gads1	Glutamate decarboxylase	-1.47	1.50	yes	-1.20	0.01	0.10	no
Carbohydrate metabolism									
cabom01g05390	Tal1	Transaldolase	-0,61	1.64	yes	-0.61	0.19	0.48	no
cabom03g08370	Fba1	Fructose-bisphosphate aldolase FBA1	1.88	1.91	yes	-1.00	0.03	0.19	no
cabom02g10110	Pgm2	Phosphoglucomutase	-0.98	1.39	yes	0.40	0.36	0.66	no
cabom02g07540	Sou1	Sorbose reductase SOU1	-1.73	1.69	yes	-0.45	0.45	0.74	no
Glycero-3-phosphate metabolism									
cabom03g09120	Gpp2	(DL)-glycerol-3-phosphatase 2	2.85	1.96	yes	-0.55	0.30	0.60	no
Sterol biosynthesis									
cabom02g09300	Erg6	Sterol 24-C-methyltransferase	1.24	1.85	yes	0.53	0.20	0.50	no

cabom03g08670	Erg20	Farnesyl pyrophosphate synthase	1.00	2.41	yes	1.13	0.01	0.08	no
cabom03g10680	Erg13	Hydroxymethylglutaryl-CoA synthase	2.05	1.38	yes	0.63	0.14	0.41	no
cabom03g11440	Erg11	Lanosterol 14-alpha demethylase	1.47	1.71	yes	0.92	0.05	0.24	no
cabom01g08730	Dap1	Cytochrome P450 regulator dap1	-1.19	1.85	yes	0.15	0.76	0.90	no
cabom03g05710	Erg27	3-keto-steroid reductase	0.51	1.52	yes	0.51	0.22	0.52	no
		Transport							
cabom01g03610	Crs1/Sfh2	Phosphatidylinositol transfer protein CSR1	-1.96	1.80	yes	-0.02	0.97	0.99	no
cabom03g16540	Gvp36	Protein GVP36	-0.53	1.82	yes	-0.90	0.06	0.27	no
cabom03g14070	Mog1	Nuclear import protein MOG1	-0.94	2.40	yes	-0.27	0.53	0.79	no
cabom01g13820	Sec22	Protein transport protein SEC22	1.80	1.33	yes	0.21	0.64	0.84	no
cabom03g16490	Tim50	Mitochondrial import inner membrane translocase subunit TIM50	1.34	1.39	yes	0.90	0.03	0.19	no
		Amino acids metabolism							
cabom01g00270	Sam1	S-adenosylmethionine synthase	1.74	1.81	yes	0,60	0.16	0.44	no
cabom01g09660	Hom6	Homoserine dehydrogenase	1.35	1.68	yes	0.28	0.50	0.77	no
cabom03g00180	Leu1	3-isopropylmalate dehydratase	2.21	2.20	yes	0.14	0.73	0.89	no
cabom02g08330	Ilv2	Ketol-acid reductoisomerase, mitochondrial	1.12	2.03	yes	-1.00	0.10	0.34	no
cabom03g14220	Ilv3	Dihydroxy-acid dehydratase, mitochondrial	1.33	1.52	yes	0.71	0.08	0.31	no
cabom03g13780	Ilv6	Acetolactate synthase small subunit, mitochondrial	1.38	1.68	yes	0.13	0.76	0.90	no
cabom02g09000	Met6	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	0.93	1.70	yes	0.60	0.17	0.45	no
cabom03g04840	Met14	Adenylyl-sulfate kinase	0.76	1.65	yes	0.75	0.16	0.45	no
cabom02g06300	Dld/Lpd1	Dihydrolipoyl dehydrogenase, protein L	-0.82	1.50	yes	-0.01	0.98	0.99	no

cabom01g06260	Sah1	Adenosylhomocysteinase	0.74	1.32	yes	0.04	0.93	0.97	No
		Stress response proteins							
cabom02g02140	Ahp1	Peroxiredoxin type-2	-1.10	1.78	yes	0.43	0.33	0.63	no
cabom01g13680	Sod2	Superoxide dismutase [Mn], mitochondrial	-1.48	1.50	yes	-0.51	0.32	0.62	no
cabom01g08540	Glr1	Glutathione reductase	-0.66	2.87	yes	0.54	0.24	0.53	no
cabom03g16530	Hyr1 (Gpx3)	Peroxiredoxin HYR1	-1.13	1.49	yes	-0.52	0.22	0.51	no
cabom01g08860	Cdc37	Hsp90 co-chaperone Cdc37	0.51	1.53	yes	-0.34	0.41	0.70	no
cabom03g15710	Tma19	Translationally-controlled tumor protein homolog TCTP	0.85	1.33	yes	1.29	0.0014	0.03	yes
cabom03g15690	Ssb1	Heat shock protein SSB	0.75	1.53	yes	-0.39	0.34	0.64	no
cabom01g11090	Prx1	Mitochondrial peroxiredoxin PRX1	1.40	1.54	yes	1.61	8.42E-05	0.0042	yes
cabom02g04670	Hsp12	12 kDa heat shock protein Hsp12	3.05	1.33	yes	2.92	2.76E-10	7.61E-08	yes
		Nucleotide metabolism							
cabom02g11200	Mpg1/Psa1	Mannose-1-phosphate guanyltransferase	1.38	1.49	yes	-0.46	0.29	0.59	no
cabom02g03880	Dpm3	Dolichol-phosphate mannosyltransferase subunit 3	0.93	1.57	yes	0.79	0.11	0.36	no
cabom03g09860	Gda1	Guanosine-diphosphatase	0.75	2.22	yes	0.42	0.30	0.60	no
cabom03g09910	Gas5	1,3-beta-glucanosyltransferase GAS5	1.28	1.43	yes	0.88	0.05	0.24	no
cabom03g11310	Ade1	Phosphoribosylaminoimidazole-succinocarboxamide synthase	1.05	1.72	yes	0.61	0.14	0.42	no
cabom03g02550	Hnt1	Hit family protein 1	1.29	1.44	yes	-0.91	0.11	0.36	no
cabom02g01320	Apa2	5',5'''-P-1,P-4-tetraphosphate phosphorylase 2	0.93	2.76	yes	0.88	0.06	0.27	no
cabom03g11130	Fmp45	SUR7 family protein	-0.97	1.59	yes	2.58	2.58E-10	7.58E-08	yes
		Protein folding and stability							

cabom03g17490	Vps70	Vacuolar protein sorting-associated protein 70	-1.13	1.54	yes	-0.53	0.29	0.59	no
cabom01g10340	Cal1	Calnexin homolog	-0.61	1.34	yes	-0.44	0.32	0.62	no
cabom01g09330	SPAC1F5/DbsA	Putative protein disulfide-isomerase, DbsA like protein, disulphide-bond oxidoreductase A	1.83	1.83	yes	0.24	0.61	0.83	no
cabom02g09020	ERJ5	ER-localized J domain-containing protein 5	1.85	1.89	yes	0.12	0.78	0.91	no
cabom02g06550	Bag102	BAG family molecular chaperone regulator 1B	2.47	1.58	yes	0.77	0.06	0.26	no
cabom02g09110	Ubc1	Ubiquitin-conjugating enzyme E2 1	0.83	1.34	yes	0.11	0.80	0.92	no
cabom01g10260	Ola1	ATPase OLA1	0.85	1.80	yes	0.31	0.47	0.75	no
cabom01g02100	Hsp10	10 kDa heat shock protein, mitochondrial	0.52	1.67	yes	-1.60	0.01	0.12	no
		Cytoskeleton							
cabom03g00380	Arp2	Actin-related protein 2	-0.59	2.50	yes	0.44	0.30	0.60	no
cabom03g05720	Sac6	Fimbrin	-1.03	1.93	yes	-0.01	0.99	0.99	no
cabom01g09500	Slm2	Phosphatidylinositol 4,5-bisphosphate-binding protein	-0.69	1.32	yes	-1.68	0.00	0.01	yes
cabom03g09780	Sla1	La protein homolog	1.18	1.43	yes	0.27	0.51	0.78	no
cabom01g07170	Cdc4	Myosin regulatory light chain cdc4	1.53	1.76	yes	0.71	0.09	0.34	no
cabom03g09100	SPAC458.02c	Uncharacterized protein C458.02c	1.03	1.72	yes	0.81	0.06	0.28	no
		Other							
cabom03g09310	Pep4	Saccharopepsin	-1.75	1.74	yes	-1.50	0.0007	0.0192	yes
cabom03g14740	Prb1	Cerevisin	-1.14	1.32	yes	-1.24	0.07	0.29	no
cabom01g04210	Aim2	Protein AIM2	-1.54	1.55	yes	0.00	1.00	1.00	no
cabom03g00870	IPR013154	Alcohol dehydrogenase GroES-like domain	-0.52	1.35	yes	-0.01	0.97	0.99	no
		Translation							

cabom01g04640	Fun12	Eukaryotic translation initiation factor 5B	1.97	1.55	yes	0.48	0.24	0.53	no
cabom02g12940	Sui3	Eukaryotic translation initiation factor 2 subunit beta	1.69	1.43	yes	0.35	0.42	0.71	no
cabom03g15340	Hcr1	Eukaryotic translation initiation factor 3 subunit J	1.04	2.38	yes	0.14	0.73	0.89	no
cabom01g12920	Int6	Eukaryotic translation initiation factor 3 subunit E	0.97	1.94	yes	0.30	0.47	0.75	no
cabom02g01540	Tif471	Eukaryotic translation initiation factor 4 gamma	1.71	1.35	yes	0.08	0.86	0.95	no
cabom03g00620	Dbp5	ATP-dependent RNA helicase DBP5	0.99	1.40	yes	0.37	0.38	0.68	no
cabom02g10960	Ala1	Alanyl-tRNA synthetase, cytoplasmic	1.09	1.42	yes	0.48	0.28	0.58	no
cabom03g13680	Efb1	Elongation factor 1-beta OS=Saccharomyces cerevisiae	0.94	1.64	yes	0.79	0.09	0.33	no
cabom02g04730	Egd1	Nascent polypeptide-associated complex subunit beta	0.89	2.04	yes	0.71	0.12	0.38	no
cabom03g04570	Hta1	Histone H2A-alpha	-0.53	2.66	yes	-0.19	0.65	0.85	no
cabom03g11380	Tif1/2	ATP-dependent RNA helicase eIF4A	0.88	1.56	yes	-0.13	0.76	0.90	no
cabom01g07150	Arc1	GU4 nucleic-binding protein 1	0.79	1.85	yes	1.00	0.01	0.13	no
cabom01g06110	YHR020W	Putative prolyl-tRNA synthetase	0.79	2.13	yes	0.22	0.61	0.83	no
cabom03g08600	Pab1	Polyadenylate-binding protein, cytoplasmic and nuclear RNA recognition motif.	1.02	1.47	yes	-0.41	0.37	0.67	no
cabom03g04260	Arb1	ABC transporter ATP-binding protein	1.43	1.96	yes	0.12	0.77	0.91	no
cabom03g17220	Rpp1B	60s Acidic ribosomal protein	1.05	2.63	yes	-0.21	0.64	0.85	no
cabom02g12160	Rpp2A	Acidic ribosomal protein P2	0.87	1.98	yes	0.37	0.48	0.76	no
cabom01g00500	Rpp0	60S Acidic ribosomal protein P0	1.36	1.93	yes	0.11	0.79	0.92	no
cabom02g10190	Rpp1A	60s Acidic ribosomal protein	1.11	2.02	yes	0.42	0.50	0.77	no

cabom02g03070	Rpl2	60S ribosomal protein L2	1.40	3.44	yes	0.04	0.93	0.97	no
cabom02g03230	Rpl3	60S ribosomal protein L3	1.24	1.47	yes	0.16	0.69	0.87	no
cabom02g00650	Rpl4A	60S ribosomal protein L4-A	1.43	1.85	yes	-0.37	0.42	0.72	no
cabom02g00920	Rpl5	60S ribosomal protein L5	1.47	1.65	yes	-0.56	0.18	0.47	no
cabom02g03990	Rpl6B	60S ribosomal protein L6-B	1.26	2.53	yes	-0.18	0.73	0.89	no
cabom02g06700	Rpl7	60S ribosomal protein L7	1.19	1.73	yes	-0.29	0.50	0.77	no
cabom02g03140	Rpl8B	60S ribosomal protein L8-B	1.45	2.22	yes	-0.11	0.79	0.91	no
cabom01g10160	Rpl9B	60S ribosomal protein L9-B	1.28	1.64	yes	0.42	0.32	0.62	no
cabom03g03060	Rpl10A	60S ribosomal protein L10a	1.22	3.20	yes	1.15	0.00	0.07	no
cabom02g04930	Rpl10	60S ribosomal protein L10	1.32	1.81	yes	-0.10	0.82	0.93	no
cabom03g12630	Rpl13	60S ribosomal protein L13	1.33	2.31	yes	0.23	0.61	0.83	no
cabom02g07780	Rpl16A	60S ribosomal protein L16-A	0.85	2.19	yes	0.08	0.86	0.95	no
cabom02g10090	Rpl19	60S ribosomal protein L19	1.02	1.59	yes	0.03	0.94	0.98	no
cabom03g13570	Rpl22	60S ribosomal protein L22	0.91	2.18	yes	1.63	0.00	0.00	yes
cabom03g12020	Rpl25	60S ribosomal protein L25	1.43	2.05	yes	-0.34	0.49	0.77	no
cabom01g07380	Rpl26B	60S ribosomal protein L26-B	1.28	2.41	yes	1.60	0.00	0.06	no
cabom02g05030	Rpl27A	60S ribosomal protein L27-A	1.29	2.25	yes	0.66	0.19	0.49	no
cabom01g03270	Rpl28E	Probable 60S ribosomal protein L28e	2.19	1.52	yes	0.48	0.27	0.57	no
cabom02g05610	Rpl31	60S ribosomal protein L31	1.28	3.24	yes	-0.03	0.95	0.98	no
cabom02g12490	Rpl35	60S ribosomal protein L35	1.45	2.47	yes	-0.86	0.16	0.45	no
cabom01g13830	Rpl50	60S Ribosomal L50	0.76	1.35	yes	0.44	0.31	0.61	no
cabom03g12530	Rps1	40S ribosomal protein S1	1.38	1.85	yes	0.57	0.16	0.45	no
cabom03g14920	Rps2	40S ribosomal protein S2	1.44	1.98	yes	-1.06	0.03	0.19	no
cabom02g00360	Rps3	40S ribosomal protein S3	1.14	2.04	yes	0.01	0.99	0.99	no
cabom01g13970	Rps4	40S ribosomal protein S4	1.49	1.48	yes	-0.15	0.71	0.88	no
cabom03g12270	Rps5	40S ribosomal protein S5	0.87	2.23	yes	-0.70	0.11	0.37	no
cabom02g06020	Rps6	40S ribosomal protein S6	1.38	2.37	yes	-0.30	0.51	0.78	no
cabom02g06060	Rps10A	40S ribosomal protein S10-A	1.34	2.98	yes	0.41	0.44	0.73	no
cabom02g00420	Rps15A	40S ribosomal protein S15-A	1.36	2.54	yes	-0.19	0.70	0.87	no
cabom01g11320	Rps17B	40S ribosomal protein S17-B	1.41	1.87	yes	-0.16	0.74	0.89	no

cabom03g01870	Rps20	40S ribosomal protein S20	1.26	2.14	yes	0.05	0.92	0.97	no
cabom01g01000	Rps21	40S ribosomal protein S21	1.22	1.58	yes	-0.17	0.73	0.89	no
cabom01g11870	Rps25B	40S ribosomal protein S25-B	1.59	1.61	yes	0.42	0.33	0.63	no
		Not well defined function/Unknown							
cabom02g03080		no-hit	1.15	2.26	yes	-0.04	0.93	0.98	no
cabom02g07760		no-hit	1.60	2.20	yes	0.25	0.63	0.84	no
cabom02g07380		no-hit HsbA	-1.64	1.51	yes	2.39	3.33E-09	6.39E-07	yes
cabom01g00680		no-hit	1.11	4.43	yes	-0.31	0.49	0.76	no
cabom02g10040		no-hit	-0.74	2.29	yes	0.44	0.34	0.64	no
cabom01g13140		no-hit	0.85	1.99	yes	0.28	0.53	0.79	no
cabom02g04090		no-hit	0.81	1.88	yes	1.08	0.01	0.11	no
cabom02g05070		no-hit	1.00	1.38	yes	0.92	0.04	0.22	no
cabom01g05030		Probably ribonuclease L-PSP	-0.68	1.45	yes	0.14	0.75	0.90	no
cabom03g15000		Uncharacterized protein C16G5.07c SPFH domain / Band 7	1.17	1.66	yes	0.74	0.07	0.29	no
cabom01g13920		TB2/DP1, HVA22 family _stress response	1.29	1.89	yes	-2.52	5.95E-06	0.00	yes

3.5 mRNA profiling

The change at the gene expression level based on mRNAseq was statistically evaluated using the Cufflinks software (with FDR <0.05). In this study, we used mainly mRNAseq data for validation of the protein ratios of the 615 proteins quantified with SILAC. Due to the differences in technology, the levels of biological information and statistics, we concentrated on the general trends in gene and protein ratio abundance (**Table 2.1**). From a Pearson correlation coefficient calculation it can be seen that mRNA-seq analysis poorly correlates to the SILAC experiment (r=0.29) (**Fig. 2.6**), but a similar observation was previously made in related studies [148]. However, when only the 147 significantly regulated proteins (that passed a t-test) are taken into account, the correlation improves to 0.47 and when only up-regulated proteins in the stationary phase are considered, the correlation coefficient increases to 0.64 (the details about quatification can be found in **Suppl. data**, Transcriptomic data).

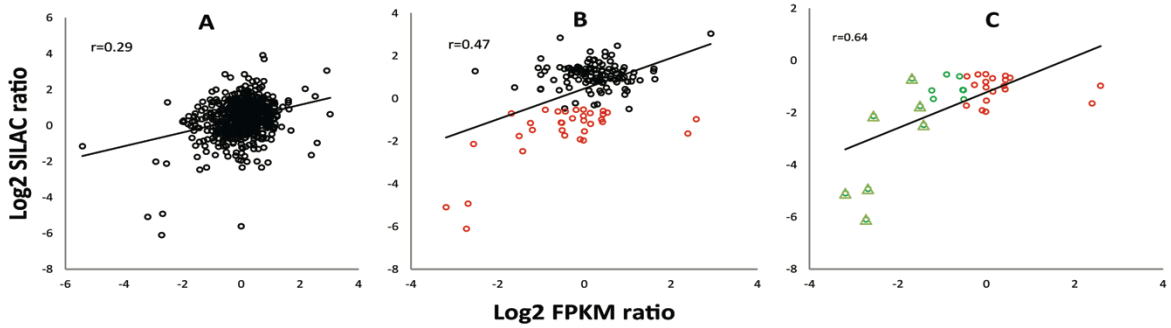


Figure 2.6. A. The overall correlation between protein and mRNA changes is r= 0.29. B. The correlation between proteins and mRNA ratio is improved when only significantly quantified proteins (147 protein, which pass a t -test) are plotted. In red, significantly up-regulated proteins during sophorolipid production. C. The correlation between 33 proteins significantly up regulated during sophorolipid production and their corresponding mRNA ratios. Marked in green are proteins which exhibit a common trend with mRNA-expression. Proteins specified in green triangles have an expression level confirmed by significant mRNA changes.

We also observe a common trend in protein and transcript abundance within GO classes at the two time points. The diagrams in **Fig. 2.7** (stationary phase) and **Suppl. Fig. 2.3** (exponential phase) represent the distribution of the log2 MS intensities (calculated from the ion intensity) and the log2 FPKM value for enriched GO terms extracted from **Fig. 2.5** [149]. For example, the diagrams representing the stationary phase (**Fig. 2.7**) show that in general the mRNA abundance for the various classes is similar as compared to the protein abundance. Highly abundant transcripts mostly represented protein classes which are highly expressed, including enzymes involved in glycolysis, ATP synthesis coupled proton transport, the mitochondrial proton-transporting ATP synthase complex and coupling factor F(o). Deviation in transcript - protein expression is typically observed in GO groups connected with endopeptidase and DNA binding activity, which have relatively lower peptide MS intensities.

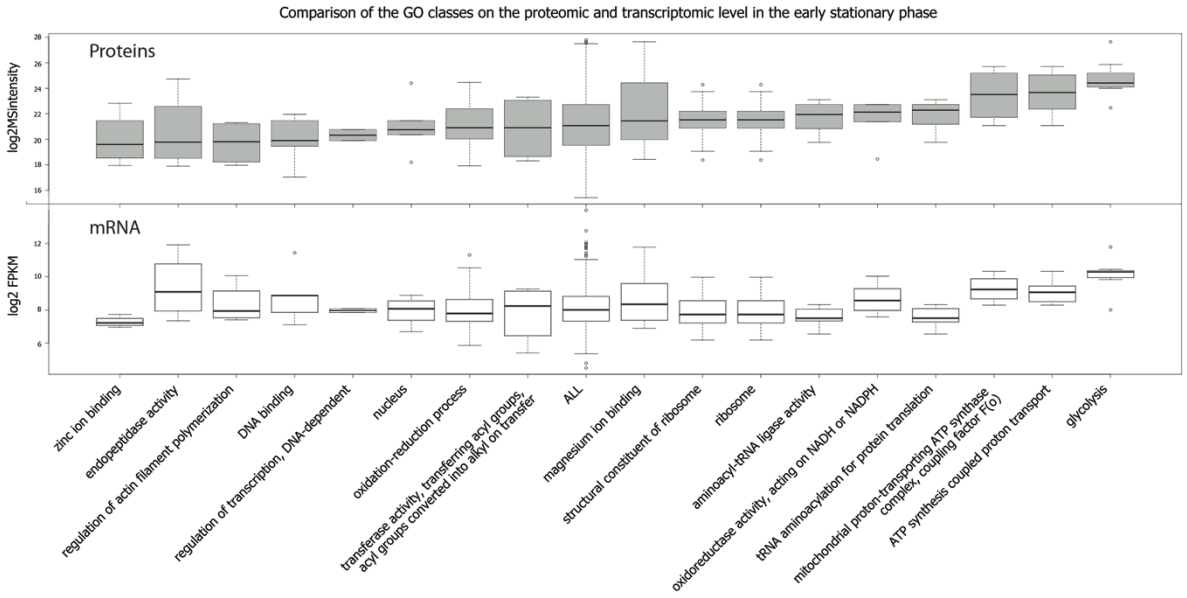


Figure 2.7. Proteins and transcript abundance within GO classes for the early stationary phase. In gray, distribution of log2 MS intensity within GO classes while in white the distribution of log2 FPKM values for transcript corresponding transcripts. (Suppl. Fig. 2.3 presents data for the exponential phase).

3.6 Glucose, ammonium and phosphate measurements

Sophorolipid synthesis is starting at the stationary phase and is generally linked to exhaustion of some nutrients like ammonium or phosphate. We have measured the level of ammonium, phosphate and glucose in the media during transitions from exponential to stationary phase (0-96h) under the conditions used for the proteomic experiments. We confirmed by TLC visualization that at OD_{600} of 10 no sophorolipids are produced while at OD_{600} of 30 they are present in the medium (**Fig. 2.3**). The time course of the measured parameters is displayed in **Fig. 2.8**. We notice a pH drop from 5.8 to around 4 at 96h, which is typical for yeast due to secretion of organic acids to the medium. Glucose concentration stays stable till 35h of growth and then a small decrease can be observed with a final concentration of 0.57 M, which is still far above limiting concentrations. The phosphate concentration drops slightly from 21 mM to 17 mM at 96h. Similarly, the free ammonium concentration reached 1.95 mM at 30h and then drops to 1.8 mM at 70h. In general, although we notice small changes in the concentration of the nutrients during the switch to the sophorolipid production state but none of the described nutrients is depleted. This is an important observation because it was earlier suggested that sophorolipid production is initiated by nitrogen [9] or/and phosphate limitation [58].

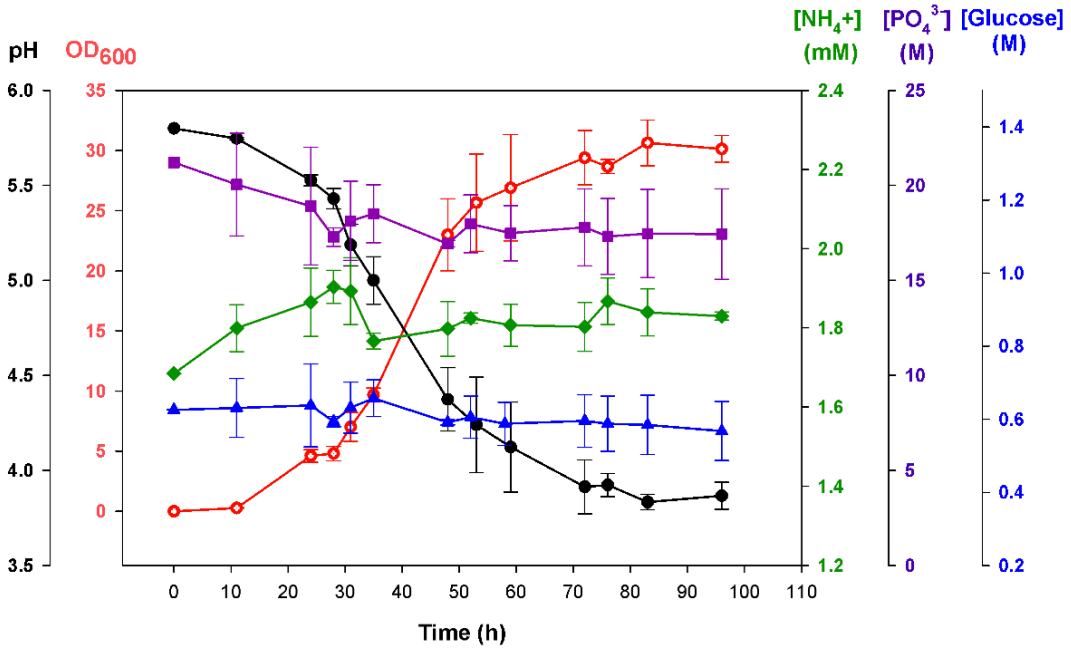


Figure 2.8. Time course of cultivation parameters (OD_{600}, pH) and the level of ammonium, phosphate and glucose. The error bars are equal to the standard deviation calculated from three biological replicates.

4 Discussion

4.1 The sophorolipid biosynthesis pathway is induced upon leaving the exponential growth phase

Sophorolipids are synthesized from two fundamental building blocks: glucose and fatty acids. In industrial production, typically vegetable oils are added to fermentation to increase SL yields [40]. However, when no such substrate is added, *S. bombicola* uses *de novo* synthesized fatty acids. In 2007, Van Bogaert [11] proposed a pathway for sophorolipid biosynthesis which is confirmed by the discovery of the genes encoding the different enzymes involved [46-49, 53] (**Fig. 2.1**). In the first step of this pathway, fatty acids are terminally or subterminally hydroxylated by a membrane-bound cytochrome P450 enzyme (Cyp2M1) [46]. In our proteomic experiment, this cytochrome P450 is exclusively identified in the stationary growth phase and therefore no SILAC ratio was obtained. In addition, our mRNAseq data confirm previous RT-PCR experiments showing that the expression of the corresponding gene is dramatically increased in the stationary phase (**Table 2.1**). We did not detect any other Cyp52 family member in our SILAC experiments, which further suggests that Cyp52M1 is the key enzyme to convert fatty acids to substrates for the consequent glycosyltransferase reactions. Interestingly, we discovered a possible enhancer for Cyp52M1 action, *i.e.* a homologue to damage resistance protein1 (Dap1) with higher abundance in the stationary phase cells. Dap1 is a haem binding protein related to cytochrome b5 that stabilizes and positively regulates cytochrome P450 proteins [150, 151] and participates in the metabolism of lipids and sterols [152].

In the next two reactions of sophorolipid biosynthesis glucose is stepwise bound to, respectively, the hydroxylated fatty acid and the resulting glucolipid (**Fig. 2.1**). As described by Saerens *et al.* [47, 48], those reactions are carried out by two glucosyltransferases UgtA1 and UgtB1, which use UDP-glucose. Both glucosyltransferases show a significant higher abundance in the early stationary phase followed by the up-regulation of their corresponding genes.

In addition, the enzyme responsible for sophorolipid acetylation (**Fig. 2.1**), acetyltransferase (AT), has the most extreme log2 fold change on both the proteomic and mRNA level. Lastly, the ABC transporter (Mdr), which was reported to function as the sophorolipid exporter [53] showed a higher abundance in all three replicate experiments, but did not pass the t-test due to variation in the actual SILAC ratios. However, on the mRNA level it is significantly overexpressed in the early stationary phase.

In general, we can conclude that the sophorolipid gene cluster is highly expressed at the early stationary phase, resulting in a set of high abundant enzymes which explains the high yield of the product. The last, yet unidentified, enzyme in the pathway is a lactonase that catalyzes the ring closure to form the lactonic sophorolipids. Our data did not reveal the nature of this enzyme. However, we emphasize that in this study a selection for more hydrophobic proteins is performed, leaving the possibility that this enzyme is lost during sample preparation.

4.2 Stationary phase related changes in protein abundance

The samples representing the sophorolipid producing cells were taken after 75 h of growth when the growth curve was flattened (**Fig. 2.8**), *i.e.* when the cellular growth was slowed down. Not surprisingly, the majority of proteins displaying significant changes in abundance can be explained in this context. The largest group of down-regulated proteins is involved in protein translation and maturation. In our experiment, 5 translational initiation factors, 37 ribosomal proteins and 10 other proteins connected with protein synthesis showed a significantly reduced abundance when compared to the exponential phase. The reduction in *de novo* protein synthesis is also reflected by the decreased presence of several enzymes involved in amino acid biogenesis. For example the enzymes participating in the production of 'pyruvate family'-amino acids (Ile2, -3, -6 and Leu1) and sulphur containing amino acid biosynthetic enzymes (Met6, Met14) were significantly underrepresented in the stationary phase. A similar observation was made for proteins participating in protein folding (DbsA, Bag102), glycosylation (Mpg1, Dpm3), sorting (Hsp10, ERJ5) and proteasomal protein degradation (Ubc1) [153, 154]. The same trend was followed by another protein involved in ubiquitination, Cdc4, which level of expression is in line with its function in cell cycle progression [155].

This dramatic reduction in protein biosynthesis is accompanied by up-regulation of protein degradative components of the lysosomal vacuoles, *i.e.* two proteases: saccharopepsin (Pep4) and cerevisin (Prb1), and glutamate carboxypeptidase (vacuole sorting protein (VSP70)). The simultaneous up-regulation of both proteases is logical since Prb1 activates pre-peptides of Pep4 [156]. It is established that the expression of this group of enzymes is a typical cellular response to starvation, so to any process resulting in deprivation of nourishment. Their function is to prevent the harmful accumulation of some proteins or to activate vacuolar enzymes necessary for survival [157].

Those proteases, are for example responsible for fructose-1,6-phosphatase degradation when cells switch from starvation to rich glucose medium [158]. The activation of vacuolar lysosomes is not necessarily linked to full exhaustion of nitrogen supply. Reduction of the availability of non-essential amino acids like leucine affects the cell cycle and can induce morphological changes such as changes in ultrastructure and the number of vacuoles [159].

Another pathway that is typically associated with the stationary phase is the glyoxylate cycle. This cycle, taking place in the peroxisome, bypasses the citric cycle and is supplied with succinate produced from acetyl-CoA. In our experiment the two enzymes specific for this cycle, *i.e.* isocitrate lyase (Icl1) and malate synthase (Mls1), are significantly up-regulated during the sophorolipid production phase (**Suppl. Fig. 2.4**). Their appearance in the stationary phase is well described in other species like *S. cerevisiae* [160] or *C. albicans* [161]. Also glyoxylate reductase (Gor1), which reduces glyoxylate to glycolate, is more abundant in early stationary phase cells. The function of this enzyme is not very clear, though it could be involved in detoxification of excess glyoxylate [162]. Associated with this pathway is the up-regulation of glutamate decarboxylase (Gad1) that converts L-glutamate into 4-aminobutyric acid. This glutamate catabolic metabolite supplies the citric acid and glyoxylate cycle for subsequent conversion into succinate aldehyde and succinate. It is typical for the stationary phase in other yeasts [161].

We report a striking atypical behaviour of heat shock protein 12 (Hsp12), which is widely documented to be induced in the stationary phase [163, 164], but is down-regulated in sophorolipid producing cells at both the mRNA and protein level. Recently, Welker et al [165] showed that Hsp12 is a membrane stabilizing protein induced in several stress responses such as osmotic and heat stress. It actually controls membrane fluidity in *S. cerevisiae* and is involved in response against desiccation stress. We assume that the production of sophorolipids protects *S. bombicola* from such stress reducing the need for Hsp12. Interestingly, knocking out Hsp12 increased the accumulation of the storage carbon source trehalose in *S. cerevisiae* [166] to which sophorolipid production in *C. apicola* and *S. bombicola* was compared [40]. Supportive for the adaptation of the cell wall structure, we observed a strong increase in abundance for another protein potentially involved in cell surface maintenance, annotated as a hydrophobic surface protein A (HsbA) [167]. However, knocking out this gene did not indicate any direct importance for the SL production.

4.3 Proteins involved in metabolic changes accompanying the sophorolipid synthesis pathway

The large production of sophorolipids by *S. bombicola*, which can yield around 400 gram/liters, requires a shift in fluxes through the metabolic pathways of carbohydrate and fatty acid metabolism. Previous studies based on biotransformation of D-[1-13C] glucose revealed that the glucose building blocks are obtained from gluconeogenetic pathways [40]. In this work, we used non-limiting glucose concentrations (120 g/L glucose) and only a small decrease in glucose content in medium was noticed after 75h of growth. As such, for the glycolysis/gluconeogenesis pathway enzymes, no significant differences in abundance were observed between the two tested conditions except for fructose-bisphosphate aldolase (Fba1), an enzyme that is not considered to be a key regulated enzyme in this pathway. A marked increase is observed in the abundance of phosphoglucomutase (Pgm2) that is needed to convert glucose-6-phosphate to glucose-1-phosphate and further to UDP-

glucose production, the substrate for the glycosyltransferases UgtA1 and UgtB12 in the sophorolipid pathway.

We also did not observe major changes in the abundance of fatty acid synthesis enzymes. Indeed, fatty acid synthase (FAS) displays a rather constitutive expression in *S. cerevisiae* while in some oleaginous yeast it can be regulated by the presence of external FA [168]. Metabolomics studies showed that cells from the stationary phase contain less FA (building blocks of membranes and cell wall) when compared to the exponential phase [169]. However, the up-regulation of the proteins participating in the creation of the buildings blocks of fatty acids such as citrate lyase and citrate synthase, is also reported in oleaginous yeast [95]. In our experiment we only observed a significant down-regulation of enzymes involved in ergosterol synthesis. The ergosterol pathway is extremely demanding in terms of acetyl-CoA and NADPH and its down-regulation could be an indication of redirecting carbon flows and reducing power into sophorolipid production. The need for NADPH is also illustrated by the observation that pentosephosphate cycle enzymes such as transaldolase (Tal1) are more abundant. Alternatively, ergosterol biosynthesis is strongly responsive to osmotic stress, via the HOG response pathway and it remains to be determined whether the change of the physicochemical properties in the cell environment, due to sophorolid production, modulates the ergosterol biosynthesis pathway [170].

Another component of the fatty acid biosynthesis pathway, (DL)-glycerol-3-phosphatase 2 (Gpp2) is down-regulated during SL production. It takes part in glycerol biosynthesis and is induced by osmotic stress [171]. The down-regulation of this enzyme leads to accumulation of G3P which can be used for phospholipid synthesis. In *Yarrowia lipolityca* activity of the GPP was not present and its metabolism was dedicated to produce G3P instead of glycerol. The increased level of G3P stimulate triglyceride (TAG) synthesis [172] and thus this atypical pathway contributes to the oleaginous character of *Y. lipolityca*. In case of *S. bombicola* G3P could participate in lipid bodies formation, which presence was observed during stationary phase by optical microscopy (results not shown).

4.4 Stress response

All aerobically growing organisms are continuously exposed to reactive oxidants and oxidative stress occurs when the concentration of those oxidants increases beyond the oxidative buffering capacity of the cell. Cells developed defence mechanisms against oxidative stress and in *S. cerevisiae* these enzymatic responses are well studied [173]. In *S. bombicola* similar enzymes are overrepresented in the stationary phase, including glutathione reductase (Glr1), glutathione peroxidase (Hyr1=Gpx3), thioredoxin peroxidase (Aph1) and mitochondrial superoxide dismutase (Sod2). Only mitochondrial thioredoxin peroxidase (Prx1) showed the reverse response. Most of these proteins are regulated mainly by the YAP1 transcription factor as a reaction to H_2O_2 stress. The YAP1 transcription factor is shown to partly cooperate with Snf1, the yeast ortholog of AMP-activated protein kinase. Snf1 is a transcription factor, which is the central component in the glucose repression signalling pathway regulating typically stationary phase metabolic pathways such as gluconeogenesis [174]. Reduced levels of other stress responsive proteins can be explained in related terms. For example, Ssb1 is shown to sequester Snf1 in an unphosphorylated state when glucose is available [175]. Tma19 is involved in cell division and is previously described to be repressed in the stationary phase of baker's yeast [176].

4.5 Cytoskeleton/vesicle transport

Cells remodel their cytoskeletons to regulate processes like endocytosis/exocytosis, cell polarity and cell morphology [177]. Endocytosis/exocytosis is a plasma membrane-originated vesicular trafficking process in which actin is used for vesicle formation and movement [178]. In SL producing conditions we detected a group of proteins connected with those processes which displayed a higher abundance. To that group belongs actin binding protein (Sac6) participating in actin patch formation [179], actin nucleator (Apr2), responsible for actin polymerization [180] and phosphatidylinositol 4,5-bisphosphate-binding protein (Slm2) involved in actin cytoskeleton polarization during cell cycle progression, cell wall integrity and nutrient receptor endocytosis [181]. Another up-regulated protein is Gvp36, a BAR domain protein, connected with vesicular traffic and nutritional adaptation. Noteworthy is the high increase of abundance of Sfh2 which belongs to the phosphatidylinositol (PI) transfer protein Sec14 family [182]. Sec14-related proteins belong to the non-classical PI phosphate transporters (PIPTs) that only bind phosphatidylinositol phosphate (PIP) [183]. It was shown that Sfh2 together with Sfh5 participate in regulating post-Golgi membrane trafficking events [184]. Those secretory processes might be essential for both translocation of the transporter in the plasma membrane as for sophorolipid vesicles movement inside of the cell, though at present the secretory pathway of sophorolipids is unknown.

4.6 What triggers sophorolipid production?

The sophorolipid producing enzymes are organized in a gene cluster which indicates co-regulation [53]. During the last years, clustered genes responsible for the biosynthesis of comparable molecules were characterized in other species producing biosurfactants, like *Ustiligo maydis* [51] and *Pseudozyma flocculosa* [52]. In both examples, the expression of the protein assembly is maintained by a transcription factor responsive to a changing situation in the environment. In case of *U. maydis,* production of the biosurfactants occurs under nitrogen limitation, which activates the Rua1 transcription factor included in the ustilagic acid production gene cluster [55]. In the sophorolipid cluster of *S. bombicola* no transcription factor was detected [53]. Our proteomic work does not give further clues in that matter, although it was not oriented to signaling processes which require phosphoproteomic experiments. Davila *et al.* [57] described that sophorolipid production in *S. bombicola* is connected with nitrogen limitation. Later, Albrecht *et al.* [58] followed nitrogen and phosphate concentration during *S. bombicola* growth and concluded that SL production occurred at total phosphate exhausted and nitrogen limitation. Our results are not consistent with this observation. We detected only a small decrease in concentration in all three nutrients (ammonium, phosphate and glucose). It is arguable that the concentration of more specific amino acids or other metabolites is modulating sophorolipid production but this requires further investigation.

5 Conclusion

Our work focused on a proteomic comparison between sophorolipid inducing and non-inducing conditions in *S. bombicola*. We conclude that the proposed enzymes for the sophorolipid pathway in *S. bombicola* are co-ordinately produced in the stationary phase at level 2 magnitudes of order

higher than in exponential phase grown cells. Membrane protein enrichment also allowed confirming overexpression of the ABC transporter discovered in SL cluster. Sophorolipid production is not strictly dependent on glucose, ammonium and phosphate limitation and is accompanied by alteration of central carbon pathways, cytoskeleton reorganization and stress responses.

Supplementary tables

Suppl. Table 2.1. Composition of SD medium for transformation experiments and preculturing

Component	Concentration [g/L]
YNB/w aa	6.8
agar	20
glucose	20
$^{12}C_6$-Lys or $^{13}C_6$-Lys	0.2
Tryptophan	0.3
Histidine	0.15
Methionine	0.3

Suppl. Table 2.2. Modified Lang medium for SILAC experiments on sophorolipid producing *S. bombicola*

Component	Concentration [g/L]
glucose•H_2O	132
YNB/w aa	6.8
3Na-citraat 2H_2O	5
NH_4Cl	1.5
KH_2PO_4	1
K_2HPO_4	0.16
$MgSO_4$•7H_2O	0.7
NaCl	0.5
$CaCl_2$•2H_2O	0.27
$^{12}C_6$-Lys or $^{13}C_6$-Lys	0.2
Tryptophan	0.3
Histidine	0.15
Methionine	0.3

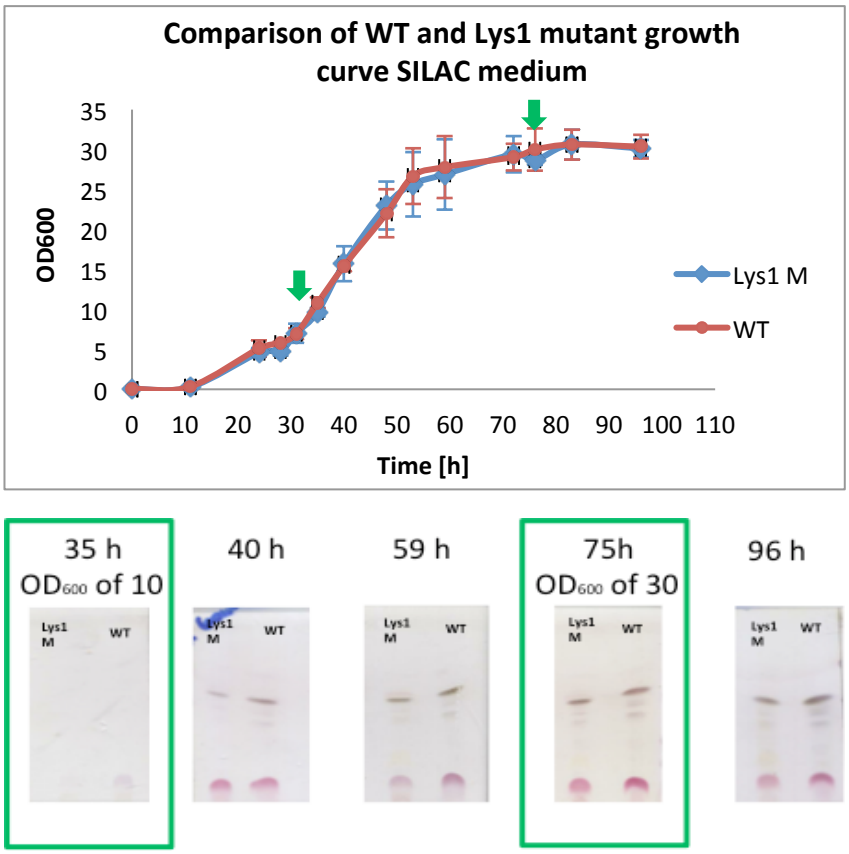
Suppl. Table 2.3. MaxQuant parameters

Parameter	Value
Version	1.1.1.36
Peptide FDR	0.01

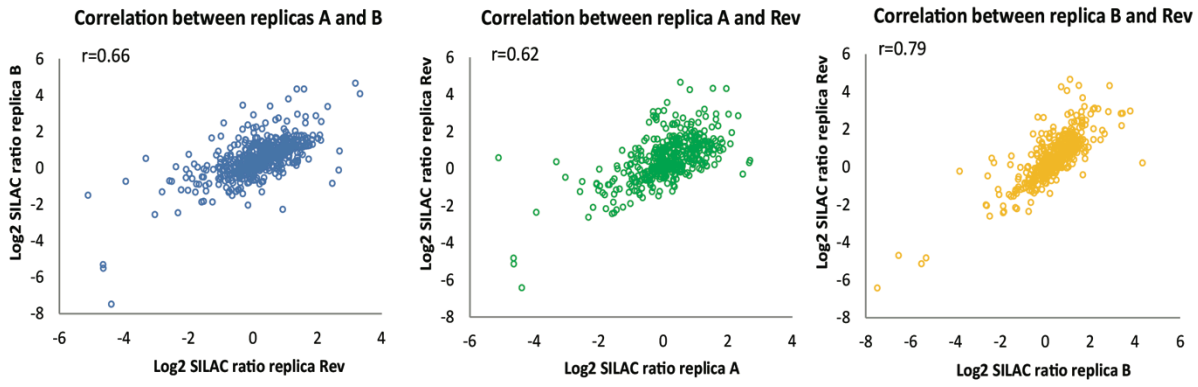
Max. peptide PEP	1
Protein FDR	0.01
Site FDR	0.01
Apply site FDR separately	TRUE
Min. peptide Length	6
Min. score	0
Min. unique peptides	0
Min. razor peptides	1
Min. peptides	1
Use only unmodified peptides and	TRUE
Modifications included in protein quantification	Oxidation (M)
Peptides used for protein quantification	Razor
Discard unmodified counterpart peptides	TRUE
Min. ratio count	2
Site quantification	Use least modified peptide
Re-quantify	TRUE
Keep low-scoring versions of identified peptides	TRUE
Label-free protein quantification	FALSE
iBAQ	FALSE
iBAQ log fit	TRUE
Match between runs	FALSE
Find dependent peptides	FALSE
Fasta file	C:\Documents and Settings\Administrator\ Desktop\fasta\augPredFunc_3cons_noTab.fasta
First search fasta file	C:\Documents and Settings\Administrator\Desktop\ fasta\augPredFunc_3cons_noTab.fasta
Experimental design file	C:\Documents and Settings\Administrator\Desktop\kasia\ Expdesign\Expdesign_silacexpstat-22062012\ jan2011_ silacexpstaM2\ combined\experimentalDesignTemplate.txt
Mass filtering	TRUE
Labeled amino acid filtering	TRUE
Site tables	Oxidation (M)Sites.txt
Calculate PIF	FALSE
Cut peaks	TRUE
Variable modifications	Oxidation (M)
Fixed modifications	

Randomize	FALSE
Special AAs	K
Include contaminants	TRUE
Multiplicity	2
First search ppm tol	20
MS/MS tol. (CID)	0.5 Da
Top MS/MS peaks per 100 Da. (CID)	6
MS/MS deisotoping (CID)	FALSE
MS/MS tol. (HCD)	20 ppm
Top MS/MS peaks per 100 Da. (HCD)	10
MS/MS deisotoping (HCD)	TRUE
MS/MS tol. (Unknown)	0.5 Da
Top MS/MS peaks per 100 Da. (Unknown)	6
MS/MS deisotoping (Unknown)	FALSE
MS/MS tol. (ETD)	20 ppm
Top MS/MS peaks per 100 Da. (ETD)	10
MS/MS deisotoping (ETD)	TRUE
AIF correlation	0.8
AIF topx	50
AIF topx corr	FALSE
AIF min mass	0
AIF SIL weight	4
AIF ISO weight	2

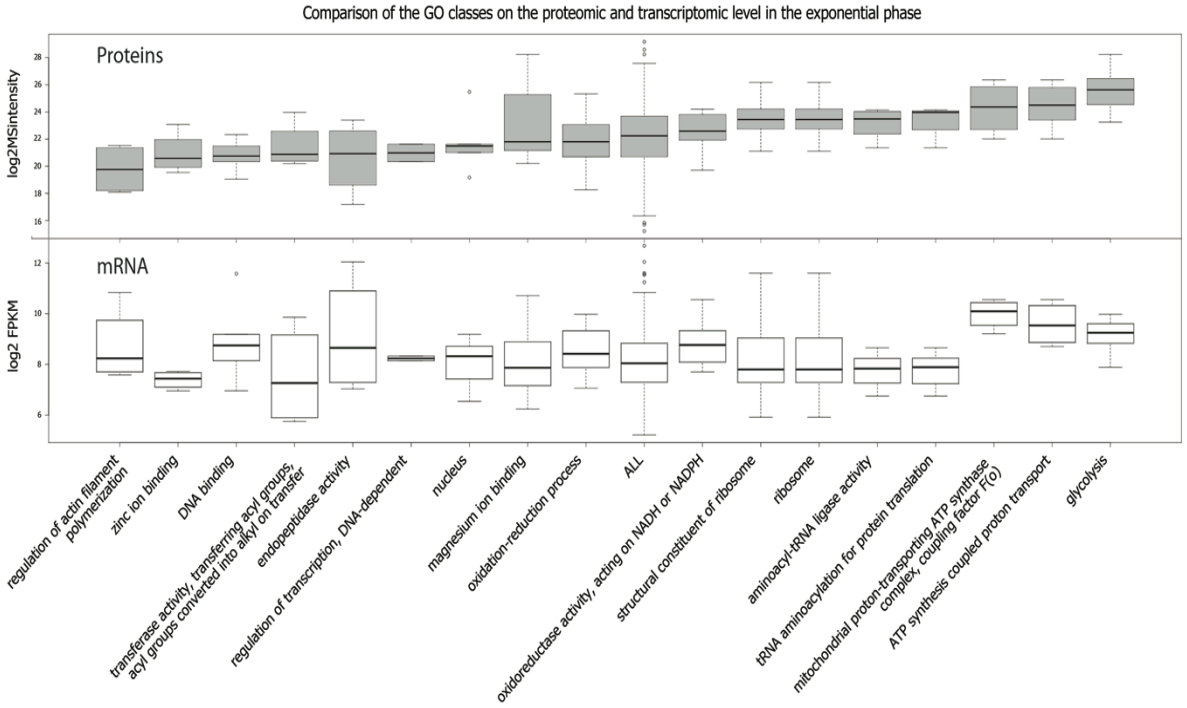
Supplementary figures



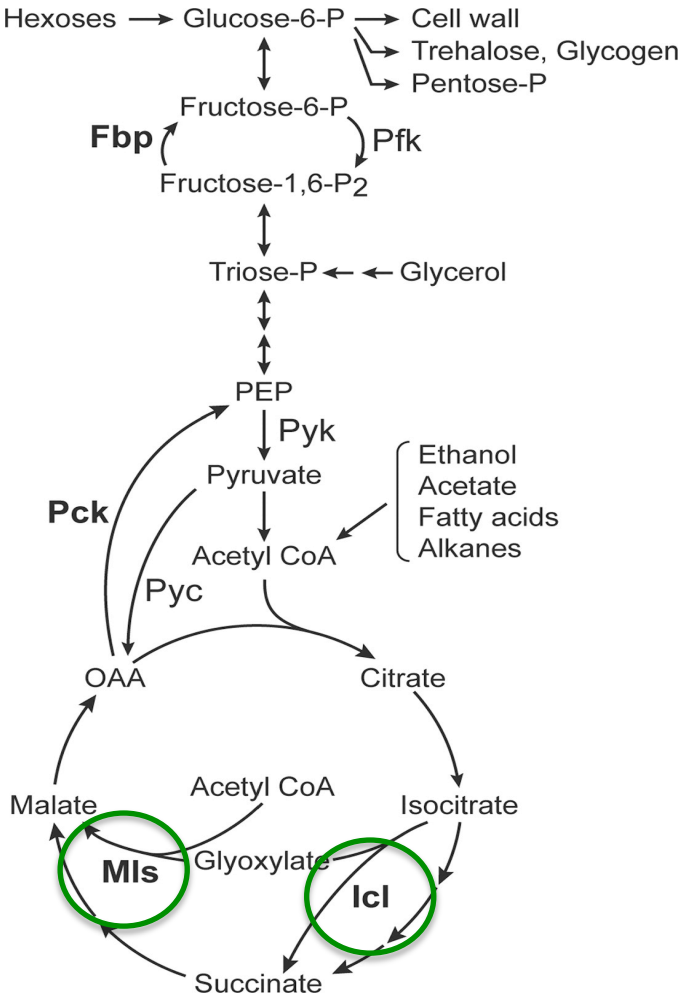
Suppl. Fig. 2.1.Comparison of the growth curves of WT and lys1 mutant. Below, the TLC pattern of extracted sophorolipids. The green frames indicate time points selected for the SILAC analysis.



Suppl. Fig. 2.2. Correlation of log2 SILAC ratio between three biological replicas. In replica Rev SILAC ratio was already inverted from L/H to H/L.



Suppl. Fig. 2.3. Protein and transcript abundance within GO classes for the exponential phase. In gray, distribution of log2 MS intensity within GO classes while in white the distribution of log2 FPKM values for transcript corresponding transcripts.



Suppl. Fig. 2.4. Glyoxylate cycle, Krebs cycle and glyconeogenesis in yeasts. Specific gluconeogenic enzymes are indicated in bold letters. Pyruvate carboxylase (Pyc), an anaplerotic enzyme, might be needed both in gluconeogenesis and in glycolysis, depending on the carbon source. Fbp, fructose-1, 6-bisphosphatase; Pfk, phosphofructokinase; Pyk, pyruvate kinase; Icl, isocitrate lyase; Mls, malate synthase; OAA, oxaloacetate [185]. In green circle proteins up regulated during SL production.

Supplementary data

Proteomic data and transcriptomic data are stored at Dropbox and can be downloaded via link https://www.dropbox.com/sh/e34ad5bnwi7heir/8hkNf4Hwnw.

Chapter 3

Exoproteome analysis of *Starmerella bombicola* results in the discovery of an esterase required for lactonization of sophorolipids.

Ciesielska K.[a]; Van Bogaert IN.[c]; Chevineau S.; Li B.[b]; Groeneboer S.[a], Soetaert, W.[c], Van de Peer Y.[b]; Devreese B.[a]

[a] Laboratory for Protein Biochemistry and Biomolecular Engineering, Gent University

[b] VIB Department of Plant Systems Biology and Department of Plant Biotechnology and Bioinformatics, Gent University

[c] Laboratory of Industrial Biotechnology and Biocatalysis , Gent University

Part of the content of this chapter is included in patent application WO 2013/092421 :" A lactonase derived from *Stramerella bombicola* and uses thereof"

Contribution: K. Ciesielska and S. Chevineau (Ms student) performed proteomic measurements and analysis of the *S. bombicola* exoproteome which led to the discovery of a lactone esterase. Dr. IN. Van Bogaert created and analysed the *S. bombicola* lactone esterase mutant strain.

Abstract

Starmerella bombicola secretes sophorolipids, a family of biosurfactants that find applications in green household products and cosmetics. Over the past years, a gene cluster was discovered that is responsible for the entire synthesis of the open (acidic) form of these molecules starting from glucose, fatty acids and acetyl-CoA building blocks. However, a significant fraction of the natural product is obtained as ring closed (lactonic). Both genetic and proteomic approaches hitherto failed to discover an enzyme responsible for the ring closure step. We hypothesized that this enzyme is extracellularly secreted. Therefore, we characterized the composition of the *S. bombicola* exoproteome at different time points of the growth and compared it with other yeast exoproteomes. We identified 44 proteins, many of them commonly found in other fungi. Curiously, we also discovered a putative lipase with homology with *Pseudozyma antarctica* lipase A. KO mutation of this lipase gene resulted in complete abolishment of the sophorolipid lactonisation proving that this is the missing enzyme in the sophorolipid biosynthetic pathway.

Keywords: biosurfactants, extracellular proteins, *Starmerella bombicola*, lipase, natural product synthesis

1 Introduction

Starmerella bombicola is a non-pathogenic fungal species isolated from honey of bumblebees [186]. It has an important biotechnological potential because it produces large amounts of biosurfactants, *i.e.* sophorolipids [24]. These biodegradable molecules are good surface tension reducers and already found application in household products [187] and in cosmetics [188]. Moreover sophorolipids are gaining interest for their antimicrobial properties [189].

Sophorolipids are glycolipids composed of a sophorose-head, which is connected at its reducing end to a terminally or subterminally hydroxylated C16 or C18 fatty acid. Natural sophorolipids can be further modified by acetyl groups at the 6' and/or 6'' position of the sophorose group. They exist in a ring-closed (lactonic) and open (acidic) form, which differ in physicochemical properties (**Fig. 1.1**). Previously, sophorolipid biosynthesis in *S. bombicola* was studied at the genetic and proteomic level. This resulted in the discovery of a sophorolipid biosynthesis gene cluster coding for a cytochrome P450, two glycosyltransferases, an acetyl transferase and an ABC transporter. However, the enzyme responsible for the sophorolipid lactonization is not included in this gene cluster [53]. Likewise, a quantitative study of the intracellular proteome, comparing a sophorolipid producing versus a non-producing growth phase did not allow us to discover a specific protein connected with lactonisation (Chapter 2). Basically, it remained obscure whether sophorolipid lactone formation really requires an enzymatic reaction or whether it is a product of the spontaneous reaction under certain culture conditions. However, it was for a long time suggested that a cell wall-bound or extracellular lipase would be responsible for the lactone formation [40]. There are a number of arguments to support this idea. First, lactonic sophorolipids display a higher antimicrobial activity compared to the open form [190], and extracellular conversion to lactones could be protective. In addition, lactonisation of the acidic unacetylated sophorolipid ester at the 6"-position was obtained *in vitro* by an extracellular lipase B from *Pseudozyma antarctica* (former *Candida antarctica*) in an anhydrous tetrahydrofuran environment [45]. Lastly, lactonic sophorolipids tend to precipitate as white crystals in the extracellular medium and such precipitates have never been observed intracellularly.

To reveal the existence of such an extracellular sophorolipid lactonase, we set up an exoproteome analysis experiment. A semiquantitative approach to compare the extracellular protein at 4 different time points during growth was selected, allowing us to discriminate specific processes in the stationary phase where sophorolipids are produced. We compared our data with reports of the composition of the exoproteome (arguably referred to as "secretome") in the literature, of other biotechnologically and medically important yeasts like *C. albicans* [112], *C. utilis* [111] or *Pichia pastoris* [191] .

2 Materials and methods

2.1 Materials

Materials for fungal growth, Lang medium and LB medium, for SDS-PAGE analysis, trypsin digestion and TLC development and visualization, were purchased from Sigma-Aldrich (USA) except for

glucose, trisodium citrate and sodium chlorate that were ordered from Merck Millipore (Germany). Acrylamide from National Diagnostics (USA) and trypsine from Wako Chemicals USA, VA. LC-MS grade acetonitrile, formic acid and acetic acid were bought from Biosolve (The Netherlands).

2.2 Strain and culture conditions for exoproteome analysis and genetic engineering

Wild type *S. bombicola* ATCC 22214 was inoculated on Yeast Extract Peptone Dextrose (YPD) agar plates, containing 2 m/v% glucose, 1 m/v% yeast extract, 2 m/v% peptone and 2 m/v% agar, and incubated at 28°C for 48h. Then, from a single colony, liquid precultures were prepared in medium described by Lang *et al* [136]. For the exoproteome experiment, the liquid cultures were started with 500 000 cells in a volume of 150mL medium in 1L bottles and incubated at 28°C, 1800 rpm. Samples were taken at different time points (sample A : 18h (OD_{600} 5), sample B : 30h (OD_{600} 10), sample C: 72h (OD_{600} 50), and sample D : 100h (OD_{600} 50)). Two biological replicas for each time point of sample collection were set up.

Starmerella bombicola PT36, an ura3 autotrophic mutant derived from *Starmerella bombicola* ATCC 22214 [135], was used to construct the lactonase knock-out strain. When sophorolipid production was intended, the medium described by [136] was used. 37.5 g/L rapeseed oil was added two days after inoculation. Yeast cultures were incubated at 30°C and 200 rpm for a total time of 10 days.

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

2.3 Sophorolipid analysis

Thin layer chromatography (TLC) was used to test the presence of sophorolipids in the culture medium. 1 ml of culture was mixed with 440 µL ethyl acetate and 11 µL acetic acid and then shaken vigorously for 5 min. After centrifugation at 9000 g for 5 min, the upper solvent layer was removed and transferred into a fresh tube. Ten µl of this sample was spotted on a silica plate (Kieselgel 60_{F254}) and separated using a mixture composed of chloroform/methanol/water (ratio 65/15/2) [9]. After 1h of solvent migration, the sophorolipids were visualized by spraying a mixture of acetic acid, sulfuric acid and anisaldehyde (ratio 100/2/1) and heating.

To evaluate the lactonase-negative mutant, HPLC analysis was used. Sophorolipid samples were daily extracted as follows: 3 mL of ethanol was added to 1 mL culture broth and shaken vigorously for 5 min. After centrifugation at 9000 g for 5 min, the supernatant was collected. At the end of the incubation period, 3 volumes of ethanol were added to the culture broth for total extraction of sophorolipids. Cell debris was removed by centrifugation at 1500 g during 10 min. Sophorolipid samples were then analysed on a Varian Prostar HPLC system using a Chromolith® Performance RP-18e 100-4.6 mm column (Merck KGaA) at 30°C with Evaporative Light Scattering Detection (Alltech). A gradient of two eluents, a 0.5 v/v% acetic acid aqueous solution and acetonitrile, was used to separate the components. The gradient started at 5% acetonitrile and linearly increased till 95% in 40 min. The solvent was kept at 95% acetonitrile for 10 min and then restored to 5% acetonitrile in 5 min. A flow rate of 1 mL/min was applied. In order to allow qualitative and quantitative analysis of the different samples, dilutions of a standard were analysed in parallel.

Liquid chromatography mass spectrometry (LC-MS) analysis was performed on a Shimadzu LC10AD vp LC system with a Quattro LC MS detector from Waters, applying electro spray ionisation (ESI). The detection range was set at m/z 200 to 800 and the negative ion mode was applied. The same column and LC conditions as for the HPLC analysis were used.

2.4 Extracellular proteins collection and concentration

The culture medium was collected and centrifuged at 4°C for 15min at 4000g. The supernatant was filtered through a 250mL Filtropur V25 vacuum filter with a pore diameter of 0.2µm (Sarstedt, Nümbrecht, Germany) and two tablets of Complete EDTA-free protease inhibitor cocktail (Roche, Basel, Switzerland) were added to the filtrate. The protein solution was then concentrated in an Amicon stirred cell (Millipore, Billerica, MU, USA) equipped with a Sartorius membrane with a 10kDa cut-off (Sartorius, Goettingen, Germany) until a volume of 30mL was reached. A second step of concentration was performed using Vivaspin 15R ultrafiltration devices with a 10 kDa cut-off (Sartorius, Goettingen, Germany). The tubes were centrifuged at 4°C and 3000 g until an appropriate concentration was obtained. The protein concentration of the samples was determined using the BCA protein assay kit (Pierce Biotechnology, Rockford, IL, USA) according to the manufacturer's protocol.

2.5 PNGase F treatment of the exracellular proteins

To perform the N-deglycosylation, a PNGase F deglycosylation kit (New England Biolabs, Ipswich, MA, USA) was used. To the volume of the protein solution corresponding to an amount of 80µg, 4µL of denaturing buffer (5 m/v% SDS, 0.4M DTT) was added along with MQ water until a final volume of 40µL. The mixture was incubated at 100°C for 10min. After that, 1µL of PNGase F (500units) was added together with 8µL of G7 reaction buffer (0.5M sodium phosphate, pH 7.5), 8µL of 10 m/v% NP-40 and MQ water till a final volume of 80µL. The reaction took place at 37°C overnight. To reduce the volume and permit the separation on SDS-PAGE, acetone precipitation was performed by adding 4 times the volume of cold acetone, vortexing the mixture and leaving it overnight at -20°C. After centrifugation for 20min at 13000g, the supernatant was discarded and the pellet was dried and resuspended in 15µL of 2 m/v% SDS. Next, this mixture together with Laemli buffer in volume of 25µL was separated on a 12.5 m/v% SDS-PAGE gel.

2.6 SDS-PAGE separation

Eighty µg of each exoproteome sample was mixed with Laemli buffer to 25 µl volume, loaded on a 12.5 m/v% SDS-PAGE gel and electrophoresis was performed at 110V. Next, the gel was incubated in a fixation solution, containing 50 v/v% ethanol and 2 v/v% phosphoric acid, for 30 min and then stained overnight under agitation in a solution, containing 34 v/v% methanol, 3 v/v% phosphoric acid, 17 m/v% ammonium sulfate and 0.2 m/v% Coomassie Brilliant Blue G-250. The gel was then destained in a solution containing 30 v/v% methanol for 1h, with the solution renewed after 30min.

2.7 In-gel trypsin digestion

The total SDS-PAGE gel lane was sliced into fragments as is presented in **Figure 3.1**. The gel pieces were destained by three 20 min incubation cycles at 30°C with 150 µl of a 50 v/v% acetonitrile (ACN)/ 200mM NH_4HCO_3 mixture. The proteins in dried gel pieces were digested by adding 8 µl of 0.002 µg/µl trypsin (Promega, Madison, WI, USA) in 50mM NH_4HCO_3 and incubating overnight at 37°C. The resulting peptides were collected from the supernatant. More peptides were extracted from the gel piece with 60 µl of 60 v/v% ACN/0.1 v/v% formic acid (FA) during 20 min incubation at 30°C. Next, the samples were vortexed for 3 min, shortly centrifuged and extraction was repeated with 20 µl of 60 v/v%ACN/0.1 v/v% FA. These extracts were pooled with the supernatant and the resulting peptide mixture was dried under vacuum centrifugation and dissolved in 15 µl of 2 v/v% ACN/0.1 v/v% FA.

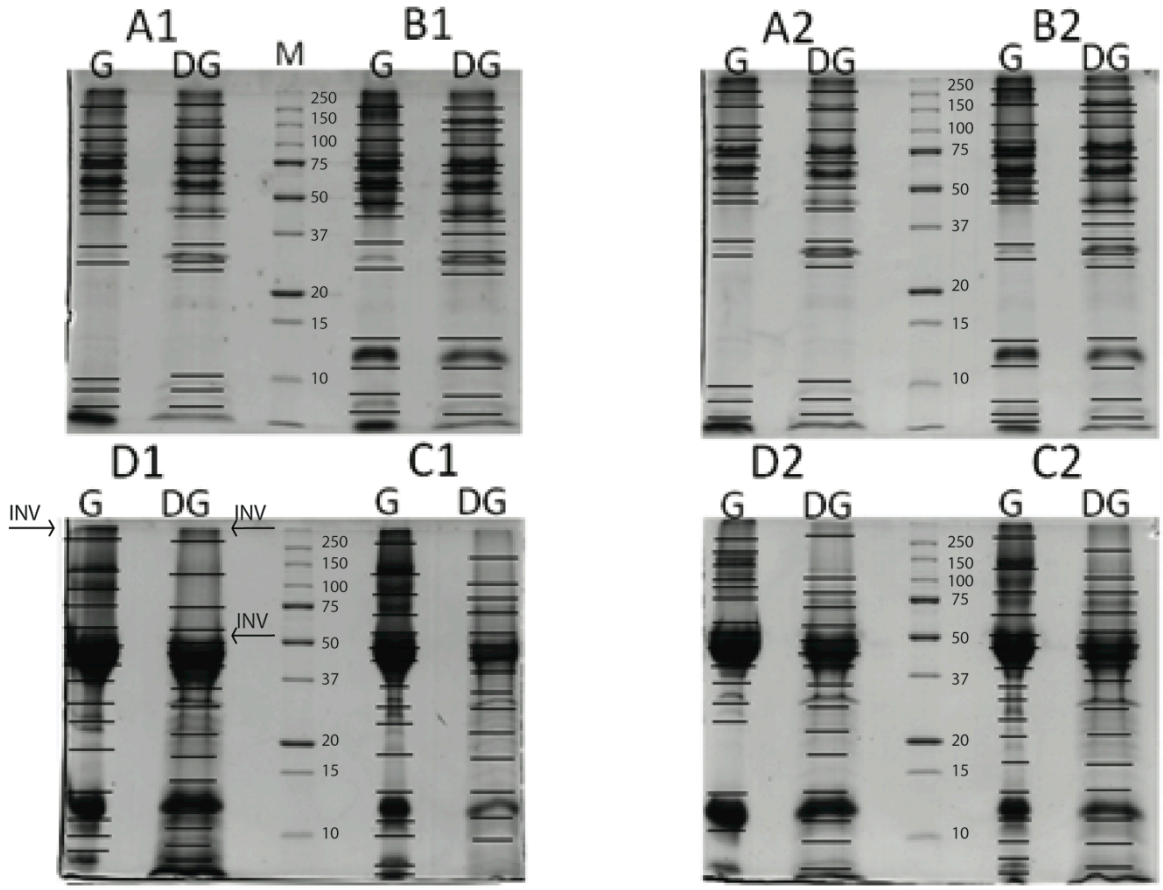


Figure 3.1 SDS-PAGE separation of the exoproteome of the *Stramerella bombicola* for four time points A, B, C, D in two biological replicas (1, 2). For time point D1, the deglycosylated (DG1) exoproteome was analysed in which invertase (INV) was identified in band 1 and band 5.

2.8 Nano LC-ESI-LTQ-FT ICR

Five µl of the extracted peptides were loaded on a Zorbax 300SB-C18 trap column 5mmx0.3mm and separated by a Zorbax 300SB-C18 analytical column 150 mm x 75 µm (Agilent) connected to an

Agilent 1200 chromatographic system (Agilent, Santa Clara, CA, USA). The LC system was coupled to an LTQ-FTUltra mass spectrometer (Thermo Fisher Scientic, Waltham, MA). The separation was performed by reversed phase chromatography using a 50 min linear gradient ranging from 2% buffer A to 80% buffer B at a 300 nl/min flow rate. The mobile phase buffer A consisted of 0.1 v/v% formic acid in water and buffer B of 0.1 v/v% formic acid in acetonitrile. The LC eluent was directly coupled to a Triversa NanoMate ESI source (Advion, Ithaca, NY), working in the nanoLC mode and equipped with D-chips to which a 1.55 kV voltage was applied.

The FT-ICR mass analyzer acquired MS scans (3 s) at a resolution of 100,000 during the LC separation. The three most intense precursor peptides (minimal intensity : 500) from each MS scan were automatically selected and fragmented in the LTQ ion trap mass analyser which was also used to collect MS/MS spectra. Mass range for precursor ion scans was from m/Z 300 to 1400, charge state 1 was rejected. The isolation width was 2 m/Z. The normalized collision energy was set at 35 with an activation Q of 0.25 and an activation time of 30 ms. The dynamic exclusion parameters were as follows : repeat count 2, repeat duration 30s, exclusion duration 90s with an exclusion mass width of ±5 ppm.

2.9 Protein identification and sequence analysis

The LC-MS/MS data were analysed with Mascot Daemon version 2.2.2 against an in-house generated *S. bombicola* protein sequence database including a decoy database. This database and the genome sequence analysis used to generate it were previously described in Chapter 2. Trypsin was specified as the enzyme with up to two missed cleavages allowed. As a variable modification, oxidation (M) and propionamide were selected. A peptide mass tolerance of 10ppm and a fragment mass tolerance of 0.3Da were chosen. The Mascot ion score cut-off was fixed at 30 and the p-value threshold was set at 0.05. Proteins with a minimum of 1 unique peptides detected were regarded as a positive identification on the condition that they were detected in more than one replicate experiment per time point. Mascot automatically calculate the emPAI factor for each identified protein [192]. The emPAI factor of proteins identified with at least 2 unique peptides was used to estimate those proteins abundance at different time points. Additionally the secretion signal, GPI anchor and cellular localization was predicted by online software respectively SignalP 4.1 Server [193], PredGPI [107] and LocTree2 [194].

2.10 Invertase assay

300 ml of *S. bombicola* extracellular medium was collected after 72h of growth, at the point where the peptide count for invertase was the highest, and concentrated as described before. Next, 1ml of this concentrated exoproteome was dialysed overnight using 10kDa cut-off membrane (Spectra Por) against 25 mM Tris, 150 NaCl pH 4.5 buffer to remove glucose and sophorolipids which could interfere with assay. The protein concentration was determinated with the BCA protein assay kit (Pierce). For the assay 40 µL of 80 mg/mL stock exoproteome concentration was used and the final exoproteome concentration used in the assay was 35.5 mg/mL. Invertase activity was determined using EnzyChrom^{TM} Invertase Assay kit (EIVT-100) from BioAssay systems. In the assay, invertase first cleaves sucrose to fructose and glucose, then the product (glucose) concentration is determinated

by colorimetric measurement of glucose-dye complex (at λ=570 nm). As a negative control the dialyzed secretome without substrate was used. The activity was measured after 26 min.

2.11 Creation of the lactonase knock-out cassette

A total fragment of 1944bp comprising the complete lactonase CDS was amplified using the primers lip2for3 and lip2rev3 (**Table 3.1**) and cloned into the pGEM-T® vector (Promega). The created vector of 4946 bp was digested with *Mfe*I en *Nar*I, in this way deleting 282 bp of the lactonase coding region.

The *Starmerella bombicola* Ura3 autotrophic marker [195] was amplified with the primers ura3MFeIFor and ura3NarIrev (**Table 3.1**), harbouring the restriction sites for respectively *Mfe*I and *Nar*I in their 5' extensions. The purified PCR fragment of 2064bp was cut with mentioned restriction enzymes and ligated into the digested vector. The resulting vector of 6717 bp was used as a template to generate the lactonase knock-out cassette with the primers lip2for3 and lip2rev3. The fragment of 3806 bp contains the ura3 marker with approximately 0.8 kb of the lactonase sequence on each site, required for homologues recombination at the lactonase locus. This linear fragment was used to transform *Starmerella bombicola* PT36 by electroporation. Transformants were selected on synthetic dextrose (SD) plates [0.67 m/v% yeast nitrogen base without amino acids (DIFCO) and 2 m/v% glucose].

All DNA sequences were determined at LGC genomics, (Berlin, Germany). *E. coli* cells were transformed as described by Inoue *et al*. [196]. Bacterial plasmid DNA was isolated with the QIAprep Spin Miniprep Kit (Qiagen).

Table 3.1 Primers used for knocking-out the *S. bombicola* sophorolipid lactonase gene. All primers were obtained from Sigma Genosys. Underlined regions mark restriction sites.

Name	Feature	Sequence
lip2for3	cloning lactonase	CAGCGCTGGGATTCATCTGCTC
lip2rev3	cloning lactonase	GCTAAGCAGCCTTGGGAGTTTC
ura3MFeIFor	amplification ura3 marker	TACAATTG-GCCTATAAGGCTAAAGAAAGTA
ura3NarIrev	amplification urea3 marker	ATGGCGCC-GATGCCGAGGAACTGTCATTGC
koLip2FlankFor	checking 5' KO genotype	CAGACGCATTGGCTGCCTTC
ura3OutBeginRev	checking 5' KO genotype	ACTGCCATCATGGTTCAACCTCAC
koLip2FlankRev	checking 3' KO genotype	TACTGCTCTGCCGATCGTTG
Ura3OutEndFor	checking 3' KO genotype	TAAAGAAACGAAGGGCCCAGCAGTC

3 Results

3.1 Exoproteome identification

To characterize the exoproteome of *S. bombicola*, the cells were cultivated in a culture medium described by Lang *et al.* [136] to four time points representing two different growth phases. The presence of sophorolipids in the stationary phase was confirmed by TLC while in the exponential phase no sophorolipids were visible (**Fig. 3.2**). **Table 3.2** presents the 44 proteins identified from all combined GeLC-MS experiments using a search against the preliminary *S. bombicola* annotated genome. 39 of the identified proteins possess a secretion signal, as predicted by the SignalP software. However, the 5 proteins without predicted secretion signal are reported to be secreted in other organisms. In 11 out of 44 identified proteins PredGPI software predicted, with high probability, a GPI anchor site which localize those proteins in the cell wall-bound compartment. LocTree2 software predicts an extracellular localization for 29 proteins, a cytosolic or ER location for 6 proteins, plasma membrane for 2 proteins and 1 protein is predicted to be peroxisomal. The small amount of intracellular proteins suggests minimal cells lysis during growth and sample processing.

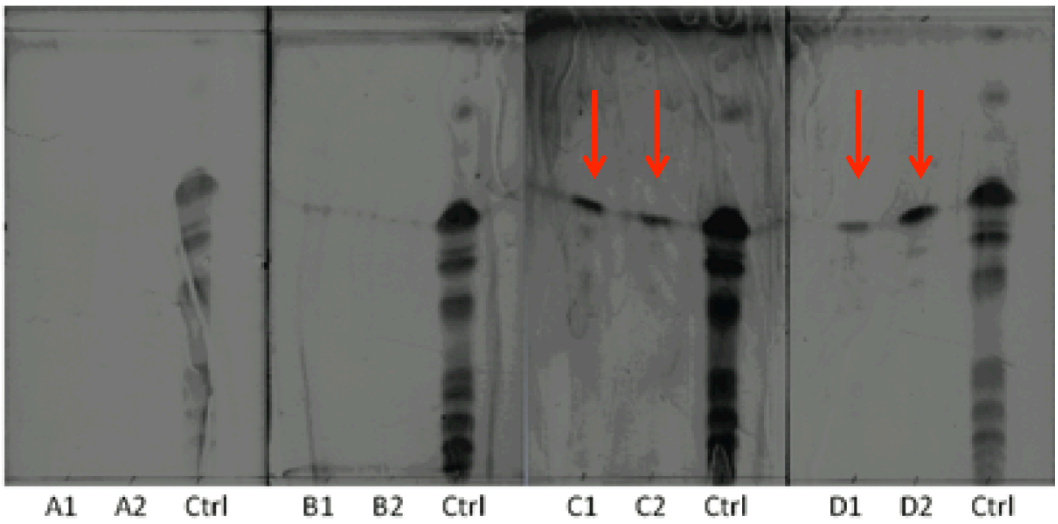


Figure 3.2 The presence of sophorolipids was tested by TLC in the medium of each time point. A1 and A2 correspond to 18h of growth time, B1 and B2 to 30h, C1 and C2 to 72h while D1 and D2 correspond to 100h. Sophorolipids are indicated with red arrows. Ctrl- control composes of the mixture of sophorolipids.

3.2 Gene ontology analysis of the exoproteome

All identified proteins were grouped according to their functional annotation (**Fig. 3.3**).The majority of identified proteins are proteases (25%). 11% is constituted of proteins involved in the cell wall organization. 9% are described as a stress related proteins and glucanases, respectively. 4% appears for glucanosyltransferases while 7% for glycosidase (like invertase). Other interesting proteins belong to lipid-related proteins (7%) and phosphatases (5%). We grouped GO classes with only a

single representing protein into the "other" group (5%). 18% of the proteins are labeled as "Unknown".

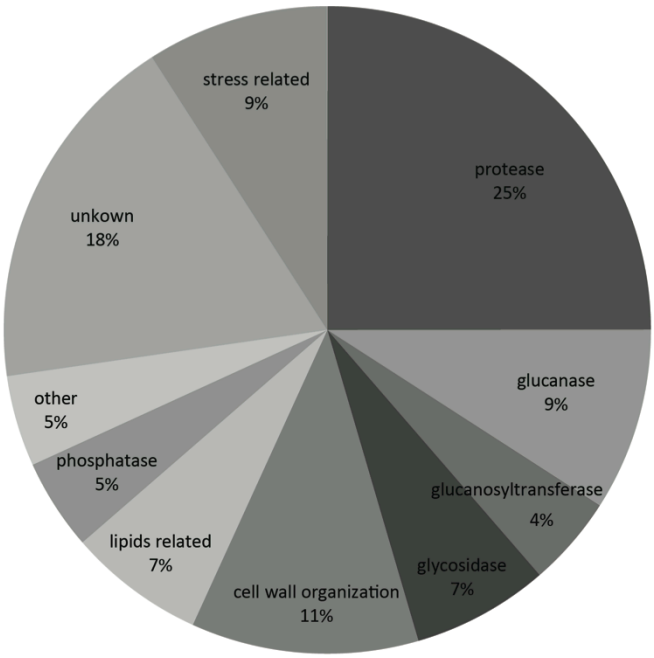


Figure 3.3 List of proteins identified in the *S. bombicola* exoproteome grouped according to their function.

3.3 Comparison between exponential and stationary phase extracellular proteins

In total, 34 proteins were identified in the exponential phase (point A and B) from which 6 were exclusively present in this growth phase (**Table 3.3**). Among these is an enzyme known to be more active in the exponential phase in baker's yeast, leucine aminopeptidase (Lap1) [197], and two proteins that are ER resident and involved in protein folding (Kar2,Pdi).

From the 38 proteins identified in the stationary phase, 10 were phase specific. Most proteins are recognize as secreted proteases: carboxypeptidase Y (Cpy1) and 4 candidapepsin homologues (Sapt1). Other proteins are known to allow the recovery of essential metabolites, e.g. phosphatase, 3-phytase B (Phy3), a nonspecific endoribonuclease (Rny1) and an invertase homologue (Suc2).

Table 3.2 Proteins identified in the exoproteome in *S. bombicola* were *in silico* analysed for secretion signal, GPI-anchor and localization. Proteins were grouped according to their predicted function. Annotation of the *S. bombicola* genes is based on similarities with genes from species: *1-Aspergillus fumigatus,* 2-*Aspergillus niger,* 3-*Aspergillus oryzae,* 4-*Bacillus subtilis,* 5-*Candida albicans,* 6-*Candida glabrata,* 7-*Candida tropicalis,* 8-*Escherichia coli (strain K12),* 9-*Humicola insolens,* 10-*Prunus dulcis,* 11-*Saccharomyces cerevisiae,* 12-*Schizosaccharomyces pombe,* 13-*Yarrowia lipolytica* (More details in Suppl. Table 3.1). The label "[a]" and "[b]" indicates proteins in common with biotechnologically or medically important yeast respectively.

Genome nr	Sort name	Protein	Predicted signal peptide (SP)	Predicted GPI anchor (GPI)	Subcellular localization (LocTree2)	Function
cabom01g03830	11Axl2	Protein AXL2	yes	no	secreted	cell wall organization
cabom01g06580	6Kre9	Cell wall synthesis protein KRE9	yes	no	secreted	cell wall organization
cabom02g08250	11ECM33[ab]	Cell wall protein ECM33	yes	yes	secreted	cell wall organization
cabom02g08390	11Tos1[a]	Protein TOS1	yes	no	secreted	cell wall organization
cabom03g08560	11SIM1[b]	Protein SIM1	yes	no	secreted	cell wall organization
cabom01g01190	11Dse4	Endo-1,3(4)-beta-glucanase 1	yes	no	er membrane	glucanase
cabom01g03300	11Bgl2[a]	Glucan 1,3-beta-glucosidase	yes	no	secreted	glucanase
cabom01g06180	5Xog1	Glucan 1,3-beta-glucosidase	yes	no	secreted	glucanase
cabom02g06210	11Scw4[ab]	Probable family 17 glucosidase SCW4	yes	no	secreted	glucanase
cabom02g02320	11Gas1[ab]	1,3-beta-glucanosyltransferase GAS1	yes	yes	secreted	glucanosyltransferase
cabom03g05170	11Gas3	Probable 1,3-beta-glucanosyltransferase GAS3	yes	yes	secreted	glucanosyltransferase
cabom01g00590	11Cts1[ab]	Endochitinase	yes	no	secreted	glycosidase
cabom01g09430	11Suc2[a]	Invertase	no	no	cytosolic	glycosidase
cabom03g02060	11Crh1[ab]	Probable glycosidase CRH1	yes	yes	secreted	glycosidase
cabom02g01720	5Lip1	Lipase 1	yes	no	secreted	lipids related
cabom02g11140	11Plb3	Lysophospholipase 3	yes	yes	plasma membrane	lipids related
cabom03g07360	11Pry2[ab]	Protein PRY2	no	no	secreted	lipids related
cabom01g03900	4ybbD	Uncharacterized lipoprotein ybbD	no	no	cytosolic	other
cabom02g06860	12SPCC285.05	Uncharacterized protein C285.05	yes	no	secreted	other

cabom02g01730	2PhyB	3-phytase B	yes	no	secreted	phosphatase
cabom03g14480	2PhyB	3-phytase B	yes	no	secreted	phosphatase
cabom01g03660	7Sapt1	Candidapepsin	yes	no	secreted	protease
cabom01g03670	7Sapt1	Candidapepsin	yes	no	secreted	protease
cabom01g03740	7Sapt1	Candidapepsin	yes	no	secreted	protease
cabom01g08280	5Cpy1	Carboxypeptidase Y	yes	no	secreted	protease
cabom01g10760	7Sapt1	Candidapepsin	yes	no	secreted	protease
cabom02g01580	11Yps3	Aspartic proteinase yapsin-3	yes	yes	er membrane	protease
cabom02g08790	11MKC7	Aspartic proteinase MKC7 (YPS2)	yes	yes	secreted	protease
cabom02g09480	11Ape3	Aminopeptidase Y	no	no	cytosolic	protease
cabom02g09720	3Lap1	Leucine aminopeptidase LAP1	yes	no	secreted	protease
cabom02g10300	5Sap5	Candidapepsin-5	yes	yes	secreted	protease
cabom03g06170	10PNGase A	Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A	yes	no	secreted	protease
cabom01g14090	13Rny1	Ribonuclease T2-like	no	no	secreted	stress related
cabom02g01450	13Kar2[b]	78 kDa glucose-regulated protein homolog	yes	no	er	stress related
cabom02g09850	9Pdi	Protein disulfide-isomerase	yes	no	er	stress related
cabom02g11610	3KatG	Catalase-peroxidase	yes	no	cytosolic	stress related
cabom01g02570		no-hit	yes	no	plasma membrane	unknown
cabom02g00130	3HsbA	Hydrophobic surface binding protein A	yes	no	secreted	unknown
cabom02g04820		no-hit	yes	no	secreted	unknown
cabom02g07380	3HsbA	Hydrophobic surface binding protein A	yes	yes	peoxisome	unknown
cabom02g09580	1AfuA	Allergen nr 7	yes	no	cytosolic	unknown
cabom02g13830	8YncE	Uncharacterized protein YncE	yes	yes	cytosolic	unknown
cabom03g09030		no-hit	yes	yes	er membrane	unknown
cabom03g11880		no-hit	yes	no	er membrane	unknown

Table 3.3 Heat map visualizing the proteins present in the transition from exponential to stationary phase. The number of identified unique peptides is shown for every replica (1, 2) at the 4 time points (A ,B, C, D). Proteins are ordered according to their abundance in the growth phases. A green colour of the cell indicates that the protein is present and identified with at least 2 unique peptides, a blue colour means that only 1 unique peptide was identified while the red colour illustrates that no peptides were detected.

Genome nr	Protein name	Exponential phase				Stationary phase			
		A		B		C		B	
		1	2	1	2	1	2	1	2
cabom02g01450	78 kDa glucose-regulated protein homolog	21	11		2				
cabom02g09480	Aminopeptidase Y	5	2						
cabom02g09720	Leucine aminopeptidase LAP1	4	7	5	6				
cabom02g09850	Protein disulfide-isomerase	3	4	1					
cabom02g13830	Uncharacterized protein YncE	3		6	5				1
cabom01g02570	no-hit	2	3	4	4			1	
cabom02g06210	Probable family 17 glucosidase SCW4	8	9	11	10		1	1	1
cabom02g08390	Protein TOS1	2	1	1	2		1	1	2
cabom02g08790	Aspartic proteinase MKC7 (YPS2)	2	1	1	1	1	1	1	1
cabom03g08560	Protein SIM1	5	6	3	5	4	8	7	5
cabom02g07380	Hydrophobic surface binding protein A	3	3	3	3	3	3	3	3
cabom02g02320	1,3-beta-glucanosyltransferase GAS1	3	3	5	2	5	5	6	5
cabom01g06180	Glucan 1,3-beta-glucosidase	11	6	16	17	17	21	17	22
cabom01g01190	Endo-1,3(4)-beta-glucanase 1	11	5	21	17	32	35	32	31
cabom03g05170	Probable 1,3-beta-glucanosyltransferase GAS3	6	4	6	4	7	7	8	8
cabom01g00590	Endochitinase	3	1	4	4	11	9	10	10
cabom02g10300	Candidapepsin-5	6	5	10	7	4	9	7	8
cabom03g02060	Probable glycosidase CRH1	5	2	5	5	7	8	7	9
cabom03g07360	Protein PRY2	4	3	3	4	3	4	3	4

cabom02g11140	Lysophospholipase 3	2	1	7	5	9	12	11	11
cabom02g06860	Uncharacterized protein C285.05	3	3	4	4	4	5	3	3
cabom02g08250	Cell wall protein ECM33	1	1	2	2	4	4	4	5
cabom03g14480	3-phytase B	2		2	1	1	2	3	1
cabom02g00130	Hydrophobic surface binding protein A	1		1	2	1	1	1	1
cabom02g09580	Allergen nr 7	2	1	2			1	1	2
cabom03g11880	no-hit	2	2	4	3		2	2	
cabom03g06170	Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A	3		4	1	1	7	9	6
cabom03g09030	no-hit	1		1	1	1	1	1	1
cabom02g11610	Catalase-peroxidase			2	1		7	7	
cabom01g03300	Glucan 1,3-beta-glucosidase	1	2	4	12		13	8	14
cabom02g01720	Lipase 1		1	4	4	4	8	5	4
cabom02g01580	Aspartic proteinase yapsin-3			2		3	3	2	2
cabom01g06580	Cell wall synthesis protein KRE9	1		1			1	1	1
cabom01g03830	Protein AXL2	1		1			2	2	1
cabom02g01730	3-phytase B						3	2	
cabom01g09430	Invertase			1			5	8	5
cabom01g14090	Ribonuclease T2-like					2	6	2	5
cabom02g04820	no-hit					1	2	2	1
cabom01g03900	Uncharacterized lipoprotein ybbD					2	1	2	1
cabom01g08280	Carboxypeptidase Y					17	16	14	17
cabom01g10760	Candidapepsin					9	8	6	9
cabom01g03670	Candidapepsin					1	3	1	2
cabom01g03660	Candidapepsin					1	2	1	2
cabom01g03740	Candidapepsin					1	1	1	1

3.4 Semiquantitative protein abundance analysis

We used the emPAI factors obtained for each protein in the different samples as a semiquantitative measure of the protein amount [192]. In our experiment, we calculated the average of the emPAI factors from both replicas per time point (A, B, C, D). We also expressed the protein content in [mol%], by dividing the average emPAI factor per protein by ∑emPAI values of all proteins identified with 2 unique peptides in that time point (**Suppl. Table 3.2**). The calculation of these values shows that glucan 1,3-beta-glucosidase (Xog1) is the most abundant protein in all 4 time points. The family 17 glucosidase Scw4 and Leucine aminopeptidase (Lap1) enrich the exoproteome at the time points representing exponential phase, while carboxypeptidase Y (Cpy1) is clearly more abundant at the stationary phase. However most of the proteins have a similar [mol %] content in both growth phases. There are a few examples presenting a small increased or decreased level which are summarized in **Figure 3.4**.

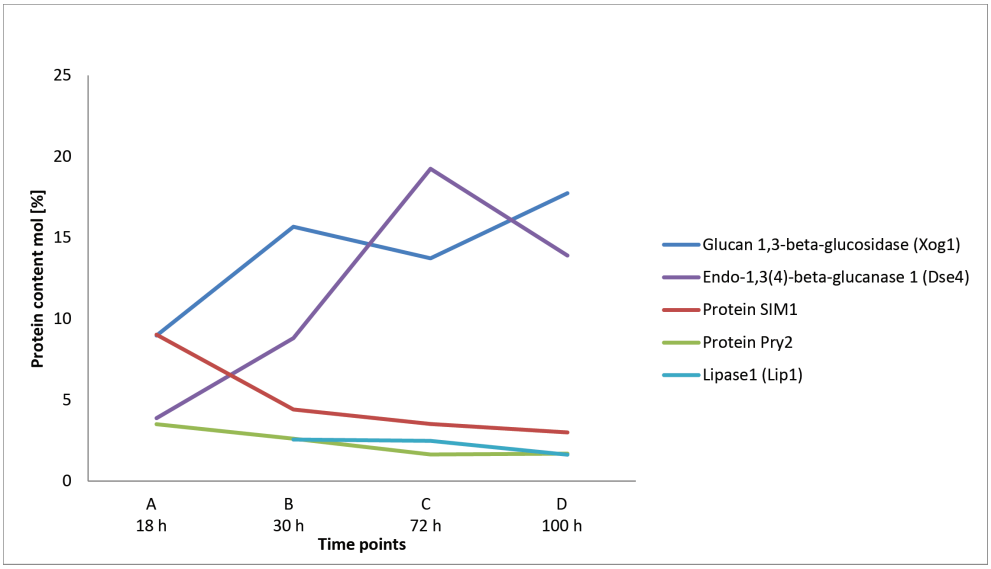


Figure 3.4 Selected examples of protein expression dynamics in the *S. bombicola* exoproteome. Most of the proteins have a stable expression during transition from exponential to stationary phase (in all 4 time point), e.g. Pry1 and Lip1. However, there are a few proteins of which abundance levels (Dse4 and Xog1) are higher in later growth phase, while other protein abundance (SIM1) decrease in function of time.

3.5 Exoproteome comparison among yeast species

Despite the fact that several yeast species that produce biosurfactants are discovered, their exoproteome is not yet described. We therefore compared the *S. bombicola* exoproteome with other examples available in literature, typically of widely used biotechnological or clinical relevant species. Over the last years the secretomes of several yeast were analysed, including *Pichia pastoris* [108], *Kluyveromyces lactis* [109, 110], *Candida utilis* [111], *C. albicans* [112] and *C. glabrata* [113]. Evidently, the composition of the exoproteome depends on the growth conditions, which were not identical in all described examples. However, in all examples glucose was the main carbon source, except the study of the *C. albicans* where sucrose was used. When we compared the *S. bombicola*

exoproteome with the exoproteomes of biotechnologically important yeast (*K. lactis*, *C. utilis* and *P. pastoris)* we found 5 proteins in common for all four species and 5 proteins were shared in 3 out of 4 those species (**Fig. 3.5 A**). These proteins are mainly connected with cell wall maintenance and include glucanases (Bgl2, Ssw4, Crh1, Dse4), chitinase (Cts1), transglucosylase (Gas1), and two cell wall proteins with unknown function (Tos1 and Pry1 or its isoform Pry2). The *S. bombicola* exoproteome has 8 proteins which also can be found in two pathogenic species: *C. albicans* and *C. glabrata* (**Fig. 3.5 B**). As in the earlier group, most of these proteins are members of the cell wall maintenance proteins and seem thus to be common for all yeasts. Only two proteins are exclusive for *Candida* and *Starmerella* species: Kar2 and Sim1. The proteins in common for each group are indicated in **table 3.2** with letters "[a]" and "[b]".

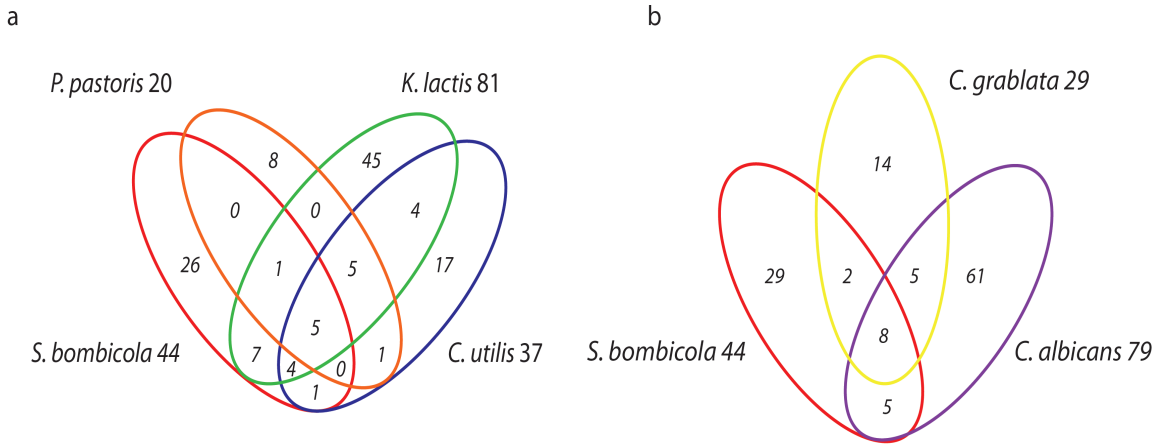


Figure 3.5 Comparison of *S. bombicola* exoproteome with other biotechnologically (a) and medically (b) important yeasts.

3.6 Invertase

In the exoproteome of *S. bombicola,* grown on glucose, we detected an invertase homologue exclusively present in the stationary phase. To demonstrate the activity of this enzyme we performed a colorimetric assay on the exoproteome collected at 72h. We detected a low activity corresponding to 0,244 U/L (specific activity=6.87 $*10^{-5}$ U/mg), where one unit of invertase catalyzes the formation of 1 µmole glucose per min at pH 4.5 and 30°C, under the assay conditions. This is a rather low activity compared to similar enzymes from other species. For example in the exoproteome of *Pichia* sp., an invertase activity of 10-40 U/L was detected [198] while in the excellent invertase producer, *S. cerevisiae Nncim 3287,* an invertase activity of 300 000 U/L and higher has been reported [199]. However, it is worth to underline that in these studies the invertase was expressed at optimized expression conditions, in contrast to our experiment. Regarding the medium composition, it was found that invertase secretion is strongly stimulated by the presence of sucrose [200].

As the identified protein does not possess a predicted secretion signal, we tried to reveal its extracellular character by confirming its glycosylation pattern. When the exoproteome, isolated from cultures collected for 72h, was treated with PNGase F the invertase was identified in bands

with a molecular weight lower than the original samples (bands 1, 5 and vs band 1 only) as shown in **Figure 3.1.**

3.7 Discovery of missing lactone esterase

The stationary and exponential exoproteomes were compared to detect proteins specific for the latter phase. Previous proteomic analysis of the intracellular protein fraction confirmed that the sophorolipid gene cluster is induced during transition to the stationary phase. Surprisingly, no protein with a potential lactonase/esterase/lipase function was found to be more abundant in the stationary phase. We then searched the datasets for proteins that have potential lipid binding or lipid metabolic functions. Three such proteins were observed, all of them with a stable protein abundance at the 4 time points. Two proteins seem not to be relevant to the SL production. The first protein is homologous to lysophospholipase 3 (Plb3), a cell wall bound phospholipase while, the second, homologous to PRY2, is a sterol binding protein involved in the export of acetylated sterols. The most interesting protein seemed to be a lipase 1 (Lip1) homolog that shows 33% identity to *Pseudozyma antarctica* lipase A. *Ps. antarctica* lipase A is an enzyme of high interest in biotechnology because it is active in a wide temperature and pH range. It prefers a long-chained carboxylic acid as a substrate and poorly recognizes short acyl groups such as acetate and butyrate [201].

To investigate the possible role of the Lip1 homolog in sophorolipid biosynthesis, we created a knock out strain of this protein. We have grown Lip1 KO strain under similar growth conditions as the wild type and collected the extracellular medium at the early and late stationary phase. Analysis of the sophorolipid extract using HPLC revealed a complete absence of any lactonic sophorolipid; exclusively acidic sophorolipids were produced (**Fig. 3.6**). This indicates that this enzyme is the 'missing' lactone esterase responsible for the lactonization of the acidic sophorolipids and that lactonization is likely to occur extracellularly as proposed by Hommel *et al*. (1994).

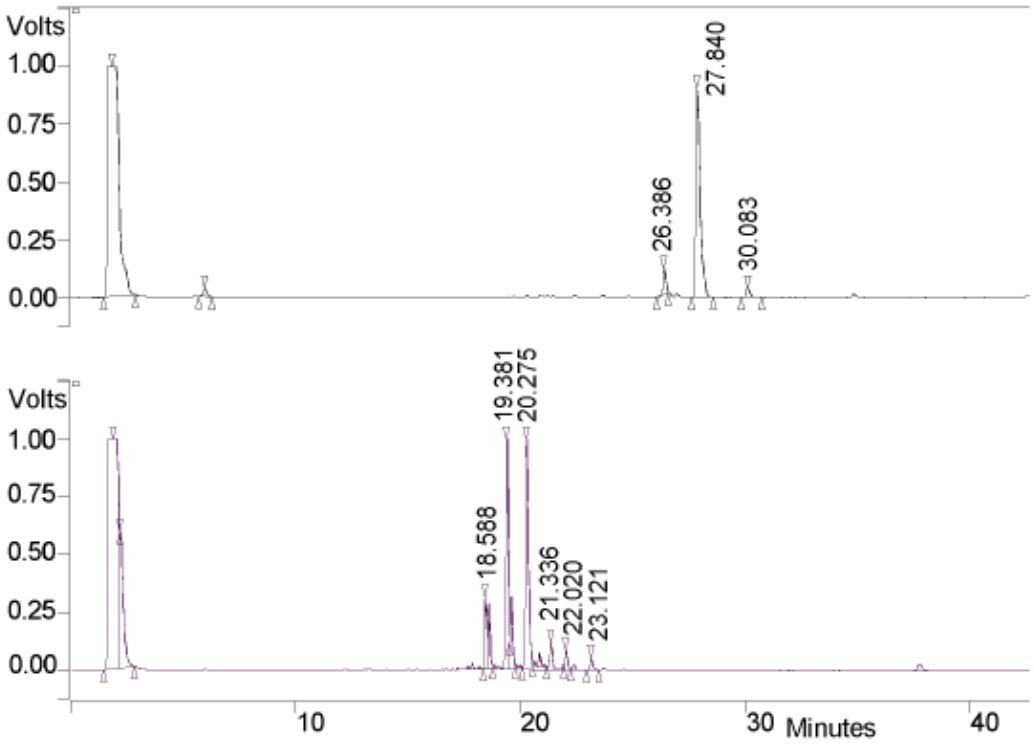


Figure 3.6 HLPC-ELSD chromatogram of sophorolipids produced by the wild-type (upper) and lactonase knock-out strain (lower) without the addition of rapeseed oil. Lactonic sophorolipids elute between 25 and 31 minutes, acidic ones between 17 and 24 minutes. LC-MS analysis identified the peaks at 19.4 and 20.3 min as non-acetylated acid sophorolipids with a C18:1 and C18:0 fatty acid chain respectively. The peaks between 20.4 and 23.6 are originate from mono-and diacetylated acidic sophorolipids. The peak at 27.8 correspondss to di-acetylated lactonic C18:1 sophorolipids.

4 Discussion

4.1 Comparison with other exoproteomes

Since the exoproteome of the biosurfactant producing strains was not yet described, it was unknown whether the extracellular protein compartment changes upon secretion of these molecules. To check this theory the *S. bombicola* exoproteome was compared with exoproteomes described in the literature of biotechnologically and medically relevant yeasts, such as *K.* lactis [110], *P.* pastoris [108] *C.* albicans [112] and *C.* utilis [111] . Although the experimental set-up (pH, temperature) and the strategy for proteomic analysis (2D-PAGE or LCMS) varied, a common core for the secretome of these species was proposed [111]. Many *S. bombicola* proteins belonging to that group were also identified. These proteins are mainly associated with cell wall construction and reorganization. We did not discover many distinct proteins in comparison with non-biosurfactant producing strains. It is possible that the same proteins have a different function or activity in the presence of biosurfactants

like sophorolipids, although this still needs further analysis. When a comparison of the *S. bombicola* exoproteome was performed in the context of its size, it was found to be mostly similar to that of the non-pathogenic yeast *C. utilis.* In general, in all described yeast species in **Table 3.4** the number of predicted secreted - containing an N-terminal secretion signal - and identified secreted proteins diverges substantially. It seems that the amount of the predicted secreted proteins is always overestimated. This may confirm the plasticity of the secretome, which can dynamically change and adapt to the new conditions. On the other hand, a subset of the predicted secreted proteins may also comprise proteins of the surfome, which are not always identified in the medium. In case of *S. bombicola* the percentage of identified to predicted proteins is quite high in comparison to other species. It can be explained by the fact that in our experiments 4 different time points were analysed, from which 16 proteins were specific only for a single growth phase, while in other studies usually only two conditions were tested.

Table 3.4. Comparison of *S. bombicola* with exoproteomes of other yeast species described in the literature.

	Pichia pastoris [108]	***Kluyveromyces lactis [110]***	***Candida utilis [111]***	***Candida albicans [112]***	***Starmerella bombicola***
Size of the genome	9.4 Mb	10.7Mb	12.5Mb	14.8Mb	9.4Mb
Predicted ORF	5450	5076	6417	6135	4617
Predicted secreted proteins	88	178	403	283	64
Number of detected proteins	20	81	37	79	44
Number of detected proteins with secretion signal	14	57	24	44	39
% of detected proteins with secretion signal	16	32	6	16	61

4.2 Characterization of the exoproteome

Most of the identified proteins in the exoproteome of *S. bombicola* have a predicted secretion signal and an extracellular localization. These proteins are thus secreted by the classical pathway, where they are transported from the ER to the Golgi apparatus and subsequently to the cell surface via Golgi-derived vesicles. A few identified proteins are predicted to possess a secretion signal for intracellular localization. Two of them are localized in the ER: 78 kDa glucose-regulated protein

homolog (Kar2) and protein disulfide-isomerase (Pdi). Kar2 is involved in correct protein folding and secretion [202] while Pdi catalyses the formation and rearrangement of disulphide bonds. Those proteins are chaperones which participate in the unfolded protein response. Identification of those proteins in secretomes was previously described for *K. lactis* and *C. albicans*, and may also originate from co-transport with their target proteins into secretory vesicles [110, 112].

We identified a few proteins that do not possess a predicted secretion signal. They could be transported to the cell surface and outside the cell via non-classical pathways where a secretion signal is not mandatory [97]. For example, we found a cell wall bound ribonuclease-T2 (Rny1) [203] and PRY2, a sterol-binding protein homolog, previously described to be secreted [204]. Another protein is aminopeptidase Y, it is a vacuolar enzyme which presence was also confirmed in the secretome of *K. lactis* [109]. Invertase in *S. cerevisiae* has two isoforms one cytoplasmic and one extracellular, both encoded by the *suc2* gene. Those isoforms are translated from two distinct, differentially regulated mRNAs, which differ only in their 5'-ends. The 1.9 kb mRNA encodes the external form and specifies a leader peptide which directs the protein product into the secretory pathway. This signal sequence is missing in the 1.8 kb mRNA encoding the internal invertase [205]. The invertase in *S. bombicola* does not have a predicted secretion signal but we demonstrated it is glycosylated (see below).

We also detected non-covalently bound cell wall proteins like Scw4 and Bgl2, which could be released in the medium by shaking culture flasks. We also identified many proteins predicted to contain a GPI-anchor, *i.e.* lysophospholipase 3(Plb3), chitin transglycosidase (Crh1), two aspartic proteases (YPS3, MKC7) and 2 glucanosyltransferases (Gas1, Gas3). Those proteins could appear in the medium due to proteolytic enzyme action [110]. In *S. cerevisiae* aspartyl proteases (yap) take part in that process [206] and in *S. bombicola* we also identified yap proteases (Yps3, MKC7). However in *C. albicans* it was suggested that GPI-anchor proteins could be released to the medium due to cell wall remodelling [112].

4.3 Phase specific proteins

In our experiment, we used a semiquantitative method to compare the exoproteome during the exponential and stationary growth phases. Yet, most of the identified proteins were found in both phases. Specific for the exponential phase, two aminopeptidases and two ER localized stress related proteins (Kar2 and Pdi), were identified. Interestingly, endoribonuclease T2 (Rny1) is found only in the stationary phase. Rny1 plays a protective role in plants [207] and can also act as a cytotoxin [208]. On the other hand, in *S. cerevisiae,* Rny1 was proposed to regulate membrane permeability and stability [203]. Specific for the *S. bombicola* stationary phase are also proteases and proteins connected with nutrient extraction, such as phytase (Phy3) and invertase (Suc2). The presence of an invertase homologue in *S. bombicola* exoproteome is intriguing since our growth medium contained a high amount of the glucose, while in the case of *S. cerevisiae*, its extracellular invertase is only expressed under glucose-limiting conditions and its expression is regulated by the carbon catabolism repression. However, in some non-conventional yeast (*K. lactis* and *P. anomala*), high osmolarity does not result in an inhibitory effect on the derepression of the invertase biosynthesis [209]. The molecular weight (MW) of the invertase from *S. bombicola* is similar to the hyperglycosylated extracellular form of *S. cerevisiae* invertase with a molecular weight (MW) of 250 kDa. After

deglycosylation the MW of the invertase from *S. bombicola* is reduced to 50kDA, similar to MW of non-glycosylated (intracellular) invertase of *S. cerevisiae*. In some strains, differently processed forms of invertase are found, but in *S. bombicola* only one form was detected.

S. bombicola is a nectarivorous yeast found in sucrose-rich environments, such as honey ,nectar of the flowering plants and bee hives [210]. It is proposed that the yeast cells are passively transported to the flowers by bumblebees. There, they proliferate and degrade complex sugars to monomers. The presence of an invertase in the exoproteome could be therefore connected with the natural habitat of *S. bombicola* where its expression is possibly utilized to extract nutrients.

4.4 GO

Classification of the identified proteins according to their functional annotation reveals that most proteins belong to proteases which is not typical for a biotechnological important species. Therefore, we believe that *S. bombicola* would not be a good host for protein expression. However, we also find many proteins typical for other yeast exoproteomes related to the carbohydrate conversion. The Probable family 17 glucosidase Scw4 possibly plays a role in cell expansion during growth and in cell-cell fusion during mating. It belongs to the glycosyl hydrolase 17 family, like the glucan 1,3-beta-glucosidase Bgl2, a major protein of the cell wall that is involved in cell wall maintenance [211]. The Sim1 protein belongs to the SUN family [112] and may play a role in cell wall morphogenesis and septation. It is also involved in aging, oxidative stress response, and in the regulation of mitochondrial biogenesis. The 1,3-beta-glucanosyltransferase (Gas1) is required for cell wall assembly and also has a role in transcriptional silencing. Additionally, endochitinase (Cts1) was detected, which is required for cell separation after mitosis [212]. Other proteins related to the cell wall include probable glycosidase Crf2, protein TOS1 and aspartic proteinase yapsin-3 Yps3. All these proteins have the particularity to be present in all analysed growth phases.

4.5 Lactone esterase

When the *Starmerella bombicola* sophorolipids cluster genes were analyzed, no enzyme responsible for the lactonisation was discovered [53]. In order to try to identify an enzyme responsible for efficient lactonization, the annotated genome of *S. bombicola* was screened for the presence of putative lipases. One of the 9 putative lipases was evaluated as a lipase belonging to Class 3 which is composed of enzymes that are not closely related to other lipases and therefore might possess other (secondary) activities. However, knocking out this gene in *S. bombicola* did not result in the partial or complete loss of the lactonizing abilities of the mutant and the mutant did not lose its ability to hydrolyze rapeseed oil triglycerides as well (personal communication with Dr. IN. Van Bogaert). Also in other species producing biosurfactants such an enzyme was not known.

A: Lactone esterase nucleotide sequence

ATGCTGGCTCTGTTTTTTTCGCTTGCGCCTCTACTTTCTCAAGCTCTCCCTTTAGGCTATACTGCGGCCCCCGCTGAATCATT
CTATTTTTGGCCAGAGAACATATCCAGCCTCCAAGCTGGCGAGATTTTTAGAAAACGGGAACTCTTAACTCTCCCAGACAT
CTTTGACTTTGGCCCTAATCTGGAAAAGGTCGTACAAGTGGCTTACAAAACCCGTCTCACCGATGGCAATGACTCGTTTTC
CATCGCCAGTATCTTTATCCCTAAGAATCCAAGCCCAGAACTCAAACTTTACTCTTATCAGACGTTTGAGGATGCCGTGCAG
CTTGATTGTGCCCCAAGCTATGCTTTAGAAGTGGGTAACAAGTCCAGCAACTATCTTCCTGTCACTAGCAATTTATCTGCCA

TCAGTCGAGAACTTGAGAAAGGACGTCACTGCATTATCCCTGATCACGAGGGCTATATTTCAGGATTCTTTGCAGGACGG
CAGGAGGGATATGCTGGTTTAGACGGAATTCGCGCTGCTCGAAACTATCTCAATGGCACCAACGAGACCCCAATTGGTAT
CTTCGGATACAGTGGAGGTGCACAAGCAACGGCCTGGATTGTTGATTTGCATGACGAGTATGCTCCTGACTTGAACTTTG
TTGGAACAGTTTCTGGAGGCACTTTGGTTGACGCTTGGGGCACTTTTCAGTATATCGACTATCCGAAGGTGTATCTAAAGG
GCAGCATTCTTATCATGTATACGGGTCTTTTTTCAGGTTATCCAGCTCAATTTGAGGTGATTTGGCCATATATTGAGCCTGT
AATTCAAGAAAACATGCTACTGCTACGTTTGGCGCCGAATGATTGTAACCAAAGCCCGATACTTCAAGGTTACAACAATTC
AATCATGGCCGGTATACATGTGGACCTTCCCGAATTCCCTGCTTCTAAGTACATATTCCAGCACGAGTCCCTCCTTGCCAAC
TACAGCGTAGTGCCAGTTTCCACACCGAAGTTTCCTCGCTACATGTACCATGGTGGATCTGATGAGTTGGCCAAATTGAGC
CTTGTCGAGCAGTATGTTGATCAACAATGGAATACCGGCGCTAATCTCACCTTCGTGGTGTATCCGGGTCTTCTTCATGAC
GAGACGGCTTACCGTGGCTTTGATGCCGCGATGGATTGGCTTGATGCCCAGCTCGATAGTGGATACCTTCCACCTGTAAA
CTCAACTCATACATGA

B: Lactone esterase amino acid sequence

MLALFFSLAPLLSQALPLGYTAAPAESFYFWPENISSLQAGEIFR**KRELLTLPDIFDFGPNLEKVVQVAYK**TRLTDGNDSFSIASIFI
PK**NPSPELK**LYSYQTFEDAVQLDCAPSYALEVGNKSSNYLPVTSNLSAISRELEKGRHCIIPDHEGYISGFFAGR**QEGYAGLDGIR**
AARNYLNGTNETPIGIFGYSGGAQATAWIVDLHDEYAPDLNFVGTVSGGTLVDAWGTFQYIDYPKVYLKGSILIMYTGLFSGYP
AQFEVIWPYIEPVIQENMLLLRLAPNDCNQSPILQGYNNSIMAGIHVDLPEFPASKYIFQHESLLANYSVVPVSTPKFPR**YMYHG**
GSDELAKLSLVEQYVDQQWNTGANLTFVVYPGLLHDETAYRGFDAAMDWLDAQLDSGYLPPVNSTHT

Figure 3.7 Nucleotide and amino acid sequence of lactone esterase from *Starmerella bombicola*. The underline amino acids representing predicted secretion signal. Bolded amino acids were experimentally detected by LC-MS analysis.

Our previous proteomic analysis of the intracellular proteome (Chapter 2) also did not reveal the identity of such a lactonic sophorolipid forming enzyme. Hommel *et al.* suggested, after studying the closely related *Candida apicola*, that such an enzyme would be a cell wall-bound lipase [40]. Indeed, there are many examples of *in vitro* esterification of hydroxylated fatty acids and even sophorolipids molecules using extracellular lipases [45, 130]. A potential candidate enzyme annotated as a putative lipase (Lip1) was found. *In silico* characterization revealed that this protein possesses a secretion signal (**Fig. 3.7**), but not a consensus sequence for GPI-anchoring to the cell wall. This enzyme shows 33% identity with a lipase form *Ps. antarctica*, lipase A (CALA). However CALA is hitherto not associated with the lactonisation of *Ps. antarctica* glycolipid products *i.e.* mannosylerythritol lipids (MEL). Surprisingly, a single knock-out of this lipase/esterase gene resulted in complete absence of any lactonic form of the sophorolipids. The missing enzyme for SL biosynthetic pathway was discovered. Moreover the lactonization was confirmed to occure extracellularly. More detailed characteristic of this enzyme can be found in the next chapter.

5 Conclusion

In recent years, the microbial exoproteome attracted increased attention from the biotechnological sector due to the interesting properties of the enzymes found there. Here, we present the first analysis of the exoproteome of a biosurfactant producer, *i.e. S. bombicola*. Besides identifying many proteins described in other yeast exoproteomic studies, two particular proteins caught our

attention: an invertase and a lactone esterase. The detection of an extracellular invertase is possibly connected to the natural 'sweet' habitats of S*. bombicola*, e.g. honey and nectar. However, in our *in vitro* assays the enzyme demonstrated a low activity and it seems unlikely that, even under optimal reaction conditions, the enzyme would perform better then the already commercialized invertase from *S. cerevisiae*. Nonetheless, the confirmation of its presence in the *S. bombicola* exoproteome is important for understanding the relationship between bumblebees and yeasts in flowering plants. It was already suggested that bees select yeast-containing flowers due to a predigesting function of yeasts, by which complex carbohydrates are transformed into more accessible forms for bees [210].

The important discovery of the elusive lactone esterase allows for a better control of industrial sophorolipid biosynthesis. Until now, it was not possible to produce sophorolipids in *S. bombicola* with a homogeneous structural composition. Here, we created a mutant that offers a one-step production technology for the fermentative synthesis of industrially important molecules, acidic sophorolipids, from cheap, renewable substrates. We anticipate that, due to the higher foaming capacity and better water solubility of the SL mixture produced by our engineered strain, this type of biosurfactants will find a broad range of applications in detergent, pharmaceutical and cosmetic industries.

Supplementary tables

Suppl. Table 3.1. Detailed information about proteins identified in the extracellular *S.bombicola* proteome.

Short name	Genome nr	Protein	Uniprot	E-Value	Description from Swiss Prot 2012/06 and Interproscan V.34.
Cts1	cabom01g00590	Endochitinase	P29029	8.00E-67	Endochitinase OS=Saccharomyces cerevisiae GN=CTS1 PE=1 SV=2 no-hit NULL no-hit
Dse4	cabom01g01190	Endo-1,3(4)-beta-glucanase 1	P53753	1.00E-141	"Endo-1,3(4)-beta-glucanase 1 OS=Saccharomyces cerevisiae GN=DSE4 PE=1 SV=1" Glyco_hydro_81 2.30E-199 Glycosyl hydrolase family 81
	cabom01g02570	no-hit			no-hit NULL no-hit no-hit NULL no-hit
Bgl2	cabom01g03300	Glucan 1,3-beta-glucosidase	P15703	9.00E-96	"Glucan 1,3-beta-glucosidase OS=Saccharomyces cerevisiae GN=BGL2 PE=1 SV=1" Glyco_hydro_17 7.60E-57 Glycosyl hydrolases family 17
Sapt1	cabom01g03660	Candidapepsin	Q00663	4.00E-42	Candidapepsin OS=Candida tropicalis GN=SAPT1 PE=1 SV=1 Asp 2.40E-51 Eukaryotic aspartyl protease
Sapt1	cabom01g03670	Candidapepsin	Q00663	2.00E-44	Candidapepsin OS=Candida tropicalis GN=SAPT1 PE=1 SV=1 Asp 1.50E-50 Eukaryotic aspartyl protease
Sapt1	cabom01g03740	Candidapepsin	Q00663	5.00E-36	Candidapepsin OS=Candida tropicalis GN=SAPT1 PE=1 SV=1 Asp 2.40E-46 Eukaryotic aspartyl protease
Axl2	cabom01g03830	Protein AXL2	P38928	4.00E-35	Protein AXL2 OS=Saccharomyces cerevisiae GN=AXL2 PE=1 SV=1 SKG6 4.10E-12 Transmembrane alpha-helix domain
ybbD	cabom01g03900	Uncharacterized lipoprotein ybbD	P40406	3.00E-87	Uncharacterized lipoprotein ybbD OS=Bacillus subtilis GN=ybbD PE=1 SV=1 Glyco_hydro_3 4.40E-48 Glycosyl hydrolase family 3 N
Xog1	cabom01g06180	Glucan 1,3-beta-glucosidase	P29717	1.00E-122	"Glucan 1,3-beta-glucosidase OS=Candida albicans GN=XOG1 PE=1 SV=4" Cellulase 9.50E-17 Cellulase (glycosyl hydrolase family 5)
Kre9	cabom01g06580	Cell wall synthesis protein KRE9	O74683	9.00E-05	Cell wall synthesis protein KRE9 OS=Candida glabrata GN=KRE9 PE=3 SV=1
Cpy1	cabom01g08280	Carboxypeptidase Y	P30574	6.00E-92	Carboxypeptidase Y OS=Candida albicans GN=CPY1 PE=2 SV=2 Peptidase_S10 5.70E-107 Serine carboxypeptidase
Suc2	cabom01g09430	Invertase	Q6BJW6	2.00E-74	Invertase 2 OS=Saccharomyces cerevisiae GN=SUC2 PE=1 SV=1 Glyco_hydro_32N 7.30E-75 Glycosyl hydrolases family 32 N-terminal
Sapt1	cabom01g10760	Candidapepsin	Q00663	7.00E-29	Candidapepsin OS=Candida tropicalis GN=SAPT1 PE=1 SV=1 Asp 1.50E-47 Eukaryotic aspartyl protease
Rny1	cabom01	Ribonuclease T2-like	Q6C	5.00E	Ribonuclease T2-like OS=Yarrowia lipolytica GN=RNY1 PE=3 SV=1 Ribonuclease_T2 8.20E-40 Ribonuclease

	g14090		AV7	-67	T2 family
HsbA	cabom02 g00130	Hydrophobic surface binding protein A			no-hit NULL no-hit HsbA 1.60E-07 Hydrophobic surface binding protein A
Kar2	cabom02 g01450	78 kDa glucose-regulated protein homolog	Q99 170	0.00E +00	78 kDa glucose-regulated protein homolog OS=Yarrowia lipolytica GN=KAR2 PE=3 SV=1 HSP70 4.80E-251 Hsp70 protein
Yps3	cabom02 g01580	Aspartic proteinase yapsin-3	Q12 303	2.00E -51	Aspartic proteinase yapsin-3 OS=Saccharomyces cerevisiae GN=YPS3 PE=1 SV=1 Asp 1.10E-64 Eukaryotic aspartyl protease
Lip1	cabom02 g01720	Lipase 1	O94 091	2.00E -28	Lipase 1 OS=Candida albicans GN=LIP1 PE=3 SV=2 LIP 9.80E-37 Secretory lipase
PhyB	cabom02 g01730	3-phytase B	P347 54	4.00E -86	3-phytase B OS=Aspergillus niger GN=phyB PE=1 SV=1 Acid_phosphat_A 2.80E-30 Histidine acid phosphatase
Gas1	cabom02 g02320	1,3-beta-glucanosyltransferase GAS1	P221 46	1.00E -139	"1,3-beta-glucanosyltransferase GAS1 OS=Saccharomyces cerevisiae GN=GAS1 PE=1 SV=2" Glyco_hydro_72 2.00E-132 Glycolipid anchored surface protein (GAS1)
	cabom02 g04820	no-hit			no-hit NULL no-hit no-hit NULL no-hit
Scw4	cabom02 g06210	Probable family 17 glucosidase SCW4	P533 34	1.00E -70	Probable family 17 glucosidase SCW4 OS=Saccharomyces cerevisiae GN=SCW4 PE=1 SV=1 Glyco_hydro_17 6.70E-07 Glycosyl hydrolases family 17
SPCC2 85.05	cabom02 g06860	Uncharacterized protein C285.05	O74 493	7.00E -52	Uncharacterized protein C285.05 OS=Schizosaccharomyces pombe GN=SPCC285.05 PE=2 SV=1 NUP 4.80E-111 Purine nucleoside permease (NUP)
HsbA	cabom02 g07380	Hydrophobic surface binding protein A			no-hit NULL no-hit HsbA 3.00E-07 Hydrophobic surface binding protein A
ECM33	cabom02 g08250	Cell wall protein ECM33	B5VE 42	8.00E -49	Cell wall protein ECM33 OS=Saccharomyces cerevisiae (strain AWRI1631) GN=ECM33 PE=3 SV=2 Recep_L_domain 2.70E-11 Receptor L domain
Tos1	cabom02 g08390	Protein TOS1	P382 88	6.00E -52	Protein TOS1 OS=Saccharomyces cerevisiae GN=TOS1 PE=1 SV=1 DUF2401 6.60E-91 Putative secretory protein (DUF2401)
MKC7	cabom02 g08790	Aspartic proteinase MKC7	P533 79	2.00E -41	Aspartic proteinase MKC7 OS=Saccharomyces cerevisiae GN=MKC7 PE=1 SV=2
Ape3	cabom02 g09480	Aminopeptidase Y	P373 02	1.00E -108	Aminopeptidase Y OS=Saccharomyces cerevisiae GN=APE3 PE=1 SV=1 Peptidase_M28 9.50E-31 Peptidase family M28
AfuA	cabom02 g09580	Allergen nr 7	O42 799	5.00E -07	Allergen Asp f 7 OS=Aspergillus fumigatus GN=AFUA_4G06670 PE=1 SV=2 DPBB_1 6.00E-13 Rare lipoprotein A (RlpA)-like double-psi
Lap1	cabom02 g09720	Leucine aminopeptidase LAP1	Q2PI T3	1.00E -114	Leucine aminopeptidase 1 OS=Aspergillus oryzae GN=LAP1 PE=1 SV=1 Peptidase_M28 1.90E-29 Peptidase family M28
Pdi	cabom02 g09850	Protein disulfide-isomerase	P550 59	7.00E -91	Protein disulfide-isomerase OS=Humicola insolens PE=1 SV=1 Thioredoxin 2.70E-63 Thioredoxin
Sap5	cabom02 g10300	Candidapepsin-5	P430 94	2.00E -07	Candidapepsin-5 OS=Candida albicans GN=SAP5 PE=1 SV=1 Asp 8.10E-35 Eukaryotic aspartyl protease

Plb3	cabom02g11140	Lysophospholipase 3	Q08108	1.00E-146	Lysophospholipase 3 OS=Saccharomyces cerevisiae GN=PLB3 PE=2 SV=1 PLA2_B 1.50E-186 Lysophospholipase catalytic domain
KatG	cabom02g11610	Catalase-peroxidase	Q67LP5	0.00E+00	Catalase-peroxidase OS=Aspergillus oryzae GN=katG PE=3 SV=1 peroxidase 2.10E-79 Peroxidase
yncE	cabom02g13830	Uncharacterized protein YncE/ecoli	P76116	1.00E-10	Uncharacterized protein YncE OS=Escherichia coli (strain K12) GN=yncE PE=1 SV=1 NHL 2.30E-06 NHL repeat
Crf1/2/Crh1	cabom03g02060	Probable glycosidase CRH1	P53301	3.00E-19	Probable glycosidase CRH1 OS=Saccharomyces cerevisiae GN=CRH1 PE=1 SV=1
Gas3	cabom03g05170	Probable 1,3-beta-glucanosyltransferase GAS3	Q03655	3.00E-92	"Probable 1,3-beta-glucanosyltransferase GAS3 OS=Saccharomyces cerevisiae GN=GAS3 PE=1 SV=1" Glyco_hydro_72 9.80E-113 Glycolipid anchored surface protein (GAS1)
PNGase A	cabom03g06170	Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A	P81898	1.00E-29	Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A OS=Prunus dulcis PE=1 SV=2 PNGaseA 4.60E-61 Peptide N-acetyl-beta-D-glucosaminyl asparaginase amidase A
Pry2	cabom03g07360	Protein PRY2	P36110	1.00E-23	Protein PRY2 OS=Saccharomyces cerevisiae GN=PRY2 PE=2 SV=1 CAP 6.60E-18 Cysteine-rich secretory protein family
Sim1	cabom03g08560	Protein SIM1	P40472	3.00E-98	Protein SIM1 OS=Saccharomyces cerevisiae GN=SIM1 PE=1 SV=1 SUN 2.20E-98 Beta-glucosidase (SUN family)
	cabom03g09030	no-hit			no-hit NULL no-hit no-hit NULL no-hit
	cabom03g11880	no-hit			No-hit NULL no-hit no-hit NULL no-hit
phyB	cabom03g14480	3-phytase B	P34754	3.00E-94	3-phytase B OS=Aspergillus niger GN=phyB PE=1 SV=1 Acid_phosphat_A 1.20E-32 Histidine acid phosphatase

Suppl. Table 3.2. emPAI factor and protein content [mol%] calculated by the Mascot per protein in every analysed time point.

Genome nr	Protein	Exponential phase				Stationary phase			
		A		B		C		D	
		emPAI	protein content [mol%]	emPAI	protein content [mol%]	emPAI	protein content [mol%]	emPAI	protein content [mol%]
cabom02g01450	78 kDa glucose-regulated protein homolog	1.15	10.60	0.05	0.26				
cabom02g09480	Aminopeptidase Y	0.17	1.57						
cabom02g09720	Leucine aminopeptidase LAP1	0.73	6.76	0.70	4.04				
cabom02g09850	Protein disulfide-isomerase	0.23	2.08						
cabom02g13830	Uncharacterized protein YncE/ecoli	0.12	1.11	0.49	2.85				
cabom01g02570	no-hit	0.22	1.99	0.33	1.92				
cabom02g06210	Probable family 17 glucosidase SCW4	2.06	19.02	2.78	16.14				
cabom02g08390	Protein TOS1	0.07	0.65	0.07	0.41			0.07	0.23
cabom02g08790	Aspartic proteinase MKC7 (YPS2)	0.10	0.88						
cabom03g08560	Protein SIM1	0.98	9.02	0.76	4.42	0.97	3.52	0.90	3.00
cabom02g07380	Hydrophobic surface binding protein A	0.84	7.73	1.10	6.40	1.24	4.52	0.74	2.47
cabom02g02320	1,3-beta-glucanosyltransferase GAS1	0.25	2.31	0.22	1.28	0.47	1.71	0.56	1.85
cabom01g06180	Glucan 1,3-beta-glucosidase	0.97	8.98	2.70	15.68	3.77	13.74	5.33	17.75
cabom01g01190	Endo-1,3(4)-beta-glucanase 1	0.42	3.89	1.52	8.81	5.28	19.24	4.17	13.90
cabom03g05170	Probable 1,3-beta-glucanosyltransferase GAS3	0.42	3.89	0.42	2.44	0.69	2.50	0.73	2.43
cabom01g00590	Endochitinase	0.09	0.83	0.25	1.45	1.24	4.52	1.24	4.13
cabom02g10300	Candidapepsin-5	0.50	4.63	0.73	4.22	0.53	1.92	0.67	2.22
cabom03g02060	Probable glycosidase CRH1	0.31	2.82	0.47	2.73	0.85	3.10	0.85	2.83
cabom03g07360	Protein PRY2	0.38	3.52	0.45	2.62	0.45	1.64	0.51	1.70
cabom02g11140	Lysophospholipase 3	0.05	0.46	0.39	2.24	0.75	2.74	0.79	2.62
cabom02g06860	Uncharacterized protein C285.05	0.33	3.05	0.38	2.21	0.49	1.79	0.24	0.80
cabom02g08250	Cell wall protein ECM33		0.00	0.16	0.93	0.44	1.61	0.57	1.90

cabom03g14480	3-phytase B	0.06	0.51	0.06	0.32	0.06	0.20	0.09	0.30
cabom02g00130	Hydrophobic surface binding protein A			0.36	2.09				
cabom02g09580	Allergen nr 7	0.12	1.11	0.12	0.70			0.12	0.40
cabom03g11880	no-hit	0.08	0.74	0.14	0.81	0.04	0.15	0.03	0.08
cabom03g06170	Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A	0.10	0.88	0.13	0.76	0.35	1.28	0.67	2.22
cabom02g11610	Catalase-peroxidase			0.04	0.20	0.16	0.58	0.14	0.47
cabom01g03300	Glucan 1,3-beta-glucosidase	0.11	0.97	1.93	11.23	2.33	8.50	5.02	16.71
cabom02g01720	Lipase 1			0.44	2.56	0.68	2.48	0.49	1.63
cabom02g01580	Aspartic proteinase yapsin-3			0.05	0.26	0.14	0.51	0.09	0.30
cabom01g03830	Protein AXL2					0.04	0.13	0.04	0.12
cabom02g01730	3-phytase B					0.09	0.33	0.06	0.2
cabom01g09430	Invertase					0.20	0.73	0.53	1.77
cabom01g14090	Ribonuclease T2-like					0.36	1.30	0.30	1.00
cabom02g04820	no-hit					0.11	0.38	0.11	0.35
cabom01g03900	Uncharacterized lipoprotein ybbD					0.05	0.18	0.05	0.17
cabom01g08280	Carboxypeptidase Y					4.33	15.80	3.77	12.56
cabom01g10760	Candidapepsin					1.20	4.36	1.10	3.65
cabom01g03670	Candidapepsin					0.14	0.49	0.09	0.28
cabom01g03660	Candidapepsin					0.11	0.38	0.05	0.17
SUM		10.805		17.19		27.5		30.065	

Suppl. Table 3.3. Detected peptides of the lactone esterase and their expect value and ion score calculated by Mascot version 2.3.01

Nr.	Score	Expect	Peptides
1	51	8.6e-006	K.NPSPELK.L
2	56	3.3e-006	K.VVQVAYK.T
3	91	1.3e-009	R.QEGYAGLDGIR.A
4	70	1e-007	R.YMYHGGSDELAK.L
5	36	0.00026	R.YMYHGGSDELAK.L
6	39	0.00014	R.YMYHGGSDELAK.L + Oxidation (M)
7	67	1.9e-007	R.YMYHGGSDELAK.L + Oxidation (M)
8	94	7.4e-010	R.ELLTLPDIFDFGPNLEK.V
9	62	1.2e-006	R.ELLTLPDIFDFGPNLEK.V
10	80	2.3e-008	K.RELLTLPDIFDFGPNLEK.V
11	73	9.4e-008	R.KRELLTLPDIFDFGPNLEK.V
12	43	6.8e-005	R.KRELLTLPDIFDFGPNLEK.V

Chapter 4

Lactone esterase of *Starmerella bombicola :* expression, purification and enzymatic properties

Ciesielska K.[a]; Van Bogaert IN.[b]; Roelands S.[b]; Vandenberghe I.[a]; Groeneboer S.[a]; Soetaert W.[b]; Devreese B.[a]

[a]Laboratory for Protein Biochemistry and Biomolecular Engineering, Ghent University, Ghent, Belgium

[b]Laboratory of Industrial Biotechnology and Biocatalysis, Ghent University, Ghent, Belgium

This chapter is included in patent application WO 2013/092421:" A lactonase derived from *Stramerella bombicola* and uses thereof"

Contribution: K. Ciesielska created a *Pichia pastoris* strain expressing lactone esterase and the lactone esterase Ser194Ala mutant. K. Ciesielska performed protein purification and enzymatic assay analysis. Dr. IN. Van Bogaert and Dr. S. Roelands prepared substrates for the enzymatic assay (sophorolipids) and performed HPLC and LC-MS analysis. I. Vandenberghe performed Edman degradation and assays with p-nitrophenyl esters.

Abstract

Lipases are enzymes that catalyse the hydrolysis of ester bonds in triglycerides. This reaction is thermodynamically favoured in an aqueous environment while in organic solvents, lipases can perform the reverse esterification reaction. In the context of lactonic sophorolipid formation, an intramolecular esterification occurs, even in an aqueous fermentation broth. The reaction mechanism for the lactonization of sophorolipids remained unclear: until recently, it was unknown whether it is the result of a spontaneous reaction or a true enzyme catalysed process. In the previous chapter, we described the discovery of a lipase-like enzyme in the *S. bombicola* exoproteome which appeared to be the missing lactone esterase in the sophorolipid pathway. Indeed, a knockout strain for this protein led to the production of only acidic sophorolipids. We thus showed that sophorolipid lactonization is an enzyme catalysed reaction but the mechanism of its action is still unclear. In order to characterize its enzymatic properties, the lactone esterase from *S. bombicola* was expressed in *Pichia pastoris* and purified in a two-step strategy. Consequently, its activity as a lipase and esterase was assayed. We also estimated its apparent K_m (app K_m) value for the original substrate, sophorolipids, and we identified the active site serine by site-directed mutagenesis. More structural analyses are required to elucidate the exact mechanism of the enzyme in detail but we here put forward a few hypotheses that could explain the catalysis of the seemingly entropically disfavoured sophorolipid lactonization.

Keywords : *Starmerella (Candida) bombicola* , sophorolipids, lactone esterase, natural product synthesis, enzymes

1 Introduction

Sophorolipids are the only glycolipidic biosurfactants which are present as a mixture of acidic and lactonized molecules (**Fig 1.1**). This variety in molecular conformation influences their biophysical and biological properties. For example, lactonic sophorolipids (SL) have a better capacity to lower the surface tension and have antimicrobial properties, whereas acidic SL display better foam production and solubility [213].

The lactonic form of a sophorolipid is a cyclic ester, resulting from an esterification reaction between the carboxyl group of the hydroxy fatty acid and a hydroxyl group of sophorose. In nature, such macrolactones (composed of more than 12 C-atoms) are rather rare because their formation is supposed to be difficult. Firstly, lactonisation has to compete with intermolecular esterification, which would result in polymerization. Secondly intramolecular esterification is entropically unfavorable when the lactone ring is composed of more than 7C-atomes [128]. Little is known about the actual enzymes that catalyse such reactions. However there are some examples of *in vitro* experiments where lipases perform esterification of hydroxy fatty acid methyl esters (see Chapter 1) and even sophorolipids [45, 214]. The best described enzymes are lipase P from *Pseudomonas sp* [123] and lipase B (CALB) from *Pseudozyma (*former *Candida) antarctica* [126]. The latter enzyme, commercially exploited as Novozyme 435, can also perform sophorolipid amide formation, selective acetylation/acylation [44] and macrolactone formation [45]. Curiously, in *in vitro* assays for sophorolipids with CALB, the esterification occurs at the 6'-position of sophorose, while during natural production most of the sophorolipids lactones are esterified at the 4'-position and only a minor fraction at the 6- or 6'-position [9]. In general, lipases evolved naturally to catalyze the hydrolysis of the ester bond in an aqueous environment, and only in organic solvents, the esterification reaction is favored. Consequently, it was completely unknown which enzyme could catalyze such a seemingly improbable reaction, sophorolipid lactonization, *in vivo* in aqueous conditions.

During a proteomic analysis, we recently discovered a lipase-like enzyme in the extracellular fraction of *S. bombicola* cultures. A *S. bombicola* knock-out strain for this enzyme was totally defective in producing lactonized sophorolipids (see Chapter 3). This enzyme shows 33% of sequence identity with another lipase form *Ps. antarctica*, lipase A (CALA). Hitherto, no CALA activity connected to sophorolipid modification has been described, in contrast to CALB. The enzymatic properties of CALA are less well explored then those of CALB but CALA was described to exhibit activity towards various long-chain linear primary alcohols and carboxylic acids, and only low activity towards very short-chain acids. Moreover, CALA is able to perform conversion of highly branched substrates where most other lipases fail to display any activity [201]. A most interesting aspect of CALA is its preference for esterification above hydrolysis [215]. CALA was found to perform esterification of fatty acids and water insoluble alcohols in a water abundant environment [216].

In this chapter the specifications of the *S. bombicola* lipase-like enzyme, which we designated as the lactone esterase (SBLE), are described. This enzyme was recombinantly expressed in *Pichia pastoris* and purified in a two-step strategy. In-depth characterization of the kinetic parameters of SBLE is hampered by the lack of pure substrates, but preliminary enzymatic assays with different forms of

sophorolipids allowed to estimate the apparent values of K_m ,V_{max} and $k_{cat.}$. Finally the predicted active pocket was confirmed by site-directed mutagenesis.

2 Material and Methods

2.1 Materials

Compounds used for enzymatic assay buffer preparation *i.e.* sodium citrate, sodium hydroxide and hydrochloric acid, were purchased from Sigma. From the same supplier come products like *Pseudozyma (Candida) antarctica* lipase B, paranitrophenyl acetate (pnpa), paranitrophenyl butyrate (pnpb) and paranitrophenyl palmitate (pnpp) and sodium dodecyl sulphate, Triton X-100 and Tris buffer. The solvents acetonitrile and tetrahydrofuran were purchased from Biosolve and Riedel de Haën, respectively. Sodium chloride was purchased from Merck.

2.2 Sophorolipids preparation

The total mixture of sophorolipids was extracted from a *S. bombicola* WT cultured in the medium described by Lang *et al.* [136]. 37.5 g/L rapeseed oil was added two days after inoculation. Yeast cultures were incubated at 30°C and 200 rpm for a total time of 10 days. At the end of the *S. bombicola* WT culturing, 3 volumes of ethanol were added to the culture broth for total extraction of sophorolipids. Cell debris was removed by centrifugation at 1500 g during 10 min. For further gravimetric analysis, the supernatant water-ethanol mixture of the total extraction was evaporated. Two volumes of ethanol were added to dissolve the sophorolipids and the residual hydrophobic carbon components. The mixture was filtered to remove the water-soluble compounds and was evaporated again. One volume of water was added and the solution was brought to pH 7. Then one volume of hexane was added and after vigorous shaking, the mixture was allowed to separate. The different fractions were collected, evaporated and the mass was determined. The hexane phase contains residual oil, while the water phase contains the sophorolipids. Samples were analysed by HPLC using and Evaporative Light Scattering Detection as described in point 2.9.

The lactonic sophorolipids were enriched by precipitation from the raw sophorolipids mixture by overnight incubation at 4°C in water (150 rpm). The resulting solution containing SL crystals was placed at -80°C and lyophilized overnight after which a white powder consisting of very pure lactonic SLs was obtained. Part of these highly purified di-acetylated lactonic SLs were chemically converted to acidic non-acetylated SLs as described by [217]. The non-, mono- and di-acetylated acidic sophorolipids were obtained by growing *Starmerella bombicola sble Δ* on glucose as previously described (see Chapter 3 point 2.2) and sophorolipids were extracted as above. For the enzymatic assays a 100 mM stock solution of sophorolipids was prepared in water.

2.3 DNA construct

Genomic DNA of *Starmerella bombicola* strain ATCC 22214 was used to isolate the gene for the lactone esterase. The primers FW_sble 5'GCTGCAGGACTCCCTTTAGGC3' and Rev_sble 5'CGTCGACTGTATGAGTTGAGT3' were used to create a PCR fragment with two additional restriction

sites for *Pst*I and *Sal*I, to allow cloning in the αpPiczB expression vector (Invitrogen), in frame with the N-terminal α-factor secretion signal and the C-terminal His-tag. After gel-purification, the PCR product was subcloned into the pCR2.1-TOPO vector (Invitrogen) and propagated in *Escherichia coli* TOP10 cells (Invitrogen). After plasmid purification by chromatography (Qiagen MIDI kit) the insert was cut out with PstI and SalI. αpPiczB was also linearized with the same restriction enzymes and purified from gel (innuPREP DOUBLEpure kit). The insert was ligated to αpPiczB using T4DNA ligase and the construct was transformed into One shot TOP10 electrocompetent *E.coli*. Plasmid DNA from transformed clones was purified with chromatography (Qiagen) and confirmed by multiple restriction digestion with XmnI+HindIII, NcoI and SfoI+EcoRV. A control sequencing of the construct (outsourced to Beckman Coulter Genomics) confirmed the αpPiczB_sble construct (**Fig. 4.1**). The linearized αpPiczB_sble construct obtained with *Sac*I, was transformed into electrocompetent *Pichia pastoris* NRRL-Y-11430 cells (obtained from Prof. Nico Callewaert).

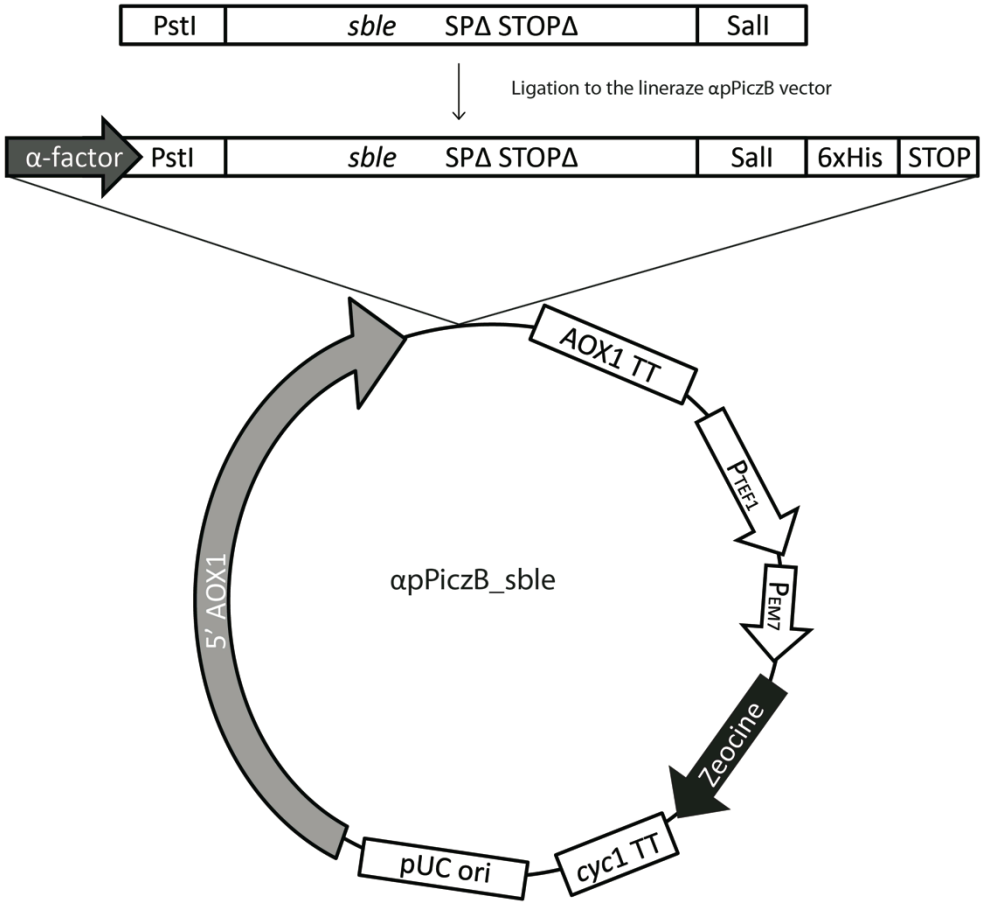


Figure 4.1. Expression cassette ligation into the linearized αpPiczB vector

2.4 Protein expression

Pichia pastoris NRRL-Y-11430 transformed with the αpPiczB_sble construct was grown in 2L of Buffered-Glycerol Complex (BMGY) medium in four 2L flasks containing 500 ml medium for 48h at 28°C, 250 rpm. Then, the BMGY medium was replaced with Buffered-Methanol Complex (BMMY)

medium containing methanol instead of glycerol. The cell pellets were washed and dissolved in 1L BMMY medium under sterile conditions. Protein expression in BMMY medium was carried out for 48h at 28°C, 250 rpm. Every 12h, 1 v/v% MeOH was added for continuous stimulation of protein expression. Finally, medium containing expressed lactone esterase was separated from the cellular fractions by centrifugation (10 min at 4000 g). To evaluate the expression levels, 20 µL of medium was mixed with 5 µL Laemmli buffer containing β-mercaptoethanol and separated on a 12.5 m/v% SDS-PAGE at 130V, followed by Western blot analysis using a HRP-coupled antibody directed against the C-terminal His-taq (Invitrogen). Femto Luminol (Thermo Scientific) was used as a chemiluminescencent substrate.

2.5 Protein purification

Collected medium (2L) of *P. pastoris* expressing SBLE was filtered through a 0.22 µm bottle top filter (Corning) and dialyzed using membranes with a 6-8 kDa cutoff (Spectra Por). The dialysis was performed against 20L of a buffer containing 50mM Na_2HPO_4; 300mM NaCl 300mM (pH 7,5) at 4°C during 24h (repeated three times). Next the dialyzed medium was filtrated through a 0.22 µm bottle top filter (Corning) and loaded onto a Talon Superflow column (Clontech) with a bed volume of 20 ml, connected to an Akta-purifier (GE Healthcare) system. This column was first equilibrated with the same buffer as used for dialysis. After loading the sample, the column was washed with equilibration buffer containing 5mM imidazole. Consequently, the protein was eluted with a buffer containing 200mM imidazole. The protein samples from two runs on the Talon column were combined and concentrated to 2 ml using Vivaspin columns with a MW cutoff of 10kDa (Sartorius).

Next, the partially purified lactone esterase, further referred to as rSBLE, was injected into a Superdex 200 column (GE Healthcare) connected the Akta system with 20mM Tris/150mM NaCl (pH 7,5) as running buffer. The fractions corresponding to lactone esterase were collected and stored at -80°C.

2.6 SDS-PAGE and trypsin digestion

20 µL of the different protein fractions were mixed with 10µL of Laemmli buffer, containing β-mercaptoethanol and separated by SDSPAGE (12.5 m/v%) at 130V. The 3 visible gel bands at 80 kDa, 30 kDa and 15 kDa (see **Fig. 4.2**) were cut from the gel and proteins were digested with 0.002 µg/µl trypsin (Promega) in 50mM $(NH_4)_2CO_3$ during overnight incubation in 37°C. The peptides were two times extracted from the gel pieces with 60 µl of a 60 v/v% ACN/0.1 v/v% FA. These extracts were pooled with the supernatant and the resulting peptide mixture was dried using vacuum centrifugation. Finally, they were dissolved in 15 µl of 2 v/v% ACN/0.1 v/v% FA.

2.7 Mass spectrometric analysis of tryptic fragments

Five µL of the extracted peptides were loaded onto a Zorbax 300SB-C18 analytical column (150 mm x 75 µm, Agilent) connected to an Agilent 1200 chromatographic system (Agilent, Santa Clara, CA, USA) coupled to an LTQ-FTUltra mass spectrometer (Thermo Fisher Scientic, Waltham, MA). The separation was performed by reversed phase chromatography using a 50 min linear gradient ranging from 2% buffer A to 80% buffer B at flow rate of 300 nl/min. The mobile phase buffer A was 99.9%

water with 0.1% formic acid. Mobile phase B was 99.9% acetonitrile with 0.1% formic acid. The LC eluent was directly coupled to a Triversa NanoMate ESI source (Advion, Ithaca, NY), working in the nanoLC mode and being equipped with D-chips upon which a 1.55 kV voltage was applied. The FT-ICR mass analyzer acquired MS scans (3 s) at 100,000 resolution during the LC separation. The three most intense precursor peptides for each MS scan were automatically selected and fragmented by the LTQ ion trap mass analyzer.

Raw LC-MS/MS data received from the FTICR MS measurements were analyzed using Mascot Daemon version 2.3.2. To identify proteins, Mascot version 2.3.01 searches were performed against the in-house available primary annotated *S. bombicola* genome containing 4617 genes, predicted from 3 contigs using the Augustus algorithm, together with its decoy database created automatically by Mascot. For identification of *Pichia pastoris* contaminanting proteins, a *P. pastoris* protein database was used. In search against both databases the following search settings were applied: data were imported with the Thermo Finnigan LCQ/DECA RAW filter, maximum 2 missed cleavages of trypsin were allowed and oxidation (M) was set as variable modification. The ESI-FTICR instrument was chosen with a possible MS/MS error tolerance of 0.5 Da and peptide error tolerance of 10 ppm. All peptides with a significance threshold higher then 0.05 and an ion score cut off lower than 30 were discarded. In case of search against *P. pastoris* database the ion score cut off was 20. We accepted proteins with two or more peptide hits within the above criteria.

2.8 N-terminal protein sequence analysis

100 µg of purified rSBLE was separated by SDS-PAGE 12.5 m/v% at 130V. Then, the protein was transferred to a PVDF membrane using electroblotting. N-terminal sequence determination was performed by automated Edman degradation on a Procise model 494 sequencer instrument, equipped with an on-line HPLC system consisting of a 140C Microgradient pump and a 785A programmable absorbance detector (all from Applied Biosytems). The analysis was performed with acid delivery in the gas phase.

2.9 HPLC and LC-MS analysis of glycolipids

The same methods were used as described in chapter 3, point 2.3.

2.10 Analysis of substrate selectivity

135 mg of dried SL were dissolved in 2 ml water to prepare a 100 mM stock solution used in the enzymatic assays. Five different rSBLE concentrations were prepared: 0.6 µg/ml; 1.8 µg/ml; 3.2 µg/ml; 6 µg/ml; 9 µg/ml (stock concentration 60 µg/ml). The enzyme was incubated with 5 mM of the different sophorolipid mixtures: a mixture obtained from WT *S. bombicola*; a mixture enriched in the lactone form by precipitation from the WT mixture, chemically prepared non-acetylated acidic SL and a mixture of non-, mono- and diacetylated acidic SL produced by the *sbleΔ* mutant. The total volume of the reaction was 1 ml. To buffer the reaction at pH 3.5, 50 mM of sodium citrate dihydrate was used. The assay after overnight incubation at 28°C on a shaking platform was stopped by adding the extraction mixture containing 440 µl ethylacetate and 11 µl acetic acid. From the organic phase, 400 µl was recovered and analyzed on HPLC. For every condition, a blank reaction

was prepared, where the enzyme was replaced by the buffer in which the enzyme was prepared: 25 mM Tris, 150 mM NaCl, pH 7.5.

2.11 Temperature and pH dependency of rSBLE

The rSBLE activity was evaluated at two temperatures, 30°C and 60°C, in an assay with 5 mM of a mixture of non-, mono- and diacetylated acidic SL. 9 µg/ml of rSBLE was used (stock concentration 60 µg/ml) in the final volume of the reaction, 500 µl. The reaction was buffered either at pH 3.5 or pH 6, by 50 mM of sodium citrate dihydrate. The following four time points of the reaction were selected: 1h, 3h, 6h, 12h. For every time point a blank was prepared. The reaction was stopped with 1.5 ml 100% EtOH. After evaporation to a volume of 120 µl, samples were analyzed on HPLC .

2.12 Preliminary kinetic analysis

No pure single substrate for the rSBLE is available. Because product formation was only observed with the mixture of the non-,mono- and diacetylated SL we selected this substrate to determine the approximate kinetic parameters of the lactone esterase. The assay was performed in a volume of 500 µl at 30°C with shaking, in 50 mM sodium citrate dihydrate buffer at pH 3.5. Different substrate concentrations were tested: 0.5 mM; 1 mM; 2.5 mM; 5 mM; 15 mM and 30 mM. The used protein concentration was 1.8 µg/ml. The samples were incubated for 10 min, 20 min, 30 min, 40 min and 60 min. Blanks were collected at time=0 and after 30 min and 60 min. The reaction was stopped with 1.5 ml 100% EtOH. After evaporation to a volume of 120 µl, samples were analyzed on HPLC . The area under the peak eluting at 27,89 min representing the diacetylated product of the lactonic C18:1 sophorolipid was used to calculate the rate of the reaction at each substrate concentration.

2.13 Evaluation of lipase activity

The Lipase Activity Assay kit (BioVision) was used to follow the hydrolysis of the triglyceride substrate. In this assay, glycerol is quantified enzymatically by monitoring a correlated change in the OxiRed probe absorbance at λ=570nm. The sensitivity of this kit is as low as 0.02 mU lipase activity per well. We followed the protocol provided by the manufacturer. We tested different concentrations of the purified lactone esterase, ranging from 0.6 µg/ml to 3 µg/ml of protein.

2.14 Evaluation of esterase activity

For the enzymatic assay on p-nitrophenyl derivatives, purified rSBLE in 150 mM NaCl, 25 mM Tris, pH 7,5 was used. A stock solution of rSBLE at 6 mg/ml, was prepared using Vivaspin concentrators (Sartorius), with a MWCO of 5 KDa. The enzyme concentration was determined using a Nanodrop 2000 spectrophotometer (Isogen). *Pseudozyma (Candida) antarctica* lipase B (CALB) work solutions were derived from a 1.2 mg/ml stock solution in the same enzyme buffer as used for rSBLE. Each of the p-nitrophenyl derivates was prepared in a different way depending on its stability and solubility. A work solution of 1 mM p-nitrophenyl acetate (ppna) was freshly prepared in 5 v/v% acetonitrile/50 mM Tris, pH 7.3, just before use. For p-nitrophenyl butyrate (pnpb) a 1 mM work solution in 50 mM Tris, pH 7.3 containing 0.2 v/v% Triton X-100 and 0.43M tetrahydrofuran (THF) was used. The p-nitrophenyl palmitate (pnpp) stock solution was prepared as follows: 0.0135 g pnpp and 0.017 g

sodium dodecyl sulphate were transferred to a 100 ml volumetric flask. In a 10 ml falcon tube, 1 g of Triton X-100 was weighed, diluted to 10 ml with water and heated (37°C) before adding to the volumetric flask. The mixture in the flask was brought to volume with water. Dissolving was stimulated by stirring at 65°C, alternated with 10 minutes of sonication until the solution was clear. 500 µl of the stock solution was two times diluted with buffer to obtain a work solution with a final concentration of 178 µM pnpp in 50 mM Tris, pH 7.3.

In a 96 well microtiter plate, 200 µl of a substrate work solution was added to 40 µl of either enzyme (sample) or enzyme buffer (blank). From these wells, 200 µl was transferred to an empty well after mixing. Spectrofotometric measurements were performed on a Microplate reader, model 680XR, (Bio-rad) using a wavelength of 405 nm, in the time drive mode, performing 60 readings with intervals of 30 or 60 seconds at an incubation temperature of 25°C. Blanking for the background of the microtiter plate wells and/or yellowish color of the concentrated rSBLE solution was done by endpoint measurements of the empty wells. Only wells with the same background were used to compare samples with blanks. *Pseudozyma (Candida) antarctica* lipase B was used as a positive control for the hydrolysis of all substrates.

2.15 Analysis of glycosylation

5 µg of purified protein was deglycosylated with 1 µl of EndoH (500 U/ µl, New England Biolabs) in 50mM HEPES pH 7.5. Reaction took place at RT overnight at 20°C. The digestion was analyzed on a 12.5 m/v% SDS-PAGE gel.

2.16 Site-directed mutagenesis and mutant analysis

The αpPiczB_sble construct was sent to Genscript where site-directed mutagenesis of one amino acid was performed. Ser(194) from the conservative motif GYSGGA encoded by an AGT codon was replaced with Ala encoded by GCT. The resulting plasmid was transferred to *P. pastoris* and the mutant protein was expressed and purified as described for the wild type. This mutant (in concentration of 6 µg/ml) was also used for the overnight enzymatic assays with 5 mM of a mixture of non-, mono- and diacetylated acidic SL. 6 µg/ml of rSBLE (prepared as described above) was used as a positive control. The reaction took place in 50 mM sodium citrate dihydrate buffer at pH 3.5 and pH6. Product extraction and its analysis were performed as described above.

3 Results

3.1 Expression in Pichia pastoris

To further analyze the characteristics of SBLE we developed a heterologous, MeOH inducible rSBLE producing system in *Pichia pastoris* NRRL-Y-11430. The expression cassette in plasmid αpPiczB_sble is composed of a mature form of SBLE inserted between the α-factor secretion signal and a C-terminal hexahistidine-tag. The resulting plasmid (**Fig. 4.1**) was cut with SacI in the AOX region and used to transform *P. pastoris*. The transformant present in lane nr 1 was selected randomly, based on Western Blot analysis using an antibody directed against the His-tag (**Fig. 4.2**). The medium of

non-induced cells was used as a negative control (**Fig. 4.2**, lane 12) while, as a positive control, an unrelated purified 20 kDa His-tagged protein that was available in-house was used (**Fig. 4.2**, lane 13). The apparent MW of the recombinant protein (rSBLE) was 75 kDa which is more than the theoretical weight for mature rSBLE (45 kDa), indicative for protein glycosylation.

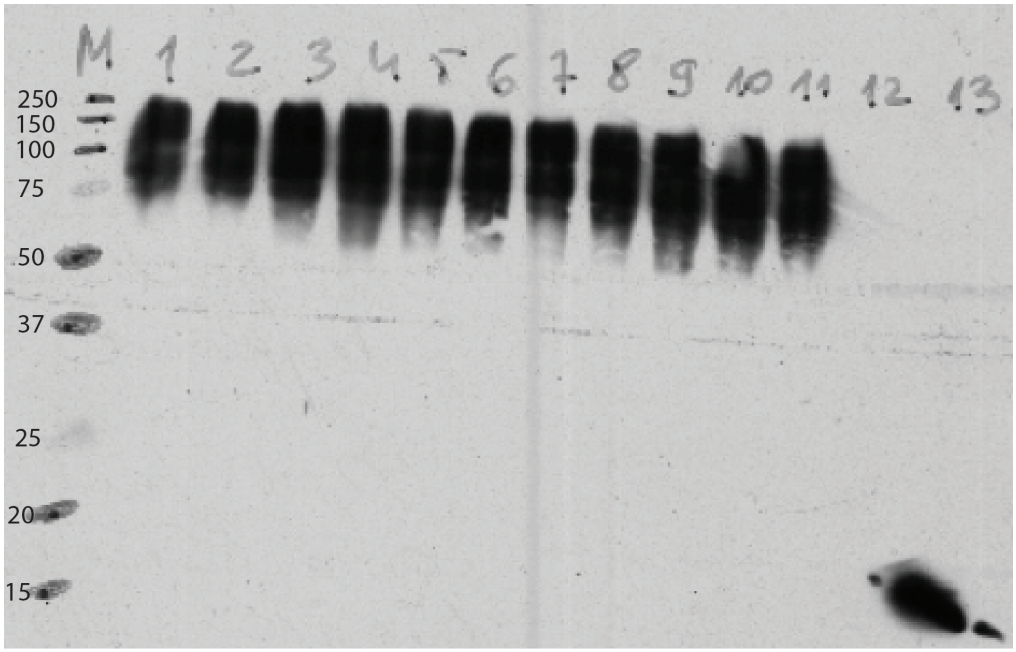


Figure 4.2. Western blot analysis on the medium of different *P. pastoris* transformants to evaluate expression of the rSBLE. M-marker. Lane 1-11: Transformants producing rSBLE; Lane 12- positive control : an unrelated 20 kDa His-Tagged protein; Lane 13: Negative control, medium from a *P. pastoris* transformant where expression was not induced with MeOH.

3.2 Purification of rSBLE

We followed a two-step purification scheme. First, the His-tagged protein was captured from the filtered medium using a Talon IMAC-column. Bound proteins were eluted with imidazole. The SDS-PAGE pattern from the retained fractions displays three bands (**Fig. 4.3A**). These bands were characterized by mass spectrometric analysis of their tryptic digests. The band with an approximate molecular weight of 75 kDa represents the recombinant lactone esterase (rSBLE). It was identified with 20 peptides, which sequences can be found in **Suppl. Table 4.1**. The two other bands around 30kDa (PC1) and 15kDa (PC2) were classified as contaminating proteins from *P. pastoris*, each identified with a single unique peptide in the peptide mass fingerprint search against the *P. pastoris* protein database. The protein at 15 kDa is probably Protein phosphotyrosine phosphatase (PAS_1g04820|PAS_chr1-4_0151) while the protein migrating at 30 kDa is Phosphatidylinositol-4-kinase(PAS_1g00540|PAS_chr1-3_0100). However, the fingerprints of these two protein bands also contained some peptides of rSBLE, indicating rSBLE degradation or a different pattern of glycosylation. The large difference in molecular weight of those contaminants prompted us to perform an additional polishing step using size exclusion chromatography. We managed to separate rSBLE from the contaminating proteins (**Fig. 4.3B**) and obtain 0.75 mg/L of pure rSBLE.

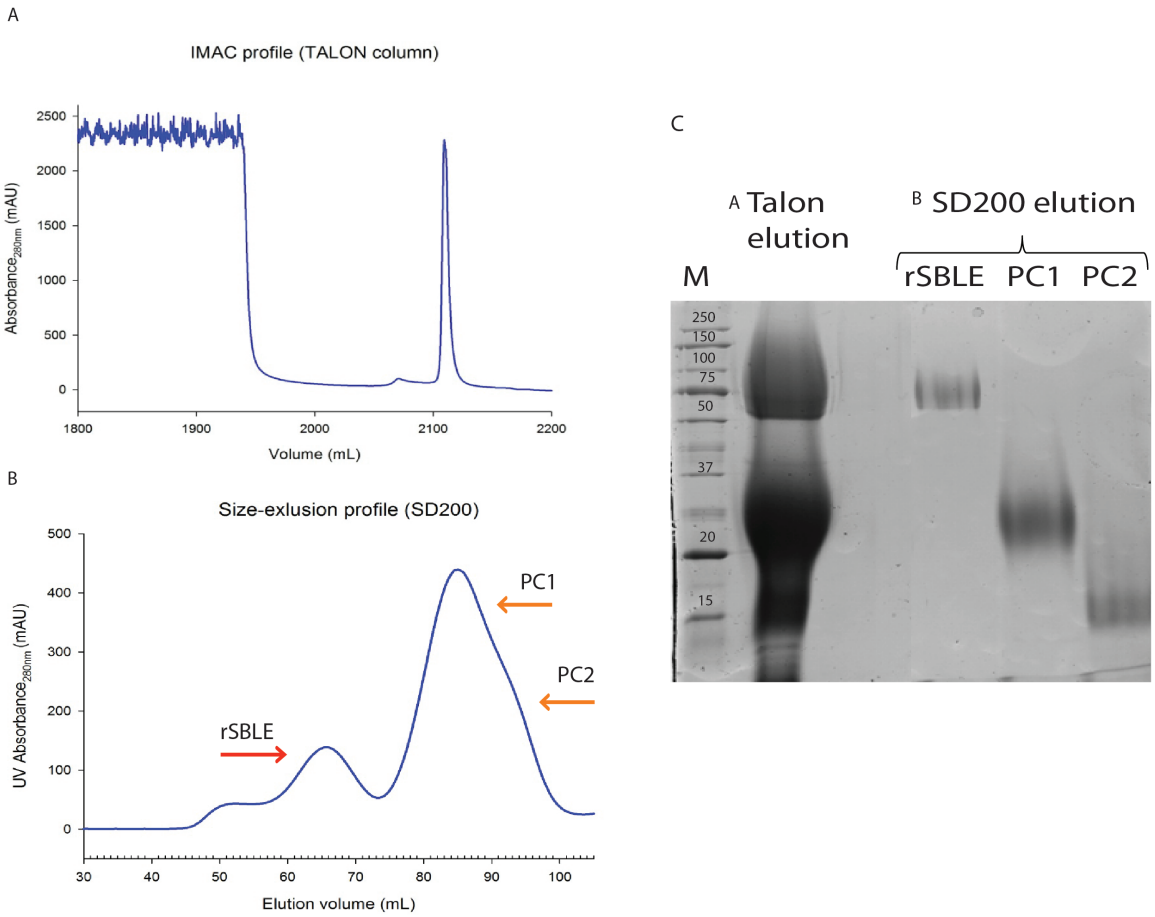


Figure 4.3. The upper chromatogram shows the result from IMAC purification using a Talon-column (A) while the lower chromatogram shows the elution profile from a Superdex 200 column (B). The red arrows indicate the rSBLE. An SDS-PAGE analysis of the peaks of the SEC-peaks is shown on the right. rSBLE is represented by the band at 75 kDA while it is 30kDa and 15kDa for the *P. pastoris* contaminant proteins, PC1 and PC2.

3.3 Amino acid sequencing and glycosylation analysis of rSBLE

S. bombicola lactone esterase (SBLE) sequence has 1233 nucleotides which are translated to 410 amino acids (**Fig. 3.7**). In order to recombinantly express SBLE in *Pichia pastoris* we picked up the mature form of SBLE, without secretion signal and cloned this fragment in frame with α-factor secretion signal in the pPiczB vector. The MW of the mature SBLE form is 43.9 kDa and its pI is 4.63, while the recombinant protein with two additional AG N-terminal amino acids (AG) and a C-terminal 6xHistag MW is supposed to have a molecular weight of 45kDa and a pI of 4.83. However, as it can be seen in **figure 4.3C** purified rSBLE has an apparent MW almost of nearly 75 kDa. This could be explained by extensive N-linked glycosylation because in the SBLE sequence, 7 potential N-glycosylation sites are present (predicted with NetNGlyc1.0.server). Asn-Xaa-Ser/Thr motifs are found at Asn34, Asn122, Asn133, Asn181, Asn325, Asn367 and Asn406 (**Fig. 4.4**). To analyze the extent of N-linked glycosylation, purified rSBLE was incubated with EndoH. After EndoH-treatment, rSBLE displayed a significant shift in the apparent MW on an SDS-PAGE gel, migrating close to its predicted MW of 45kDa (**Fig. 4.5**).

We also performed an N-terminal sequence analysis (Edman degradation) to confirm that the protein is secreted in the correct form. We detected mainly the predicted N-terminal sequence of the protein rSBLE (AGLPLGYTAAPA). The sequence data, however, revealed the presence of two minor components with as N-terminal sequence: *EAEA*APLGYT and *EA*APLGYTAA, which originate from incomplete removal of the signal sequence (α-factor) by proteases Kex-2 and Ste13.

3.4 Homology with other lipases

SBLE is most similar to lipase A from *Pseudozyma antarctica* (CALA) (33% identity). Generally CALA's sequence identity with other typical lipases is low. CALA-like lipases were also found in two other glycolipid producing strains *Kurtzmanomyces* sp. [218] and *U. maydis* [219]. However, those enzymes are not claimed to be associated with the lactonisation of their glycolipid products, mannosylerythritol lipids (MEL). We created a multisequence aligment between SBLE and other lipases (**Fig. 4.4**). The 3-D structure of CALA is solved [220] and we used this structure to obtain insights into the secondary structure of SBLE. The sequence similarity is high enough to suppose that SBLE has the typical α/β hydrolase fold with an ordered appearance of α-helix and β-sheets (**Fig. 1.16**). Moreover the typical conserved motif for lipases, GYSGGA can be found at position 192-197 in SBLE (172-177 CALA). Ser194 (Ser174 CALA) in that motif forms the catalytic triad together with Asp 346 (CALA Asp 324) and His 378 (CALA His 356) (**Fig. 4.4**). These residues are conserved in all four aligned lipases. We also noticed a conserved disulfide bridge between Cys112 and Cys286 (Cys91 and Cys263 in CALA) connecting the α/β hydrolase domain with the lid, which is thought to form a substrate tunnel for hydrophobic substrates and which is possibly the primary determinant of substrate specificity on the acyl site [220]. This α-helical lid (between β7 and β8) in CALA is composed of 6 helices (α6-α11) and is likely to be conserved in SBLE (**Fig. 4.4**). CALA prefers straight long-chain carboxylic acids rather than smaller acyl groups like acetate and butyrate. It also prefers a *trans* rather than *cis* conformation of the fatty acids. However, SBLE is expected to act on the *cis* conformers which is found in most SL structures [9]. CALA has been shown to display a moderate interfacial activation [221]. It is suggested that the solved structure shows CALA in its closed, inactive form and that several conformational changes have to occur in order to produce the active form. Most notably, it was hypothesized that the β10 and β11 strands at the C-terminus of CALA form an active-site flap that mediates interfacial activation. These C-terminal strands are rich in aromatic residues which can flip out and expose the active site when they interact with the lipid phase. Notably, the SBLE sequence is shorter than other compared lipases and lacks α15 and β11. It is not clear if and how interfacial activation occurs in SBLE.

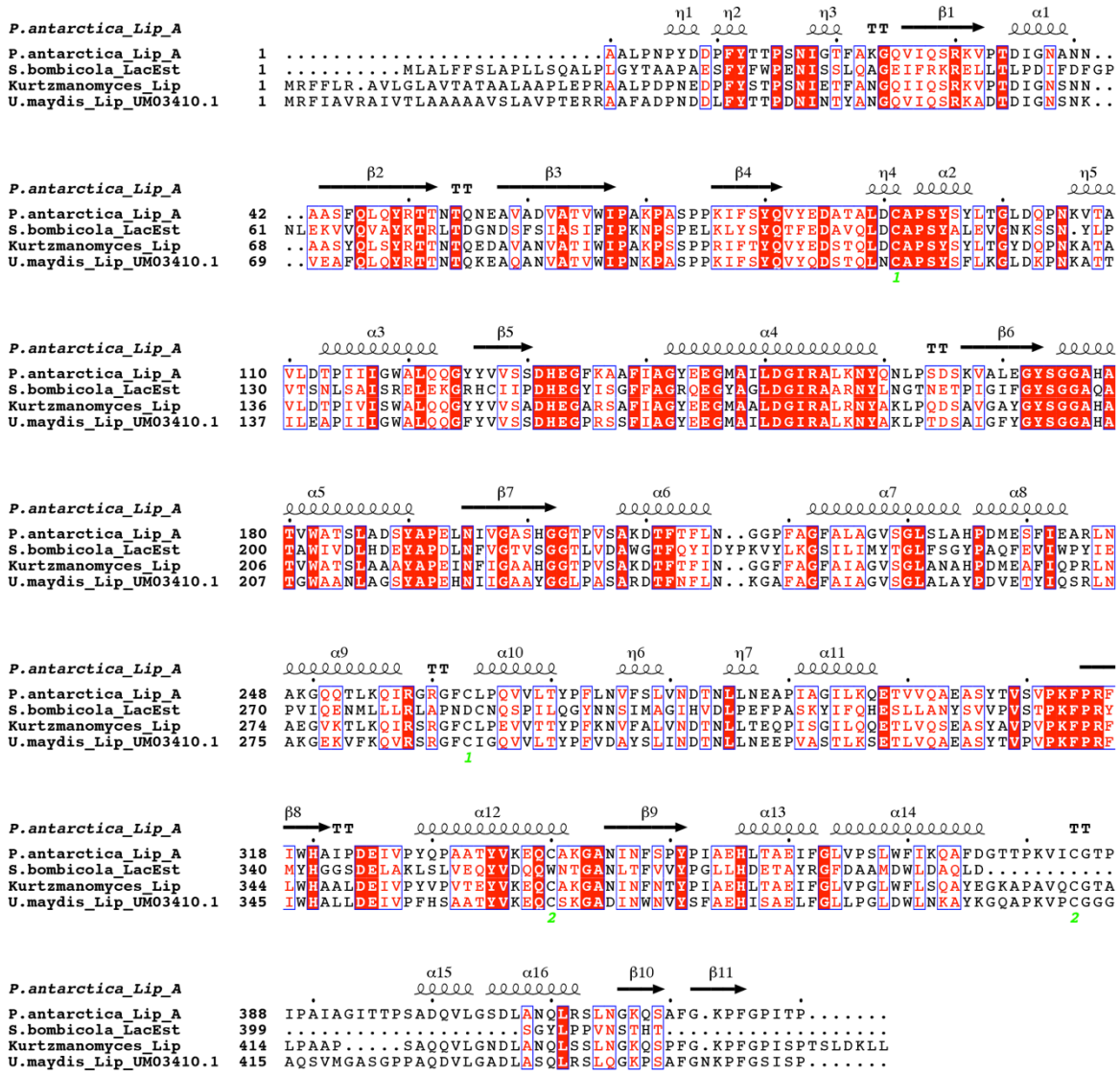


Figure 4.4 Multisequence alignment between *P. antartica* lipase A (LipA), *S. bombicola* lactone esterase (LacEst), *Kurtzmanomyces* lipase (Lip) and *U. maydis* lipase UMD3410.1 (Lip). The secondary structure elements of Lip A from *P.antartica* (pdb 2VEO) are indicated. Red letters-similar AA; Red background-identical AA; green 1, 2 – conserved disulphide bridges.

3.5 Substrate (SL) specifity

Firstly, we tested enzymatic activity with different sophorolipid mixtures (SL) under conditions mimicking the *S. bombicola* growth conditions during SL secretion, namely 28°C and pH 3.5. The positive results were only obtained in an assay with using a non-, mono- and di acetylated acidic SL mixture. In the HPLC chromatogram we observed a peak eluting at 27.7 min which represents diacetylated lactonic SL (data not shown). Furthermore, no lactonic products were obtained in an assay with chemically prepared non-acetylated acidic SL, which indicates the enzyme only accepts the acetylated precursor. We also saw that rSBLE does not hydrolyse lactonic SL into acidic SL. As this

substrate was selected for further characteristic of rSBLE like optimal pH and temperature, the detailed description of the results (HPLC and LCMS spectras) will be discussed in next paragraph.

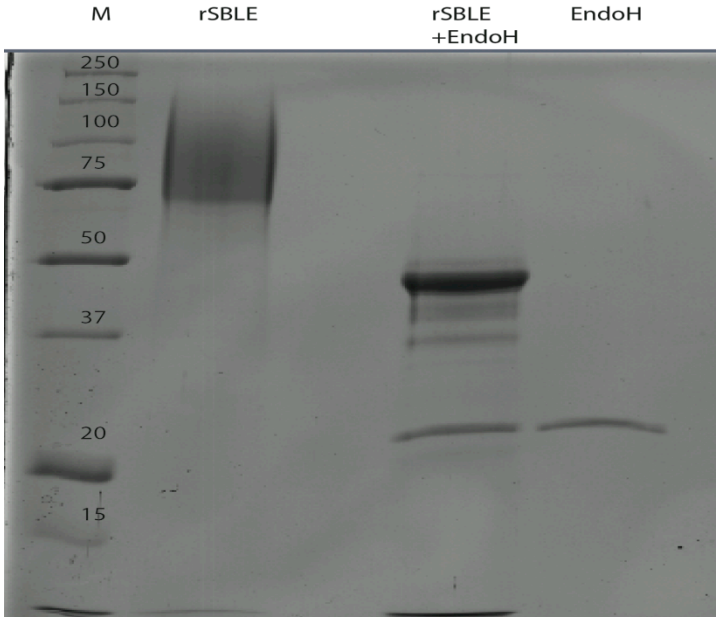


Figure 4.5. Deglycosylation of rSBLE with EndoH. Following EndoH-treatment, the apparent MW of rSBLE shifts from around 75 kDa to 45kDa. EndoH (25kDa) alone was loaded as control.

3.6 pH and temperature dependency of rSBLE

With the substrate selected above, we compared an enzyme assay at pH 3.5 and pH 6. These are the extreme pH values to which *S. bombicola* is exposed during growth. Indeed, the initial growth medium has a pH 5.8, but the pH decreases to 3.5 when the cells reach the stationary phase, the time point where SL are produced. The assay at both pH-values shows a positive result, since the main product peak at 27.7 min corresponding to diAc lactonic SL (C18:1) (MW=688) with a subterminally hydroxylated fatty acid (FA) chain was observed in "Sample" in **Figure 4.6**. The smaller peak eluting at 28 min also represents a diAc lactonic SL (C18:1) but with a terminally hydroxylated fatty acid chain. In the same HPLC chromatogram we additionally noticed some minor peaks eluting at 26.1 min which correspond to monoAc lactonic SL C18:1 (MW=646) or diAc lactonic SL C18:2 (MW=686) while peaks eluting at 24.1-24.2 min can also be monoAc lactonic SL C18:1 (MW=646) and/or diAc acidic SL C18:1 (MW=706), C18:0 (MW=708) (summary in **Table 4.1**). MonoAc lactonic SL can elute at different retention times due to the difference in position of the acetyl group and location of the fatty acid hydroxyl-group.

To confirm the identity of the products, now only based on retention time, we also performed an LC-MS analysis. **Figure 4.7** presents the BPI chromatogram of the sample and the blank from the assay performed at pH3.5 (**Fig. 4.6**). After detailed screening of the XIC chromatograms (**Fig. 4.7.1** and **4.7.2**) we observed four new lactonic products in the sample. The detected mass corresponded with the expected molecular weight for the diAc lactonic SL: C18:2 (MW=686), C18:1 (MW=688), C18:0 (MW=690) and monoAc lactonic C18:1 (MW 646). Surprisingly, nonAc lactonic SL were detected.

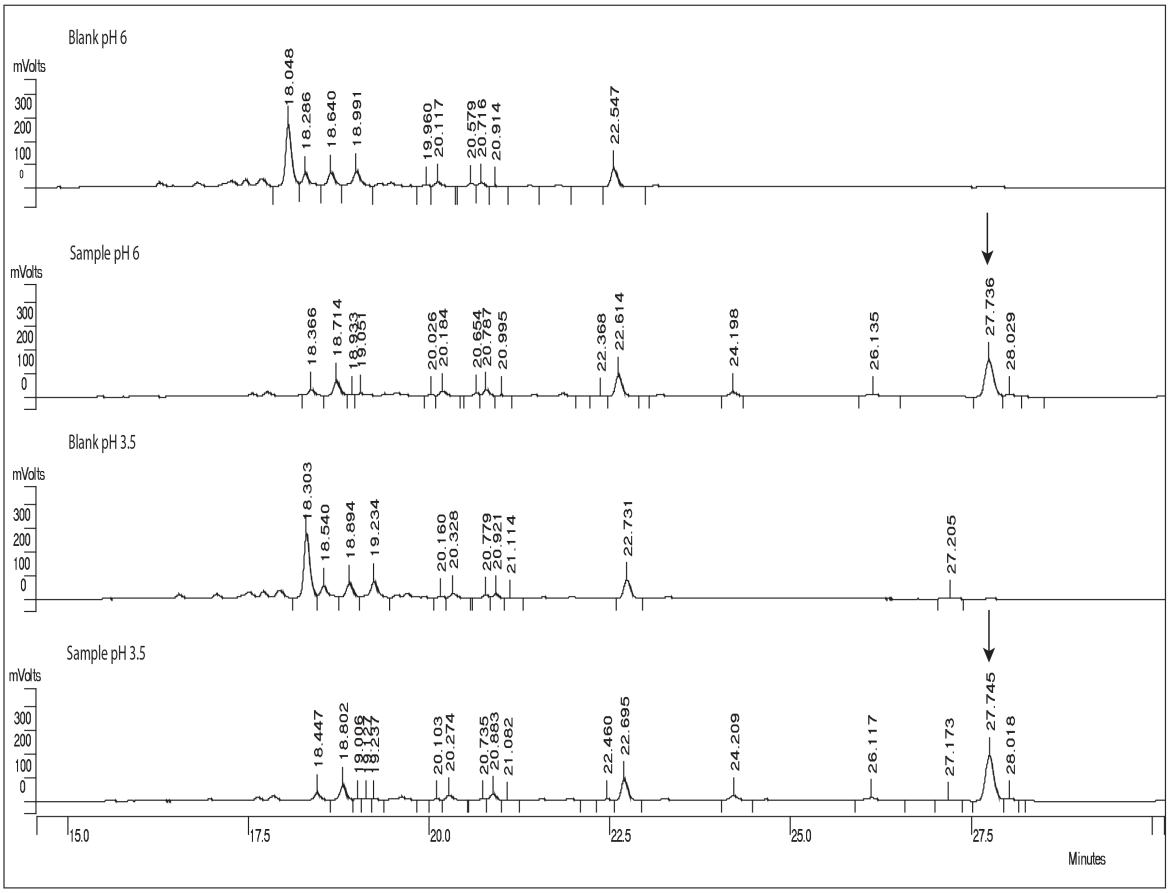


Figure 4.6. Example of the rSBLE assay at different pH-values (3.5 and 6), with 5mM acidic SL mix after time 1h incubation in 30°C. Sample- purified rSBLE at concentration of 9µg/ml; Blank-buffer in which rSBLE was purified. The arrow indicates the main product of the assay: diAc lactonic SL C18:1. The LCMS analysis of assay at pH 3.5 is presented in figure 4.7, 4.7.1 and 4.7.2.

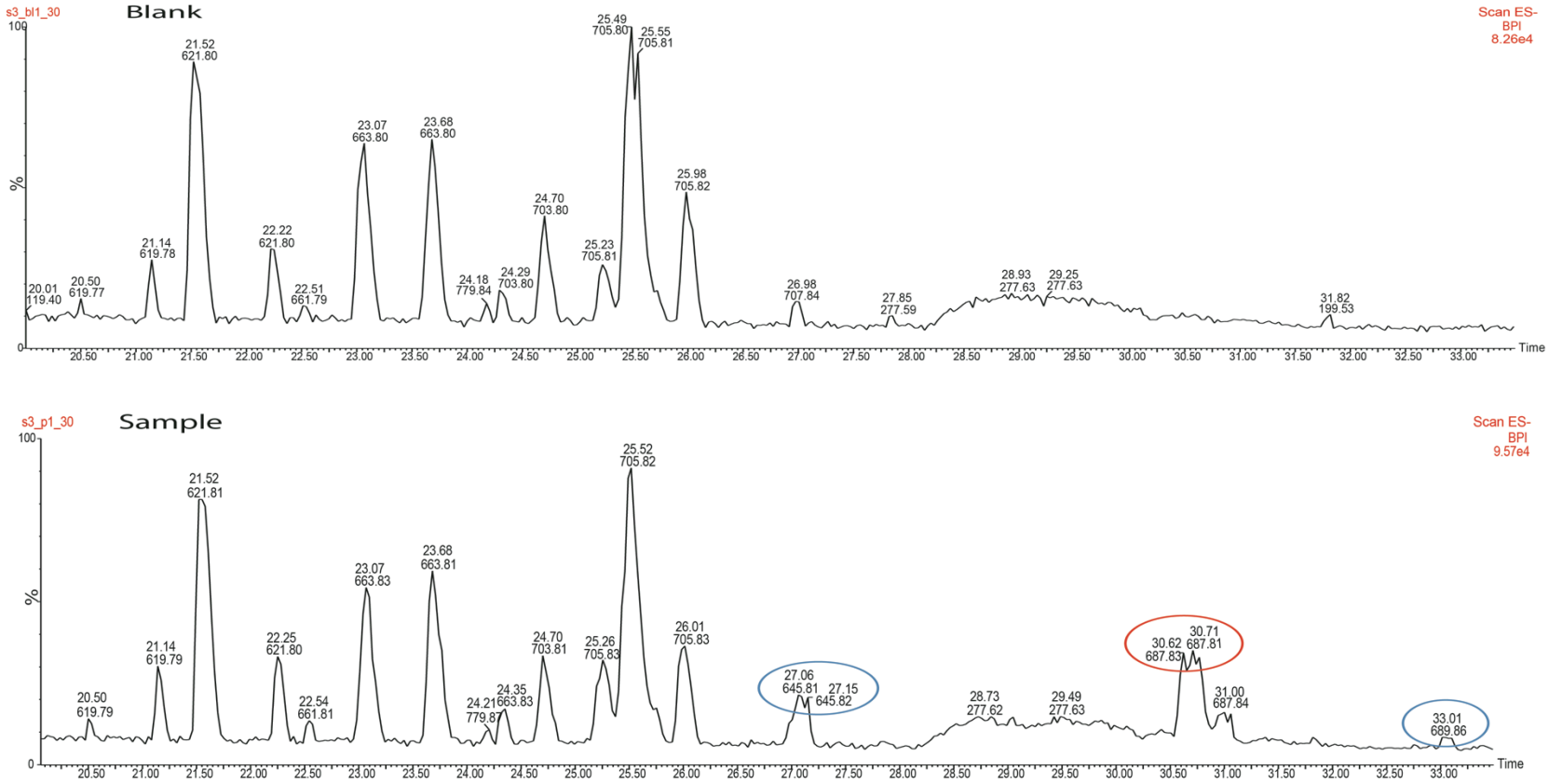


Figure 4.7. Base peak intensity (BPI) chromatogram from LCMS analysis of sample and blank from rSBLE assay with 5mM mix of the acidic SL at 30°C and pH3.5 after 1h. Sample-purified rSBLE at concentration of 9µg/ml; Blank-buffer in which rSBLE was purified. In circles are pointed a products of the assay. In red is market the diAc lactonic SL C18:1 (MW688) while in blue diAc C18:2 (MW690) and monoAc C18:1 (MW646) lactonic SL. The detailed extracted ion chromatograms (XIC) of those products elution time are presented below in 4.7.1 and 4.7.2.

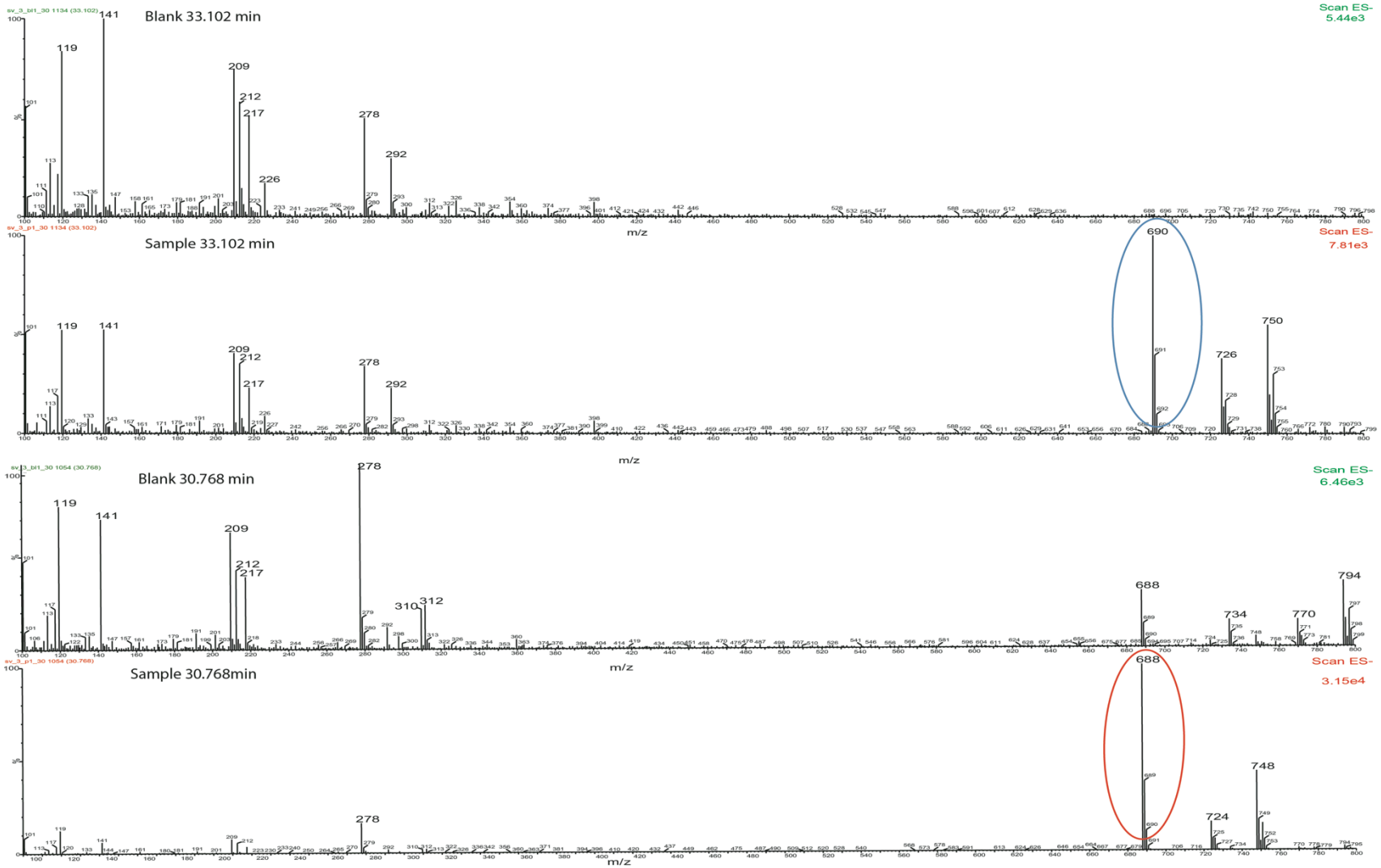


Figure 4.7.1 The upper XIC presents diAc lactonic SL C18:0 with MW 690 eluting at 33.102 minutes. This form of the SL is only visible in the sample in comparison to the blank. The bottom XIC presents diAc lactonic SL C18:1 with MW 688 eluting at 30.768 minutes. This is the main product of the rSBLE enzymatic reaction visible in HPLC chromatograms. In sample the intensity of this peak is an order of magnitude greater compare to blank.

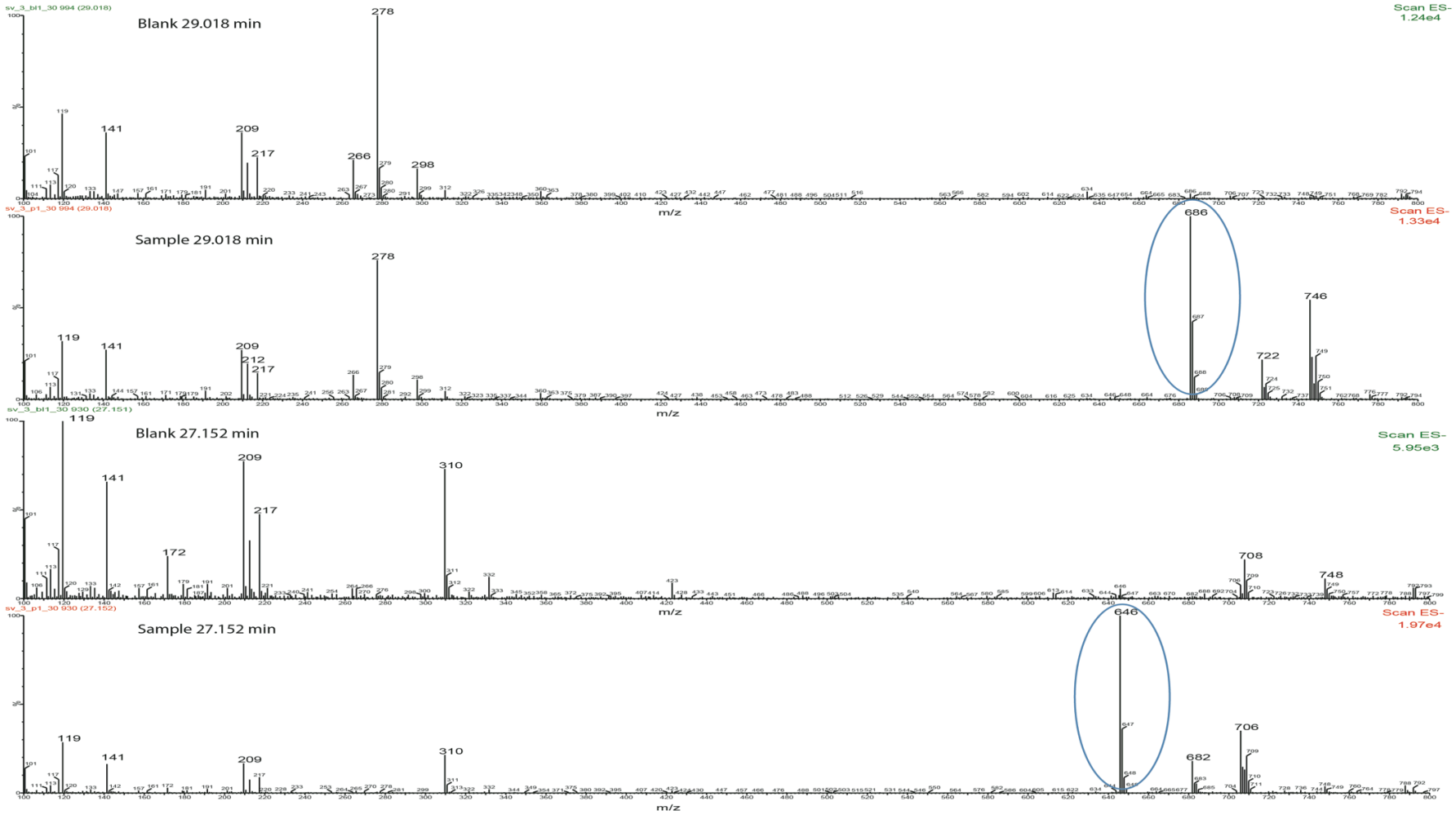


Figure 4.7.2 The upper XIC presents diAc lactonic SL C18:2 with MW 686 eluting at 29.018 minutes (this product was not visible on BPI chromatogram). This form of the SL is only visible in the sample in comparison to the blank. The bottom XIC presents monoAc lactonic SL C18:1 with MW 646 eluting at 27.152 minutes. In the sample, the intensity of this peak is one order of magnitude higer compare to blank.

Table 4.1 Summary of the major peaks detected in LCMS (Fig. 4.7; 4.7.1 and 4.7.2) and its corresponding elution time from HPLC (Fig. 4.6) analysis. The informations about identification, MW and their corresponding elution time of differnet forms of the sophorolipids were taken from the PhD thesis of Dr. S.Roelants [222] and were supported by the personal communication with Dr. IN.Van Bogaert. In red is marked the main product of the rSBLE assay, diAc lactonic C18:1 SL, while in blue are marked other lactonic product found in that assay.

HPLC elution time	LCMS elution time	MW	Identity	Chain length	Acetylation
17.3-17.8	20.5-21.14	620	SL open	C18:2	non
18.03-18.3	21.52-22.25	622	SL open	C18:1	non
20.19-20.3	22.54	662	SL open	C18:2	mono
20.6-21.09	23.07-24.35	664	SL open	C18:1	mono
22.6-22.8	24.29-24.70	704	SL open	C18:2	di
22.6-23.1; 24.1-24.2	25.26-26.01	706	SL open	C18:1	di
24.1-24.2	26.98	708	SL open	C18:0	di
24.1-24.2; 26.1	**27.06-27.15**	**646**	**SL closed**	**C18:1**	**mono**
26.1	**29.02**	**686**	**SL closed**	**C18:2**	**di**
27.70; 28.03	**30.62-31.00**	**688**	**SL closed**	**C18:1**	**di**
29.8-30.00	**33.01**	**690**	**SL closed**	**C18:0**	**di**

Also assay at 60°C was tested but only a minimal amount of main product (at 27.7 min) was produced compared to the assay with 30°C (**Fig. 4.8**). For the rSBLE kinetic study, we therefore chose a pH-value of 3.5 and a temperature of 30°C.

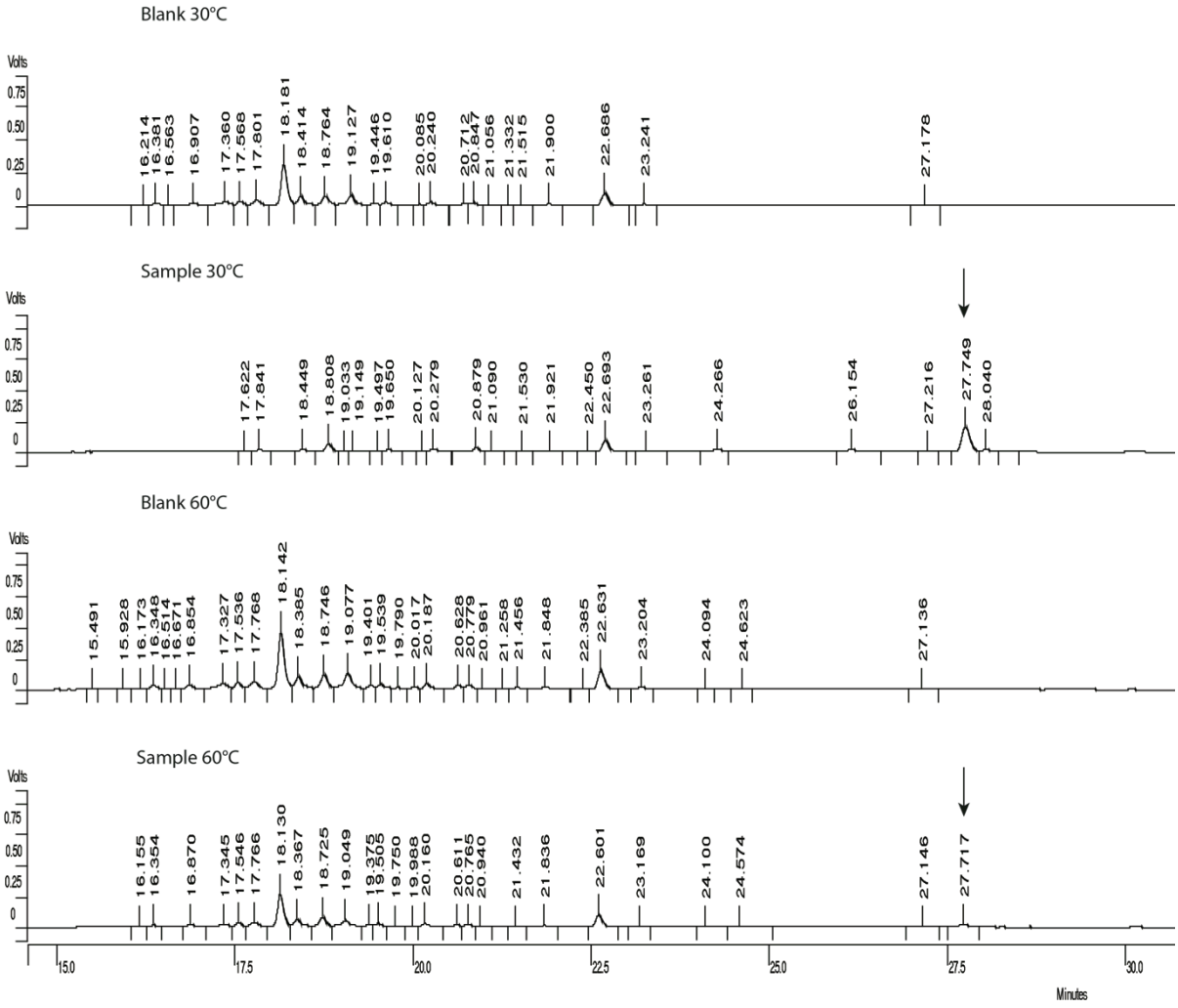


Figure 4.8. Example of the rSBLE assay at different temperatures (30°C and 60°C), with 5mM acidic SL mix after time 3h incubation at pH 3.5. Sample purified rSBLE at concentration of 9µg/ml; Blank-buffer in which rSBLE was purified. The arrow indicates the main product of the assay: diAc lactonic SL C18:1.

3.7 Preliminary kinetic analysis

We performed an enzymatic assay by determining product formation at six different concentrations of SL substrate at five time points. **Figure 4.9** shows a few of the corresponding chromatograms showing product accumulation in the function of time. From the area under the peak eluting at 27.7 min, representing the main product, the amount of product in mM was recalculated based on a standard curve (**Table 4.2**). Next, curves were generated to derive the reaction rate for six different substrate concentrations (**Fig. 4.10**) and finally we prepared a graph representing the relation of initial velocity with the substrate concentration. From this graph an apparent (app) K_m and V_{max} could be obtained (**Fig. 4.11**). Those values are formally assigned as apparent because the used substrate is in fact a mixture of many molecules and we know that the enzyme could react with at least 4 of them as 4 products were detected in LCMS: diacetylated SL acidic C18:2, C18:1, C18:0 and monoacetylate acidic SL C18:1 (**Fig. 4.8**). We should take into account that the enzyme/substrate interactions can be disturbed by other substrates, which can act as "competitive" inhibitors. Further experiments using pure substrates are required to derive the real K_m ,V_{max} and $k_{cat.}$

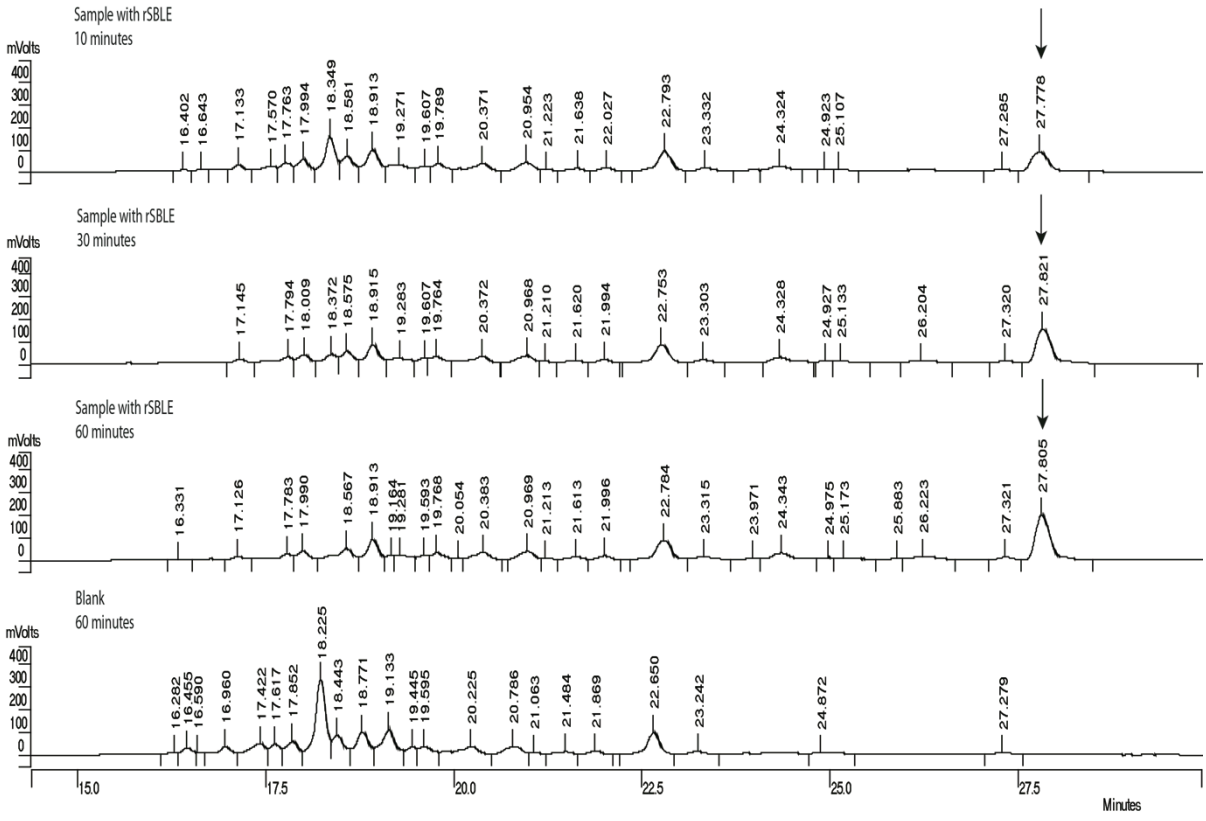


Figure 4.9. HPLC chromatogram from assay of rSBLE with 2.5 mM acidic SL mix in conditions described in paragraph 2.12. The arrow indicate increasing in time product, diAC lactonic SL. The area under this peak was used to calculate app. K_m ,V_{max} and $k_{cat.}$

Table 4.2. Data obtaiend during an rSBLE kinetic experiment at different concentrations of the acidic SL mixture at pH3.5 and at 30°C after five different time points. The product of the reaction was analyzed by the HPLC. The area under the peak 27.7 min, representing diAc lactonic SL C18:1, from HPLC chromatograms was recalculated to the product concentration according to a standard curve: Area under the peak [mV*s]= 2038.3*SL concentration [mM]. This standard curve was created based on measurements of the peak area of the acidic nonAc SL C18:1 (peak 18.3 min) in different concentrations. The product concentration in the reaction volume (500 µl) was calculated: SL[mM] in the reaction= SL[mM]/4.167.

Subst. con.	**0.5mM**			**1mM**			**2.5mM**		
Time	**Area under a peak 27.7**	**diAc lactonic SL C18:1**	**diAc lactonic SL C18:1 in the reaction**	**Area under a peak 27.7**	**diAc lactonic SL C18:1**	**diAc lactonic SL C18:1 in the reaction**	**Area under a peak 27.7**	**diAc lactonic SL C18:1**	**diAc lactonic SLC18:1 in the reaction**
[min]	[mV*sec]	[mM]	[mM]	[mV*sec]	[mM]	[mM]	[mV*sec]	[mM]	[mM]
0	0	0	0	0	0	0	0	0	0
10	69.4	0.034	0.008	121	0.059	0.014	585	0.287	0.069
20	107	0.052	0.013	176	0.086	0.021	1332	0.653	0.157
30	156	0.077	0.018	250	0.123	0.029	1492	0.732	0.176
40	214	0.105	0.025	407	0.200	0.048	1672	0.820	0.197
60	168	0.082	0.020	443	0.217	0.052	2171	1.065	0.256
Subst. con.	**5mM**			**15mM**			**30mM**		
Time	**Area under a peak 27.7**	**diAc lactonic SL C18:1**	**diAc lactonic SL C18:1 in the reaction**	**Area under a peak 27.7**	**diAc lactonic SL C18:1**	**diAc lactonic SL C18:1 in the reaction**	**Area under a peak 27.7**	**diAc lactonic SL C18:1**	**diAc lactonic SLC18:1 in the reaction**
[min]	[V*sec]	[mM]	[mM]	[mV*sec]	[mM]	[mM]	[mV*sec]	[mM]	[mM]
0	0	0	0	0	0	0	0	0	0
10	565	0.277	0.067	1432	0.703	0.169	1450	0.711	0.171
20	1301	0.638	0.153	1540	0.756	0.181	2110	1.035	0.248
30	2434	1.194	0.287	2580	1.266	0.304	2590	1.271	0.305
40	3080	1.511	0.363	3230	1.585	0.380	3410	1.673	0.401
60	2664	1.307	0.314	5255.	2.578	0.619	4740	2.325	0.558

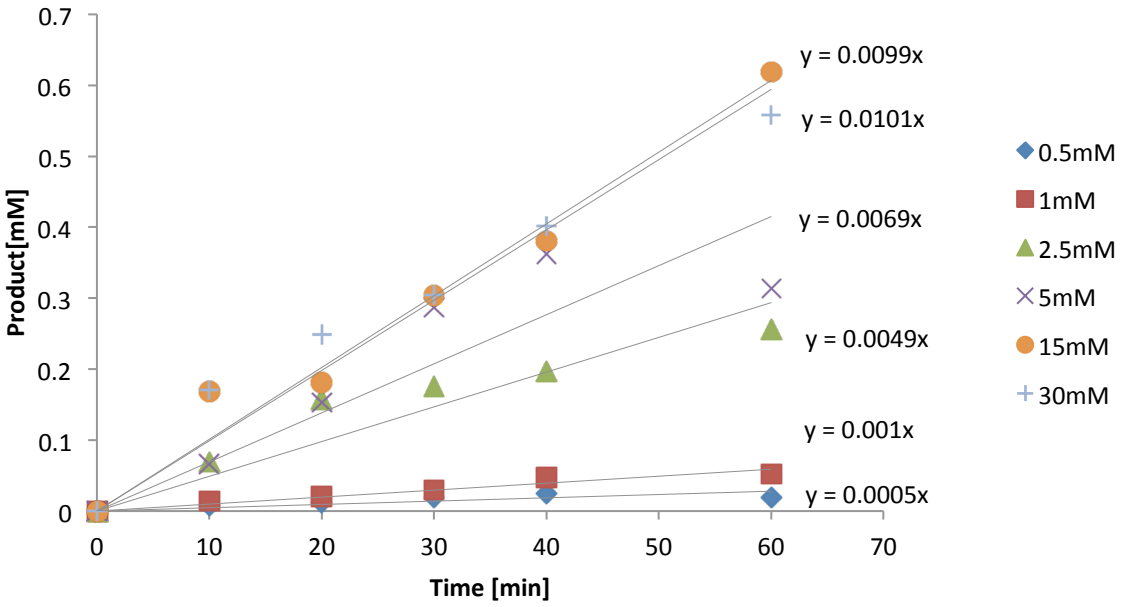


Figure 4.10. Product formation (diAc lactonic SL C18:1) vs. time of six independent rSBLE experiments with different concentration of acidic SL mix (0.5-30mM) during five time points. The assays were performed discontinuous.

Two models were used to fit our data, *i.e.* the typical Michaelis-Menten model and the Hill equation. From the MM model, an app. K_m of 4.38 mM and k_{cat} of 5.00 1/s are determined, while according to the Hill equation an app. K_m of 2.9 mM and $k_{cat\ of}$ 4.30 1/s were obtained. We are not able to predict which model is correct, multiple replicas of this experiment should be performed to learn more about rSBLE kinetic. However Hill model could indicate the substrate activation/interface activation of the enzyme which is typical for some lipases.

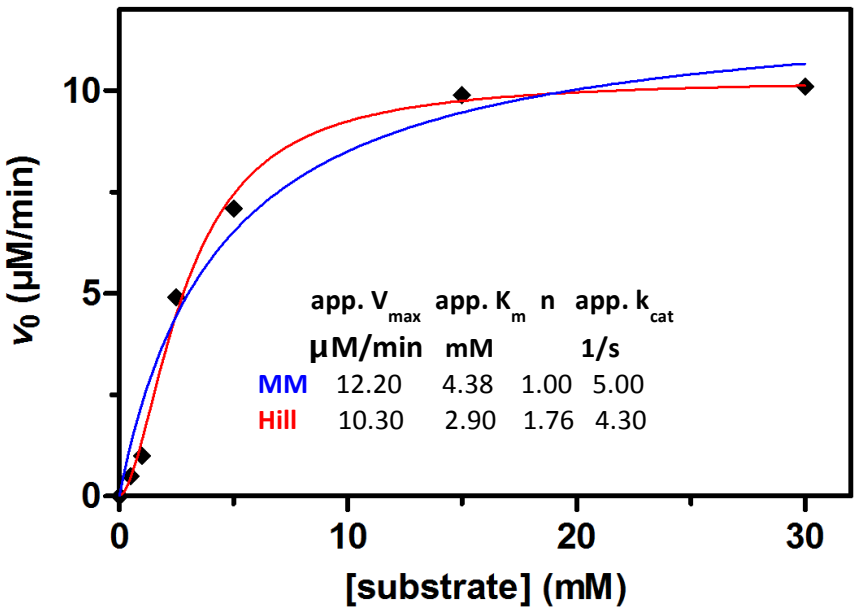


Figure 4.11. rSBLE initial velocity in function of the concentration of the acidic SL mix.

3.8 Esterase activity

We considered the capacity of rSBLE to act as an esterase or as a classical lipase. We first used the BioVision Lipase Activity kit which allows to monitor the hydrolysis of the ester bond in triglycerides. No activity could be observed. To test whether SBLE has a simple esterase activity, we performed assays of purified rSBLE with paranitrophenyl esters. The positive control, CALB, showed activity with all paranitrophenyl esters used in this test. rSBLE did not exhibit any activity towards pnp-palmitate (**Fig. 4.12B**) even at high enzyme concentration (1 mg/ml). We observed some minor activity of the enzyme with pnp-butyrate and pnp-acetate (**Fig. 4.12A** and **4.13A**). However, we also used the rSBLE Ser194Ala mutant (see below), which should be inactive due to removal of the active site Ser, showed equal activity as rSBLE with pnp-acetate (**Fig. 4.13B**). Moreover, when a negative control sample was prepared from the empty α-pPiczB plasmid expression in *P. pastoris* (dialyzed and purified according to the same methods rSBLE) we still saw some activity with pnp-acetate (**Fig. 4.13C**). Therefore we believe that that rSBLE does not hydrolyze esters bond in both long and short chained substrates.

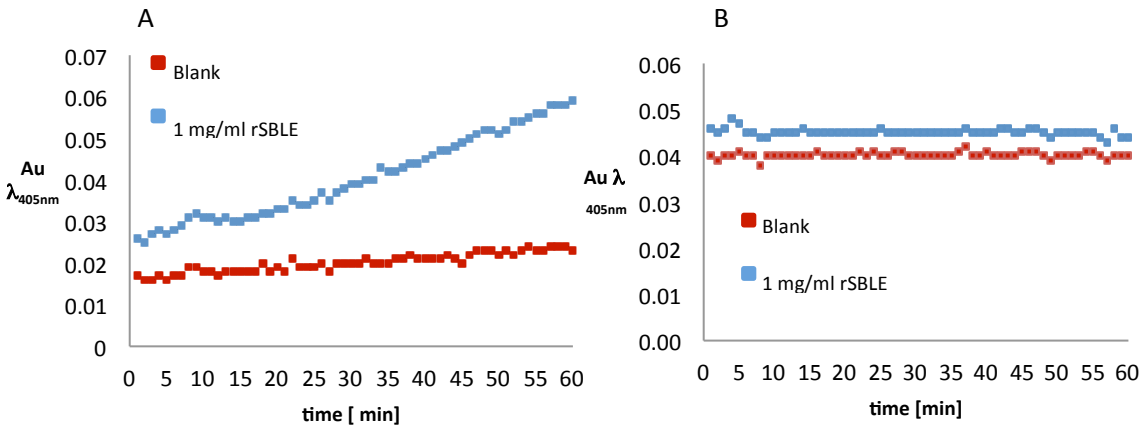


Figure 4.12 Esterase assay of rSBLE with (A) 833µM paranitrophenyl butyrate (pnpb) and (B) 149 µM paranitrophenyl palmitate (pnpp).

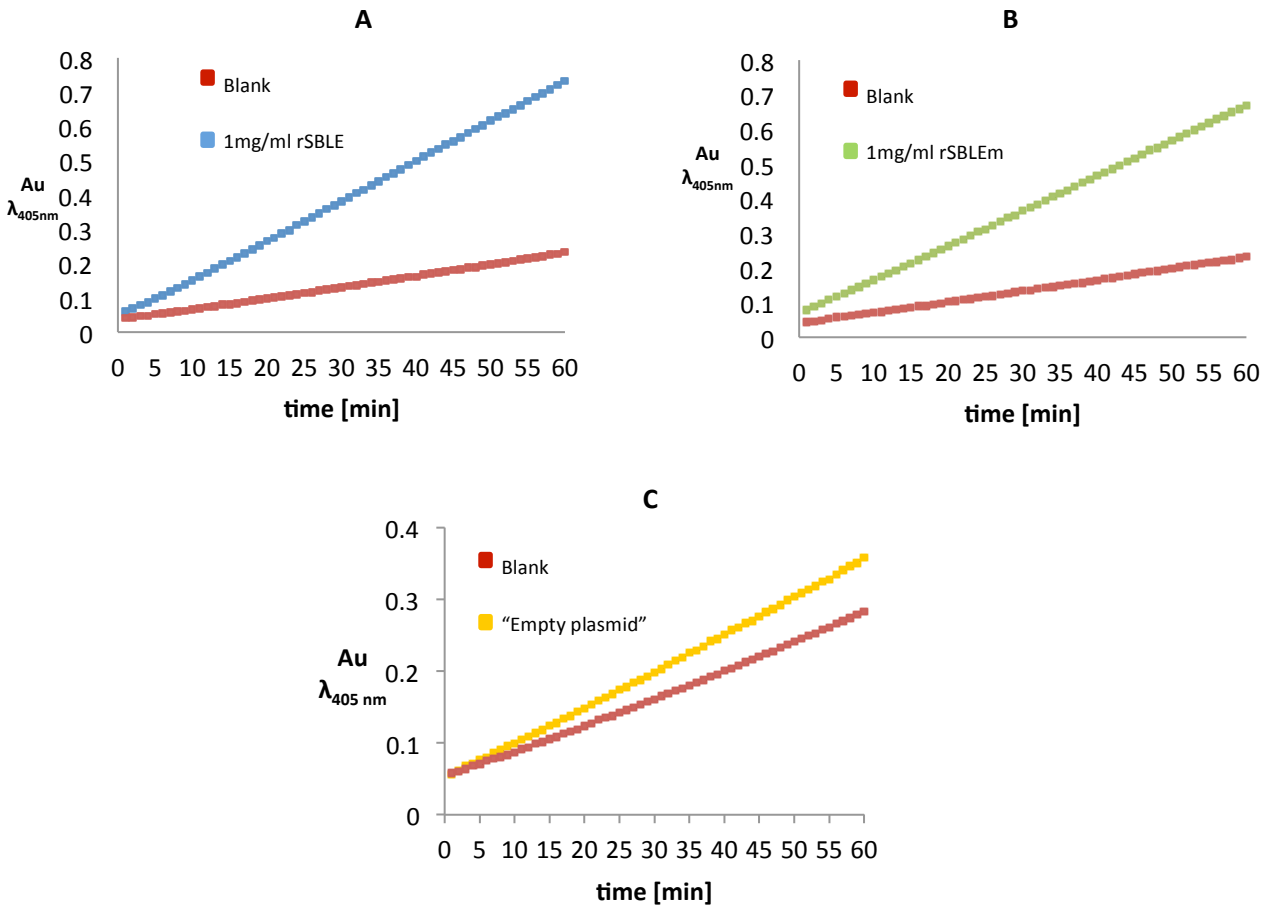


Figure 4.13. Esterase assay of rSBLE (A), rSBLE Ser194Ala mutant (rSBLEm) (B) and medium collected from a negative control using an empty plasmid in *P. pastoris* ("Empty plasmid") (C) with 833 µM paranitrophenyl acetate (pnpa).

3.9 Site-directed mutagenesis

Based on the homology with CALA, we predicted that Ser 194 in the GYSGG motif represents the active serine residue of SBLE. To verify this, we performed site-directed mutagenesis replacing this Ser by Ala. The mutant rSBLE (rSBLEm) was expressed and purified in the same way as the rSBLE. The purified rSBLEm was used to perform an enzymatic assay with the acidic SL mix. As expected, we did not observe any lactonic SL in the HPLC separation of the resulting product (**Fig. 4.14**).

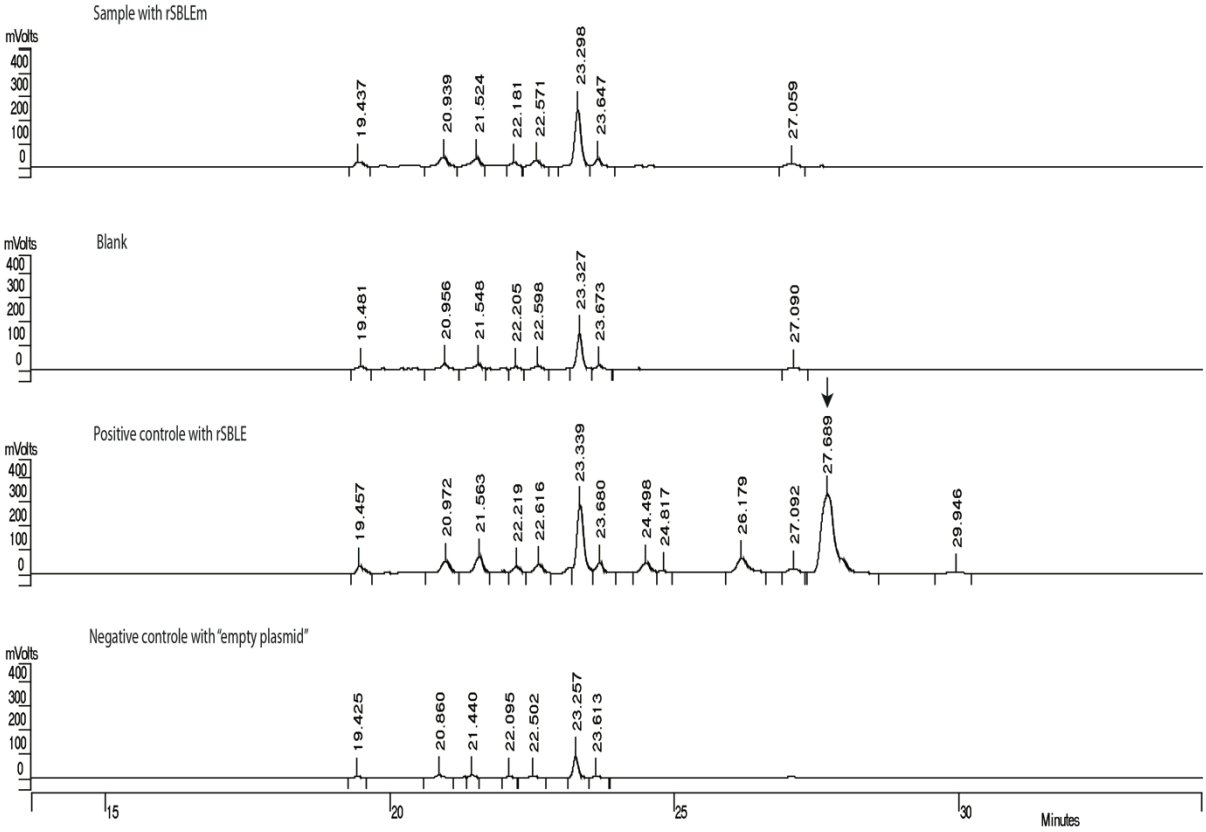


Figure 4.14. HPLC chromatogram representing the products of the assay with 5 mM mixture non-, mono- and diacetylated acidic SL (as described in point 2.16). rSBLEm-pure Ser mutant of the recombinant SBLE; rSBLE-pure recombinant SBLE; Blank- assay with buffer in which rSBLE and rSBLEm were purified; "empty plasmid"- concentrated medium from the αpPiczB vector in *P. pastoris*. The product , diacetylated lactonic form of SL, is indicated with the arrow (peak 27.435 min) is only present in assay with positive control.

4 Discussion

Earlier studies on the sophorolipid (SL) biosynthetic pathway showed that the formation of acidic SL takes place inside the cell. The enzymes responsible for the fatty acid hydroxylation, sophorose head formation and its acetylation were previously identified and described in detail. In the previous chapter, we described how the missing link in the biochemical pathway leading to lactonic sophorolipids, a lactone esterase (SBLE) was discovered. Its involvement in sophorolipid ring closure was demonstrated by the creation of a knock-out strain. The sophorolipids produced by that mutant contain only acidic SL while the lactone SL are lacking.

In order to further characterize the enzyme, we set up a heterologous expression system. We selected *P. pastoris* as an expression host, because it was successfully used for the expression of other lipases [223]. The purified rSBLE was obtained in a two step purification strategy based on using IMAC and size exclusion chromatography and we obtained 0.75 mg/L (1.5mg) protein. By

repeating the protein production process a few times, we generated sufficient amount of rSBLE to perform the described enzymatic assays. The final yield is actually quite low compared to expression yield of 60 mg/L of *Rhizopus oryzae* lipase also expressed in *P. pastoris* [224]. To improve it, we could create strains that contain multiple copies of an expression cassette or use different strains of *P. pastoris* strain, for example a protease-deficient strain [225]. The expression of rSBLE was confirmed by Western blotting and Edman degradation. We also observed a rich glycosylation of the rSBLE which probably reflect the natural properties of SBLE as a secreted protein but it is also known that *P. pastoris* widely glycosylates secreted proteins [225]. Further analysis of the glycosylation patterns of the natural SBLE and recombinant rSBLE enzyme are necessary to determine whether glycosylation has any influence on the enzyme activity.

SBLE does not show significant sequence similarity with enzymes described to have lactonisation activity towards hydroxyl fatty acids or sophorolipidis *in vitro*, like lipase B from *Pseudozyma antarctica* [126] and lipase P from *Pseudomonas* [123]. The closest homolog of SBLE appears to be another lipase from *Ps. antarctica,* lipase A (CALA), which does not posses the described above activities. From the sequence alignment with CALA we notice that SBLE has a GYSSGG motif typical for the serine hydrolase type-lipases. To test if Ser from this motif is necessary for the SBLE behavior, we created a rSBLE mutant where we replaced the predicted catalytic triad Ser with Ala. The enzyme lost its activity for the SL lactonization. We therefore propose that SBLE operates via the common mechanism in which the catalytic serine residue is assisted by a His/Asp pair. Although SBLE and CALA share many structural features including the catalytic triade and helical lid domain, there are some notable differences such as the absence in SBLE of the C-terminal β-sheet loop which covers the active site and may be involved in interfacial activation. Therefore structural studies are necessary to reveal detais about the mechanism in SBLE.

Despite the sequence similarity, SBLE and CALA display significant differences at several levels. CALA is highly thermostabile [201] while we proved that rSBLE in an assay at 60°C produce clearly a lower yield than in an assay at 30°C. CALA tolerates a wide range of pH with optimum at around pH7, while SBLE has an optimal pH in the acidic and neutral range, which is also recorded for some other lipases [226]. It would be interesting to test a wider range of the pHs and temperatures to learn more about the optimal conditions for rSBLE. Moreover, CALA has a preference to *trans* fatty acids while most of the lipases prefer *cis* fatty acids. This preference can be connected with the structure of those proteins [227]. We predict that SBLE will have more preferences to *cis* configured fatty acid tails corresponding to the structure found in most of the SL [9].

CALA hydrolyses triglycerides with a preference to position sn-2 in contrast to other lipases which mainly prefer the 1,3-sn position. In rSBLE, we did not manage to prove any hydrolysis activity, either for the triglycerides, lactones SL or paranitrophenyl esters. Interestingly, the *S. bombicola* KO strain for SBLE described in the previous chapter, is still able to produce SL when oil is added to medium so this enzyme does not have a main lipase/hydrolase function in *S. bombicola* (Dr. IN. V.B. personal communication). The same mutant was also analyzed in the SL catabolism and still a lactone ring opening was observed [222]. The observations above support our primary results, that SBLE does not have a hydrolysis activity. On the other hand, our study confirmed that SBLE prefers to create an ester bond in acidic sophorolipids.

The ability to lactonize hydroxylated fatty acids or sophorolipids *in vitro* was claimed not to be a common features of lipases but is a specific characteristic for some of them [123]. It is expected that special structural features are necessary to catalyse the synthesis of the macrolactone. Moreover, monolactonization is successful only when the substrate has a specific chain length (between C15-C20) and an organic solvent is used to create the anhydrous environment favouring the intramolecular estrification while the presence of water is promoting hydrolysis [123]. Also, a high substrate concentration creates a competition between the reactive groups of the same molecule (monolactone) or different substrate molecule (multilactones), near the enzyme active site. This is unfavourable for the –OH and -COOH from the same molecule and favorable for reaction between hydroxyl and carbonyl groups of different substrate molecules. Thus, a small substrate concentration can offer a better condition for intramolecular esterification [126]. However, in our assays the substrate concentration is usually quite high (0.5-30mM in assays) and lactonic SL are still formed even in an aqueous environment.

This extraordinary behaviour of SBLE could be possibly explained by specific structural feature. Its closest homolog, CALA, was already described to favor esterification and alcoholysis compared to hydrolysis [215]. For example, the esterification of fatty acids in water abundant systems (formation of 2-ethyl-1-hexyl palmitic acid ester) was described [216]. To explain this behavior, this study focused on the preferences of lipase to water. Apparently, many hydrolases have a separate entrance to the active site for the water: such water tunnel were detected in many of them during molecular modeling. It was suggested that the presence/absence of that water tunnels could influence the rate of hydrolysis versus esterification [228]. In case of CALA, it is suggested that in the presence of water and the substrates, the enzyme deactivates a specific interaction between water and the active site to avoid hydrolysis. Possibly in *S. bombicola* lactone esterase preference for the cyclic esterification could also be explained by such a structural feature but further structural work is required to explain the mechanism of the reaction.

Apart from the structural features of SBLE, sophorolipds themselves could create a pro-lactonisation condition. Firstly, the structure of the sophorolipids has definitely a high importance on the efficiency of lactonisation. The fatty acids built in SL typically have a “curved” *cis* conformation [9] which locates the carboxylic acid group close to the sophorose head (**Fig. 1.1**). Moreover, the acetylation of the SL can direct the reaction to 4”-C while in *in vitro* experiments, nonacetylated SLs rather are lactonized at 6”-C sophorose [45]. We believe that acetylation of the sophorolipids also prevents the polymerization process with which lactonisation competes. Secondly, the properties of sophorolipids as biosurfactants can have a great impact on the activity of the enzyme. There are many examples where surfactants are added to *in vitro* lipase assays. Surfactants in concentration above CMC can form micellar aggregates that may also interact with hydrophobic parts of lipases in solution [229]. These various interactions have a significant effects on the conformation of enzymes and they can either promote or decrease the activity of lipases. It was proved that proteins interact differently with micelles as compared to monomers, e.g. some proteins can be denatured by surfactant monomers while other proteins can only be denatured by surfactant micelles [230]. **Figure 4.15A** represents a different interaction of lipase and surfactants structures.

Characterization of the rSBLE took place with SL in a concentration above CMC and at a pH where micelle assembly was described. We believe that the presence of that assembly could influence SBLE action. Our suggestions can be supported by the description of the enzymatic esterification process in aqueous miniemulsion mechanism [231] where catalysis, leaded by the water-soluble enzyme, takes place at the interface between the hydrophobic droplet and aqueous phase (schema in **Fig. 4.15B**).

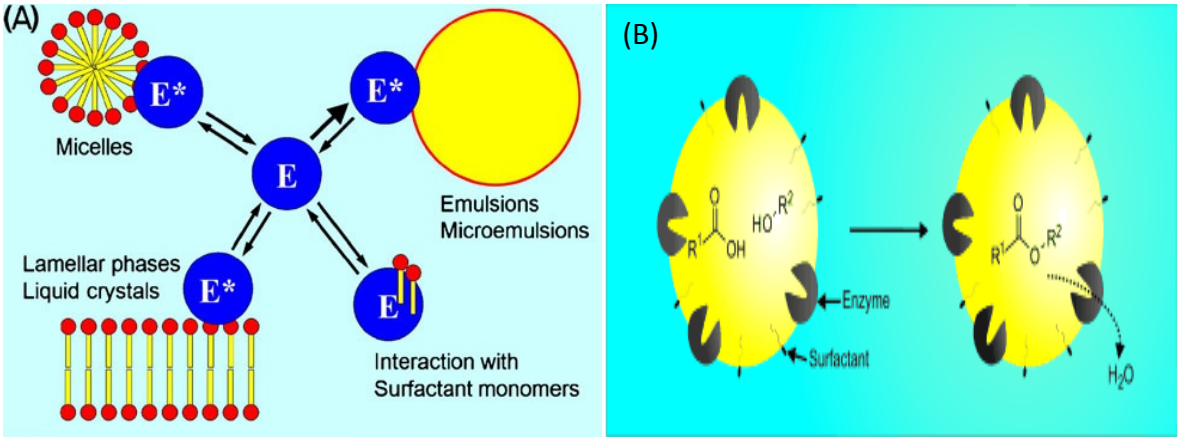


Figure 4.15. A. Possible partitioning of lipase [229] between the various phases/structures dispersed in solution in the presence of surfactants and/or lipids. Arrow's size indicates the favoured interactions. E, lipase in solution; E*, lipase present at the interface between water and emulsions/microemulsions, mixed micelles or lamellar structures (vesicles, liposomes).
B. Scheme representing the esterification in the miniemulsion [231].

5 Conclusion

Lactonization is a complicated reaction that requires specific conditions to be fulfilled in order to successfully produce monolactones. We show here that a lactone esterase, which was identified in the *S. bombicola* exoproteome, is an enzyme that can lactonize acidic sophorolipids specifically. We suggest that SBLE possibly operates via a mechanism typical for serine hydrolases. However, this does not explain the ability of the SBLE to perform intramolecular esterification in aqueous solutions. The complete mechanism of this striking reaction is thus not understood. Possibly the environment of sophorolipid lactonization, namely the pH and SL concentration, is essential for such a process. Sophorolipids as biosurfactants assemble themselves in micelles which can create prolactonization conditions in water. Further knowledge about this system would be undiscernibly profitable for the industry and science. Indeed, the ability to perform esterification in an aqueous environment reduces time and energy. It is fascinating that such a complicated arrangement was created by nature.

Supplementary tables

Suppl. Table 4.1 : Detected peptides of the recombinant lactonase by MS analysis. Every peptide is described by sequence, its expect value and ion score as calculated by Mascot version 2.3.01

Nr	Score	Expect	Peptides
1	42	6.20E-05	K.NPSPELK.L
2	55	5.80E-06	K.VVQVAYK.T
3	50	1.00E-05	K.VVQVAYKTR.L
4	81	1.50E-08	R.QEGYAGLDGIR.A
5	46	2.60E-05	R.YMYHGGSDELAK.L
6	64	3.60E-07	R.YMYHGGSDELAK.L + Oxidation (M)
7	46	6.30E-05	R.QEGYAGLDGIRAAR.N
8	91	1.60E-09	K.SSNYLPVTSNLSAISR.E
9	73	8.50E-08	K.SSNYLPVTSNLSAISR.E
10	83	1.40E-08	R.LTDGNDSFSIASIFIPK.N
11	77	5.00E-08	R.ELLTLPDIFDFGPNLEK.V
12	57	2.40E-06	R.HCIIPDHEGYISGFFAGR.Q + Propionamide (C)
13	59	1.50E-06	R.HCIIPDHEGYISGFFAGR.Q + Propionamide (C)
14	49	3.40E-05	K.RELLTLPDIFDFGPNLEK.V
15	42	9.50E-05	R.KRELLTLPDIFDFGPNLEK.V
16	70	1.20E-07	R.KRELLTLPDIFDFGPNLEK.V
17	44	8.00E-05	K.GRHCIIPDHEGYISGFFAGR.Q + Propionamide (C)
18	61	1.80E-06	R.LTDGNDSFSIASIFIPKNPSPELK.L
19	43	7.80E-05	R.ELLTLPDIFDFGPNLEKVVQVAYK.T
20	54	7.80E-06	K.TRLTDGNDSFSIASIFIPKNPSPELK.L

Chapter 5

General conclusions and future perspectives

Intracellular proteome

In our study on the intracellular proteome of *S. bombicola* we focused on a proteomic comparison between sophorolipid producing and non-producing conditions. We confirmed that the proposed enzymes for the sophorolipid production pathway are co-ordinately produced in the stationary phase at 2 magnitudes of order higher than in exponentially grown cells. Membrane protein enrichment also allowed us to confirm the overexpression of the ABC transporter which was previously discovered in the SL cluster. However, the last enzyme in the SL production pathway, a lactone esterase, which catalyzes the ring closure to form lactonic sophorolipids, was not identified in this intracellular proteome search. The fact that we were able to detect SL cluster proteins with the applied metabolic labelling strategy (SILAC) is encouraging to perform similar studies in other uncharacterized biosurfactant producers. Moreover this proteomic study provides the first insights in the proteome of this highly important industrial strain. The acquired knowlegde could be used to define targets for genetic modifications in order to increase the yield of SL or to produce a new-to-nature biosurfactants.

Among the changes in protein abundances during the transition from the exponential growth phase to the SL producing stationary phase, the majority of the differences can be explained, not surprisingly, in the context of the different growth phases. This probably masks some aspects of the direct impact of sophorolipid production on the cells. If instead of the exponential phase, a non SL producing strain would be used as a reference, definitely the detected changes would be more specific for SL secretion. However, only recently we have access to such a strain, which was created by knocking out the cytochrome P450 gene, encoding the first enzyme in the SL pathway. On the other hand, in the current study we were able to compare the metabolic flows in the two growth phases. We found that the largest group of down-regulated proteins in the stationary phase is involved in protein translation and maturation and amino acid biogenesis. A similar observation was made for proteins participating in protein folding, glycosylation, sorting and degradation. In contrast, the enzymes from the glyoxylate cycle, which bypasses the citric cycle, were significantly up-regulated during the stationary phase. We also report a striking atypical behaviour of heat shock protein 12 (Hsp12), which is widely documented to be induced in the stationary phase [163, 164], but is down-regulated in sophorolipid producing cells. Hsp12 could have a function in membrane stabilization [165] and control of membrane fluidity in response to stress. We hypothesize that the production of sophorolipids protects *S. bombicola* from such stress reducing the need for Hsp12. There are still other puzzling observations that require further investigation. For example we did not observe major changes in the abundance of fatty acid synthesis enzymes. Typically, fatty acid synthase is constitutively expressed but in general, the amount of FA decreases in the stationary phase, resulting from a reduced need for membrane and cell wall biosynthesis [169]. However, in the case of sophorolipids production, the up regulation of enzymes creating building blocks for FA synthesis such as citrate lyase and citrate synthase was expected, similarly to oleaginous yeast [95]. We noticed only a significant decrease of proteins participating in the ergosterol synthesis. Again, the ergosterol biosynthesis is strongly responsive to osmotic stress and it remains to be determined whether the change of the physicochemical properties in the cell environment, due to sophorolid production, modulates the ergosterol biosynthesis pathway [170]. It would be highly interesting to employ a metabolomics strategy to expand our knowledge about the fatty acid metabolism and general glycolysis/gluconeogenesis flux. Such an integration of two powerful techniques would give

us a more complete view of the protein functions *in vivo*. We believe that the combined information can also help us to elucidate how SL production is regulated. Moreover, we propose a time dependent experiment, using iTRAQ or a similar technology, where proteins expression levels will be monitored for a larger number of time points.

In the sophorolipid gene cluster of *S. bombicola,* no transcription factor was detected [53]. Our proteomic work does not give further clues on the regulation of sophorolipid production. In the literature, sophorolipid production in *S. bombicola* was described to be connected with nitrogen and phosphate limitation [58] [57]. Our results are not consistent with this observation. We detected only a small decrease in the concentration of all three nutrients (ammonium, phosphate and glucose). It is possible that sophorolipid production levels are increased under nutrient deprivation but its induction is regulated by a hitherto unknown trigger that needs further investigation.

Extracellular proteome characterization

Recently, the interest on the microbial extracellular proteome increased in the biotechnological sector due to the important activities of the enzymes present there. We are the first to characterize the exoproteome of the biosurfactant producer *S. bombicola*. This research generally aimed to characterize the extracellular proteome, and more specifically, to reveal the existence of a protein responsible for sophorolipid lactonization. The exoproteome was collected at four time points representing the exponential and stationary phase, which are respectively non-producing and SL producing phases. The majority of the identified proteins were present in both phases and their expression level was rather stable. Most of them have a predicted secretion signal and an extracellular localization. In the exoproteome of the exponential phase, some specific proteins were connected with protein folding/secretion while in the stationary phase proteins connected to pathways like protein degradation and nutrient extraction were found. The identified *S. bombicola* exoproteome is similar in size and content to that of other biotechnologically and medically important yeasts like *C. albicans* [112], *C. utilis* [111] or *Pichia pastoris* [191]. By means of functional annotation we could divide the proteins into different groups e.g. proteins related to cell wall synthesis and maintenance, nutrients extraction, etc. Unexpectedly, the biggest group of proteins is represented by proteases. This is different from the exoproteome from other species of biotechnological interest, and refers more to pathogenic species. Anyway, we suspect that *S. bombicola* would not be a good host for heterologous protein expression. The discovery of the invertase homologue in the *S. bombicola* exoproteome is intriguing since our growth medium contained a high amount of glucose, while in the case of *S. cerevisiae*, extracellular invertase is only expressed under glucose-limiting conditions. However, in some non-conventional yeast (*K. lactis* and *P. anomala*), high osmolarity does not result in an inhibitory effect on the derepression of the invertase biosynthesis [209]. *S. bombicola* is a nectarivorous yeast found in sucrose-rich environments, such as honey, nectar of the flowering plants and bee hives [210]. The presence of an invertase in the exoproteome could be therefore connected with the natural habitat of *S. bombicola* where its expression is possibly utilized to extract nutrients. We found that the activity of the identified invertase is low and it is not clear if it is more efficient, even under optimized medium conditions, then the already commercialy available *S. cerevisiae* invertase. However, confirmation of its presence in the *S. bombicola* exoproteome could be important for bee-yeast interaction studies.

It was already suggested that some bees select flowers containing yeast due to their predigesting function for complex sugars to more accessible forms for bees [232].

Our main discovery in the exoproteome was the discovery of the missing enzyme (SBLE) in the SL biosynthetic pathway, responsible for the lactonization of acidic SL. A single knock-out of this lipase/esterase gene resulted in complete absence of any lactonic form of sophorolipids. Moreover, lactonization was confirmed to take place in the extracellular compartment (**Fig. 5.1**). The discovery of the lactone esterase allows to better control sophorolipid biosynthesis. Thanks to the knock-out mutant we are now able to produce a single form of SL, namely acidic SL. Until now, it was not possible to produce such a sophorolipid mixture in a natural way. Due to the higher foaming capacity and better water solubility of the SL mixture produced by the mutant, these compounds have unique properties and show better performances for several applications such as their use as detergents, in pharmaceutical applications, in cosmetic applications, etc. The use of the KO strain and this enzyme has been patented (WO 2013/092421).

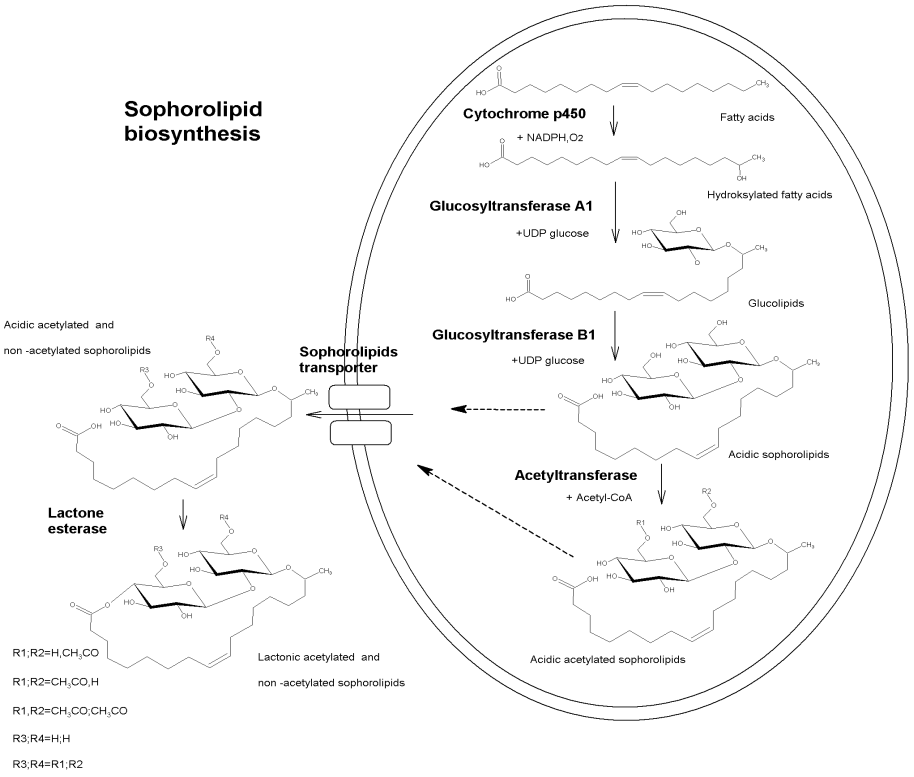


Figure 5.1 Biosynthesis of the sophorolipids. Thanks to a proteomic study validated by molecular biology approaches we demonstrated that acidic sophorolipid lactonization takes place outside of the cell.

It would be interesting to analyse the exoproteome in other environments containing, for example, other carbon sources or oils, to identify proteins involved in their conversion to sophorolipids. After the discovery of the SBLE we realized that the exoproteome could be a source of many interesting proteins. We are screening some proteins identified in the *S. bombicola* exoproteome collected during SL catabolism (collaboration with Dr. S.Roelants and Prof. Soetaert). Additionally, the structure/function of other enzymes in the exoproteome of *S. bombicola* is worth analysing from a biophysical point of view. Sophorolipids are produced in large amounts by this organism, even in

concentrations above the CMC. At this condition SL can create a different form of assemblies like micelles or vesicles. Together with the acidic extracellular surroundings this creates an environment that could potentially unfold and deactivate proteins. How the extracellular enzymes evolved to persist such an environment is unclear and studies using circular dichroism or related techniques to follow secondary structure changes could be performed to follow the proteins' interaction with sophorolipids.

Lactone esterase

During the analysis of the extracellular proteome we thus discovered the missing lactone esterase (SBLE) from the SL pathway, which is responsible for ring closure of acidic SL. In general, lactonization is a complex reaction, which demands fulfilment of multiple conditions. We certainly proved by the enzymatic assays that the identified SBLE can produce the lactone form. We confirmed its action at only two pH-values and two temperatures while it would be still interesting to test a wider range of those parameters to learn more about the optimal conditions for SBLE action. Moreover, we showed that the enzyme does not have any hydrolysis activity, either for the triglycerides, lactonic SL or paranitrophenyl esters and we demonstrate SBLE has a higher preference for esterification then hydrolysis, similar as its closest homolog CALA. We also assume that SBLE's preferred substrate will be *cis* fatty acids, because most of the SL have a *cis* FA tail [9]. We still need to confirm this, because it closest homolog CALA has specific preferences to trans fatty acids. From the analysis of the products, di- and monoacetylated lactonic SL, we predict that the favored substrate needs to have a sugar head decorated with one or two acetyl group. However, SBLE activity towards other substrates should still be tested. The best candidates would be hydroxylated fatty acids, whose lactones are of high importance in the perfume and pharmaceutical industry. These assays should be performed both in water and organic solvent and a possible addition of SL should be considered.

In general the ability of a lipase to lactonize hydroxylated fatty acids or sophorolipids *in vitro* was claimed as an uncommon feature of lipases, but a specific characteristic for some of them [123]. It was expected that special structural features are necessary to catalyse the synthesis of macrolactones. However to understand these and gain deeper insights, a molecular structure is required. We proposed that SBLE can operate via the common mechanism of serine hydrolases in which the catalytic serine residue is assisted by the conserved His/Asp residues. However, this mechanism does not explain the preference of esterification above hydrolysis. Moreover, when the sequence of CALA was compared with SBLE, we predicted the presence of a discrete helical lid in SBLE. This lid could be potentially involved in interfacial activation. The structures of closed and open states of the protein would be key to explain the molecular mechanism of that process.

Secondly, the chemical environment of the lactonizing enzyme could be a crucial factor for the success of the reaction. We expect that sophorolipids, as biosurfactants, can have a great impact on SBLE. In general surfactants can compete with oils for adsorption at the oil-protein interface or they can form micellar aggregates that may interact with hydrophobic parts of lipases in solution [229]. These various interactions have significant effects on the conformation of enzymes, and as such they can either promote or decrease the activity of lipases. In most cases the activation of a lipase is triggered by conformational changes in the presence of lipids and/or surfactants. Proteins interact

differently with micelles as compared to monomeric surfactants, e.g. some proteins can be denatured by surfactant monomers while other proteins can only be denatured by surfactant micelles [230]. The interaction between protein and SLs could be tested, with monomers and micelles, mimicking the extracellular environment during the SL production phase. Understanding such a system at a molecular level would be undiscernibly profitable for the industry and science. The ability to perform esterification in an aqueous environment reduces time and energy, and at the same time it will be more economical and most importantly, environmental friendly.

Finally, to perform all mentioned future studies we need to improve the yield of SBLE expression. We could posibbly improve our expression strategy in *P. pastoris* for example by increasing the copy number of the expression cassette or by modifying our purification strategy. This would also be an important aspect in the case of the eventual commercialization of SBLE.

Chapter 6

Summary

Surfactants are present in our everyday life as the active agent of soaps, shampoos and detergents. Most of the surfactants are produced chemically, often with the use of by-products from oil refinery processes. However, such compounds have a low biodegradability, leading to significant problems in aqueous ecosystems. A growing awareness for the ecological consequences resulted in higher demands for novel bio-based products. Environmental friendly substitutes for surfactants are termed biosurfactants, as they are produced by microorganisms. Biosurfactants are preferred over their chemically produced counterparts because they can be produced under mild conditions, display a lower toxicity, higher biodegradability and environmental compatibility. Especially the glycolipid biosurfactants, which are composed of sugar and fatty acids, caught the attention of the academic and industrial world. They can be produced in high yields using renewable resources, which is in line with the current policy towards a more sustainable industry. According to 'Transparency Market Research', the global biosurfactants market has grown fast over the last few years and has a great prognosis for the future [4].

In this thesis, we focused on a specific class of glycolipid biosurfactants: sophorolipids, composed of sophorose linked to C-16 or C-18 hydroxylated fatty acids. Over the last decade, the unique properties of sophorolipids found many applications in different areas. Their ability to efficiently lower the surface tension, connected with their biodegradability made them attractive compounds to produce a new generation of environmental friendly cleaning products. Sophorolipids are also claimed to have protective effects on the skin, hair and nails, and, consequently, they are interesting substrates for cosmetic and dermatological products [64]. Moreover, sophorolipids recently found an application in nanotechnology as capping molecules for nanoparticles which improves their polarity and solubility in solvents [68], and could therefore find great applications in carrying various biomolecules [69]. The aim of this thesis was to discover new players in natural SL biosynthesis and provide an overview of the cellular processes that participate in SL production. In this work we performed a detailed proteomic study on the best SL producer, the non-pathogenic yeast *Starmerella bombicola* ATCC 22214. This strain is already used in the industry and, therefore, it was selected as a model to obtain more knowledge on the mechanism of sophorolipids biosynthesis. Moreover, we could rely on a genomic database for *S. bombicola*, resulting from the research of an interdisciplinary research platform at Ghent University, including the research groups of Prof. Yves Vandepeer and Prof. Wim Soetaert. They selected this strain as a platform for the production of new-to-nature molecules.

To have full control over sophorolipid production in an industrial setting, the biochemistry underlying its biosynthesis needs to be completely understood. Therefore, in this thesis we used a quantitative proteomic technology, SILAC, to compare *S. bombicola*'s intracellular proteome between a sophorolipid producing and non-producing state. This study provides the first insights in the proteome of this highly important industrial strain. The acquired information can be used for introducing genetic modifications in order to increase the yield of SL or to produce new-to-nature biosurfactants. We confirmed that the proposed enzymes for the sophorolipid production pathway are co-ordinately produced in the stationary phase at 2 magnitudes of order higher than in exponentially grown cells. Membrane protein enrichment also allowed us to confirm the overexpression of the ABC transporter which was previously discovered in the SL cluster. The fact that we were able to detect SL cluster proteins with the applied strategy is encouraging to perform similar studies in other uncharacterized biosurfactant producers. We also observed a lot of changes

in protein abundances during the transition from the exponential growth phase to the SL producing stationary phase, and not surprisingly the majority of these differences can be explained in the context of changing growth phases. The largest group of down-regulated proteins in the stationary phase is involved in protein translation and maturation and amino acid biogenesis. The enzymes from the glyoxylate cycle, which bypasses the citric cycle, were significantly up-regulated during that phase. We did not collect much information about fatty acid synthesis enzymes except for the significant decrease of proteins participating in the ergosterol synthesis. There are still other puzzling observations that warrant further investigation. It would be highly interesting to employ a metabolomics strategy to expand our knowledge about the fatty acid metabolism and general glycolysis/gluconeogenesis fluxes in *S. bombicola* during SL production.

In particular, we focussed our studies on the extracellular proteome. This research aimed to reveal the existence of a protein responsible for sophorolipid lactonization and generally to characterize the extracellular proteome. The examined *S. bombicola* exoproteome is similar in size and content to other biotechnologically and medically important yeasts like *Candida albicans* [112], *Candida utilis* [111] or *Pichia pastoris* [191]. However in contrast to those reports, *S. bombicola*'s exoproteome contains a large amount of proteases. An interesting discovery was the detection of an invertase homologue in the *S. bombicola* exoproteome since our growth medium conditions were not conductive for its expression. This enzyme is possibly not attractive from a commercial point of view but may be good example to illustrate the concept of bee-yeast cooperation. Its presence in the exoproteome can be connected with the natural habitat of *S. bombicola* where its expression is possibly utilized to extract nutrients. However our main discovery in the exoproteome was the missing enzyme (SBLE) in the SL biosynthetic pathway responsible for the lactonization of acidic SL. A single knock-out of this lipase/esterase gene resulted in the complete absence of any lactonic form of sophorolipids in favour of acidic SL. While the wild-type strain produces a mixture of both structures, the insertion of a second copy of SBLE in the genome resulted in the production of mainly lactonic sophorolipids [222]. Our discovery thus allows for a better control over SL biosynthesis. The homogeneity in SL structures is especially advantageous for large-scale productions in fermentors. The time and energy necessary for the chemical post-fermentational modifications can thereby be significantly reduced, while the unique properties of the obtained SL molecules, like foaming capacity or water solubility, will definitely show better performances for several applications such as their use in detergents or in pharmaceutical applications.

Finally, we showed by enzymatic assays that the enzyme identified in the *S. bombicola* exoproteome lactone esterase (SBLE) can produce the SL lactone form. From the analysis of the product, di- and monoacetylated lactonic SL, we predict that the favored substrates need to have a sugar head decorated with one or two acetyl group. We believe that SBLE belongs to a narrow group of enzymes able to perform internal esterification. In the near future SBLE activity towards other substrates will be tested, namely hydroxylated fatty acids, whose lactones are of high importance in the perfume and pharmaceutical industry. However the reaction mechanism of SBLE is not yet known. In addition, we did not manage to show any hydrolysis activity of SBLE, either on triglycerides, lactonic SL or paranitrophenyl esters and we suspect SBLE has a higher preference for esterification then hydrolysis, contrary to its closest homolog CALA. This extraordinary behaviour, namely esterification in an aqueous solution, could be possibly explained by unusual structural features. On the other hand, SL as a biosurfactant can also create a pro-lactonic environment. In general, surfactants can

activate or deactivate/unfold proteins [230]. Moreover, SL form a different type of aggregates depending on the pH and their concentration which can also influence the lactonization reaction. For example, micellar aggregates may interact with hydrophobic parts of lipases in solution [229]. Insight into that molecular mechanism would be undiscernibly profitable for the industry and science. The ability to perform esterification in an aqueous environment reduces time and energy, and at the same time it will be more economical and, most importantly, environmental friendly.

This thesis illustrates how proteomic data can be useful to tackle problems in biotechnology. The presented data can guide future molecular biology studies and boost the creation of new mutants in *S. bombicola*. While the story of the lactone esterase just begins and opens the door to the next level of understanding sophorolipids lactonization, further enzymatic and structural studies should be undertaken as this enzyme has a commercial potential.

Samenvatting

Surfactanten zijn de actieve bestanddelen van dagdagelijkse producten als zepen, shampoos en detergenten. De meeste surfactanten worden chemisch geproduceerd, vaak als bijproducten van olieraffinaderijen. Deze chemische surfactanten zijn echter weinig biologisch afbreekbaar en veroorzaken dan ook significante problemen in waterige ecosystemen. Een toenemende bewustwording van deze problematiek resulteerde in een hogere vraag naar milieuvriendelijke 'biosurfactanten'. De term 'bio' refereert hierbij naar het feit dat deze surfactanten geproduceerd worden door micro-organismen en dus van biologische oorsprong zijn. Een van de voordelen van biosurfactanten is dat ze een eenvoudiger productieproces kennen t.a.v. hun chemische tegenhangers. Bovendien vertonen ze een lagere toxiciteit en een hogere biodegradeerbaarheid waardoor ze dus meer compatibel zijn met het milieu. Vooral de glycolipide biosurfactanten, die opgebouwd zijn uit suikers en vetzuren genieten veel interesse van de academische en industriële wereld. Dit type biosurfactanten kan namelijk efficiënt geproduceerd worden op basis van hernieuwbare grondstoffen. Volgens het gerenommeerde marktonderzoeksbureau 'Transparency Market Research' heeft de globale markt van biosurfactanten, die de voorbije jaren sterk gegroeid is, dan ook gunstige vooruitzichten [4].

In deze thesis werd gefocust op een specifieke klasse van glycolipide biosurfactanten, genaamd sophorolipiden (SL). SL zijn opgebouwd uit sophorose gelinkt aan C-16 of C-18 gehydroxyleerde vetzuren. In de voorbije tien jaar hebben sophorolipiden toepassingen gevonden in verschillende domeinen. Hun vermogen om de oppervlaktespanning te verlagen, in combinatie met hun biodegradeerbaarheid, maakt van sophorolipiden aantrekkelijke verbindingen voor de productie van een nieuwe generatie van milieuvriendelijke schoonmaakproducten. Sophorolipiden zouden ook een gunstig effect hebben op de huid, haar en nagels, en het zijn dan ook interessante substraten voor cosmetische en dermatologische producten [64]. Sophorolipiden vonden recent ook een toepassing in de nanotechnologie. Door ze te gebruiken als 'capping' moleculen kan de polariteit en oplosbaarheid van nanopartikels verhoogd worden [68], . Daardoor kunnen sophorolipiden een toepassing vinden als 'carrier' partikels van biologische moleculen [69]..

Dit doctoraal onderzoek had tot doel de cellulaire processen die verband houden met SL-productie te karakteriseren en mogelijks nieuwe moleculaire spelers te ontdekken die een rol spelen in de biologische productie van SL. In dit werk werd daartoe een gedetailleerde proteomische studie uitgevoerd op de niet-pathogene gist, *Starmerella bombicola* ATCC 22214. Deze streng is gekend als de beste natuurlijke produceerder van SL. In dit onderzoek konden we beroep doen op een genomische databank van S. *bombicola*. De beschikbaarheid van deze databank is het resultaat van een interdisciplinair onderzoeksplatform aan de Universiteit van Gent, waarbij de onderzoeksgroepen van Prof. Yves Vandepeer en Prof. Wim Soetaert betrokken zijn. Zij hadden deze gist namelijk geselecteerd voor de productie van nieuwe, natuurlijke surfactanten.

Om een optimale controle van de industriële productie van sophorolipiden te bekomen, dient de biochemie die aan de basis ligt van de biosynthese volledig begrepen te worden. Hiertoe vergeleken wij het intracellulaire proteoom van *S. bombicola* onder niet-producerende omstandigheden met het proteoom van SL-producerende cellen d.m.v. de SILAC methode. Deze studie vormt tegelijkertijd ook de eerste karakterisering van het proteoom van dit belangrijk industrieel organisme. De kennis

die uit deze studie kan voortvloeien kan dan aangewend worden om genetische modificaties in *S. bombicola* te introduceren die de opbrengst van SL verhogen of leiden tot de productie van nieuwe, biologisch geproduceerde SL.

Wij konden bevestigen dat de voorgestelde enzymen van de SL-biosynthese pathway gecoördineerd geëxpresseerd worden in de stationaire groeifase, tot niveaus die twee orden van grootte hoger zijn dan in de exponentiele groeifase. Bovendien bevestigde een experiment met een membraaneiwit-aangereikte fractie ook de overexpressie van de ABC transporter die deel uitmaakt van de SL gencluster. Het feit dat wij in staat waren om de eiwitten van de SL cluster te detecteren in onze strategie is aanmoedigend om een soortgelijk onderzoek uit te voeren in ongekarakteriseerde biosurfactant-producerende organismen.

We observeerden ook vele veranderingen in eiwitabundanties gedurende de transitie van de exponentiële naar de stationaire groeifase. De differentiële expressie van vele van deze eiwitten kan toegeschreven worden aan de metabolische veranderingen die de cel ondergaat bij deze transitie. De grootste groep van downgereguleerde eiwitten in de stationaire fase is betrokken bij eiwittranslatie en -maturatie en aminozuursynthese. De enzymen van de glyoxylaatcyclus, die zorgen voor een omzeiling van de citroenzuurcyclus, waren daarentegen significant opgereguleerd. Behalve de significante downregulatie van enzymen betrokken bij ergosterolsynthese, observeerden we eerder onverwacht geen grote verandering in de abundantie van enzymen betrokken bij vetzuursynthese. Ter uitbreiding van deze studie, zou het zeer interessant zijn om via metabolomics de vetzuursynthese en flux van glycolyse/gluconeogenese bij de overgang naar de SL-producerende fase te bestuderen.

Een van onze proteomische studies was expliciet gericht op het extracellulaire proteoom van *S. bombicola*. Deze strategie had naast een algemene karakterisering van het extracellulaire proteoom als doel een extracellulair eiwit te identificeren dat verantwoordelijk is voor de lactonisatie van sophorolipiden. Onze bevindingen wijzen erop dat het *S. bombicola* exoproteoom wat betreft grootte en inhoud vergelijkbaar is met dat van andere biotechnologisch en medisch interessante gisten zoals *Candida albicans* [112], *Candida utilis* [111] en *Pichia pastoris* [191]. Opvallend was dat het exoproteoom van *S. bombicola* veel proteasen bevat.

Wij konden ook een invertase homoloog in het *S. bombicola* exoproteoom detecteren. Dit was eerder onverwacht omdat onze groeicondities niet inducerend zijn voor de expressie ervan. Hoewel onze enzymatische karakterisering een beperkt commercieel potentieel voor dit invertase voorspelt, illustreert dit enzym wel mogelijks de rol van gisten bij het plant-bij mutualisme. De aanwezigheid in het exoproteoom kan in verband staan met de natuurlijke habitat van *S. bombicola* en de expressie ervan kan geïnduceerd worden onder de hoge osmotische druk veroorzaakt door de suikers in de natuurlijke leefomgeving.

Onze belangrijkste ontdekking in het exoproteoom was zonder meer het enzym SBLE dat verantwoordelijk is voor de lactonisatie van zure, open SL. Een knockout van dit lipase/esterase gen resulteerde in de volledige afwezigheid van de SL-lactonvorm ten voordele van open SL. Terwijl de wild type streng een mengsel van beide SL structuren produceert, resulteerde de insertie van een tweede kopij van het SBLE-gen in het genoom van *S. bombicola* in de productie van vnl. lacton-SL [222]. Onze bevindingen laten dus een betere controle toe over de SL biosynthese. De verhoogde

homogeniteit van de SL-structuren is vooral voordelig bij grootschalige producties in fermentors. De tijd en energie nodig voor de chemische post-fermentatieve modificaties kan daardoor significant verminderd worden, terwijl de uniforme eigenschappen van de verkregen SL moleculen, zoals schuimcapaciteit en wateroplosbaarheid, zeker toepassingen zullen vinden in de detergent- en farmaceutische industrie.

Tenslotte toonden we d.m.v. enzymatische *in vitro* assays aan dat het SBLE-enzym verantwoordelijk is voor de productie van lacton-SL. Uit de analyse van de producten, di- en monogeacetyleerde lactonische SL, blijkt dat de substraten bij voorkeur een suikergroep hebben met 1 of 2 acetylgroepen. Wij geloven dat SBLE behoort tot een kleine groep van enzymen die interne esterificatie kunnen uitvoeren. In de nabije toekomst zal de SBLE-activiteit naar andere types substraten toe zoals gehydroxyleerde vetzuren getest worden. De lactonesters van deze laatste vetzuren zijn namelijk uitermate belangrijk in de parfum- en farmaceutische industrie.

Het werkingsmechanisme van SBLE blijft voorlopig nog onduidelijk. Wij konden geen hydrolyse-activiteit van SBLE aantonen t.a.v. triglyceriden, lacton-SL of paranitrophenylesters. We vermoeden dus dat SBLE een sterke voorkeur heeft voor de esterificatiereactie, in tegenstelling tot het dichtste SBLE-homoloog, CALA, dat hydrolyse prefereert. De buitengewone eigenschap van SBLE, namelijk esterificatie in een waterige omgeving, kan mogelijks verklaard worden door unieke structurele eigenschappen. Anderszijds, kunnen SL, gezien het potentieel van surfactanten om de activiteit van enzymen te beinvloeden [230], ook een pro-lacton omgeving creëren door te interageren met SBLE. Hierbij is het interessant op te merken dat SL verschillend types aggregaten vormen, afhankelijk van de pH en hun concentratie. Micel-aggregaten zouden bv. kunnen interageren met hydrofobe regio's van lipases in oplossing [229]. Een gedetailleerde kennis van het exacte mechanisme zou ten gunste komen van de industrie en wetenschap. De mogelijkheid om esterificatie in een waterige omgeving uit te voeren spaart tijd en energie uit, en tegelijkertijd is dit proces meer economisch en milieuvriendelijk.

Deze thesis illustreert hoe proteomische data kan gebruikt worden om biotechnologische problemen te benaderen. De bekomen kennis over de SL biosynthese kan toekomstige moleculair biologische studies sturen en de creatie van nieuwe mutanten van *S. bombicola* versnellen. Terwijl het verhaal van het SBLE lacton-esterase net begonnen is en de deur opent naar een volgend niveau in ons begrip van SL lactonisatie, zouden - gezien het hoge commerciële potentieel van dit enzym - verdere enzymatische en structurele studies op SBLE ondernomen moeten worden.

References

1. Laurier L. Schramm, E.N.S.a.D.G.M., *Surfactants and their applications.* Annu. Rep. Prog. Chem., Sect. C,, 2003. **99**: p. 3-48.
2. Ron, E.Z. and E. Rosenberg, *Natural roles of biosurfactants.* Environ Microbiol, 2001. **3**(4): p. 229-36.
3. Pacwa-Plociniczak, M., et al., *Environmental applications of biosurfactants: recent advances.* Int J Mol Sci, 2011. **12**(1): p. 633-54.
4. (PRWEB), A.N. *Global Biosurfactants Market Analyzed by Transparency Market Research* http://www.prweb.com/releases/2012/11/prweb10158903.htm. 2012.
5. Kitamoto, D., H. Isoda, and T. Nakahara, *Functions and potential applications of glycolipid biosurfactants--from energy-saving materials to gene delivery carriers.* J Biosci Bioeng, 2002. **94**(3): p. 187-201.
6. Syldatk, C., et al., *Chemical and physical characterization of four interfacial-active rhamnolipids from Pseudomonas spec. DSM 2874 grown on n-alkanes.* Z Naturforsch C, 1985. **40**(1-2): p. 51-60.
7. Kitamoto D, H.K., Nakahara T, Tabuchi T, *Production of mannosylerythritol lipids by Candida antarctica from vegetable-oils.* Agric Biol Chem, 1990. **54**: p. 37-40.
8. Saharan, B.S., RK.; Sharma, D.;, *A Review on Biosurfactants: Fermentation, Current Developments and Perspectives.* Genetic Engineering and Biotechnology Journal, 2011. **29**.
9. Asmer, H.J., et al., *Microbial-Production, Structure Elucidation and Bioconversion of Sophorose Lipids.* Journal of the American Oil Chemists Society, 1988. **65**(9): p. 1460-1466.
10. Davila, A.M., R. Marchal, and J.P. Vandecasteele, *Kinetics and Balance of a Fermentation Free from Product Inhibition - Sophorose Lipid Production by Candida-Bombicola.* Appl Microbiol Biotechnol, 1992. **38**(1): p. 6-11.
11. Van Bogaert, I.N., et al., *Microbial production and application of sophorolipids.* Appl Microbiol Biotechnol, 2007. **76**(1): p. 23-34.
12. Van Bogaert, I.N.A., J.X. Zhang, and W. Soetaert, *Microbial synthesis of sophorolipids.* Process Biochemistry, 2011. **46**(4): p. 821-833.
13. Lang, S.K., E.; Wagner, F.; , *Antimicrobial effectes of biosurfactants.* Fett Wiss Technol- Fat Sci Technol, 1989. **91**: p. 363-6.
14. Shah, V., et al., *Sophorolipids, microbial glycolipids with anti-human immunodeficiency virus and sperm-immobilizing activities.* Antimicrob Agents Chemother, 2005. **49**(10): p. 4093-100.
15. Romsted, L.S., *Introduction to Surfactant Self-Assembly.* Supramolecular Chemistry: From Molecules to Nanomaterials, 2012.
16. Nagarajan, R., *Theory of Surfactant Self-Assembly: A Predictive Molecular Thermodynamic Approach* Langmuir, 1991(7): p. 2934-2969
17. Zhou, S., et al., *Supramolecular assemblies of a naturally derived sophorolipid.* Langmuir, 2004. **20**(19): p. 7926-32.
18. Baccile, N., et al., *Unusual, pH-induced, self-assembly of sophorolipid biosurfactants.* ACS Nano, 2012. **6**(6): p. 4763-76.
19. Penfold, J., et al., *Solution self-assembly of the sophorolipid biosurfactant and its mixture with anionic surfactant sodium dodecyl benzene sulfonate.* Langmuir, 2011. **27**(14): p. 8867-77.
20. Tulloch AP, S.J., *Fermentation of long-chain compounds by Torulopsis apicola. IV. Products from esters and hydrocarbons with 14 and 15 carbon atoms and from methyl palmitoleate.* Can J Chem, 1986. **46**: p. 1523–1528.
21. Spencer JFT, G.P., *Torulopsis bombicola sp.n.* Antonie Van Leeuwenhoek, 1970. **36**: p. 29–133.
22. Pekin, G.V.-S., F.; Kosaric, N., *Production of sophorolipids from Candida bombicola ATCC 22214 using Turkish corn oil and honey.* Engineering in Life science, 2005. **5**: p. 357-362.

23. Rosa, C.A. 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.* Int J Syst Bacteriol, 1998. **48 Pt 4**: p. 1413-7.
24. Kurtzman, C.P., et al., *Production of sophorolipid biosurfactants by multiple species of the Starmerella (Candida) bombicola yeast clade.* FEMS Microbiol Lett, 2010. **311**(2): p. 140-6.
25. Price, N.P., et al., *Structural characterization of novel sophorolipid biosurfactants from a newly identified species of Candida yeast.* Carbohydr Res, 2012. **348**: p. 33-41.
26. Konishi, M., et al., *Production of new types of sophorolipids by Candida batistae.* J Oleo Sci, 2008. **57**(6): p. 359-69.
27. Imura, T., et al., *Enzymatic conversion of diacetylated sophoroselipid into acetylated glucoselipid: surface-active properties of novel bolaform biosurfactants.* J Oleo Sci, 2010. **59**(9): p. 495-501.
28. Sipiczki, M., *Starmerella caucasica sp. nov., a novel anamorphic yeast species isolated from flowers in the Caucasus.* J Gen Appl Microbiol, 2013. **59**(1): p. 67-73.
29. Li, S.L., et al., *Starmerella jinningensis sp. nov., a yeast species isolated from flowers of Erianthus rufipilus.* Int J Syst Evol Microbiol, 2013. **63**(Pt 1): p. 388-92.
30. Kurtzman, C.P., *Candida kuoi sp. nov., an anamorphic species of the Starmerella yeast clade that synthesizes sophorolipids.* Int J Syst Evol Microbiol, 2012. **62**(Pt 9): p. 2307-11.
31. Herrera, C.M.P., M.I.; Medrano, M., *Yeasts in nectar of an early-blooming herb: sought by bumble bees, detrimental to plant fecundity.* Ecology, 2013. **94**: p. 273-279.
32. Chen, J., et al., *Sophorolipid produced from the new yeast strain Wickerhamiella domercqiae induces apoptosis in H7402 human liver cancer cells.* Appl Microbiol Biotechnol, 2006. **72**(1): p. 52-9.
33. Thaniyavarn, J., et al., *Production of sophorolipid biosurfactant by Pichia anomala.* Biosci Biotechnol Biochem, 2008. **72**(8): p. 2061-8.
34. Tulloch AP, S.J., Deinema MH, *A new hydroxy fatty acid sophoroside from Candida bogoriensis.* Can J Chem, 1986. **46**(345–348).
35. Ashby, R.N., A.; Solaiman, DKY.; Foglia,TA.;, *Sophorolipid Biosynthesis from a Biodiesel Co-product Stream.* AOCS, 2005. **82**(9).
36. Solaiman, D.K., et al., *Simplified soy molasses-based medium for reduced-cost production of sophorolipids by Candida bombicola.* Biotechnol Lett, 2007. **29**(9): p. 1341-7.
37. Daverey, A. and K. Pakshirajan, *Sophorolipids from Candida bombicola using mixed hydrophilic substrates: production, purification and characterization.* Colloids Surf B Biointerfaces, 2010. **79**(1): p. 246-53.
38. Fleurackers, S., *On the use of waste frying oil in the synthesis of sophorolipids.* European Journal of Lipid Science and Technology, 2006. **108**: p. 5-12.
39. Davila, A.M., R.; Vandecasteele, JP., *Sophorose lipid production from lipidic precursors-predictive evaluationof idustrail substrates.* Journal of Industrial Microbiology, 1994. **13**: p. 249-257.
40. Hommel, R.K., et al., *Production of Sophorose Lipid by Candida (Torulopsis) Apicola Grown on Glucose.* Journal of Biotechnology, 1994. **33**(2): p. 147-155.
41. Van Bogaert, I.N., et al., *Sophorolipid production by Candida bombicola on oils with a special fatty acid composition and their consequences on cell viability.* Biotechnol Lett, 2010. **32**(10): p. 1509-14.
42. Van Bogaert, I., et al., *Production of new-to-nature sophorolipids by cultivating the yeast Candida bombicola on unconventional hydrophobic substrates.* Biotechnol Bioeng, 2011. **108**(4): p. 734-41.
43. Azim, A., et al., *Amino acid conjugated sophorolipids: A new family of biologically active functionalized glycolipids.* Bioconjug Chem, 2006. **17**(6): p. 1523-9.
44. Singh, S.K., et al., *Regioselective enzyme-catalyzed synthesis of sophorolipid esters, amides, and multifunctional monomers.* J Org Chem, 2003. **68**(14): p. 5466-77.

45. Bisht, K.S., R.A. Gross, and D.L. Kaplan, *Enzyme-Mediated Regioselective Acylations of Sophorolipids.* J Org Chem, 1999. **64**(3): p. 780-789.
46. Van Bogaert, I.N., et al., *Importance of the cytochrome P450 monooxygenase CYP52 family for the sophorolipid-producing yeast Candida bombicola.* FEMS Yeast Res, 2009. **9**(1): p. 87-94.
47. Saerens, K.M., et al., *Identification of the UDP-glucosyltransferase gene UGTA1, responsible for the first glucosylation step in the sophorolipid biosynthetic pathway of Candida bombicola ATCC 22214.* FEMS Yeast Res, 2011. **11**(1): p. 123-32.
48. Saerens, K.M., et al., *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, 2011. **28**(4): p. 279-92.
49. Saerens, K.M., L. Saey, and W. Soetaert, *One-step production of unacetylated sophorolipids by an acetyltransferase negative Candida bombicola.* Biotechnol Bioeng, 2011. **108**(12): p. 2923-31.
50. Keller, N.P. and T.M. Hohn, *Metabolic pathway gene clusters in filamentous fungi.* Fungal Genetics and Biology, 1997. **21**(1): p. 17-29.
51. Hewald, S., et al., *Identification of a gene cluster for biosynthesis of mannosylerythritol lipids in the basidiomycetous fungus Ustilago maydis.* Appl Environ Microbiol, 2006. **72**(8): p. 5469-77.
52. Teichmann, B., et al., *Identification of a biosynthesis gene cluster for flocculosin a cellobiose lipid produced by the biocontrol agent Pseudozyma flocculosa.* Molecular Microbiology, 2011. **79**(6): p. 1483-95.
53. Van Bogaert, I.N., et al., *The biosynthetic gene cluster for sophorolipids: a biotechnological interesting biosurfactant produced by Starmerella bombicola.* Mol Microbiol, 2013.
54. Van Bogaert, I.N., et al., *The biosynthetic gene cluster for sophorolipids: a biotechnological interesting biosurfactant produced by Starmerella bombicola.* Mol Microbiol, 2013. **88**(3): p. 501-9.
55. Teichmann, B., et al., *Activation of the ustilagic acid biosynthesis gene cluster in Ustilago maydis by the C2H2 zinc finger transcription factor Rua1.* Appl Environ Microbiol, 2010. **76**(8): p. 2633-40.
56. Hommel, R.K., Stegner,S., Kleber, H. and Weber, L., *Effect of ammonium ions on glycolipid production by Candida (Torulopsis) apicola.* Appl Microbiol Biotechnol, 1994. **42**(2-3): p. 192-197.
57. Davila, A.M.M., R.; Vandecasteele, J. P. , *Kinetics and balance of a fermentation free from product inhibition: sophorose lipid production by Candida bombicola.* Appl Microbiol Biotechnol, 1992. **38**(1): p. 6-11.
58. Albrecht, A., U. Rau, and F. Wagner, *Initial steps of sophoroselipid biosynthesis by Candida bombicola ATCC 22214 grown on glucose.* Appl Microbiol Biotechnol, 1996. **46**(1): p. 67-73.
59. Baviere, M.D., D.; Lecortier, J.;, *Process for washing solid particles comprising sophoroside solution.* US Patent 5326407, 1994.
60. Marchal, R.L., J.; Sulzer, C.; Davila, AM.;, *Production of the sophorolipid acetate acids form oils or esters.* US Patent 5900366, 1999.
61. Ducreusx, J.B., D.; Baviere, M.; Bocard, C.; Monin, N.;, *Composition containinga surface active compound and glycolipids and decontamination process for a porous medium polluted hydrocarbons.* US Patent 5654192 1997.
62. Mulligan, C.N., R.N. Yong, and B.F. Gibbs, *Heavy metal removal from sediments by biosurfactants.* J Hazard Mater, 2001. **85**(1-2): p. 111-25.
63. Akari, S.A., Y.;, *Methods of modyfing quality of wheat flour product.* Japanese Patent 61205449, 1987.
64. Morya, V.K., et al., *Medicinal and Cosmetic Potentials of Sophorolipids.* Mini Rev Med Chem, 2013.

65. Borzeix, C., *Use of the sophorolipids comprising diacetyl lactones as agents stimulating skin fibroplast metabolism.* World Patent 99/62479, 1999.
66. Maingault, M., *Utilization of the sophorolipids as therapeutically active substances or cosmetic product, in particular for the treatment of the skin.* . US Patent 5981497, 1999.
67. Chen, J.S., X.; Zhang, H.; Qu, YB.; Miao, YJ.;, *Production, structure elucidation and anticancer properties of sophorolipidform Wickerhamiella domerqiae.* Enzyme Microb Technol, 2006. **39**: p. 501-506.
68. Kasture, M., et al., *Multiutility sophorolipids as nanoparticle capping agents: synthesis of stable and water dispersible Co nanoparticles.* Langmuir, 2007. **23**(23): p. 11409-12.
69. Baccile, N., et al., *Sophorolipids-functionalized iron oxide nanoparticles.* Phys Chem Chem Phys, 2013. **15**(5): p. 1606-20.
70. Baccile, N.N., N.; Malfatt, IL.; Van Bogaert, ING.; Soetaert, W.; Babonneau, F.; , *Sophorolipids: a yeast-derived glycolipid as a greener structure directing agent for self-assembled nanomaterials.* Green Chem, 2010. **12**: p. 1564-7.
71. Inoue, S.M., N.;, *Process of producing a hydroksy fatty acid esters.* US Patent 4197166, 1980.
72. Wilkins, M.R., et al., *From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis.* Biotechnology (N Y), 1996. **14**(1): p. 61-5.
73. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.
74. Goodacre, R., *Metabolomics – the way forward.* Metabolomics, 2005. **1**(1): p. 1-2.
75. de Godoy, L.M., et al., *Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system.* Genome Biology, 2006. **7**(6): p. R50.
76. Elliott, M.H., et al., *Current trends in quantitative proteomics.* J Mass Spectrom, 2009. **44**(12): p. 1637-60.
77. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. **17**(10): p. 994-9.
78. Ong, S.E., L.J. Foster, and M. Mann, *Mass spectrometric-based approaches in quantitative proteomics.* Methods, 2003. **29**(2): p. 124-30.
79. DeSouza, L.V. and K.W. Siu, *Mass spectrometry-based quantification.* Clin Biochem, 2013. **46**(6): p. 421-31.
80. Oda, Y., et al., *Accurate quantitation of protein expression and site-specific phosphorylation.* Proc Natl Acad Sci U S A, 1999. **96**(12): p. 6591-6.
81. Bachi, A. and T. Bonaldi, *Quantitative proteomics as a new piece of the systems biology puzzle.* J Proteomics, 2008. **71**(3): p. 357-67.
82. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.
83. Ong, S.E. and M. Mann, *A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC).* Nat Protoc, 2006. **1**(6): p. 2650-60.
84. Zhu, W., J.W. Smith, and C.M. Huang, *Mass spectrometry-based label-free quantitative proteomics.* J Biomed Biotechnol, 2010. **2010**: p. 840518.
85. Ishihama Y, O.Y., Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M., *Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.* Mol Cell Proteomics., 2005. **4**(9): p. 1265-72.
86. Botstein, D., S.A. Chervitz, and J.M. Cherry, *Yeast as a model organism.* Science, 1997. **277**(5330): p. 1259-60.
87. Goffeau, A., et al., *Life with 6000 genes.* Science, 1996. **274**(5287): p. 546, 563-7.
88. Goto, S., et al., *Expression and localization of the Na-K-2Cl cotransporter in the rat cochlea.* Brain Res, 1997. **765**(2): p. 324-6.

89. Shevchenko, A., et al., *Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels.* Proc Natl Acad Sci U S A, 1996. **93**(25): p. 14440-5.
90. Picotti, P., et al., *A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis.* Nature, 2013. **494**(7436): p. 266-70.
91. Kniemeyer, O., et al., *Identification of virulence determinants of the human pathogenic fungi Aspergillus fumigatus and Candida albicans by proteomics.* Int J Med Microbiol, 2011. **301**(5): p. 368-77.
92. Trabalzini, L., et al., *Proteomic response to physiological fermentation stresses in a wild-type wine strain of Saccharomyces cerevisiae.* Biochem J, 2003. **370**(Pt 1): p. 35-46.
93. Smart, K.A., *Brewing yeast genomes and genome-wide expression and proteome profiling during fermentation.* Yeast, 2007. **24**(11): p. 993-1013.
94. Dragosits, M., et al., *The response to unfolded protein is involved in osmotolerance of Pichia pastoris.* BMC Genomics, 2010. **11**: p. 207.
95. Liu, H., et al., *The proteome analysis of oleaginous yeast Lipomyces starkeyi.* FEMS Yeast Res, 2011. **11**(1): p. 42-51.
96. Hammami, W., et al., *Proteomic analysis of the metabolic adaptation of the biocontrol agent Pseudozyma flocculosa leading to glycolipid production.* Proteome Sci, 2010. **8**: p. 7.
97. Nombela, C., C. Gil, and W.L. Chaffin, *Non-conventional protein secretion in yeast.* Trends Microbiol, 2006. **14**(1): p. 15-21.
98. Nickel, W. and C. Rabouille, *Mechanisms of regulated unconventional protein secretion.* Nat Rev Mol Cell Biol, 2009. **10**(2): p. 148-55.
99. Ding, Y., et al., *Unconventional protein secretion.* Trends Plant Sci, 2012. **17**(10): p. 606-15.
100. Duran, J.M., et al., *Unconventional secretion of Acb1 is mediated by autophagosomes.* J Cell Biol, 2010. **188**(4): p. 527-36.
101. Kuchler, K., R.E. Sterne, and J. Thorner, *Saccharomyces cerevisiae STE6 gene product: a novel pathway for protein export in eukaryotic cells.* EMBO J, 1989. **8**(13): p. 3973-84.
102. Abrahamsen, H. and H. Stenmark, *Protein secretion: unconventional exit by exophagy.* Curr Biol, 2010. **20**(9): p. R415-8.
103. Truesdell, P.F., et al., *Molecular characterization of a cDNA from the true armyworm Pseudaletia unipuncta encoding Manduca sexta allatotropin peptide(1).* Insect Biochem Mol Biol, 2000. **30**(8-9): p. 691-702.
104. Dranginis, A.M., et al., *A biochemical guide to yeast adhesins: glycoproteins for social and antisocial occasions.* Microbiol Mol Biol Rev, 2007. **71**(2): p. 282-94.
105. Insenser, M.R., et al., *Gel and gel-free proteomics to identify Saccharomyces cerevisiae cell surface proteins.* J Proteomics, 2010. **73**(6): p. 1183-95.
106. Cappellaro, C., V. Mrsa, and W. Tanner, *New potential cell wall glucanases of Saccharomyces cerevisiae and their involvement in mating.* J Bacteriol, 1998. **180**(19): p. 5030-7.
107. Pierleoni, A., P.L. Martelli, and R. Casadio, *PredGPI: a GPI-anchor predictor.* BMC Bioinformatics, 2008. **9**: p. 392.
108. Mattanovich, D., et al., *Genome, secretome and glucose transport highlight unique features of the protein production host Pichia pastoris.* Microb Cell Fact, 2009. **8**: p. 29.
109. Madinger, C.L., et al., *The effect of carbon source on the secretome of Kluyveromyces lactis.* Proteomics, 2009. **9**(20): p. 4744-54.
110. Swaim, C.L., et al., *Physical and computational analysis of the yeast Kluyveromyces lactis secreted proteome.* Proteomics, 2008. **8**(13): p. 2714-23.
111. Buerth, C., et al., *Growth-dependent secretome of Candida utilis.* Microbiology, 2011. **157**(Pt 9): p. 2493-503.
112. Sorgo, A.G., et al., *Mass spectrometric analysis of the secretome of Candida albicans.* Yeast, 2010. **27**(8): p. 661-72.

113. Stead, D.A., et al., *Impact of the transcriptional regulator, Ace2, on the Candida glabrata secretome.* Proteomics, 2010. **10**(2): p. 212-23.
114. Pandey, A., et al., *The realm of microbial lipases in biotechnology.* Biotechnol Appl Biochem, 1999. **29 (Pt 2)**: p. 119-31.
115. Pfeffer, J.C., *The Lipases from Candida antarctica: Cloning, Expression and their Application in the Synthesis of Structured Lipids* Universität Stuttgart , PhD thesis, 2008.
116. Svendsen, A., *Sequence comparisons within the lipase family. .* Woolley,P., Petersen,S.B. (Eds.), Lipases-their structure, biochemistry, and application, Cambridge University Press, Camridge, . 1994: p. 1-21.
117. Akoh, C.C., et al., *GDSL family of serine esterases/lipases.* Prog Lipid Res, 2004. **43**(6): p. 534-52.
118. Ollis, D.L., et al., *The alpha/beta hydrolase fold.* Protein Eng, 1992. **5**(3): p. 197-211.
119. Jaeger, K.E., B.W. Dijkstra, and M.T. Reetz, *Bacterial biocatalysts: molecular biology, three-dimensional structures, and biotechnological applications of lipases.* Annu Rev Microbiol, 1999. **53**: p. 315-51.
120. Corey, J.E.N., K.C., *Efficient and mild lactonization method for the synthesis of macrolides.* J. Am. Chem. Soc., 1974. **96**(17): p. 5614-5616.
121. Mukaiyama, T.N., K.; Kikuchi, K., *An efficient method for the synthesis of the macrocyclic lactone.* Chem. Letters, 1977: p. 441-444.
122. Inanaga, J.H., K.; Saeki, H.; Katsuki, T.; Yamaguchi, M. , *A Rapid Esterification by Means of Mixed Anhydride and Its Application to Large-ring Lactonization.* Bull. Chem. Soc. Jpn. , 1979. **52**: p. 1989–1993. .
123. Makita, A.N., T.; Yamada, Y., *Lipase catalysed synthesis of macrocyclic lactones in organic solvents.* Tetrahedron Letters, 1987. **28**(7): p. 805-808.
124. Guo Zhi-wei; Ngooi, T.K.S., A.; Fulling, G.; Sih,C.J., *Macrocylic lactones via biocatalysis in non-aqueous media.* Tetrahedron Letters, 1988. **29**(44): p. 5583-5586.
125. Ihara, F., et al., *Purification, characterization, and molecular cloning of lactonizing lipase from Pseudomonas species.* J Biol Chem, 1991. **266**(27): p. 18135-40.
126. Robinson, G.K.A., M.J.; Knowles, C.J.; Cheetham, P.S.J.; Motion, K.R. , *An investigation into the factors influencing lipase-catlyzed intramolecular lactonization in microaqueous systems.* Enzyme Microb Technol, 1994. **16**: p. 855-863.
127. Pozo, M.P., R.; Gotor, R., *Vinyl carbonates as novel alkoxycarbonylation reagents in enzymetic synthesis of carbonates.* Tetrahedron, 1992. **48**: p. 6477-6484.
128. Rees, G.D., B.H. Robinson, and G.R. Stephenson, *Macrocyclic lactone synthesis by lipases in water-in-oil microemulsions.* Biochim Biophys Acta, 1995. **1257**(3): p. 239-48.
129. Jaeger, D.A.I., J.T., *Effect of inverse micelles on the competition between lactonization and polymerization reactions of an .omega.-hydroxy carboxylic acid.* journal of organic chemistry, 1981. **46**: p. 4964-4968
130. Gargouri, M., P. Drouet, and M.D. Legoy, *Synthesis of a novel macrolactone by lipase-catalyzed intra-esterification of hydroxy-fatty acid in organic media.* Journal of Biotechnology, 2002. **92**(3): p. 259-66.
131. Aad, G., et al., *Observation of a centrality-dependent dijet asymmetry in lead-lead collisions at sqrt[S(NN)] =2.76 TeV with the ATLAS detector at the LHC.* Phys Rev Lett, 2010. **105**(25): p. 252303.
132. Lang, S. and D. Wullbrandt, *Rhamnose lipids - biosynthesis, microbial production and application potential.* Appl Microbiol Biotechnol, 1999. **51**(1): p. 22-32.
133. Shah, V., M. Jurjevic, and D. Badia, *Utilization of restaurant waste oil as a precursor for sophorolipid production.* Biotechnol Prog, 2007. **23**(2): p. 512-5.
134. de Godoy, L.M.F., et al., *Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system.* Genome Biology, 2006. **7**(6).

135. Van Bogaert, I.N., et al., *Development of a transformation and selection system for the glycolipid-producing yeast Candida bombicola.* Yeast, 2008. **25**(4): p. 273-8.
136. Lang, S., et al., *Production of native and modified sophorose lipids.* Chimica Oggi-Chemistry Today, 2000. **18**(10): p. 76-79.
137. de Graaf, D.C., et al., *Insights into the venom composition of the ectoparasitoid wasp Nasonia vitripennis from bioinformatic and proteomic studies.* Insect Mol Biol, 2010. **19 Suppl 1**: p. 11-26.
138. Stanke, M., et al., *Using native and syntenically mapped cDNA alignments to improve de novo gene finding.* Bioinformatics, 2008. **24**(5): p. 637-44.
139. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.
140. *Reorganizing the protein space at the Universal Protein Resource (UniProt).* Nucleic Acids Res, 2012. **40**(Database issue): p. D71-5.
141. Goujon, M., et al., *A new bioinformatics analysis tools framework at EMBL-EBI.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W695-9.
142. Sterck, L., et al., *ORCAE: online resource for community annotation of eukaryotes.* Nat Methods, 2012. **9**(11): p. 1041.
143. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.
144. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-11.
145. Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nat Biotechnol, 2010. **28**(5): p. 511-5.
146. Conesa, A., et al., *Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.* Bioinformatics, 2005. **21**(18): p. 3674-6.
147. Gawronski, J.D. and D.R. Benson, *Microtiter assay for glutamine synthetase biosynthetic activity using inorganic phosphate detection.* Anal Biochem, 2004. **327**(1): p. 114-8.
148. de Godoy, L.M., et al., *Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast.* Nature, 2008. **455**(7217): p. 1251-4.
149. Lundberg, E., et al., *Defining the transcriptome and proteome in three functionally different human cell lines.* Mol Syst Biol, 2010. **6**: p. 450.
150. Debose-Boyd, R.A., *A helping hand for cytochrome p450 enzymes.* Cell Metab, 2007. **5**(2): p. 81-3.
151. Hughes, A.L., et al., *Dap1/PGRMC1 binds and regulates cytochrome P450 enzymes.* Cell Metab, 2007. **5**(2): p. 143-9.
152. Mallory, J.C., et al., *Dap1p, a heme-binding protein that regulates the cytochrome P450 protein Erg11p/Cyp51p in Saccharomyces cerevisiae.* Mol Cell Biol, 2005. **25**(5): p. 1669-79.
153. Hu, S.H., et al., *Structure of TcpG, the DsbA protein folding catalyst from Vibrio cholerae.* J Mol Biol, 1997. **268**(1): p. 137-46.
154. Briknarova, K., et al., *Structural analysis of BAG1 cochaperone and its interactions with Hsc70 heat shock protein.* Nat Struct Biol, 2001. **8**(4): p. 349-52.
155. Feldman, R.M., et al., *A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p.* Cell, 1997. **91**(2): p. 221-30.
156. Wolff, A.M., N. Din, and J.G. Petersen, *Vacuolar and extracellular maturation of Saccharomyces cerevisiae proteinase A.* Yeast, 1996. **12**(9): p. 823-32.
157. Teichert, U., et al., *Lysosomal (vacuolar) proteinases of yeast are essential catalysts for protein degradation, differentiation, and cell survival.* J Biol Chem, 1989. **264**(27): p. 16037-45.

158. Horak, J., J. Regelmann, and D.H. Wolf, *Two distinct proteolytic systems responsible for glucose-induced degradation of fructose-1,6-bisphosphatase and the Gal2p transporter in the yeast Saccharomyces cerevisiae share the same protein components of the glucose signaling pathway.* J Biol Chem, 2002. **277**(10): p. 8248-54.
159. Cakar, Z.P., et al., *Vacuolar morphology and cell cycle distribution are modified by leucine limitation in auxotrophic Saccharomyces cerevisiae.* Biol Cell, 2000. **92**(8-9): p. 629-37.
160. Wu, J., et al., *Global analysis of nutrient control of gene expression in Saccharomyces cerevisiae during growth and starvation.* Proc Natl Acad Sci U S A, 2004. **101**(9): p. 3148-53.
161. Kusch, H., et al., *A proteomic view of Candida albicans yeast cell metabolism in exponential and stationary growth phases.* Int J Med Microbiol, 2008. **298**(3-4): p. 291-318.
162. Rintala, E., et al., *The ORF YNL274c (GOR1) codes for glyoxylate reductase in Saccharomyces cerevisiae.* Yeast, 2007. **24**(2): p. 129-36.
163. Praekelt, U.M. and P.A. Meacock, *HSP12, a new small heat shock gene of Saccharomyces cerevisiae: analysis of structure, regulation and function.* Molecular and General Genetics MGG, 1990. **223**(1): p. 97-106.
164. Michan, C. and C. Pueyo, *Growth phase-dependent variations in transcript profiles for thioredoxin- and glutathione-dependent redox systems followed by budding and hyphal Candida albicans cultures.* FEMS Yeast Res, 2009. **9**(7): p. 1078-90.
165. Welker, S., et al., *Hsp12 is an intrinsically unstructured stress protein that folds upon membrane association and modulates membrane function.* Mol Cell, 2010. **39**(4): p. 507-20.
166. Shamrock, V.J.S.V.J. and G.G.L.G.G. Lindsey, *A compensatory increase in trehalose synthesis in response to desiccation stress in Saccharomyces cerevisiae cells lacking the heat shock protein Hsp12p.* Canadian journal of microbiology, 2008. **54**(7): p. 559-568.
167. Ohtaki, S., et al., *Novel hydrophobic surface binding protein, HsbA, produced by Aspergillus oryzae.* Appl Environ Microbiol, 2006. **72**(4): p. 2407-13.
168. Meyer, K.H. and E. Schweizer, *Control of fatty-acid synthetase levels by exogeneous long-chain fatty acids in the yeasts Candida lipolytica and Saccharomyces cerevisiae.* Eur J Biochem, 1976. **65**(2): p. 317-24.
169. Rupcic, J., C. Milin, and V. Maric, *Effect of growth phase on the content and composition of ceramides of the hydrocarbon-assimilating yeast Candida lipolytica.* Syst Appl Microbiol, 1999. **22**(3): p. 486-91.
170. Montanes, F.M., A. Pascual-Ahuir, and M. Proft, *Repression of ergosterol biosynthesis is essential for stress resistance and is mediated by the Hog1 MAP kinase and the Mot3 and Rox1 transcription factors.* Mol Microbiol, 2011. **79**(4): p. 1008-23.
171. Pahlman, A.K., et al., *The yeast glycerol 3-phosphatases Gpp1p and Gpp2p are required for glycerol biosynthesis and differentially involved in the cellular responses to osmotic, anaerobic, and oxidative stress.* J Biol Chem, 2001. **276**(5): p. 3555-63.
172. Dulermo, T. and J.M. Nicaud, *Involvement of the G3P shuttle and beta-oxidation pathway in the control of TAG synthesis and lipid accumulation in Yarrowia lipolytica.* Metab Eng, 2011. **13**(5): p. 482-91.
173. Jamieson, D.J., *Oxidative stress responses of the yeast Saccharomyces cerevisiae.* Yeast, 1998. **14**(16): p. 1511-27.
174. Wiatrowski, H.A. and M. Carlson, *Yap1 accumulates in the nucleus in response to carbon stress in Saccharomyces cerevisiae.* Eukaryot Cell, 2003. **2**(1): p. 19-26.
175. von Plehwe, U., et al., *The Hsp70 homolog Ssb is essential for glucose sensing via the SNF1 kinase network.* Genes Dev, 2009. **23**(17): p. 2102-15.
176. Rinnerthaler, M., et al., *MMI1 (YKL056c, TMA19), the yeast orthologue of the translationally controlled tumor protein (TCTP) has apoptotic functions and interacts with both microtubules and mitochondria.* Biochim Biophys Acta, 2006. **1757**(5-6): p. 631-8.
177. Moseley, J.B. and B.L. Goode, *The yeast actin cytoskeleton: from cellular function to biochemical mechanism.* Microbiol Mol Biol Rev, 2006. **70**(3): p. 605-45.

178. Engqvist-Goldstein, A.E. and D.G. Drubin, *Actin assembly and endocytosis: from yeast to mammals.* Annu Rev Cell Dev Biol, 2003. **19**: p. 287-332.
179. Kaksonen, M., C.P. Toret, and D.G. Drubin, *A modular design for the clathrin- and actin-mediated endocytosis machinery.* Cell, 2005. **123**(2): p. 305-20.
180. Moreau, V., et al., *The yeast actin-related protein Arp2p is required for the internalization step of endocytosis.* Mol Biol Cell, 1997. **8**(7): p. 1361-75.
181. Schmidt, A., J. Kunz, and M.N. Hall, *TOR2 is required for organization of the actin cytoskeleton in yeast.* Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13780-5.
182. Saito, K., L. Tautz, and T. Mustelin, *The lipid-binding SEC14 domain.* Biochim Biophys Acta, 2007. **1771**(6): p. 719-26.
183. Kearns, M.A., et al., *Novel developmentally regulated phosphoinositide binding proteins from soybean whose expression bypasses the requirement for an essential phosphatidylinositol transfer protein in yeast.* EMBO J, 1998. **17**(14): p. 4004-17.
184. Routt, S.M., et al., *Nonclassical PITPs activate PLD via the Stt4p PtdIns-4-kinase and modulate function of late stages of exocytosis in vegetative yeast.* Traffic, 2005. **6**(12): p. 1157-72.
185. Jardon, R., C. Gancedo, and C.L. Flores, *The gluconeogenic enzyme fructose-1,6-bisphosphatase is dispensable for growth of the yeast Yarrowia lipolytica in gluconeogenic substrates.* Eukaryot Cell, 2008. **7**(10): p. 1742-9.
186. Spencer, J.F., P.A. Gorin, and A.P. Tulloch, *Torulopsis bombicola sp.n.* Antonie Van Leeuwenhoek, 1970. **36**(1): p. 129-33.
187. Develter, D., Lauryssen, L., *Properties and industrial applications of sophorolipids.* European Journal of Lipid Science and Technology, 2010(112): p. 628-638.
188. Lourith, N. and M. Kanlayavattanakul, *Natural surfactants used in cosmetics: glycolipids.* Int J Cosmet Sci, 2009. **31**(4): p. 255-61.
189. Sleiman, J.N., et al., *Sophorolipids as antibacterial agents.* Ann Clin Lab Sci, 2009. **39**(1): p. 60-3.
190. Gross, R.S., V. , *Antimicrobial propertis of various forms of the sophorolipids.* World Patent 2004/044215 A1 2004.
191. Huang, C.J., et al., *A proteomic analysis of the Pichia pastoris secretome in methanol-induced cultures.* Appl Microbiol Biotechnol, 2011. **90**(1): p. 235-47.
192. Ishihama, Y., et al., *Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.* Mol Cell Proteomics, 2005. **4**(9): p. 1265-72.
193. Petersen, T.N., et al., *SignalP 4.0: discriminating signal peptides from transmembrane regions.* Nat Methods, 2011. **8**(10): p. 785-6.
194. Goldberg, T., T. Hamp, and B. Rost, *LocTree2 predicts localization for all domains of life.* Bioinformatics, 2012. **28**(18): p. i458-i465.
195. Van Bogaert, I.N., et al., *Cloning, characterization and functionality of the orotidine-5'-phosphate decarboxylase gene (URA3) of the glycolipid-producing yeast Candida bombicola.* Yeast, 2007. **24**(3): p. 201-8.
196. Inoue, H., H. Nojima, and H. Okayama, *High efficiency transformation of Escherichia coli with plasmids.* Gene, 1990. **96**(1): p. 23-8.
197. Trumbly, R.J. and G. Bradley, *Isolation and characterization of aminopeptidase mutants of Saccharomyces cerevisiae.* J Bacteriol, 1983. **156**(1): p. 36-48.
198. Younes Ghasemi, M.M., Abdollah Ghasemian, Sara Rasoul-Amini *Experimental design of medium optimization for invertaseproduction by Pichia sp.* Journal of Food Science and Technology 2011.
199. Vernekar, M.a.S., P., *Statistical media optimization for b-fructofuranosidase production by Saccharomyces cerevisiae Nncim 3287* Journal of Cell and Tissue Research, 2011. **11**(2): p. 2827-2832.

200. A. E. Maioranol,, C.A.O.R.C.-F.R.M.P.R.M.B.G.-M.E.S.d.S.M.F.A.R., *Media optimization for β-Fructofuranosidase production by Aspergillus oryzae* Braz. J. Chem. Eng., 2012. **29**(1).
201. Kirk, O.a.W., Christensen, M. , *Lipases from Candida antarctica: Unique Biocatalysts from a Unique Origin.* Org. Proc. Res. Dev., 2002. **6**(4): p. 446-451.
202. Morrow, M.W., et al., *The Candida albicans Kar2 protein is essential and functions during the translocation of proteins into the endoplasmic reticulum.* Curr Genet, 2011. **57**(1): p. 25-37.
203. MacIntosh, G.C., et al., *Characterization of Rny1, the Saccharomyces cerevisiae member of the T2 RNase family of RNases: unexpected functions for ancient enzymes?* Proc Natl Acad Sci U S A, 2001. **98**(3): p. 1018-23.
204. Choudhary, V. and R. Schneiter, *Pathogen-Related Yeast (PRY) proteins and members of the CAP superfamily are secreted sterol-binding proteins.* Proc Natl Acad Sci U S A, 2012. **109**(42): p. 16882-7.
205. Carlson, M. and D. Botstein, *Two differentially regulated mRNAs with different 5' ends encode secreted with intracellular forms of yeast invertase.* Cell, 1982. **28**(1): p. 145-54.
206. Miller, K.A., L. DiDone, and D.J. Krysan, *Extracellular secretion of overexpressed glycosylphosphatidylinositol-linked cell wall protein Utr2/Crh2p as a novel protein quality control mechanism in Saccharomyces cerevisiae.* Eukaryot Cell, 2010. **9**(11): p. 1669-79.
207. Ye, Z.H. and D.L. Droste, *Isolation and characterization of cDNAs encoding xylogenesis-associated and wounding-induced ribonucleases in Zinnia elegans.* Plant Mol Biol, 1996. **30**(4): p. 697-709.
208. Schneider, R., et al., *Identification of a structural glycoprotein of an RNA virus as a ribonuclease.* Science, 1993. **261**(5125): p. 1169-71.
209. Turkel, S., T. Turgut, and N. Kayakent, *Effect of osmotic stress on the derepression of invertase synthesis in nonconventional yeasts.* Lett Appl Microbiol, 2006. **42**(1): p. 78-82.
210. Herrera, C.M., et al., *Inhospitable sweetness: nectar filtering of pollinator-borne inocula leads to impoverished, phylogenetically clustered yeast communities.* Proceedings of the Royal Society B-Biological Sciences, 2010. **277**(1682): p. 747-754.
211. Kalebina, T.S., et al., *Deletion of BGL2 results in an increased chitin level in the cell wall of Saccharomyces cerevisiae.* Antonie Van Leeuwenhoek, 2003. **84**(3): p. 179-84.
212. Kuranda, M.J. and P.W. Robbins, *Chitinase is required for cell separation during growth of Saccharomyces cerevisiae.* J Biol Chem, 1991. **266**(29): p. 19758-67.
213. Otto, R.T., et al., *Production of sophorolipids from whey. II. Product composition, surface active properties, cytotoxicity and stability against hydrolases by enzymatic treatment.* Appl Microbiol Biotechnol, 1999. **52**(4): p. 495-501.
214. Gargouri, M., P. Drouet, and M.D. Legoy, *Synthesis of a novel macrolactone by lipase-catalyzed intra-esterification of hydroxy-fatty acid in organic media.* J Biotechnol, 2002. **92**(3): p. 259-66.
215. Brenneis, R.B., B.; Kley, G., *Alcoholysis of waste fats with 2-ethyl-1-hexanol using Candida antarctica lipase A in large-scale tests.* European Journal of Lipid Science and Technology, 2004. **106**(12): p. 809-814.
216. Brenneis, R. and B. Baeck, *Esterification of fatty acids using Candida antarctica lipase A in water-abundant systems.* Biotechnol Lett, 2012. **34**(8): p. 1459-63.
217. Saerens, K., et al., *Production of glucolipids and specialty fatty acids from sophorolipids by Penicillium decumbens naringinase: Optimization and kinetics.* Biotechnol J, 2009. **4**(4): p. 517-24.
218. Kakugawa, K., et al., *Cloning, characterization, and expression of cDNA encoding a lipase from Kurtzmanomyces sp. I-11.* Biosci Biotechnol Biochem, 2002. **66**(6): p. 1328-36.
219. Brundiek, H., et al., *The short form of the recombinant CAL-A-type lipase UM03410 from the smut fungus Ustilago maydis exhibits an inherent trans-fatty acid selectivity.* Appl Microbiol Biotechnol, 2012. **94**(1): p. 141-50.

220. Ericsson, D.J., et al., *X-ray structure of Candida antarctica lipase A shows a novel lid structure and a likely mode of interfacial activation.* J Mol Biol, 2008. **376**(1): p. 109-19.
221. Martinelle, M., M. Holmquist, and K. Hult, *On the interfacial activation of Candida antarctica lipase A and B as compared with Humicola lanuginosa lipase.* Biochim Biophys Acta, 1995. **1258**(3): p. 272-6.
222. Roelants, S., *Starmerella bombicola as a platform organism for the production of biobased compounds.* PhD-thesis, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium. , 2013.
223. Eom, G.T., et al., *High-level extracellular production and characterization of Candida antarctica lipase B in Pichia pastoris.* J Biosci Bioeng, 2013. **116**(2): p. 165-70.
224. Minning, S., C. Schmidt-Dannert, and R.D. Schmid, *Functional expression of Rhizopus oryzae lipase in Pichia pastoris: high-level production and some properties.* Journal of Biotechnology, 1998. **66**(2-3): p. 147-56.
225. Li, P., et al., *Expression of recombinant proteins in Pichia pastoris.* Appl Biochem Biotechnol, 2007. **142**(2): p. 105-24.
226. Ghosh, P.K., et al., *Microbial lipases: production and applications.* Sci Prog, 1996. **79 (Pt 2)**: p. 119-57.
227. Domínguez de María, P.C.-O., C.; Tuin, B.; Bargeman,G.; van der Meer, A.; van Gemert, R., *Biotechnological applications of Candida antarctica lipase A: State-of-the-art.* Journal of Molecular Catalysis B: Enzymatic, 2005. **37**(1-6): p. 35-46.
228. Larsen, M.W., et al., *Suppression of water as a nucleophile in Candida antarctica lipase B catalysis.* Chembiochem, 2010. **11**(6): p. 796-801.
229. Delorme, V., et al., *Effects of surfactants on lipase structure, activity, and inhibition.* Pharm Res, 2011. **28**(8): p. 1831-42.
230. Otzen, D., *Protein-surfactant interactions: a tale of many states.* Biochim Biophys Acta, 2011. **1814**(5): p. 562-91.
231. Aschenbrenner, E.M., C.K. Weiss, and K. Landfester, *Enzymatic esterification in aqueous miniemulsions.* Chemistry, 2009. **15**(10): p. 2434-44.
232. Canto, A., et al., *Pollinator foraging modifies nectar sugar composition in Helleborus foetidus (Ranunculaceae):An experimental test.* Am J Bot, 2008. **95**(3): p. 315-20.

Curriculum vitae

CURRICULUM VITAE

PERSONAL INFORMATION

Name:	Katarzyna Ciesielska		
Nationality:	Polish	**Languages:**	Polish (native)
Phone:	+32472735613		English (fluent)
Email:	katarzyna.ciesielska84@gmail.com		Dutch (intermediate)

EDUCATION

Nov 2008 – Sep 2013	PhD, Ghent University, Faculty of Sciences
	Unit for Mass spectrometry & Proteomics,
	Laboratory for Protein Biochemistry and Biomolecular Engineering
	PhD thesis: "Proteomic study of the sophorolipid producer *Stramerella bombicola"*
	Promotor: Prof. Bart Devreese
Oct 2003 – Jul 2008	M.Sc., Technical University of Lodz, Faculty of Biotechnology and Food Science
	Institute of Technical Biochemistry (Poland)
Feb 2008 - Jul 2008	Erasmus program at Ghent University (Belgium)
	Master thesis: "Adenylate kinase from *Plasmodium falciparum*"
	Promotor: Prof. Savvas Savvides
	Guiding scientist: Dr. Bjorn Vergauwen
Sep 1999 - Jul 2003	Mathematics – IT class
	I High School of J. Chelmonski in Łowicz

COURSES AND WORKSHOPS

- Introduction to R (UGent and CvS) (Ghent 2013)
- Ghent Bio-Economy summer school (Ghent 2012)
- Advanced academic English (UGent): Writing skills (2011) and conference skills (2012)
- Quantitative proteomics and data analysis workshop (London 2011)
- Mascot training course (London 2010)
- 4th EU Summer School in Proteomic Basics (Brixen 2010)
- Dutch language course (Ghent 2010-2012)

PUBLICATIONS

Ciesielska, K., Li, B., Groeneboer, S., Van Bogaert, I., Lin, Y.-C., Soetaert, W., Van de Peer, Y., Devreese, B. 2013. SILAC-based proteome analysis of *Starmerella bombicola* sophorolipid production. Journal of proteome research; Aug. 2013

In progress:

Ciesielska, K. and Van Bogaert, I., Chevineau, S., Li, B., Groeneboer, S., Soetaert, W., Van de Peer, Y., Devreese, B. Exoproteome analysis of *Starmerella bombicola* results in the discovery of an esterase required for lactonization of sophorolipids.

Ciesielska, K., Van Bogaert, I., Roelands, S., Groeneboer, S., Vandenberghe, I., Soetaert, W., Devreese, B. Lactone esterase of *Starmerella bombicola:* expression, purification and enzymatic properties.

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

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

Ciesielska, K. and Roelants, S., Verweire, S., Groeneboer, S., Verweire, S., De Maeseneire, S., Van Bogaert, I., Devreese, B. and Soetaert, W. Combined techniques for the determination of protagonists involved in sophorolipid catabolism by its producer *Starmerella bombicola*.

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

PATENTS

Van Bogaert, I, **Ciesielska, K.,** Roelants, S., Soetaert, W., Devreese, B. A lactonase derived from *Candida bombicola* and uses thereof." UGent. PCT/EP2012/075571. (July 2013)

TECHNICAL SKILLS

- SILAC-stable isotope labeling of cells (quantitative proteomics)
- iTRAQ- chemical labeling (quantitative proteomics)
- Protein extraction
- Protein and peptide separation using HPLC
- Mass spectrometry analysis with LT-FTICR and MALDI
- Proteomic data analysis (Mascot, MaxQuant)
- Comparison of genomic and proteomic data
- Yeast exoproteome analysis
- Molecular biology (cloning, qPCR, SDS-PAGE, Western blot analysis)
- Protein expression (mainly in *Pichia pastoris*) and purification (IMAC, SEC)
- Designing and carrying out enzymatic assays (lipase)