Development of tools to unravel key metabolic features of *Pseudozyma spp*. in the production of mannosylerythritol lipids from D-glucose and D-xylose

Ana Catarina Amiguinho Reto

Thesis to obtain the Master of Science Degree in

Biotechnology

Supervisors: Dr. César Simões da Fonseca

Dr. Frederico Castelo Alves Ferreira

Examination Committee:

Chairperson: Prof. Leonilde de Fátima Morais Moreira

Supervisor: Dr. César Simões da Fonseca

Members of the Committee: Prof. Nuno Gonçalo Pereira Mira

November 2015

Acknowledgments

I would like to thank to my advisor, Dr. César Fonseca, for the support during my research work, for his motivation, enthusiasm, and knowledge and for the opportunity to collaborate at LNEG. His assistance helped me to perform my lab work and writing of this thesis. I would like to thank my co-advisor Prof. Frederico Ferreira for his guidance and his full-time availability.

Thank you to Dr. Nuno Faria for his guidance and for gave me access to the laboratory and research facilities at LNEG.

I would like to thank all of the IBB and LNEG members for their help and support in, especially to Marisa Santos, Maria José Leandro and Catarina Figueiredo.

I also would like to acknowledge my friends and family for their support and encouragement.

I also would like to thank for the funding which allowed the development of the work. This work was co--funded by Fundação para Ciência e Tecnologia by National with the Exploratory Project "TAKE--OFF" through FEDER funds “Programa Operacional Factores de Competitividade” – COMPETE (FCOMP--01--0124--FEDER--041191) and national (FCT) funds (EXPL/AAG--TEC/0610/2013).

Abstract

The work performed in this thesis intended to develop the necessary tools to understand key metabolic features of *Pseudozyma spp*. in the production of mannosylerythritol lipids from D-glucose and D-xylose, and get the first insights in the respective biochemical and molecular assets. Firstly, it was confirmed that the carbon sources differently affects MEL production by *Pseudozyma spp.. P.aphidis* produced higher MEL titres from D-glucose, while *P.antarctica* produced similar MEL titres from D-glucose and D-xylose. Comparing the two strains, *P. antarctica* was the best MEL producer in the conditions tested. Along with the carbon source, the influence of the nitrogen source in MEL production was also assessed, but ammonia was prejudicial for MEL production. The activity of the enzymes involved in metabolic pathways potentially contributing for MEL production, through the consumption and production of NADPH, were implemented and evaluated. The oxidative PPP is, apparently, the major source of NADPH in the production of MEL both from D-glucose and D-xylose. However, the induction of this pathway by D-xylose, in response to the higher demand for NADPH, is much more effective in *P. antarctica*, justifying the relative high MEL titres in D-xylose medium. From the analysis of the *Pseudozyma spp* genomes, the genes related with MEL production (assembly of building blocks and export) and with the consumption and the production of NADPH were identified and gene-specific primer pairs were designed to be used in Real-Time RT-PCR. In accordance with enzyme activities, in *P. antarctica* the transcript levels of the genes involved in NADPH production are significant higher in D-xylose than in D-glucose. Moreover, *P.antarctica* EMT1 (assembling mannosylerythritol) is also highly overexpressed in D-xylose, which supports the relative high MEL titres in D-xylose medium even if it requires carbon recycling through oxidative PPP.

Resumo

O trabalho desenvolvido nesta tese teve como objetivo desenvolver ferramentas necessárias para compreender as características metabólicas relevantes de *Pseudozyma spp.* na produção de manosileritritolípidos (MEL), e obter as primeiras informações relevantes do ponto de vista bioquímico e molecular. Primeiro foi confirmado que a fonte de carbono influencia de forma diferente a produção MEL por *Pseudozyma spp.*. *P. aphidis* produziu maiores concentrações de MEL em D-glucose, enquanto *P. antarctica* produziu MEL em concentrações semelhantes a partir de D -glucose e D-xilose. Comparando as duas estirpes, *P. antarctica* foi a melhor produtora de MEL nas condições testadas. Juntamente com a fonte de carbono, também foi avaliada a influência da fonte de azoto na produção de MEL, a amónia foi prejudicial para a produção MEL. Foram implementadas e avaliadas as atividades das enzimas envolvidas em vias metabólicas que contribuem para a produção de MEL, através da produção e consumo de NADPH. A parte oxidativa da via das pentoses fosfato é, aparentemente, a maior fonte de NADPH, tanto em D-glucose como em D-xilose. No entanto, a indução desta via por D-xilose, em resposta à maior necessidade de NADPH, é muito mais efetiva em *P. antarctica*. Da análise do genoma de *Pseudozyma spp,* foram identificados genes envolvidos na produção de MEL e no consumo e produção de NADPH e desenhados pares de oligonucleótidos iniciadores para uso em RT-PCR em tempo real. Tal como verificado com as atividades enzimáticas, em *P. antarctica*, os níveis de transcrição dos genes envolvidos na produção de NADPH são significativamente superiores em D-xilose do que em D-glucose. Também, o EMT1 de *P. antarctica* (síntese de manosileritritol) é significativamente sobre-expresso em D-xilose, o que justifica as relativas elevadas concentrações de MEL em D-xilose, mesmo requerendo reciclagem de carbono através da via oxidativa das pentoses fosfato.

Contents

Figure Index

Table Index

Abbreviations

ACC	Acetyl-CoA carboxylase
ACP	Acyl carrier protein
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
CDW	Cell dry weight
FAS	Fatty acid synthase
$FADH_2$	Flavin adenine dinucleotide
FID	Flame ionization detector
F6P	Fructose-6-phosphate
GAP	Glyceraldehyde- 3-phosphate
GC	Gas Chromatography
G6P	Glucose 6- Phosphate
HPLC	High-performance liquid chromatography
K_{cat}	Turnover number
Km	Michaelis-Menten constant
LC	Liquid Chromatography
NAD^+	Nicotinamide adenine dinucleotide (oxidized form)
NADH	Nicotinamide adenine dinucleotide (reduced form)
$NADP^+$	Nicotinamide adenine dinucleotide phosphate (oxidized form)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
MDH	Malate dehydrogenase
MEL	Mannosylerythritol lipid
NMR	Nuclear magnetic resonance
PCR	Polymerase chain reaction

PPP Pentose Phosphate Pathway

RT-PCR Reverse transcription polymerase chain reaction

R5P Ribose-5-phosphate

Ru5P Ribulose-5-phosphate

TCA Tricarboxylic acid cycle

$V_{máx}$ Maximal Velocity

Xu5P Xylulose-5-phosphate

1 Introduction

1.1 Motivation

In the twentieth century major part of research, namely chemical and chemical engineering, emphasis was given for the development of petroleum-, coal-, and natural gas-based refinery to exploit the cheaply available fossil feedstock. These feedstocks are used in a variety of industries, for example, fuel, fine chemicals and pharmaceuticals.[1] Currently, the fossil resources are not regarded as sustainable from the economic, social and environmental point of views. In order to not compromising future generations it is necessary to invest in a sustainable development. Nowadays a new concept is emerging: the bioeconomy, that pursues the vision of producing chemicals and fuel from renewable feedstock.[2] The concept of the bioeconomy initially started from the life sciences and biotechnology fields, which has then been extended to incorporate other ideas such as the biorefinery concept[3]. Biorefinery is a concept that supports the conversion of plant-based biomass to chemicals, energy and materials that run our civilization, replacing the needs of petroleum, coal, natural gas, and other non-renewable energy and chemical sources.[4] Biomass is renewable, in that the plant synthesizes chemicals by drawing energy from the sun, carbon dioxide and water from the environment, while releasing oxygen. Fractionation and conversion of the lignocellulosic biomass can produce a vast array of chemicals, energy and materials. Developing capabilities to converting variety of plant-based biomass to chemicals, energy and materials is the key to move the world economy back to be based on the sustainably renewable biomass[4].

The main plants components are carbohydrates (simple sugars, cellulose, hemicelluloses, starch, etc.) and lignin. The polysaccharides cellulose and hemicellulose, together with lignin, constitute the lignocellulose. It is organized by rigid cellulose fibers bound to lignin and hemicellulose. By a hydrolysis process they can give rise to fermentable monomer sugars, as D-glucose and D-xylose, while lignin is very resistant to break down and usually is not a substrate for fermentation processes. Despite lignin cannot be hydrolyzed into fermentable sugars, it can be integrated in biorefineries, for instance, for energy production by combined heat and power, increasing the overall lignocellulosic material valorization.[5] The utilization of cellulose and hemicellulose fractions of lignocellulosic materials allows the production of fuels (e.g. ethanol)[6,7] and chemicals (e.g. lactic acid) through biochemical conversion processes[3]. The biological conversion of pentoses generated from hemicellulose into value-added compounds has been the target of research over the past years[6]. The conversion of biomass into fuels or chemicals is a process that can be improved both through metabolic and bioprocess engineering, and preferentially through their integration. Systems biology approaches are beneficial to improve the knowledge of specific biochemical pathways, which can be used to increase the production of specific bio-based products.[5]

In this work, the production of mannosylerythritol lipids (MEL) by *Pseudozyma* yeast is studied through physiological, biochemical and molecular approaches. MEL is a glycolipid with surfactant properties and can be applied in a variety of applications due to its versatile physico-chemical features.[5] Recently, it

was shown, in our time, that this molecule can be produced from lignocellulosic sugars, besides the most common substrate, soybean oil.[6] The knowledge of biochemical routes for sugar (pentoses and hexoses) utilization in *Pseudozyma* spp. and the respective metabolic and molecular features can give new insights for future improvement of MEL production and the cost-competitiveness of the biological process.

1.2 Objectives and Challenges

The aim of this study was to develop tools to understand key metabolic features of *Pseudozyma spp*. in the production of mannosylerythritol lipids from lignocellulosic sugars (D-glucose and D-xylose), by combining physiological, biochemical and molecular approaches. The genomes of *Pseudozyma spp*., like *P. aphidis* and *P. antarctica*, were recently released, thus allowing systems biology studies, namely through transcriptomics approaches[8], or more specific molecular biology studies (e.g. transcript analysis of specific pathways). As described in the literature, MEL production requires three main building blocks: mannose, erythritol and medium-chain fatty acids. To yield mannose and erythritol from lignocellulosic monosaccharides, D-glucose and D-xylose are metabolized through the metabolic pathways, glycolysis and PPP[5]. Fatty acid biosynthesis is required to generate the lipidic moiety of MEL and it is dependent of high NADPH availability. Thus, apart from the enzymes and respective coding genes for MEL biosynthesis, it is relevant to identify those related with NADPH production and consumption, in order to fully understand the main metabolic pathways involved in the conversion of D-glucose and D-xylose into MEL. Overall, to improve MEL production the requirements for lipid accumulation under excess of carbon source and nitrogen limitation need to be clarified[9].

1.3 Research questions

This thesis intends to address the following queries:

- Which carbon source (D-glucose and D-xylose) and nitrogen sources ($NaNO_3$ and NH_4Cl) provides higher MEL titres?
- What are the main differences in MEL production by *P. aphidis* or *P. antarctica* from different carbon and nitrogen sources?
- What are the important metabolic pathways and the requirements of *Pseudozyma spp.* to describe MEL production?
- Which tools can be used to study the metabolic pathways involved in MEL production?
- Which are the relevant genes for MEL production?
- What are the differences in the expression levels of relevant genes for MEL production between different conditions established and strains studied?

- Which are the key enzymes crucial to describe MEL production?
- Can a relation be established between enzyme activity, transcript analysis and bioconversion conditions for MEL production?
- Is the information obtained in the thesis useful to plan systems biology studies?

1.4 Research strategy

To achieve the objectives and address the research questions, the following approaches were followed:

- Microorganism: *Pseudozyma aphidis* PYCC 5535^T and *Pseudozyma antarctica* PYCC 5048^T strains were studied as they were identified by our time, as the best MEL producers from D-glucose and D-xylose[6].
- Carbon source: D-glucose and D-xylose were used as carbon sources, since they are the main monosaccharides obtained from lignocellulosic materials and were previously reported by our time as suitable carbon sources for MEL production[6].
- Nitrogen source: an alternative nitrogen source (NH_4Cl instead of $NaNO_3$) was tested in order to evaluate different NAPDH requirements for nitrogen assimilation.
- Bioinformatics: NCBI[10] and KEGG[11] databases allowed the identification of the main genes involved in MEL production, nitrogen metabolism and NADPH production and consumption in *Pseudozyma spp.* genomes. Whenever necessary, *Ustilago maydis* annotated genome was used as reference, since the *Pseudozyma spp.* are anamorphic basidiomycetous yeasts belonging to Ustilaginomycetes, which includes the smut fungus *U. maydis*[5]. *Ustilago maydis* is known to secrete large amounts of mannosylerythritol lipids under conditions of nitrogen starvation.[12]
- Enzymatic assays: Using crude extracts of *Pseudozyma spp.*, the activity of enzymes directly involved in NADPH production or consumption were assessed, namely those from the oxidative PPP, citrate shuttle and nitrate and xylose assimilation.
- Transcript analysis: The relevant cDNA sequences of *P. antarctica* and *P. aphidis*, those directly related to MEL biosynthesis (assembling of building blocks) and those related to NADPH production and consumption, were aligned using ClusterIW Multiple Alignment algorithm and searched for conserved regions. Those regions were used to design gene-specific primers (suitable for both *Pseudozyma* species) to be used in Real-Time RT-PCR experiments.

2 Literature review and state of the art

2.1 Bio-Based products and Bio economy

The building up of a bio-economy has been, since a while, an objective of the modern society, with an increased emphasis on the sustainable use of natural resources, competitiveness, socioeconomic and environmental issues. The concept of a bio-based economy integrates the full range of natural and renewable biological resources- land and sea resources, biodiversity and biological materials (plant, animal and microbial), through to the processing and the consumption of these bio-resources. The bio-economy encompasses the agriculture, forestry, fisheries, food and biotechnology sectors, ranging from the production of energy and chemicals to a wide range of industries and transportation.[13]

To building up a bio-economy, it is important to promote development and commercialization in products wholly or partly derived from biomass, such as plants, trees or animals (the biomass can have undergone physical, chemical or biological treatment). The bio-based products could provide additional product functionalities, less resource intensive production and efficient use of all natural resources.[14]

2.2 Lignocellulosic materials

Lignocellulosic feedstock can offer the potential to provide sources of organic carbon and offer interesting carbon sources for the sustainable production of bio-based chemicals. The lignocellulose is present in many waste streams from different industries and various other sources, including agricultural, forestry and municipal solid wastes.[5] Biomass derived from trees, agro-forest residues, grasses, plants, aquatic plants and crops are versatile and important renewable feedstock for chemical industry.[1]

Biomass is the largest renewable energy resource currently being used, comprising 47% of the total renewable energy consumption.[15]

The cellulose and hemicellulose, together with lignin, constitute the lignocellulose. The cellulose is a homopolymer of D- Glucose and hemicellulose is composed by xylans and arabinoxylans, where the D-Xylose and L-Arabinose are the most abundant sugars. The upgrading of both the cellulose and the hemicellulose fractions of lignocellulosic materials is essential for the production of cost-competitive bio-based products.[6]

2.3 Biosurfactants and Mannosylerythritol lipids

The biosurfactants has aroused great interest, since they have unique properties including structural diversity, higher biodegradability, lower toxicity, higher foaming ability, higher selectivity and specific activity at extreme conditions (pH, temperature and ionic strength) and mild production conditions when compared to the chemical surfactants.[16] Biosurfactants are amphiphilic molecules with polar and non-polar domains, and are produced by plants, animals as well microorganisms. This compounds display important biological activities, including antibiotic, antifungal, antiviral and antitumoral activities.

However, these compounds have another type of applications, for example, in food, cosmetics, and pharmaceutical industries. They are also used in removing soil contaminants including heavy metals, oils and other toxic pollutants. [16]

Biosurfactant containing 4-O-β-D-mannopyranosylmeso-erythritol as the hydrophilic group and a fatty acid and/or an acetyl group as the hydrophobic moiety is known as mannosylerythritol lipid (MEL). MEL is reported to be secreted by *Ustilago sp*. (as a minor component), *Schizonella melanogramma* (shizonellin) and *Pseudozyma spp*. (as a major component). The synthesis of MEL is reported as being not growth-associated and MEL can also be produced using resting (stationary phase) yeast cells.[16]. It has been suggested that MEL acts as an energy storage material in the yeast cells similar to triacylglycerols and it has been shown to reduce the surface tension of water to less than 30 mN m^{-1}.[16]

Soybean oil has been described as the substrate leading to higher MEL titres and productivities when used in *Pseudozyma* spp. cultures. However, due to soybean oil-increasing prices, cultivation methodologies and the inefficient MEL recovery from oil-containing cultures, soybean oil will hardly be sustainable. [17].

Recent studies from our group show the ability of the *Pseudozyma* strains to use pentoses and D glucose/ D -xylose mixtures, as alternative substrates towards the sustainable production of MEL. Lignocellulosic biomass may arise as an alternative raw material to the production of MEL since D -glucose and D -xylose are the main sugars found in lignocellulosic biomass.[6]

2.4 Overview of central carbon metabolism in the yeast

Sugar metabolism in yeasts is performed through their conversion into pyruvate via glucose-6-phosphate and/or fructose-6-phosphate and fructose-1,6-bisphosphate, with the generation of 2 mol of ATP per mol of sugar.[18] In the first stage, glucose is phosphorylated and cleaved to yield two triose molecules,

glyceraldehyde-3-phosphate, with the use two ATP molecules as a kind of energy investment.[19] In the second stage, the two molecules of glyceraldehyde-3-phosphate are converted into pyruvate, with concomitant generation of four ATP molecules. Therefore, glycolysis has a net profit of two ATPs per molecule of sugar.[20]

One of the important points in this pathway is that NAD^+ is the primary oxidizing agent of glycolysis. The NADH produced by this process (Eq. 1) must be continually reoxidized to keep the pathway supplied with NAD^+. In the case of the *Saccharomyces cerevisiae*, pyruvate is decarboxylated to yield CO_2 and acetaldehyde and the latter is reduced by NADH to yield NAD^+ and ethanol. However, in yeasts belonging to *Pseudozyma* genus, which are consider non-fermentative and obligate aerobes, the mitochondrial oxidation of each NADH to NAD^+ under aerobic conditions yields 2.5 ATP molecules.[20]

The glycolysis is conserved among yeasts however the metabolic destiny of pyruvate is different depending on the yeast species.[18]

$$Glucose + 2NAD^+ + 2ADP + 2P_i \rightarrow 2pyruvate + 2NADH + 2H^+ + 2ATP + 2H_2O \qquad Eq.1$$

In contrast to glucose, the conversion of pentoses, like xylose, into a central intermediate, like triose phosphate, requires other metabolic pathway with great importance for yeast, the Pentose Phosphate Pathway (PPP) (Figure 2-1). The catabolism of D-xylose requires its conversion into a PPP intermediate, D-xylulose-5-phosphate, through redox reactions using NAD(P)(H) as cofactors.[18] Usually, the use of D-xylose as carbon source requires NADPH, resulting from the initial NADPH-linked reduction of xylose to xylitol (Eq. 2)[16], through an inducible xylose reductase (EC 1.1.1.307).[21]

$$D - Xylose + NAD(P)H + H^+ \rightarrow Xylitol + NAD(P)^+ \qquad Eq.2$$

Then xylitol is oxidized to D-xylulose by an NAD^+-dependent xylitol dehydrogenase (EC 1.1.1.113) (Eq. 3).

$$Xylitol + NAD(P)^+ \rightarrow D - Xylulose + NAD(P)H + H^+ \qquad Eq.3$$

D-Xylulose is then phosphorylated by xylulose kinase (EC 2.7.1.17) and enters the PPP.[18] The partition of sugar metabolism between the glycolysis and the PPP occurs at the level of glucose-6-phosphate, with glucose-6-phosphate dehydrogenase (EC: 1.1.1.49) (Eq. 4), which directs the carbon flux to the PPP by catalyzing the oxidation of glucose-6-phosphate into 6-phosphoglucono-1,5-lactone, with the generation of NADPH. The lactone is hydrolyzed into the free acid 6-phosphogluconate by a specific lactonase, which undergoes oxidation and decarboxylation, by 6-phosphogluconate dehydrogenase (EC. 1.1.1.44), to form the keto pentose ribulose 5-phosphate. This reaction generate a second molecule of NADPH[19]. (Eq. 5).

The oxidative part of PPP is a major source of NADPH, which is essential, for instance, for biosynthetic processes such as fatty-acid synthesis and the assimilation of inorganic nitrogen in yeast.[22]

$$D-glucose\ 6-phosphate\ +\ NADP^{+} \rightarrow\ 6-phospho-D-glucono-1{,}5-lactone\ +\ NADPH + H^{+}\ \ Eq.4$$

$$6-phospho-D-gluconate\ +\ NADP^{+} \rightarrow D-ribulose\ 5-phosphate + CO_2 +\ NADPH\ + H^{+}\ \ Eq.5$$

In the non-oxidative PPP, ribulose-5-phosphate is first epimerized into xylulose-5-phosphate and then, in a series of rearrangements of the carbon skeletons, six five-carbon sugar phosphates are converted into the equivalent to five six-carbon sugar phosphates, completing the cycle and allowing continued oxidation of glucose-6-phosphate with the production of NADPH. Continued recycling leads ultimately to the conversion of glucose-6-phosphate to six CO_2.[20] Each molecule of ribulose-5-phosphate not required for nucleotide biosynthesis, together with two molecules of xylulose-5-phosphate, are converted into two molecules of fructose-6-phosphate and one molecule of glyceraldehyde-3-phosphate via the sequential action of transketolase (EC 2.2.1.1), transaldolase (EC 2.2.1.2), and, again, transketolase.[20] The main products of the PPP are ribulose-5-phosphate and NADPH, while fructose-6-phosphate and glyceraldehyde-3-phosphate may be metabolized through glycolysis, recycled to PPP, or used in other pathways (e.g. gluconeogenesis). If NADPH is in excess, the PPP may be reversed to synthesize ribulose-5-phosphate from the glycolytic intermediates.

The flux through the oxidative PPP and thus the rate of NADPH production is controlled by the rate of the glucose-6-phosphate dehydrogenase reaction, this enzyme is regulated by the $NADP^+$ concentration. When the cell consumes NADPH, the $NADP^+$ concentration rises, increasing the rate of the glucose-6-phosphate dehydrogenase reaction, thereby stimulating NADPH regeneration.[20]

The PPP can play an important role in the production of MEL from sugars, because this pathway allows the direct assimilation of D-xylose and the production of NADPH, necessary for the fatty acid biosynthesis.

It has been observed, in yeasts that the nitrogen source of the medium influences the amount of sugar directed to the PPP.[18] Using nitrate as nitrogen source, an increase of the carbon flux through the PPP shall be expected due to the increased NADPH requirement caused by the operation of nitrate and nitrite reductase (EC: 1.7.1.2 , EC: 1.7.1.2)[18] (Eq. 6 and 7)

$$NO_3^- + NAD(P)H + H^+ \rightarrow NO_2^- + H_2O + NAD(P)^+ \qquad Eq.6$$

$$NO_2^- + 3NADPH + 3H^+ \rightarrow NH_4^+ + 2H_2O + 3NADP^+ \qquad Eq.7$$

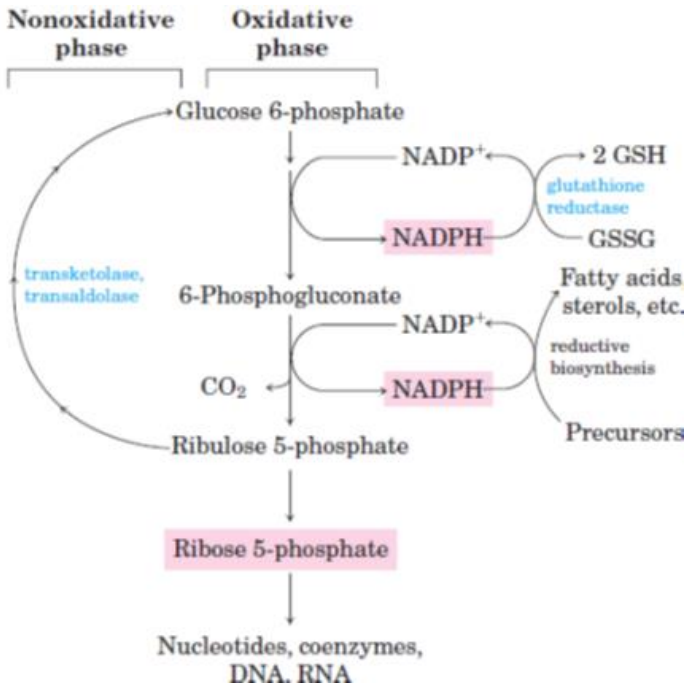


Figure 2-1:General Scheme of Pentose Phosphate Pathway[19]

In the context of this work it is important to address tricarboxylic acid (TCA) cycle (Figure 2-2), the common mode of oxidative degradation in eukaryotes and prokaryotes, because *Pseudozyma spp.* are

non-fermentative yeasts and thus pyruvate is more likely oxidized to CO_2 through the tricarboxylic acid (TCA) cycle. In this regard, glucose is converted into pyruvate in glycolysis, which can enter the mitochondria for respiration.[23] Pyruvate is, than, converted to acetyl-CoA by pyruvate dehydrogenase complex (PDH) and acetyl-CoA is further oxidized by TCA[23]. TCA is a series of reactions that oxidizes the acetyl group of acetyl-CoA to two molecules of CO_2 in a manner that conserves the liberated free energy for utilization in ATP generation. In this cycle 1 acetyl-CoA is converted in to 2 CO_2 molecules, yielding 3 NADH, 1 $FADH_2$, and 1 GTP (or ATP) molecules. The NADH and $FADH_2$ are oxidized by O_2 in the electron-transport chain with the concomitant synthesis of around 11 more ATPs, yielding a total of about 12 ATPs for one turn of the citric acid cycle.[20]

In this cycle are present several enzymes, like citrate synthase (EC 2.3.3.1), which catalyzes the condensation of acetyl-CoA and oxaloacetate to yield citrate. Another important enzyme in this cycle is Isocitrate dehydrogenase (EC: 1.1.1.42) (Eq. 8), which oxidizes isocitrate to the α-keto acid intermediate oxalosuccinate with the coupled reduction of $NADP^+$ into NADPH. Oxalosuccinate is then decarboxylated, yielding α-ketoglutarate. This is the first step in which oxidation is coupled to NADPH production and also the first CO_2-generating step.

$$Isocitrate + NADP^+ \rightarrow 2-oxoglutarate + NADPH + CO_2 \qquad Eq.8$$

The reaction involves the reduction of a second NAD^+ into NADH and the generation of a second molecule of CO_2. At this point in the cycle, two molecules of CO_2 have been produced, so that the net oxidation of the acetyl group is complete. Note, however, that it is not the carbon atoms of the entering acetyl-CoA that have been oxidized. The last enzyme in this cycle is malate dehydrogenase that reforms oxaloacetate by oxidizing malate's secondary alcohol group to the corresponding ketone with concomitant reduction of a third NAD^+ to NADH. Acetyl groups are thereby completely oxidized to CO_2.[20]

In TCA the flux is controlled by citrate synthase, isocitrate dehydrogenase, and α-ketoglutarate dehydrogenase. Their activities are controlled by substrate availability, product inhibition, inhibition by cycle intermediates, and activation by Ca^{2+}.[20]

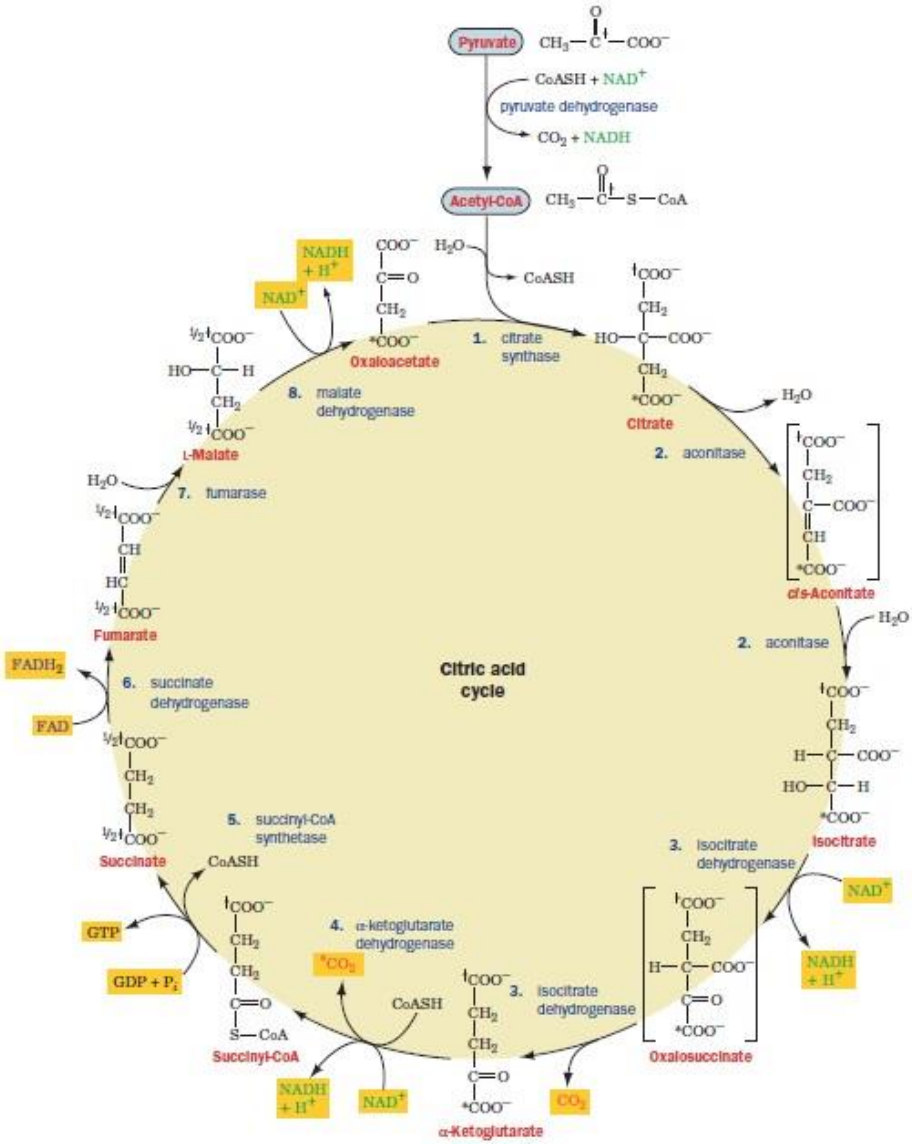


Figure 2-2: Reaction of citric acid cycle (Adapted)[20]

2.5 The fatty acid biosynthesis in yeast

The biosynthesis of fatty acids (Figure 2-3) occurs from acetyl-CoA by a cytosolic complex of six enzyme activities plus acyl carrier protein (ACP). This pathway requires the participation of a three-carbon intermediate, malonyl-CoA. One of the most important steps is the formation of malonyl-CoA from acetyl-CoA, which is an irreversible process catalyzed by acetyl-CoA carboxylase.[19]

Fatty acids are highly reduced materials and their synthesis requires the supply of NADPH to reduce acyl-CoA molecules. The continuous supply of acetyl-CoA directly in the cytosol serves an important role in ensuring provision of cytosolic acetyl-CoA for the production of fatty acids[23]. The acetyl-CoA is the

precursor for fatty acid synthase (FAS). Without availability of these two species (NADPH and Acetyl-CoA), the yeast cannot produce fatty acids.[9] In this biosynthetic pathway, the condensation reaction (condenses acyl and malonyl groups) is coupled to the hydrolysis of ATP, thereby driving the reaction to completion. This process involves two steps, one of them is ATP dependent: the carboxylation of acetyl-CoA (derived from pyruvate) by acetyl-CoA carboxylase (ACC) to form malonyl-CoA, and the second step is the exergonic decarboxylation of the malonyl group catalyzed by FAS.[20]

The synthesis of fatty acids from acetyl-CoA and malonyl-CoA involves seven enzymatic reactions that yield mainly palmitic acid. In each passage through the cycle, the fatty acyl chain is extended by two carbons and when the chain length reaches a 16-carbon product, palmitic acid, the most abundant fatty acid in eukaryotes, leaves the cycle. Chains with less of 16 carbons are generally not found. Fatty acid biosynthesis in almost all organisms culminates in the formation of either C_{16} or C_{18} saturated fatty acids[9].

The synthesis of 1 mol of C_{18} fatty acid requires 16 mols of NADPH. Two mols of NADPH are needed to reduce each 3-keto-fatty acyl group arising after every condensation reaction of acetyl-CoA with malonyl-CoA as part of the standard FAS complex into the saturated fatty acyl chain, which then undergoes a further cycle of chain lengthening.[9]

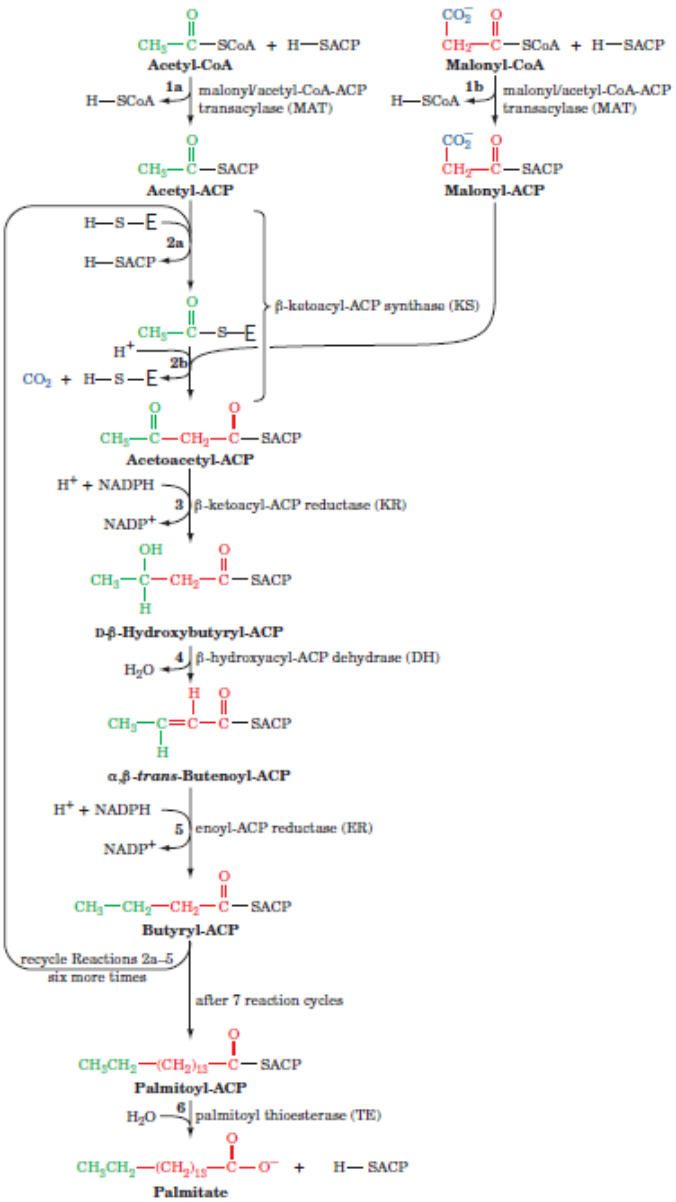


Figure 2-3: Reaction cycle for the biosynthesis of fatty acids. The biosynthesis of palmitate requires seven cycles of C2 elongation followed by a final hydrolysis step. (Adapted)[20]

Some yeasts are named, oleaginous species, because they can accumulate intracellularly relative high contents of oils, however in order to achieve lipid accumulation within the cell, they usually need to be grown in a medium with an excess of carbon substrate and a limiting amount of nitrogen.[16]

The process of fatty acid biosynthesis is crucial for accumulation of large quantities of oil. As described above, is necessary a continuous supply of acetyl-CoA directly in the cytosol, and in oleaginous yeasts, the formation of acetyl-CoA has been attributed to the presence of ATP: citrate lyase (Eq.10), which does not seems to occur in the majority of non-oleaginous species.[9] Studies prove that ACL gene was present in *P. antarctica* [24], so this species can be classified as oleaginous yeast.

$$Citrate + CoA + ATP \rightarrow acetly - CoA + oxaloacetate + ADP + P_I \qquad Eq.10$$

In oleaginous microorganisms, the citric acid accumulation is achieved by the activity of isocitrate dehydrogenase, which, as a component of the TCA cycle, is dependent on the presence of AMP. Nitrogen limitation in the cultivation of an oleaginous microorganism induces a cascade of reactions leading to the formation of acetyl-CoA.[9] In nitrogen-limiting conditions, there is an increase of AMP deaminase activity, which decreases the cellular content of AMP including its content in the mitochondria; the lower content of AMP in the mitochondria stops the isocitrate dehydrogenase activity. As a consequence isocitrate cannot be metabolized and thus it accumulates and is then readily equilibrated with citric acid. Therefore, citrate also accumulates in the mitochondria and it is efficiently exported through a citrate efflux system of the mitochondrial membrane. Citrate enters the cytosol and it is cleaved by ACL to give acetyl-CoA and oxaloacetate. While Acetyl-coA is directed to the fatty acid, oxaloacetate is converted back to pyruvate to re-entry into the mitochondria. This can occurs in two steps: oxaloacetate reacts with NADH to produce malate, catalyzed by cytosolic malate dehydrogenase (mdh) and then malate reacts with $NADP^+$ to produce pyruvate and NADPH, catalyzed by malic enzyme.[9] This pathway is named citrate shuttle (Figure 2-4) and plays an important role, not only on cytosolic acetyl-CoA production, but also in the proton regulation and NADPH generation.[5,9]

When considering synthesis of palmitate, one supplier of NADPH for fatty acid biosynthesis is considered to be malic enzyme[9] (Eq.9).

$$Malate + NADP^+ \rightarrow pyruvate + CO_2 + NADPH \qquad Eq.9$$

Malic enzyme (ME; $NADP^+$-dependent; EC 1.1.40) provides NADPH for lipid biosynthesis in oleaginous microorganisms. [25] Malic enzyme activity has been found in most oleaginous microorganisms.[9]

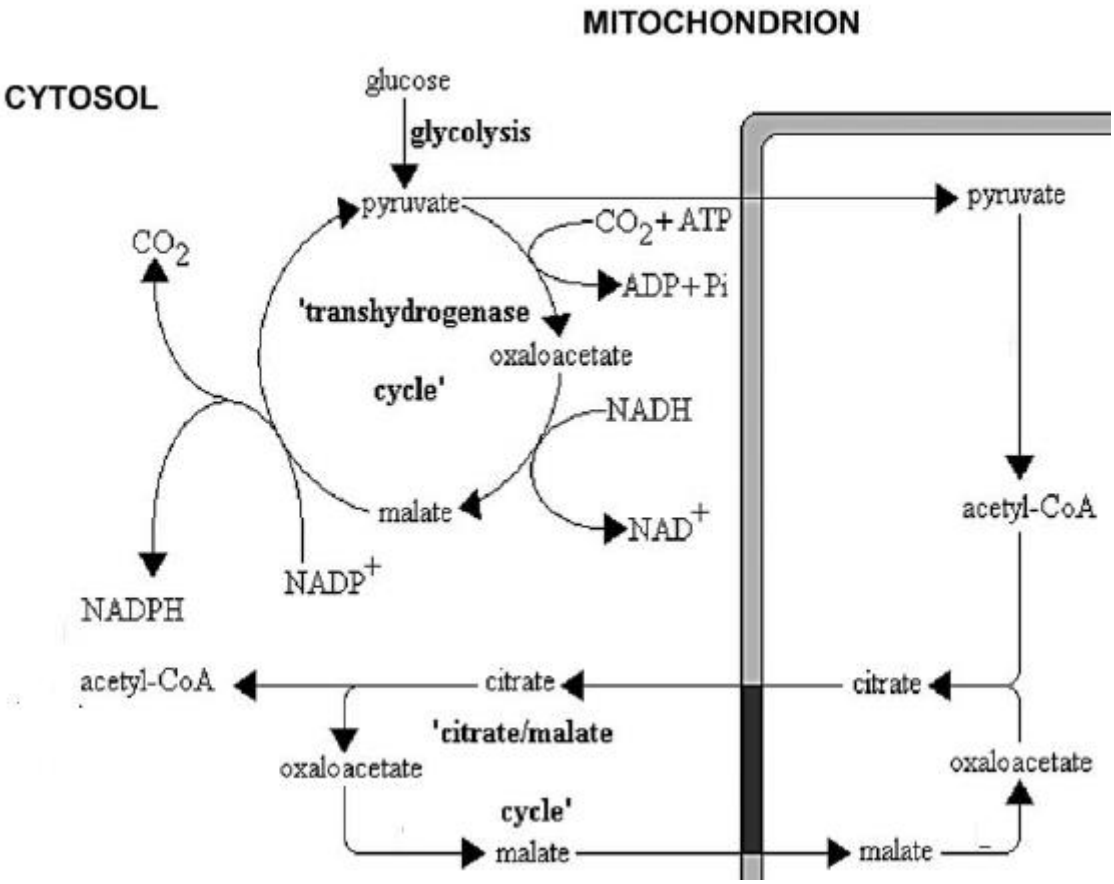


Figure 2-4: Citrate shuttle and "transhydrogenase" cycle (Adapted)[9]

2.6 Lipid β-oxidation in yeast

The fatty acid breakdown occurs, in peroxisome, by a mechanism known as β-oxidation in which the fatty acids C_β atom (the carbon next to α-carbon is β-carbon and is not directly attach to a functional group) is oxidized.[20] Before fatty acids can be oxidized, they must be "primed" for reaction in an ATP-dependent acylation reaction to form fatty acyl-CoA, this reaction is catalyzed by a family of at least 3 acyl-CoA synthases. Once fatty acids entered in β-oxidation pathway they are usually completed oxidized into an acetyl-CoA without accumulation or leakage of the intermediate species.[16] In each reaction step of the β-oxidation sequence, one molecule of acetyl-CoA, two pairs of electrons, and four protons (H^+) are removed from the long-chain fatty acyl–CoA, shortening it by two carbon atoms.

Supplementation of yeast growth medium with fatty acids as a sole carbon and energy source elicits the transcriptional up-regulation of genes encoding enzymes of the β-oxidation process. In *Saccharomyces cerevisiae,* this response is additionally accompanied by a remarkable proliferation of the peroxisomal compartment in which β-oxidation is exclusively housed.[26] In peroxisomes, the chain shortening of acyl-CoA esters between carbons 2 and 3 is catalyzed, yielding as products chain-shortened acyl-CoA and acetyl-CoA or propionyl-CoA, depending on substrates.[27] The degradation of straight-chain saturated fatty acids requires the participation of the four activities comprising the core of the β-oxidation cycle in peroxisomes, acyl-CoA oxidase, enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase, and 3-ketoacyl-CoA thiolase.[20] In peroxisomes, the energy released in the first oxidative step of fatty acid breakdown is not conserved as ATP, but is dissipated as heat.

In previous studies it was suggested that most of the fatty acids assembled in MEL are the result of β-oxidation intermediates, and that a new type of chain-shortening pathway (partial β-oxidation) (Figure 2-5) would contribute to the medium-chain lipids in the yeast, more specifically in the strain *P. antarctica*.[28] Therefore, it has been hypothesized, for *Pseudozyma spp*, that there are two distinct β-oxidation systems in the yeast, one for generating both cellular energy and acetyl-CoA and the other for medium-chain acids (C_8 to C_{12}) to synthesize MEL.[28]

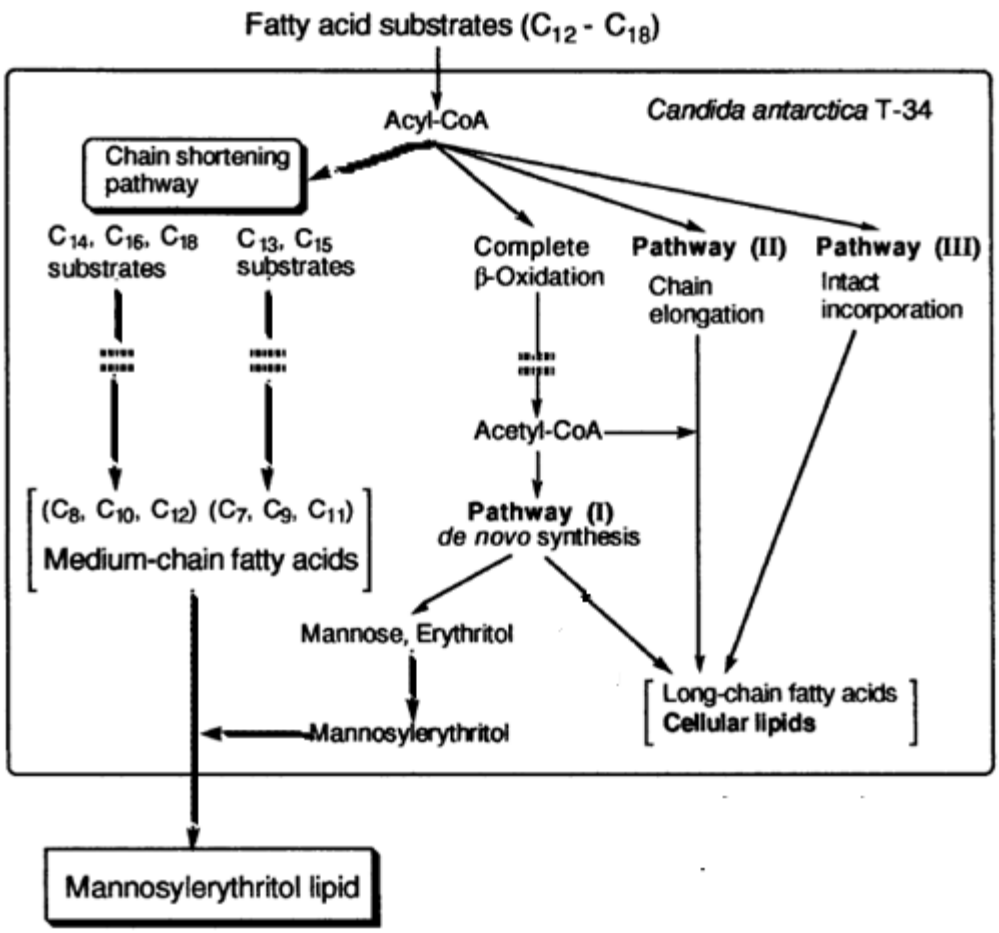


Figure 2-5: Presumptive biosynthetic pathway of MEL in Candida antarctica T-34 (Adapted) [28]

2.7 Specific pathways for MEL biosynthesis

Mannosylerythritol lipids (MELs) are glycolipids containing a 4-O-β-D-mannopyranosyl-meso-erythritol as the glycosidic/hydrophilic moiety and two short-chain fatty acids (usually C_8-C_{12}) as the hydrophobic groups.[6]

MEL is produced by the *Pseudozyma* genus, including the species *P. antarctica*, *P. rugulosa*, *P. aphidis* and *P. parantarctica*.[6] Also, the taxonomic related basidiomycetous fungus *Ustilago maydis* also has the ability to secrete large amounts of mannosylerythritol lipids (MELs) under conditions of nitrogen starvation.[12] Therefore, to characterize the MEL biosynthetic pathway is important to consider the information available for this fungus. To produce MEL is important to be gathered two conditions: lipid accumulation that only occurs under excess of carbon and nitrogen limitation; and the availability of NADPH.[9]

For MEL assembly there are essential three blocks mannose, erythritol and medium-chain fatty acids (C_8 a C_{10}). To yield mannose and erythritol from sugars it is necessary that such sugars (e.g glucose and xylose) are metabolized through the metabolic pathways described above, glycolysis and PPP. GDP-Mannose is produce through the glycolysis intermediate fructose-6-phosphate.[5] On the other hand, the other building block, erythritol is produced through the oxidative part of the PPP, or through the non-oxidative part of the PPP from glycolysis intermediates (fructose-6-phosphate and glyceraldeyde-3-phosphate).The PPP plays an important role in MEL production because as stated earlier, generate reducing equivalents in form of NADPH and ribose-5-phosphate and also generate erythrose-4-phosphate, used in synthesis of aromatic amino acids, but also as the precursor of erythritol.[5] For MEL production is crucial the presence of a medium-chain fatty acids. Studies suggested that most of the fatty acids incorporated in MEL are β-oxidation intermediates of C_{16}-C_{18} fatty acids, and that a specific type of chain-shortening pathway (partial β-oxidation) would contribute to produce medium-chain fatty acids in these yeasts.[28] Decrease of MEL synthesis was related with the inhibition of chain-shortening. Thus MEL synthesis is closely related to the operation of chain-shortening of longer fatty acids. Fatty acids are produced from pyruvate through cytosolic acetyl-CoA.

The genes coding proteins that are related with biosynthesis and export of MEL to outside of the cell had been identified.[24] For MEL assembly is necessary the formation of a hydrophilic moiety, mannosylerythritol, by the mannosylation of erythritol, which is most probably catalyzed by a glycosyltransferase, erythritol-mannosyl transferase (coded by *Emt1*). *Emt1* is essential for MEL biosynthesis and its expression is strongly induced by nitrogen limitation.[12] Then, MEL is produced via the reactions catalyzed by mannosyl acyl transferases (coded by *mac1* and *mac2*). These enzymes are supposed to catalyze the transfer of acyl groups to mannosylerythritol, so they are essential for MEL production.[12] Another of the proteins involved in this biosynthesis is mannosyl acetyl transferase that catalyze the transfer of acetyl groups (coded by *mat1*) (Figure 2-6). The acylation reactions seems to be very important for MEL secretion because while *mat1* mutants were still able to produce a significant amount of secreted MEL, deletion of either *mac1* or *mac2* resulted in complete loss of MEL production.[12]

With the information derived from several studies is possible to propose a biosynthetic pathway leading to MEL production from sugars. This pathway begins with the generation of mannosylerythritol, like preview described. Then this disaccharide is subsequently acylated with fatty acids of various lengths by the putative acyltransferases at positions C-2 and C-3. The final step of MEL biosynthesis is the acetylation of the fully acylated mannosylerythritol lipids and, at the end of this pathway, MEL is secreted by a putative MEL exporter (coded by *mmf1*).[12]

To add information relating to the already known about MEL biosynthesis, genes expression was evaluated as a function of the carbon source used, both with *P. antarctica* and *U. maydis*. Studies show that *P. antarctica* may be able to produce MELs using vegetable oils as substrates due to the high expression of the gene cluster for MEL biosynthesis, while the gene cluster expression of *U. maydis* is

sensitive to various nutrients, including carbon and nitrogen sources.[24] Also, *P. antarctica* has, potentially, expanded lipid metabolism (higher number of genes related with lipid metabolism) when compared with *S. cerevisiae* and is able to produce large amounts of MELs with soybean oil and also produce MELs from glucose.[24] In addition, it is already known that besides producing MEL from glucose, *Pseudozyma spp.* also has the capacity to produce MEL from pentoses and pentose/hexose mixtures, although without any metabolic information to compare pentoses with glucose utilization.[6]

More recent studies evaluate the transcriptomes of *P. antarctica* in the presence or absence of soybean oil using RNA-Seq.[8] Were found four genes organized in a cluster highly homologous to the known MEL clusters. This cluster encodes four enzymes of the MEL pathway *PaEMT1*, *PaMAC2*, *PaMAC1*, *PaMAT1*, and a major facilitator *PaMMF1*, which has been suggested to be responsible for the secretion of MEL. The 5 genes of the cluster are highly expressed in *P. antarctica* regardless of the carbon source, vegetable oil or glucose, whereas in *U. maydis* these genes were suppressed by the presence of vegetable oil.[24]. *P. antarctica* may be able to produce MELs culture media enriched with oil, due to the high expression of the gene cluster for MEL biosynthesis.[24].

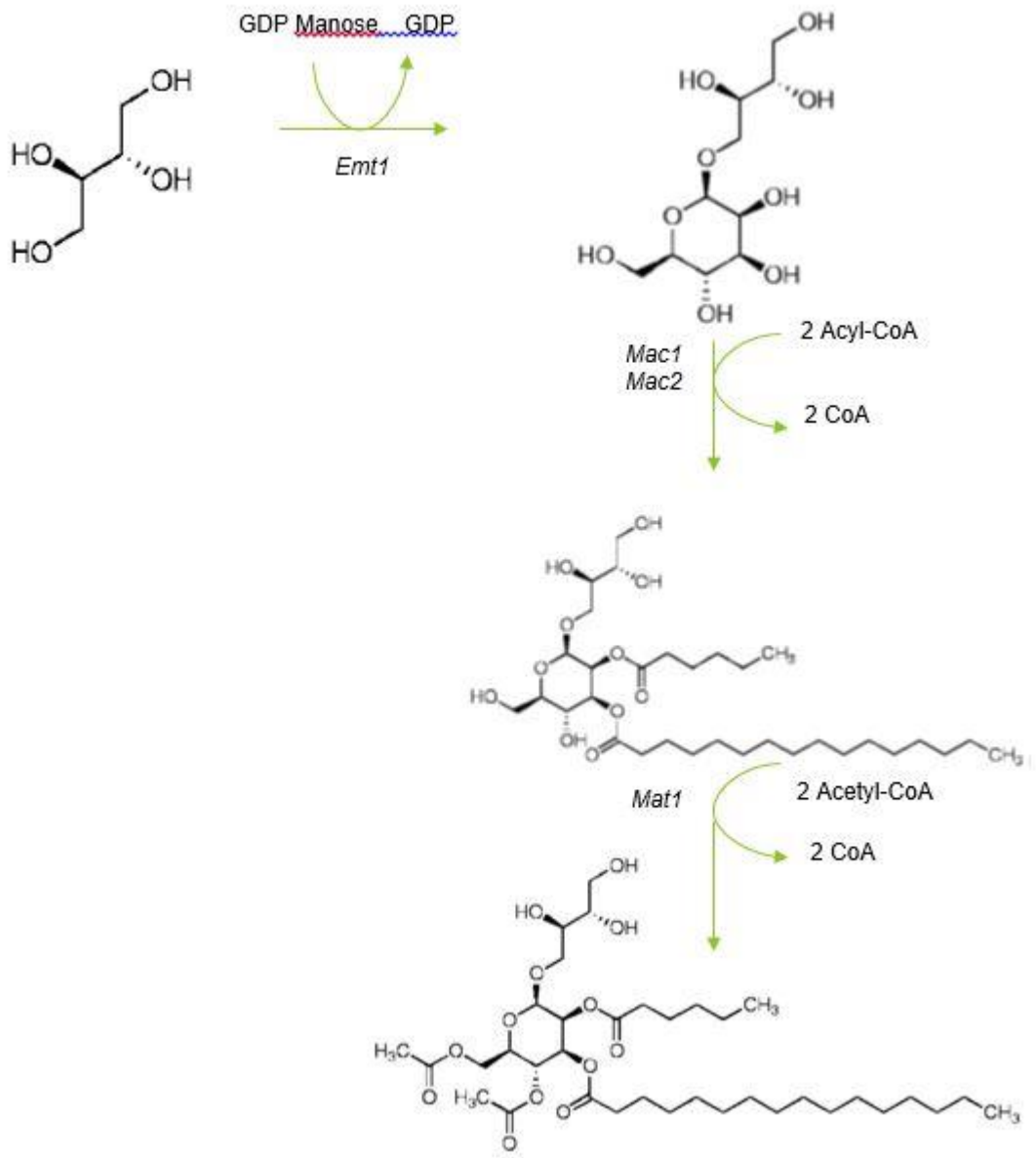


Figure 2-6: Biosynthetic pathway of MEL (adapted from Hewald et al. 2006). emt1, glycosyltransferase gene, mac1 and mac2, putative acyltransferases; mat1, acetyltransferase gene[12]

2.8 Energetics and redox balances in yeast: the role of NADPH

In yeasts is possible nominate NADPH-consuming and NADPH-producing metabolic processes (Figure 2-7).[21] In the context of this work, is important to highlight that this yeast generally requires NADPH to assimilate both nitrate and pentoses, as well as for fatty acid production.

To assimilate pentoses like D-xylose, is necessary a reduction to xylitol by an NAD(P)H dependent xylose reductase followed by an oxidation to xylulose by an NAD^+-dependent xylitol dehydrogenase.[18] To assimilate xylose, xylose reductase was considered to be strictly NADPH dependent, as in the majority of xylose-utilizing yeasts[29], the metabolic NADPH balance for xylose conversion into MEL is 1 mol NADPH per 1 mol D-xylose.[5]

Nitrate assimilation increase NADPH requirements caused by the activity of NADPH-dependent nitrate and nitrite reductase. When nitrate is used as the nitrogen source this compound is reduced via cytoplasmic nitrate and nitrite reductases with NADPH (Eq. 5 e 6).[21] The high demand for NADPH in MEL biosynthesis from pentose/nitrate mixture is higher than that from glucose/nitrate.[6] Due to the high NADPH demand when using nitrate and pentoses, it is not surprising that these metabolites are described as inducers of NADPH production in yeasts. [30]

For the MEL production it is essential the presence of NADPH because fatty acids, are highly reduced materials and to achieve their synthesis NADPH is essential as a hydrogen donor for fatty acid synthase (FAS) enzymatic complex for carbon condensation reactions.[5,31]

The two main routes to produce NADPH in the yeast cytosol are considered to be the PPP, and the citrate shuttle[21]. In yeasts, a high NADPH requirement must result in increased flows over the PPP pathway.[21] Usually oleaginous yeasts present higher values for PPP fluxes than the non-oleaginous yeasts due to their greater need of NADPH for fatty acid synthesis when compared with other yeasts (non-oleaginous).[5] In the other hand, malic enzyme activity, present in the citrate shuttle, has been found only in oleaginous microorganisms, where it is proposed to form an integrated metabolon complex that combines with ACL (Citrate Lyase) and fatty acid synthase (FAS) to ensure a direct channeling of acetyl-CoA into fatty acids.[9]

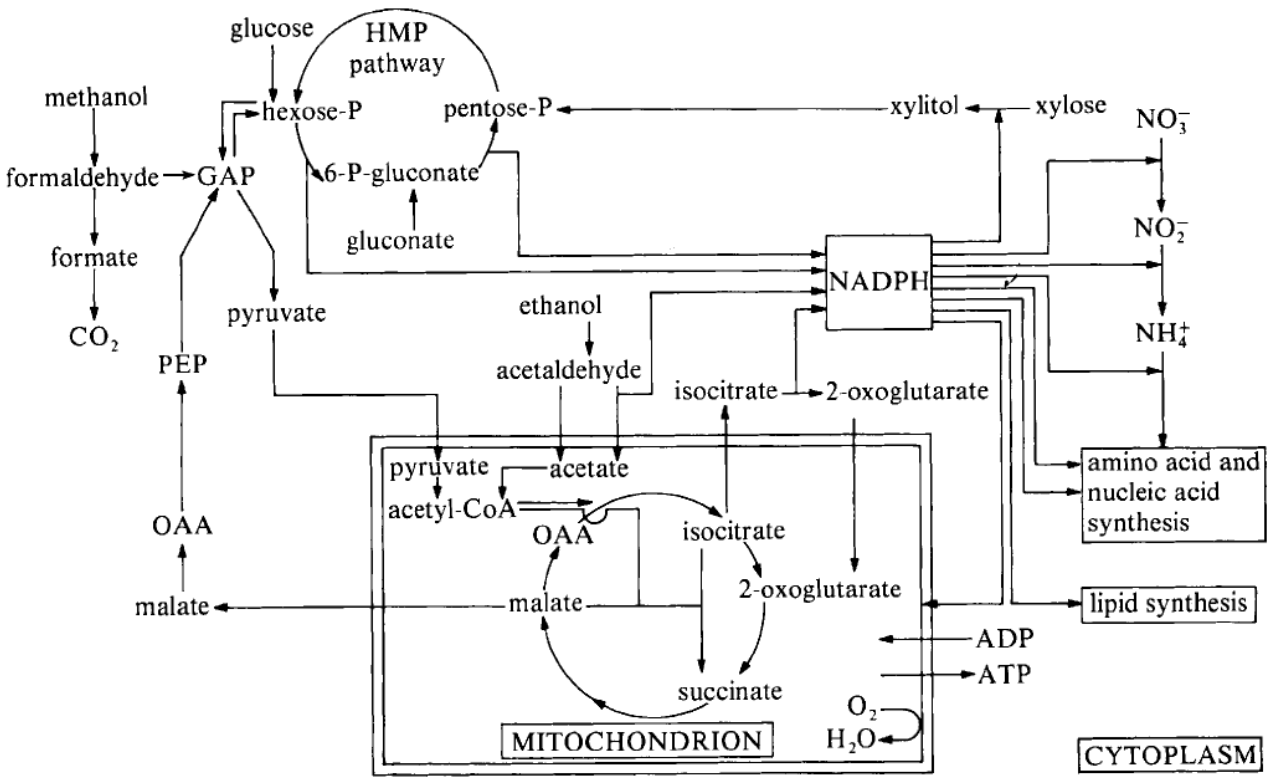


Figure 2-7: Schematic representation of NADPH-producing and NADPH-consuming processes and their localization in the yeast cell[21]

2.9 Tools to study metabolic pathways

Nowadays it is impossible to study a biological system as a whole. Systems biology is used to address many research questions.[32] Living systems are dynamic and complex, their behavior may be hard to predict from the properties of individual parts. To understand a biological system, is necessary the elucidation of the molecular basis of biological functions. This field of molecular biology is named functional genomics.[33] The functional genomic focuses on the way the components interact, for example regulation of the expression of the genes in a given functional condition.[33] For that reason is necessary to use quantitative measurements of the behavior of groups of interacting components, systematic measurement technologies such as genomics, transcriptomics, proteomics, metabolomics and bioinformatics, which combines the data from the other techniques to develop mathematical and computational models and to describe and predict dynamical behavior.[34]

Next generation sequencing (NGS) allowed the complete DNA sequencing of thousands of species, which combined with computational methods, revealed significant details about intra- and inter-specific properties.[33]

More recently, this technology has been used for the sequencing of total cDNA, in order to observe the complete transcriptome in a specific condition. With this approach the RNA-seq experiment, can derive not only an accurate, quantitative measure of individual gene expression but also be used to discover novel transcribed regions in an unbiased manner. In addition, this methodology enables a global survey of the usage of the alternative splice sites. In this approach exists high levels of sensitivity.[35]

Besides RNA-Seq, DNA-chips are also used from transcriptome analysis. This is a microarray technique based on hybridization of nucleic acids, often used with the total open reading frames from a sequenced organism. The cDNA arrays are used mostly for large-scale screening and expression studies. This tool is useful for the detection of mutations and expression monitoring, gene discovery and mapping.

In the study the enzymes or proteins it is also important to apply techniques that allow the study of proteins, i.e. their production by each strains under several conditions.

2D-PAGE is one of these techniques, where proteins were resolved on a gel using isoelectric focusing (IEF), which separates proteins in a first dimension according to their isoelectric point, followed by electrophoresis and in a second dimension, (in the presence of sodium dodecyl sulfate (SDS)), according to their molecular mass. The advantages of 2D-PAGE are that allow the resolution of large numbers of proteins, and staining of these proteins, and the determination of the relative abundances of the proteins to be quantified. Although 2D-PAGE has been limited by its inability to resolve membrane proteins[36].

There are several techniques to obtain information on metabolic products. Nuclear Magnetic Resonance is a technique where the nuclei of certain atoms are submitted to a static magnetic field and

exposed to a second oscillating magnetic field, allowing to identify many biological molecules through resonance of protons or labeled ^{13}C. Note that contrary to the carbon-12 nucleus, carbon-13 (C-13) nucleus has a nuclear spin, which, due to the presence of an unpaired neutron, can be detect by NMR technique.

^{13}C nuclear magnetic resonance (NMR) can be used to complement the previous biochemical studies and to investigate the *in vivo* carbon flux distribution through the catabolic pathways in the yeast strains. This technique allows the understanding of the differences between the catabolic pathways, in this case the glycolysis and PPP. Furthermore it allow gathering information on the central carbon metabolism.

NMR is powerful technique for the elucidation of metabolic pathways because it's unique analytical and nondestructive features. The use of ^{13}C-enriched compounds allows tracing the fate of specific carbon atoms through different metabolic routes. Furthermore, positional isotopic information derived from ^{13}C NMR can be used to characterize the flux distribution[29].

Isotopic labeling experiments coupled to ^{13}C-nuclear magnetic resonance (NMR) have been proven to be useful in determining metabolic fluxes. The distribution of the ^{13}C label among the different carbons of these intermediates and of the end products allows to follow pathways in considerable detail *in vivo* and to evaluate the relative rates of different reactions. However is important to note that these NMR experiments uniquely determined *in vivo* the extent of labeling among intermediates that may have rapid turnover rates, and thus they avoid the uncertainties encountered in extracting such metabolites[37].

On the other hand, the GC (Gas Chromatography) technique allows the study of metabolic products because, provides excellent separation, and allows efficient and precise separation of structurally similar volatile metabolites and metabolites that can be converted into volatile derivatives through derivatization. To identify compounds that are separated is possible to use different detectors, such as FID (Flame Ionization detector), thus allowing an identification of metabolites more effective and efficient. GC allows identification of many metabolites using the retention indices of retention times against standards. In the context of this work it is important to identify the size of chains of MEL that yeast produces.

One of the techniques related to the GC is LC (Liquid Chromatography) that has features making it applicable for analysis of a wide range of low-to high-polarity metabolites, including nonvolatile compounds that cannot be analyzed by GC and metabolites that are difficult to derivatize. One drawback of this technique is that LC does not achieve the excellent resolution possible with GC. However, is possible the use of HPLC (High Performance Liquid Chromatography) that has higher speed analysis and higher resolution compared to conventional systems.

Like in GC this technique combined with a mass spectrometer detector greatly contributes to achieving high sample throughput in metabolome analysis.[38,39]

These system biology or –omics tools provide enormous amount of data, but are still expensive. In the scope of this thesis, the research approach followed genomic studies and transcript and protein analysis

in order to get the first insights on the molecular basis for MEL production by *Pseudozyma* spp. from glucose and xylose, for future system biology approaches and detailed studies of specific pathways of interest. The genomic analysis allows the identification, measurement or comparison of genomic features. Through genomic analysis is possible the identification of gene as well as understand its structure and function. This analysis is possible thanks to the development of tools and databases that provide information about genomes, including sequences, maps, chromosomes, assemblies, and annotation.

Transcript analysis was performed by Real-Time RT-PCR, which is one of the most widely used methods for mRNA quantitation, since it has a large dynamic range, high sensitivity, and sequence-specity. As for RNA-seq, Real-Time RT-PCR experiments also require RNA isolation to obtain cDNA. This technique of collecting data throughout the PCR process combines amplification and detection into a single step, this is possible thought the use of a variety of different fluorescent chemistries that correlate PCR product concentration to fluorescence intensity.

For the aim of this work, Real-Time RT-PCR allow the quantification of gene expression, for example of genes involved in MEL biosynthesis, in two yeasts and at different culture conditions (e.g. carbon source).[40]

Enzymatic assays of crude cell extracts allow the detection and quantification of active enzymes. This technique allows, to measure many parameters, for example, k_{cat}, $V_{máx}$, k_m and specificity One of the most important advantages is that enzymes can be identified by their catalyzed reactions, in contrast to the other components of the cell.[41] Usually, in an enzyme assay, the main goals is to identify a particular enzyme, to prove its presence or absence or to determine the amount of the enzyme present.[42] There are different kinds of enzymatic assays which need to be adapted according with the features of the enzyme targeted.

Nowadays exist a variety of methods that allow the study of the enzymes, in the simplest case an enzyme reaction can be observed by the appearance (or disappearance) of a colored compound, applying a colorimeter or a photometer and quantifying the color intensity against a calibration curve. Another method is Fluorimetry that is more sensitive but only a few enzymatic substrates or products emit fluorescence, such as NAD(P)H, tryptophan and some artificial substrate analogues.[41] As previously shown in the biosynthetic pathway MEL, NADPH (it is possible the transformation of NADH into NADPH)[43] plays an important role, and it is involved in many metabolic pathways. Among the enzymes involved in MEL biosynthesis are those related with NADPH production and consumption.

3 Materials and Methods

3.1 Microorganisms and maintenance

The two *Pseudozyma* yeasts strains *P. antarctica* PYCC 5048^T (CBS 5955), *P. aphidis* PYCC 5535^T (CBS 6821) were provided by the Portuguese Yeast Culture Collection (PYCC), CREM, FCT/UNL, Caparica, Portugal. Stains were plated in Yeast Malt Agar (YM-agar) medium (yeast extract 3 g/l, malt extract 3 g/l, peptone 5 g/l D-glucose 10 g/l and agar 20 g/l) and incubated for 3 days at 25°C. Cultures were kept at 4°C and renewed every 2 weeks and stored at -80°C in 20% glycerol to be recovered when necessary.

3.2 Media and cultivation conditions

Inoculum was prepared by incubation of stock cultures of *P. antarctica* and *P. aphidis* at 28°C, 140 rpm, for 48 h, in liquid media containing D-glucose or D-xylose (40 g/l), $NaNO_3$ (3 g/l) or NH_4Cl (1.89 g/l), $MgSO_4$ (0.3 g/l), KH_2PO_4 (0.3 g/l) and yeast extract (1 g/l). Cells were harvested and transferred to fresh medium to obtain an initial Optical Density (OD) at 640 nm of approx. 0.5 in Erlenmeyer flasks containing 200 ml of medium (headspace to volume ratio of 5:1) and incubated under the same conditions[17,44]. Alternatively, inoculum was prepared in medium containing D-glucose and $NaNO_3$ and cells were harvested and transferred to the different combinations of carbon and nitrogen source Table 3-1.

Table 3-1: Conditions tested for MEL production in each strain; inoculum and culture media in each fermentation.

Condition tested in both strains *P. aphidis* and *P.antarctica*	Fermentation I	Fermentation II	
	Inoculum and culture media	Inoculum	Culture media
Glucose + $NaNO_3$	Glucose + $NaNO_3$	Glucose + $NaNO_3$	Glucose + $NaNO_3$
Glucose + NH_4Cl	Glucose + NH_4Cl		Glucose + NH_4Cl
Xylose + $NaNO_3$	Xylose + $NaNO_3$		Xylose + $NaNO_3$
Xylose + NH_4Cl	Xylose + NH_4Cl		Xylose + NH_4Cl

3.3 Sampling and growth and biomass determination

To analyze the profile of yeast growth (biomass), sugar consumption and MEL and fatty acid production, samples were collected periodically and processed according to each goal. Culture samples of 1 mL were used for biomass and sugar quantification. Cells were harvested at 10,000 g, for 10 min, at 4ºC, washed with deionized water and dried at 100°C for 24 h to determine cell dry weight (CDW)[6]. The supernatants were stored at -20ºC for sugar analysis. For MEL (and fatty acid) quantification, culture samples of 3 mL were collected, freeze-dried and weighted for further analysis.

3.4 Sugar quantification

Supernatants were diluted (1/10) with a sulfuric acid 0.05 M solution (1:1) and centrifuged at 10,000 g, for 1 min, and analyzed for glucose or xylose quantification by high performance liquid chromatography (HPLC) system (Merck Hitachi, Darmstadt, Germany, or Agilent 1100 Series HPLC System, Germany) equipped with a refractive index detector and a Rezex ROA-Organic acid H^+ column (300 mm ~ 7.8 mm, Phenomenex), at 40°C[6]. Sulfuric acid (0.005 M) aqueous solution was used as mobile phase at 0.5 ml/min.

3.5 MEL and fatty acid quantification

The quantification of MEL and fatty acids in biological samples was determined by methanolysis and GC analysis of methyl esters[6,45]. The samples dissolved in water were stored and freeze-dried, and the tubes containing the samples were weighed before and after freeze-drying. The freeze-dried mass

sample was transferred to one reaction tube (Schott AG, Germany). Pure methanol (20 ml) was cooled down to 0°C and 1 ml acetyl chloride was added, which generated a water-free HCl/methanol solution. Two ml of HCl/methanol solution was added to the freeze-dried mass and incubated for 1 h at 80°C for reaction into methyl esters. Heptanoic acid (4% in hexane) was used as internal standard**.** The resulting product was extracted with hexane (1 ml) and 1 µl of the organic phase was injected in a GC system (Hewlett-Packard, HP5890), equipped with a FID detector and a SUPELCOWAX® 10 capillary column (L × I.D. 60 m × 0.32 mm, df 0.25 µm). The oven was programmed from 90°C (held for 3 min) to 200°C at 15°C/min. Nitrogen gas was used at a flow rate of 50 ml/h. The quantification of MELs were based in the amount of C_8, C_{10} and C_{12} fatty acids considering a molecular weight between 574 and 676 g/mol depending on the length of the two acyl chain (C_8 to C_{12}). The quantification of glycolipids based on a specific moiety was previously described[6,45].

3.6 Enzymatic activity assays

Several enzymes activities involved in NADPH production and consumption were assessed. After 4 days of incubation, a culture sample of 10 mL was harvested at 10.000 g for 10 min. The supernatant was discarded and the cell pellet was washed with demineralized water. The cells were disrupted with 1 mL Y-PER™ (Yeast Protein Extraction Reagent, Pierce, Thermo Scientific, USA) per 0.6 g cell wet weight, for 50 min shaking at room temperature. The protein content of the supernatant was determined using Pierce™ BCA protein assay kit (Thermo Scientific, USA). The enzyme activities were performed following methodologies previously reported for other yeasts[30], through spectrophotometric assays carried out in a Shimadzu UV 2401PC Spectrophotometer, at 25ºC, with a total reaction volume of 250 µL, where the production and consumption of NADPH were followed at $340_{nm.}$during 10 minutes.

The Glucose-6-phosphate dehydrogenase (EC. 1.1.1.49) activity was assayed with 50 mM Tris/HCl buffer (pH 8-0), 5 mM MgCl and 0.4 mM $NADP^+$. The reaction was started with 5 mM glucose 6-phosphate.

The 6-Phosphogluconate dehydrogenase (EC: 1.1.1.44) activity was assayed with 50 mM glycylglycine buffer (pH 8.0) and 0.4 mM $NADP^+$. The reaction was started with 2 mM 6-phosphogluconate.

The isocitrate dehydrogenase (EC 1.1.1.42) activity was assayed with 100 mM potassium phosphate buffer (pH 7.0), 5 mM MgCl and 0.4 mM $NADP^+$. The reaction was started with 1 mM trisodium isocitrate.

The malic enzyme (EC: 1.1.1.40) activity was assayed with 50 mM triethanolamine buffer (pH 7.5), 1 mM $MnCl_2$ and 0.4 mM $NADP^+$. The reaction was started with 2 mM malate.

The xylose reductase (EC 1.1.1.307) activity was assayed with 50 mM potassium phosphate buffer (pH 6.0) 0.15 mM NADPH. The reaction was started with 100 mM xylose.

The nitrate reductase (EC 1.7.1.2**)** activity was assayed with 50 mM potassium phosphate buffer (pH 7) and 0.15 mM NADPH. The reaction was started with 10 mM $KN0_3$.

The nitrite reductase (EC 1.7.2.1) activity was assayed with 50 mM potassium phosphate buffer (pH 7), 10 mM $MgS0_4$ and 0.15 mM NADPH. The reaction was started with 2.5 mM $NaNO_2$.

3.7 Molecular biology tools

3.7.1 DNA extraction and purification

Cultures of *P. antarctica* and *P. aphidis*, grown in liquid medium under the conditions described above, were used for DNA extraction and purification.

To 1 mL of *P. antarctica* or *P. aphidis* culture, 100 µL of lysis buffer with urea (8 M urea, 0.5 M NaCl, 20 mM Tris, 20 mM EDTA and 2% SDS) was added and the mixture incubated at 65°C for 1 h. Then, 200 µL of chloroform-isoamylalcohol (4°C) was added and incubated 30 min on ice. A centrifugation step was performed for 10 minutes at 13,200 rpm and 4°C, the supernatant was removed, added to an equal volume of isopropanol (85 µL) and incubated 30 min at room temperature. After centrifugation for 10 min, at 14,800 rpm and 4°C, the supernatant was discarded and the pellet washed with 200µL of 70% ethanol (-20°C) following the same centrifugation procedure. Ethanol was evaporated and the pellet ressuspended in 1 µL of RNase and incubated 30 min at 55°C. Finally, the DNA extracted was ressuspended in 30 µL of sterile water and stored at -20°C. The DNA was analyzed in a NanoDropTM spectrophotometer (2000c Thermo Scientific), for quantification and quality control (260/230 nm and 260/280 nm ratios).

3.7.2 RNA Extraction and purification

Total RNA extraction and purification was performed with *P. antarctica* and *P. aphidis* cultures grown in 40 g/l D-glucose or D-xylose and 3 g/l $NaNO_3$ for 4 days at 28°C and 140 rpm. One ml culture was frozen in liquid nitrogen and kept at -70ºC until the RNA extraction procedure. The RNA extraction was performed using the Direct-Zol MiniPrep (Zymo Research)[46], with a few modifications. The samples were thaw in ice-bath, centrifuged at 12,000 rpm, for 1 min at 4°C, and the supernatant discarded. Cell lysis was achieved by adding 1 ml of the TRI Reagent, vortexing and incubated for 5 min at room temperature. After centrifugation at 12,000 rpm, for 1 min at 4°C, the supernatant was recovered. The mixture was placed into a Zymo-Spin™ IIC Column within a collection tube, centrifuged for 1 min and the column was transferred into a new collection tube. The column was washed with 400 µl RNA Wash Buffer and centrifuged for 30 s. The samples were treated, on-column, with DNase, for 35 min. After centrifugation at 12,000 rpm, for 30 s,

400 µl Direct-zol™ RNA PreWash was added to the column, followed by centrifugation for 1 min. The flow-through was discarded and this washing step was repeated. After that, 700 µl of RNA Wash Buffer was added to the column, followed by centrifugation for 1 min. The flow-through was discarded and, to ensure complete removal of the wash buffer, another centrifugation was performed in a new collection tube. The column was carefully transferred into an RNase-free tube and 50 µl of DNase/RNase-Free Water was added directly to the column matrix, followed by centrifugation for 1 min. After purification, all RNA samples were analyzed in a NanoDrop™ spectrophotometer (2000c Thermo Scientific) for quantification and quality control (260/230 nm and 260/280 nm ratios), and then stored at -70°C.

3.7.3 cDNA Synthesis

cDNA was prepared with RNA samples from *P. antarctica* and *P. aphidis* grown under the conditions described above. Maxima H Minus First Strand cDNA Synthesis Kit (with dsDNase)[47] was used. After genomic DNA elimination, the following reagents were added into the same tube: 1 µl of oligo (dT) primer, 1 µl of 10 mM dNTP Mix and 13 µl of water, nuclease-free, 4 µl of 5X RT Buffer and 1 µl of Maxima H Minus Enzyme Mix. The mixture were centrifuged for 1 min at 10,000 rpm and incubated for 30 min at 50 °C. Finally the reaction was terminated by heating at 85°C for 5 minutes. The cDNA was stored at -20 °C.

3.7.4 Selection of genomes and DNA/cDNA sequences from databases

Taking into account that the *P. aphidis* and *P. antarctica* strains are sequenced and the partial genome available at the NCBI database, the DNA or cDNA sequences of interest were directly accessed and, when the sequences were not directly identified, the *Ustilago maydis* genome was used as reference applying the NCBI blastn, blastp or blastx tool.

The selected DNA and cDNA are those related with MEL biosynthesis and NADPH production and consumption (Table 4-5 and Table 4-6), including those from the oxidative PPP, citrate shuttle, nitrate and xylose metabolism.

3.7.5 Design of primers for Real-Time RT-PCR

Bioedit software[48] was used to align the DNA/cDNA sequences from *P. antarctica* and *P. aphidis* through the ClusterIW Multiple Alignment algorithm and conserved regions between the genes across the two strains were identified. The primers were designed with the following criteria: cDNA amplicons of 150-200 bp; amplicon in the 5'-half of the cDNA; primer length of 20±1 bp; primer GC content of 50-60%;

avoidance of 3' primer dimers. Reference gene chosen were *ACT1* (actin) and *TAF10* (RNA polymerase II, TATA box binding protein (TBP)-associated factor).

3.7.6 Amplification of DNA and cDNA fragments by PCR

The designed primers were tested by PCR with the DNA and cDNA obtained as described above. PCR was performed using a mix for each primer tested containing 24.6 µL of sterile water, 4 µL of PCR buffer, 4 µl of 2 mM deoxynucleoside triphosphates (dNTPs), 1.6 of 10 µM of each primer, and 0.2 µl (2.U/µl) of Dream Taq. Template DNA or cDNA of *P. antarctica* and *P. aphidis* (1 µl) was added to a 0.5-ml thin-walled PCR tube, followed by the addition of 9 µl of PCR mix in each tube to give a final volume of 10 µL. The cycling parameters were (i) 98°C for 4.0 min, 39 cycles: 30 s 98°C, 30 s 55°C, 30 s 72°C, 8 min 72°C, 4°C. The amplification products were analyzed by agarose gel electrophoresis using 2% Agarose (Bioline) containing 3 µL of gel red (Biotarget). The GeneRuler 100 bp DNA Ladder[49] was used in electrophoresis to estimate the molecular mass of double-stranded DNA amplicons. The running parameters of the electrophoresis gel were 70 V and 400 mA. Gels were visualized using a Doc XR+ gel documentation system (Bio-Rad Laboratories, Inc).

3.7.7 Transcript analysis

Pseudozyma antarctica and *P. aphidis* RNA samples were used as template for one-step Real-Time RT-PCR reactions (Applied Biosystems)[50] to determine relative mRNA levels of genes related with MEL production and NADPH consumption and production against *ACT1* and *TAF10* mRNA levels, used as reference genes. Real-Time RT-PCR was performed with 50 ng of total RNA per condition, with the following RT program: Power SYBR Green RNA-to-C_T™ 1-Step Kit (30 min 48°C; 10 min 95°C; 40 cycles: 15 s 95°C; 1min 60°C). All reactions were performed in triplicate and specificity of amplification products was confirmed by the melting curve analysis.

4 Results and Discussion

4.1 Physiology of *P. aphidis* and *P. antarctica* in MEL production: influence of carbon and nitrogen sources and inoculum conditions

4.1.1 Influence of the carbon source in *P. aphidis* and *P. antarctica*, using $NaNO_3$ as nitrogen source

First, the two strains, *P. aphidis* and *P. antarctica*, were firstly cultivated in a medium containing D-glucose or D-xylose as carbon source, and $NaNO_3$ as nitrogen source. The inoculum was pre-cultivated in D-glucose supplemented with $NaNO_3$ as nitrogen source, even in the cases D-xylose was the carbon source tested. These conditions were previously described, in our group, as conditions in which both yeasts can produce MEL[6]. Growth, sugar consumption and MEL production under these conditions were evaluated and the main physiological parameters determined (Figure 4-1and Table 4-1).

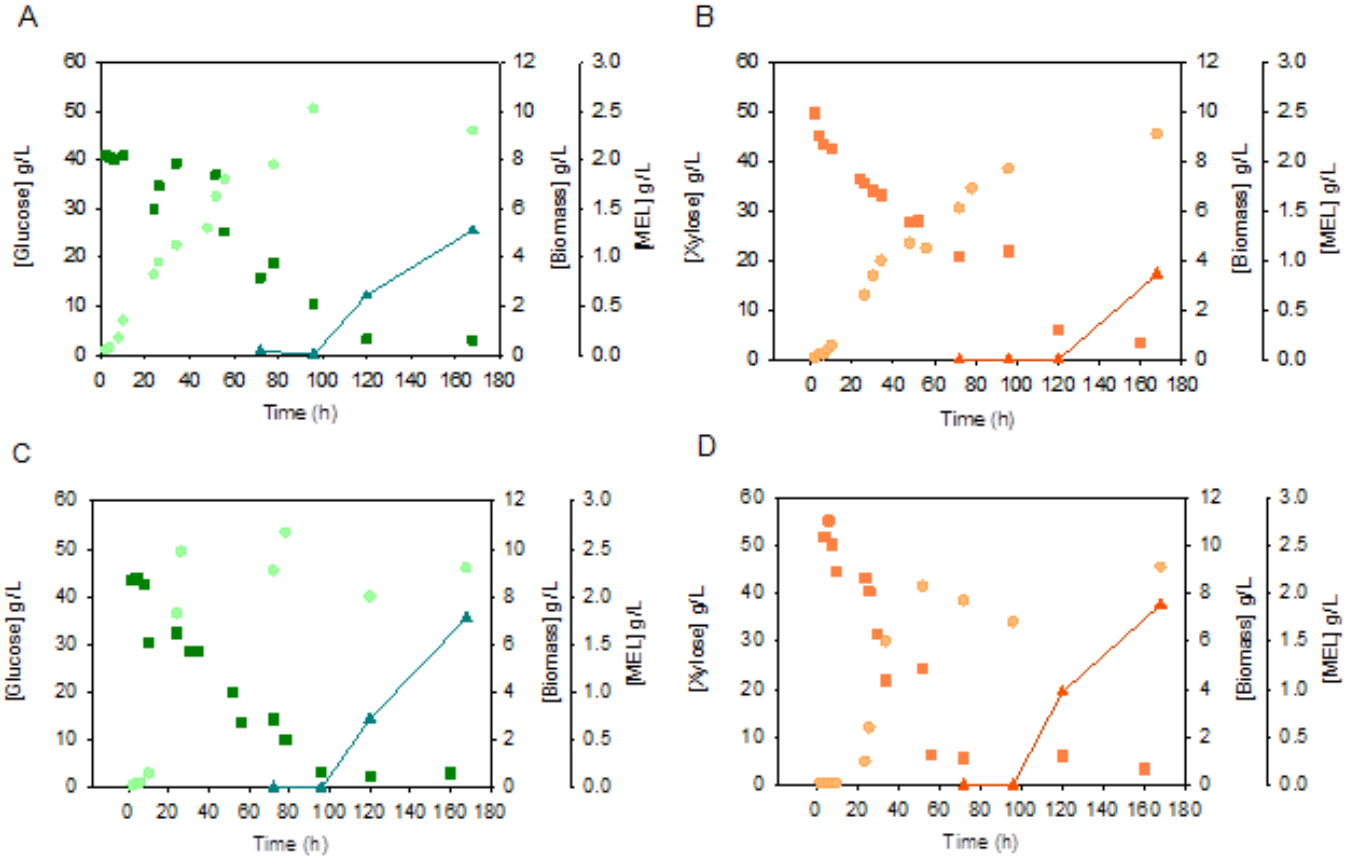


Figure 4-1: Sugar consumption (■) and MEL (▲) and biomass production (●) with P. aphidis (A, B) and P. antarctica (C, D), using D-glucose (A, C) or D-xylose (B, D) as carbon source and $NaNO_3$ as nitrogen source, at 28 ºC and 140 rpm.

Table 4-1: Physiological parameters obtained from P. aphidis and P. antarctica cultivations in D-glucose or D-xylose, using $NaNO_3$ as nitrogen source, at 28ºC and 140 rpm.

Strain	*P. aphidis*		*P. antarctica*	
Carbon source	D-Glucose	D-Xylose	D-Glucose	D-Xylose
μ (h^{-1})	0.24	0.21	0.21	0.19
$Biomass_{max}$ (g/l)	10.1	9.1	10.7	8.5
Sugar consumption rate (g/l/h)	0.38	0.45	0.43	0.45
$Y_{Xmax/S}$ (g/g)	0.25	0.23	0.27	0.21
MEL_{max} (g/l)	1.28	0.86	1.70	1.88

Pseudozyma aphidis and *P. antarctica* were able to grow in media containing D-glucose or D-xylose as sole carbon and energy source with a specific growth rate between 0.19 and 0.24 h^{-1}. For both strains, the specific grow rate is higher in D-glucose, with 0.24 h^{-1} for *P. aphidis* and 0.21 h^{-1} for *P. antarctica*. The lowest specific growth rate was found for *P. antarctica* in D-xylose (0.19 h^{-1}). Sugar consumption rates were lower for *P. aphidis* cultured in glucose and xylose (0.38 and 0.33 g/l/h, respectively) when compared with *P. antarctica* cultures (0.43 and 0.45 g/l/h, respectively). Maximum biomass cell dry weight was higher in D-glucose, over 10 g/l with both strains, 10.1 g/l for *P. aphidis* and 10.7 g/l for *P. antarctica*. In D-xylose, biomass formation was slightly lower in both strains.

In the previous studies both strains tested were able to grow on minimal medium using D-glucose and D-xylose, with a specific growth rate of 0.21 h^{-1} for D-glucose for both *P. aphidis* and *P. antarctica*, 0.17 h^{-1} and 0.13 h^{-1} for D-xylose in *P. aphidis* and *P. antarctica*, respectively.[6] Comparing these results with the ones obtained in this study it is possible to conclude that the trend was similar, as the results in D-glucose were similar in both strains and higher than in D-xylose, with higher D-xylose growth in *P. aphidis*. The previous studies reported that sugar consumptions were similar among the strains tested (0.3–0.4 g/l/h)[6]. Although a slight difference on sugar consumption rate observed between strains tested in this work, the interval observed (0.33–0.45 g/l/h) is close to what has been reported. In case of *P. aphidis*, the D-glucose consumption rate was lower when compared with the one reported in literature.

In previous studies, MEL was produced from D-glucose and D-xylose with the *Pseudozyma* strains tested, reaching MEL concentrations between of 2.8 and 1.2 g/l in D-glucose and D-xylose, respectively for

P. aphidis, and 5.4 and 4.8 g/l in in D-glucose and D-xylose, respectively for *P. antarctica*[6], both in low nitrate concentration. In media with nitrate, as in this study, MEL titres were in the range of 2-5 g/l for *P. antarctica*[6]. The results already published show that *P. antarctica* PYCC 5048^T was the most efficient MEL producer from either D-glucose or D-xylose. In this study, MEL production did not reached concentrations as high as the ones reported in literature, partially because the media used contained 3 g/l nitrate but also because in the present study, the MEL titres were measured up till day 7, while maximum MEL titres in previous studies were reported between day 7 and 14.

Still, based on the values obtained it was possible to confirm that *P. antarctica* is the most efficient MEL producer reaching similar titres in D-glucose and in D-xylose.

4.1.2 Influence of ammonia as nitrogen source in *P. aphidis* and *P. antarctica*

In order to evaluate the influence of a different nitrogen source in *P. aphidis* and *P. antarctica*, $NaNO_3$ was replaced by NH_4Cl. Still, the inoculum was pre-cultivated in D-glucose supplemented with $NaNO_3$ as nitrogen source. Growth and sugar consumption were followed and the main physiological parameters determined (Figure 4-2,Table 4-2).

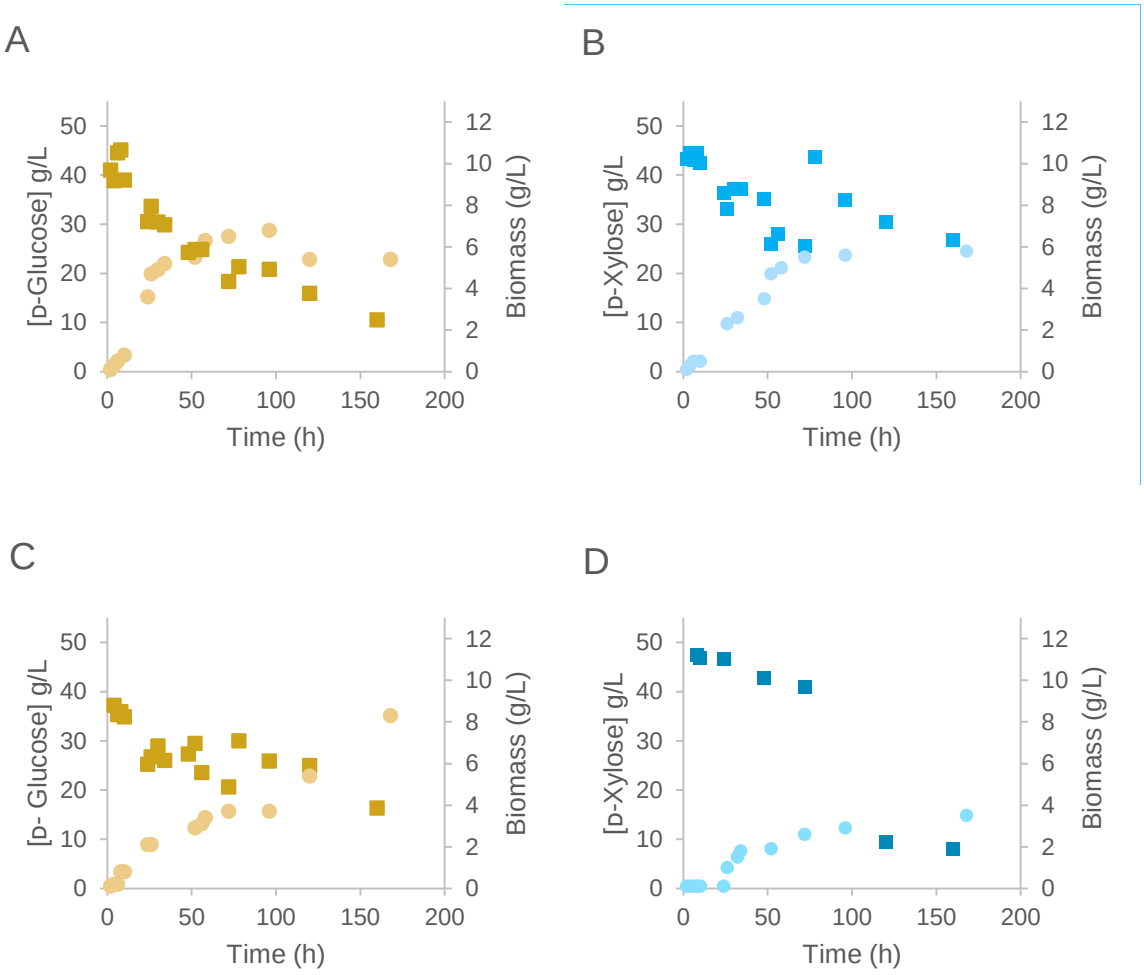


Figure 4-2: Sugar consumption (■) and biomass production (●) with P. aphidis (A, B) and P. antarctica (C, D), using D-glucose (A, C) or D-xylose (B, D) as carbon source and NH_4Cl as nitrogen source, at 28 ºC and 140 rpm..

Table 4-2: Physiological parameters obtained from P. aphidis and P. antarctica cultivations in D-glucose and D-xylose, using NH_4Cl as nitrogen source, at 28ºC and 140 rpm.

Strain	*P. aphidis*		*P. antarctica*	
Carbon source	D -Glucose	D -Xylose	D -Glucose	D -Xylose
μ (h^{-1})	0.24	0.11	0.25	0.07
$Biomass_{max}$ (g/l)	6.8	5.8	8.3	3.5
Sugar consumption rate (g/l/h)	0.31	0.27	0.39	0.11
$Y_{Xmax/S}$ (g/g)	0.17	0.16	0.21	0.08
MEL_{max} (g/l)	0	0	0	0

Pseudozyma aphidis and *P. antarctica* were able to grow in media containing NH_4Cl as nitrogen source, with a specific growth rate between 0.07 and 0.25 h^{-1}. Both strains grew on D-glucose when NH_4Cl was used as nitrogen source, at comparable specific growth rates (>0.20 h^{-1}) of those obtained when using $NaNO_3$. However, the growth in D-xylose was poor, with specific growth rates of 0.11 and 0.07 h^{-1} for *P. aphidis* and *P. antarctica*, respectively. Comparing the sugar consumption rates, it varies between 0.11 to 0.39 g/l/h. The maximum biomass achieved in both strains is higher in the presence of $NaNO_3$, 9.1 g/l for *P. aphidis* and 8.5 g/l for *P. antarctica*. The growth in D-glucose was faster than in D-xylose, as the sugar consumption and the biomass production was higher. When comparing both strains in each carbon source with NH_4Cl as nitrogen source, they presented a similar behavior between each other.

Nitrate assimilation pathway is known to consume NADPH to convert nitrate in nitrite and for further nitrite conversion. On contrary, NH_4Cl does not require NADPH for assimilation. Thus, the use of NH_4Cl could increase the amount of NADPH available leading to higher MEL production. However, MEL was not detected in any condition using NH_4Cl as nitrogen source.

No publication was found regarding the use of NH_4Cl as nitrogen source for MEL production with *Pseudozyma spp.*.

4.1.3 Influence of pre-inoculum conditions on the physiology of *P. aphidis* and *P. antarctica*

The effect of the pre-inoculum medium in *P. aphidis* and *P. antarctica* was evaluated in relation to the growth, the sugar consumption and the MEL production. In the previous sections, the composition of the pre-inoculum medium was set at 40 g/l D-Glucose, 3 g/l $NaNO_3$, 0.3 g/l KH_2PO_4, 0.3 g/l $MgSO_4$ and 1 g/l of yeast extract, and then transferred to each cultivation medium, which included D-glucose or D-xylose as carbon source and $NaNO_3$ or NH_4Cl as nitrogen source (Table 4-1 and Table 4-2). In this section the pre-inoculum media was the same of the one for the cultivation media. Growth, sugar consumption and MEL production were followed Figure 4-3 and Figure 4-4, Table 4-3).

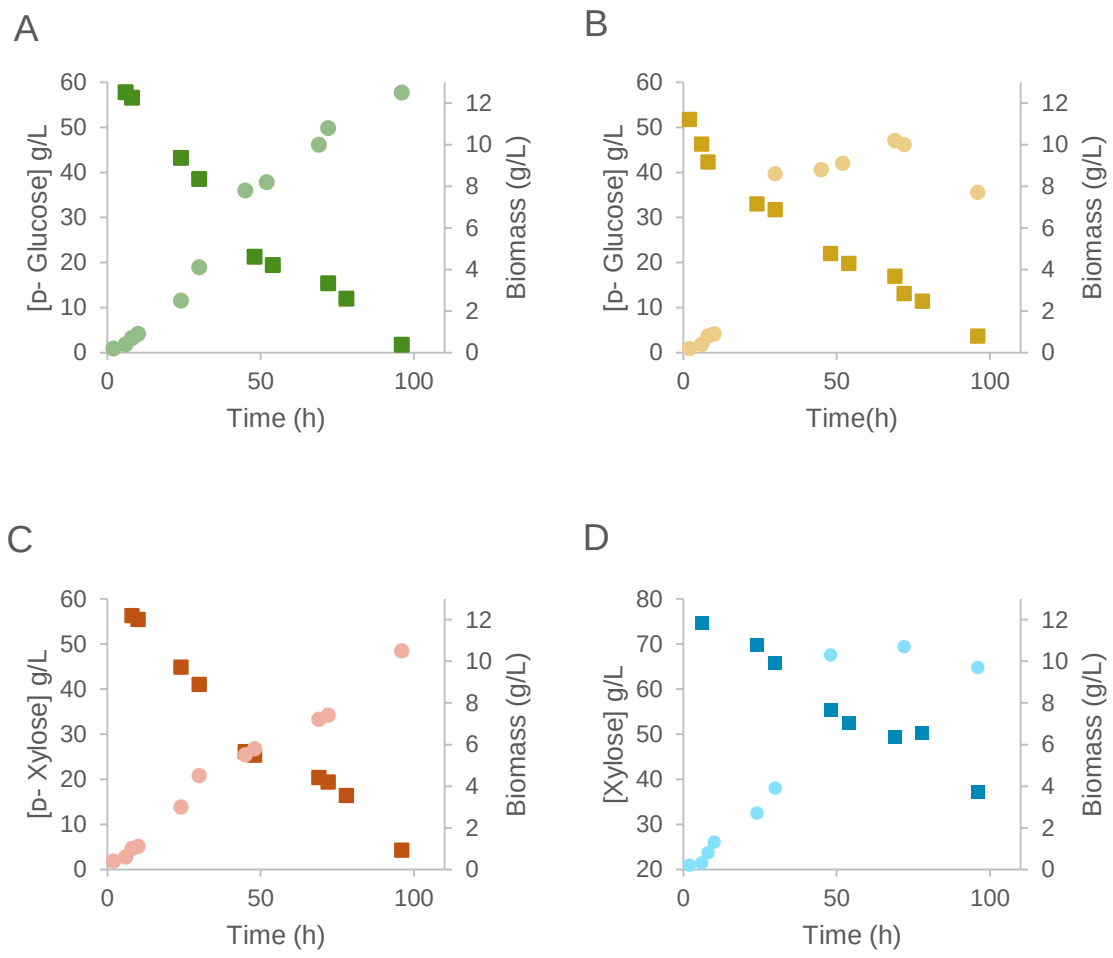


Figure 4-3: Sugar consumption (■) and biomass production (●) with P. aphidis, using D-glucose (A, B) or D-xylose (C, D) as carbon source, and $NaNO_3$ (A, C) or NH_4Cl (B, D) as nitrogen source, at 28 ºC and 140 rpm. Pre-inoculum medium was similar to the respective cultivation condition.

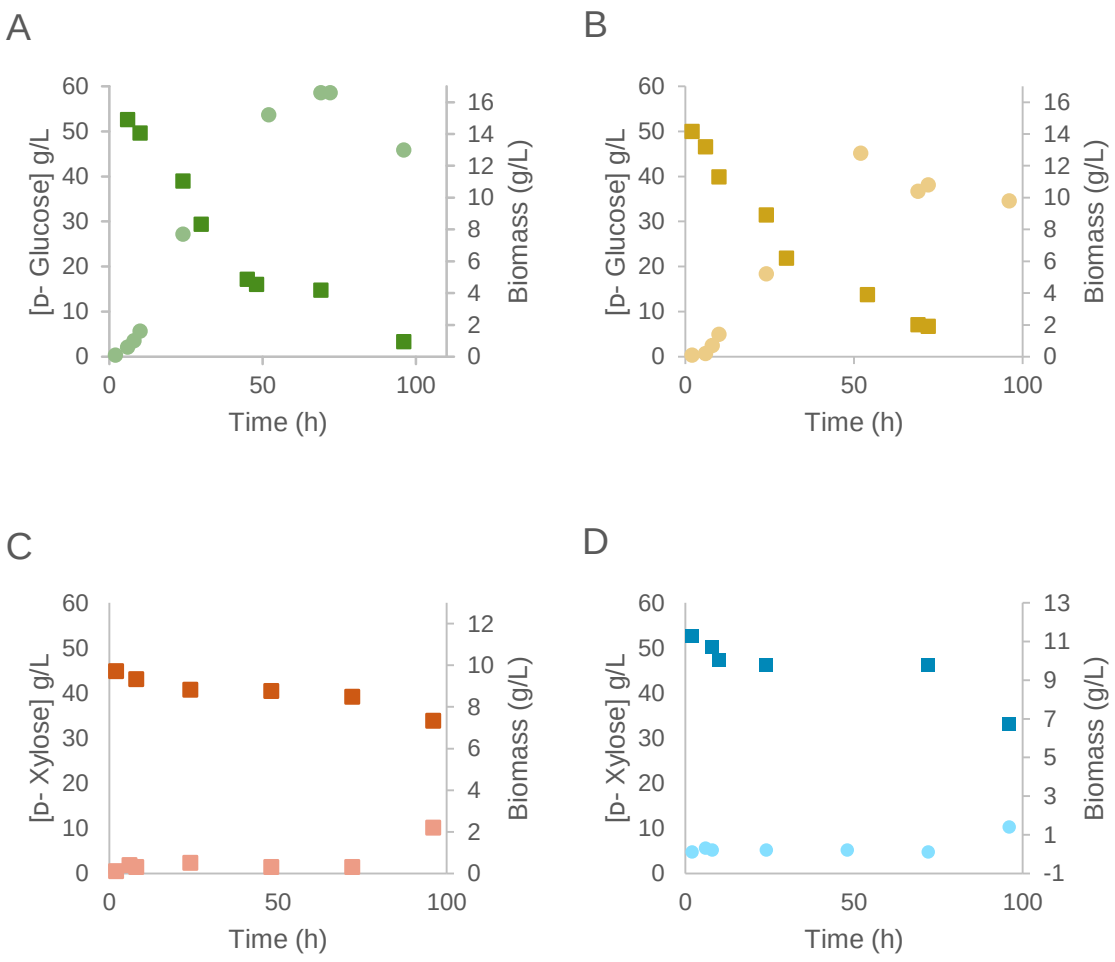


Figure 4-4: Sugar consumption (■) and biomass production (●) with P. antarctica, using D-glucose (A, B) or D-xylose (C, D) as carbon source and $NaNO_3$ (A, C) or NH_4Cl (B, D) as nitrogen source, at 28 ºC and 140 rpm. Pre-inoculum medium was similar to the respective cultivation condition.

Table 4-3: Physiological parameters obtained from P. aphidis and P. antarctica cultivations in D-glucose or D- xylose, using $NaNO_3$ or NH_4Cl as nitrogen source, at 28ºC and 140 rpm. Pre-inoculum medium was similar to the respective cultivation condition.

Strain	*P. aphidis*				*P. antarctica*			
Carbon source	D-Glucose		D-Xylose		D-Glucose		D-Xylose	
Nitrogen source	$NaNO_3$	NH_4Cl	$NaNO_3$	NH_4Cl	$NaNO_3$	NH_4Cl	$NaNO_3$	NH_4Cl
μ (h^{-1})	0.20	0.20	0.14	0.24	0.25	0.33	0.18	0.12
Biomass $_{max}$ (g/l)	12.5	10.2	10.5	10.7	16.3	12.8	2.2	1.4
Sugar consumption rate (g/l/h)	0.68	0.49	0.80	0.41	0.90	0.92	0.18	0.19
$Y_{Xmax/S}$ (g/g)	0.31	0.26	0.26	0.31	0.41	0.32	0.06	0.04
MEL_{max} (g/l)	0.72	0.92	2.58	0.72	4.46	3.68	2.30	0

Specific growth rate observed in *P. aphidis* cultures with $NaNO_3$ was lower than the reported in the literature[6] and in section 4.1.1. The specific growth rate was 0.20 h^{-1} and 0.14 h^{-1} in D-glucose and D-xylose, respectively. However, sugar consumption rate for both D-glucose and D-xylose was higher than described in the previous section 4.1.1 and described in the literature[6]. When using NH_4Cl, similar maximum biomass but lower sugar consumption rate is observed if compared with same conditions in $NaNO_3$. *Pseudozyma antarctica* cultured in D-xylose did not reach, both with $NaNO_3$ or NH_4Cl, comparable maximum biomass values to D-glucose cultures (2.2 and 1.4 g/l, respectively, in D-xylose cultures, against 16.3 and 12.8 g/l, respectively, in D-glucose cultures), showing also lower sugar consumption rate and specific growth rate. The specific growth rate and the sugar consumption rate determined in D-glucose cultures was higher than those reported in the literature[6] and in section 4.1.1. Comparing the results between the two strains it is possible to conclude that for the case of D-xylose culture use of the same media in pre-inoculum and cultivation condition led to poor *P. antarctica* grow in D-xylose, with low sugar consumption rate and maximum biomass.

In the case of *P. aphidis* grown on D-glucose, the MEL titre obtained using $NaNO_3$ was lower (0.74 g/l) that the one reported in the literature (2.80 g/l)[6] and in the previous section 4.1.1(1.28 g/L).

The MEL titre obtained for the same strain, using D-xylose and $NaNO_3$, was lower (0.45 g/L) that the one reported in the literature [6](1.2 g/L) and in the previous section 4.1.1(0.86 g/L). However, this maximum MEL titre (2.58 g/L) was obtained in a single measurement and is thus prone to error due to the indirect quantification process. Therefore is necessary to performed further assays to confirm this result. Excluding 2.58 g/l as maximum MEL titre (see Annex Figure 6-1), the highest value determined during the cultivation of *P. aphidis* in D-xylose and $NaNO_3$ was 0.45 g/l. When using NH_4Cl, MEL production was observed in both carbon sources (0.92 and 0.72 g/l for D-glucose and D-xylose, respectively), while no MEL was observed in NH_4Cl cultures as reported in section 4.1.2

For *P. antarctica* in D-glucose, the concentration of MEL, when using $NaNO_3$ as nitrogen source, was higher (4.46 g/L) than the reported in the previous section 4.1.1 (1.77 g/L). In the case of D-xylose the concentration obtain using $NaNO_3$, was similar (2.30 g/L) to the one reported in the literature in a pre inoculum with glucose and $NaNO_3$ (2.2 g/L) and slightly higher than the reported in the previous section 4.1.1 (1.88 g/L). When using NH_4Cl as nitrogen source, MEL titre in D-glucose and D-xylose cultures was higher from the reported in the previous section 4.1.1, but the maximum value does not represent a tendency so it is necessary to performed further assays to confirm this result. Excluding 3.68 g/l as maximum MEL titre (Annex Figure 6-1) obtain for culture in D-glucose and NH_4Cl, the highest value determined during cultivation of 1.48 g/l was obtained and this was higher than the reported in section 4.1.1.

Analyzing all the results, it was possible to conclude that the use of similar pre-inoculum and cultivation media did not favor sugar metabolism in *P. antarctica* cultured in D-xylose (regardless the nitrogen source) once the growth and sugar consumption were low.

Overall, the use similar inoculum and cultivation media did not improve sugar consumption rate neither MEL production.

Therefore, for the following experiments the inoculum was prepared in medium containing D-glucose and $NaNO_3$, as carbon and nitrogen sources, respectively, as previously used in MEL[6].

4.2 Enzymatic activities in *P. aphidis* and *P. antarctica*

Several enzyme activities with potential relevance for MEL production, through the consumption and production of NADPH, were assessed: D-glucose-6-phosphate dehydrogenase (EC. 1.1.1.49) (G6PDH), 6-phosphogluconate dehydrogenase (EC1.1.1.44) (6PGDH), malic enzyme (EC: 1.1.1.40) (ME), isocitrate dehydrogenase (EC 1.1.1.42) (IDH), xylose reductase (EC 1.1.1.307) (XR), nitrate reductase (EC 1.7.1.2) (NR) and nitrite reductase (EC 1.7.2.1) (NiR). These first enzymatic assays were performed with cell crude extracts of *P.antarctica* grown in D-glucose or D-xylose, using $NaNO_3$ as nitrogen source (Figure 4-5).

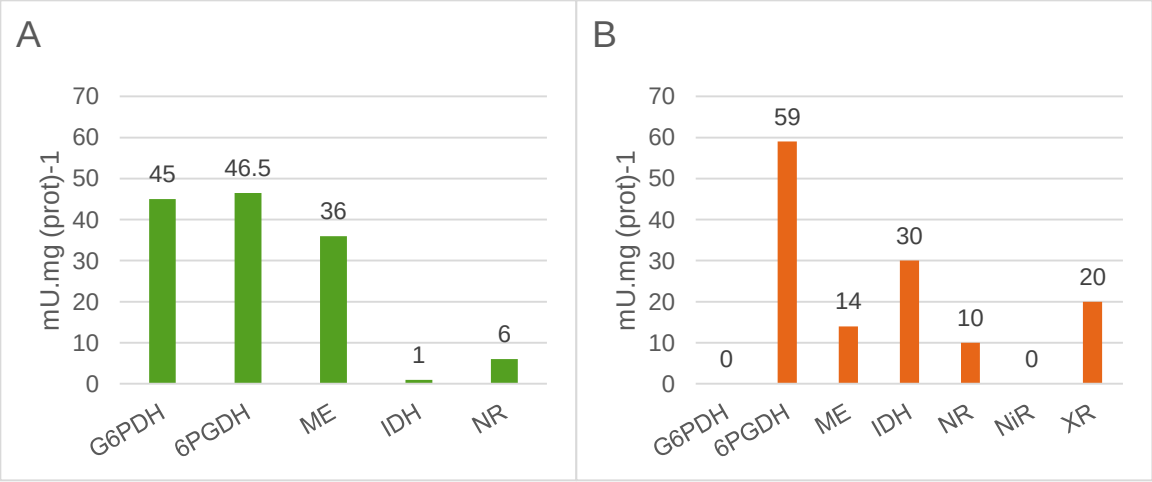


Figure 4-5: Enzymatic activity of, D-glucose-6-phosphate dehydrogenase (G6PDH), 6-phosphogluconate dehydrogenase (6PGDH) malic enzyme (ME), isocitrate dehydrogenase (IDH), nitrate reductase (NR), nitrite reductase (NiR) and xylose reducatse (XR) in crude extracts of P. antarctica grown in D-glucose (A) or D-xylose (B), using $NaNO_3$ as nitrogen source. Cells were harvested after 4 days incubation at 28ºC and 140 rpm.

The protocols to assess the enzyme activities potentially related to the consumption or production of NADPH during MEL biosynthesis were implemented. The results with *P. antarctica* cultures reveal that the selected enzymes have shown activity in D-glucose and/or D-xylose medium. In D-glucose, the enzymes showing higher activity were those of the oxidative PPP, G6PDH and 6PGDH, and the malic enzyme (ME). These enzymes and respective routes have been previously reported as the major sources of NADPH in

oleaginous yeasts when grown in D-glucose medium[21,51].Interestingly, xylose reductase was determined in D-xylose cultures, but not in D-glucose cultures. Enzymes associated to NADPH production, like 6PGDH and isocitrate dehydrogenase (IDH), presented a higher activity in D-xylose than in D-glucose, which may reflect the induction generated in D-xylose medium, in which NADPH is most probably needed for substrate consumption, as denoted by the activity of xylose reductase with NADPH.[29] However, to further confirm these results, the experimental assays need to be performed in biological duplicates and at different time points to understand the fluctuation of enzyme activities throughout the bioconversion process.

One of the most important routes to produce NADPH is the oxidative part of PPP, as denoted by the enzyme activities determined above and as previously reported for other yeasts[52,30], so the two respective enzymes, G6PDH and 6PGDH, were evaluated in *P. aphidis* and *P. antarctica* in D-glucose or D-xylose as carbon source and $NaNO_3$ or NH_4Cl as nitrogen source (Figure 4-6).

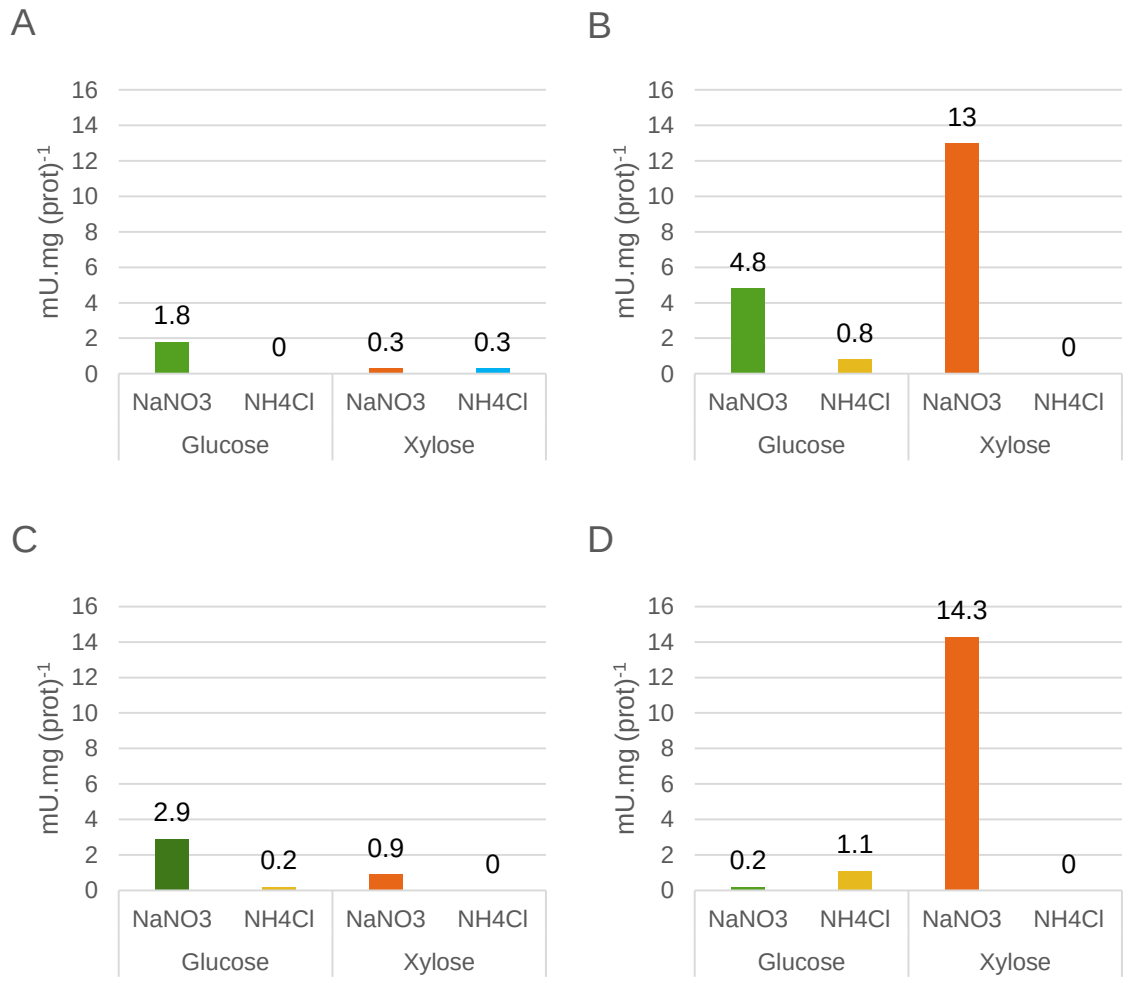


Figure 4-6: Enzymatic activity of D-glucose-6-phosphate dehydrogenase (G6PDH) (A, B) and 6-phosphogluconate dehydrogenase (6PGDH) (C, D) in crude extracts of P. aphidis (A, C) or P. antarctica (B, D) grown in D-glucose or D-xylose, using $NaNO_3$ or NH_4Cl as nitrogen source. Cells were harvested after 4 days incubation at 28ºC and 140 rpm.

In *P. aphidis* with D-xylose it was observed that oxidative PPP enzymes had very low activity (below 1 mU.mg (prot)$^{-1}$) using both nitrogen sources. While, in media with D-xylose and NH_4Cl growth was poor for both substrates, with $NaNO_3$ the oxidative PPP also seems to have no major relevance in D-xylose metabolism. On the other hand, in D-glucose, the activity of oxidative PPP enzymes is higher (between 1.5-3.0 mU.mg (prot)$^{-1}$), but only with $NaNO_3$ as nitrogen source, justifying the higher substrate consumption rate in this combination of carbon and nitrogen source.

In *P. antarctica* with D-glucose the G6PDH revealed higher enzymatic activity (4.8 mU.mg (prot)$^{-1}$) when the $NaNO_3$ is the nitrogen source than when using NH_4Cl (0.7 mU.mg (prot)$^{-1}$). Interestingly, in the case of *P. antarctica* with D-xylose and $NaNO_3$ the activity is significant higher (13 mU.mg (prot)$^{-1}$), revealing the response (induction of the oxidative PPP enzymes) to the requirement of NAPDH for both D-xylose and $NaNO_3$ assimilation. Accordingly, D-xylose consumption is even faster than D-glucose when $NaNO_3$ is used as nitrogen source (see Table 4-1) with similar MEL titres with both carbon sources.

Studies already performed with *Candida utilis,* show that G6PDH activity using D-glucose as carbon source was 3-fold higher in $NaNO_3$ than in NH_4Cl[30]. Moreover, the replacement of D-glucose by D-xylose further increases G6PDH in the presence of $NaNO_3$[30]. The 6PGDH revealed the same trend[30].

While *P. antarctica* seems to follow the same induction of oxidative PPP enzymes as *C. utilis*, in response to the higher NADPH requirements for D-xylose and $NaNO_3$ assimilation, *P. aphidis* does not respond with such efficiency to the need of NADPH, namely for the use of D-xylose, as denoted by the lower D-xylose consumption rate. Both G6PDH and 6PGDH are NADPH producers through the oxidative PPP[21,52], and the results revealed that this pathway is an important NADPH provider for MEL synthesis from D-glucose and D-xylose. In *P. aphidis* the production of NADPH is higher in D-glucose/$NaNO_3$, which can explain the higher MEL concentration (1.28 g/L) obtain when compared with that in D-xylose/$NaNO_3$ (Table 4-1). On the other hand, *P. antarctica* apparently has a higher carbon flux through oxidative PPP in D-xylose/$NaNO_3$, which can compensate the higher NADPH requirement for the utilization of these carbon and nitrogen sources. This result is consistent with the similar MEL concentration obtained under this condition (1.88 g/L) and under the one with D-glucose/$NaNO_3$ (Table 4-1).

In yeasts, it has been shown that the need for NADPH induces increased fluxes through the oxidative PPP[21] and it has been demonstrated that carbon recycling through this pathway is an effective way to generate NADPH during pentose utilization.[29]

4.3 Genomes and genes from *P. aphidis* and *P.antarctica*

In order to initiate the molecular study of *P. aphidis* and *P.antarctica* on the production of MEL from D-glucose and D-xylose, the correspondence between the strains available for this work and the ones with nucleotide and amino acid sequences deposited in databases, was established (Table 4-4). From the *Pseudozyma spp.* genomes available at NCBI database it was possible to identify the match with the yeast strains used in this study and in previous studies on the utilization of D-glucose and D-xylose for MEL production.

Table 4-4: Correspondence of P. aphidis and P. antarctica strains in different yeast culture collections

Species	*PYCC No*	*CBS No*	*Other reference (at NCBI)*
P. aphidis	5535^T	6821^T	DSM 70725^T
P. antarctica	5048^T	5955^T	JCM 10317^T

PYCC – Portuguese Yeast Culture Collection, Portugal

CBS – Centraalbureau voor Schimmelcultures (Fungal Biodiversity Centre), The Netherlands

JCM – Japanese Collection of Microorganisms, Japan

DSM - Deutsche Sammlung von Mikroorganismen, Germany

NCBI – National Center for Biotechnology Information, USA

Based on the Literature Review (Chapter 2), several genes were identify as relevant for the molecular study of the MEL biosynthesis from D-glucose and D-xylose, including those related with the assembly of the MEL building blocks (Table 4-5)[8], and those related with NADPH production and consumption for carbon (Table 4-6) or nitrogen (Table 4-7) metabolism.

Table 4-5: Genes of Pseudozyma spp. involved in MEL assembly and export

Gene	*Biological Function/Pathway*	*Enzymatic Reaction*
EMT1	MEL biosynthesis	Glycosyltransferase (mannose and erythritol assembling)
MAC1 and MAC2	MEL biosynthesis	Acyltransferase (mannosylerythritol acylation)
MAT1	MEL biosynthesis	Acetyltransferase (MEL acetylation)
MMF1	MEL Export	Major facilitator

Table 4-6: Genes of Pseudozyma spp. involved in carbon metabolism and NADP(H)-dependent

Gene	*Biological Function/Pathway*	*Enzymatic Reaction (cofactor dependence)*
ZWF1	Oxidative PPP	Glucose-6-phosphate dehydrogenase ($NADP^+$-dependent)
GND1	Oxidative PPP	6-phosphogluconate dehydrogenase $NADP^+$-dependent
XYL1	Xylose catabolism	Xylose reductase (NADPH-dependent reduction of xylose to xylitol)
IDP1	TCA	Isocitrate dehydrogenase ($NADP^+$-dependent isocitrate oxidative decarboxylation
ME1	Fatty acid biosynthesis/ Citrate Shuttle	Malic enzyme $NADP^+$-dependent (S)-malate oxidative decarboxylation

Table 4-7: Genes of Pseudzyma spp. involved in nitrogen metabolism and NADP(H)-dependent

Gene	*Biological Function/Pathway*	*Enzymatic Reaction (cofactor dependence)*
YNR1	Nitrate assimilation	Nitrate reductase (NADPH-dependent reduction of nitrate to nitrite)
YNI1	Nitrate/Nitrite assimilation	Nitrite reductase (NADPH-dependent reduction of nitrite to ammonia)
GLT1	Glutamate and nitrogen metabolism	Glutamate synthase ($NAD(P)^+$-dependent synthesis of glutamate from glutamine and alpha-ketoglutarate)

After the identification of putative relevant genes according to their biological function, the bioinformatics platform at NCBI[10] was used to directly identify the respective nucleotide and amino acids in sequenced *Pseudozyma spp.* genomes. Whenever necessary, the BLAST tool at NCBI was used with annotated sequences from the close related fungus *U. maydis*. All the genes and proteins identify in *P. aphidis* and *P. antarctica* genomes and their respective accession numbers are presented in Table 6-1and Table 6-2 (Annexes). Examples of blastp results using *U. maydis* ZWF1, GND1 and EMT1 are presented in Table 4-8

Table 4-8: Identification of protein sequences of ZWF1, GND1, and EMT1 from P.aphidis and P.antarctica through Blastp (NCBI) results using the respective sequence from U. maydis.

Gene	Description	Max score	Total Score	Query Score	E value	Ident	Accession
ZWF1	*P. antarctica*	890	890	100%	0.0	85%	GAK68546.1
	P. aphidis	891	891	100%	0.0	85%	ETS65312.1
GND1	*P. antarctica*	908	908	100%	0.0	90%	GAK64517.1
	P. aphidis	900	900	100%	0.0	90%	ETS64853.1
EMT1	*P.antarctica*	890	890	100%	0.0	85%	GAK68006.1
	P. aphidis	891	891	100%	0.0	85%	ETS61959.1

.

.

4.4 Design of primers for Real-Time RT-PCR using the selected nucleotide sequences from *P. aphidis* and *P. antarctica*

After selecting, in *P. aphidis* and *P. antarctica*, the nucleotide sequences of the genes of interest, it was possible to pair and align them and identify conserved regions between the two DNA (or mRNA) sequences in order to be able to design the gene-specific primers for Real-Time RT-PCR.

In some cases, the comparison between the DNA and mRNA sequences available in NCBI[10] database revealed the presence of introns, as illustrated for EMT1 (Figure 4-7).

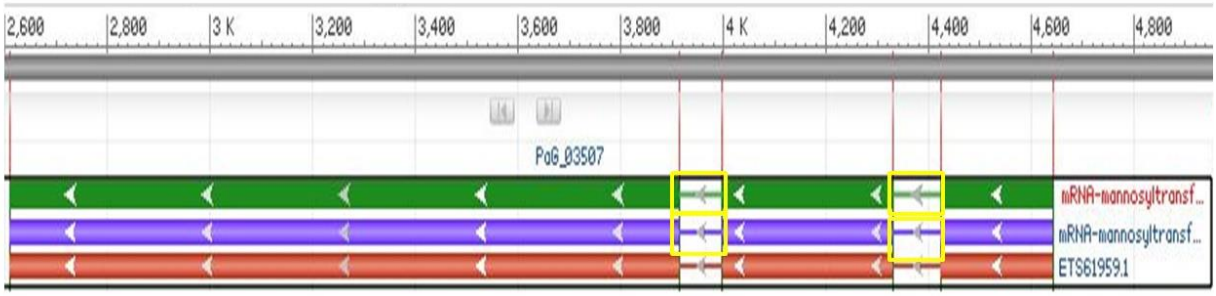


Figure 4-7: EMT1 gene in P. aphidis. The green line correspond to pre RNA (1600bp), the blue line correspond to the mRNA (1600bp) complement, the red line to protein and the grey line to the DNA (2000bp).

EMT1 gene, of 2000 bp, apparently originates a pre-mRNA with two introns, leading to a mature mRNA of 1600 bp. This analysis of the sequences allowed the design of primers in conserved and introns-free regions.

Figure 4-8 illustrates the alignment between the two EMT1 sequences from *P. aphidis* and *P. antarctica* and the conserve regions selected to design gene-specific primer pair with the criteria described in the Materials and Methods Chapter 3. The procedure described was followed for the design of gene-specific primer pairs for all the selected genes.

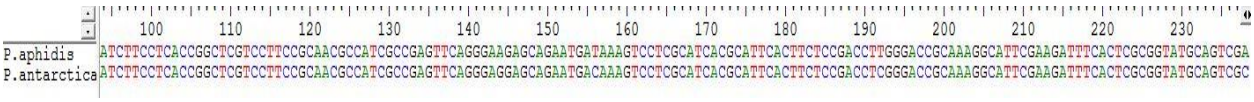


Figure 4-8: The alignment betwwen P.antarctica and P.aphidis illustrate conserved regions used for the primers desing..

The criteria defined for primer pair design, namely the PCR product length, amplification region and primer length, GC content and 3'-dimer-free, were necessary to ensure that the primers used in Real-Time RT-PCR have similar properties, including the same amplification efficiency.

The complete list of gene-specific primer pairs designed is presented in Table 4-9.

Table 4-9: List of gene-specific primer pairs designed for Real-Time RT-PCR

Gene	*Primer Forward (5'-3')*	*Primer Reverse (5'-3')*
ACT1	TACGGACACTCGTCTATCA	CGATGAGGAAGCGCCGTTCT
TAF10	GCGACAAGAACCGCAAGCTG	GTCGGGCGTGATGACGGTAT
EMT1	AAGTCCTCGCATCACGCATT	GCGGAACTCTTGCTCGGTAA
MAT1	TTCATCCGATCGCAGCCAAG	GCGTCGCTCCAGACGGTAG
MAC1	CGACGTTCCGAGTGCATCAT	GAAGCACGCGGTACGTGAAT
MAC2	GAAGGTGCTCAACTACACAG	AGCGCATGATCGAGGTCAA
ME1	CTCGCTGACCATCGCATCAT	TGCTTCTCGAGCGGTGTGGA
ZWF	CAGCCGCCTCTTCTACATGG	TCGATGCGGAAGGTCTCCTC
GND1	TCACTGTCTGCGCGTACAAC	GCCTTGACGAGGAGGATCAT
YNR1	CACCGCGAGATCGAGTTCAG	TCGAGAATGTCGGCGAGGTA
YNI1	TCGGTCAAGGCGCTCACTTA	CCTTCTTGGCTCCGATGATG
MMF1	GGACGCAAGAACTCGGTGAA	AGCGTGGTGAGATCGGACAT
XYL1	ACGAGCCGAAGCTGCCATCA	TCGAGCAGCGACGTCTTGGT
XYL1a	ATCGGCCTCGGCACTTGGCT	GAGGTAGAGGTCGAGGTAGTC
IDP1	TTCGGCCTCAAGAAGATGTG	AAGTTCTGGACTCGGTACTG
GLT1	GGTCTCCAGAACTGGTTCTA	GAAGGTGTTGGTGGAGAATC

4.5 Testing primer pairs with DNA and cDNA templates

The primer pairs were tested with genomic DNA from the two different strains, *P. aphidis* and *P. antarctica*, in order to evaluate if the designed primers were suitable to amplify the gene region envisaged in both strains. Moreover, using cDNA, produced from total RNA, it was possible to confirm the presence of introns, comparing the size of fragments between the PCR products generated from genomic DNA and from cDNA for each gene. The primers test with cDNA was performed starting from total RNA from *P.antarctica* cultivated in 40 g/l D-glucose and 3 g/l $NaNO_3$ for 4 days. The design of the primer pair for EMT1 envisaged the amplification of fragments containing 200-300 bp when amplified from genomic DNA and around 190 bp when amplified from cDNA, thus has allowing the identification of one of the EMT1 introns.

The sizes of these amplification products were consistent with the predicted sizes. All the primer pairs evaluated successfully amplified the regions envisaged with the genomic DNA from *P. antarctica* (Figure 4-9. In the case of the MAC1, MAC2, GLT1, MDH, GND1 and XYL1 genes it is not possible to observe amplification in cDNA tested. One of the possible explanations for this fact is that, under the tested conditions, these genes are not expressed or have a very low expression level. However, some of these genes (MAC1, MDH and GND1) were still evaluated, under the same conditions, in terms of expression level through Real-Time RT-PCR. In the case of EMT1, it is possible to conclude that this gene has introns, because the size of the fragment in DNA is about 300bp and the cDNA is about 200bp. The MAT1 primer pair amplified two fragments from cDNA, one with 200bp, which confirm the information present in GenBank, and another fragment with approx. 100 bp, which may suggest alternative splicing in that region (non-retained intron) or unspecific amplification, probably resulting from primer dimers. The gene-specific primers for MMF1, IDP, ZWF1, YNR1, YNI1 and XYL1a amplified similar fragments (approx. 200 bp) from DNA and cDNA, which confirm the information obtained from the databases.

Based on the PCR results obtained with gene-specific primers and cDNA from *P. antarctica* the following genes were selected for Real-Time RT-PCR experiments: EMT1, MAC1, MAT1, MMF1, GND1, ZWF1, MDH,IDP, YNR1, YNI1 and XYL1a.

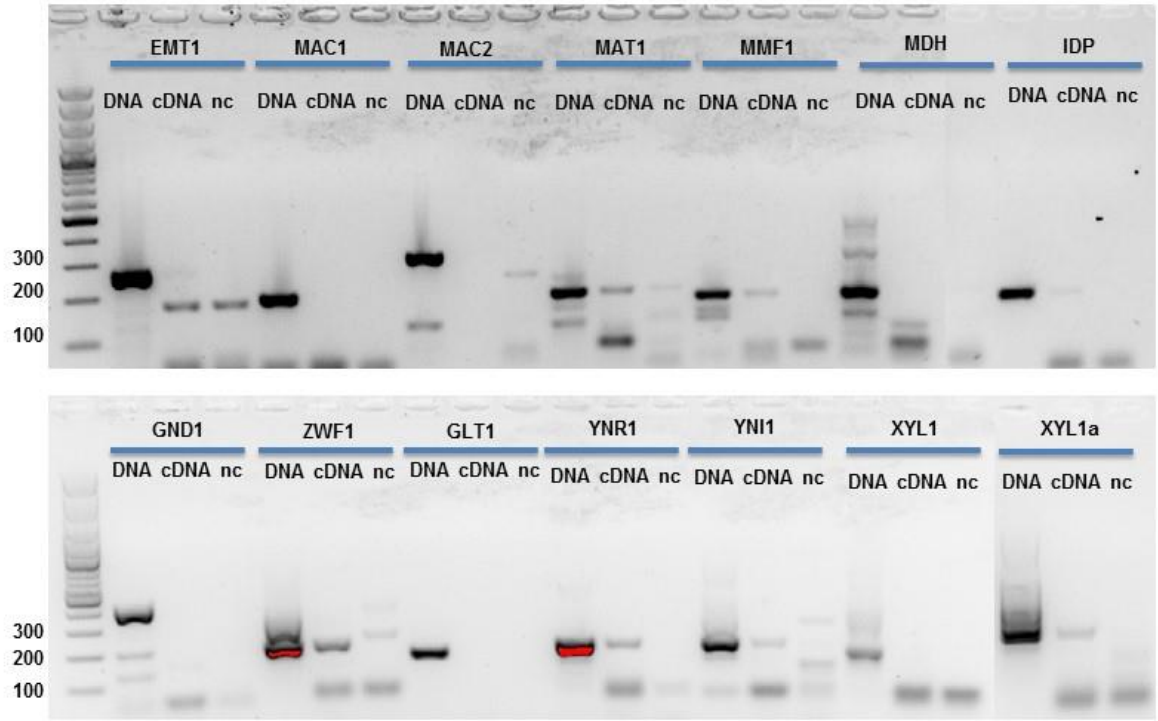


Figure 4-9: Gel electrophoresis (70V and 400mA) of PCR products with the gene-specific primers selected using as template DNA or cDNA (glucose/nitrate, day 4) from P. antarctica. Molecular weight estimated by comparison with Gene Ruler 100bp DNA Ladder Plus.

This approach, apart from allowing the testing of gene-specific primers with the amplification of DNA fragments, was also important to confirm that the RNA extracted has quality to be used in Real-Time RT-PCR experiments. However, some of the negative controls (water instead of DNA or cDNA) apparently generated DNA amplification. This issue requires further attention. Also, some unspecified amplification were observed with some of the primers, which can be related with primer dimers, as stated above.

4.6 Transcript analysis of *P. aphidis* and *P. antarctica* in D-glucose and D-xylose

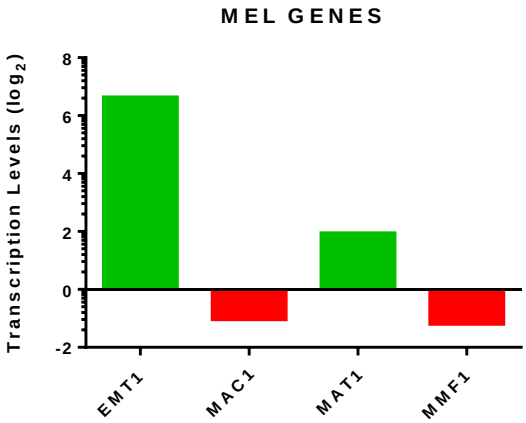
Total RNA from *P. aphidis* and *P. antarctica* was extracted from 4-days cultures in D-glucose or D-xylose, using $NaNO_3$ as nitrogen source. The expression of genes involved in MEL assembly and export, and in NADPH production and consumption were studied. The results obtained were triplicates from a single biological experiment. The analysis the results obtained were represented as the transcription level in D-xylose in function of that of D-glucose, using ACT1 as reference gene. The differences between the two reference genes used (ACT1 and TAF10) were minimal (data not shown).

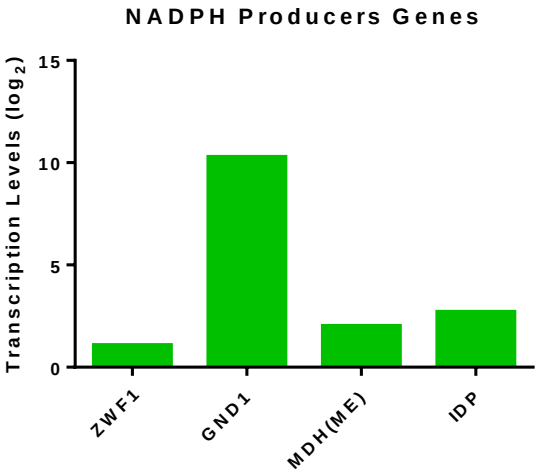
In *P. antarctica* the genes involved in production and consumption of NADPH were overexpress in D- xylose when compared with D-glucose. The GND1 gene revealed to be much higher expressed in D-xylose than in D-glucose and was seen in enzymatic activity where .14.3 mU.mg (prot)$^{-1}$ was obtain in same conditions The gene YNR1 was more express than the other genes with the same function (consuming NADPH). Also XYL1a is overexpress in D- xylose, which was expected since may be required for the carbon source assimilation.

Overall comparing the genes that are involved in NADPH production and those who were involved in NADPH consumption, it was possible conclude both group of genes are overexpress in *P.antarctica* when the carbon source is D- xylose.

The MEL genes, those related with the MEL assembly and export (Table 4-5), were expressed differently in *P. antarctica*. The EMT1 and MAT1 were overexpressed in D-xylose compared to D-glucose, while the MAC1 and MMF1 were underexpressed in D-xylose compared to D-glucose. EMT1 is highly expressed compared with other MEL genes.

Further biological replicates are required in order to confirm, or not, these observations.





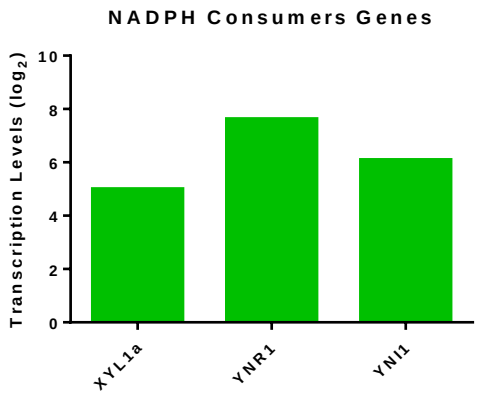
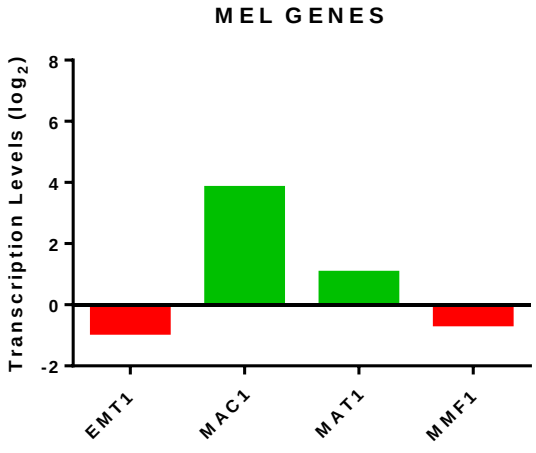


Figure 4-10: Relative transcript levels of EMT1, MAC1, MAT1, MMF1, ZWF1, GND1, MDH(ME), IDP, XYL1a, YNR1 and YNI1 in D-xylose in function of D-glucose. The assessment was performed with P. antarctica cultivated during 4 days at 28ºC and 140 rpm.

In *P. aphidis* the genes involved in NADPH consumption were not assessed. However, it was possible to evaluate the expression of MEL genes and of the genes involved in the production of NADPH. The gene GND1 were overexpressed in D-xylose when compared with D-glucose. On contrary, the genes ZWF1 and MDH(ME), also involved in NADPH production, were underexpressed in D- xylose.

The MEL genes were expressed differently in *P. aphidis*. The MAC1 and MAT1 were overexpressed in D-xylose comparing with D-glucose, while the EMT1 and MMF1 were underexpressed in D-xylose comparing with D-glucose. The difference in the expression levels between D-xylose and D-glucose -were smaller than the observed for *P. antarctica*.

Further biological replicates are required in order to confirm, or not, these observations.



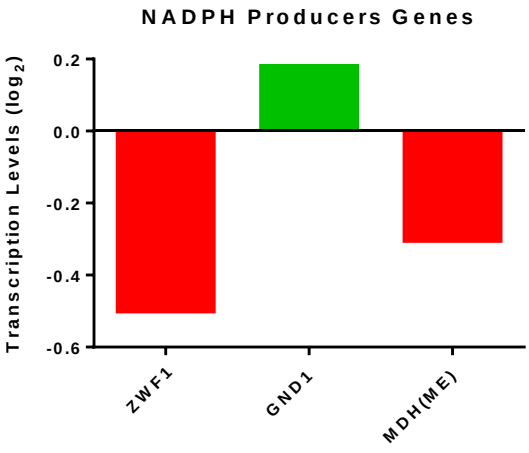


Figure 4-11: Relative transcript levels of EMT1, MAC1, MAT1, MMF1, GND1, ZWF1, MDH (ME), in D-xylose in function of D-glucose. The assessment was performed with P. aphidis cultivated during 4 days at 28ºC and 140 rpm.

The results of genes transcription levels can be related with the MEL production capacity of both strains using D-glucose or D-xylose as carbon sources. In *P. antarctica* the EMT1 and MAT1 were overexpressed in D-xylose comparing with D-glucose, which support the results obtained on MEL production, which is slightly higher in D-xylose (1.88 g/L) than in D-glucose (1.7 g/L). The EMT1 is a glycosyltransferase, essential for MEL biosynthesis[12]. Besides that, the genes related with the production of NADPH, were overexpressed in D-xylose compared to D- glucose. Although the use of pentoses demands NADPH, the overexpression of NADPH producing genes can balance that need as well as can contribute to fatty acid biosynthesis, ultimately to be incorporated in MEL after carbon chain-shortening[6,29,21,18].

On contrary, in *P. aphidis,* the differences in expression levels between conditions (D-glucose or D-xylose) were not significant. These results can explain the higher MEL production from D-glucose (1.28 g/L) than from D-xylose (0.86 g/L). Besides that, the genes related with the production of NADPH were underexpressed in D-xylose comparing with D- glucose (except the GND1), which can indicate that less NADPH was available for fatty acid production and consequent MEL biosynthesis. The results obtained through Real-Time RT-PCR were consistent with the previous reported, where *P. antarctica* PYCC 5048^{T} proved to be the best MEL producer.[6]

The transcription levels of the genes ZWF1 and GND1 can be related with the enzymatic activity measured in both strains in the same conditions. In *P.antarctica* the results of RT-PCR showed that the transcription levels of genes related with the production of NADPH were overexpress in D-xylose comparing with D- glucose, which was in the agreement with the results obtain in enzymatic activities. In fact, both enzymes presented a higher activity in D-xylose than in D-glucose (Figure 4-5 and Figure 4-6). For example, D-glucose 6-phosphate dehydrogenase (coded by ZWF1) revealed significant higher (3-fold) activity in D-xylose, 13.0 mU.mg (prot)$^{-1}$, than in D-glucose 4.8 mU.mg (prot)$^{-1}$, with the corresponding gene expression being 6-fold higher expressed in D-xylose. The results from both Real-Time RT-PCR and enzymatic activities were consistent, and support the previous observations where *P. antarctica* is consider the best MEL producer using both carbon sources (Table 4-1), once is able to produce the NADPH necessary for the MEL production.

In *P. aphidis,* the enzymatic activity of D-glucose 6-phosphate dehydrogenase (coded by ZWF1) was lower (0.3 mU.mg (prot)$^{-1}$) in D-xylose than in D- glucose (1.8 mU.mg (prot)$^{-1}$), which is in concordance with the results obtained with Real-Time RT-PCR. Since the results were obtained from a single experience, further replicates are required in order to confirm, or not, these observations.

Comparing the two strains, *P. aphidis* and *P. antarctica* in terms of transcription levels of genes it was possible to conclude that *P. antarctica*, but not *P. aphidis*, has the ability to direct carbon flux through the oxidative PPP, to provide the necessary NADPH, both for carbon assimilation and MEL production (fatty

acid biosynthesis)[5]. Also, *P. antarctica* overexpresses EMT1 gene in D-xylose, unlike *P. aphidis*, being this gene essential for MEL biosynthesis.

5 Conclusion and Future Perspectives

This work presents a set of new information that allow a better understanding of MEL production process, as well as possible metabolic pathways involved in its production. This study made use of several approaches (physiology, biochemistry and molecular biology) to reveal the key metabolic features of *Pseudozyma spp*. in the production of MEL from D-glucose and D-xylose.

The use of NH_4Cl instead of $NaNO_3$ as nitrogen source was expected to contribute to increase MEL production, since the spending of NADPH necessary to assimilate nitrate (Eq 6) is not necessary when ammonium was used (Eq 7). However, the study maintained $NaNO_3$ as the best nitrogen source for MEL production, which suggests that the yeast has mechanisms to compensate the greater need in NADPH in the presence of nitrate.

This work confirmed some of the results from previous studies[6] on the MEL production from pentoses by *P. aphidis* and *P .antarctica* , with the later confirmed as the best MEL producer under the conditions testes, with MEL titres of 1.7-1.9 g/L. Knowing that for the MEL production one of the building blocks is fatty acids[6,31], and for that biosynthesis is necessary NADPH, the production and consumption of NADPH was studied. First, enzyme activities involved in these reactions were assessed. The oxidative PPP seems to play a major role on NADPH production, mainly in *P. antarctica*, through the high activity of glucose-6-phosphate dehydrogenase (G6PDH, EC. 1.1.1.49) and 6-phosphogluconate dehydrogenase (6PGDH, EC1.1.1.44), which can produce NADPH in both D-glucose and D-xylose cultures. In *P. aphidis* the production of NADPH is higher in D-glucose/$NaNO_3$ medium, but in *P. antarctica* a higher enzymatic activity was found in D- xylose/$NaNO_3$ medium. The results obtained are consistent but require further replicates.

At last, transcript analysis was used to evaluate the expression of genes involved in MEL assembly and export, and genes involved in production and consumption of NADPH. Confirming the results obtained with enzymatic activities, *P. antarctica* revealed overexpression of genes involved in NADPH production when D-xylose was used as substrate instead of D-glucose, which could indicated that in presence of D-xylose *P. antarctica* was the ability to shift carbon flux through the oxidative PPP, for the production of NADPH[5]. NADPH is supplied for the metabolism of the carbon source and also for fatty acid biosynthesis towards MEL production. Moreover, *P. antarctica* overexpressed EMT1 in D-xylose, further contributing for MEL production.

This work allowed the development of tools to understand the metabolic pathways involved in MEL production and provided the first molecular insights on yeast metabolism during MEL production from D-glucose and xylose. The planning of systems biology approaches will now benefit from the information gather in this thesis. These approaches can include transcriptomics (e.g. RNAseq) and metabolomics (^{13}C-

labeled carbon sources followed by NMR). RNAseq was already applied to evaluate the transcriptomes of *P. aphidis* in the presence or absence of soybean oil.[8] The results in RNA-Seq, showed the strong inducing effect of the hydrophobic carbon source on the expression of the MEL cluster genes with exception of the acetyltransferase gene PaGMAT1. A similar approach can be followed using *P. aphidis* and *P. antarctica* cultivated in D-glucose and D-xylose. Another approach suggested is metabolomics analysis, for example by ^{13}C-NMR to investigate *in vivo* of carbon flux distribution through metabolic pathways. The NMR allow the quantification of fluxes using ^{13}C-labeled substrates[29]. In this specific case the metabolism of ^{13}C-D-glucose and ^{13}C-D-xylose can be followed and further understand the complexity of metabolic fluxes for MEL production.

6 Annexes

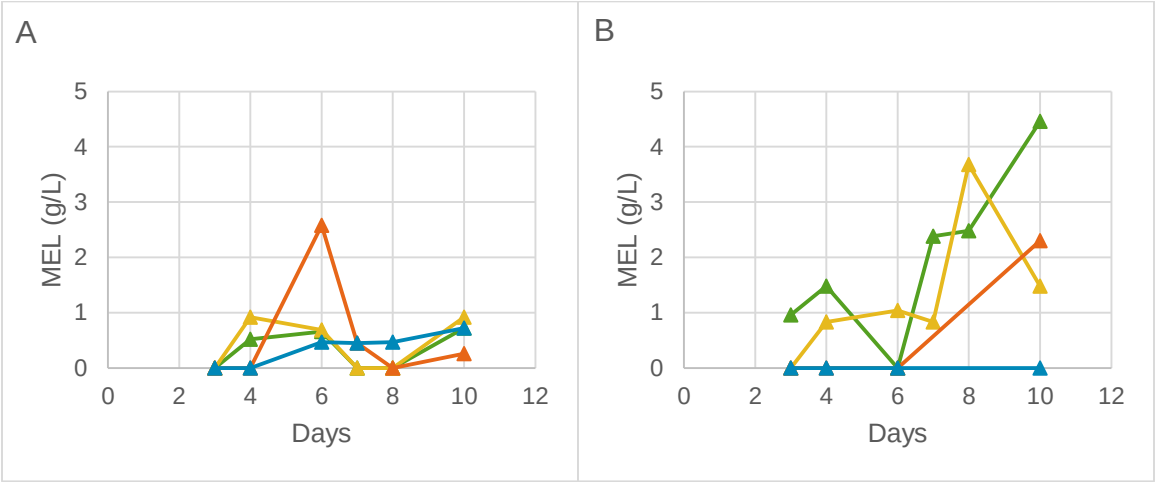


Figure 6-1: MEL (▲) production with P. aphidis (A) and P. antarctica (B), using D glucose/ $NaNO_3$ (▲), D- glucose /NH_4Cl (▲),D xylose / $NaNO_3$ (▲) and D xylose /NH_4Cl (▲), at 28 ºC and 140 rpm.

Table 6-1: Blastp results using protein sequences of XYL1, IDP1, MDH, GLT1, YNR1 and YNI1 from U.maydis with the identification of P.aphidis and P.antarctica respective sequences

Genes	Description	Max score	Total Score	Query Score	E value	Ident	Accession
XYL1	*Pseudozyma antarctica T-34*	706	706	100%	0.0	100%	ETS65124.1
	Pseudozyma aphidis DSM 70725	608	608	89%	0.0	97%	GAC75481.1
IDP1	*Pseudozyma antarctica*	571	571	91%	0.0	98%	GAK67473.1
	Pseudozyma aphidis DSM 70725	597	597	96%	0.0	99%	ETS60339.1
MDH	*Pseudozyma antarctica*	1770	1770	100%	0.0	100%	GAK64078.1
	Pseudozyma aphidis DSM 70725	1407	1660	100%	0.0	95%	ETS60818.1
GLT1	*Pseudozyma antarctica*	4258	4258	99%	0.0	95%	GAK66158.1
	Pseudozyma aphidis DSM 70725	4250	4250	99%	0.0	94%	ETS64084.1
YNR1	*Pseudozyma antarctica*	1319	1319	96%	0.0	67%	GAK66154.1
	Pseudozyma aphidis DSM 70725	1306	1306	96%	0.0	67%	ETS64087.1
YNI1	*Pseudozyma antarctica*	1858	1858	99%	0.0	78%	GAK66155.1
	Pseudozyma aphidis DSM 70725	1874	1874	100%	0.0	78%	ETS64086.1

Table 6-2: Blastp results using protein sequences of MAC1, MAC2, MAT1, and MMF1 from U.maydis with the identification of P.aphidis and P.antarctica respective sequences.

Genes	Description	Max score	Total Score	Query Score	E value	Ident	Accession
MAC1	*Pseudozyma antarctica*	683	683	94%	0.0	59%	GAK68003.1
	Pseudozyma aphidis DSM 70725	683	683	90%	0.0	60%	ETS61961.1
MAC2	*Pseudozyma antarctica*	527	587	94%	5e-178	52%	GAK68004.1
	Pseudozyma aphidis DSM 70725	536	594	91%	0.0	54%	ETS61960.1
MAT1	*Pseudozyma antarctica*	577	577	98%	0.0	53%	GAK68001.1
	Pseudozyma aphidis DSM 70725	570	570	98%	0.0	54%	ETS61963.1
MMF1	*Pseudozyma antarctica*	794	794	89%	0.0	72%	ETS61962.1
	Pseudozyma aphidis DSM 70725	786	786	83%	0.0	78%	GAK68002.1

7 References

1. Naik, S. N., Goud, V. V., Rout, P. K. & Dalai, A. K. Production of first and second generation biofuels: A comprehensive review. *Renew. Sustain. Energy Rev.* **14,** 578–597 (2010).
2. Axelsson, L. *et al.* Perspective: Jatropha cultivation in southern India: Assessing farmers' experiences. *Biofuels, Bioprod. Biorefining* **6,** 246–256 (2012).
3. McCormick, K. & Kautto, N. The Bioeconomy in Europe: An Overview. *Sustainability* **5,** 2589–2608 (2013).
4. Liu, S., Abrahamson, L. P. & Scott, G. M. Biorefinery: Ensuring biomass as a sustainable renewable source of chemicals, materials, and energy. *Biomass and Bioenergy* **39,** 1–4 (2012).
5. Ricardo, N. & Faria, T. Yeast mannosylerythritol lipids from lignocellulose : a novel strategy for the production of jet biofuel. (INSTITUTO SUPERIOR TÉCNICO).
6. Faria, N. T. *et al.* Production of glycolipid biosurfactants, mannosylerythritol lipids, from pentoses and d-glucose/d-xylose mixtures by Pseudozyma yeast strains. *Process Biochem.* **49,** 1790–1799 (2014).
7. Gírio, F. M. *et al.* Hemicelluloses for fuel ethanol: A review. *Bioresour. Technol.* **101,** 4775–4800 (2010).
8. Günther, M. *et al.* The transcriptomic profile of Pseudozyma aphidis during production of mannosylerythritol lipids. *Appl. Microbiol. Biotechnol.* **99,** 1375–1388 (2015).
9. Ratledge, C. Fatty acid biosynthesis in microorganisms being used for Single Cell Oil production. *Biochimie* **86,** 807–15 (2004).
10. http://www.ncbi.nlm.nih.gov/.
11. http://www.genome.jp/kegg/No Title.
12. Hewald, S. *et al.* Identification of a gene cluster for biosynthesis of mannosylerythritol lipids in the basidiomycetous fungus Ustilago maydis. *Appl. Environ. Microbiol.* **72,** 5469–77 (2006).
13. Food, I. *Bio-based economy in Europe : state of play and future potential - Part 2*.
14. http://www.cen.eu/work/areas/chemical/biobased/Pages/default.aspx.
15. Pittman, C. U. & Steele, P. H. Pyrolysis of Wood / Biomass for Bio-Oil : A Critical Review. (2006). doi:10.1021/ef0502397
16. Arutchelvi, J. I., Bhaduri, S., Uppara, P. V. & Doble, M. Mannosylerythritol lipids: a review. *J. Ind. Microbiol. Biotechnol.* **35,** 1559–70 (2008).
17. Morita, T., Konishi, M., Fukuoka, T., Imura, T. & Kitamoto, D. Physiological differences in the formation of the glycolipid biosurfactants, mannosylerythritol lipids, between Pseudozyma antarctica and Pseudozyma aphidis. *Appl. Microbiol. Biotechnol.* 307–315 (2007). doi:10.1007/s00253-006-0672-3
18. Petit, T., Gancedo, C., Flores, C. & Rodr, C. Carbohydrate and energy-yielding metabolism in non-conventional. *FEMS Microbiol.* **24,** (2000).

19. Nelson, D. & Cox, M. *Lehninger Principles of Biochemistry*. (W. H. Freeman Publisher, 2011).
20. Voet, D. & Voet, J. *Biochemistry*. *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki* (John Wiley & Sons, 2004).
21. Bruinenberg, P. M., Van Dijken, J. P. & Scheffers, W. a. A Theoretical Analysis of NADPH Production and Consumption in Yeasts. *Microbiology* **129,** 953–964 (1983).
22. Kruger, N. J. The oxidative pentose phosphate pathway : structure and organisation . Curr Opin Plant Biol The oxidative pentose phosphate pathway : structure and organisation. 236–246 (2003). doi:10.1016/S1369-5266(03)00039-6
23. Nielsen, J. Synthetic Biology for Engineering Acetyl Coenzyme A Metabolism in Yeast. **5,** 14–16 (2014).
24. Morita, T. *et al.* Genome and transcriptome analysis of the basidiomycetous yeast Pseudozyma antarctica producing extracellular glycolipids, mannosylerythritol lipids. *PLoS One* **9,** e86490 (2014).
25. Ratledge, C. The role of malic enzyme as the provider of NADPH in oleaginous microorganisms: A reappraisal and unsolved problems. *Biotechnol. Lett.* **36,** 1557–1568 (2014).
26. Hiltunen, J. K. *et al.* The biochemistry of peroxisomal L -oxidation in the yeast Saccharomyces cerevisiae. **27,** (2003).
27. Poirier, Y., Antonenkov, V. D., Glumoff, T. & Hiltunen, J. K. Peroxisomal B-oxidation-A metabolic pathway with multiple functions. *Biochim. Biophys. Acta - Mol. Cell Res.* **1763,** 1413–1426 (2006).
28. Yanagishita, H., Haraya, K. & Kitamoto, H. K. Contribution of a chain-shortening pathway to the biosynthesis of the fatty acids of mannosylerythritol lipid (biosurfactant) in the yeast Candida antarctica : Effect of ℳ -oxidation inhibitors on biosurfactant synthesis. *Biotechnol. Lett.* **20,** 813–818 (1998).
29. Fonseca, C. *et al.* Use of in vivo 13C nuclear magnetic resonance spectroscopy to elucidate L-arabinose metabolism in yeasts. *Appl. Environ. Microbiol.* **74,** 1845–55 (2008).
30. Bruinenberg, P. M., Van Dijken, J. P. and S. A. An Enzymic Analysis of NADPH Production and Consumption in. *J. Gen. Microbiol.* 965–971 (1983).
31. Wakil, S. J., Stoops, J. K. & Joshi, V. C. Fatty acid synthesis and its regulation. *Annual review of biochemistry* **52,** 537–579 (1983).
32. Chuang, H.-Y. et all. A Decade of Systems Biology. *Cell Dev Biol* (2010). doi:10.1146/annurev-cellbio-100109-104122.A
33. Werner, T. Next generation sequencing in functional genomics. **11,** (2010).
34. https://sysbio.med.harvard.edu/.
35. Wilhelm, B. T. & Landry, J.-R. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* **48,** 249–57 (2009).
36. Issaq, H. & Veenstra, T. Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives. *Biotechniques* **44,** 697–8, 700 (2008).
37. Ugurbil, K., Brown, T. R., den Hollander, J. a., Glynn, P. & Shulman, R. G. High-resolution 13C

nuclear magnetic resonance studies of glucose metabolism in Escherichia coli. *Proc. Natl. Acad. Sci.* **75,** 3742–3746 (1978).

38. Ogura, T. & Sakamoto, Y. *Application of Metabolomics Techniques using LC / MS and GC / MS Profi ling Analysis of Green Tea Leaves*. **10,**
39. Pongsuwan, W. *et al.* Prediction of Japanese green tea ranking by gas chromatography/mass spectrometry-based hydrophilic metabolite fingerprinting. *J. Agric. Food Chem.* **55,** 231–236 (2007).
40. Wong, M. & Medrano, J. Real-time PCR for mRNA quantitation. *Biotechniques* **39,** 75–85 (2005).
41. Bisswanger, H. Enzyme assays. *Perspect. Sci.* **1,** 41–55 (2014).
42. Goddard, J.-P. & Reymond, J.-L. Recent advances in enzyme assays. *Trends Biotechnol.* **22,** 363–70 (2004).
43. Singh, R., Lemire, J., Mailloux, R. J. & Appanna, V. D. A novel strategy involved anti-oxidative defense: The conversion of NADH into NADPH by a metabolic network. *PLoS One* **3,** (2008).
44. Rau, U. *et al.* Formation and analysis of mannosylerythritol lipids secreted by Pseudozyma aphidis. *Appl. Microbiol. Biotechnol.* **66,** 551–9 (2005).
45. Faria, N. *et al.* Conversion of cellulosic materials into glycolipid biosurfactants, mannosylerythritol lipids, by Pseudozyma spp. under SHF and SSF processes. *Microb. Cell Fact.* **13,** 155 (2014).
46. Manual, I. Instruction Manual. (1921).
47. Scientific, T. Thermo Scientific Maxima H Minus First Strand cDNA Synthesis Kit. 1–6
48. Hall, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41,** 95–98 (1999).
49. Zilinskiene, J. GeneRuler_100bp_DNALadder. 2–3
50. T, P. S. G. R. Power SYBR® Green RNA-to-CT™1-Step Kit Protocol (PN 4391003C).
51. Ratledge, C. The role of malic enzyme as the provider of NADPH in oleaginous microorganisms: A reappraisal and unsolved problems. *Biotechnol. Lett.* **36,** 1557–1568 (2014).
52. Wasylenko, T. M., Ahn, W. S. & Stephanopoulos, G. The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in Yarrowia lipolytica. *Metab. Eng.* **30,** 27–39 (2015).