Master's thesis

2019

Margit Ertresvåg Jakobsen

NTNU
Norwegian University of
Science and Technology
Faculty of Natural Sciences
Biotechnology and Food Science
Department of

Margit Ertresvåg Jakobsen

ATP-Dependent Citrate Lyase and Glycerol 3-Phosphate Dehydrogenase in *Aurantiochytrium* T66: An Attempted *In Vivo* Analysis and Heterologous Expression

May 2019

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU

Norwegian University of
Science and Technology

ATP-Dependent Citrate Lyase and Glycerol 3-Phosphate Dehydrogenase in *Aurantiochytrium* T66: An Attempted *In Vivo* Analysis and Heterologous Expression

Margit Ertresvåg Jakobsen

MBIOT5
Submission date: May 2019
Supervisor: Helga Ertesvåg
Campus: Trondheim

Norwegian University of Science and Technology
Department of Biotechnology and Food Science

Preface

For the last 5 years, I have been a student of biotechnology at the Norwegian University of Science and Technology. I have learned about the rules of chemistry, and about the guidelines usually applying to biological systems. I have realized that understanding a few basic principles unlocks the door to vast new information. And I have understood that no matter how much you study, nature will always find new ways to surprise you. All you can do as a scientist, is to learn more about nature and its marvelous tricks and apply this knowledge to achieve your goals.

This thesis is the culmination of my master's degree in biotechnology. The work was performed at the Institute of Biotechnology at NTNU. I would like to express my gratitude towards my mentor and supervisor, Dr. Scient. and Research Professor Helga Ertesvåg. I am grateful for the chance you have given me to do independent research and experiments, while all the time providing the support and guidance I needed. I would also like to extend my gratitude towards all my colleagues in the labs in K3 for your help with the practical lab work. I would like to thank my parents, Janne Marit and Arve, for your support and love. I am also grateful to have enjoyed the company of all of my friends at NTNU both at school and during our spare time. And last, but not least, Runar, thank you for being my best friend, my biggest supporter and my rock during these years.

My years at NTNU have thought me a lot about the world of science. I will always treasure the experiences I've had at NTNU. I look forward to what comes next, whatever that is.

Abstract

Thraustochytrids are industrially interesting microorganisms because of their ability to produce large amounts of lipids, especially the omega-3 fatty acid docosahexaenoic acid (DHA). The metabolism of these organisms is not completely understood. The unravelling of the underlying mechanisms of lipid production in these organisms is of great importance in order to achieve better yields of DHA in industrial fermentation. The Auromega project, of which this thesis is a part, attempts to understand and model the metabolism of the thraustochytrid *Aurantiochytrium* T66.

The object of this thesis was to identify the presence of the enzyme ATP-dependent citrate lyase (ACL) and the cofactor requirements glycerol 3-phosphate dehydrogenase (G3PDH) in *Aurantiochytrium* T66. Five genes of interest possibly encoding these enzymes were provided by the Auromega project group. Gene T66006817.1, T66010760.1, and T66009329.1 were candidate genes for encoding ACL, while gene T66008081.2 and T66006468.1 were candidate genes for encoding G3PDH. None of the candidate genes for ACL were annotated as ACL, but as enzymes catalyzing similar reactions. The absence of good ACL candidates could mean that *Aurantiochytrium* T66 does not have the ACL enzyme at all. ATP-dependent citrate lyase is an enzyme involved in providing acetyl-CoA for the lipid synthesis in the cytosol. The absence of this enzyme would mean that some other mechanism provides acetyl-CoA for lipid synthesis. The cofactor requirements of G3PDH was interesting to investigate, because the requirements of either NADH or NADPH would be important for modelling of the metabolism. G3PDH is involved in the production of glycerol 3-phosphate, which is the backbone for triacylglycerols. Fatty acids are commonly stored as triacylglycerols in lipid accumulating organisms.

The genes were cloned into expression vectors based on the RK2 plasmid using a *Pm*/XylS promoter system. The vectors contained a sequence encoding a C-terminal His-tag, thus enabling purification by immobilized metal-affinity chromatography (IMAC). The plasmids were transferred to *Escherichia coli*. Recombinant protein was attempted purified by IMAC, but the genes of interest could not be expressed using the initial set up. Different strategies to improve expression were then applied. Growth temperature in *E. coli* was lowered. Expression was attempted in *Pseudomonas putida* with different protocols with regards to cell density at induction and inducer concentration. The use of the signal peptide PelB was attempted in both *E. coli* and *P. putida*. None of these strategies yielded a significant yield of the proteins of interest when SDS-PAGE was performed.

A second approach was to assay activity in *Aurantiochytrium* T66. *Schizochytrium* S21 was used as a control where ACL activity had been previously reported. Growth curves for both organisms were made, and samples for enzyme assays were taken in the exponential faze and in the transition to the stationary faze. The culture samples were attempted lysed using sonication. Lysis of the cells proved difficult, and the protein yields were low. ACL and G3PDH assays were nevertheless performed on the protein extracted. For G3PDH, both assays with NADH and NADPH as cofactors were performed. The assays did not show any significant activity of neither ACL nor G3PDH. The assay protocols were verified using commercially produced enzymes. Since *Schizochytrium* S21 has been reported to have ACL activity and the assumed G3PDH genes in *Aurantiochytrium* T66 showed good homology to genes encoding characterized enzymes, the failure of the assays was probably caused by the inefficient lysis.

Sammendrag

Thraustochytrider er industrielt interessante mikroorganismer på grunn av deres evne til å produsere store mengder lipider, spesielt omega-3 fettsyren dokosaheksaensyre (DHA). Metabolismen til disse organismene er ikke fullstendig kartlagt. Utredningen av mekanismene bak fettsyreproduksjonen i disse organismene er viktig for å kunne forbedre utbyttet av DHA ved industrielle fermenteringer. Auromegaprosjektet, som dette oppgaven er en del av, jobber for å bedre forstå og modellere metabolismen til thraustochytriden *Aurantiochytrium* T66.

Formålet med denne oppgaven var å identifisere tilstedeværelsen av enzymet ATP-avhengig sitrat lyase (ACL) og kofaktoravhengigheten til de to ulike glyserol 3-fosfat dehydrogenasene (G3PDH) i *Aurantiochytrium* T66. Fem kandidatgener som kunne kode for disse enzymene ble valgt ut av Auromegagruppen. Gen T66006817.1, T66010760.1 og T66009329.1 var kandidatgener for ACL, mens gen T66008081.2 og T66006468.1 var kandidatgener for G3PDH. Ingen av kandidatgenene for ACL var annotert som dette enzymet. Disse var annotert som enzymer som katalyserer lignende reaksjoner. Fraværet av gode kandidatgener for ACL kunne bety at *Aurantiochytrium* T66 ikke har dette enzymet. ACL er et enzym som er involvert i å skaffe acetyl-CoA til lipidsyntese i cytosol. Hvis dette ensymet ikke er til stede, vil det bety at en annen mekanisme benyttes for å forsyne fettsyresyntesen med substrat. Kofaktoravhengigheten til G3PDH var interessant å se nærmere på, fordi avhengigheten av enten NADH eller NADPH vil være viktig informasjon for modelleringen av organismens metabolisme. G3PDH er involvert i produksjonen av glyserol 3-fosfat, som utgjør en sentral del av triglyserider. Fettsyrer lagres ofte som triglyserider i organismer som akkumulerer lipider.

Kandidatgenene ble klonet inn i ekspresjonsvektorer basert på RK2 plasmidet med et *Pm*/XylS promotorsystem. Vektorene inneholdt en sekvens som kodet for en C-terminal His-tag, hvilket gjorde utrensing med immobilisert metallaffinitetskromatografi (IMAC) mulig. Plasmidene ble overført til *Escherichia coli*. Rekombinant protein ble forsøkt utrenset med IMAC, men genene kunne ikke uttrykkes med dette oppsettet. Ulike strategier ble forsøkt for å forbedre uttrykket. Dyrkningstemperaturen i *E. coli* ble senket. Uttrykk i *Pseudomonas putida* med ulike protokoller i forhold til celletetthet ved induksjon og induksjonssubstanskonsentrasjon ble forsøkt. Bruken av signalpeptidet PelB ble forsøkt i både *E. coli* og *P. putida*. Ingen av disse strategiene ga signifikant utbytte av peptidene når SDS-PAGE ble utført.

En alternativ strategi var å måle enzymaktivitet i *Aurantiochytrium* T66. *Schizochytrium* S21 ble brukt som en kontroll hvor ACL aktivitet var tidligere rapportert. Vekstkurver for begge organismene ble laget, og prøver for enzymaktivitetsmålinger ble tatt i både eksponentiellfase og i overgangen til stasjonærfase. Cellekulturprøvene ble forsøkt lysert ved sonikering. Lysering viste seg å være vanskelig, og proteinutbyttet var lavt. ACL og G3PDH aktivitetsmålinger ble likevel utført. For G3PDH ble målinger med både NADH og NADPH utført. Målingene viste ingen signifikant aktivitet for hverken ACL eller G3PDH. Aktivitetsmålingsprotokollene ble verifisert ved å teste kommersielt produserte enzymer. Siden ACL aktivitet i *Schizochytrium* S21 var tidligere rapportert og G3PDH-kandidatene i *Aurantiochytrium* T66 viste god homologi med karakteriserte enzymer, ble det antatt at aktivitetsmålingene var negative på grunn av lyseringsproblemene.

Table of Contents

1. Introduction

Export of seafood is an important source of income for Norway. In 2018, seafood for 99 billion Norwegian kroner were exported, of which 72 % were produced by aquaculture. The export value is rapidly growing, with an increase of 122 % in the last decade (NorwegianSeafoodCouncil, 2019).

However, this industry faces a substantial challenge; there is a lack of long-chain omega-3 fatty acids for the fish feed. The current available amount of omega-3 fatty acids is insufficient both for the advised human consumption and for the growth of aquaculture. Omega-3 fatty acids, especially eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) have been shown to have positive effects on human health, while the lack of these will have negative health affects (Salem and Eggersdorfer, 2015). A reduced amount of these fatty acids in fish feed does not seem to have a negative effect on the growth of the fish itself. However, in some cases fish fed with reduced amounts of omega-3 fatty acids have been more susceptible to pathogens compared to fish fed with higher amounts of omega-3 fatty acids (Bransden et al., 2003). Nevertheless, a reduced amount of omega 3 fatty acids in the fish feed will substantially decrease the nutritional value for human consumption. This is why the replacement of the EPA and DHA containing fish oil with plant oil, which does not contain these components, will not provide a viable solution for production of fish feed (Gasco, 2018).

Several solutions have been proposed to produce more omega-3 fatty acids. One solution is to genetically engineer plants and animals to increase their omega-3 content. Even though this has been successfully performed, the resistance towards genetically modified organisms is substantial in the Western world. However, another possibility is also present; the use of omega-3 producing algae. The process is relatively expensive, but it runs on cheap and readily available raw materials. It is even vegan, halal and kosher, and thus it can be used to supply omega-3 fatty acids for people of particular nutritional regimes related to different beliefs (Salem and Eggersdorfer, 2015).

Microbial oils have been studied for more than a century. Advances were made on the identification and exploitation of oil producing microorganisms during the first half of the 20th century, particularly in Germany during the two wars where food sources were scarce. However, the development stopped when agricultural oil production gained efficiency after

World War 2. In the early 1980s, microbial oils once again became interesting because some of these oils have a valuable lipid composition (Ratledge and Wynn, 2002).

Several organisms have been identified and selected for commercial production of microbial oils. Within the thraustochytrid family, several species have proven to be able microbial oil production organisms (Ratledge and Wynn, 2002). The mechanisms of lipid production in the different thraustochytrid species is not completely understood. By developing a more detailed understanding of the organism, strain improvement by genetic engineering may be performed. This may again give a more economically viable process, enabling the oil produced to be accessible for low price range markets such as aquaculture (Aasen et al., 2016).

The Auromega project is a cooperation project between NTNU and SINTEF, funded by Digital Life Norway. The project started in 2017 and aims to develop a model of thraustochytrids to better exploit their lipid-producing abilities. In order to do this, experimental molecular genetics, mathematical modelling, and large-scale omics analyses will be used. This master thesis was a part of this project. The focus of this thesis was to investigate the potential presence and activity of a few enzymes involved in the lipid synthesis of a strain of thraustochytrids; *Aurantiochytrium* T66. This was done by cloning and heterologous expression of the candidate genes in bacteria.

1.1 Thraustochytrids

Thraustochytrids were first described in 1934. These organisms are unicellular, heterotrophic eukaryotes (Aasen et al., 2016). They usually have a uniform, spherical shape with a diameter between 30 and 100 µm (Gupta et al., 2012). Thraustochytrids are marine organisms found in a wide variety of marine environments, usually within the salinity range of 20-34 ‰. Even though these organisms are found in both hypo- and hypersaline environments, they are considered obligately dependent on sodium. They usually have numerous mitochondria, and might have several lipid bodies in their cytoplasm (Fossier Marchan et al., 2018).

The taxonomy of these organisms has been widely debated. In a review written by Fossier Marchan et al. (2018), the taxonomic order in Table 1.1 was proposed.

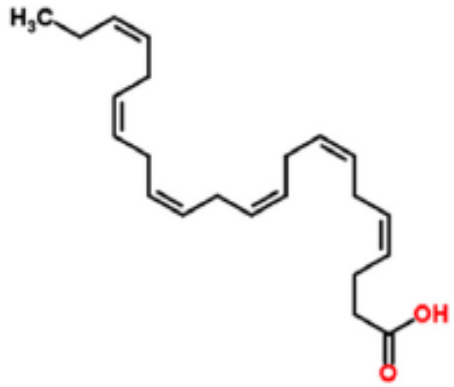
Table 1.1: ***Taxonomic classification of the Thraustochytriidae family****. Classification proposed by Fossier Marchan et al. (2018).*

Kingdom: Chromista
Subkingdom: Harosa
Superphylum: Heterokonta
Phylum: Bygira
Subphylum: Sagenista
Order: Thraustochytrida
Family: Thraustochytriidae

In the Thraustochytriidae family there is currently recognized nine genera; *Thraustochytrium, Japonochytrium, Schizochytrium, Aurantiochytrium, Ulkenia, Sicyoidochytrium, Parietichytrium, Botryochytrium, and Monorhizochytrium* (Fossier Marchan et al., 2018).

Thraustochytrids have gained interest during the last few decades, in particular due to their docosahexaenoic acid (DHA) producing properties (Gupta et al., 2012). They are also found to be oleaginous organisms, which means that they are able to store lipids in amounts that surpasses 20% of their dry weight (Ratledge and Wynn, 2002). Jakobsen et al. (2008) reported that during laboratory fermentation trials the thraustochytrid *Aurantiochytrium* T66 reached a total lipid content of 54-63 % (w/w) of dry weight. In the same report, the maximum fraction of DHA was found to be 36-52 % of the total lipids. They also showed that the increase in lipid storage in the thraustochytrid was a consequence of nitrogen and/or oxygen limitations.

DHA is a polyunsaturated fatty acid (PUFA). It is commonly denoted 22:6ω,3, meaning that the carbon chain consists of 22 carbon atoms and has 6 double bonds, of which the first is at position number 3 counted from the methyl end (Gupta et al., 2012). The structure of DHA is shown in Figure 1.1.



Docosahexaenoic acid (DHA)
$C_{22}H_{32}O_2$

*Figure 1.1**: Structure of docosahexaenoic acid (DHA)**. (Rapoport and Taha, 2014).*

Intake of DHA has been proven important to ensure correct embryonic development, as well as being a prevention against cardiovascular diseases in adults (Fossier Marchan et al., 2018). DHA-containing oils have been industrially produced by thraustochytrids for the last couple of decades. The company originally called OmegaTech, today incorporated into DSM, started their production in the 1990s (Aasen et al., 2016). DHA-containing oils produced by thraustochytrids have been extensively tested and approved as safe by both American and European authorities. In some cases, the DHA content in food have been increased by either adding the oil directly to the product (e.g. choritzo sausage) or feeding it to the animal producing the product (e.g. hens). The aquaculture industry is currently looking for ways to acquire DHA without having to capture wild fish stocks. DHA from thraustochytrids might provide an alternative DHA source, either by the direct addition of the organisms in the feed, by feeding thraustochytrids to shrimps and other small organisms used as fish feed, or by extracting the DHA from the thraustochytrids and then incorporating the oil in the fish feed (Fossier Marchan et al., 2018).

1.2 Synthesis of Fatty Acids

Glucose is a compound that is used to supply cells with both energy and precursors for cellular components. It can be oxidized via the glycolysis to pyruvate in the cytosol of the cell. The pyruvate may be further used to provide energy for the cell by degradation through the tricarboxylic acid cycle (TCA) in the mitochondria (Nelson and Cox, 2013b). Alternatively, the pyruvate can be used to make fatty acids (Ratledge and Wynn, 2002). An overview of these processes is shown in Figure 1.2.

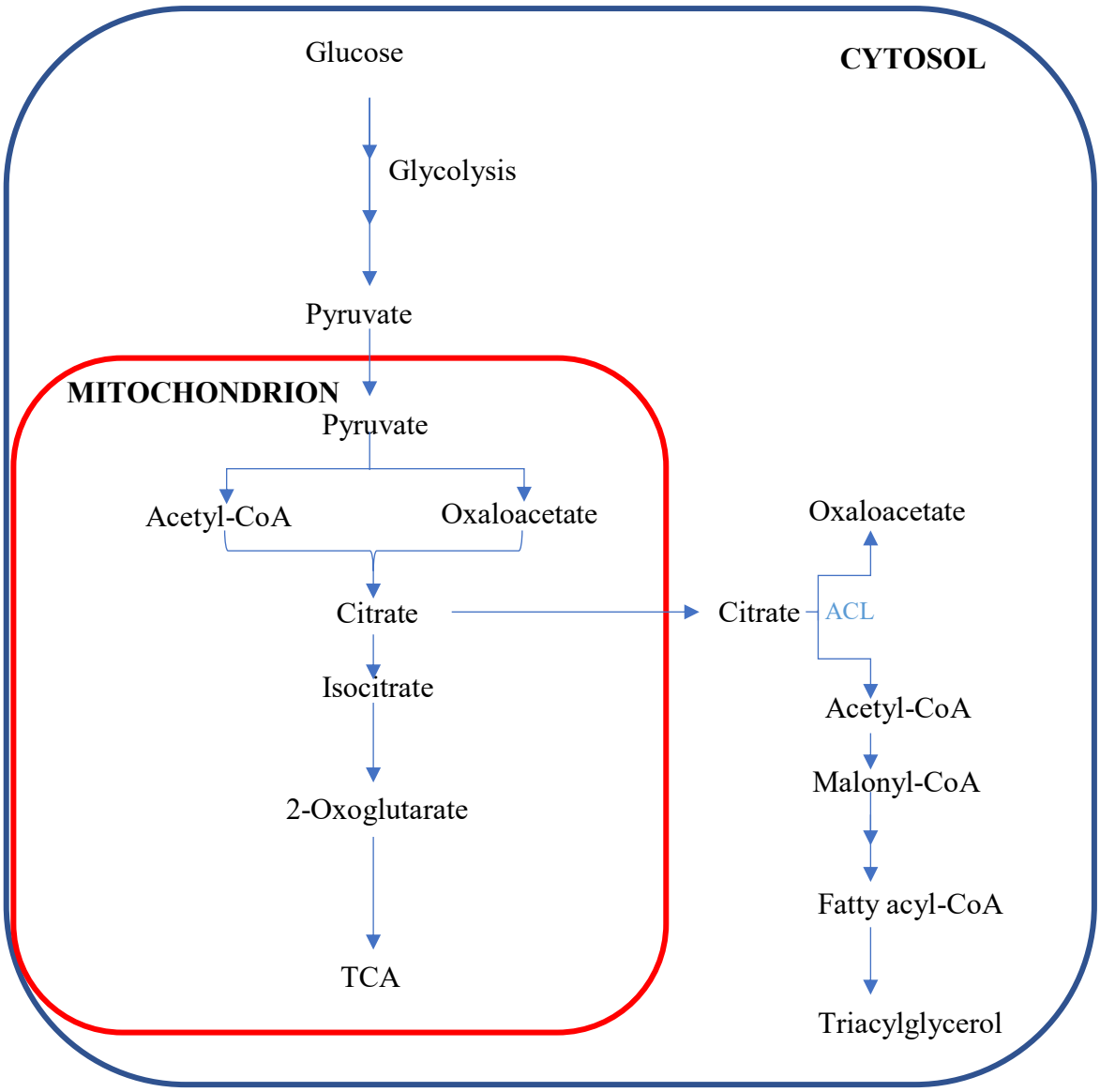


Figure 1.2: ***Overview of the breakdown of glucose through the TCA cycle or its conversion to triacylglycerols****. Glucose may be degraded to pyruvate through the glycolysis. The pyruvate molecules may be further degraded to provide energy through the TCA cycle in the mitochondria or to provide precursors for production of triacylglycerols (Nelson and Cox, 2013b). The figure is adapted from Ratledge and Wynn (2002). The ATP-dependent citrate lyase (ACL) enzyme is indicated in blue.*

Fatty acids in eukaryotes are synthesized from malonyl-coenzyme A (CoA) molecules (Ratledge, 2004). Malonyl-CoA is made from acetyl-CoA and a converted CO_2 molecule by the enzyme acetyl-CoA carboxylase at the cost of one adenosine triphosphate (ATP) (Figure 1.3). The synthesis of fatty acids is a four-step process. This process is catalyzed by either

separate enzymes or a complex, which is either way called fatty acid synthase (FAS) (Nelson and Cox, 2013c).

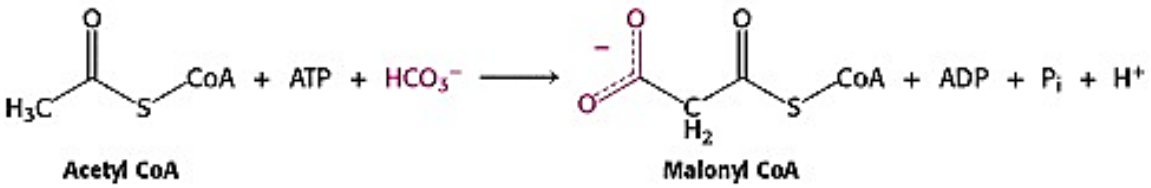


Figure 1.3: ***Synthesis of malonyl-CoA****. An acetyl-CoA molecule is converted to a malonyl-CoA molecule by addition of a converted CO_2 molecule (HCO_3^-). The reaction is catalyzed by the enzyme acetyl-CoA carboxylase, containing a biotin prosthetic group, at the cost of one adenosine triphosphate (ATP) molecule. The figure is copied from Berg et al. (2002b).*

A thiol group of the FAS system needs to be activated before the synthesis can start. The activation is achieved by the transfer of an acetyl group from acetyl-CoA, via acyl carrier protein (ACP), to a cysteine -SH group. An ACP thiol group is then activated by the transfer of a malonyl group from malonyl-CoA to the protein. ACP may be either a domain of the FAS complex or a separate peptide, depending on the organism. It serves as a transfer aid, shuttling intermediates of the reaction between different domains or peptides (Nelson and Cox, 2013c).

The first step of the reaction is a condensation of the activated acyl group attached to the FAS and two carbons from the malonyl-CoA attached to the ACP. The condensation releases a CO_2 molecule, and yields a β-keto product. The next step is a reduction of the β-keto group to yield an alcohol at the cost of one NADPH molecule. A water molecule is then eliminated, leaving a double bond. The double bond is reduced at the cost of another NADPH molecule, yielding a saturated fatty acyl group. The first round of the four-step reaction is shown in Figure 1.4. These four steps may then be repeated to yield palmitate, which is a saturated fatty acid containing 16 carbon atoms. The final fatty acid is released from ACP by enzymatic hydrolysis. (Nelson and Cox, 2013c).

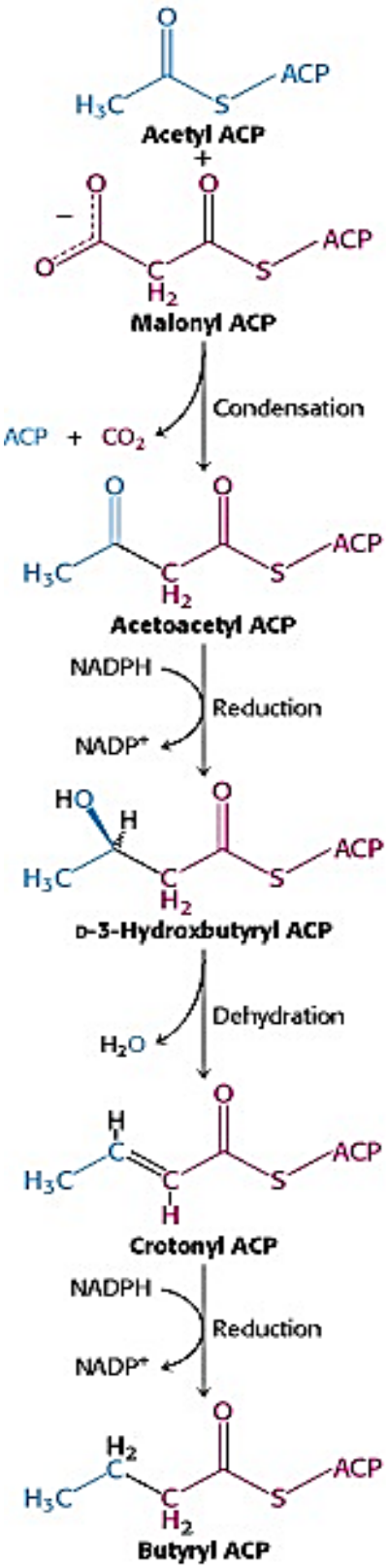


Figure 1.4: ***The first four-step round in the synthesis of a fatty acid****. The reaction is catalyzed by fatty acid synthase (FAS). The first step of the reaction is a condensation of an activated acyl group attached to the FAS via acyl carrier protein (ACP) and two carbons from the malonyl-CoA attached to an ACP. The condensation releases a CO_2 molecule, and*

yields the β-keto product acetoacetyl-ACP. The next step is a reduction of the β-keto group to yield an alcohol (D-3-hydroxybutyryl-ACP) at the cost of one NADPH molecule. A water molecule is then eliminated, leaving a double bond (crotonyl-ACP). The double bond is then reduced at the cost of another NADPH molecule, yielding a saturated fatty acyl group (butyryl-ACP). The figure is copied from Berg et al. (2002b).

Acetyl-CoA is made in the mitochondria of nonphotosynthetic eukaryotes, while the fatty acid synthesis takes place in the cytosol. One way of overcoming this locational difference, is converting the acetyl group to citrate by citrate synthase, transport of the citrate across the mitochondrial membrane, and then reconversion to acetyl-CoA by ATP-dependent citrate lyase (ACL) in the cytosol (Nelson and Cox, 2013c). The activity of ACL is further described in Section 1.4.1.

Fatty acid synthesis in thraustochytrids has been proposed to use two similar pathways; the FAS system described above and the polyketide synthase (PKS) pathway (Aasen et al., 2016). The FAS pathway generates short, saturated fatty acids such as 14:0 and 16:0, which have been detected in thraustochytrids (Hauvermale et al., 2006). These fatty acids may be converted to unsaturated, longer fatty acids by enzymes such as desaturases and elongases (Aasen et al., 2016). Longer, unsaturated fatty acids have been found in thraustochytrids (Yokoyama and Honda, 2007). However, it has been proposed that DHA is produced by the PKS pathway in thraustochytrids. The pathway is not completely known, but it presumably uses similar steps as the FAS pathway. The main difference is that in most cycles, the double bond from the newly added malonyl-CoA is not removed. This will save reducing power (see Figure 1.5) (Aasen et al., 2016).

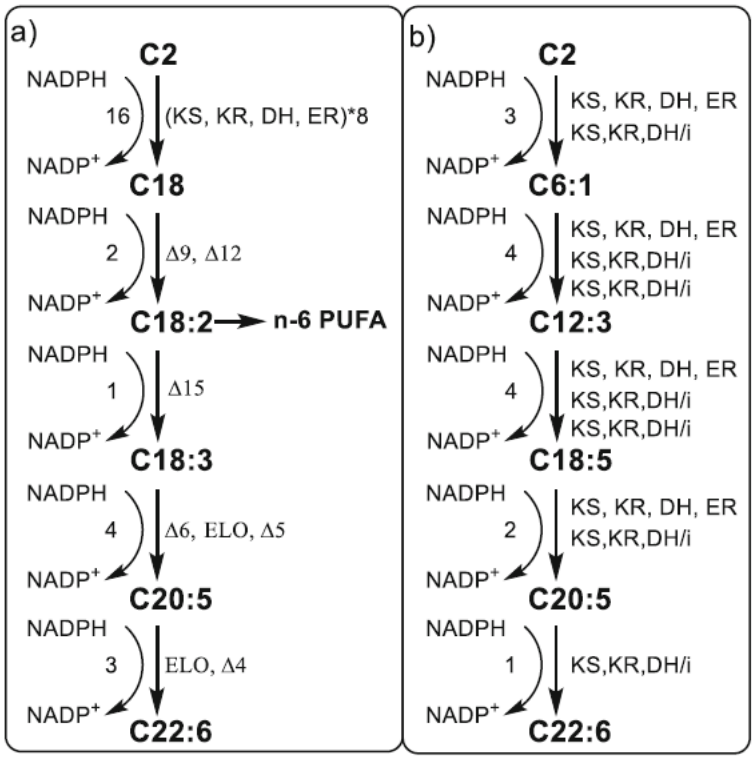


Figure 1.5: ***Comparison of the FAS and the PKS pathways****. These pathways are two possible pathways for synthesis of DHA. a) The FAS pathway uses malonyl-CoA and NADPH to produce an 18:0 fatty acid. Elongases and desaturases may then further transform the fatty acid into DHA (C22:6). The ω-3 pathway is shown, while the branching point to ω-6 is indicated. b) The PKS pathway uses the same precursors as FAS to generate DHA, but a different enzymatic combination. The amount of NADPH used for each step in the two pathways is indicated on the left side of each pathway, while the enzymes involved are indicated on the right side. KS is ketoacyl transferase, KR is keto reductase, DH is dehydratase, DH/i is a combined dehydratase and trans-cis isomerase capable of removing double bonds, and ER is enoyl reductase. Δ signifies a desaturase, where the number explains the position of the introduced double bond from the carboxyl end. ELO is a multifunctional elongase incorporating the activities of KS, KR, DH and ER. The figure is copied from Aasen et al. (2016).*

1.3 Synthesis of Triacylglycerols

Synthesized fatty acids are usually either incorporated into the phospholipids found in membranes or used to make triacylglycerols. In order to make either of these molecules, a backbone in the form of phosphatidic acid is needed. This molecule is made from L-glycerol 3-phosphate by the attachment of two fatty acids, which are carried as fatty acyl groups by coenzyme A. The L-glycerol 3-phosphate itself can be made from dihydroxyacetone phosphate from the glycolysis by glycerol 3-phosphate dehydrogenase (G3PDH) or from glycerol by glycerol kinase (Nelson and Cox, 2013c). A general outline of the production of phosphatidic acid is shown in Figure 1.6.

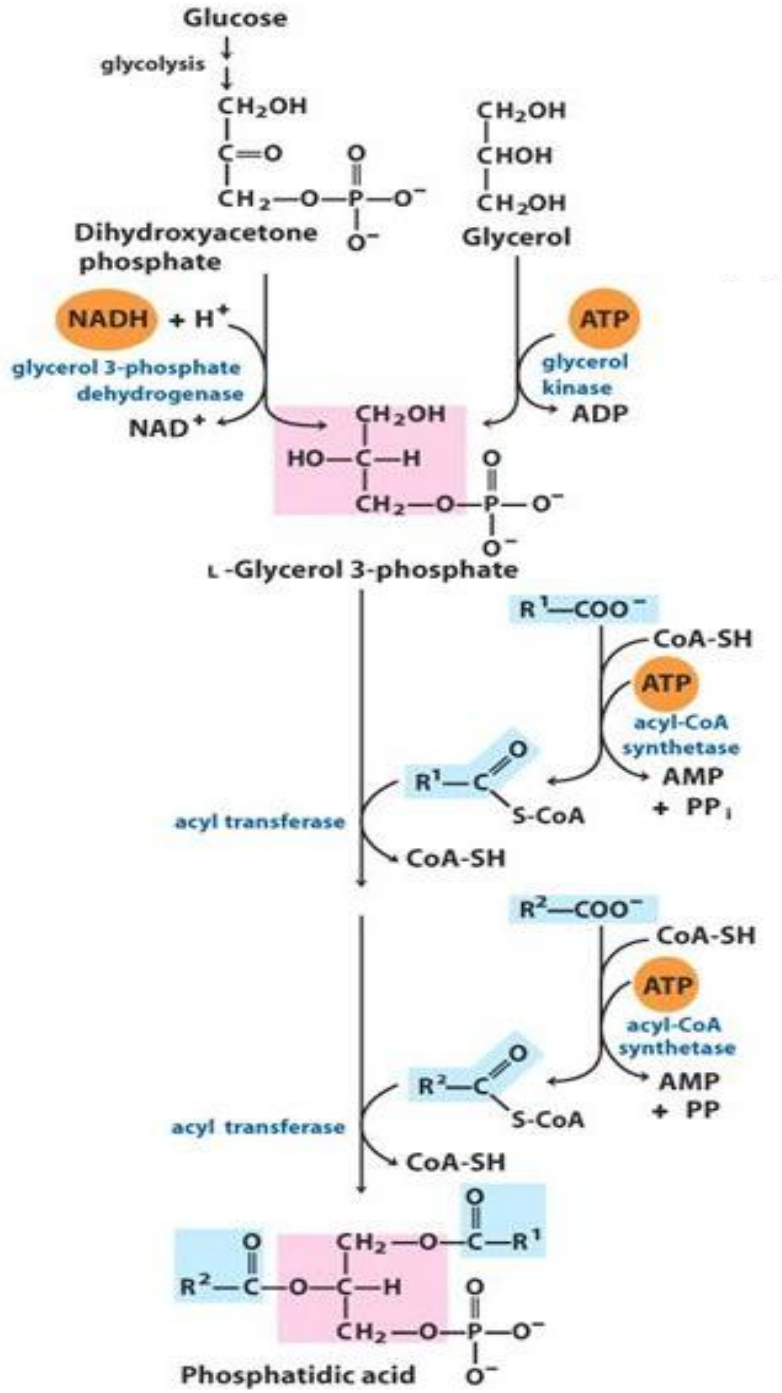


Figure 1.6: ***Overview of phosphatidic acid synthesis****. L-glycerol 3-phosphate can be made both from glycerol by glycerol kinase or from dihydroxyacetone phosphate from the glycolysis by glycerol 3-phosphate dehydrogenase. Fatty acids may then be attached by ester linkages by acyl transferase. The fatty acids are carried by coenzyme A, to which they are attached by acyl-CoA synthetase. The cost of ATP or NADH molecules are indicated by orange circles.*

The R^1 and R^2 in the formulas represents fatty acids. The figure is copied from Nelson and Cox (2013c).

The phosphatidic acid may be used to synthesize either a triacylglycerol or a glycerophospholipid by the action of different enzymes. For glycerophospholipids, a head group is attached to the phosphate group. Phosphatidic acid phosphatase can convert phosphatidic acid to 1,2-diacylglycerol, which in turn may receive a third fatty acyl group transported by coenzyme A by the activity of acyl transferase. This will yield a triacylglycerol (Nelson and Cox, 2013c). This process is outlined in Figure 1.7.

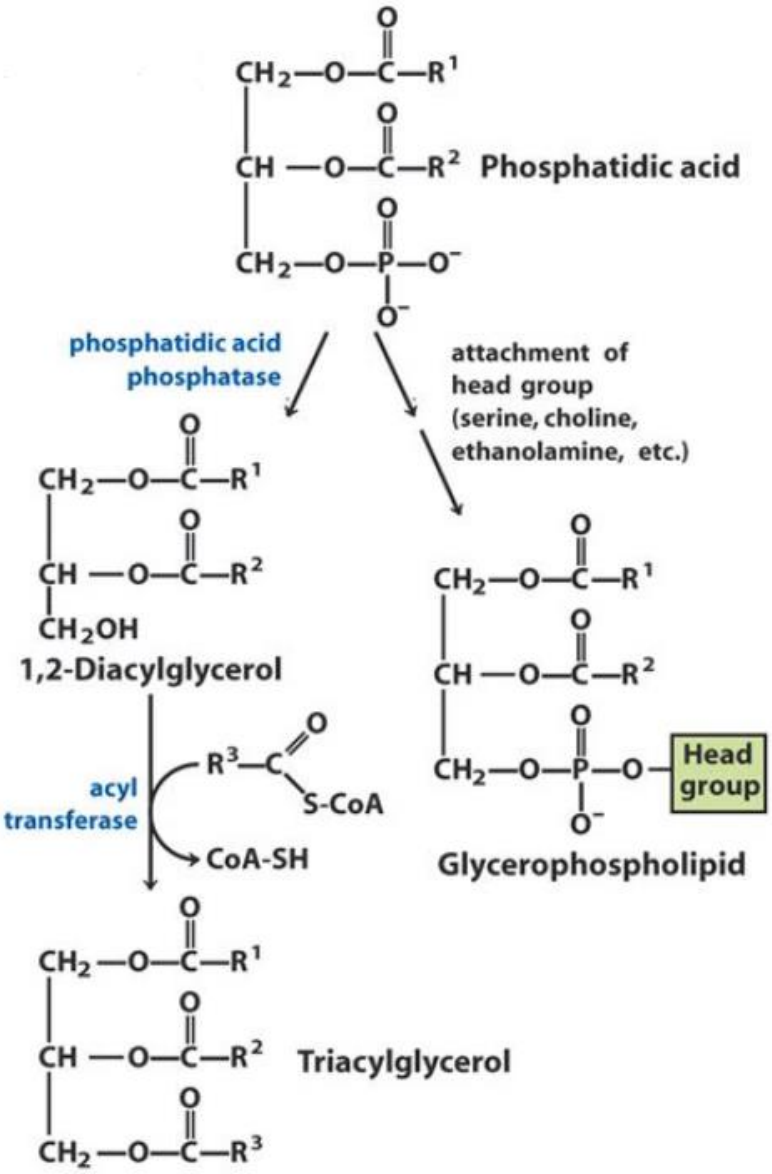


*Figure 1.7:**Overview of the synthesis of triacylglycerols and glycerophospholipids**. These two groups of lipids are both synthetized from phosphatidic acid. For glycerophospholipids, a*

head group is attached to the phosphate group. Alternatively, phosphatidic acid phosphatase can convert phosphatidic acid to 1,2-diacylglycerol, which in turn may receive a third fatty acyl group transported by coenzyme A by the activity of acyl transferase. This will yield a triacylglycerol. The figure is copied from Nelson and Cox (2013c).

1.4 Enzymes of Interest

As the Auromega project is focused on the lipid synthesis of thraustochytrids, it was interesting to investigate enzymes involved in these pathways. ATP-dependent citrate lyase (ACL) was chosen for its role in the indirect transport of acetyl-CoA from the mitochondria to the cytosol of the cells. Glycerol 3-phosphate dehydrogenase was chosen for its role in providing the substrate glycerol 3-phosphate for lipid synthesis, as well as its role in regenerating NAD+ for the glycolysis in the cytosol.

1.4.1 ATP-Dependent Citrate Lyase (ACL)

ATP-dependent citrate lyase is an enzyme usually found in eukaryotes. It catalyzes the cleaving of citrate, using ATP and CoA. The products are acetyl-CoA, oxaloacetate adenosine diphosphate (ADP) and inorganic phosphate (P_i) (Srere, 1972). The reaction is shown in Eq. 1.1.

$$Citrate^{3-} + ATP + CoA = acetyl\text{-}CoA + oxaloacetate + ADP + P_i \qquad Eq.\ 1.1$$

Mg^{2+} has been found to be the most effective cofactor, but the reaction may proceed with a slower rate when Mn^{2+} or Co^{2+} is replacing Mg^{2+} (Srere, 1972).

The ACL enzyme has been pointed out by Ratledge and Wynn (2002) to be an essential enzyme for oleaginous yeasts.

Increased lipid production in oleaginous yeasts can be caused by a nitrogen limitation. Growth under these conditions have been shown to lead to an abrupt drop in adenosine monophosphate (AMP) levels in the cells. This has been proposed to be due to the nitrogen dependence of the enzyme AMP deaminase, which catalyzes the conversion of AMP to inosine monophosphate (IMP). The low AMP concentration may then reduce the activity of the isocitrate dehydrogenase (ICDH) enzyme that converts isocitrate to 2-oxoglutarate using NAD^+. This will increase the concentration of isocitrate, which may then be converted to citrate via aconitate by aconitase. The surplus citrate will then provide more substrate for the ACL, which will then provide more acetyl-CoA for lipid synthesis (Ratledge and Wynn,

2002). Figure 1.8 illustrates these processes, how they relate to the glycolysis and the citric acid cycle (TCA), and where in the cell the reactions take place.

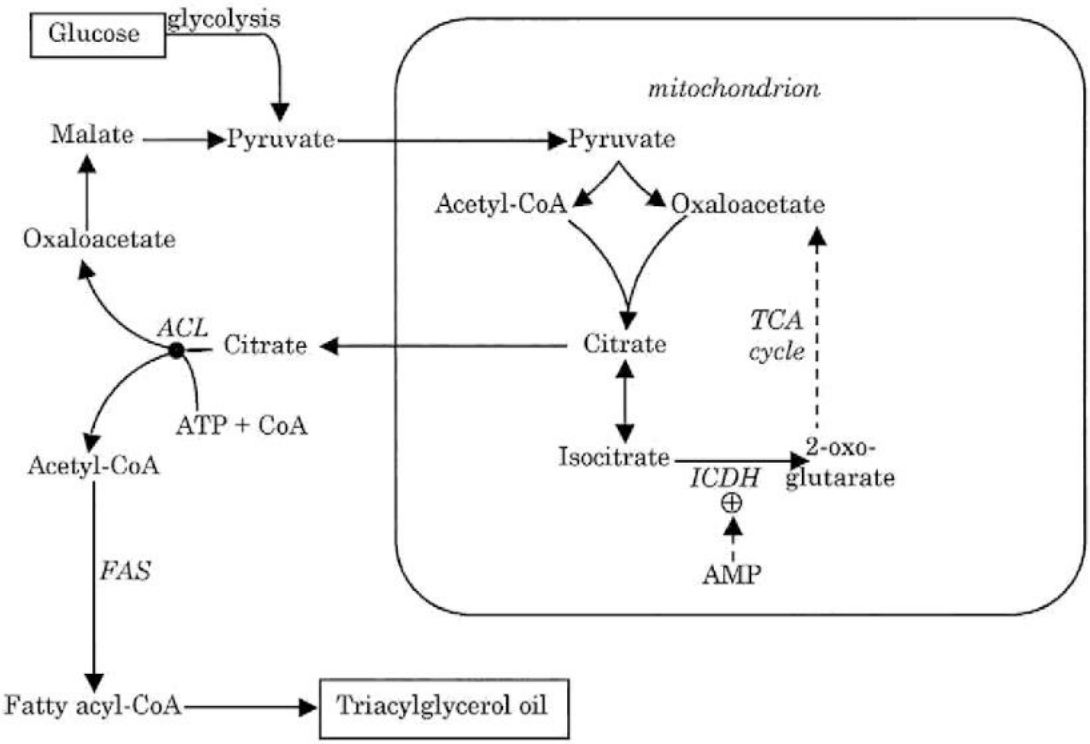


Figure 1.8: ***Scheme of the processes involved in lipid accumulation in oleaginous yeasts****. Formation of triacylglycerol may be triggered by the reduction of nitrogen availability. The resulting processes are described in the text above. The enzymes shown in the figure are ATP-dependent citrate lyase (ACL), isocitrate dehydrogenase (ICDH), and fatty acid synthase (FAS). The tricarboxylic acid cycle (TCA) is indicated as a dotted line. The figure is copied from Ratledge and Wynn (2002).*

1.4.2 Glycerol 3-Phosphate Dehydrogenase (G3PDH)

In Section 1.2 the role of G3PDH in the conversion of DHAP to G3P in the cytosol is described. However, another variant of this enzyme (mG3PDH) is found embedded on the surface of the inner mitochondrial membrane. This enzyme, together with the cytosolic version, forms a shuttle system of electrons across the mitochondrial membrane.

During glycolysis, NAD^+ is consumed in the cytosol. NADH then needs to be reoxidized to NAD^+ by the respiratory chain. This takes place in the mitochondria, but the NADH molecules cannot pass the inner mitochondrial membrane. However, when cytosolic G3PDH converts DHAP to G3P, an NADH molecule is oxidized to NAD^+ (Figure 1.9). The

mitochondrial isozyme can then regenerate DHAP by reducing flavin adenine dinucleotide (FAD) to $FADH_2$. The mG3PDH then catalyzes the regeneration of FAD by transferring the two reducing units to ubiquinone in the mitochondrial membrane. The electrons will then enter the respiratory chain, where they will generate 1.5 ATP molecules. This is 1 ATP less that what the direct use of NADH would give (Berg et al., 2002a).

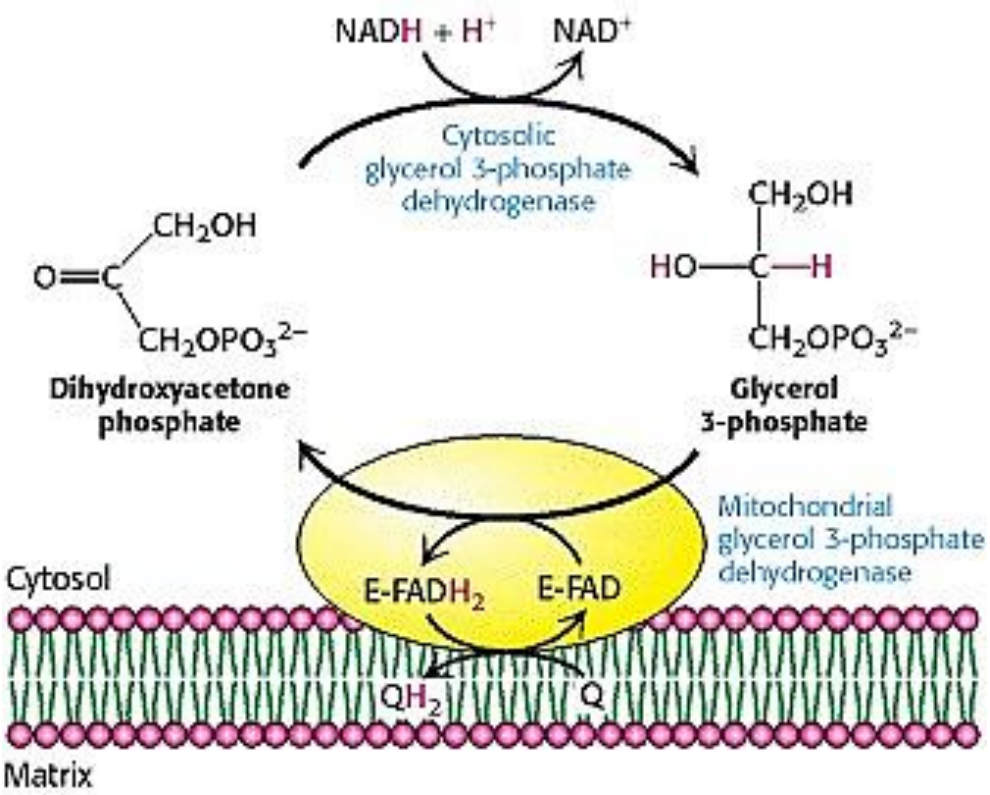


Figure 1.9: ***The glycerol 3-phoshate shuttle****. This is a mechanism regenerating NAD^+ in the cytosol. NADH cannot itself pass the mitochondrial membrane to be reoxidized by oxidative phosphorylation. However, just as cytosolic glycerol 3-phosphate dehydrogenase can reduce dihydrocyacetone phosphate to glycerol 3-phosphate (G3P), the isozyme embedded in the inner mitochondrial membrane can oxidize G3P by reducing FAD. FAD may then be regenerated by transferring the electrons to ubiquinone in the mitochondrial membrane. The electrons will then be used to generate ATP by oxidative phosphorylation. The figure is copied from Berg et al. (2002a).*

1.5 Heterologous Gene Expression

Genes may be expressed in other organisms than the one from which they originate. A practical way to do this, is to clone the gene into an expression vector (a type of plasmid). A host organism may then produce the protein encoded by the gene. An expression vector

usually includes a promotor which controls the expression of the gene, a peptide tag, a selection gene, and restriction sites. Like other plasmids, it should also have an origin of replication where the replication may be initiated (Clark and Pazdernik, 2013a).

1.5.1 Origin of Replication and Copy Number

The synthesis of DNA in all organisms start from a specific area in the DNA, an origin of replication (*ori*). The replication is the result of the coordination of several enzymes. The process starts by the binding of a replication initiation protein or an initiation complex to the origin of replication. The binding separates the DNA strands and promotes the other components needed for the replication machinery (the replisome) to function, namely the helicase, the primase and the polymerase. In the case of the bacterial chromosome origin (*oriC*), the initiation protein is DnaA (Wawrzycka et al., 2015). The RK2 plasmid carries its own origin of vegetative replication, the *oriV*. This origin of replication is made up of two clusters of a total of eight 17-base pair (bp) repeats (iterons) (Kolatka et al., 2010). The RK2 plasmid also encodes the TrfA protein, its own initiation protein. This protein interacts with the iterons, and assisted by amongst other host encoded proteins, DnaA, initiates replication (Wawrzycka et al., 2015). Even though RK2 is a low copy number plasmid present in 4-8 copies per chromosome in *E. coli* (Kolatka et al., 2010), Durland et al. (1990) showed that mutated versions of the *trfA* gene such as *trfA cop271C* will yield a significantly higher copy number.

All the pMEJ plasmids not based on a TOPO® vector backbone made in this thesis contain the RK2 *oriV* origin of replication and encode the TrfA cop271C initiation protein.

1.5.2 Promotor

Genes to be expressed in expression vectors may be placed under the control of a wide range of promotors (Clark and Pazdernik, 2013a). One such promotor is the *Pm* promotor, which in turn is regulated by the XylS regulator encoded by the *xylS* gene. The XylS regulator is in turn controlled by its own promotor, which usually creates a constitutive low expression of the regulator. The activation of the *Pm* promotor may be achieved by the addition of a wide range of benzoic acid derivatives, that are both relatively cheap and easily diffuses across the bacterial membrane. An example of one such inducer is m-toluate (3-methylbenzoate). When the inducer binds to XylS, it will form a dimer. This dimer may in turn bind to the *Pm* promotor sequence and thus induce the transcription of the gene controlled by the promotor.

The *Pm* promotor vectors also contains an efficient ribosome binding site upstream of the cloning site at the start codon (Gawin et al., 2017).

1.5.3 Signal Peptide

Numerous proteins that are synthesized inside cells, are destined to become extracytoplasmic proteins. The preproteins of these are often marked with a code for exportation, which is a short peptide sequence referred to as a signal peptide (Auclair et al., 2012). A typical signal peptide is approximately 20 amino acids (aas) long. It consists of a few positive amino acids, followed by a hydrophobic domain (Pines and Inouye, 1999). Upon reaching the site of exportation, the signal peptide is cleaved from the remaining peptide chain by a signal peptidase, resulting in a mature protein (Auclair et al., 2012). Correct cleavage may require aas such as glycine or alanine following the hydrophobic domain. *E. coli* may accumulate as much as 20% of its total protein in the periplasm, which makes it a favored producer of heterologous proteins. Additionally, translocation to the periplasm may only require a signal sequence, whereas membrane targeting is dependent on the sequence of the protein itself (Pines and Inouye, 1999).

Sletta et al. (2007) showed that the fusion of a *pelB* signal sequence to the 5' end of a gene controlled by the *Pm/xylS* promoter gave a significantly higher yield of heterologous expressed protein in *E. coli* compared to the absence of a signal sequence. The increased expression was largely attributed to improved translation of the recombinant peptides. *pelB* was chosen as the signal sequence used in this thesis.

1.5.4 His-tag

An effective way to purify recombinant proteins, is to use immobilized metal-affinity chromatography (IMAC). This separation technique exploits the affinity of histidine amino acid residues for metal ions such as Ni^{2+}, Co^{2+}, and Cu^{2+} (Bornhorst and Falke, 2000). When the technique is used to purify proteins, it is also referred to as fast protein liquid chromatography (FPLC).

When this technique is used, a sequence encoding approximately 6 histidine residues in a row is genetically engineered into either the 5' or the 3' end of the gene encoding the protein to be purified. The sequence must be placed in such a way that it will be translated in the right reading frame to encode histidine residues. Quite often, this sequence is included as part of the expression vector. A protein extract from the production organism may then be loaded onto a column with a stationary phase containing a divalent cation. The electrons in the

double bonds in the imidazole ring of the histidine residues will then form coordination bonds with the cations in the stationary phase. If the mobile phase transporting the peptides also contains a small amount of imidazole, this will prevent unspecific binding of other peptides containing a few histidine residues in a row. This will cause the His-tagged proteins to stay bound to the column, while the other proteins will be eluted. If another mobile phase containing a higher concentration of free imidazole is then loaded onto the column, the imidazole molecules will take the place of the His-tagged proteins, and these will then be eluted (Bornhorst and Falke, 2000).

1.5.5 Optimization of Heterologous Expression

Heterologous protein expression usually needs to be optimized. Several factors might be optimized in order to overcome different challenges.

Two main challenges were identified by Papaneophytou and Kontopidis (2014). Either there can be a problem to express significant amounts of the protein, or the protein expressed might aggregate and form inclusion bodies. If proteins are caught in inclusion bodies, their biological activity might be lost even if the aggregates are dissolved (Kaur et al., 2018). In order to express significant amounts of protein, both transcription and translation needs to be optimized (Hannig and Makrides, 1998).

Temperature optimization has been shown to be a possible solution to both of the main problems related to heterologous expression. Translation of heterologously expressed genes has in some cases been shown to be temperature dependent (Al-Hendy et al., 1991). Also, lowering of growth temperature can in some cases prevent formation of inclusion bodies (Kaur et al., 2018). A too high growth temperature might also cause plasmid loss (Papaneophytou and Kontopidis, 2014).

The basic design of DNA, with 4 different bases read as codons of 3 bases, gives the possibility of 64 different codons. This is more than enough to encode the 20 amino acids and stop codon needed. Consequently, one amino acid can usually be encoded by several different codons. Different organisms tend to prefer specific codons. This is referred to as the codon bias (Athey et al., 2017). This seems to be particularly important at the beginning of the translation (Hannig and Makrides, 1998). The GC-content of the host organism might have an influence on the expression of heterologous proteins, because it is related to the codon bias. This is mainly due to the preference, or at least larger proportion, of C or G as the third base in the codons. The GC-content of an organism is related to the proportion of tRNAs available

for each codon. The difference in codon preference and consequently tRNA composition between the host organism and the organism the gene originates from may be overcome by two different solutions. The gene may be codon optimized to encode codons more common to the host organism. Alternatively, the genes for the tRNAs posing the problem might be introduced and induced along with the genes to be expressed, either in the expression vector or in a separate vector (Kaur et al., 2018).

Proteolysis can also be an important factor when optimizing heterologous expression. This process is a natural part of the host metabolism, where proteins are degraded. The process is not random, and proteins that are for instance recognized as misfolded will be degraded. The translocation of a heterologously expressed protein to the periplasm of the host organism may solve the proteolysis problem. Numerous other solutions to this problem also exists, such as using fusion proteins, using host cells with lower amounts of proteases, and coexpressing chaperones (Hannig and Makrides, 1998).

Translocation to the periplasm might also solve several other expression problems. The cytoplasm of bacteria is a reducing environment. Some proteins might need the more oxidizing conditions found in the periplasm to achieve the correct formation of disulfide bonds. In some cases it might also facilitate recuperation of the protein after production (Kaur et al., 2018).

1.6 Purpose of the Thesis

The sequence of the *Aurantiochytrium* T66 have been determined by Illumina-sequencing, and these sequences are published at the NCBI online database with the accession code LNGJ00000000 (Liu et al., 2016). These sequences have been annotated by Basic Local Alignment Search Tool (BLAST) comparison. In this thesis, some of the open reading frames (ORFs) were investigated further to determine if they could be more confidently annotated. The OFRs chosen as genes of interest are shown in Table 1.2. The parts of the gene codes indicated in bold will be used as the short forms of the gene codes throughout this thesis. None of these genes were found to contain introns. The information described in this section and in the table was provided by the Auromega project group.

The Auromega project group selected five ORFs suspected to encode either G3PDH or ACL activity. No genes in *Aurantiochytrium* T66 were annotated as ACL, but genes that were annotated as succinate-CoA ligase or citrate lyase were chosen to study further. Succinyl-CoA is an enzyme involved in the citric acid cycle, and it catalyzes the conversion of succinyl-CoA

to succinate while generating an ATP-molecule and liberating CoA (Nelson and Cox, 2013a). The molecules involved in this reaction bear resemblance to the molecules involved in the reaction catalyzed by ACL, although the reaction described in Section 1.4.1 is similar to the reverse reaction described for succinyl-CoA. It is thus easy to assume that the enzymes involved in the two reactions might bear resemblance, and that a computer-based annotation might confuse the two.

*Table 1.2: **Overview of genes of interest in this thesis**. The code of the genes used by the Auromega project, the preliminary annotation of the genes, and the suspected product of the genes are indicated. Also, the scaffold number, their GenBank accession code, and the region on the scaffold where the gene can be found are listed.*

Gene code	Annotation	Suspected product	Scaffold GeneBank	Accession code	Region
T6600**6817**.1	Succinate--CoA ligase [ADP-forming] subunit alpha	ACL	Scaffold_0008	KQ758883.1	773623..774588
T660**10760**.1	Succinate--CoA ligase [ADP-forming] subunit beta	ACL	Scaffold_0005	KQ758821.1	1050626..1051903
T6600**9329**.1	Citrate lyase subunit beta-like protein	ACL	Scaffold_0002	KQ758900.1	1387734..1388699
T6600**6468**.1	glycerol-3-phosphate dehydrogenase	G3PDH	Scaffold_0007	KQ758880.1	Complement (959159..961102)
T6600**8081**.2	glycerol-3-phosphate dehydrogenase, mitochondrial precursor	G3PDH	Scaffold_0025	KQ758896.1	216096..218339

The purpose of this thesis was to see whether the ORFs of interest could produce functional enzymes when cloned into bacteria, and thereby be able to evaluate the correctness of the previous annotation. Also, no clear ACL homologue was found in the genome search performed by the Auromega group. Whether or not this enzyme is found in *Aurantiochytrium* T66 would be of interest in the quest to understand and model its metabolism. The existence of G3PDH in *Aurantiochytrium* T66 was assumed certain based on the annotations done by the Auromega group. However, the cofactor used by the G3PDH was not known, and the knowledge of this would be important for modelling.

2 Materials and Methods

2.1 Bioinformatic Tools and Research

Based on the information provided by the Auromega group described in Section 1.6 the preliminary computer-based research described in Section 2.1.1, 2.1.2, and 2.1.3 were performed. Other bioinformatic tools used in this thesis are subsequently described.

2.1.1 Basic Local Alignment Search Tool (BLAST)

Basic Local Alignment Search Tool (BLAST) is a widely used bioinformatic tool that aligns sequences based on similarity. It may align many combinations of protein and nucleotide sequence data, depending on the chosen variant of BLAST. It searches for short similar segments, and thus performs alignments from these "hot spots". It returns the results of the alignment along with statistical data (Johnson et al., 2008).

As part of the preliminary research, a protein BLAST (BLASTP 2.9.0+) (Altschul et al., 1997) was performed by entering the letter sequence of the protein encoded by the genes of interest. The Protein Data Bank (PDB) protein database was chosen for the search because these proteins are in general experimentally verified.

2.1.2 Swiss-Model Homology Modelling

When His-tags are to be used, it is an advantage to know whether the ends of the proteins to be tagged are embedded inside the protein or not. The SWISS-MODEL homology modelling tool was used to make 3D models of the proteins.

SWISS-MODEL takes an amino acid sequence, for instance in a FASTA format, as input. First, a template search will then be performed, where the input sequence is compared to evolutionary related protein sequences in the SWISS-MODEL template library. The template library is based on input from the PDB. The templates will then be ranked according to the

expected quality of the model based on the Global Model Quality Estimate (GMQE) and the Quaternary Structure Quality Estimate (QSQE). For each template proposed by SWISS-MODEL, a 3D model may be generated by the ProMod3 modelling engine (Waterhouse et al., 2018).

3D models were made for all the proteins of interest using the highest ranked template. Based on these models, the availability of the N-terminal and C-terminal ends for carrying a His-tag was evaluated.

2.1.3 Signal Peptide Search

In order to investigate whether the proteins encoded by the genes of interest encoded signal peptides, the SignalP v.4.1 signal peptide search program was used. This program takes a FASTA sequence input and compares the sequence to a database of experimentally verified signal peptide sequences. It returns a yes or no prediction of whether there is a signal peptide encoded by the peptide, and if yes; its location in the peptide chain and the probable cutoff site. One may choose to search within either eukaryotic, Gram positive, and Gram negative databases, and the first was chosen for the investigations in this thesis (Nordahl Petersen et al., 2011).

2.1.4 The Codon Bias

A database called the High-performance Integrated Virtual Environment-Codon Usage Tables (HIVE-CUTs) collects genome data from GenBank and NCBI's RefSeq database, analyses the codon usage, and returns the number of appearances of a given codon per 1000 codons. Thus, a low scoring codon will likely not be the preferred codon for the organism in question (Athey et al., 2017).

The first ~39 codons of each gene of interest were compared to the codon preference data provided by HIVE-CUTs for *Pseudomonas putida* (NCBI taxonomic id 303) and *E. coli* (NCBI taxonomic id 562). The beginning of the gene was chosen in the belief that the codon bias might be the most relevant in the start of the translation of the genes.

2.1.5 Benchling

Benchling is an online molecular biology tool for saving and processing data. It includes features such as the possibility to annotate parts of sequences, performing virtual digests, make new DNA molecules by digestion followed by ligation, and creation of alignments (*Benchling Molbio Overview*, 2019). Benchling have been used to process the DNA related information used in this thesis.

2.2 Primers

The primers used to amplify the genes of interest by the Polymerase Chain Reaction (PCR) were made using the program Clone. The primers were designed so that the product would have a NdeI cut site at the 5' end and a NotI, PspOMI, NheI or AvrII cut site at the 3' end. The reverse primer was given a tail that would give the right reading frame for the His-tag in the plasmid used later. Special primers were also ordered for gene 6468 and 8081 to sequence the center of the PCR products formed. Primers for M13 (forward and reverse) were used for sequencing. The primers were ordered from Sigma-Aldrich. The sequences of the primers are shown in Table 2.1. The primers used to incorporate the AvrII restriction site in the vector itself are indicated "AvrII" in the table.

Table 2.1: ***Primers used for PCR reactions and sequencing****. All primers are written in 5'-3' direction. F, R, and S indicates "forward primer," "reverse primer," and "sequencing primer", respectively. The bold bases indicate bases different from the template DNA. The underlines mark restriction enzyme cut sites, and the specific enzyme is indicated behind the primer sequence.*

Gene		**Primer sequence (5'-3')**	
6817	F	**CAT**ATGCTCGCGCGTACC	(NdeI)
	R	**GCGGC**CGC**AAGCATG**GCCTTGCCAAGAGCCTTC	(NotI)
10760	F	**CAT**ATGTCGGCCCTGGGAG	(NdeI)
	R	**GGGCCCGTAGCAT**GGCAGCGAGCTTGACAG	(PspOMI)
9329	F	**CAT**ATGACGATGTCGCCCGCCCGGAC	(NdeI)
	R	**CCTAGGAGAAG**AGCGCGCGCCTTGATAC	(AvrII)
6468	F	**CAT**ATGCAGTCCCTCCGGCGCGCATC	(NdeI)
	R	**GGGCCCGTAGCATC**TCTTCGGGGAGCTTCATGGTC	(PspOMI)
	S	CAAGGACGAGGAAGGAAAGC	
8081	F	**CAT**ATGGCGGCGAGGATGAC	(NdeI)
	R	**G**CTAGC**A**CAAACATGACGCCCGAGGAG	(NheI)
	S	TTATGAAGGCCAGCACAACG	
AvrII	F	CATAT**G**CTCGC**GC**G**TACC**	(NdeI)
	R	G**C**GGC**C**GC**A**AG**CA**TGGC**CTT**G**C**C**AAGAGC**C**TTC	(NotI)
M13	F	CGCCAGGGTTTTCCCAGTCACGAC	
M13	R	AGCGGATAACAATTTCACACAGGA	

2.3 Plasmids and Strains

The plasmids made in this thesis were based on other plasmids. The plasmids that made the starting point for the cloning in this thesis are listed in Table 2.2.

Table 2.2: ***Overview of plasmids used as starting points for cloning****. The reference/source of the plasmid is given as well as relevant information about the plasmid itself.*

Plasmid	Reference/Source	Characteristics
pMV20	Vilnes (2018)	RK2 *oriV*, *kan*^{r}, *trfA cop271, lac* promotor
pMV23	Vilnes (2018)	RK2 *oriV*, *bla*, *trfA cop271, Pm* promotor
pCR™II-Blunt-TOPO®	Invitrogen	TOPO® cloning vector, *kan*^{r}
pMSM2	Morteza Moghadam (unpublished)	*alyB* with His-tag
pJB861	Blatny et al. (1997)	*kan*^{r}
pVB1 mCherry Kan-271	Vectron Biosolutions	RK2 *oriV*, *kan*^{r}, *trfA cop271, Pm* promotor
pVB1 A1B1	Vectron Biosolutions	*pelB, Pm* promotor

During the work with this thesis, numerous plasmids were made. An overview of these is shown in Table 2.3.

Table 2.3: ***Overview of plasmids made in this thesis****. Backbone and insert for each plasmid is indicated, as well as the restriction enzymes used to cut each component.*

Plasmid	Backbone	Insert	Restriction enzymes for insert	Restriction enzymes for backbone
pMV20α	pMV20	-	-	SalI + XbaI
pMEJ1	pMV23	pMV20α	NdeI + PspOMI	NdeI + NotI
pMEJ2	pMEJ1	pMEJ1 with AvrII cut site (PCR R)	NdeI + NotI	NdeI + NotI

pMEJ3T	pCR™II-Blunt-TOPO®	T66006817.1		
pMEJ4T	pCR™II-Blunt-TOPO®	T66010760.1		
pMEJ5T	pCR™II-Blunt-TOPO®	T66006468.1		
pMEJ6T	pCR™II-Blunt-TOPO®	T66009329.1		
pMEJ7T	pCR™II-Blunt-TOPO®	T66008081.2		
pMEJ3	pMEJ1	pMEJ3T	NdeI + NotI	NdeI + NotI
pMEJ4	pMEJ1	pMEJ4T	NdeI + PspOMI	NdeI + NotI
pMEJ5	pMEJ1	pMEJ5T	NdeI + PspOMI	NdeI + NotI
pMEJ6	pMEJ2	pMEJ6T	NdeI + AvrII	NdeI + AvrII
pMEJ7	pMEJ2	pMEJ7T	NdeI + NheI	NdeI + AvrII
pMEJ8	pMEJ3	Kan^r gene from pJB861	SbfI	PstI
pMEJ9	pMEJ4	Kan^r gene from pJB861	SbfI	PstI
pMEJ10	pMEJ5	Kan^r gene from pJB861	PstI	SbfI
pMEJ11	pMEJ6	Kan^r gene from pJB861	PstI	SbfI
pMEJ12	pMEJ7	Kan^r gene from pJB861	PstI	SbfI
pMEJ13	pVB1 mCherry Kan-271	pVB1 A1B1	AgeI + NdeI	AgeI + NdeI
pMEJ14	pMEJ13	pMEJ3	NdeI + DraIII	NdeI + DraIII
pMEJ15	pMEJ13	pMEJ4	NdeI + SalI	NdeI + XhoI
pMEJ16	pMEJ13	pMEJ5	NdeI + DraIII	NdeI + DraIII
pMEJ17	pMEJ13	pMEJ6	NdeI + DraIII	NdeI + DraIII
pMEJ18	pMEJ13	pMEJ7	NdeI + SalI	NdeI + XhoI

Several organisms, and in some cases strains of these, were used in this thesis. These are shortly described in Table 2.4.

Table 2.4: ***Overview of organisms and strains****. Organisms and strains used in this thesis are listed. Comments on their properties as well as references to relevant information concerning their characteristics are given.*

Species	Strain	Comment	Reference
E. coli	DH5α	Recombinant strain from K-12 collection, commonly used for cloning and plasmid amplification	Kostylev et al. (2015)
E. coli	RV308	Recombinant protein production strain from K-12 collection, GC-content ~50%	Kujau et al. (1998), Kaur et al. (2018)
P. putida	KT2440	Bacterium with high stress-tolerance and ability be used as a production organism in industrial biotechnology, high GC-content (61.5%)	Belda et al. (2016)
Aurantiochytrium	T66	See Section 0, high GC-content (62.8%)	Liu et al. (2016)
Schizochytrium	S31 (ATCC 20888)	Thraustochytrid species, produces large amount of lipids	Fossier Marchan et al. (2018), Chang et al. (2013)

2.4 Media and Solutions

2.4.1 LB Medium

10 g/L NaCl

10 g/L Tryptone (Oxoid)

5 g/L Yeast Extract (Oxoid)

RO water

For LB agar plates, Agar Bacteriologial (Oxoid) was added to a final concentration of 15 g/L.

pH was adjusted to 7.0 with 1 M HCl or 5 M NaOH. The medium was autoclaved for 20 minutes at 120 ˚C.

When 3 times concentrated LB (3XLB) was made, the concentration of tryptone and yeast extract was tripled, while the concentration of NaCl was kept constant.

2.4.2 Antibiotic Solutions and Concentrations

200 µg/mL ampicillin (stock 100 mg/mL)

50 µg/mL kanamycin (stock 50 mg/mL)

200 µg/mL streptomycin/ampicillin (stock 60 mg/mL)

Antibiotic solutions in the stock concentrations listed above were prepared by weighing out the antibiotic and dissolving it in RO water. The solution was sterilized by filtration before it was frozen in 1 mL aliquots. Upon use, the antibiotics were thawed and added to growth media just before inoculation to give the concentrations listed in front of each antibiotic. In the case of agar plates, the antibiotics were added just before the plates were poured and after the medium temperature was below 60 ˚C.

2.4.3 YPD Medium

20 g/L Bacteriological Peptone (Oxoid)

10 g/L Yeast Extract (Oxoid)

RO water

For YPD agar plates, Agar Bacteriologial (Oxoid) was added to a final concentration of 24 g/L.

The mixture was autoclaved for 20 minutes at 120 ˚C.

The following ingredients were added to 500 mL of the mixture above in a sterile environment:

50 mL of 40 % (w/v) glucose (autoclaved separately),

200 mL of sterile RO water,

250 mL of 7% Tropic Marin® Sea Salt Classic (sterile filtrated separately).

The liquid was mixed.

2.4.4 Psi Medium

5 g/L Yeast Extract (Oxoid)

20 g/L Tryptone (Oxoid)

5 g/L $MgSO_4$

RO water

pH was adjusted to 7.6 using KOH. The medium was autoclaved for 20 minutes at 120 ˚C.

2.4.5 SOC Medium

20 g/L Tryptone (Oxoid)

5 g/L Yeast Extract (Oxoid)

0.5 g/L NaCl

2.5 mM KCl

3.6 g/L Glucose

5.08 g/L $MgCl_2*7H_2O$

RO water

pH was adjusted to 7.0 using NaOH. The medium was sterilized by filtration and frozen in aliquots of 1 mL.

2.4.5 TFB1 Buffer

30 mM K-Acetate

100 mM RbCl

10 mM $CaCl_2*2H_2O$

50 mM $MnCl_2*4H_2O$

15 % v/v Glycerol

RO water

pH was adjusted to 5.8 with dilute acetic acid before the solution was sterile filtered.

2.4.6 TFB2 Buffer

10 mM MOPS

75 mM $CaCl_2*2H_2O$

10 mM RbCl

15 % v/v Glycerol

RO water

pH was adjusted to 6.5 with dilute NaOH before the solution was sterile filtered.

2.4.7 TAE Buffer

242 g Tris(hydroxymethyl)aminomethane (Tris base)

57.1 mL 100% Acetic acid

100 mL 0.5 M EDTA (pH 8.0)

500 mL RO water

The ingredients were mixed and stirred until completely dissolved. RO water was added until the total volume was 1000 mL. The solution was autoclaved for 20 min at 120 ˚C. The solution was diluted to 1/50 with RO water before use.

2.4.8 Agarose Gels

0.8 % SeaKem® LE Agarose

TAE buffer

0.005% GelGreen or GelRed Nucleic Acid Stain (Biotium)

Agarose and TAE buffer were mixed and heated in a microwave oven until the agarose was dissolved. GelGreen or GelRed was then added, and the solution was stored in a heating

cabinet at 60 ˚C until used. Gels were made 30-60 minutes prior to use by pouring the hot solution in a cast with a comb.

2.4.9 FPLC Buffers

Buffer A – Binding buffer

0.5 M NaCl

45 mM Imidazole

50 mM MOPS buffer (pH 7.5)

RO water

The solution was sterile filtered and stored cold.

Buffer B – Elution buffer

0.5 M NaCl

0.5 M Imidazole

50 mM MOPS buffer (pH 7.5)

RO water

The solution was sterile filtered and stored cold.

2.4.10 Transfer Buffer

5.8 g/L Tris base

2.9 g/L Glycine

0.4 g/L Sodium dodecyl sulfate (SDS)

20 % Ethanol

RO water

pH was adjusted to 8.55 using concentrated HCl.

2.4.11 Tris-Buffered Saline with Tween-20 (TBST)

3.03 g/L Tris base

8.77 g/L NaCl

0.5 mL/L Tween-20

RO water

pH was adjusted to 6.98 using concentrated HCl.

2.4.12 Trace Metal Solution 1 (TMS 1)

5 g/L $FeSO_4*7H_2O$

390 mg/L $CuSO_4*5H_2O$

440 mg/L $ZnSO_4*7H_2O$

150 mg/L $MnSO_4*H_2O$

10 mg/L $Na_2MoO_4*2H_2O$

20 mg/L $CoCl_2*6H_2O$

50 mL/L Concentrated HCl

RO water

The solution was sterile filtered and stored at room temperature (RT).

2.4.13 Vitamin Mix 1

0.05 g/L Thiamine HCl

0.005 g/L B_{12}/Cobalmine

RO water

The solution was sterile filtered and stored at RT.

2.4.14 Media for Lipid Accumulation

40 g/L Glucose

0.7 g/L NH_4Cl

0.3 g/L Yeast Extract (Oxoid)

0.3 g/L KH_2PO_4

18 g/L Na_2SO_4

0.25 g/L $MgSO_4*7H_2O$

0.2 g/L $CaCl_2*2H_2O$

0.4 g/L KCl

6.1 g/L Tris-base

5.8 g/L Maleic acid

RO water

pH was adjusted to 7.0 before the media was sterile filtered.

Just before inoculation, the following components were added;

TMS 1 (5 mL/L),

Vitamin Mix 1 (1 mL/L).

2.5 Methods - Cloning

2.5.1 Growth of Microorganisms

The growth temperatures and shaking speeds for growth listed in this section are the standard conditions used. If other conditions were used, this will be stated along with the results.

2.5.1.1 Growth of *Aurantiochytrium* T66

Aurantiochytrium T66 was taken from a frozen glycerol stock and plated on an YPD agar plate. The plate was incubated at 22-25 ˚C for ≥ 2 days. Colonies from the plate were then incubated in liquid YPD media at 25 ˚C and 170 rpm over night (ON).

2.5.1.2 Growth of *E. coli*

E. coli was plated on LA agar plates. The plates were incubated at 37 ˚C ON. A colony from a plate was then incubated in liquid LA medium at 37˚C at 225 rpm ON.

2.5.1.3 Growth of *P. putida*

P. putida KT2440 was plated on LA agar plates. The plates were incubated at 30 ˚C ON. A colony from a plate was then incubated in liquid LA medium at 30˚C at 225 rpm ON.

2.5.2 Glycerol Stocks

Glycerol stocks were prepared by mixing 800 µL of a culture broth with 400 µL autoclaved 60% glycerol in a cryo tube. The tube was then frozen at -80 ˚C.

2.5.3 Isolation of Genomic DNA from T66

The isolation of genomic DNA from T66 was performed using a MasterPure™ Complete DNA and RNA Purification Kit from Epicentre Biotechnologies.

0.5 mL T66 culture that had been incubated ON in a liquid YPD medium were centrifuged at RT for 60 seconds at 13,000 rpm. The pellet was resuspended by 10 seconds of vortexing, followed by pipetting up and down. 1 µL of Proteinase K was added to 300 µL of Tissue and Cell Lysis Solution. The Tissue and Cell Lysis Solution was designed to rupture the cells, while the Proteinase K is an enzyme that should degrade the protein in the cell lysate. 300 µL of this mixture were added to the resuspended pellet before mixing by pipetting. The sample was incubated at 65 ˚C for 15 minutes, during which time it was vortexed every 5 minutes. It was then cooled to 37 ˚C by incubation at this temperature for 5 minutes. 1 µL of RNase A was added to the sample before mixing thoroughly. The purpose of this was to degrade the RNA present in the sample. The sample was incubated at 37 ˚C for 30 minutes.

The sample was placed on ice for approximately 5 minutes. 150 µL of MCP Protein Precipitation Reagent were added to the lysed sample. This reagent was designed to precipitate the protein by desalting it. The sample was vortexed for 10 seconds. This was then centrifuged at 4 ˚C at 10,000 xg for 10 minutes. The supernatant was transferred to a new tube, and the pellet containing the protein discarded. 500 µL of isopropanol was added to the supernatant before it was inverted approximately 40 times. The DNA was then pelleted by centrifugation at 4 ˚C at 10,000 xg for 10 minutes. The isopropanol was removed by decantation. The pellet was washed twice with 70 % ethanol. After decanting the ethanol off, the pellet was left to dry at RT for about 1 hour. The DNA was then resuspended by the addition of 50 µL of TE buffer followed by ON incubation at 30 ˚C in a shaker. The sample was then frozen.

2.5.4 Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR) is a method used to amplify a specific sequence of DNA (the target sequence) from the original DNA sample (the template). Primers are made to match each end of the target sequence. These are usually between 15 and 20 bp long and may include special restriction sites to facilitate the ligation of the products into plasmids. A heat stable DNA polymerase is used to perform the elongation of the DNA from the primers by adding new nucleotides according to the template sequence. PCR machines, or more appropriately thermocyclers, are used to control temperature during the reaction. The thermocycler first heats the template DNA to about 90 ˚C to separate the two DNA strands of the double helix from each other. The temperature is then lowered to allow the primers to bind to each end of the target sequence. This temperature depends on the melting temperature of the primers, but are usually around 50-60 ˚C. The temperature is then increased to the

optimum temperature of the DNA polymerase (around 70 ˚C) and kept there for the time needed for the polymerase to synthesize the length of the target sequence. The cycle then starts over again. For each cycle the number of copies of the target sequence will increase exponentially (Clark and Pazdernik, 2013b).

2.5.4.1 Preparation of PCR Samples

25.5 µL autoclaved RO water

10 µL Q5 5X Reaction Buffer

10 µL Q5 5X High GC Enhancer

1 µL dNTP (10 mM)

0.5 µL purified genomic DNA or purified plasmid

1.25 µL forward primer (20 pmol/ µL)

1.25 µL reverse primer (20 pmol/ µL)

0.5 µL Q5® High-Fidelity DNA polymerase

The Q5 related compounds as well as the nucleotide mixture was provided by New England Biolabs, Inc (NEB).

The ingredients above were added to a PCR tube to give a total volume of 50 µL. The tubes were kept on ice during the preparation. The solution was carefully mixed and spun down before running the PCR.

2.5.4.2 PCR Conditions

The PCR reaction was set up using a thermocycler from Eppendorf. The initial denaturation temperature was 98 C˚ for 3 minutes. 35 cycles of 10 s denaturation at 98 C˚, 15 s of primer annealing at the temperatures indicated in Section 3.3.1, and elongation at 72 C˚ for the durations indicated in Section 3.3.1 were run. Finally, an elongation at 72 C˚ for 8.5 min was performed.

2.5.5 Agarose Gels

Agarose gels are made of agarose, a polysaccharide from red algae. These gels are used to separate DNA fragments according to size. Each nucleotide in a DNA fragment will have a negative charge due to the phosphate group. When DNA molecules are loaded in wells on one side of the gel and an electric current is sent through the gel, the DNA will move through the

gel towards the positive pole. The larger the fragment, the slower it will move, due to the impediment caused by the gel matrix. The size of the fragments may be evaluated by the loading of a marker containing fragments of known size (Alberts et al., 2015).

A 0.8 % agarose gel was made as described in Section 2.4.8. The gel was placed in a tray with TAE buffer (Section 2.4.7), and the DNA samples were loaded in the wells. A GelDoc station from BioRad was connected to provide the electric field. A strength of a 100 V was used, and the gel was run until sufficient separation was achieved. A Molecular Imager® ChemiDoc™ XRS+ Imaging System from BioRad was used to develop the image.

2.5.6 Purifying DNA from an Agarose Gel

The purification was performed using a Monarch DNA Gel Extraction kit from NEB.

All centrifugation steps were carried out at 13,000 rpm on a 120 mm diameter centrifuge or at 16,000 xg.

The DNA fragment was excised from the agarose gel and transferred to a 1.5 mL Eppendorf tube. 4 times the gel volume of Gel Dissolving Buffer was added to the tube. This was incubated at 54 ˚C, while periodically vortexing, until the gel was dissolved.

The sample was loaded onto a column in a collection tube. This was centrifuged for 1 minute to insert the DNA in the filter of the column and to remove the liquid by discarding the flow-through.

The column was reinserted into the collection tube, and 200 µL of DNA Wash Buffer was added to wash the sample. This was centrifuged for 1 minute before the flow-through was discarded. The washing with DNA Wash Buffer followed by centrifugation was repeated.

The column was transferred to an Eppendorf tube, and 6-20 µL of DNA Elution Buffer was added to the filter. This was incubated at RT for 1 minute and then centrifuged for 1 minute to elute the DNA from the filter. The column was discarded, and the concentration of the eluted DNA was measured.

2.5.7 PCR Product Purification

This protocol was performed using the description, buffers and spin columns provided with the Monarch PCR & DNA Cleanup kit.

All centrifugations were carried out at 16,000 xg.

The PCR sample was diluted in a 5:1 ratio of buffer to sample with DNA Cleanup Binding Buffer. The solution was mixed thoroughly by pipetting up and down. The sample was then loaded onto a column in a collection tube and then centrifuged for 60 seconds. This step left the DNA in the column, while the liquid was discarded. 200 µL DNA Wash Buffer was then added to the column, which was then centrifuged for 1 minute to remove non-DNA molecules that might be in the column. This step was repeated to ensure a pure product. The column was then transferred to a sterile Eppendorf tube. 20 µL of DNA Elution buffer was then added to the center of the column. This was incubated for 1 minute at RT before centrifugation for 1 minute to elute the DNA.

2.5.8 TOPO® Cloning

TOPO® cloning was performed using a Zero Blunt® TOPO® PCR Cloning kit manufactured by Invitrogen.

The principle for this cloning was that the plasmid supplied with the kit was linear with a Topoisomerase I enzyme bound to each 3' end of the molecule by a phospho-tyrosyl bond. This bond may be broken by the reaction with a 5' hydroxyl group, which will then be bound to the DNA molecule. When another linear DNA molecule is added to the solution, each 5' hydroxyl group might remove a Topoisomerase I enzyme. The result will thus be a circular plasmid composed of the TOPO® vector and the fragment added (*Zero Blunt® TOPO® PCR Cloning Kit User Guide*, 2014).

2 µL PCR product, 0.5 µL Zero Blunt® TOPO® PCR Cloning Salt Solution and 0.5 µL pCR™II-Blunt-TOPO® vector was carefully mixed in an Eppendorf tube and incubated for 5 minutes at RT. The mixture was then transformed into competent *E. coli* cells (see section 2.5.10).

2.5.9 Making Chemically Competent *E. coli* Cells

Transformation, the introduction of foreign DNA into a bacterial cell, mainly consists of three steps. The first step is a preparation step, where the cells are prepared (usually weakened) in such a way that they will more easily allow DNA to enter their cytoplasm. The cells are then shocked in a way that allows DNA to cross their membrane. Finally, they are allowed to recover in a rich medium. In the case of chemical transformation, the cations are the main active components in the preparation step. Divalent cations such as Mn^{2+} and Ca^{2+} have been shown to promote transformation. The effectiveness of the transformation may be further increased by including monovalent cations such as Cs^{+}, Na^{+}, and Rb^{+} and using a slightly

acidic buffer. The actual transformation in chemical transformation is commonly achieved by giving the cells a heat-shock (Aune and Aachmann, 2010).

E. coli cells were inoculated from a frozen glycerol stock in 10 mL Psi-medium. This was incubated at 225 rpm at 37 ˚C ON.

1 mL from the ON culture was inoculated in 100 mL Psi-medium in a shake flask (1% inoculate). This was incubated at 37 ˚C at 225 rpm until OD_{600} was 0.40-0.43. The culture was then incubated on ice for 15 minutes. The cells were pelleted by centrifugation at 4000 rpm at 4 ˚C for 5 minutes in a Falcon-tube. The supernatant was removed by decantation.

The preparation step was then performed by the washing with different buffers. First, the cells were resuspended in 40 mL cold TFB1 by carefully rotating the tube manually. The solution was incubated on ice for 5 minutes. The solution was centrifuged at 4000 rpm at 4 ˚C for 5 minutes. The cells were resuspended in 3 mL cold TFB2 by carefully rotating the tube manually.

The solution was then aliquoted in Eppendorf tubes by adding 100 µL to each tube while keeping both the tubes and the solution on ice. The cells were then frozen immediately on liquid nitrogen and stored at -80 ˚C.

2.5.10 Heat-Shock Transformation of Chemically Competent Cells

The competent cells (see Section 2.5.9) were thawed on ice. 10 µL ligation mix, 2 µL purified plasmid, or 3 µL of TOPO® cloning mixture was added to 100 µL cells, and this was mixed by careful pipetting. The mixture was kept on ice for 30-60 minutes. The cells were then heat-shocked at 37 ˚C for 2 minutes. Afterwards, they were incubated on ice for 2 minutes. 900 µL of 37 ˚C SOC medium was added. This was incubated at 225 rpm at 37 ˚C for 1-2 hours. 200 µL of the mixture was plated on an LA plate with a selective antibiotic for the specific plasmid. The rest of the mixture was plated on another plate of the same kind, as a safety precaution in case of low efficiency of the transformation. A negative control was made in the same way, but with addition sterile RO water instead of DNA. The plates were incubated at 37 ˚C ON.

2.5.11 Making Electrocompetent *E. coli* Cells

Electroporation is another technique for introducing foreign DNA into bacterial cells. Even though the precise mechanism underlying this method is not completely understood, the main explanation is that giving the cells a right amount of electrical shock will cause formation of

temporary pores. DNA may then enter the cells through these pores. During electroporation, it is crucial to wash the cells properly before introducing the electrical shock to avoid any ions disturbing the electrical field (Aune and Aachmann, 2010).

E. coli cells from a frozen glycerol stock was inoculated in 3 mL LB at 37 ˚C and 225 rpm ON. 1 mL of the ON culture was then added to 200 mL of fresh LB in a baffled shaking flask to give ~1:200 dilution. These cells were incubated at 37 ˚C and 225 rpm until the OD_{600} reached ~0.5. The culture was cooled on ice for 10-15 minutes before it was transferred to cool 250 mL centrifuge bottles. The culture was spun at 5000 rpm at 4 ˚C for 8 minutes.

The supernatant was discarded, and the cell pellet was resuspended in 200 mL cold sterile RO water. The solution was spun at 5000 rpm at 4 ˚C for 8 minutes. The cell pellet was resuspended in 100 mL cold sterile RO water and spun at 5000 rpm at 4 ˚C for 12 minutes.

The cell pellet was resuspended in 4 mL cold sterile 10% glycerol and transferred to a 38 mL centrifuge flask. The cells were spun for 14 minutes at 7830 rpm at 4 ˚C, and the supernatant was discarded.

The cells were resuspended in 0.8 mL cold sterile 10% glycerol and aliquoted into cold sterile Eppendorf tubes with 40 µL in each. The aliquots were flash frozen by incubation for 10 seconds in liquid nitrogen. The cells were kept at -80 ˚C.

2.5.12 Electroporation of Electrocompetent *E. coli* Cells

A 40 µL aliquot of electrocompetent cells was thawed on ice. 2 µL plasmid was mixed with the cells before incubation for ≥ 1 minute on ice. The solution was then transferred to a cold 0.2 cm electroporation cuvette. This was inserted into a Gene Pulser Xcell™ Electroporation System. The cells were then given an electrical pulse using the Bacterial 2 program, which gave a pulse of 2500 V and 200 Ω for about 5 milliseconds. 1 mL of SOC preheated to RT was added to the cells immediately after the pulse. This solution was mixed and transferred to a 1.5 mL Eppendorf tube. This was incubated at 37 ˚C and 225 rpm for 1 hour, and then the solution was plated on LA plates containing a selective antibiotic for the plasmid.

2.5.13 Plasmid Purification

Two different plasmid purification kits with individual protocols were used during this thesis, depending on which kit was in the storage of the group at the time of each experiment. Both kits were assumed to be of high quality. The two kits provided plasmid preps of similar quality.

2.5.13.1 Zymo Plasmid Purification Protocol

This protocol was performed using the description, buffers and spin columns provided with the ZR Plasmid Miniprep™ – Classic kit.

All centrifugations were carried out at 11,000-16,000 xg.

1-5 mL liquid bacteria culture was added to a 1.5 mL Eppendorf tube or a 15 mL Falcon tube and spun for 15 seconds. The supernatant was then removed. The pellet was resuspended in 200 µL P1 buffer by pipetting. 200 µL P2 Buffer was added, and the solution was carefully mixed by inversion before 1 minute of incubation at RT. 400 µL P3 Buffer was then added, and the solution was mixed carefully by inverting until the pink color of buffer P2 disappeared. The combination of buffers P1, P2, and P3 is an alkaline lysis procedure intended to rupture the cells. The sample was incubated at 1 minute before it was spun for 3 minutes to precipitate protein and other large components of the cell.

The supernatant was transferred to a spin column in a collection tube, which was spun for 30 seconds. This was performed to leave the DNA in the spin column filter, while the liquid was removed in the flow-through and then discarded. 200 µL of Endo-Wash Buffer was added to the column before it was spun for 30 seconds to wash the DNA. 400 µL of Plasmid Wash Buffer was added before centrifugation for 1 minute to wash the DNA sample once more. The column was transferred to a sterile 1.5 mL Eppendorf tube. 35 µL of DNA Elution Buffer was then added to the column, which was spun for 30 seconds to elute the DNA.

2.5.13.2 Monarch Plasmid Purification Protocol

Alternatively, plasmids were purified using a Monarch Plasmid DNA Miniprep kit from NEB.

1.5-5 mL of liquid *E. coli* culture grown over night at 37 ˚C in LB medium was centrifuged for 1 minute at 22 ˚C and 7830 rpm or 30 seconds at RT and 16,000 xg. The supernatant was discarded, and the pellet resuspended in 200 µL Plasmid Resuspension Buffer. This was pipetted until the cells were completely resuspended. 200 µL of Plasmid Lysis Buffer were added to lyse the cells. The tube was immediately and gently inverted a few times until the color of the entire solution was pink. This was incubated for 1 minute at RT. The lysate was neutralized by the addition of 400 µL of Plasmid Neutralization Buffer. The tube was gently inverted a few times before it was incubated for 2 minutes at RT. The mixture was then centrifuged at 23 ˚C at 13,000 rpm or 16,000 xg for 5 minutes.

The supernatant was carefully transferred to a spin column, which was centrifuged for 1 minute at 13,000 rpm or 16,000 xg at 23 ˚C. The flow-through was discarded, and the column reinserted. 200 µL of Plasmid Wash Buffer 1 was added to the column, which was centrifuged for 1 minute at 23 ˚C at 13,000 rpm or 16,000 gx. 400 µL of Plasmid Wash Buffer 2 was then added to the column, followed by 1 minute of centrifugation at 23 ˚C at 13,000 rpm or 16,000 xg. The column was transferred to a clean microfuge tube. 35-50 µL DNA Elution Buffer was added to the center of the column matrix. This was incubated for 1 minute at RT, followed by centrifugation for 1 minute at 23 ˚C at 13,000 rpm or 16,000 xg.

2.5.14 Restriction Reactions

All enzymes and buffers used for restrictions were provided by NEB.

DNA and autoclaved RO water was mixed to a total volume of 17 µL. 2 µL of buffer adapted to the enzymes used were added to the solution. 0.5 µL of each enzyme was then added. The mixture was then incubated for ≥ 1 h at a temperature suitable for the enzymes in question.

2.5.15 Klenow Reactions

The Klenow fragment of the DNA polymerase enzyme and the nucleotide mixture was provided by NEB.

The Klenow fragment is an enzyme used to create blunt ends on a DNA fragment. 0.5 µL of Klenow enzyme and 0.5 µL of 10 mM dNTP mix was added to a restriction reaction mixture. This was incubated at 37 ˚C for 15 minutes.

2.5.16 Ligation

The enzyme and the buffer were provided by NEB.

Backbone DNA and insert DNA were mixed with water in amounts giving a 1:3 molar concentration of backbone to insert. This mixture was adjusted to a volume of 17 µL with sterile RO water. 2 µL 10 X T4 ligase buffer and 1 µL of T4 ligase was added to the mixture. The reaction was incubated at 16 ˚C ON.

2.5.17 Electroporation of *P. putida*

P. putida KT2440 from either a plate or a glycerol stock was inoculated in 3 mL LB at 30 ˚C and 225 rpm ON. 200 µL of the over night culture was added to 20 mL LB. This was incubated at 30 ˚C and 225 rpm until OD_{540} reached 0.5 ± 0.1. The culture was cooled on ice for 30 min, before it was centrifuged at 4000 xg for 15 min at 4 ˚C. The pellet was washed by

dissolving it twice in 20 mL and once in 10 mL 300 mM cold sucrose before centrifugation as before. The washed pellet was then dissolved in 200 µL cold 10% glycerol.

Aliquots of 40 µL cells was mixed with 4 µL plasmid before incubation on ice for 30 min. The cells were then electroporated using a Gene Pulser Xcell Electroporation System (Bio-Rad) and 2 mm cold cuvettes. The cells were electroporated at 2,5 kV, 25 µF and 200 Ω. Immediately after electroporation, 900 µL SOC medium at RT was added. The cells were incubated at 30 ˚C and 225 rpm for 1.5 – 2 h before they were plated on agar plates with a selective antibiotic for the plasmid.

2.6 Methods – Protein Expression

2.6.1 Protein Production and Purification from *E. coli*

Bacterial cells containing a plasmid with the gene of interest from either a plate of a frozen glycerol stock were inoculated in 10 mL 3XLB with a selective antibiotic at 37 ˚C and 225 rpm ON. 2 mL of this culture was then inoculated in 200 mL 3XLB with a selective antibiotic for 3-4 hours at 37 ˚C and 225 rpm. OD_{600} was measured and the expression of the gene of interest was induced by the addition of 400 µL 0.25 M m-toluate in 50 % ethanol (0.5 mM in culture solution). The cells were then grown for 4 hours. Volume and OD_{600} of the culture were measured. The cells were then spun down by centrifugation for 10 min at 7000 rpm. The supernatant was discarded, and the cell pellet was frozen at -20 ˚C.

The cell pellet was dissolved in 20 mL cold 50 mM MOPS with pH 7.5. From this point, the cells were kept on ice. A FisherBrand Model 505 Sonic Dismembrator from Fisher Scientific delivering 500 W with a 3.2 mm tip was used to sonicate the cells. The cells were sonicated for 3x 2 min (20 s pulse, 30 s pause) with an amplitude of 20 %. The sonicated solution was centrifuged for 30 min at 30,000xg at 4 ˚C. The supernatant was filtered with a Filtropur® syringe filter of 0.20 µm from Sarstedt.

The proteins of interest were separated from the other proteins in the lysate using IMAC (Section 1.5.4). The chromatographs used were ÄKTA Start from GE Healthcare Life Sciences and ÄKTA P-920 with an UPC-900 detector from Amersham Biosciences. The mobile phases are described in Section 2.4.9. A 1mL His-Trap™ HP column with Ni^{2+} ions was used.

The system was rinsed with sterile filtered RO water, buffer B, and buffer A prior to separation. A flow rate of 1 column volume (CV) per min was used during the procedure. An

equilibration with buffer A was performed using 5 CV of buffer. The sample was then loaded on to the column, which was subsequently washed with 5 CV to elute unbound protein. 20 CV of either an equal mix of buffer A and B or of a gradient from 100% buffer A to 100% buffer B were used to elute the sample. 20 CV of buffer B was then run thorough the column to elute eventual remaining protein. The column was then equilibrated with 5 CV of buffer A before either loading the next sample or doing a final wash with water and 20 % ethanol.

2.6.2 Lysis of Cells with B-PER

The OD_{600} and volume of the cultures to be lysed were measured, and the cells were pelleted by centrifugation at 7000 rpm for 10 min. The pellets were then either frozen or kept on ice until they were dissolved in a volume of 0.9 % NaCl to get a concentration of 3 g cell dry weight (CDW) per L. This was calculated by using the relationship CDW (g/L) = 0.36 * OD_{600} (Ren et al., 2013).

1 mL of each of the dissolved pellets were then used to make a soluble and an insoluble protein fraction with B-PER™ Complete Bacterial Protein Extraction Reagent from Thermo Scientific using the protocol provided with the product.

The cells were pelleted by centrifugation at 5000 xg for 10 min at RT. 65.4 µL B-PER Complete was used to dissolve the pellet, which was subsequently incubated for 15 min at gentle shaking (~400 rpm) at RT. The samples were then centrifuged for 20 min at 16,000xg at RT. The supernatant was transferred to a new tube, and the pellet was dissolved in 65.4 µL sterile RO water. This resulted in the two fractions used further; the soluble fraction (supernatant) and the insoluble fraction (pellet).

2.6.3 SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Proteins may be separated based on molecular weight much like DNA is separated on agarose gels. However, proteins may have both positive, negative, and no charge. Boiling protein samples in SDS will give the proteins a negative charge proportional to their size, because the hydrophobic tails of SDS will coil around the peptide backbones with their negative head groups pointing out. β-mercaptoethanol or similar substances may be added to destroy disulfide bonds stabilizing tertiary structures or holding together subunits. The SDS treated samples may then be separated on a gel using an electric field like the one used on agarose gels, where the smallest proteins will move the fastest towards the positive pole. Due to the smaller sizes of proteins than of DNA, polyacrylamide gels are used. The proteins may be visualized with compounds such as Coomassie Blue (Clark and Pazdernik, 2013c).

20 µL of protein sample was mixed with 10 µL 3X SDS Sample Buffer from NEB and 1 µL 1 M DTT. This solution was boiled for 10 min at 100 ˚C. 10 µL of each sample was then loaded onto a RunBlue SDS Gel 4-12% from Expedeon. 3 µL Prestained Protein Ladder, Broad Range from NEB were used (see Appendix A – Ladders). The gel was stained using InstantBlue staining from Expedeon. A Molecular Imager® ChemiDoc™ XRS+ Imaging System from BioRad was used to develop the image.

2.6.4 Western Blot

Western Blotting is a technique used to identify a specific protein among many proteins using the specific binding properties of antibodies. The proteins are first separated based on size using SDS-PAGE. The proteins are then transferred to a membrane. The membrane is incubated with a specific antibody binding to the protein of interest. A detection system coupled to the antibody, either directly to the primary antibody or to a secondary antibody, is used to identify the presence of the protein of interest (Clark and Pazdernik, 2013c).

Before the Western blot procedure, the proteins were separated by SDS-PAGE as described in Section 2.6.3. The protein gel was removed from the plastic covers and overlaid by a nitrocellulose membrane (0.45 micron) from Thermo Scientific. The membrane had been saturated with transfer buffer (Section 2.4.10) for 10 minutes. On both sides of the gel and membrane, double Whatman™ filter paper (grade 3 mm, CHR) from GE Healthcare was laid. The transfer from the gel to the membrane was made in a Dual Cool Electrophoresis System DCX-700 from C.B.S. Scientific. The system was set up as described in the protocol following the transfer system. Approximately 1 L of cold transfer buffer was added, along with a stirring magnet and frozen cooling elements. The transfer was performed at 80 V for 50 minutes.

The following membrane incubations and washings were performed with shaking at RT with ~20-30 mL liquid. The membrane was saturated with a blocking buffer (3% Skim Milk Powder from Fluka Analytical in TBST, see Section 2.4.11) for 1 hour. Subsequently, the membrane was washed twice with TBST for 10 minutes. The antibody was attached by 1 hour of incubation with a HisProbe-HRP Working solution (1:5000 dilution of the 4 mg/mL stock solution of HisProbe™-HRP from Thermo Scientific). The membrane was then washed 3 times with TBST. 1 mL of SuperSignal™ West Pico Chemiluminescent Substrate Working Solution (50% Luminol/Enhancer Solution and 50% Stable Peroxide Buffer) from Thermo Fisher was carefully spread out on the membrane. A clear plastic sheet was used to keep the

membrane moist during 5 min of incubation and subsequent development of the image. The image taken after 300 s of exposure was used for evaluation. A Molecular Imager® ChemiDoc™ XRS+ Imaging System from BioRad was used.

2.7 Methods – Cultivation of Thraustochytrids

2.7.1 Fermentation Setup

Aurantiochytrium T66 and *Schizochytrium* S21 (ATCC 20 888) from a YPD plate or a glycerol stock, respectively, were plated on YPD agar plates with 200 µg/mL ampicillin and streptomycin. These were left at RT for a few days until clearly visible colonies appeared. Cells from these plates were then transferred to 50 mL YPD media with 200 µg/mL ampicillin and streptomycin in 250 mL baffled shake flasks. The flasks were incubated at 25 ˚C and 170 rpm for ~23 h.

4% of the T66 YPD-culture (OD_{600}=0.860) or 2% of the S21 YPD-culture (OD_{600}=1.728) were used as inoculum for the fermentation cultures. The fermentations were set up as three biological parallels with 200 mL lipid accumulation media (Section 2.4.14) with 5 mL/L TMS1 (Section 2.4.12) and 1 mL/L Vitamin Mix 1 (Section 2.4.13) in 1 L baffled shake flasks. The fermentation was carried out at 25 ˚C and 170 rpm. Samples taken during the fermentations are further described below.

2.7.2 OD_{600} Measurements

Optical density at 600 nm were measured in order to monitor the growth of the cultures. The OD_{600} were measured using a spectrophotometer. Samples were vigorously vortexed immediately before measurements were taken. RO water was used to dilute the samples and as a blank. This was considered acceptable, since RO water had an OD_{600} of -0.008 when the spectrophotometer was blanked with the lipid accumulation media containing the TMS1 and Vitamin Mix 1.

2.7.3 Nitrogen Samples

1 mL sample was taken out from each biological parallel. These samples were centrifuged at 16,000xg for 1 min. The centrifugation time was increased to 5 min for the last 3 sample points due to apparent high viscosity of the supernatant. The samples were stored at -20 ˚C.

Ideally, a kit to measure the ammonium content in the samples should have been used to analyze the samples. However, the delivery of the kit was delayed, and the measurements could thus not be performed.

2.7.4 The Bradford Method

The protein concentration of a solution may be evaluated by using the dye Coomassie Brilliant Blue G-250. The absorption maximum of the dye is usually at 465 nm, but upon binding to protein it shifts to 595 nm. The amount of absorbance at 595 nm is linearly dependent on the amount of protein bound to the dye (Bradford, 1976). This method is commonly referred to as the Bradford Method.

The Bradford assay was performed using a Coomassie (Bradford) Protein Assay Kit (Number 23200) from Thermo Scientific and the protocol supplied with the product.

5 µL of each sample was added to a well in a clear 96-well microplate. 250 µL Coomassie Reagent was added to the well, and the sample and the reagent were mixed by pipetting. The sample was incubated for ~10 min before the absorption at 595 nm was measured using a SpectraMax Plus 384 spectrophotometer from Molecular Devices. Three parallels were performed for each sample, including the blank sample (only buffer). The absorbance was calculated as the average absorbance of the three parallels minus the average absorbance of the blank samples. These values were compared with a standard curve to estimate the concentration of the samples.

See Appendix B – Bradford Standard Curve for details about the standard curve.

2.7.5 Enzyme Samples

8 mL sample was taken out from each fermentation flask. The samples were centrifuged at 7197 xg for 15 min at 4 ˚C. The cells were washed by dissolving them in 8 mL 0.1 M TrisHCl (pH 7.4) by vortexing and manual shaking. The cells were pelleted by spinning at 7197 xg for 10 min at 4 ˚C. The washing step was repeated. The final pellet was dissolved in 8 mL 0.1 M TrisHCl (pH 7.4). The solution was passed through a 0.4 mm*19mm Microlance^{TM} 3 syringe from Becton Dickinson to remove the largest cell aggregates. The solution was kept on ice until sonication.

Sonication Test

Before the first enzyme samples were taken out, a sonication test was performed. The samples were washed and dissolved in Tris-HCl buffer as described above.

A FisherBrand Model 505 Sonic Dismembrator from Fisher Scientific delivering 500 W was used with a 3.2 mm tip. The samples were sonicated at 40% amplitude (max. for the tip chosen) with 4 s pulse followed by 10 s rest. The samples were homogenized for a longer

period, while 100 µL protein sample were taken out periodically. The protein samples were centrifuged at 5000 xg for 15 min at 4 ˚C. The protein content of the supernatant was then evaluated with the Bradford Method (Section 2.7.4).

Sonication

The same sonicator and tip as in the test was used for the sonication of the samples. Due to difficulties lysing the cells, the samples were sonicated at different conditions. These are described in the result section.

Centrifugation of Lysate

The lysate of the cells from the sonication was centrifuged for 12 min at 16 000xg at 4 ˚C. The supernatant was then transferred to a new tube. For the 2nd set of enzyme samples, the supernatant was filtered through a 0.45 µm Filtropur® filter from Sarstedt. The samples were then stored at 4 ˚C until the assays described in Section 2.8 were performed within 28 hours of the initial sampling.

2.7.6 Total Lipid Samples

10 mL sample was taken out and centrifuged at 5000 xg for 5 minutes at 4 ˚C. The pellet was washed by dissolving it in 10 mL 0.9% NaCl before spinning it at 5000 xg for 10 minutes at 4 ˚C. The pellet was then washed with 10 mL StakPure filtered water and spun down at 5000 xg for 10 minutes at 4 ˚C. The pellet was flash frozen by submersion in liquid N_2 for ~10s. The pellet was then stored at -20 ˚C.

The frozen samples were lyophilized with a VirTis BenchTop Pro for ~24 h at 12-19 mTorr and -103 ˚C. ~20 mg of each freeze-dried sample was transferred to a 2 mL Precellys vial with sleeve and caps with O-rings. These vials also contained ~0.5 g 1.4 mm Zirconium oxide beads from Bertin Technologies. 1000 µL of 2:1 methanol and chloroform were added to the vials. The samples were homogenized for 3*30 s at 6500 rpm with 15 s pause while cooling with N_2 on a Precellys 24 by Bertin Technologies. 333 µL chloroform was added, and the samples were vortexed for ~20 s. Subsequently, 333 µL RO water was added before ~20 s vortexing. The samples were then centrifuged for 14,000 rpm for 5 min at ~15 ˚C. 600 µL of the lipid faze were transferred to a pre-weighed 1.5 mL transparent HPLC vial. The samples were left in a fume hood for ~22 h in order to let the chloroform evaporate. The vials were then weighed again. The lipid yield was calculated based on the assumption that the total chloroform volume was 666.3 µL.

2.8 Methods - Enzyme Assay Protocols

2.8.1 Glycerol 3-P Dehydrogenase (G3PDH) Assay

This protocol is based on the protocol described by Blomberg and Adler (1989).

Reaction Solution Reagents

20 mM imidazole-HCl (pH 7.0)

1 mM 1,4-dithiothreitol (DTT) (freshly prepared)

1 mM $MgCl_2$

0.67 mM dihydroxyacetone phosphate (DHAP) (lithium salt)

0.09 mM β-NADH (disodium salt hydrate) or β-NADPH (reduced tetrasodium salt hydrate)

Protocol

Two master mixes with imidazole-HCl, DTT, $MgCl_2$ and either NADH or NADPH were made. 140 µL were added to each well in a clear 96-well microplate. 50 µL protein sample were added to each well. Three parallels were run for each sample. The reaction was started by the addition of 10 µL 0.01 M DHAP. The mixture was briefly mixed by pipetting, and absorbance measuring was started within 10 s. The absorbance was measured at 340 nm every 10th second for 10 minutes.

Additional Theory

DTT and 1,4-dithioerythritol (DTE), usually referred to as Cleland's reagent, are compounds having a reducing potential that is often exploited when buffers are made. These compounds may reduce free disulfides but are not able to reach disulfides imbedded within a protein, which is where native disulfide bonds are usually found. These compounds may thus ensure that the structure of a protein remains intact, even in a solution rich in oxygen (Hermanson, 2013).

2.8.2 ATP-Dependent Citrate Lyase (ACL) Assay

This protocol is based on the protocols described by Chávez-Cabrera et al. (2015) and Chen et al. (2014).

Reaction Solution Reagents

200 mM Tris-HCl (pH 7.4)

0.2 mM Coenzyme A (sodium salt hydrate)

0.1 M β-NADH (reduced disodium salt hydrate)

10 mM $MgCl_2$

10 mM DTT (freshly prepared)

20 mM Sodium Citrate

5 U/mL L-Malate dehydrogenase (MDH) (EC: 1.1.1.37) from pig heart (mitochondrial) from Roche

10 mM ATP (disodium salt)

Protocol

A master mix with Tris-HCl, CoA, NADH, $MgCl_2$, DTT, sodium citrate and MDH was made. 133 µL were added to each well in a clear 96-well microplate. 50 µL protein sample were added to each well. Three parallels were run for each sample. The reaction was started by the addition of 16.6 µL 0.091 M ATP. The mixture was briefly mixed by pipetting, and absorbance measuring was started within 10 s. The absorbance was measured at 340 nm every 10th second for 10 minutes.

Additional Theory

The activity of ACL can be determined by measuring the NADH consumption of the coupled reaction with the malate dehydrogenase (MDH) enzyme (Chávez-Cabrera et al., 2015). When ACL breaks down citrate, it produces oxaloacetate in a 1:1 stoichiometric relationship (Srere, 1959). The oxaloacetate (OAA) may then be transformed by MDH, which needs NADH as a cofactor in a 1:1 stoichiometric relationship relative to OAA. The amount of NADH consumed will then be directly proportional to the citrate made by ACL (Chávez-Cabrera et al., 2015). This theory assumes that these reactions are so fast that the time of the coupling is negligible.

3 Results

3.1 Bioinformatic Research

A heterologous expression of the ORFs of interest was to be performed. However, before the experimental part was initiated, it was of interest to investigate analogs to the proteins encoded by the ORFs of interest. The purpose of this was both to be able to better plan the experimental work, but also to better know what to expect as experimental results.

3.1.1 BLAST

As part of the preliminary research, a BLASTp was performed for the letter sequence of the proteins. The sequences were compared with entries in the protein data base (PDB). The best match for each protein where a publication on the protein structure identified existed was used for comparison (Table 3.1).

The products of gene 6817 and 10760 matched with the α- and β-units of succinate-CoA ligase, respectively. This further supported the interest to investigate whether these genes might encode ACL, as this coincided with the annotation previously done by the Auromega group. Gene 9329 had the best match with malyl-CoA lyase, which is an enzyme of the citrate lyase-like family (González et al., 2017). These enzymes catalyze similar reactions, and it is thus possible that these enzymes might be confused by annotation tools.

The product of gene 8081 had the best match with FAD-dependent alpha-glycerophosphate oxidase, while the product gene 6468 had the best match with FAD-dependent glycerol 3-phosphate dehydrogenase. Since the sequences of these two enzymes have been reported to be similar (Colussi et al., 2008), they could have been confused during annotation.

Table 3.1: ***Results from BLASTp of genes of interest****. The letter sequence of the proteins encoded by the genes of interest were entered in the BLASTp search tool (Altschul et al., 1997), where they were compared to proteins from the Protein Data Base. The best match of proteins where the structures are published are reported in the table. Their PDB accession code, the query cover, the percent identity and the E-value (statistical significance) are reported.*

Gene	Match	PDB accession	Reference	Query cover	Percent identity	E-value
6817	Chain A, Succinyl-CoA Synthetase [*Escherichia coli*]	1CQI_A	Joyce et al. (2000)	98%	64.18%	1e-98
10760	Chain B, Succinyl-CoA ligase [GDP-forming], mitochondrial [*Sus scrofa*]	5CAE_B	Huang and Fraser (2016)	96%	52.35%	4e-103
9329	Chain A, Malyl-CoA Lyase [*Methylorubrum extorquens* AM1]	5UGR_A	González et al. (2017)	94%	31.82%	4e-26
8081	Chain A, Alpha-Glycerophosphate Oxidase [*Streptococcus sp.*]	2RGH_A	Colussi et al. (2008)	81%	27.33%	2e-33
6468	Chain A, Glycerol-3-Phosphate Dehydrogenase [*Bacillus halodurans*]	3DA1_A	Colussi et al. (2008)	92%	30.43%	1e-49

3.1.2 Signal Peptide Search

In order to evaluate where a His-tag could be attached, a signal peptide search was performed. The search indicated the presence of a signal peptide in only one of the proteins encoded by

the genes of interest, gene 6468. This signal peptide was predicted to consist of the first 27 amino acids, with a cutoff site between aa 27 and 28. As the signal peptide is often a zip code for export from the cytoplasm, this could indicate that the protein produced from gene 6468 was meant to be exported to another compartment than the cytosol.

3.1.3 Homology Modelling

In order to know where a His-tag could be attached, 3D models were made of all the proteins of interest. A summary of the modelling is listed in Table 3.2. The C-terminal end seemed freely available for all genes except 8081, for which neither end could be modelled. The N-terminal end could not be modelled for any of the genes. Genes 10760 and 9329 seemed to form homodimers and homotrimers, respectively, of which only the latter could be completely modelled. This left the availability of the C-terminal end of 10760 in the dimeric state unknown.

*Table 3.2: **3D homology modelling results for proteins encoded by genes of interest**. The 3D structure of the proteins of interest were modelled using the SWISS-MODEL homology modelling tool. Comments are made on the ability to model and eventual availability of the N-terminal and C-terminal ends of the proteins, as well as their assumed polymerization state.*

Gene	Structure modelling	Polymerization
6817	Unable to model N-terminal. C-terminal seemed available.	Monomer
10760	Unable to model N-terminal. C-terminal seemed available.	Homodimer, could only model monomer
9329	Unable to model N-terminal. C-terminal seemed available.	Homotrimer, could model all
6468	Unable to model N-terminal. C-terminal seemed available.	Monomer
8081	Could not model either end.	Monomer

3.2 Construction of RK2-Based Expression Vector

In order to express the genes of interest in a bacterial host, an expression vector was needed. A broad-spectrum vector was chosen in case the expression host had to be changed. This expression vector should have a His-tag and restriction sites enabling a ligation of the genes

of interest in the correct reading frame. In addition, a *Pm* promoter and a relatively high copy number was desired. To achieve a vector with restriction sites that could be used to insert the gene of interest without fragmenting it, two separate vectors were made; pMEJ1 and pMEJ2.

An overview of the process of making the expression vectors pMEJ1 and pMEJ2 is shown in Figure 3.1. To achieve this construction, the backbone of pMV23 was used. This fragment included the desired promotor and the *oriV* origo of replication and the *trfA cop271C* gene from RK2. This back bone also contained a *bla* gene that could be used to select for the plasmid. The desired restriction enzyme cut frame and the His-tag was added by the insertion of a modified fragment from pMV20. In the case of pMEJ2, PCR was used to add an extra restriction site.

The restriction enzyme cut sites were designed so that the gene of interest could be ligated into the expression vector behind the promoter using NdeI to cut the 5'-end and NotI (pMEJ1) or AvrII (pMEJ2) to cut the 3'-end of the expression vector. The overhang at the 3'-end of pMEJ1 would then be compatible with the overhang from fragments cut with either PspOMI or NotI, and the corresponding overhang from pMEJ2 would be compatible with fragments cut with either SpeI, XbaI, NheI or AvrII. In the case of both pMEJ1 and pMEJ2, an XhoI restriction site was situated between the NotI site and the His-tag. This restriction site could also be used for the cutting of the 3'-end of the expression vector, and the overhang would be compatible with the overhang of fragments cut with either XhoI, SalI, PspXI, or SgrDI. This construction of restriction sites would give a large flexibility for the insertion of different fragments.

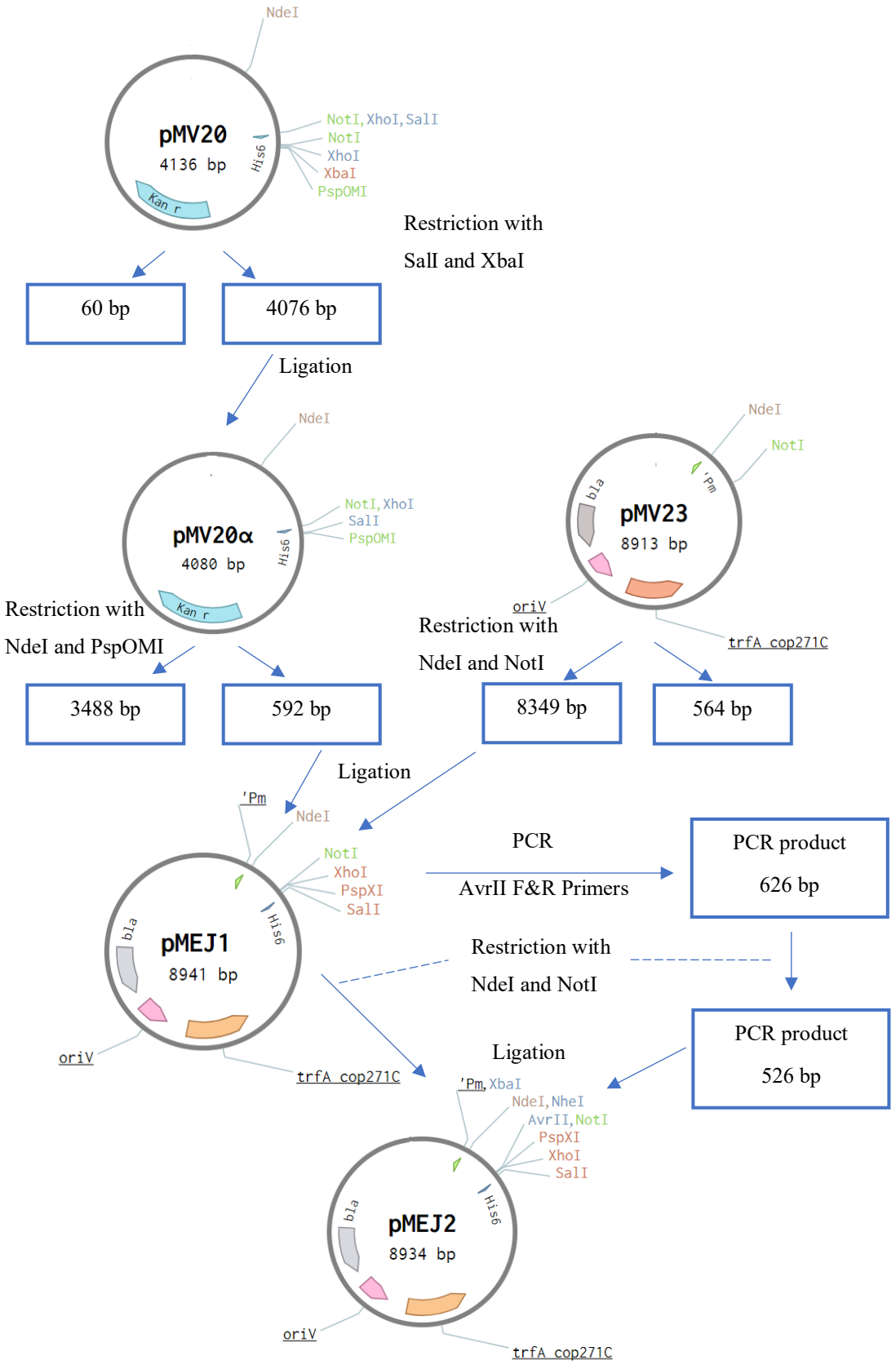


Figure 3.1: ***Overview of the construction of expression vectors pMEJ1 and pMEJ2***. *The restriction enzymes used and which fragments were joined by ligation are indicated.*

3.2.1 Construction of pMV20α

The NotI and XhoI restriction sites between bases 980 and 1000 in pMV20 had to be removed to ensure that later cutting with NotI would not remove the His-tag.

The pMV20α plasmid was restricted with SalI and XbaI, and overhangs were blunted using a Klenow polymerase before ligation. Plasmids purified from transformant colonies of *E. coli* DH5α were restricted. The fragments from the digest were separated on an agarose gel (Figure 3.2). The fragment assumed to be a 4076 bp was cut out and purified. Two different ladders were used for this, as well as for other agarose gels in this thesis; λ HindIII and λ PstI. See Appendix A – Ladders for information about the ladders. Further discussions will name plasmid candidates from different colonies as c1 (colony 1), c2 (colony 2) and so forth.

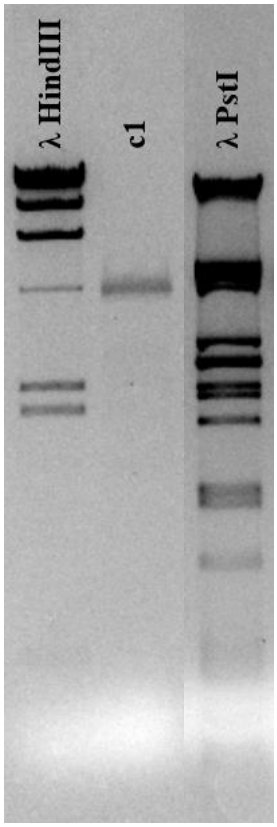


Figure 3.2: ***Digest of pMV20 before cloning****. Purified plasmids from* E. coli *colonies containing pMV20 were digested with SalI and XbaI, treated with Klenow enzyme, and then separated on an agarose gel. Theoretically, this would give a 4076 bp DNA fragment and a 60 bp fragment.*

The purified fragment from pMV20 was then religated to give the pMV20α plasmid. This plasmid was heat-shock transformed into *E. coli* DH5α. The newly formed plasmids were purified from four different colonies. The purified pMV20α candidates were cut with PstI and

SgrAI to verify whether the plasmids were correct (Figure 3.3, left side). Table 3.3 indicates the expected sizes of the fragments. As seen on the figure, all four candidates seemed correct.

Further, the purified pMV20α plasmids were also cut with NdeI and PspOMI (Figure 3.3, right side). According to the figure, all four colonies appeared correct (see Table 3.3). It was chosen to use the plasmids from candidate 1, and the 592 bp fragment from this colony was cut out from the gel and purified.

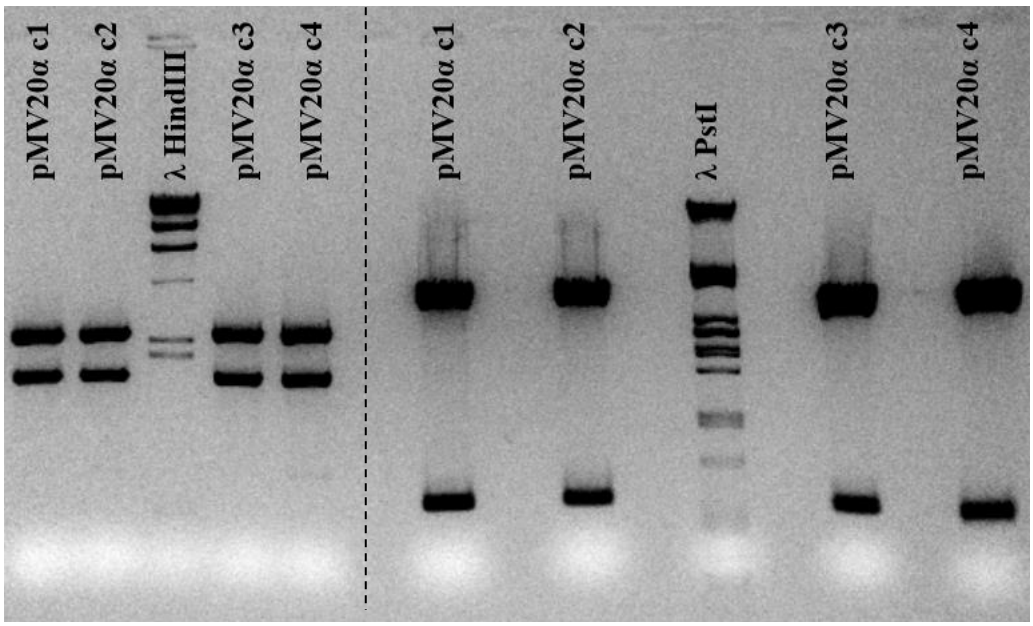


Figure 3.3: ***Digests before cloning and for verification of pMV20α****. Plasmids from four different* E. coli *colonies containing pMV20α were digested with PstI and SgrAI (left side of figure) or NdeI and PspOMI (right side of figure).*

Table 3.3: ***Expected fragment sizes from digest of pMV20α and parent plasmids****. Theoretical fragment sizes from the digest of pMV20α and its parent plasmid. The restriction enzymes used for each digest are indicated in the table. Parent plasmids are marked green and placed directly underneath the new plasmid.*

Plasmid	Restriction enzymes	Expected fragments (bp)
pMV20α	PstI + SgrAI	2442 + 1638
pMV20	PstI + SgrAI	2000 + 1638 + 498
pMV20α	NdeI + PspOMI	3488 + 592
pMV20	NdeI + PspOMI	3488 + 648

3.2.2 Construction of pMEJ1

The pMV23 plasmid was cut with NdeI and NotI. This would theoretically give an 8349 bp fragment and a 564 bp fragment. The fragments were separated on an agarose gel (Figure 3.4). The ~8 kbp fragment of pMV23 was excised from the gel and purified.

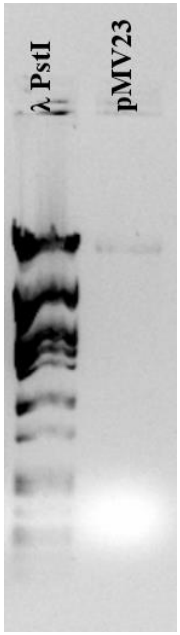
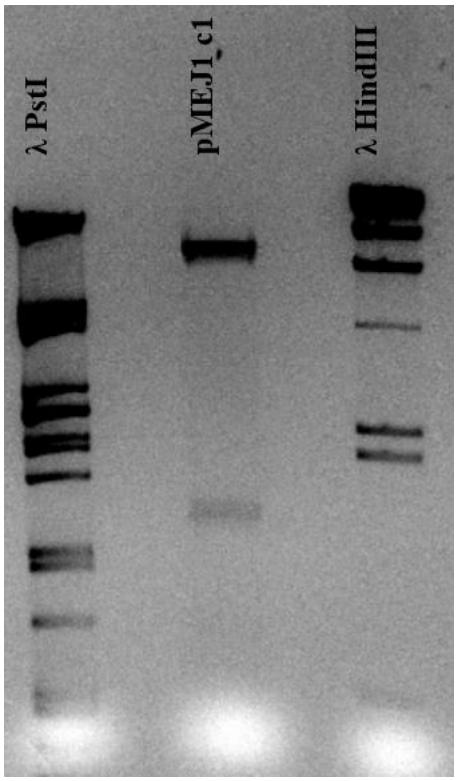


Figure 3.4: ***Digest of pMV23 before cloning****. The pMV23 plasmid was digested with NdeI and NotI and separated on an agarose gel.*

The 592 bp fragment from pMV20α was ligated to the 8349 bp fragment from pMV23. The resulting plasmid was transformed into *E. coli* DH5α. Subsequently, the new plasmid was purified from the transformed bacteria. Verification of the new plasmid was done by restriction with PstI followed by gel electrophoresis (Figure 3.5).

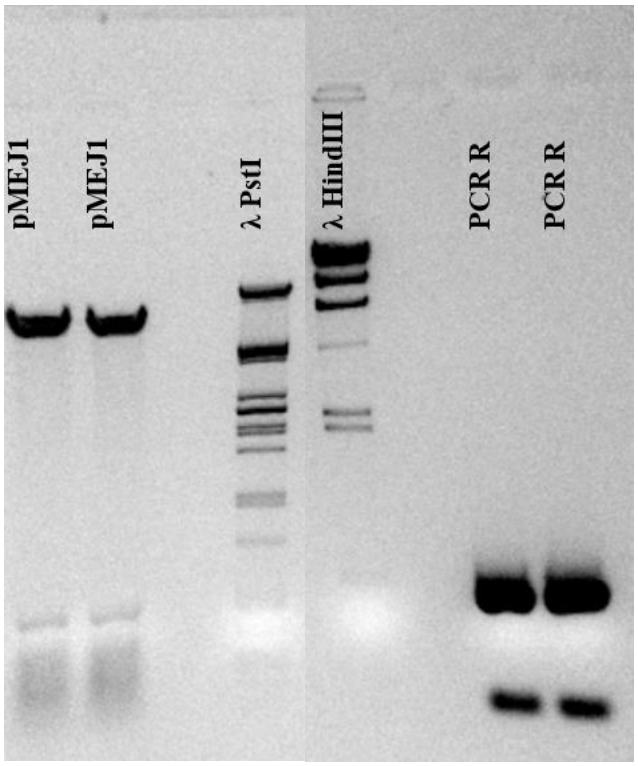


Plasmid	Expected fragments (bp)
pMEJ1	7538 + 1403
pMV20α	4080
pMV23	8913

*Figure 3.5: **Verification digest of pMEJ1**. The pMEJ1 plasmid from* E. coli *DH5α colony 1 was digested with PstI and separated on an agarose gel. The expected fragment sizes are shown to the right. Parent plasmids are marked green and placed underneath the new plasmid.*

3.2.3 Construction of pMEJ2

Two of the genes of interest contained restriction sites for both NotI and PspOMI. Therefore, an expression vector where another enzyme could be used to restrict downstream of the inserted gene was needed. The pMEJ2 plasmid was made to be an equivalent of the pMEJ1 plasmid, where the gene fragment could be inserted by cutting with NdeI and AvrII instead of with NdeI and NotI/PspOMI. A DNA fragment was made by amplification with PCR of a region of the pMEJ1 plasmid using primers with a NdeI restriction site upstream of the gene and an AvrII restriction site just upstream of a NotI restriction site downstream of the gene. This fragment is hereby referred to as PCR R. Both PCR R and the pMEJ1 plasmid was digested with NdeI and NotI and separated on an agarose gel (Figure 3.6). The restriction seemed correct according to the gel.

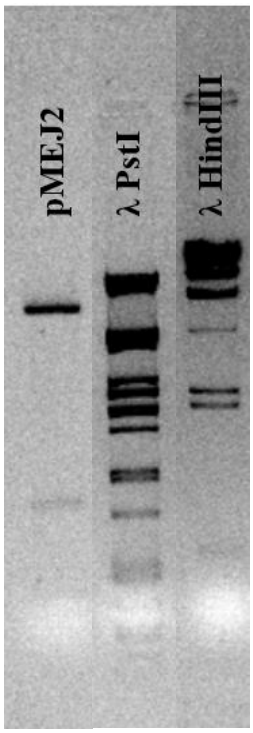


Plasmid	Expected fragments (bp)
pMEJ1	8408 + 533
PCR R	526 + 94 + 6

Figure 3.6: ***Digest of pMEJ1 and PCR R before cloning****. The pMEJ1 plasmid and a PCR product of this plasmid, referred to as PCR R, was digested with NdeI and NotI. The fragments were separated on an agarose gel. The figure consists of two parts of the same gel. Theoretical fragment sizes from the digest is shown to the right.*

The 526 bp fragment from PCR R and the 8408 bp fragment were cut out from the gel, purified and ligated. The product of this ligation should be plasmid pMEJ2, which was subsequently heat-shock transferred into competent *E. coli* DH5α.

The pMEJ2 plasmid was verified by digestion with PstI and AvrII. The fragments were separated on an agarose gel (Figure 3.7). The two largest fragments seemed to be correct, while the smallest fragment was not clearly observed. This was assumed due to the "blocking" of the band by the loading stain. Still, Figure 3.7 show that pMEJ2 contained the inserted AvrII restriction site.



Plasmid	Expected fragments (bp)
pMEJ2	7528 + 986 + 420
pMEJ1	7538 + 1403

*Figure 3.7: **Verification digest of pMEJ2.** The pMEJ2 plasmid was restricted with PstI and AvrII and separated on an agarose gel. The figure is composed of several parts of the same gel. Expected fragment sizes are shown to the right. The parent plasmid is marked green and placed underneath the new plasmid.*

3.3 Cloning of Gene 6817, 10760, 9329, 8081, and 6468

In order to clone the genes of interest, gene 6817, 10760, 9329, 8081, and 6468, into the expression vectors, these were first amplified using PCR. The amplified DNA fragments were then cloned into TOPO® vectors and sequenced. The genes of interest verified by sequencing were then cloned into either expression vector pMEJ1 or pMEJ2.

3.3.1 PCR of Genes of Interest

The genes of interest were amplified by PCR with genomic *Aurantiochytrium* T66 DNA as template. The genetic material could be used directly due to the absence of introns (see Section 1.6). Specific primers were used for each gene of interest (Section 2.2).

All the primers attaching to the start of the genes of interest were designed to insert a NdeI restriction site. This site was located so that a NdeI digest would yield a fragment beginning with the start codon ATG. The primer attaching to the end of gene 6817 was made to incorporate a NotI restriction site just after the gene. The reverse primers of genes 10760 and 6468 yielded PspOMI restriction sites at the end of the genes. An AvrII and a NheI restriction

site was engineered into the primers binding to the end of gene 9329 and 8081, respectively. All the reverse primers were designed so that the His-tags in the expression vectors would be in the correct reading frame. Table 3.4 shows the different elongation times and the annealing temperatures used for the amplification of each gene, as well as the expected product length in bp for each gene.

Table 3.4: ***PCR parameters for genes of interest****. The table indicates the final annealing temperatures and elongation times used for the PCR of the different genes of interest, as well as the expected product lengths in bp for each reaction.*

Gene	Temperature	Time	Product length (bp)
6817	60.0 ˚C	45 s	980
10760	58.0 ˚C	45 s	1291
9329	58.0 ˚C	45 s	977
8081	61.8 ˚C	1.5 min	2258
6468	67.0 ˚C	2 min	1957

The PCR products were verified by separating them from eventual other DNA fragments on agarose gels.

The separation of genes 6817, 10760, 9329, and 8081 are shown in Figure 3.8. In the case of gene 10760 and 9329, two different concentrations of template DNA in the PCR reactions were used to optimize the procedure. The bands marked with blue arrows in the figure were excised from the gel and purified.

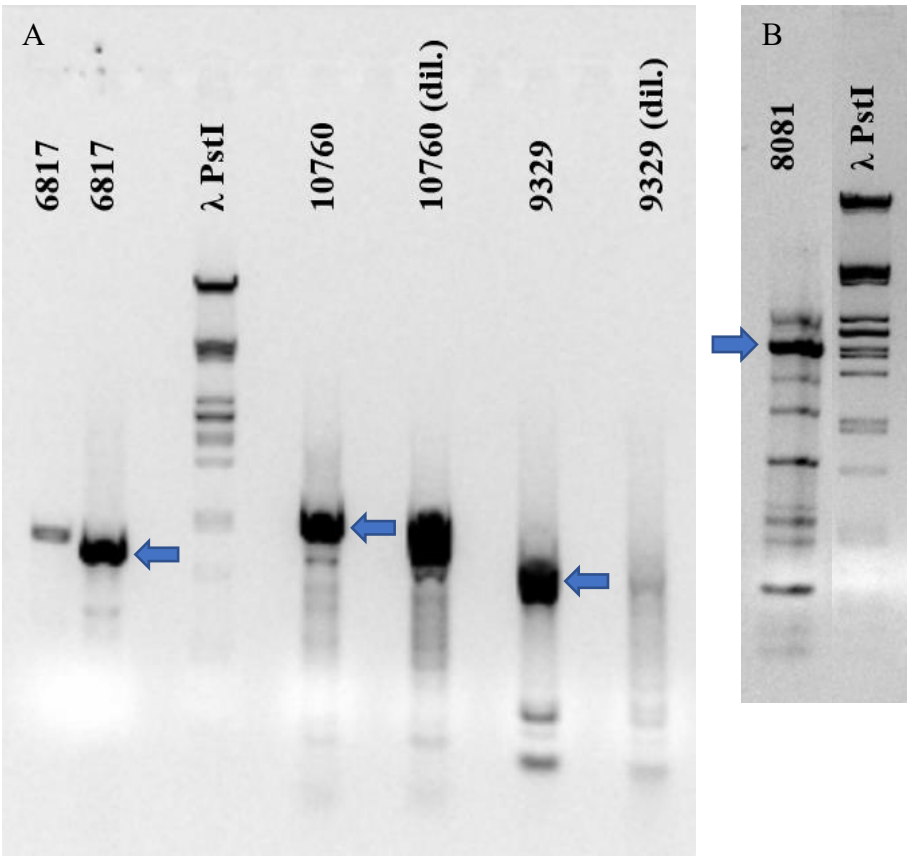


Figure 3.8: ***Verification of PCR products of gene 6816, 10760, 9329, and 8081.*** *The PCR products of gene 6817, 10760, 9329, (A) and 8081 (B) were separated on agarose gels. For gene 10760 and 9329, two different concentrations of template DNA in the PCR reaction were used, there the lowest concentrations are marked as diluted (dil.). The fragments that were cut out from the gels are indicated with blue arrows.*

For gene 6468, the annealing temperature used in the PCR reactions had to be optimized. Figure 3.9 shows the agarose gel of the PCR samples of this gene at different annealing temperatures. The largest band (~1957 bp) of the left sample of 67.0 ˚C annealing temperature was excised and purified.

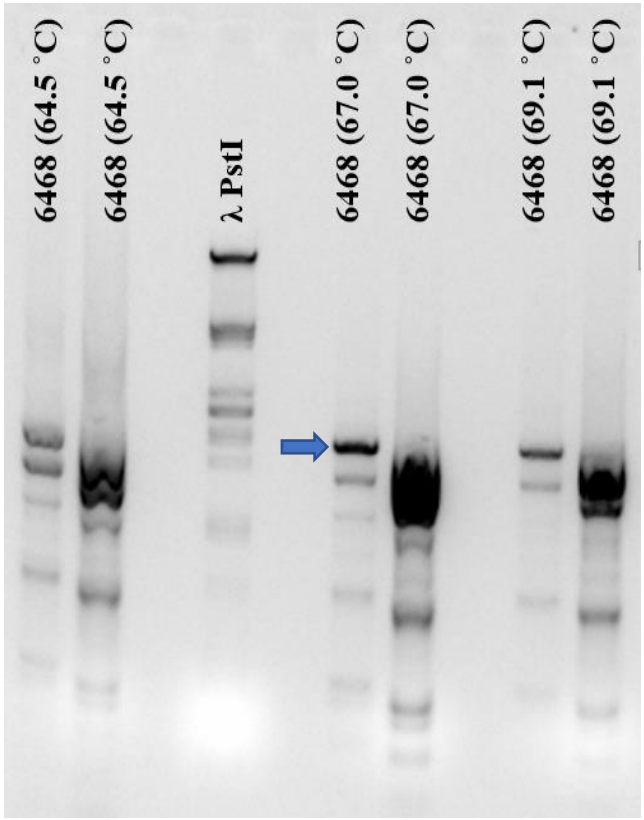


*Figure 3.9: **Annealing temperature effect on PCR of gene 6468**. The PCR products of gene 6468 at different annealing temperatures were separated on an agarose gel. The fragment that was cut out and purified is indicated with a blue arrow.*

3.3.2 TOPO® Cloning and Sequencing

The purified PCR fragments were TOPO® cloned into chemically competent *E. coli* DH5α cells. The insertion of the PCR fragments into the TOPO® vectors were verified by restriction cutting of the plasmid purified from one or two colonies, followed by gel electrophoresis (Figure 3.10 and Table 3.5).

According to the gel, c1 of the vector with 6468 had a 5'-3' insertion, while c2 had a 3'-5' insertion. For gene 6817, both candidates seemed to have a 3'-5' insertion. For gene 9329, both colonies had a 3'-5' insertion. In the case of gene 10760, c1 contained a 3'-5' insertion, while c2 contained a 5'-3' insertion.

For the vector containing gene 8081 no restriction cutting was made to evaluate the direction of the insert. However, the gel in Figure 3.10 shows the expected bands (2244 bp and 3524 bp) for the restriction of the vector with NdeI and NheI.

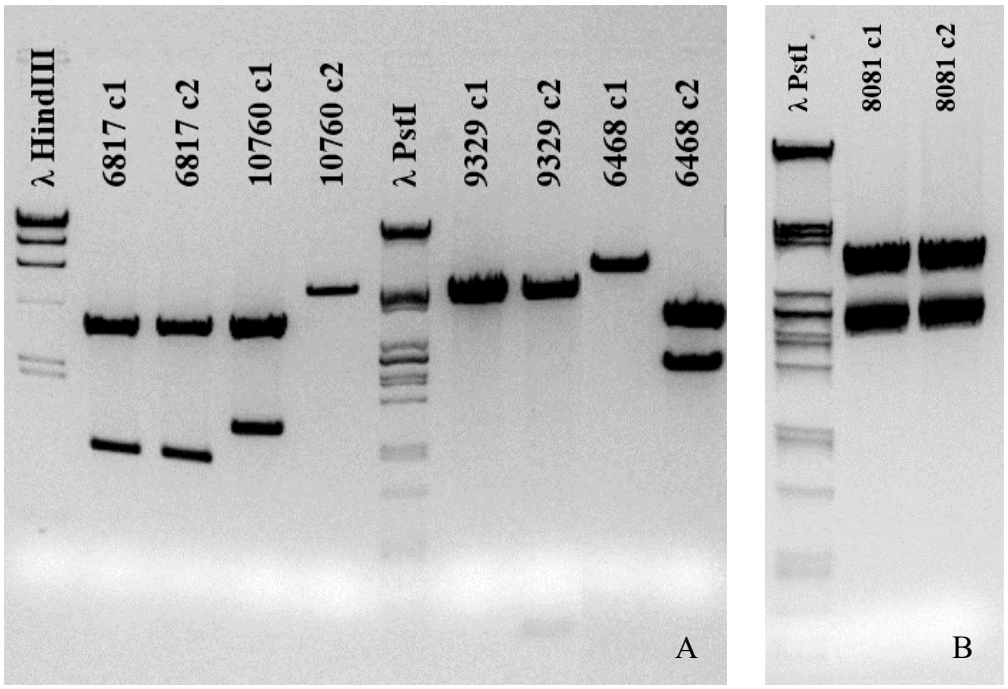


Figure 3.10: ***Verification digest of TOPO® vectors containing genes 6817, 10760, 9329, and 6468 (A) and gene 8081 (B)****. TOPO® vectors containing the PCR fragments of the genes 6817, 10760, 9329, 6468 (A) and 8081 (B) were digested with different restriction enzymes and separated on an agarose gel. The enzymes used to digest the different vectors are indicated in Table 3.5.*

Table 3.5: ***Fragment sizes for digests of variations of gene inserts in TOPO® vectors.***

Gene	Restriction enzyme	Direction of insert	Fragment sizes (bp)
6468	PspOMI	No insert	3519
		5'-3'	5413 + 63
		3'-5'	3462 + 2014
6817	NotI	No insert	3519 bp
		5'-3'	4459 + 40
		3'-5'	3487 + 1012
9329	PspOMI	No insert	3519
		5'-3'	3525 + 971
		3'-5'	4370 + 126
10760	PspOMI	No insert	3519
		5'-3'	4747 + 63
		3'-5'	3462 + 1348

The TOPO® vectors containing the inserts of the genes of interest were sent to sequencing to the commercial agent Eurofins Genomics, GATC service (LIGHTRUN). The sequencing results can be found in Appendix D – Sequencing of Genes of Interest.

The sequencing of gene 6817 showed that both colonies contained inserts in a 3'-5' direction. Sequencing failed for c1 of gene 9329, while the sequencing of c2 indicated a 3'-5' insert. The sequencing of gene 8081 was only completely performed for c1, which seemed to be in a 3'-5' direction. No synthesis mistakes were detected in gene 6817, 9329 and 8081.

The sequencing of gene 10760 indicated that c1 contained an insert in 3'-5' direction and that c2 contained an insert in 5'-3' direction. A sequence mismatch was found in base 820 in the TOPO® vector with a 5'-3' insert. The mismatch was present in both sequenced colonies. A cytosine (C) was exchanged for an adenine (A), giving a change in the amino acid from proline to threonine. This could affect enzyme activity. The protein sequence of the protein encoded including the mismatch was compared to the protein sequence of succinate-CoA ligase, ADP-forming β-subunit (mitochondrial) encoded by *Hondaea fermentalgiana* (Sequence ID: GBG32698.1). A threonine rather than a proline was found at the site here discussed. This led to the assumption that the original sequencing of the *Aurantiochytrium* T66 gene at this site might have been wrong, and that the correct and active protein could be produced from this gene.

The sequencing of gene 6468 indicated that c1 contained a 5'-3' insert, while c2 contained a 3'-5' insert. A mismatch was found at base 2269 in the TOPO® vector with a 3'-5' insert, where an A was exchanged for a C. The mismatch was present in both sequenced colonies. This would change the amino acid from serine to alanine. This mismatch was assumed to be caused by a mistake in the forward primer because of its presence in both sequenced colonies. However, due to this area being presumed a signal sequence (see Section 3.1.2), this change was assumed not to affect enzyme folding and activity.

The direction of the inserts in the TOPO® vectors of gene 6864, 6817, 10760, and 9329 confirmed the directional assumptions based on the gel in Figure 3.10.

Table 3.6 indicates which candidates were chosen for further use for each of the TOPO® vectors with inserted genes.

Table 3.6: ***TOPO® vectors with inserted gene fragments chosen for further use****. The names given to these plasmids are also indicated.*

Gene of interest	Candidate chosen	Name of plasmid
6817	c1	pMEJ3T
10760	c2	pMEJ4T
6468	c2	pMEJ5T
9329	c2	pMEJ6T
8081	c1	pMEJ7T

3.3.3 Construction of Expression Vectors Containing the Genes of Interest

The genes of interest were cut out from the sequenced TOPO® vectors with restriction enzymes. In the case of gene 6817, 10760, 9329, and 6468, restriction reactions were set up to yield fragments with the genes of interest (Figure 3.11). However, in the case of gene 8081, the verification digest shown in Figure 3.10 yielded the gene fragment in sufficient quantity to be cut directly from the gel.

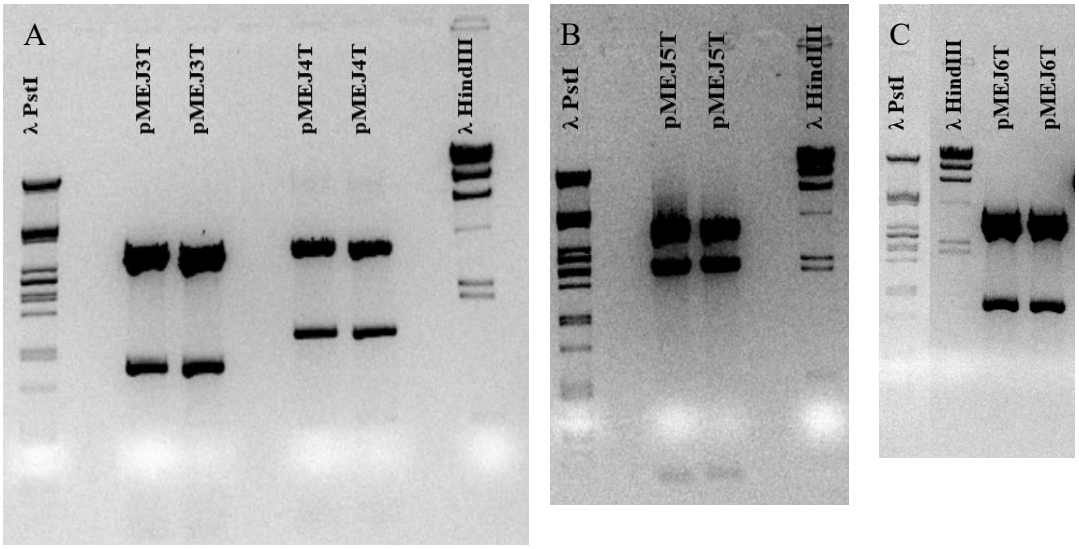


Figure 3.11: ***Digest of pMEJ3T and pMEJ4T (A), pMEJ5T (B) and pMEJ6T (C) before cloning****. A) pMEJ3T and pMEJ4T were digested with NdeI and NotI or NdeI and PspOMI, respectively. B) pMEJ5T was digested with NdeI and PspOMI. C) pMEJ6T was digested with NdeI and AvrII. The fragments from all the digests were separated on agarose gels.*

Purified pMEJ1 and pMEJ2 plasmids were cut with NdeI and NotI or NdeI and AvrII, respectively. The restriction mixtures were separated on an agarose gel, shown in Figure 3.12. The largest bands from each plasmid was cut out from the gel and purified.

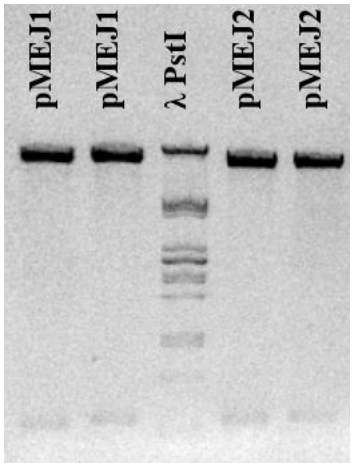


Figure 3.12: ***Digest of pMEJ1 and pMEJ2 before cloning****. Purified pMEJ1 and pMEJ2 plasmid was digested with NdeI and NotI or NdeI and AvrII, respectively. The restriction mixture was separated on an agarose gel.*

The gene fragment from pMEJ3T, pMEJ4T and pMEJ5T was ligated into pMEJ1, while the fragments from pMEJ6T and pMEJ7T were ligated into pMEJ2. The ligations were heat-shock transformed into competent *E. coli* DH5α cells.

The plasmids were then purified from one or more colonies of the transformed bacteria, digested with restriction enzymes, and separated on agarose gels (Figure 3.13 and Table 3.7). All plasmids seemed correct according to the gels. Candidate 1 (c1) for all the plasmids where several candidates were tested were chosen for later procedures.

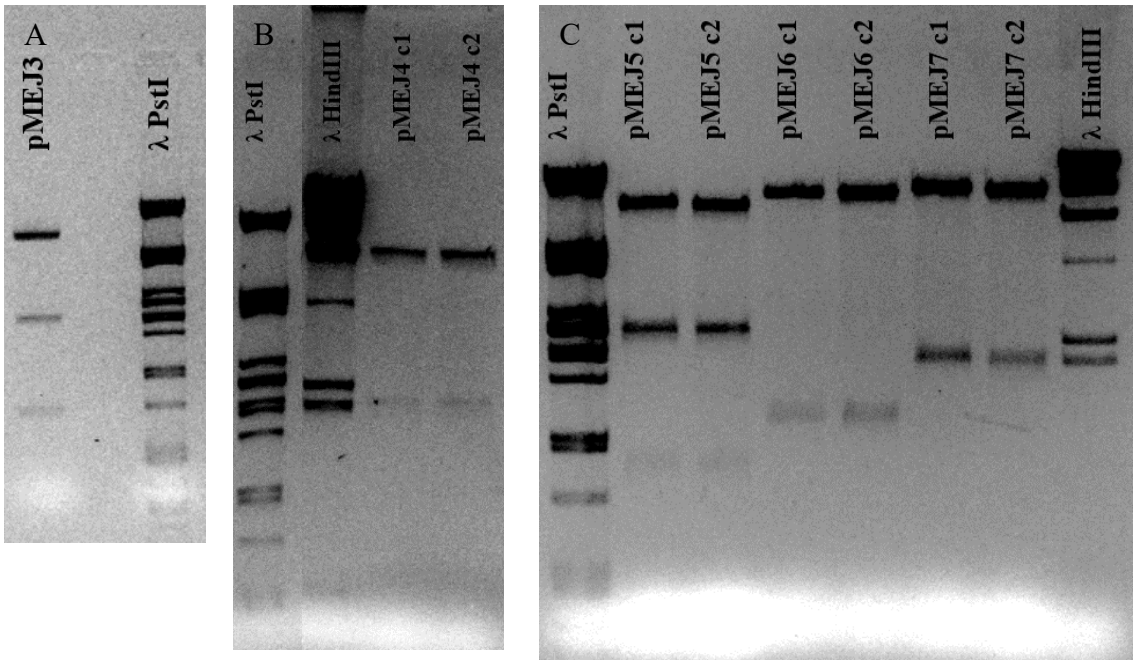


Figure 3.13: ***Verification digest of pMEJ3 (A), pMEJ4 (B), and pMEJ5, pMEJ6, and pMEJ7 (C)****. One candidate for the pMEJ3 plasmid (A), two candidates for the pMEJ4 plasmid (B), and two candidates for each of the plasmids pMEJ5, pMEJ6, and pMEJ7 (C) were digested with the restriction enzymes indicated by Table 3.7. The fragments were separated on an agarose gel.*

*Table 3.7:****Expected fragment sizes from digest of pMEJ3, pMEJ4, pMEJ5, pMEJ6, and pMEJ7 and parent plasmids****. Theoretical fragment sizes from digests of pMEJ3, pMEJ4, pMEJ5, pMEJ6, and pMEJ7. Restriction enzymes are indicated. Parent plasmids for each new plasmid are marked green and placed directly underneath the new plasmid.*

Plasmid	Restriction enzymes	Expected fragments (bp)
pMEJ3	SalI + HindIII	6400 + 2134 + 752
pMEJ1	SalI + HindIII	6657 + 2134 + 94 + 56
pMEJ4	SacI + SbfI	6960 + 2049 + 683
pMEJ1	SacI + SbfI	8250 + 683
pMEJ5	SacI + NotI	7005 + 2386 + 967
pMEJ1	SacI + NotI	7280 + 1661
pMEJ6	BglII	8055 + 1331
pMEJ2	BglII	8934
pMEJ7	BglII	8730 + 1928
pMEJ2	BglII	8934

3.4 Protein Production in *E. coli*

The plasmids containing the genes of interest were electroporated into *E. coli* RV308. These strains were then grown to produce the proteins encoded by the different plasmids, which were subsequently purified using FPLC with a stepwise elution. A culture with the pMSM2 plasmid carrying an alginate lyase gene (*alyB*) from *Azotobacter vinelandii* not known to require any cofactor (Ertesvåg, unpublished data) was also grown as a negative control. The control protein also contained a His-tag.

The chromatogram of the purification of gene 6468 is shown in Figure 3.14, of gene 8081 in Figure 3.15, and of the negative control in Figure 3.16. The blue line showing the absorption at 280 nm in mAU indicates that the bulk of the proteins was eluted during loading, while the small peak just after 50 mL elution volume indicates a small portion of the protein being eluted after the first change of buffer composition. The peaks of gene 6468 and 8081 are significantly smaller than the peak of the negative control, indicating a lower protein concentration in these samples. This was confirmed by protein concentration measurements using a NanoDrop One from ThermoFisher Scientific, where the concentration of the presumed proteins of interest were about 1/15th of the concentration of the negative control.

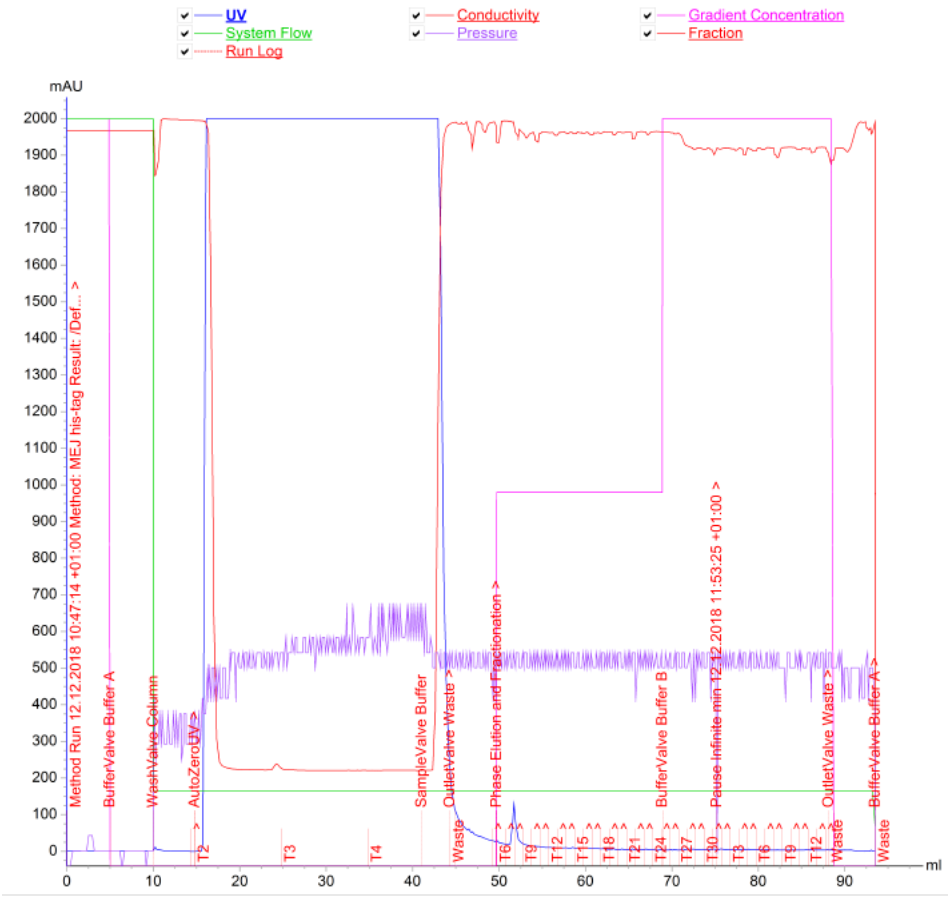


Figure 3.14: ***Chromatogram from purification of gene 6468.*** E. coli *RV308 cells containing a plasmid with the gene 6468 were grown with an inducer to produce the enzyme encoded by the gene. The enzyme was then purified using FPLC (ÄKTA start from GE Healthcare Life Sciences) with a His-Trap HP column (1 mL). A stepwise gradient of the elution buffer was used, where 0 % buffer was applied first, followed by 50 and 100 %. The concentration is shown by the purple line in the chromatogram. The x-axis shows the elution volume and the y-axis shows the absorbance (blue line) measured in mAU.*

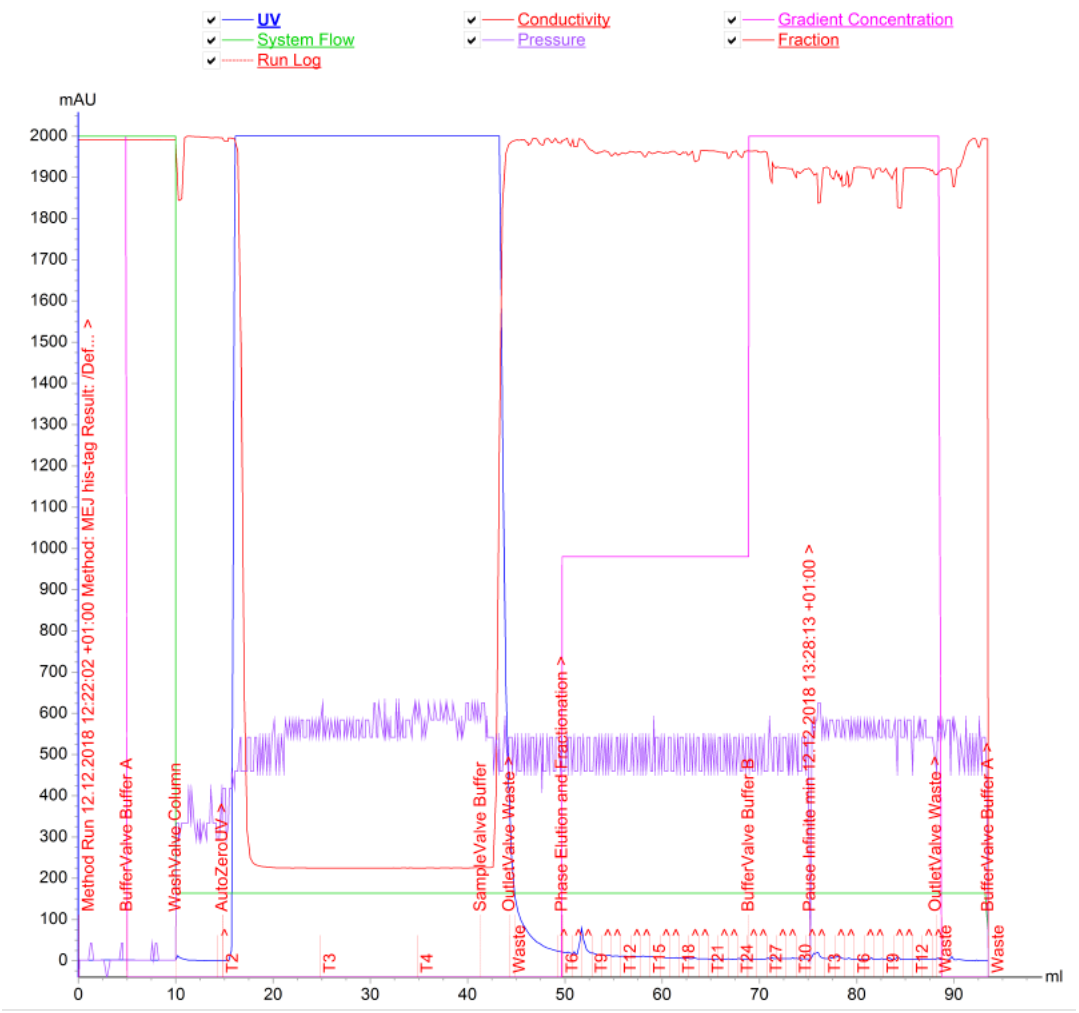


Figure 3.15: ***Chromatogram from purification of gene 8081.*** E. coli *RV308 cells containing a plasmid with the gene 8081 were grown with an inducer to produce the enzyme encoded by the gene. The enzyme was then purified using FPLC (ÄKTA start from GE Healthcare Life Sciences) with a His-Trap HP column (1 mL). A stepwise gradient of the elution buffer was used, where 0 % buffer was applied first, followed by 50 and 100 %. The concentration is shown by the purple line in the chromatogram. The x-axis shows the elution volume and the y-axis shows the absorbance (blue line) measured in mAU.*

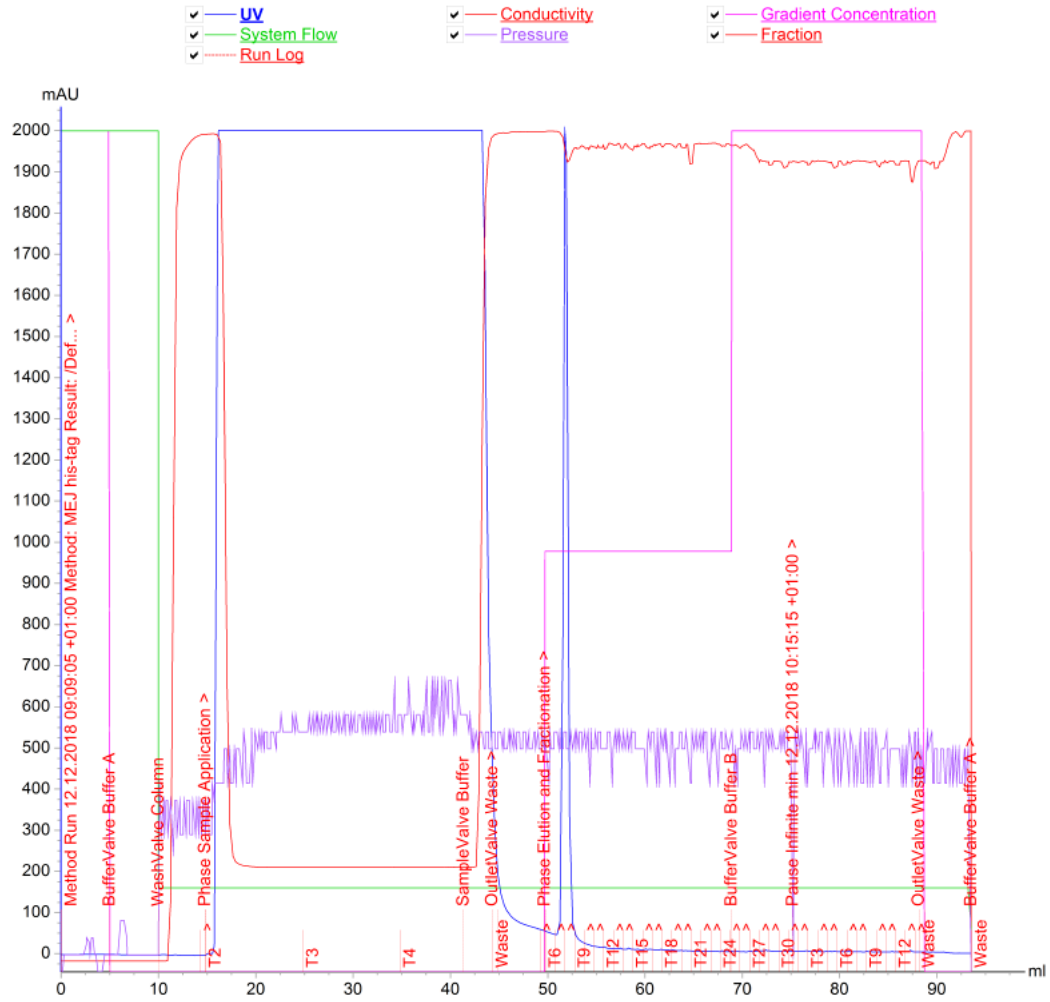


Figure 3.16: ***Chromatogram from purification of control.*** E. coli *RV308 cells containing a plasmid with an* alyB *gene were grown with an inducer to produce the enzyme encoded by the gene. The enzyme was then purified using FPLC (ÄKTA start from GE Healthcare Life Sciences) with a His-Trap HP column (1 mL). A stepwise gradient of the elution buffer was used, where 0 % buffer was applied first, followed by 50 and 100 %. The concentration is shown by the purple line in the chromatogram. The x-axis shows the elution volume and the y-axis shows the absorbance (blue line) measured in mAU.*

The fractions containing the peak eluted after the buffer composition change were pooled. These were run on FPLC again with the same parameters except for the gradient, which was now linear. This operation was done in order to see whether the peaks were indeed single peaks, or if they were a collection of peaks eluted at the same time. The chromatogram of this double purification of gene 6468 is shown in Figure 3.17. This figure shows that the concern of there being several peaks appearing as one was seemingly unfounded. The same was observed for the double purification of gene 8081 and the control.

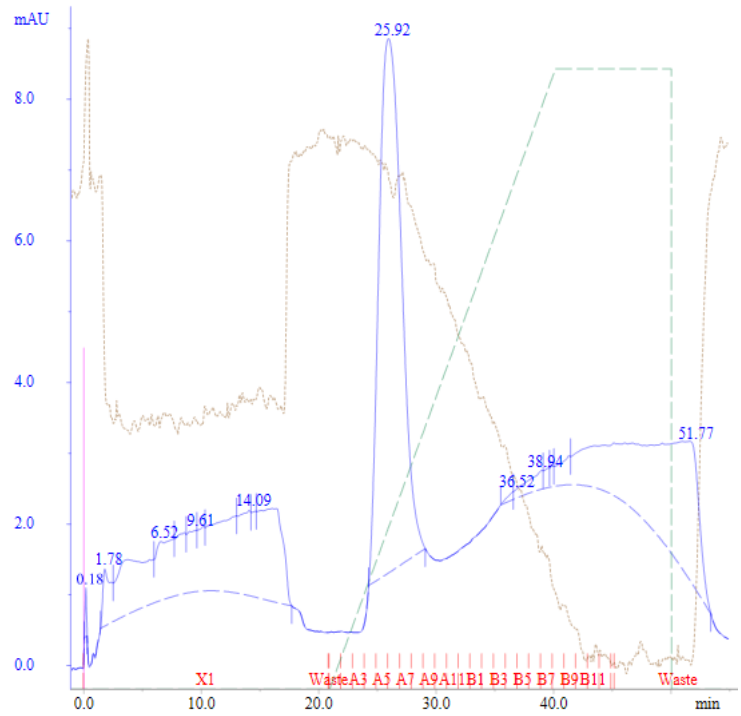


Figure 3.17: ***Chromatogram from double purification of gene 6468.*** E. coli *RV308 cells containing a plasmid with the gene 6468 were grown with an inducer to produce the enzyme encoded by the gene. The enzyme was then purified using FPLC (ÄKTA P-920 from Amersham Biosciences) with a His-Trap HP column (1 mL). For the first purification, a stepwise gradient was used (see Figure 3.14). The fractions belonging to the presumed protein peak containing the His-tagged protein was then pooled. This sample was then purified with a linear gradient of buffer A and B, from 0 to 100% buffer B. The concentration is shown by the green dotted line in the chromatogram. The x-axis shows the elution time and the y-axis shows the absorbance (blue line) measured in mAU.*

G3PDH assays were ran on the protein produced by *E. coli* following the protocol described in Section 2.8.1 and purified with FPLC. 10 µL purified protein sample was added instead of 50 µL supernatant, which is the case described in the protocol. No enzyme activity was detected.

3.5 Growth Temperature Experiment with *E. coli*

As the lowering of temperature might improve heterologous gene expression by preventing formation of inclusion bodies or otherwise improve transcription (see Section 1.5.5), a new protocol with growth at a lower temperature (20 ˚C) for *E. coli* was tested.

For this experiment, culture volumes of 25 mL in 250 mL baffled shake flasks were used. Two parallels of *E. coli* RV308 precultures containing plasmids pMEJ3-7 were inoculated, grown until induction and induced with m-toluate as described in Section 2.6.1. One parallel of each expression vector was inoculated at 37 ˚C for 4 h after induction, while the other parallel was inoculated at 20 ˚C for 20 h. Samples from each of the parallels were lysed and separated into a soluble and an insoluble fraction using B-PER.

Protein gels with the insoluble and soluble fractions (Figure 3.18) were used to separate the protein extracts. A Prestained Protein Ladder from NEB was used for this and subsequent protein gels (Appendix A – Ladders).

An overview of the molecular weights of the genes of interest as well as in which plasmid they were encoded is shown in Table 3.8. The molecular weight of the control protein was 82.3 kDa. A quite large band (marked on the figure) was observed around this size in the insoluble fractions of the control. For the soluble fractions, two large bands were seen where only one large band was found in the other parallels at ~80 kDa. Only the control at 37 ˚C is shown. However, neither of the other fractions showed any large bands around the expected area for the proteins encoded by the genes of interest. This indicated that the proteins were not produced by *E. coli* RV308 at neither 20 ˚C nor at 37 ˚C. The control however confirmed that the procedure itself at 37 ˚C was working, because the heterologously expressed proteins of the control were indeed produced, purified and shown on a protein gel.

*Table 3.8: **Molecular weights of proteins**. The molecular weights in kDa of each of the genes encoded by the plasmids containing a gene of interest is shown.*

Plasmid	Gene encoded	Molecular weight (kDa)
pMEJ3	6817	35.1
pMEJ4	10760	47.5
pMEJ5	6468	72.2
pMEJ6	9329	36.0
pMEJ7	8081	83.9

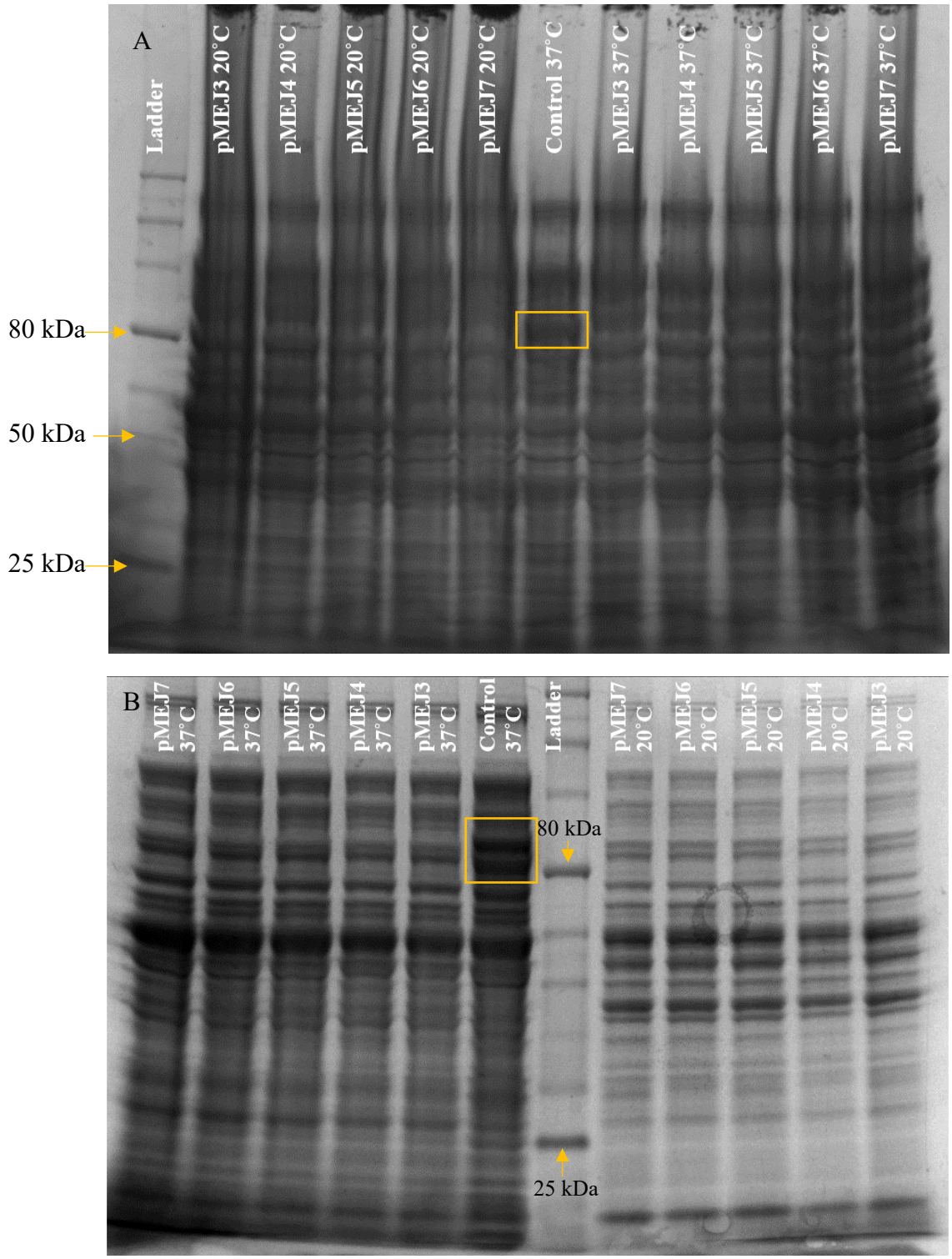


Figure 3.18: ***Protein gel of insoluble (A) and soluble (B) protein from temperature effect on growth experiment with*** **E. coli**. *The insoluble (A) and insoluble (B) protein extracted from* E. coli *RV308 cells grown at 20 ˚C or 37 ˚C after induction were boiled in SDS and separated on SDS-PAGE gels. The ladder used was a Prestained Protein Ladder from New England Biolabs. The bands of the control protein are marked yellow.*

3.6 Vectors for Growth in *P. putida*

Due to the apparent inability of *E. coli* RV308 to produce the desired enzymes, it was decided to try expression in *P. putida*. These bacteria have a more similar GC-content to the *Aurantiochytrium* than the *E. coli*, and it was also a hope that the differences in the metabolic machinery of *P. putida* from that of *E. coli* might be better for expression of the genes of interest. The vectors containing the genes of interest were adapted for growth in *P. putida*. Because of the natural resistance of *P. putida* towards ampicillin, another antibiotic resistance gene giving resistance against an antibiotic able to restrain *P. putida* had to be incorporated into the vectors with the genes of interest. This was done by incorporating a gene for kanamycin resistance from vector pJB861. An overview of the new plasmids made is indicated in Table 2.3.

The kanamycin gene was cut out from pJB861 using either PstI or SbfI. The reason for the use of two different enzymes was caused by an enzyme shortage in the lab at the time. The choice of enzyme was not important for the final outcome. The fragments were separated on an agarose gel (Figure 3.19 and Table 3.9). Both digests seemed correct on the gel. In both cases, the 1240 bp fragments contained the kanamycin gene. These fragments were purified from the gel.

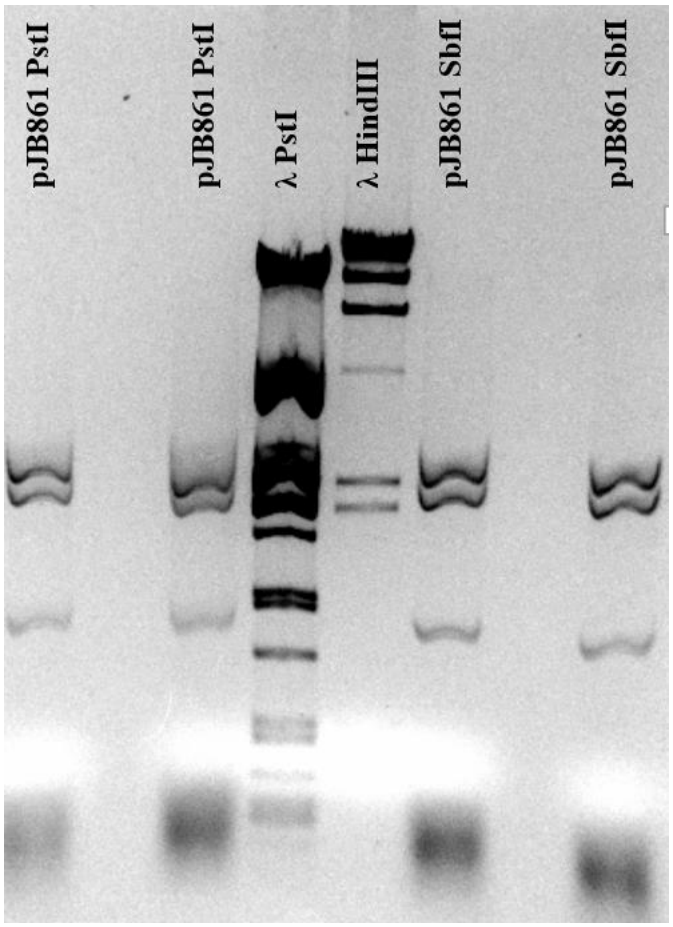


Figure 3.19: ***Digest of pJB861 before cloning****. Some of plasmid pJB861 was digested with SbfI and some was digested with PstI. The fragments were separated on an agarose gel.*

Table 3.9: ***Expected fragment sizes from digests of pJB861.*** *Theoretical fragment sizes from the digest of pJB861 with the indicated restriction enzymes.*

Plasmid	Restriction enzymes	Expected fragments (bp)
pJB861	PstI	3341 + 2698 + 1240
pJB861	SbfI	3323 + 2680 + 1240 + 18 + 18

A single cut was made in the backbones of plasmids pMEJ3-pMEJ7. pMEJ3 and pMEJ4 were digested with PstI, and pMEJ5, pMEJ6, and pMEJ7 were digested with SbfI. The restriction mixtures were then separated on an agarose gel (Figure 3.20). As the restriction was supposed to make one cut in the plasmids, the resulting bands on the gel should be a single band at around 9-10 kbp. The expected bands were observed on the gel, but a smaller band around 5 kbp was also observed. It was assumed that this was a result of uncut DNA. The largest bands were purified from the gel.

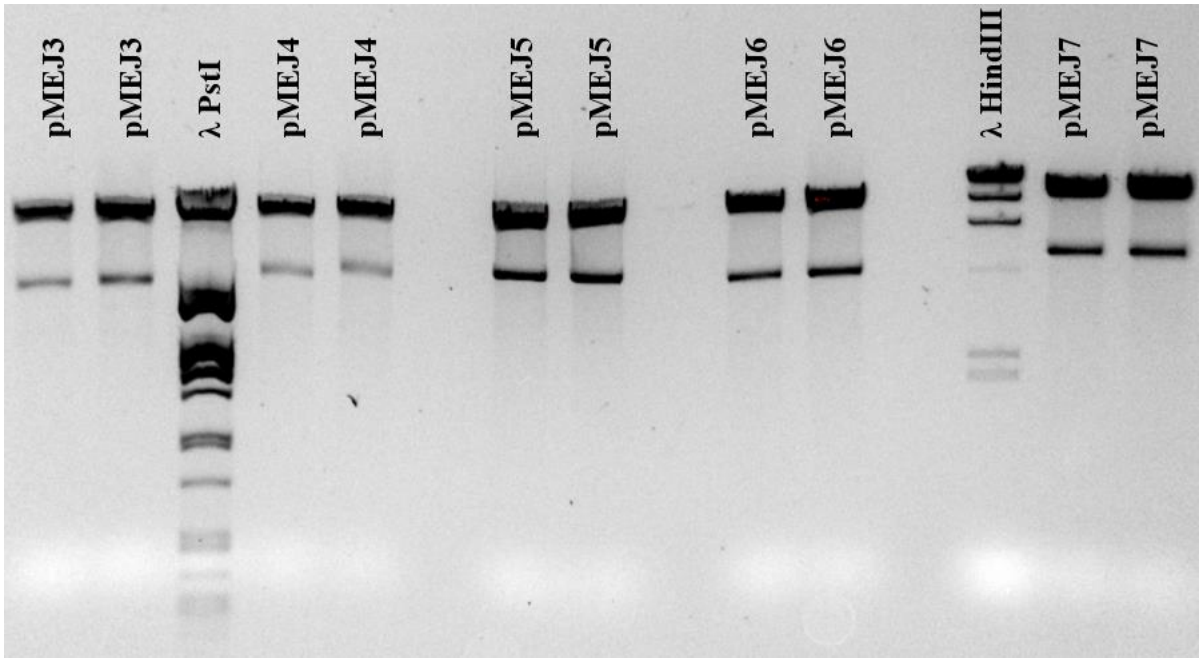


Figure 3.20: ***Digest of pMEJ3, pMEJ4, pMEJ5, pMEJ6, and pMEJ7 before cloning****. Purified plasmids pMEJ3 and pMEJ4 were digested with PstI, while pMEJ5, pMEJ6, and pMEJ7 were digested with SbfI. The fragments were separated on an agarose gel.*

The kanamycin gene containing fragments from pJB861 were ligated into the original expression vectors according to Table 2.3. The ligation mixtures were then heat-shock transformed into *E. coli* DH5α and plated on LA with kanamycin. The plasmids were then purified from selected *E. coli* colonies.

pMEJ8 and pMEJ10 were verified by digestion with HindIII and PspOMI. pMEJ12 was verified by digestion with PspOMI and EcoRI. pMEJ9 and pMEJ11 were verified by restriction with PspOMI and HindIII. The fragments were separated on an agarose gel (Figure 3.21 and Table 3.10). The candidate of pMEJ8 seemed correct. Candidate 4 of pMEJ12, candidate 3 of pMEJ9 and candidate 5 of pMEJ11 were chosen. Assumptions on the direction of the insert of the kanamycin resistance gene were not made due to the inaccuracy of the gels. The candidate of pMEJ10 was first assumed correct and chosen for further use. This assumption seems to have been wrong after more critical evaluation of the gel, and the later results from this plasmid must thus be ignored.

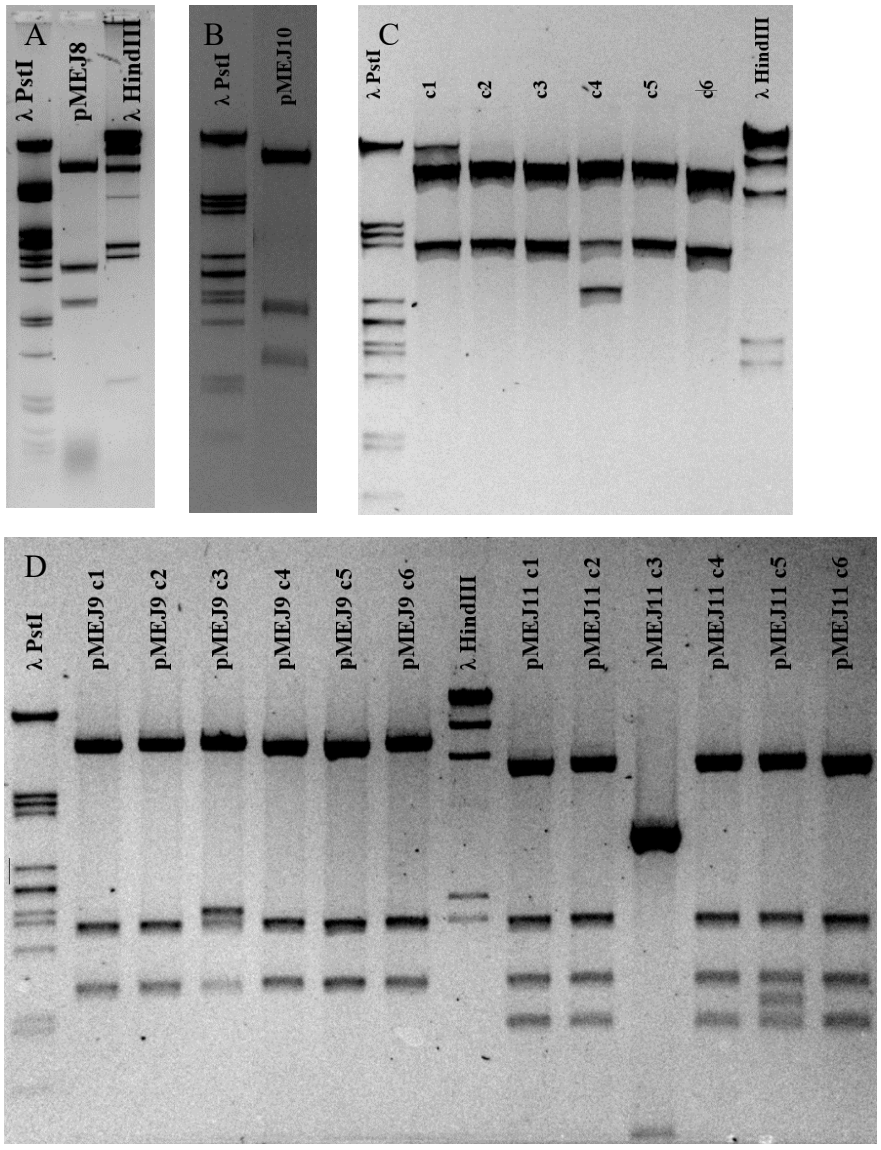


Figure 3.21: ***Verification digest of pMEJ8 (A), pMEJ10 (B), pMEJ12 (C) and pMEJ9 and pMEJ11 (D)****. A) One candidate for pMEJ8 was digested with HindIII. B) One candidate for pMEJ10 was digested with HindIII and PspOMI. C) Six candidates for pMEJ12 were digested with PspOMI and EcoRI. D) Six candidates for pMEJ9 and pMEJ11 were digested with PspOMI and HindIII. All fragments were separated on agarose gels.*

Table 3.10: ***Expected fragment sizes from digest of pMEJ8, pMEJ9, pMEJ10, pMEJ11, and pMEJ12 and parent plasmids****. Theoretical fragment sizes from digests of pMEJ8, pMEJ9, pMEJ10, pMEJ11, and pMEJ12. Restriction enzymes are indicated. Parent plasmids for each new plasmid are marked green and placed directly underneath the new plasmid.*

Plasmid	Restriction enzymes	Expected fragments (bp)
pMEJ8	HindIII	7246 + 1979 + 1395
pMEJ3	HindIII	7246 + 2134
pJB861	HindIII	3927 + 3352
pMEJ10	HindIII + PspOMI	5005 + 3219 + 1867 + 1507
pMEJ5	HindIII + PspOMI	5005 + 3219 + 2134
pJB861	HindIII + PspOMI	3927 + 3352
pMEJ12	PspOMI + EcoRI	5631 + 4200 + 2067
pMEJ7	PspOMI + EcoRI	5631 + 2960 + 2067
pJB861	PspOMI + EcoRI	7279
pMEJ9 5'-3' insert	PspOMI + HindIII	5005 + 2553 + 1867 + 1507
pMEJ9 3'-5' insert	PspOMI + HindIII	5005 + 2553 + 1979 + 1395
pMEJ4	PspOMI + HindIII	5005 + 2553 + 2134
pJB861	PspOMI + HindIII	3927 + 3352
pMEJ11 5'-3' insert	PspOMI + HindIII	5005 + 1867 + 1507 + 1188 + 1059
pMEJ11 3'-5' insert	PspOMI + HindIII	5005 + 1979 + 1396 + 1188 + 1059
pMEJ6	PspOMI + HindIII	5005 + 2134 + 1188 + 1059
pJB861	PspOMI + HindIII	3927 + 3352

3.7 Protein Production in *P. putida*

The plasmids described in Section 3.6 were electroporated into *P. putida* KT2440.

During the first protocol tested, herby referred to as Protocol 1, 25 mL of LB with kanamycin was inoculated with 1 % of *P. putida* preculture. These cultures were grown for 3 h. OD_{600} was measured before the cultures were induced with 50 µL of 0.25 M m-toluate in 50% ethanol. The cultures were then inoculated for ~20 h. OD_{600} and volume was measured.

The protein was extracted in a soluble and an insoluble fraction using B-PER. These fractions were separated on a protein gel. The gel can be seen in Figure 3.22. The molecular weights

should be the same as described in Table 3.8. The desired bands did not particularly stand out in either fraction, even though the insoluble fraction was a bit overloaded. It was nevertheless interpreted as a failure to produce the desired proteins using Protocol 1.

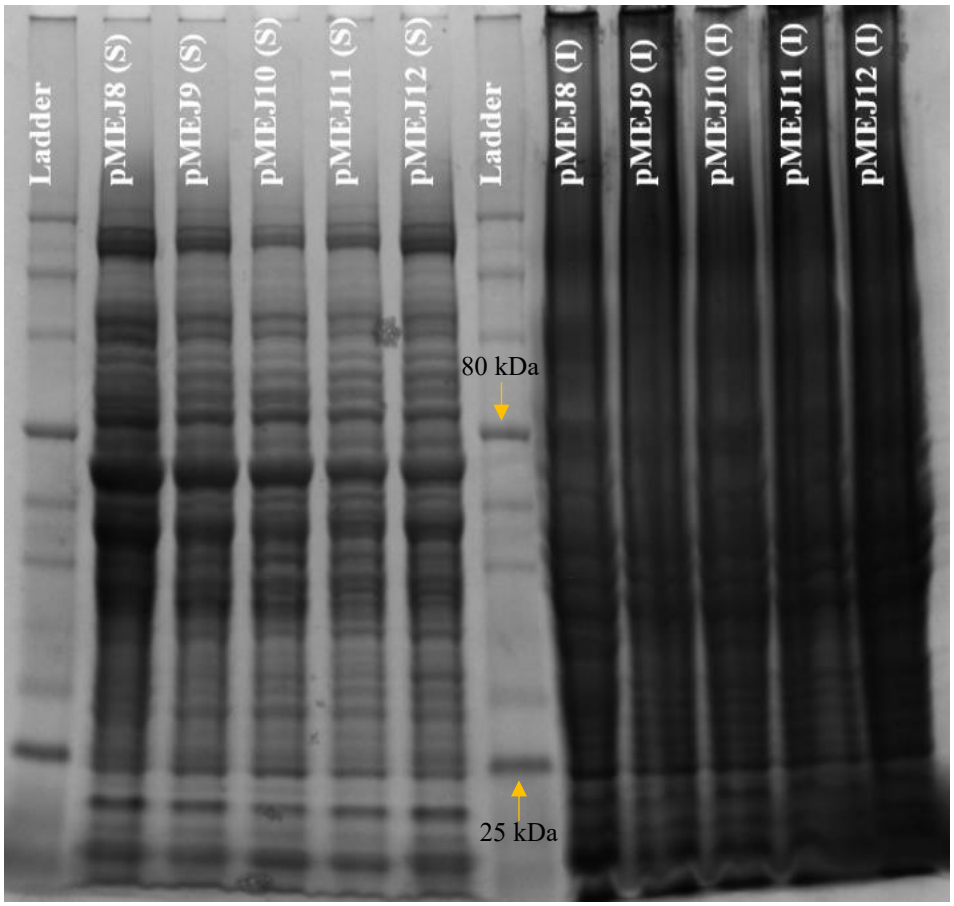


Figure 3.22: ***Protein gel of protein produced in*** **P. putida** ***using Protocol 1****. Protein was produced in* P. putida *using Protocol 1 and extracted in an insoluble (I) and a soluble (S) fraction. These proteins were separated on a RunBlue SDS protein gel (4-12%) from Expedeon.*

In the second protocol tested, hereby referred to as Protocol 2, 25 mL of LB with kanamycin was inoculated with 1 % preculture. These cultures were grown until OD_{600} was ~2. The cultures were induced with 200 µL of 0.25 M m-toluate in 50% ethanol. The cultures were then inoculated for ~15-18 h (ON). OD_{600} and volume was measured.

Like previously, the protein produced was extracted using B-PER and separated into an insoluble and a soluble fraction. A protein gel was used to separate the proteins extracted. In this case, only $1/10^{th}$ of the insoluble fraction was loaded on the gel compared to the soluble fraction. The gel can be seen in Figure 3.23. The gel did not show any clear bands in the

expected area for any of plasmids. The conclusion for Protocol 2 was thus the same as the one for Protocol 1; the desired proteins were not produced.

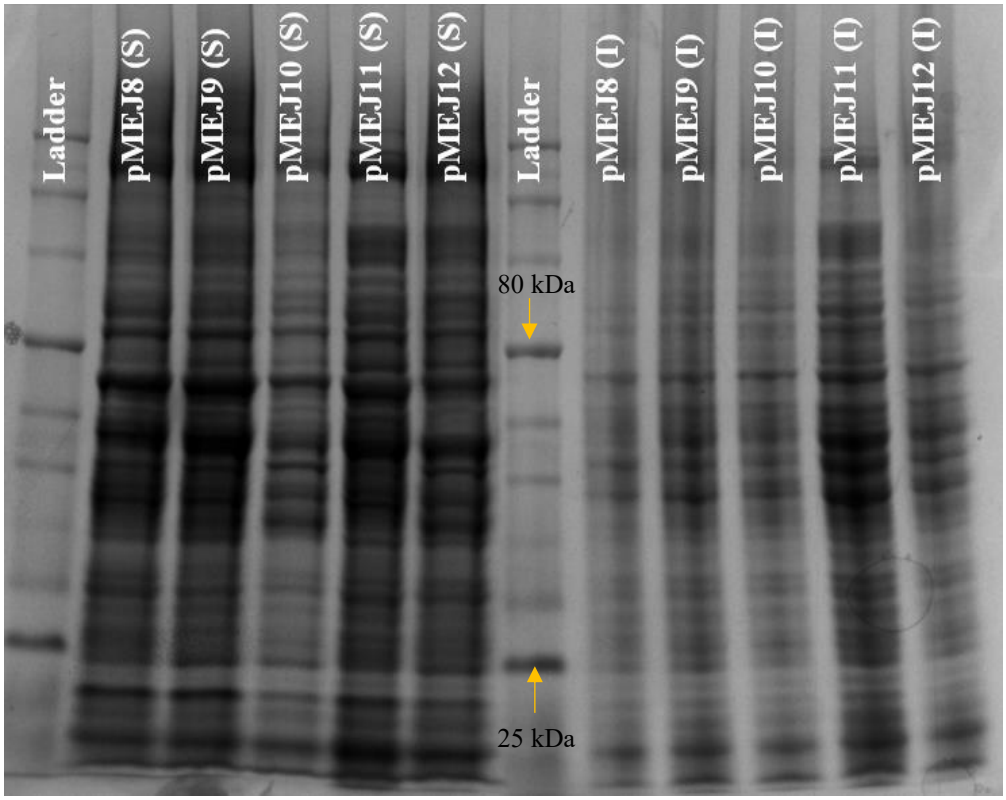


Figure 3.23: ***Protein gel of protein produced in*** **P. putida** ***using Protocol 2****. Protein was produced in* P. putida *using Protocol 2 and extracted in an insoluble (I) and a soluble (S) fraction. The proteins were separated on a RunBlue SDS protein gel (4-12%) from Expedeon.*

3.8 Expression Vector Containing Signal Sequence

As described in Section 1.5.3, a signal sequence might improve heterologous protein expression by both avoiding host proteases, facilitate transcription initiation and ensure correct protein folding. It was decided to make a new expression vector for production of the proteins in *P. putida* containing the signal sequence PelB and a kanamycin resistance gene. An mCherry gene was included behind the signal sequence, which was later to be replaced with the genes of interest. The expression vector was made by combining two vectors kindly provided by Vectron Biosolutions; pVB1 A1B1 and pVB1 mCherry Kan-271.

Both vectors were grown in and purified from *E. coli* DH5α. They were digested with AgeI and NdeI, and the fragments were separated on an agarose gel (see Figure 3.24). The smaller fragments were difficult to see on the gel, but the presumed 379 bp fragment of pVB1 A1B1 was cut out from the gel as well as the presumed 7702 bp fragment of pVB1 mCherry Kan-

271. These two fragments were purified from the gel and ligated to become the pMEJ13 plasmid. The ligation mixture was heat-shock transformed into *E. coli* DH5α.

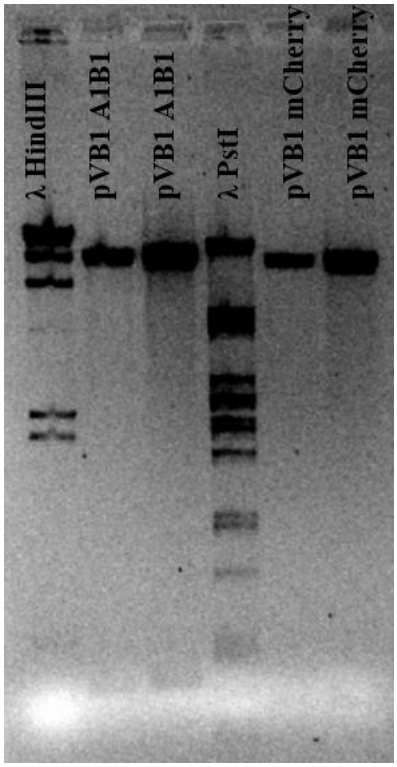


Figure 3.24: ***Digest of pVB1 A1B1 and pVB1 mCherry Kan-271 before cloning****. Plasmids pVB1 A1B1 and pVB1 mCherry Kan-271 were purified from* E. coli *DH5α and digested with AgeI and NdeI. The fragments were separated on an agarose gel. Theoretically, pVB1 A1B1 was expected to yield fragments of 8502 bp and 379 bp, while pVB1 mCherry Kan-271 was expected to yield fragments of 7702 bp and 310 bp.*

Four candidates for the newly formed pMEJ13 plasmid were purified from colonies of *E. coli* DH5α. The plasmid was verified by digestion with NheI and BamHI. The fragments were separated on an agarose gel (Figure 3.25 and Table 3.11). The gel indicates that all four colonies contained the correct plasmid, and c1 was chosen for further use.

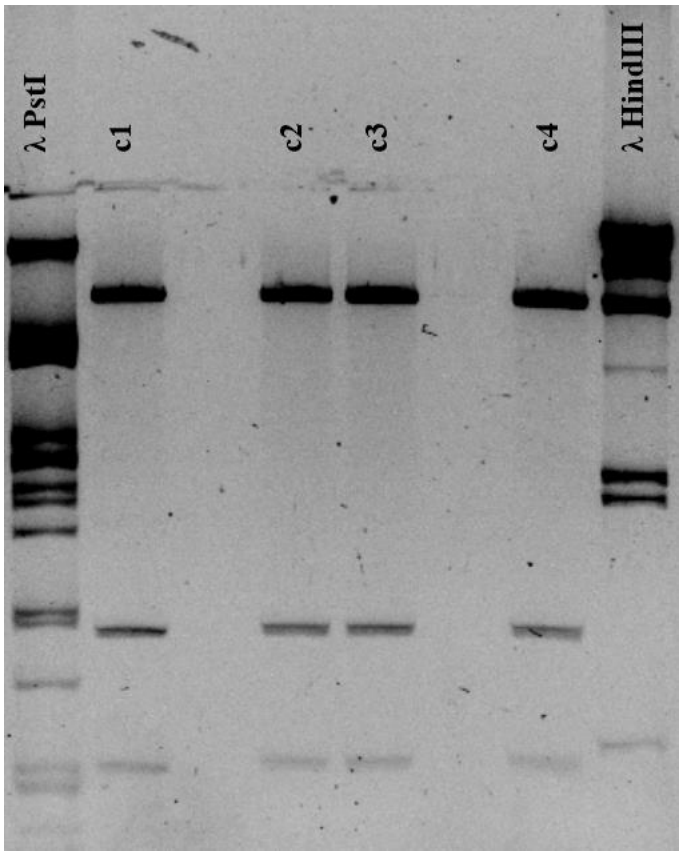


Figure 3.25: ***Verification digest of pMEJ13****. Plasmid pMEJ13 purified from four colonies of* E. coli *DH5α were digested with NheI and BamHI. The fragments were separated on an agarose gel.*

Table 3.11: ***Expected fragment sizes from digest of pMEJ13.*** *Theoretical fragment sizes from digest of pMEJ13 and its parent plasmids. Restriction enzymes are indicated. Parent plasmids for the new plasmid are marked green and placed directly underneath the new plasmid.*

Plasmid	Restriction enzymes	Expected fragments (bp)
pMEJ13	NheI + BamHI	6562 +1025 + 494
pVB1 A1B1	NheI + BamHI	6605 + 956 + 671 + 495 + 154
pVB1 mCherry Kan-271	NheI + BamHI	7518 + 494

This plasmid was also verified by sequencing using a primer adapted to start upstream of the promoter. Sequencing was performed by GATC as previously described. This verified the correct insertion of the promoter, the signal sequence and the start of the mCherry gene.

3.9 Vectors with Signal Sequence and Genes of Interest

The genes of interest were cut out from the set of vectors made for the first expression in *E. coli* (pMEJ3-7) and ligated into the new expression vector, pMEJ13.

pMEJ13 was digested with NdeI and XhoI or NdeI and DraIII. The fragments were separated on an agarose gel (Figure 3.26A and Table 3.12). The smaller fragments were difficult to see on the gel. The larger fragments were thus cut out, and the gel was exposed to more UV light (Figure 3.26B). At this higher degree of exposure, weak and broad bands at ~750 bp were observed for both digests. Even though the gel did not show clear bands, it was chosen to continue with the excised fragments and trust that later verification digests would uncover eventual mistakes made during this restriction.

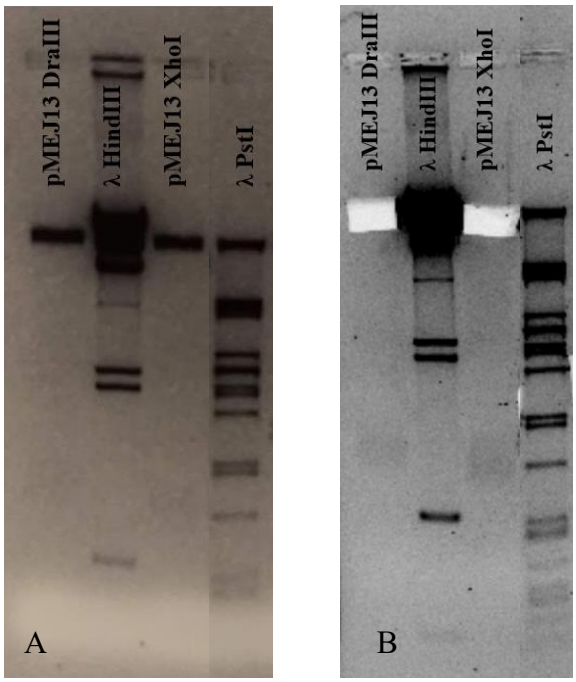


Figure 3.26: ***Digest of pMEJ13 before cloning****. Plasmid pMEJ13 was cut with NdeI and either DraIII and XhoI. The fragments were separated on an agarose gel. The figure is composed of two parts of the same gel. Part A shows the gel exposed to a low amount of UV light before the largest fragments were excised. Part B shows the same gel exposed to a higher amount of UV light.*

Table 3.12: ***Expected fragment sizes from digest of pMEJ13.*** *Theoretical fragment sizes from digest of pMEJ13. Restriction enzymes are indicated.*

Plasmid	Restriction enzymes	Expected fragments (bp)
pMEJ13	NdeI + DraIII	7301 + 780
pMEJ13	NdeI + XhoI	7365 + 716

pMEJ3, pMEJ5, and pMEJ6 were digested with NdeI and DraIII. pMEJ4 and pMEJ7 were digested with NdeI and SalI. The fragments were separated on agarose gels (Figure 3.27 and Table 3.13). This seemed correct according to the gels. The fragments containing the genes of interest (marked red in Table 3.13) were ligated with fragments from pMEJ13 according to Table 2.3. These plasmids were then heat-shock transformed into *E. coli* DH5α.

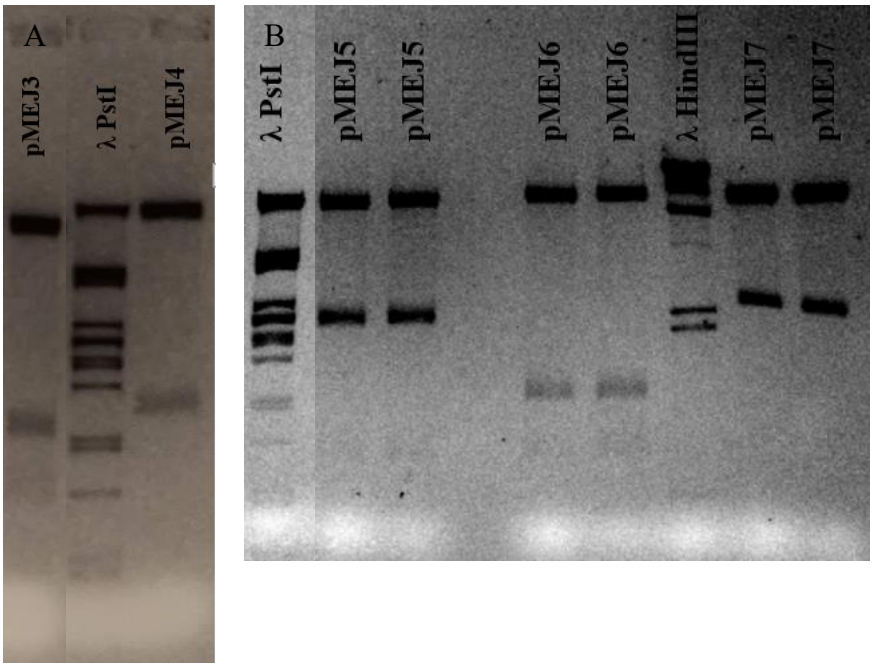


Figure 3.27: ***Digest of pMEJ3, pMEJ4, pMEJ5, pMEJ6, and pMEJ7 before cloning****. A) pMEJ3 was digested with NdeI and DraIII, and pMEJ4 was digested with NdeI and SalI. B) pMEJ35 and pMEJ6 were digested with NdeI and DraIII, and pMEJ7 was digested with NdeI and SalI. All of the digests were separated on agarose gels.*

Table 3.13: ***Expected fragment sizes from digest of pMEJ3, pMEJ4, pMEJ5, pMEJ6, and pMEJ7.*** *Theoretical fragment sizes from digest in bp. Restriction enzymes are indicated. The fragments containing the genes of interest are marked red.*

Plasmid	Restriction enzymes	Expected fragments (bp)
pMEJ3	NdeI + DraIII	6990 + 1149 + 777 + 464
pMEJ4	NdeI + SalI	8359 + 1333
pMEJ5	NdeI + DraIII	6990 + 2127 + 777 + 464
pMEJ6	NdeI + DraIII	6990 + 1155 + 777 + 464
pMEJ7	NdeI + SalI	8359 + 2299

To verify the newly formed plasmids, plasmids were purified from four different colonies of *E. coli* DH5α. The plasmid candidates were digested by enzymes and separated on agarose gels (Figure 3.28 and Table 3.14). All candidates seemed correct according to the gels, and candidate 1 was chosen for all the plasmids for further use.

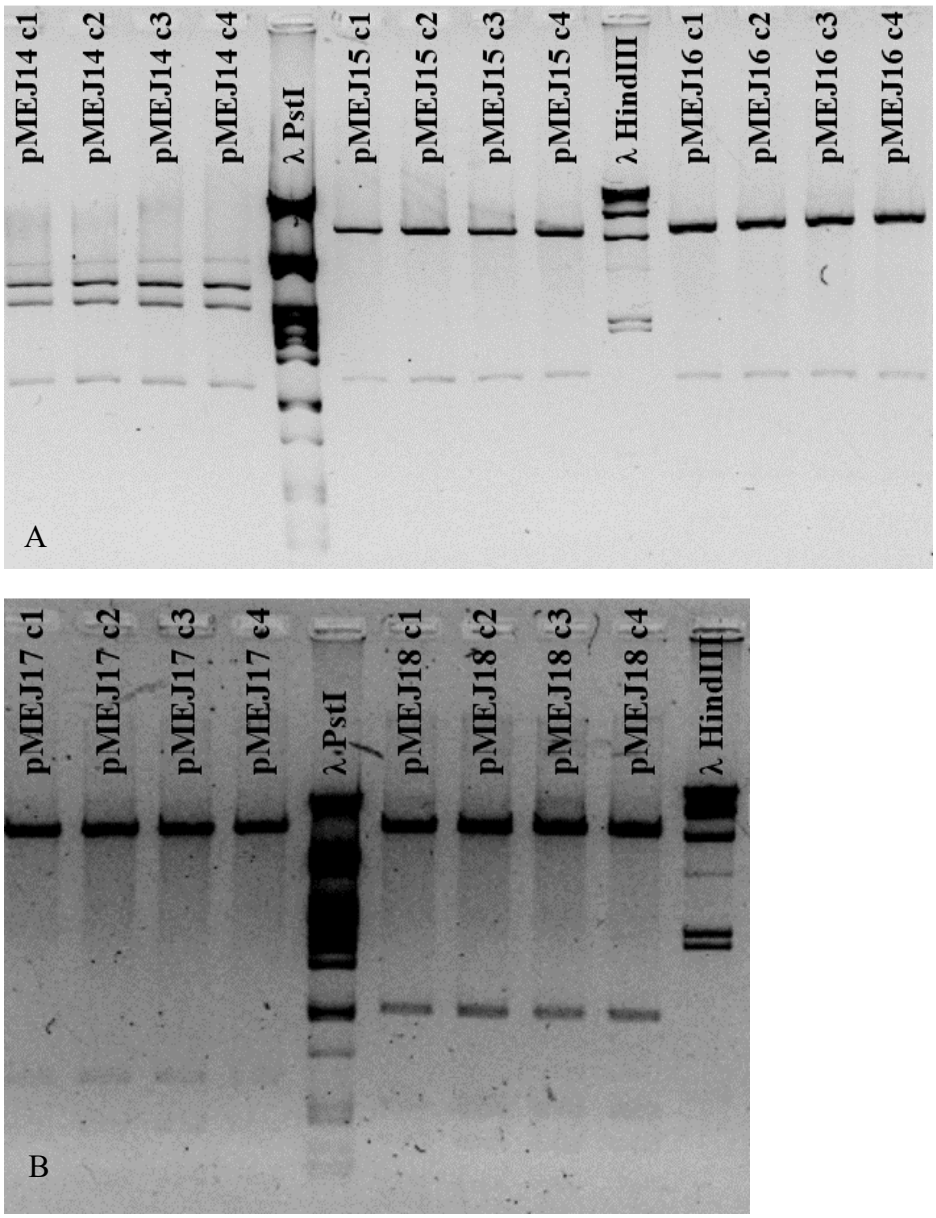


Figure 3.28: ***Verification digests of pMEJ14, pMEJ15, pMEJ16, pMEJ17, and pMEJ18****. A) Four candidates for pMEJ14 were digested with HindIII and XbaI, and for candidates for pMEJ15 and pMEJ16 were digested with NheI and XhoI. B) Four candidates for pMEJ17 and pMEJ18 were digested with NheI and XhoI. All the digests were separated on agarose gels.*

Table 3.14: ***Expected fragment sizes from digest of pMEJ14, pMEJ15, pMEJ16, pMEJ17, and pMEJ18 and their parent plasmids.*** *Theoretical fragment sizes from digested plasmids and their parent plasmids. Restriction enzymes are indicated. Parent plasmids for the new plasmid are marked green and placed directly underneath the new plasmid. pMEJ13 was one of the parent plasmids for all the new plasmids.*

Plasmid	Restriction enzymes	Expected fragments (bp)
pMEJ14	HindIII + XbaI	3394 +2634 + 1177 + 1245
pMEJ3	HindIII + XbaI	6071 + 2134 + 1175
pMEJ13	HindIII + XbaI	3394 + 2634 + 1177 + 876
pMEJ15	NheI + XhoI	7205 + 1340 + 153
pMEJ4	NheI + XhoI	9539 + 153
pMEJ13	NheI + XhoI	7164 + 917
pMEJ16	NheI + XhoI	7269 + 1287 + 458 + 243 + 171
pMEJ5	NheI + XhoI	8657 + 1287 + 243 + 171
pMEJ17	NheI + XhoI	7269 + 632 + 392 + 164
pMEJ6	NheI + XhoI	8591 + 631 + 164
pMEJ18	NheI + XhoI	7205 + 1169 + 526 + 375 + 185 + 123 + 81
pMEJ7	NheI + XhoI	9368 + 526 + 375 + 185 + 123 + 81

3.10 Protein Production in *P. putida* with Vectors Containing PelB

Plasmids pMEJ13, pMEJ14, pMEJ15, pMEJ16, and pMEJ17 were electroporated into competent *P. putida* cells. The attempt to electroporate pMEJ18 into *P. putida* was unsuccessful. pMEJ13 was included as a control. However, this control was not codon optimized for *P. putida*, and would presumably be more successfully expressed in *E. coli*.

The *P. putida* containing the plasmids were grown following Protein Production Protocol 2. The protein was extracted using B-PER. Both the soluble and insoluble fractions were separated on an SDS-PAGE gel (Figure 3.29). $1/10^{th}$ of the insoluble fraction was loaded on the gel compared to the amount of soluble fraction. The expected molecular weights were the same as indicated by Table 3.8 when it is taken into account which genes each plasmid encoded (see Table 2.3). The expected molecular weight of mCherry was 26.7 kDa. According to the gel, neither the proteins of interest nor mCherry was produced in significant amounts.

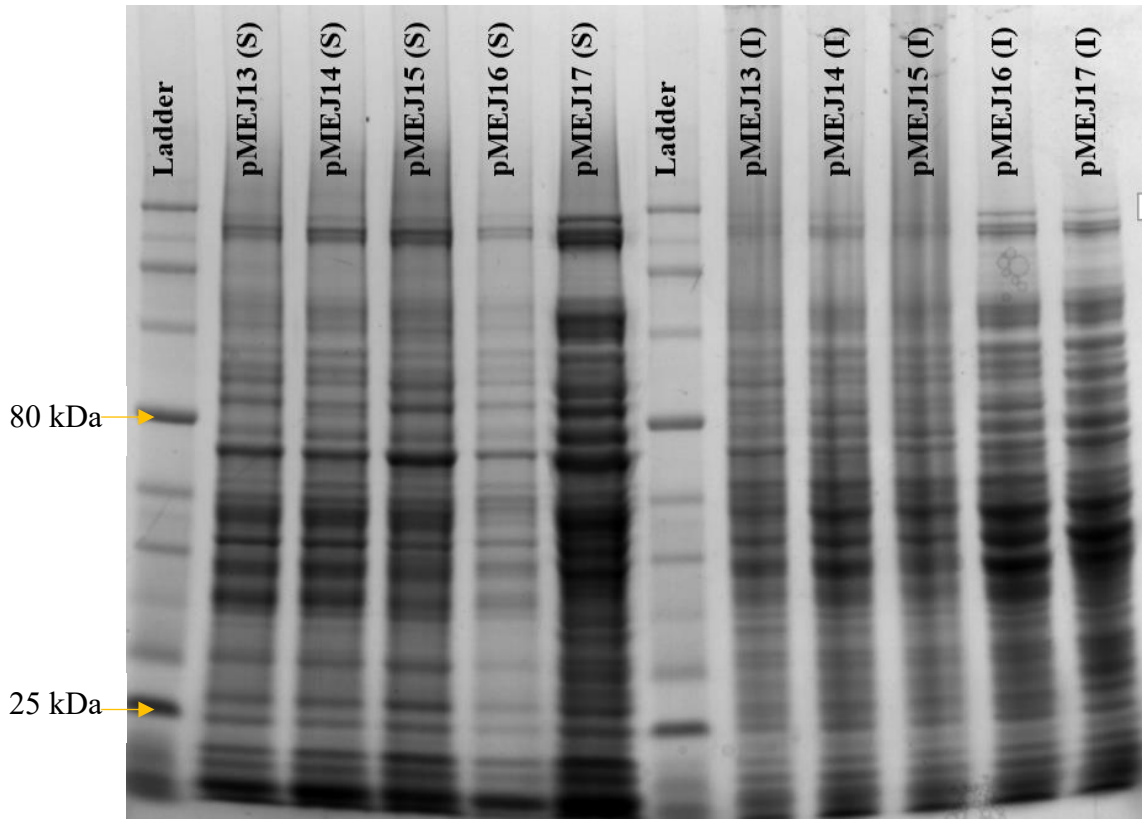


Figure 3.29: ***Protein gel of protein produced in* P. putida**. *Protein was produced in* P. putida *and extracted in an insoluble (I) and a soluble (S) fraction. These proteins were separated on a RunBlue SDS protein gel (4-12%) from Expedeon.*

3.11 Protein Production in *E. coli* RV308 with Vectors Containing PelB

pMEJ13, pMEJ14, pMEJ15, pMEJ16, pMEJ17, and pMEJ18 were electroporated into *E. coli* RV308. The resulting strains were cultivated using the growth and purification protocol for *E. coli*. The protein was subsequently extracted using B-PER.

Two identically loaded SDS-PAGE gels with the soluble fractions were run. However, even though the loading of the gels were the same, the gels were not run at the same length of time. As positive control an extract from *E. coli* RV308 with plasmid pMSM2 encoding a His-tagged protein of 82.3 kDa was used. One gel was stained using InstantBlue as previously (Figure 3.30). The other was used for Western Blotting (Figure 3.31).

The proteins of interest were not detected by neither electrophoresis nor by Western Blotting, but a band appeared in the column where the pMSM2 control protein was loaded.

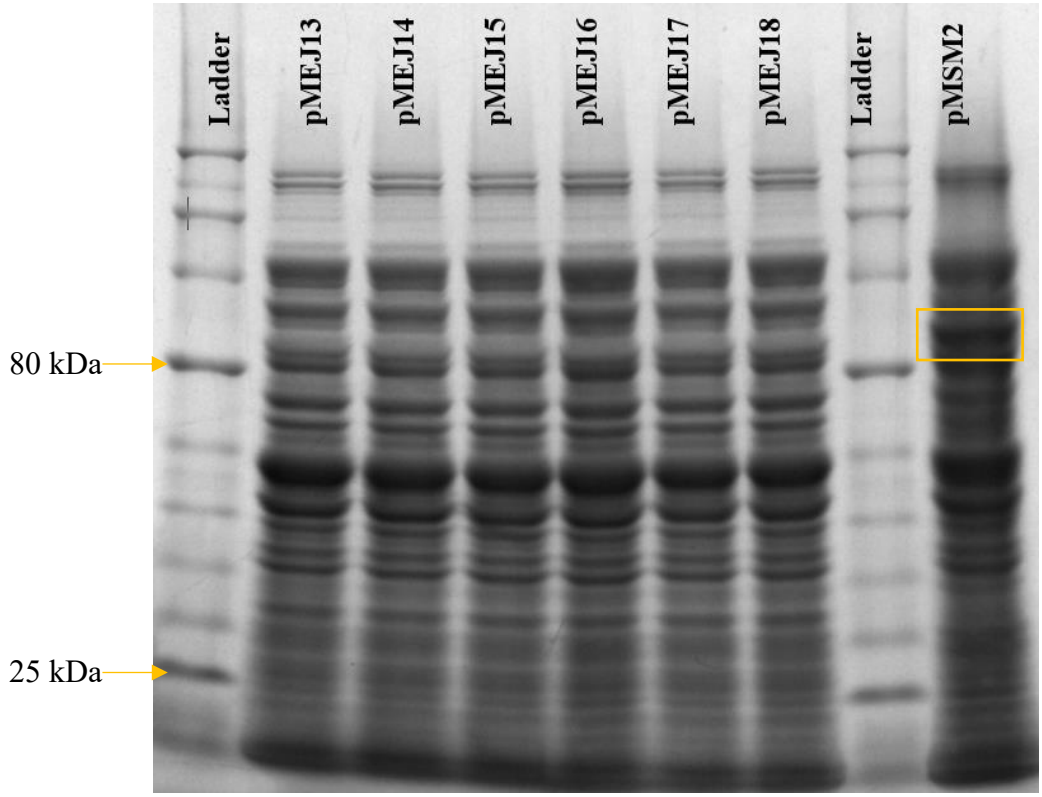


Figure 3.30: ***Protein gel of protein produced in*** **E. coli** ***RV308****. Protein was produced in* E. coli *containing pMEJ13, pMEJ14, pMEJ15, pMEJ16, pMEJ17, pMEJ18 or pMSM2. The soluble fractions were separated on a RunBlue SDS protein gel (4-12%) from Expedeon. The band of the control protein is marked yellow.*

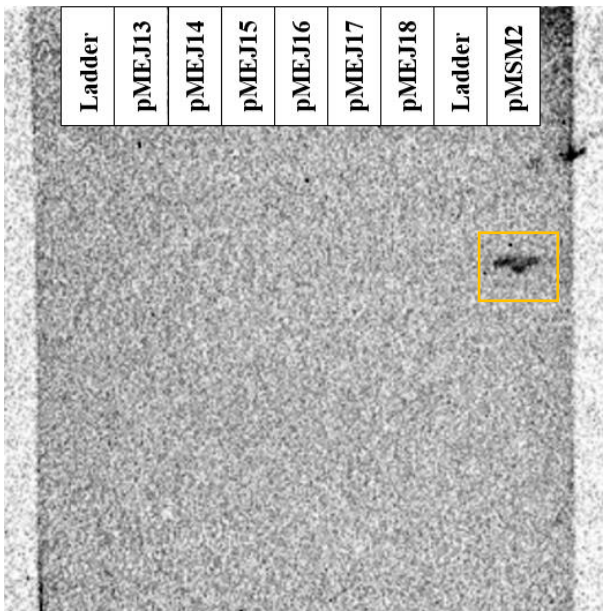


Figure 3.31: ***Western Blot of protein from E. coli RV308****. A Western Blot was performed using an antibody binding to His-tags, HisProbe™-HRP.*

3.12 Codon Bias Evaluation

Because of the problems in expressing proteins in *E. coli* and *P. putida*, it was investigated whether the codon bias (see Section 2.1.4) could be the cause of these problems. This was evaluated for the translation of the first part of the genes of interest with regards to *E. coli* (Table 3.15) and *P. putida* (Table 3.16). The frequency of appearance of each codon per 1000 codons are shown. This number will be low both in the case of rare codons and of rare amino acids. However, rare amino acids could also be relevant constraints, as this might also lead to low levels of the tRNAs related to the translation of these (Section 1.5.5). As shown in the tables, some codons could potentially pose problems. However, especially in the case of gene 6817 and 9329 there should be no codon bias in either host organism.

Table 3.15: ***Codon bias evaluation in* E. coli *for genes of interest****. The first codons in the sequence of the genes of interest are shown. Codons marked yellow appear less than 1/1000, codons marked red appear 1-2 times per 1000 codons, and codons in green appear 2-5 times per 1000 codons.*

Gene	Start of gene sequence
8081	ATGGCGGCGAGGATGACGCTGCGGAACCTGGGCCGCGCG
	AGAACGCCGCCCTCGCGGCGGCTCCTGTCCACGCGGACG
	CCCGCCGCCGCCGCCGCCGCCGCGCCAGGGACGTCGGCG
10760	ATGTCGGCCCTGGGAGGATTCTCGCGTGTGGCGTCGAGC
	GTCGGACGTTCGGCGCGCGGCGCGCAGCTGCTCCGCCAC
	ACCCAGACGCGCAAGCTCAACATCCATGAGTACGCGAGC
6817	ATGCTCGCGCGTACCGTGACCAAGACCCGCCAGGCGGCG
	ACGCTGACGGCGACCCGTGCCATGTCGGCCACGGCCAAC
	GTGTGGGTGGACAAGAACACCAAGGTCCTCTGCCAGGGC
6864	ATGCAGTCCCTCCGGCGCGCATCAAGGCCCGCCGGCGTG
	GCCGCCGTGTCTTTCGTGGCGGCAGCAGGCGTGGCCGTG
	GCCGCCTTTGGCGAGCGCAGCGAAGACGACACGGGCAAT
9329	ATGACGATGTCGCCCGCCCGGACCGTGGCGGCCGCCGCG
	GGCCTCTCGACGGCGCCGGGCGGGCCCGGGCTCGCCGCC
	GAGGACGCCTCGGTGGCGTCGGTGCGCCGCAGCGTGCTG

Table 3.16: ***Codon bias evaluation in* P. putida *for genes of interest****. The first codons in the sequence of the genes of interest are shown. Codons marked yellow appear less than 1/1000, codons marked red appear 1-2 times per 1000 codons, and codons in green appear 2-5 times per 1000 codons.*

Gene	Start of gene sequence
8081	ATGGCGGCGAGGATGACGCTGCGGAACCTGGGCCGCGCG
	AGAACGCCGCCCTCGCGGCGGCTCCTGTCCACGCGGACG
	CCCGCCGCCGCCGCCGCCGCCGCGCCAGGGACGTCGGCG
10760	ATGTCGGCCCTGGGAGGATTCTCGCGTGTGGCGTCGAGC
	GTCGGACGTTCGGCGCGCGGCGCGCAGCTGCTCCGCCAC
	ACCCAGACGCGCAAGCTCAACATCCATGAGTACGCGAGC
6817	ATGCTCGCGCGTACCGTGACCAAGACCCGCCAGGCGGCG
	ACGCTGACGGCGACCCGTGCCATGTCGGCCACGGCCAAC
	GTGTGGGTGGACAAGAACACCAAGGTCCTCTGCCAGGGC
6864	ATGCAGTCCCTCCGGCGCGCATCAAGGCCCGCCGGCGTG
	GCCGCCGTGTCTTTCGTGGCGGCAGCAGGCGTGGCCGTG
	GCCGCCTTTGGCGAGCGCAGCGAAGACGACACGGGCAAT
9329	ATGACGATGTCGCCCGCCCGGACCGTGGCGGCCGCCGCG
	GGCCTCTCGACGGCGCCGGGCGGGCCCGGGCTCGCCGCC
	GAGGACGCCTCGGTGGCGTCGGTGCGCCGCAGCGTGCTG

3.13 Fermentation Experiment

Due to the failure to heterologously express the enzymes of interest, it was decided to measure the enzyme activity in *Aurantiochytrium* T66. A fermentation experiment was set up as described in Section 2.7.1. Three biological parallels of both *Aurantiochytrium* T66 (flask 1, 2, and 3) and *Schizochytrium* S21 (flask 4, 5, and 6) were inoculated.

3.13.1 Growth Curve, Nitrogen Content and Total Lipid Content

OD_{600} was measured approximately twice a day during the fermentation. The growth curve for each flask can be seen in Figure 3.32. The time points for enzymatic assay sampling are indicated by blue arrows, where the first one is taken in the exponentiental faze (28 h) and the second in the shift from exponentiental towards stationary faze (66 h).

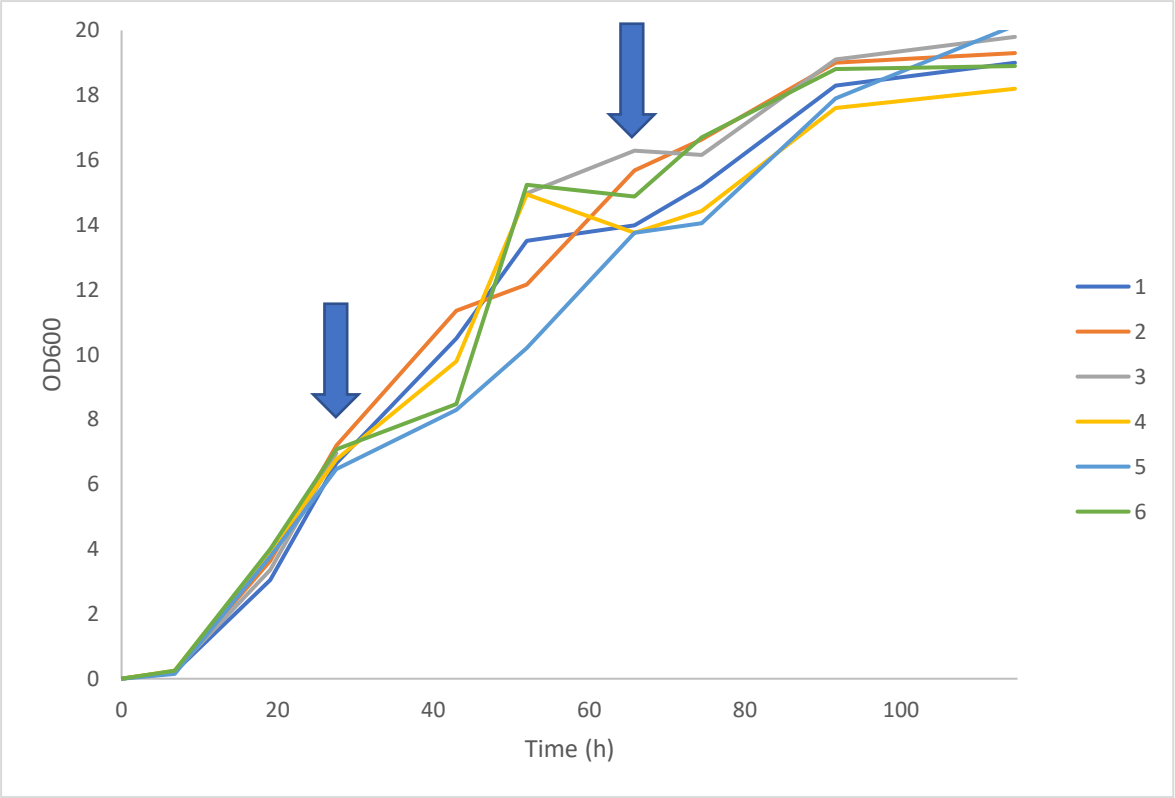


Figure 3.32: ***Growth curve of flasks with S21 and T66****. The growth curves are based on OD measurements at 600 nm as a function of time (hours). Flask 1, 2, and 3 contained T66, while flask 4, 5, and 6 contained S21. The blue arrows indicate the time points where enzyme samples were taken.*

The nitrogen content samples could not be processed due to the delayed delivery of the kit intended to measure this. It was thus decided to look to the assimilable nitrogen data in fermentation of *Schizochytrium* S21 presented by Chang et al. (2013). They reported that the nitrogen was completely spent after 18 h of fermentation while using a similar growth media as the one used for the fermentation experiment in this thesis.

The lipid content at the end of the fermentations was calculated to be an average 59.0 ± 1.3 (w/w) % for T66 and 64.3 ± 1.5 (w/w) % for S21 of the dry weight for the three parallels.

3.13.2 Sonication Test

A sonication test was performed as described in Section 2.7.5. The test samples were the precultures of T66 and S21 in YPD media (OD_{600} = 6.37 and 8.86, respectively) and a culture of S21 grown in the lipid accumulation media with TMS1 and Vitamin Mix 1 (OD_{600} = 25.8). The protein concentration as a function of sonication time, by which the actual pulse time is meant, is shown in Figure 3.33. By the flattening of the curves of the cultures grown in the

YPD media, the S21 cells seemed completely lysed around 300 s, while the T66 cells seemed completely lysed around 500 s. On the other hand, the very low protein yield of S21 grown in lipid accumulation media despite its significantly higher OD than the other cultures indicated that a complete lysis of the cells were not achieved under these conditions.

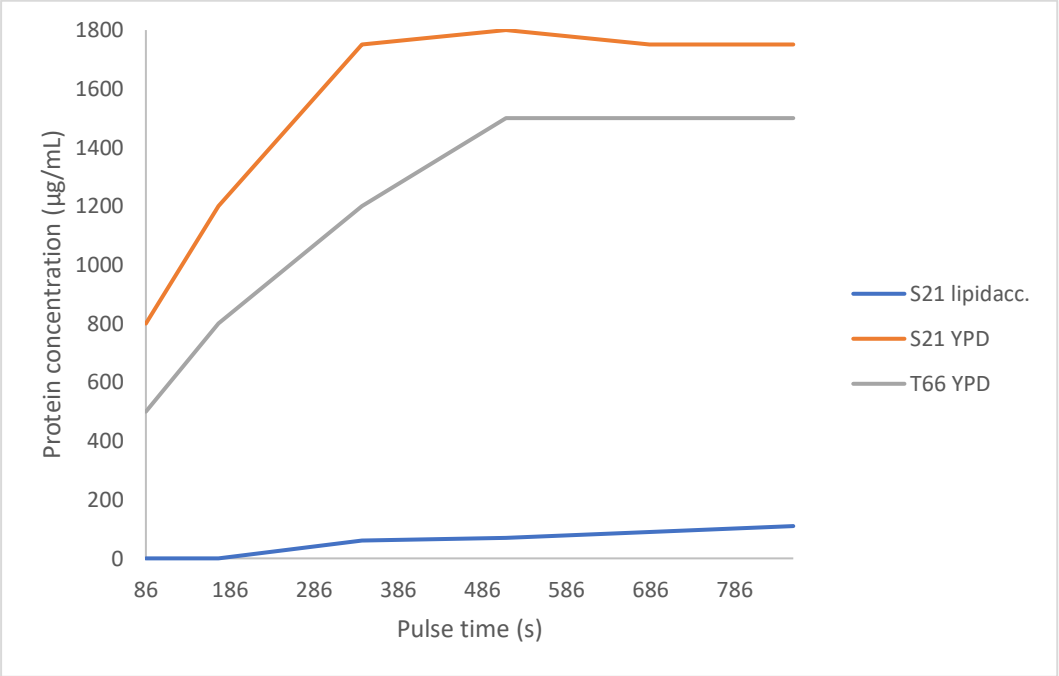


Figure 3.33: ***Protein concentration as a function of sonication time for sonication trial****. The protein concentration was measured using the Bradford Method at several points during the sonication of the samples.*

3.13.3 Sonication of Enzyme Assay Samples

Samples for enzyme assays were taken at 28 h and 66 h of fermentation and prepared as described in Section 2.7.5. Because challenges with sonication were detected but not solved, the samples were sonicated unequally in an attempt to achieve better protein yields. During the sonication of the samples, protein concentration was measured for some samples. The total pulse time and final protein concentrations of the samples from the different flasks are indicated in Table 3.17 (28 h) and Table 3.18 (66 h). For the samples taken at 28 h, the resting time between sonication was varied as indicated by the table. For the samples taken at 66 h, some samples were sonicated in a diluted state (diluted to ½ of the original concentration) for 50% or 100% of the sonication time, as indicated by the table.

The yields of the samples taken at 28 h were around 200 µg/mL, while the yields from 66 h were 150 µg/mL or less.

Table 3.17: ***Protein concentration in samples for enzyme assays taken at 28 h****. Protein concentration in µg/mL of the sonicated enzyme assay samples from flasks 1-6 were estimated by the Bradford Method. Total pulse time as well as time of each pulse ("pulsation time") and the rest time between each pulse is given in seconds.*

Flask	Total pulse time (s)	Pulsation time (s)	Rest time (s)	Protein concentration (µg/mL)
1	1350	4	10	200
2	900	4	4	170
3	1200	4	4	200
4	900	4	10	190
5	900	4	4	210
6	1200	4	4	210

Table 3.18: ***Protein concentration in samples for enzyme assays taken at 66 h****. Protein concentration in µg/mL of the sonicated enzyme assay samples from flasks 1-6 were estimated by the Bradford Method. Some samples were sonicated in a diluted state (diluted to ½ the concentration) for 50 % or 100 % of the sonication time.*

Flask	Total pulse time (s)	Protein concentration (µg/mL)	Percent of sonication as diluted
1	2700	150	
2	1800	120	
3	1800	70	50 %
4	1800	70	100 %
5	1800	130	
6	1800	140	

3.13.4 Enzyme Activity

Control assays were performed with commercially available enzymes to verify the assay protocols (Appendix C – Enzyme Assay Controls). These data are only intended to be of a qualitative and not quantitative nature, and the experiments were only performed for one technical parallel.

G3PDH

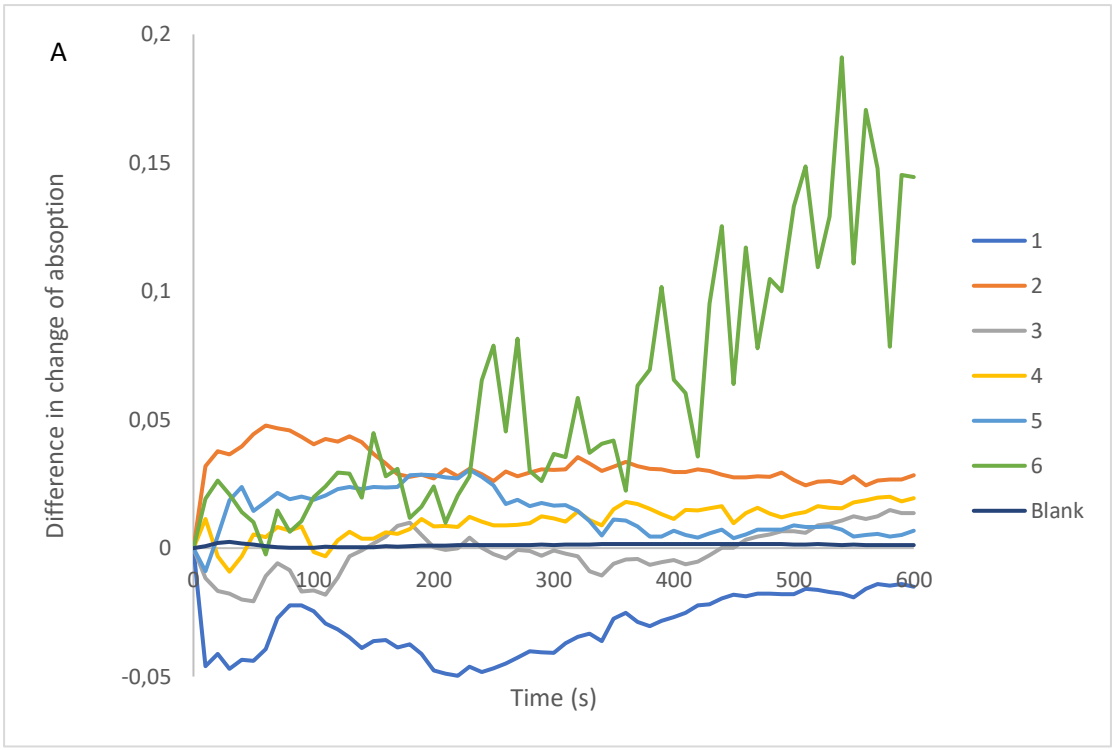
The G3PDH assays were performed as described in Section 2.8.1 with either NADH or NADPH as cofactors. All assays were performed in 3 technical replicates, including blank

samples containing only 0.1 M Tris-HCl (pH 7.4) buffer. Assays were performed both with and without the substrate (DHAP).

The average value of the three replicates was calculated. In some cases, one parallel was removed due to large variation from the other two. The average change in absorbance as a function of time was calculated by subtracting the average absorption at time zero from all the later absorbance measurements. The average change in absorbance without substrate was subtracted from the average change in absorbance with substrate. This gave the difference in change of absorption as a function of time. An increasingly negative function would indicate a larger consumption of cofactor in the samples with substrate than in the ones without substrate, while a function parallel with the x-axis would indicate no increased cofactor consumption in the presence of the substrate. Increasingly negative trends would thus indicate a detected G3PDH activity.

The difference in the change of absorption for G3PDH assays with NADH at 28 h and 66 h is shown in Figure 3.34. The results from flask 6 at 28 h, one of the S21 parallels, showed large variations in the measurements during the entire duration of the assay. All the parallels showed large variation in absorbance during the first 30-50 s. This variation was not similar in the flasks containing the same organisms; thus it was suspected that the mixing at initiation was not completely achieved. Apart from the measurements done just after initiation, the difference in change in absorption seemed to be parallel with the x-axis. This indicated no observable G3PDH activity.

The scale on the y axis is the same for the assays at the two different time points. The variation in the data at the first time point is thus observed to be larger than the variation at the second time point. The decrease in variance might be due to the additional filtration of the raw extract before the measurements at the second time point. The blank samples, where only pure buffer was added to the reaction mixture, does not exhibit this large variation. This indicates that the variation is caused by something in the enzyme samples.



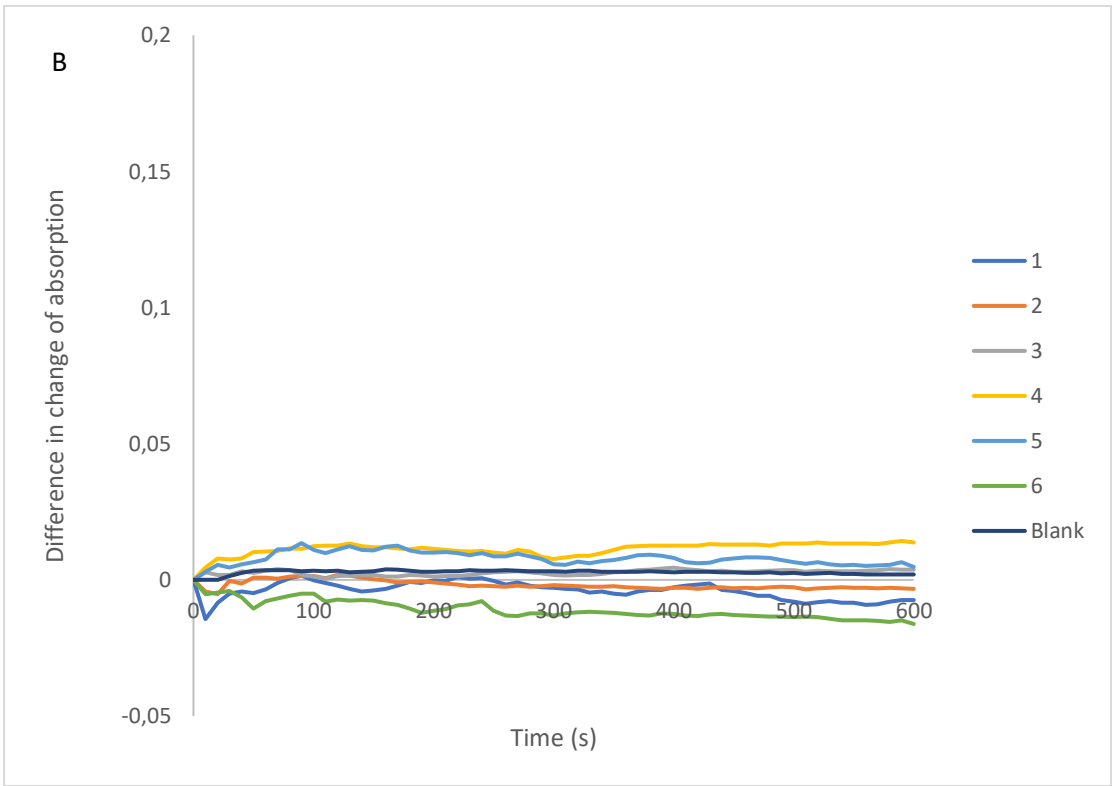


Figure 3.34: ***Difference in change of absorption in G3PDH assay with NADH at 28 h (A) and 66h (B)****. The average change in absorbance without substrate was subtracted from the average change in absorbance with substrate for the enzyme samples from flask 1-6 and for the blank samples.*

The difference in the change of absorption for G3PDH assays with NADPH at 28 h and 66 h is shown in Figure 3.35. Similar observations were made for these assays as for the assays with NADH. This was both with regards to the irregular results for flask 6 at 28 h, the large variation in absorbance just after initiation, the x-axis parallel functions after the first 30-50 s and the significantly smaller variation in the data from 66 h.

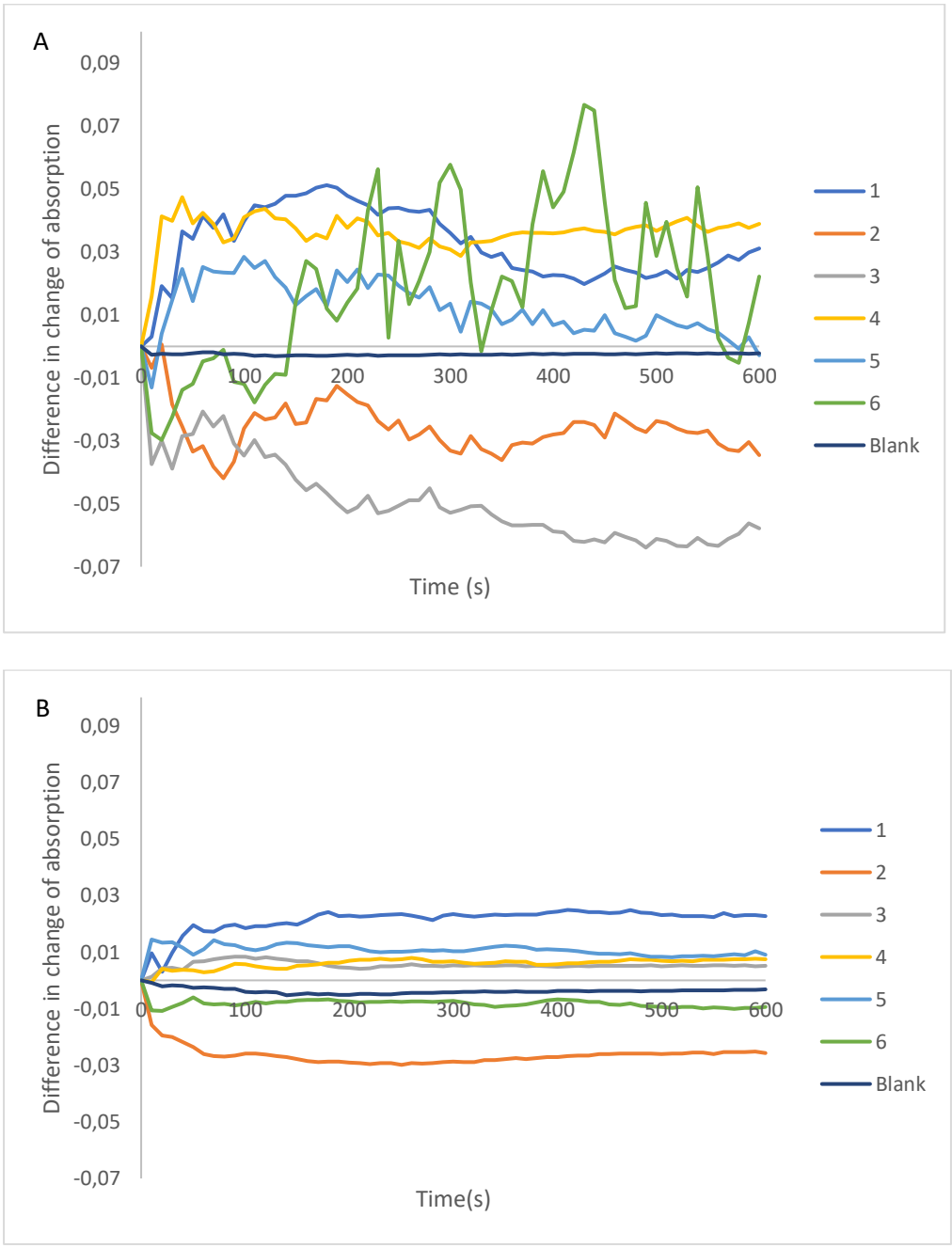


Figure 3.35: ***Difference in change of absorption in G3PDH assay with NADPH at 28 h (A) and 66h (B)****. The average change in absorbance without substrate was subtracted from the average change in absorbance with substrate for the enzyme samples from flask 1-6 and for the blank samples.*

ACL

The ACL assays were performed as described in Section 2.8.2. All assays were performed in 3 technical replicates, including blank samples containing only 0.1 M Tris-HCl (pH 7.4) buffer. Assays were performed both with and without the substrate (citrate).

Difference in change of absorption was calculated as for the G3PDH results. The difference in change of absorption for ACL assays at 28h and 66h is shown in Figure 3.36.

For the results from 28h, the variation observed in the dataset is too large to point out any apparent enzyme activity. This variation does not appear for the blank samples, and was therefore probably caused by something related to the raw extract. The results for 66h exhibited, like the ones described for the G3PDH assays, considerably less variation. This applies to all samples apart from the ones from flask 1. A slight negative trend could be observed at 66h, which could indicate an ACL activity for both the T66 and the S21 samples.

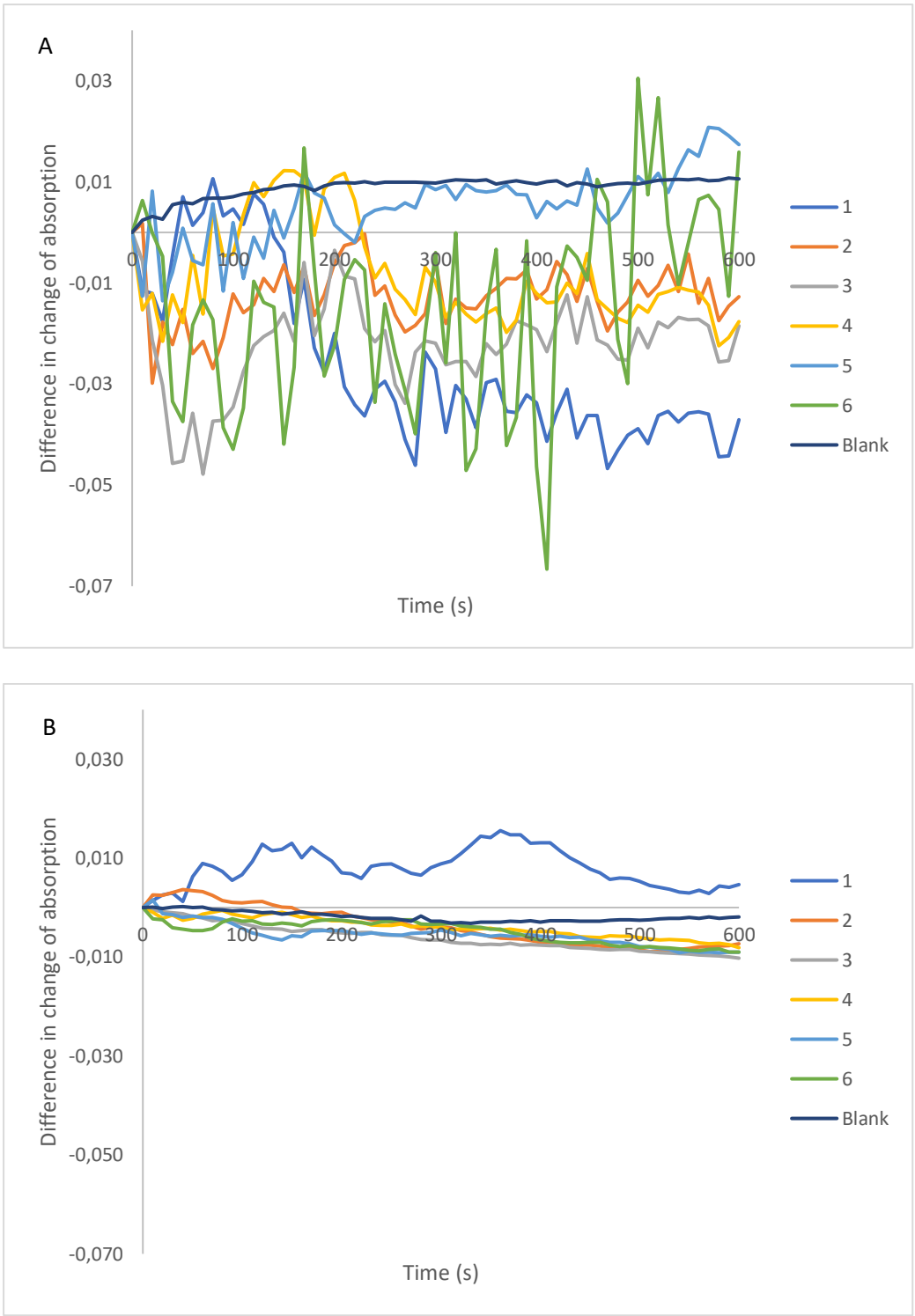


Figure 3.36: ***Difference in change of absorption in ACL assay at 28 h****. The average change in absorbance without substrate was subtracted from the average change in absorbance with substrate for the enzyme samples from flask 1-6 and for the blank samples.*

4 Discussion

Even though heterologous gene expression in theory is rather straight forward, this thesis is an example of how difficult it can be. As emphasized by Kaur et al. (2018), heterologous expression needs to be optimized for each protein to be expressed. In this thesis it was aimed to express five genes encoding proteins. In one way, this was expected to be a difficult task, as it could mean that five different optimized protocols needed to be developed. On the other hand, it would increase the chances of successfully managing to express at least one of them, assuming the eventual expression problems were not exactly the same for all five peptides.

The first attempt at heterologous expression was to use an expression system proven efficient in the first choice of host organism; the NTNU developed *Pm*/XylS promotor system in *E. coli* RV308. The production and purification of the control protein were successfully performed, thus confirming that the protocol itself was not the cause of the expression challenges. When expression failed at the normally used growth temperature of *E. coli*, 37 ˚C, a decrease in temperature after induction was attempted. The lowering of growth temperature in *E. coli* can be a method to avoid he formation of inclusion bodies. The aggregate can however be solubilized with compounds such as SDS (Kaur et al., 2018). This would indicate that even if the proteins were expressed and caught in inclusion bodies preventing His-tag purification, the proteins should nevertheless appear on a protein gel where samples were denatured with SDS. It was probably not an inclusion body formation causing the failure to express and purify the proteins, because the proteins did not show up on the protein gels in neither the soluble nor the insoluble fractions. The lowering of temperature could also have improved translation efficiency. If this was in fact improved during the experiment, the process must still have been impaired at another stage and thus preventing protein formation.

The GC-content of *E. coli* is relatively low (Kaur et al., 2018). On the other hand, the GC-content of *Aurantiochytrium* T66 is high. *P. putida* was chosen as the next attempted host organism, partly because of its similarity with regards to GC-content to T66. *P. putida* has also been shown to have a metabolism quite different from the metabolism of *E. coli*, for instance with regards to their regulation of the stress response factors RpoS and Crp (Belda et al., 2016). Toxicity of heterologously expressed protein is a known problem (Kaur et al., 2018), and the use of host organisms with differences in stress response and other metabolic features could be helpful to overcome this.

Two different protocols were used to test the expression in *P. putida*. One used a high OD at the time of induction and a high concentration of the inducer, while the other used a lower OD and a lower concentration of the inducer. It was suggested by Kaur et al. (2018) that a high OD might be advantageous if the protein were toxic to the organism. The inducer concentration was also mentioned as an important factor, where a low concentration could give a low yield and a high concentration could give inclusion body formation. However, neither of the protocols tested seemed to be effective. Because the proteins did not show up on the protein gel, inclusion body formation was not likely the problem. A more thorough study could have been performed, where several ODs and inducer concentrations could have been tested. However, due to a limited amount of time, it was chosen to attempt another method.

The translocation of a heterologously expressed protein to the periplasm of the host organism by attachment of a signal peptide may solve several expression problems, for instance degradation by peptidases. In this case, PelB was used. Translation may also be improved by the use of a signal peptide. Unfortunately, the use of a signal peptide did not yield any positive results in neither *E. coli* nor in *P. putida*.

The codon bias was further evaluated as a possible reason for the expression problems. As mentioned above, using a host with a more similar GC-content did not yield recombinant proteins. It was thus decided to look more closely at the specific codons used in the beginning of each peptide. Even though the beginning of at least two of the peptides apparently should not have caused problems in either host, translation of the peptide could be impaired at a later stage in translation. If a codon is rare for the host, it might not have the corresponding tRNAs available in sufficient amounts to translate the mRNAs from the heterologously expressed genes. For future attempts to express these genes, codon optimization of the genes where rare codons were found could have been attempted. Another possibility could be to coexpress the genes encoding the tRNAs rarely used by *E. coli* and *P. putida*.

The experiments conducted in this thesis, could not determine at which stage in the protein expression cycle the problems occurred. However, the formation of inclusion bodies was excluded. The remaining points of challenge would thus be transcription, translation, and degradation. The question about whether the genes were in fact transcribed could be further investigated by a transcriptional analysis. This could be performed by purifying the RNA from a host cell lysate, and then run RT-PCR with primers for each gene of interest. If this

indeed was the problem, the efforts to express the genes could be concentrated on transcription. Experiments focused on post-transcriptional optimization could then be put on hold until the transcription hurdle was overcome.

Chaperones might be both the cause of and solution to the expression challenges. Proteins can be marked for degradation by chaperones (Saibil, 2013). The use of mutant prokaryotic hosts lacking some chaperones have been shown to yield better expression of eukaryotic proteins. However, these mutant hosts may themselves suffer from the lack of these chaperones and accumulate inclusion bodies of both the recombinant protein and of native host proteins (Nishihara et al., 1998). Because the intent of this study was to verify expression and activity of the enzymes in question, not to produce large amounts of the enzymes, the mutant strains might still be viable enough to produce the amounts of recombinant protein needed to perform the assays.

Programs have been developed that are to a certain extent able to predict chaperone requirements of proteins based on their sequences (Klus et al., 2014). Some computer models can also identify the binding sites of the *E. coli* chaperone component DnaK on a protein based on its sequence. DnaK is part of the native quality control system in *E. coli* (Van Durme et al., 2009). For future attempts to express the genes of interest, such bioinformatic studies could be relevant. Chaperones might also be expressed along with the recombinant protein (Nishihara et al., 1998). Some types of chaperones that bind non-specifically are found in different versions in a wide range of organisms. However, some proteins are very specific in their chaperone and cochaperone requirements. An example of this is the substrates of the chaperonins, a chaperone family that encloses the peptide inside itself, providing a special environment ensuring the correct folding. These are typically found in eukaryotes (Saibil, 2013), which might pose problems when attempting expression in prokaryotic hosts.

As some proteins have very specific chaperone requirements, it could be imagined that a chaperone from *Aurantiochytrium* T66 itself was needed to ensure the correct folding of the expressed genes. Courtesy of the bioinformatic part of the Auromega project (in particular Vetle Simensen), 120 ORFs possibly encoding chaperones have been identified in the sequenced genome of *Aurantiochytrium* T66. It would be a substantial task to attempt the coexpression of all of these, especially since chaperone function often depends on the presence of several peptides (Saibil, 2013). The number of chaperone candidates to investigate could be substantially decreased by investigation of cotranscription of the

suspected chaperones with the ORFs of interest in T66. Such investigations could also provide further data concerning the growth faze where the ORFs of interest were expressed.

Another suggestion to achieve protein expression is to coexpress the peptide of interest with another protein, known as a fusion protein (Hannig and Makrides, 1998). A coexpressed peptide can be cleaved off by designed peptidase cleavage sites between the two peptides (Kaur et al., 2018). This could be a solution if the problem was that the peptides expressed were recognized as foreign and degraded by host proteases.

In some cases, prokaryotic hosts will not be suited for expression of eukatyotic proteins. Sometimes, eukaryotic hosts should be chosen due to their ability to facilitate certain posttranslational modification such as glycosylation (Kaur et al., 2018). Eukaryotic cells might also contain the chaperones necessary for the correct folding, thus avoiding degradation. Various eukaryotic expression hosts and systems have been developed, such as expression in the yeast *Saccharomyces cerevisiae* (Sampaio et al., 2011) or the use of the balculovirus expression system in insects (King, 1992). However, these systems exhibit large variations from the system described in this thesis. The attempt to use eukaryotic hosts would thus not be feasible in the work on this thesis, but it would certainly be a promising choice for further attempts to express the peptides in question.

It is worth noting that Chávez-Cabrera et al. (2015) successfully expressed active ACL from the yeast *Phaffia rhodozyma* in *E. coli*, thus proving that some eukaryotic ACL enzymes could be expressed in prokaryotes. This makes the question about the failed expression of presumed ACL from T66 in prokaryotes even more intriguing. Could it be that all of the gene candidates for ACL were wrongly suspected of ACL activity, and thus it would make no sense to compare the expression of these genes to the expression of ACL? Or does the expression fail due to some feature of thraustochytrid genes? Could it be that the expression failed due to coincidental problems with all five of the chosen peptides?

The answer to the question concerning expression of thraustochytrid genes can be found in the literature. Genes from various *Aurantiochytrium* species have been reported to be heterologously expressed, for instance a lysophospholipid acyltransferase from *Aurantiochytrium limacinum* F26-b in *Saccharomyces cerevisiae* (Abe et al., 2014) and a squalene synthase from *Aurantiochytrium* sp. KRS101 in *E. coli* (Hong et al., 2013). Thus, *Aurantiochytium* genes can in fact be heterologously expressed in both prokaryotes and

eukaryotes. This leaves the question concerning specific features related to the expression of the five ORFs of interest to future studies.

If indeed the proteins had been successfully expressed, a detail discovered in the preliminary research concerning the ACL candidates might have become significant. The gene sequences for the peptides attempted to be expressed matched with either β- or α-subunits of proteins. Could it be that a combination of the peptides encoded by the genes of interest was needed to make a functional enzyme? If so, which combination? Or could it be that one of the genes indeed encoded one of the ACL subunits, but that another required subunit was not among the sequences investigated and expressed? If a combination of the peptides chosen and expressed were needed, the purified peptides could have been combined *in vitro* before assays were run. Alternatively, operons expressing several subunits could have been made. However, the risk of there being a need of chaperones or a specific environment to correctly combine the subunits would have been present.

Even though the expression of these peptides did not give the desired results, the expression vectors made did not have any apparent flaws. In particular, pMEJ1 and pMEJ2 could prove to be useful expression vectors for the expression of many peptides in *E. coli* due to the large flexibility in the insertion of the gene to be expressed. It would also be rather straight forward to combine either pMEJ1 or pMEJ2 with pMEJ13 to get a new expression vector with both the flexibility with regards to restriction of insert and the signal sequence *pelB*.

By the time all of these different adjustments were attempted to achieve protein expression, the time of the deadline for this thesis was approaching quickly. It was thus decided to move on to a direct measurement of enzyme activity in protein extracts from *Aurantiochytrium* T66 itself. However, this experiment would only partly be able to answer the initial question; Does T66 produce the enzymes G3PDH and ACL, and if so, by which genes are they encoded? A measurement of enzyme activity in the lysate of T66 should theoretically be able to detect enzyme activity. However, the genes encoding the enzymes would still not be known. Another challenge in the enzyme assay in the T66 fermentation experiment, was the indirect assays designed to measure enzyme activity. Both assays were based on measurements of NADH or NADPH consumption. In a mixture of metabolic enzymes, many enzymes would likely contribute to the consumption of NAD(P)H. Some might even form NAD(P)H from the $NA(P)D^+$ generated by the NAD(P)H-consuming reactions. An attempt to segregate all of these other NAD(P)H-based reactions from the specific ones in question was made by

running assays both with and without the specific enzyme substrates. However, once again, one could not be certain that no other reactions were consuming the substrates. Thus, these enzyme assays would have to be considered as rough estimates needing to be properly verified by heterologously expressed enzymes or other more specific methods.

Regarding the assays described for G3PDH, it should be noted that these assays would not be able to measure the activity of the FAD-dependent G3PDH enzyme (mG3PDH) involved in the glycerol 3-phosphate shuttle (Section 1.4.2). This enzyme is a membrane bound enzyme. The signal peptide search indicated a signal peptide in gene 6468, which was one of the suspected G3PDH enzymes. This could indicate that this was a G3PDH enzyme destined for the mitochondrial membrane. If the genes selected for G3PDH activity were FAD dependent, different assays would have to be applied to detect these.

A rather surprising problem occurred during the preparation of the enzyme samples from the fermentation; the cells proved quite difficult to lyse. The initial sonication test was based on the sonication conditions described by Chang et al. (2013) in their fermentation of *Schizochytrium* S21, which was the control in this experiment. In their case, 900 s of sonication with a strength of 455 W was used to lyse the cells. As showed by the sonication test and the sonication of the samples, the cells of both species tested easily lysed when grown in YPD media, but not when they were grown in the lipid accumulation media. The media described by Chang et al. (2013) was similar to the lipid accumulation media used in this experiment. The main difference was that glycerol was the main carbon source in the media of Chang et al. (2013), opposed to glucose in the lipid accumulation media. However, the YPD media also contained glucose as the main carbon source. The sonication challenges were in fact discovered by the lucky coincidence that cultures grown in the two different media were kept in the incubator at the time of the sonication test. A large increase in the protein concentration was not observed when the lipid accumulation cultures were sonicated more than ~300 s. However, when cultures grown in the YPD media where sonicated, the protein yield was about five times higher than the lipid accumulation cultures with comparable OD. This indicated in fact that more protein should be present, and that the cultures of the lipid accumulation media were not successfully lysed. The problem could easily have been overlooked if the YPD cultures had not been included in the test, since there was no significant increase in protein concentration when sonication time was increased beyond ~500 s. The difference in susceptibility to sonication might have been explained if the samples discussed were from the late stages of fermentation, where the cultures in the lipid

accumulation media should have a large percentage of lipids. However, when samples with OD around 6-8 were compared, the cells should be in the exponential phase (see Figure 3.32), and lipid production was expected to be low in both media.

Because of the challenges concerning lysis of the cells, it took a long time to prepare the enzyme samples produced during the fermentation trial. They were also subjected to a large degree of sonication. This might harm the enzymes even if the membranes of the thraustochytrids were not destroyed. It could render the enzymes inactive, thus being a possible explanation to the lack of enzyme activity detection in the assays performed on the samples. The samples from the different fermentation flasks were also unequally sonicated in the attempt to lyse at least some of the samples. This meant that the comparison of the flasks might not be completely correct, due to the different treatment of the samples.

Because of the challenges with sonicating the cells, other cell lysing methods should be considered for future experiments. In this thesis, two other methods that might work are described. The chemical lysis with the commercially available B-PER™ Protein Extraction Reagent was used to lyse the prokaryotic host cells. Versions of this reagent adapted for yeast cells and plant cells, Y-PER™ and P-PER™, respectively, are available from the same producer (Thermo Scientific). These are described by the producer to yield active protein lysates, but this should probably be verified for each particular enzyme. Another lysis method described in this thesis, is the mechanical breakage of the cells using zirconium oxide beads and a Precellys 24 shaker which was used during the lipid extraction procedure. This method would presumably be able to overcome though membranes, but the yield of active enzymes might be low due to the possible mechanical destruction of the protein.

The enzyme assays performed did not show any clear activity, except from a small hint in the measurement of ACL activity at 66h of fermentation. The negative results were not necessarily due to the actual absence of the enzymes. As discussed above, the challenges faced while preparing the samples might have rendered the enzymes inactive. Also, the large variations and noise in the datasets achieved might have been decreased by optimizing the procedure further. By allowing the reactions to proceed for a longer time, the trend detected in the ACL assay might have been clear enough to present as a positive detection of enzyme activity.

The fermentation experiment did however yield some useful results. A growth curve of *Aurantiochytrium* T66 was made. This growth curve was very similar to the one produced by

the control organism *Schizochytrium* S21. This showed that S21 could be a useful control organism for future research, because the coordination of fermentations could be performed with relative ease. The lipid measurements also showed a high yield of lipids in the lipid accumulation media, thus confirming that the media was suited for lipid production.

No enzyme activity for neither ACL nor G3PDH was detected during the experiments described in this thesis, even though the enzyme assay protocols were verified. The enzyme activity of the heterologosly expressed enzymes could not be detected due to challenges concerning the expression. The enzyme activity in the fermentation experiment quite likely remained undetected due to sample preparation challenges. Thus, neither the presence nor the absence of either of the enzymes in question could be verified. However, the presence of G3PDH was assumed relatively certain, and the important question concerning this enzyme was the cofactor requirements. Regarding ACL, the presence was a more intriguing question. If in fact the enzyme was absent, the acetyl-CoA needed for lipid production would have to be provided in some other fashion. This knowledge could be important for the improved understanding of the *Aurantiochytrium* T66 metabolism, and thus enhance the ability to manipulate its lipid production.

5 Conclusion and Further Work

The *Aurantiochytrium* T66 genes 6817, 10760, 9329, 8081, and 6468 were cloned into RK2-based vectors. Expression was attempted in *E. coli* and *P. putida* both with and without the signal sequence *pelB*. No recombinant protein was detected for either of these experiments. A lowered growth temperature was also tested for *E. coli,* but this did not improve expression. ACL and G3PDH activity were also measured *in vivo* at several timepoints of *Aurantiochytrium* T66 fermentations. No activity was detected for either of the enzymes, but this was suspected due to sample preparation challenges and not due to the absence of the enzymes.

In order to express the genes of interest in this thesis, the use of a eukaryote host could be attempted. Expression vectors encoding several of the genes could also be constructed, as these genes likely encode subunits of the enzymes. The need to develop good protocols for enzyme extraction from *Aurantiochytrium* T66 cultures have also been demonstrated in this thesis, as well as the need to test these protocols for several different growth media.

References

Aasen, I. M., Ertesvåg, H., Heggeset, T. M., Liu, B., Brautaset, T., Vadstein, O. and Ellingsen, T. E. (2016) 'Thraustochytrids as production organisms for docosahexaenoic acid (DHA), squalene, and carotenoids', *Appl Microbiol Biotechnol,* 100(10), pp. 4309-21.

Abe, E., Ikeda, K., Nutahara, E., Hayashi, M., Yamashita, A., Taguchi, R., Doi, K., Honda, D., Okino, N., Ito, M. and Schunck, W.-H. (2014) 'Novel Lysophospholipid Acyltransferase PLAT1 of Aurantiochytrium limacinum F26-b Responsible for Generation of Palmitate-Docosahexaenoate-Phosphatidylcholine and Phosphatidylethanolamine', *PLoS ONE,* 9(8), pp. e102377.

Al-Hendy, A., Toivanen, P. and Skurnik, M. (1991) 'The effect of growth temperature on the biosynthesis of Yersinia enterocolitica 0 : 3 lipopolysaccharide: temperature regulates the transcription of the rfb but not of the rfa region', *Microbial Pathogenesis,* 10(1), pp. 81-86.

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K. and Walter, P. (2015) 'Molecular Biology of the Cell'. Sixth ed: Garland Science, pp. 439-528.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic acids research,* 25(17), pp. 3389-3402.

Athey, J., Alexaki, A., Osipova, E., Rostovtsev, A., Santana-Quintero, L. V., Katneni, U., Simonyan, V. and Kimchi-Sarfaty, C. (2017) 'A new and updated resource for codon usage tables', *BMC bioinformatics,* 18(1), pp. 391-391.

Auclair, S. M., Bhanu, M. K. and Kendall, D. A. (2012) 'Signal peptidase I: Cleaving the way to mature proteins', *Protein Science,* 21(1), pp. 13-25.

Aune, T. E. V. and Aachmann, F. L. (2010) 'Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed', *Applied Microbiology and Biotechnology,* 85(5), pp. 1301-1313.

Belda, E., van Heck, R. G. A., José Lopez-Sanchez, M., Cruveiller, S., Barbe, V., Fraser, C., Klenk, H.-P., Petersen, J., Morgat, A., Nikel, P. I., Vallenet, D., Rouy, Z., Sekowska, A., Martins dos Santos, V. A. P., de Lorenzo, V., Danchin, A. and Médigue, C. (2016) 'The revisited genome of Pseudomonas putida KT2440 enlightens its value as a robust metabolic chassis', *Environmental Microbiology,* 18(10), pp. 3403-3424.

Benchling Molbio Overview (2019): Benchling Inc. Available at: https://main.bnchcdn.com/static/docs/overviews/benchling-molbio-overview.pdf (Accessed: 08.04. 2019).

Berg, J. M., Tymoczko, J. L. and Stryer, L. (2002a) 'Section 18.5 Many Shuttles Allow Movement Across the Mitochondrial Membranes', *Biochemistry.* 5th ed. New York: W. H. Freeman and Company.

Berg, J. M., Tymoczko, J. L. and Stryer, L. (2002b) 'Section 22.4 Fatty Acids Are Synthesized and Degraded by Different Pathways', *Biochemistry.* 5th ed. New York: W. H. Freeman and Company.

Blatny, J. M., Brautaset, T., Winther-Larsen, H. C., Karunakaran, P. and Valla, S. (1997) 'Improved Broad-Host-Range RK2 Vectors Useful for High and Low Regulated Gene Expression Levels in Gram-Negative Bacteria', *Plasmid,* 38(1), pp. 35-51.

Blomberg, A. and Adler, L. (1989) 'Roles of glycerol and glycerol-3-phosphate dehydrogenase (NAD+) in acquired osmotolerance of Saccharomyces cerevisiae', *Journal of Bacteriology,* 171(2), pp. 1087-1092.

Bornhorst, J. A. and Falke, J. J. (2000) 'Purification of proteins using polyhistidine affinity tags', *Methods in enzymology,* 326, pp. 245-254.

Bradford, M. M. (1976) 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Anal Biochem,* 72, pp. 248-54.

Bransden, M. P., Carter, C. G. and Nichols, P. D. (2003) 'Replacement of fish oil with sunflower oil in feeds for Atlantic salmon (Salmo salar L.): effect on growth performance, tissue fatty acid composition and disease resistance', *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology,* 135(4), pp. 611-625.

Chang, G., Luo, Z., Gu, S., Wu, Q., Chang, M. and Wang, X. (2013) 'Fatty acid shifts and metabolic activity changes of Schizochytrium sp. S31 cultured on glycerol', *Bioresource Technology,* 142, pp. 255 - 260.

Chávez-Cabrera, C., Marsch, R., Bartolo-Aguilar, Y., Flores-Bustamante, Z. R., Hidalgo-Lara, M. E., Martínez-Cárdenas, A., Cancino-Díaz, J. C., Sánchez, S. and Flores-Cotera, L. B. (2015) 'Molecular cloning and characterization of the ATP citrate lyase from carotenogenic yeast Phaffia rhodozyma', *FEMS Yeast Research,* 15(6), pp. fov054-fov054.

Chen, H., He, X., Geng, H. and Liu, H. (2014) 'Physiological characterization of ATP-citrate lyase in Aspergillus niger', *Journal of Industrial Microbiology & Biotechnology,* 41(4), pp. 721-731.

Clark, D. P. and Pazdernik, N. J. (2013a) 'Cloning Genes for Analysis', *Molecular Biology.* 2 ed: Academic Press, pp. 194-226.

Clark, D. P. and Pazdernik, N. J. (2013b) 'Polymerase Chain Reaction', *Molecular Biology.* 2 ed: Academic Press, pp. 163-192.

Clark, D. P. and Pazdernik, N. J. (2013c) 'Proteomics: The Global Analysis of Proteins', *Molecular Biology.* 2 ed: Academic Press, pp. 459-491.

Colussi, T., Parsonage, D., Boles, W., Matsuoka, T., Mallett, T. C., Karplus, P. A. and Claiborne, A. (2008) 'Structure of alpha-glycerophosphate oxidase from Streptococcus sp.: a template for the mitochondrial alpha-glycerophosphate dehydrogenase', *Biochemistry,* 47(3), pp. 965-77.

Durland, R. H., Toukdarian, A., Fang, F. and Helinski, D. R. (1990) 'Mutations in the trfA replication gene of the broad-host-range plasmid RK2 result in elevated plasmid copy numbers', *Journal of Bacteriology,* 172(7), pp. 3859.

Fossier Marchan, L., Lee Chang, K. J., Nichols, P. D., Mitchell, W. J., Polglase, J. L. and Gutierrez, T. (2018) 'Taxonomy, ecology and biotechnological applications of thraustochytrids: A review', *Biotechnology Advances,* 36(1), pp. 26-46.

Gasco, L. (2018) *Feeds for the Aquaculture Sector : Current Situation and Alternative Sources* Cham: Springer International Publishing. *Chemistry of Foods*.

Gawin, A., Valla, S. and Brautaset, T. (2017) 'The XylS/*Pm* regulator/promoter system and its use in fundamental studies of bacterial gene expression, recombinant protein production and metabolic engineering', *Microbial biotechnology,* 10(4), pp. 702-718.

González, J. M., Marti-Arbona, R., Chen, J. C. and Unkefer, C. J. (2017) 'Structure of Methylobacterium extorquens malyl-CoA lyase: CoA-substrate binding correlates with domain shift', *Acta Crystallogr F Struct Biol Commun,* 73(Pt 2), pp. 79-85.

Gupta, A., Barrow, C. J. and Puri, M. (2012) 'Omega-3 biotechnology: Thraustochytrids as a novel source of omega-3 oils', *Biotechnology Advances,* 30(6), pp. 1733-1745.

Hannig, G. and Makrides, S. C. (1998) 'Strategies for optimizing heterologous protein expression in Escherichia coli', *Trends in Biotechnology,* 16(2), pp. 54-60.

Hauvermale, A., Kuner, J., Rosenzweig, B., Guerra, D., Diltz, S. and Metz, J. G. (2006) 'Fatty acid production in Schizochytrium sp.: Involvement of a polyunsaturated fatty acid synthase and a type I fatty acid synthase.', *Lipids,* 41(8), pp. 739.

Hermanson, G. T. (2013) 'Chapter 2 - Functional Targets for Bioconjugation', in Hermanson, G.T. (ed.) *Bioconjugate Techniques (Third Edition)*. Boston: Academic Press, pp. 127-228.

Hong, W. K., Heo, S. Y., Park, H. M., Kim, C. H., Sohn, J. H., Kondo, A. and Seo, J. W. (2013) 'Characterization of a Squalene Synthase from the Thraustochytrid Microalga Aurantiochytrium sp. KRS101', *Journal of Microbiology and Biotechnology,* (6), pp. 759-765.

Huang, J. and Fraser, M. E. (2016) 'Structural basis for the binding of succinate to succinyl-CoA synthetase', *Acta Crystallogr D Struct Biol,* 72(Pt 8), pp. 912-21.

Jakobsen, A. N., Aasen, I. M., Josefsen, K. D. and Strøm, A. R. (2008) 'Accumulation of docosahexaenoic acid-rich lipid in thraustochytrid Aurantiochytrium sp. strain T66: effects of N and P starvation and O2 limitation', *Applied Microbiology and Biotechnology,* 80(2), pp. 297.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T. L. (2008) 'NCBI BLAST: a better web interface', *Nucleic acids research,* 36(Web Server issue), pp. W5-W9.

Joyce, M. A., Fraser, M. E., James, M. N. G., Bridger, W. A. and Wolodko, W. T. (2000) 'ADP-Binding Site of Escherichia coli Succinyl-CoA Synthetase Revealed by X-ray Crystallography', *Biochemistry,* 39(1), pp. 17-25.

Kaur, J., Kumar, A. and Kaur, J. (2018) 'Strategies for optimization of heterologous protein expression in E. coli: Roadblocks and reinforcements', *International Journal of Biological Macromolecules,* 106, pp. 803-822.

King, L. A. (1992) *The Baculovirus Expression System : A laboratory guide* Dordrecht: Springer Netherlands

Klus, P., Bolognesi, B., Agostini, F., Marchese, D., Zanzoni, A. and Tartaglia, G. G. (2014) 'The cleverSuite approach for protein characterization: predictions of structural properties, solubility, chaperone requirements and RNA-binding abilities', *Bioinformatics (Oxford, England),* 30(11), pp. 1601-1608.

Kolatka, K., Kubik, S., Rajewska, M. and Konieczny, I. (2010) 'Replication and partitioning of the broad-host-range plasmid RK2', *Plasmid,* 64(3), pp. 119-134.

Kostylev, M., Otwell, A. E., Richardson, R. E. and Suzuki, Y. (2015) 'Cloning Should Be Simple: Escherichia coli DH5α-Mediated Assembly of Multiple DNA Fragments with Short End Homologies', *PloS one,* 10(9), pp. e0137466-e0137466.

Kujau, M. J., Hoischen, C., Riesenberg, D. and Gumpert, J. (1998) 'Expression and secretion of functional miniantibodies McPC603scFvDhlx in cell-wall-less L-form strains of Proteus mirabilis and Escherichia coli: A comparison of the synthesis capacities of L-form strains with an E. coli producer strain', *Applied Microbiology and Biotechnology,* 49(1), pp. 51-58.

Liu, B., Ertesvåg, H., Aasen, I. M., Vadstein, O., Brautaset, T. and Heggeset, T. M. (2016) 'Draft genome sequence of the docosahexaenoic acid producing thraustochytrid Aurantiochytrium sp. T66', *Genome Data,* 8, pp. 115-6.

Nelson, D. L. and Cox, M. M. (2013a) 'The Citric Acid Cycle', *Lehninger Principles of Biochemistry.* 6th ed. New York: W. H. Freeman and Company, pp. 633-665.

Nelson, D. L. and Cox, M. M. (2013b) 'Glycolysis, Gluconeogenesis, and the Pentose Phosphate Pathway', *Lehninger Principles of Biochemistry.* 6th ed. New York: W. H. Freeman and Company, pp. 543-585.

Nelson, D. L. and Cox, M. M. (2013c) 'Lipid Biosynthesis', *Lehninger Principles of Biochemistry.* 6th ed. New York: W. H. Freeman and Company, pp. 833-880.

Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. and Yura, T. (1998) 'Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and

GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in Escherichia coli', *Applied and environmental microbiology,* 64(5), pp. 1694-1699.
Nordahl Petersen, T., Brunak, S., Heijne, G. and Nielsen, H. (2011) 'SIGNALP 4.0: discriminating signal peptides from transmembrane regions', *Nature Methods,* 8, pp. 785-6.
NorwegianSeafoodCouncil (2019) *Sjømateksport for 99 milliarder i 2018*: Norwegian Seafood Council. Available at: https://seafood.no/aktuelt/nyheter/sjomateksport-for-99-milliarder-i-2018-/ (Accessed: 23.04 2019).
Papaneophytou, C. P. and Kontopidis, G. (2014) 'Statistical approaches to maximize recombinant protein expression in Escherichia coli: A general review', *Protein Expression and Purification,* 94, pp. 22-32.
Pines, O. and Inouye, M. (1999) 'Expression and secretion of proteins in E. coli', *Molecular Biotechnology,* 12(1), pp. 25.
Rapoport, S. I. and Taha, A. (2014) 'Chapter 22 - Imaging Brain DHA Metabolism in Vivo, in Animals, and Humans', in Watson, R.R. & De Meester, F. (eds.) *Omega-3 Fatty Acids in Brain and Neurological Health*. Boston: Academic Press, pp. 265-275.
Ratledge, C. (2004) 'Fatty acid biosynthesis in microorganisms being used for Single Cell Oil production', *Biochimie,* 86(11), pp. 807 - 815.
Ratledge, C. and Wynn, J. P. (2002) 'The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms', *Advances in Applied Microbiology*. New York, pp. 1-51.
Ren, Q., Henes, B., Fairhead, M. and Thöny-Meyer, L. (2013) 'High level production of tyrosinase in recombinant Escherichia coli', *BMC biotechnology,* 13, pp. 18-18.
Saibil, H. (2013) 'Chaperone machines for protein folding, unfolding and disaggregation', *Nature reviews. Molecular cell biology,* 14(10), pp. 630-642.
Salem, N. and Eggersdorfer, M. (2015) 'Is the world supply of omega-3 fatty acids adequate for optimal human nutrition?', *Current Opinion in Clinical Nutrition and Metabolic Care,* 18(2), pp. 147-154.
Sampaio, P. N., Sousa, L., Calado, C. R. C., Pais, M. S. and Fonseca, L. P. (2011) 'Use of chemometrics in the selection of a Saccharomyces cerevisiae expression system for recombinant cyprosin B production', *Biotechnology Letters,* 33(11), pp. 2111-2119.
Sletta, H., Tøndervik, A., Hakvåg, S., Aune, T. E. V., Nedal, A., Aune, R., Evensen, G., Valla, S., Ellingsen, T. E. and Brautaset, T. (2007) 'The Presence of N-Terminal Secretion Signal Sequences Leads to Strong Stimulation of the Total Expression Levels of Three Tested Medically Important Proteins during High-Cell-Density Cultivations of Escherichia coli', *Applied and Environmental Microbiology,* 73(3), pp. 906.
Srere, P. A. (1959) 'The Citrate Cleavage Enzyme: I. DISTRIBUTION AND PURIFICATION', *Journal of Biological Chemistry,* 234(10), pp. 2544-2547.
Srere, P. A. (1972) 'The Citrate Enzymes: Their Structures, Mechanisms, and Biological Functions', *Current Topics in Cellular Regulation*: Academic Press, pp. 229-283.
Van Durme, J., Maurer-Stroh, S., Gallardo, R., Wilkinson, H., Rousseau, F. and Schymkowitz, J. (2009) 'Accurate Prediction of DnaK-Peptide Binding via Homology Modelling and Experimental Data', *PLOS Computational Biology,* 5(8), pp. e1000475.
Vilnes, M. (2018) *Genetisk modifisering av lipidmetabolismen i Rhodococcus opacus PD630.* Master, NTNU, Trondheim.
Waterhouse, A., Rempfer, C., Heer, F. T., Studer, G., Tauriello, G., Bordoli, L., Bertoni, M., Gumienny, R., Lepore, R., Bienert, S., de Beer, T. A P. and Schwede, T. (2018) 'SWISS-MODEL: homology modelling of protein structures and complexes', *Nucleic Acids Research,* 46(W1), pp. W296-W303.

Wawrzycka, A., Gross, M., Wasaznik, A. and Konieczny, I. (2015) 'Plasmid replication initiator interactions with origin 13-mers and polymerase subunits contribute to strand-specific replisome assembly', *Proceedings of the National Academy of Sciences,* 112(31), pp. E4188.

Yokoyama, R. and Honda, D. (2007) 'Taxonomic rearrangement of the genus Schizochytrium sensu lato based on morphology, chemotaxonomic characteristics, and 18S rRNA gene phylogeny (Thraustochytriaceae, Labyrinthulomycetes): emendation for Schizochytrium and erection of Aurantiochytrium and Oblongichytrium gen. nov.(Author abstract)(Report)', *Mycoscience,* 48(4), pp. 199.

Zero Blunt® TOPO® PCR Cloning Kit User Guide. (2014). Sciences, I.b.L., 24.02.2014 [Protocol].

Appendix A – Ladders

DNA Ladders

The ladders used for the agarose gels were made by digesting λ DNA (New England Biolabs Inc.) with the restriction enzymes PstI and HindIII. The result of a virtual digest of the λ DNA with either PstI or HindIII is shown in Figure A.1. This was used as a basis for evaluating the size of the fragments in the samples on the agarose gels made during the experiments of this thesis.

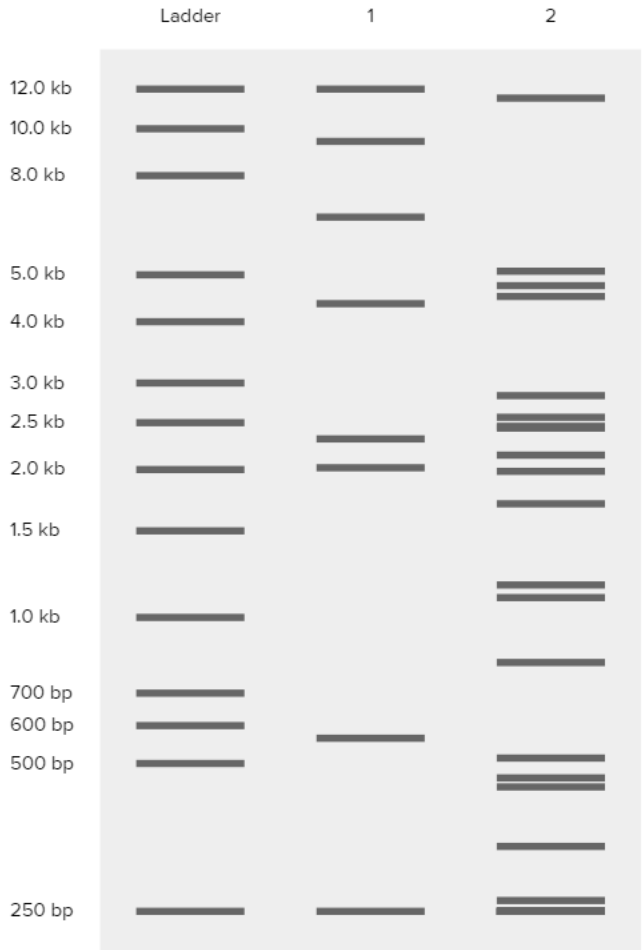


Figure A.1: ***DNA ladders****. A virtual digest was performed in Benchling. The DNA digested was λ DNA (GenBank: J02459.1). The ladder (left) is a virtual ladder indicating fragment sizes. 1 (middle) is a digest with HindIII and 2 (right) is a digest with PstI.*

Protein Ladders

For the protein gels, a Prestained Protein Ladder #7710, Broad Range (10-230 kDa), from New England Biolabs Inc. was used. The bands and corresponding molecular weights made by this ladder is indicated in Figure A.2.

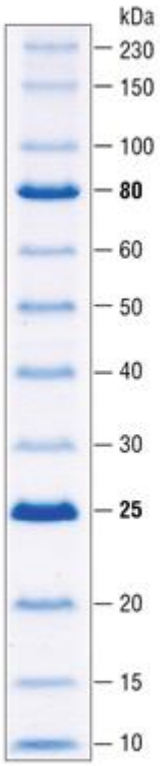


Figure A.2: ***Protein ladder****. The bands and corresponding molecular weights resulting from a protein gel with the Prestained Protein Ladder #7710, Broad Range (10-230 kDa) from New England Biolabs Inc.*

Appendix B – Bradford Standard Curve

The Bradford assay was performed using a Coomassie (Bradford) Protein Assay Kit (Number 23200) from Thermo Scientific and the protocol supplied with the product.

A standard curve (Figure B.1) for the Bradford assay was made by making a series of solutions with different concentrations of bovine serum albumin (BSA) in 0.1 M Tris-HCl (pH 7.4). 5 µL BSA sample was added to each well in a blank 96-well microplate. 250 µL Coomassie Reagent was added to each well, and the sample and the reagent was mixed by pipetting. The sample was incubated for ~10 min before the absorption at 595 nm was measured using a SpectraMax Plus 384 spectrophotometer from Molecular Devices. Three parallels were performed for each sample, including the blank sample (only buffer). The absorbance was calculated as the average absorbance of the three parallels minus the average absorbance of the blank samples. These values were plotted against the concentration to form the standard curve. Thermo Scientific recommended using a point-to-point plotting rather than using linear regression for assays performed in microplates in their protocol, thus this was chosen for the standard curve.

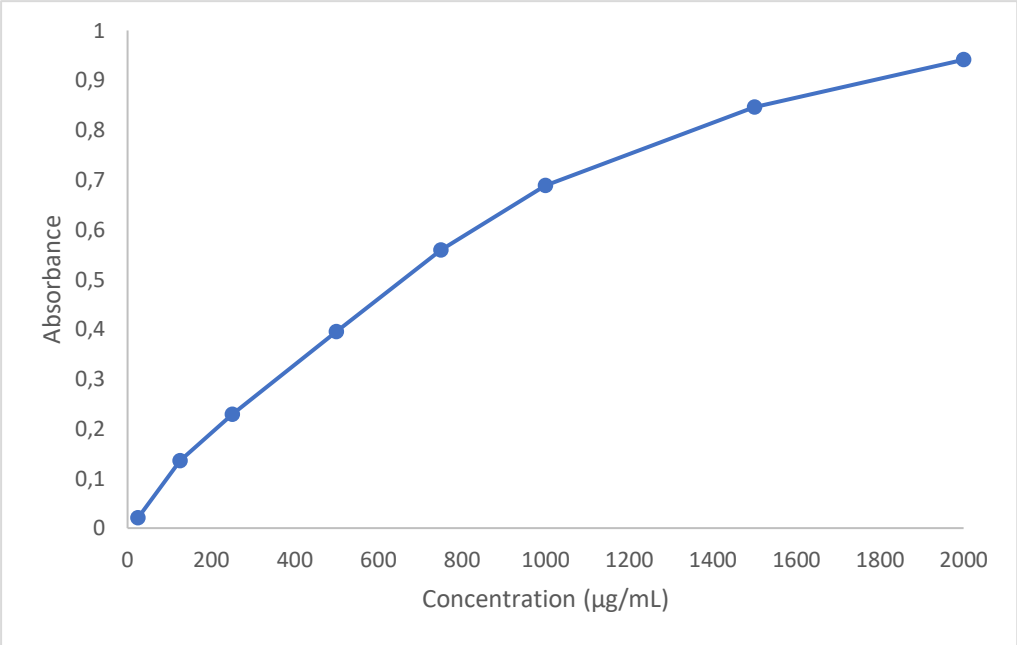


Figure B.1: ***Bradford assay standard curve****. A standard curve was made by measuring the absorbance of a Coomassie Reagent with different concentrations of BSA at 595 nm, subtracting these values from the absorbance of samples with only buffer, and finally plotting these values against the BSA concentrations.*

Appendix C – Enzyme Assay Controls

The assays were verified by testing them on commercially available enzymes.

ACL Assay

To verify the ACL assay, citrate lyase from *Klebsiella pneumoniae* from Sigma-Aldrich (product number C0897, EC: 4.1.3.6) was used. According to Srere (1972), the citrate lyase catalyzes a reaction quite similar to the reaction catalyzed by ACL. The main difference between the two, is the independence of citrate lyase of ATP. Due to this difference, ATP could not be used as the initiating agent in the assay. The enzyme itself was given this role instead. The enzyme assay was otherwise performed as described in Section 2.8.2. Instead of the protein sample, 50 µL 0.1 M Tris-HCl (pH 7.4) was used. The concentration of citrate lyase in the reaction mixture was 0.2 U/mL, where 1 U was defined as the amount of enzyme needed to convert 1.0 µmole of citrate to oxaloacetate per min at pH 7.6 at 25 °C.

One parallel of the control assay is shown in Figure C.1. The absorption indicated is the absorption after the absorption for the flattened part of the curve (calculated as average between 200 and 300 s) was subtracted. A clear decrease in absorption at 340 nm is observed, indicating a diminishing amount of NADH. This indicates that the assay set up was correct.

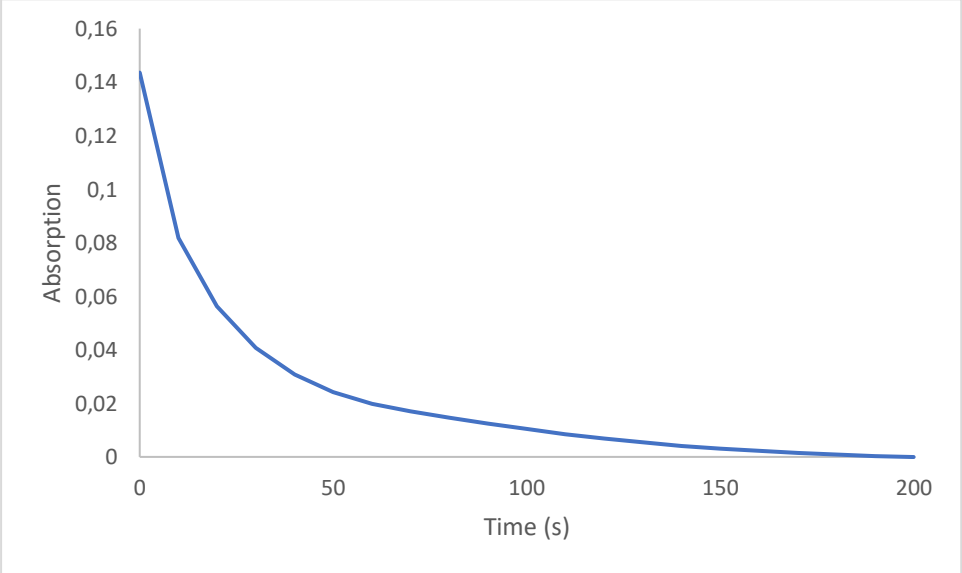


Figure C.1: ***Assay result from control assay with citrate lyase****. The assay developed for ACL was performed on commercially produced citrate lyase. The figure shows the absorption minus the absorption of the reaction after the flattening of the curve as a function of time in s.*

G3PDH Assay

To verify the G3PDH assay, G3PDH from rabbit muscle from Roche (EC: 1.1.1.8) was used. The protocol was performed as described in Section 2.8.1. Instead of the protein sample, 50 µL 0.1 M Tris-HCl (pH 7.4) was used. The assay was performed for both NADH and NADPH, even though the commercial G3PDH was described as a NADH-using G3PDH. The concentration of G3PDH in the NADH assay reaction mixture was 0.032 U/mL and in the NADPH reaction mixture it was 0.630 U/mL. 1 U was defined as the amount of enzyme needed to catalyze 1.0 µmole of DHAP per min at pH 8.0-8.5 at 25 °C.

For both assays, the raw absorption of two parallels were averaged and shown as functions of time. One parallel without enzyme was run for each assay. Figure C.2 shows the results for the G3PDH assay with NADPH, while Figure C.3 shows the corresponding figure with NADH. Both the figures show a decrease in absorption for the parallels where enzyme was added, but none in the parallels without enzyme. This indicated that the decrease in absorption is due to the consumption of NADH/NADPH by the enzymes. The NADH curve is significantly steeper than the NADPH curve, even though the enzyme concentration is 20 times higher in the NADPH parallels. This indicated that even though the enzyme could use both NADPH and NADH, the preferred cofactor was NADH.

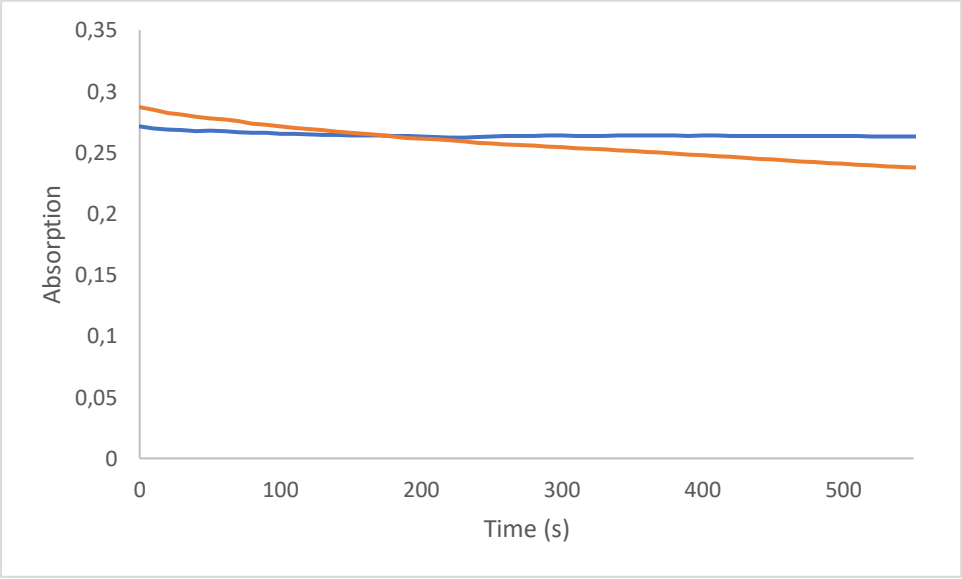


*Figure C.2: **Assay result from control assay for G3PDH with NADPH**. The assay for G3PDH was performed on a commercially produced G3PDH. The absorption is shown as a function of time in seconds. The orange line is the average of two parallels performed with the enzyme, while the blue line is a parallel performed without the enzyme.*

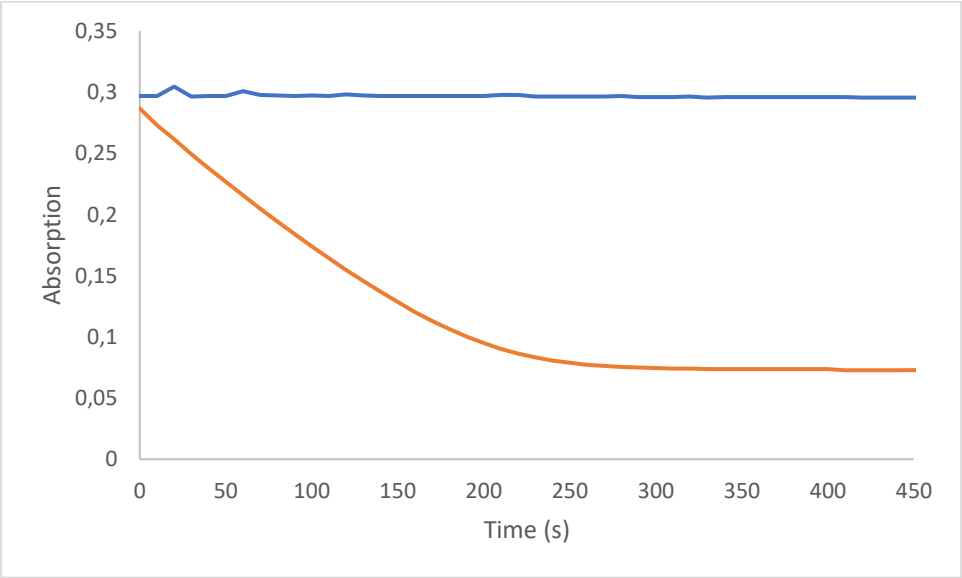


Figure C.3: ***Assay result from control assay for G3PDH with NADH****. The assay for G3PDH was performed on a commercially produced G3PDH. The absorption is shown as a function of time in seconds. The orange line is the average of two parallels performed with the enzyme, while the blue line is a parallel performed without the enzyme.*

Appendix D – Sequencing of Genes of Interest

The PCR products of the ORFs of interest were verified by sequencing the TOPO® vectors they were cloned into (see Table 2.3). The sequencing was performed by the commercial agent Eurofins Genomics, GATC service (LIGHTRUN). All ORFs were sequenced using the M13F (forward) and M13R (reverse) primers, and for gene 8081 and 6468 custom designed primers (seq) to sequence the middle of the genes were used (Table 2.1).

The sequences were aligned and evaluated in Benchling, but they are here displayed using Clone Manager 9. The dots (.) signify that the base of the sequenced fragment is the same as in the template, which is the theoretical TOPO® vectors with the gene inserts. The hyphen (-) signifies that no base was determined for this point in the fragment. If another base than the one in the template was detected at a specific point, the base detected was shown. Only the parts of the alignments where the sequencing where considered trustworthy are shown. The sequences of the genes of interest are highlighted yellow, while deviances are marked red.

Sequence Alignment of Gene 6817

```
pMEJ3T          235 agctatttaggtgacactatagaatactcaagctatgcatcaagcttggt
M13R             23 ..................................................

pMEJ3T          285 accgagctcggatccactagtaacggccgccagtgtgctggaattcgccc
M13R             73 ..................................................
M13F           1114 ----------------------------..........................

pMEJ3T          335 ttgcggccgcaagcatggccttgccaagagccttcatggcctcgagcatg
M13R            123 ..................................................
M13F           1093 ..................................................

pMEJ3T          385 gtgacgccgagcttggcgggggactcggtgacgtggacaccggcatcgcg
M13R            173 ..................................................
M13F           1043 ..................................................

pMEJ3T          435 gagggcctccatcttggcctttgcggtaccttgccgccggcaatgatgg
M13R            223 ..................................................
M13F            993 ..................................................

pMEJ3T          485 cgccggcgtggcccatgcggcggccgggcggggcggtgaggccggcgatg
M13R            273 ..................................................
M13F            943 ..................................................

pMEJ3T          535 aagctgacgacaggcttgtccgggtttccgttctgcttgagccattctgc
M13R            323 ..................................................
M13F            893 ..................................................

pMEJ3T          585 ggcctcctcctcggcggtgccgccaatctcaccgatcatgatgatgccct
M13R            373 ..................................................
M13F            843 ..................................................

pMEJ3T          635 cggtctcggggtcgttggtgaagcgctcgaggcagtcgatgaagttggtg
M13R            423 ..................................................
M13F            793 ..................................................
```

```
pMEJ3T          685 ccgttgaagggatcgccgccaatgccgacgaccgtggactggccgagacc
M13R            473 ..................................................
M13F            743 ..................................................

pMEJ3T          735 ggtctcggtggtctgggccacggcctcgtacgtgagggtgccggagcggg
M13R            523 ..................................................
M13F            693 ..................................................

pMEJ3T          785 agacgacgccgatcttgcccggggtgtggatgtagcccggcatgatgccg
M13R            573 ..................................................
M13F            643 ..................................................

pMEJ3T          835 atcttgcactcgttgggcttgatgatgccggggcagttggggccgatgag
M13R            623 ..................................................
M13F            593 ..................................................

pMEJ3T          885 gcgggtggtggtctggcgctggaggtgccacttgaccttgagcatgtcct
M13R            673 ..................................................
M13F            543 ..................................................

pMEJ3T          935 gctgcgggatgccctcggtgatgcacacgacgaggggcacctcggcctcg
M13R            723 ..................................................
M13F            493 ..................................................

pMEJ3T          985 acggcctcaagaatggccttgccggccacgggggggcggcacgaagatgac
M13R            773 ..................................................
M13F            443 ..................................................

pMEJ3T         1035 cgaggcgtcgacaccgacggcgtcctgggcctccttcacggtgttgaaca
M13R            823 ..................................................
M13F            393 ..................................................

pMEJ3T         1085 cggggagtccgaggtgctcctggccgcccttgccggggctcgtgccgccg
M13R            873 ..................................................
M13F            343 ..................................................

pMEJ3T         1135 atcatcttggttccgtacgcgatggcctgctcgctatggaaggtgccctg
M13R            923 ..................................................
M13F            293 ..................................................

pMEJ3T         1185 cttgccggtaaagccctggcagaggaccttggtgttcttgtccacccaca
M13R            973 ..............c--------------------------------
M13F            243 ..................................................

pMEJ3T         1235 cgttggccgtggccgacatggcacgggtcgccgtcagcgtcgccgcctgg
M13F            193 ..................................................

pMEJ3T         1285 cgggtcttggtcacggtacgcgcgagcatatgaagggcgaattctgcaga
M13F            143 ..................................................

pMEJ3T         1335 tatccatcacactggcggccgctcgagcatgcatctagagggcccaattc
M13F             93 ..................................................

pMEJ3T         1385 gccctatagtgagtcgtattacaat
M13F             43 ........................
```

Sequence Alignment of Gene 10760

```
pMEJ4T          235 agctatttaggtgacactatagaatactcaagctatgcatcaagcttggt
M13R             23 ..................................................

pMEJ4T          285 accgagctcggatccactagtaacggccgccagtgtgctggaattcgccc
M13R             73 ..................................................

pMEJ4T          335 ttcatatgtcggccctgggaggattctcgcgtgtggcgtcgagcgtcgga
M13R            123 ..................................................

pMEJ4T          385 cgttcggcgcgcggcgcgcagctgctccgccacacccagacgcgcaagct
M13R            173 ..................................................

pMEJ4T          435 caacatccatgagtacgcgagcatgcaggtcatgaaggactacaatgtcc
M13R            223 ..................................................

pMEJ4T          485 ccgtgcccaagaacagcgtggccacctcgcccgaggaagcccaaaaggtc
M13R            273 ..................................................

pMEJ4T          535 tacgacgccggcgagctcggcgaggatgtggtcataaaggcccaggtcct
M13R            323 ..................................................

pMEJ4T          585 ggccggtggccgcggccgcggaaccttcaagaacggcttcgagggcggcg
M13R            373 ..................................................

pMEJ4T          635 tccagatggtgtccaagtctgaggaggccaaggagatcgccggcaagatg
M13R            423 ..................................................

pMEJ4T          685 ctcggccagctcctcgtcaccaagcagactggggaggagggcaagccctg
M13R            473 ..................................................
M13F           1040 ----------..........................................

pMEJ4T          735 tgagcgcgtcctgctcatggagcgcctcaagctcaagcgcgagacctact
M13R            523 ..................................................
M13F           1005 ..................................................

pMEJ4T          785 ttagtatcctccttgaccgctcgcacgtcggcgtccccctcgtcggctcg
M13R            573 ....................................a.............
M13F            955 ....................................a.............

pMEJ4T          835 cccgccggtggcatgagcatcgaggacgtcgccgctgagaccccggagaa
M13R            623 ..................................................
M13F            905 ..................................................

pMEJ4T          885 gatcttcaccgagaagattgacatcgtgagcggcctccaggacgaacagg
M13R            673 ..................................................
M13F            855 ..................................................

pMEJ4T          935 tcgagcgcctcgccaagaacctcggctacgagggcgacgcgcacaccaag
M13R            723 ..................................................
M13F            805 ..................................................

pMEJ4T          985 tgcgttcaggtcatcaagtcgctctacgacctttttatcgactgcgactg
M13R            773 ..................................................
M13F            755 ..................................................

pMEJ4T         1035 caccctcgtcgagatcaaccctcttgcggagaccgaggctggcgacgtct
M13R            823 ..................................................
M13F            705 ..................................................
```

```
pMEJ4T          1085 tctgctgcgacgccaagctcaacttcgacgacaacgcggagttccgccag
M13R             873 ..................................................
M13F             655 ..................................................

pMEJ4T          1135 gagaagatctttgccttccgcgaccgcaaccaggaggacaagcgcgaggt
M13R             923 ..................................................
M13F             605 ..................................................

pMEJ4T          1185 cgaggcctccaagtacggcctcaactacatcggcctcgacggcaccgtcg
M13R             973 ..................................................
M13F             555 ..................................................

pMEJ4T          1235 gctgcatggttaacggcgccggtctcgccatgagcaccctcgatatcgtc
M13R            1023 ..................................................
M13F             505 ..................................................

pMEJ4T          1285 accctccacggcggttcggccagcaacttcctcgaccttggcggcggcgc
M13R            1073 ..................................................
M13F             455 ..................................................

pMEJ4T          1335 caccaaggacgctgtgcagaaggcctttgagcttctcaacgaggacgaga
M13R            1123 ..................--------------------------------
M13F             405 ..................................................

pMEJ4T          1385 aggtcaagtccatcctcgtcaacatcttcggtggcatcatgcgctgcgac
M13F             355 ..................................................

pMEJ4T          1435 atcatcgccgagggcatggtcgccgccgccaaggagatcggcctcgagaa
M13F             305 ..................................................

pMEJ4T          1485 gcccctcgtcgtccgcctcaagggtaccaacgtggagcgcgcccgcgaga
M13F             255 ..................................................

pMEJ4T          1535 tcctcggtagctccgacgtcaagctcaccatgctcgacgacctcgacgag
M13F             205 ..................................................

pMEJ4T          1585 gccgccgcgaccgctgtcaagctcgctgccatgctacgggcccaagggcg
M13F             155 ..................................................

pMEJ4T          1635 aattctgcagatatccatcacactggcggccgctcgagcatgcatctaga
M13F             105 ..................................................

pMEJ4T          1685 gggcccaattcgccctatagtgagtcgtattacaat
M13F               55 ..................................
```

Sequence Alignment of Gene 9329

```
pMEJ6T           235 agctatttaggtgacactatagaatactcaagctatgcatcaagcttggt
M13R              24 ..................................................

pMEJ6T           285 accgagctcggatccactagtaacggccgccagtgtgctggaattcgccc
M13R              74 ..................................................

pMEJ6T           335 ttcctaggagaagagcgcgcgccttgatacgctcgacaatggccagcgtg
M13R             124 ..................................................

pMEJ6T           385 cgcttggcgttgcggtagtgaaggttttccacgagcttgccgtcgaccgt
M13R             174 ..................................................

pMEJ6T           435 cgcgacgccagacccggtgcgaacggcctcctcgtgggcctcgattaccc
M13R             224 ..................................................
```

```
M13F                988 ---.............................................

pMEJ6T              485 tattggcgtggtcaatgtcggcctcgctgggcgagaagcactcatttgcc
M13R                274 .................................................
M13F                941 .................................................

pMEJ6T              535 ttttggacagtcttgggatggatgagcgtcttgccgtcaaagccgaggtc
M13R                324 .................................................
M13F                891 .................................................

pMEJ6T              585 gcggccctgatggcattcgtcgataaagccttggtcgttctgcaggtcga
M13R                374 .................................................
M13F                841 .................................................

pMEJ6T              635 cgtagacgccgtcaagggcgctggcgccggcagcgcgggccgcaatcacg
M13R                424 .................................................
M13F                791 .................................................

pMEJ6T              685 cagcgctgcagcgcgtactgcaagttgaagcgcccgggggccatgggctt
M13R                474 .................................................
M13F                741 .................................................

pMEJ6T              735 gcagcgcaggtcgttggcgaggtcgacggtgcccatgagcagggccgtga
M13R                524 .................................................
M13F                691 .................................................

pMEJ6T              785 cgcgcggatgcatggcgagctcgtcgacgttgctaacaccgcgcggcgtc
M13R                574 .................................................
M13F                641 .................................................

pMEJ6T              835 tccaccatgcaccatacgtccatggagggtttgccgcccgcggcgtcgag
M13R                624 .................................................
M13F                591 .................................................

pMEJ6T              885 gcggtcgacgagatccgcgacgtcctcggcggactcggccttgggcagga
M13R                674 .................................................
M13F                541 .................................................

pMEJ6T              935 ggacggcgtccgcgccgctcgaggcgaccgcctggacatcgtcggcgtac
M13R                724 .................................................
M13F                491 .................................................

pMEJ6T              985 cagggcgttgtgcggccgttgacgcggatgcagatctcggacatgccgta
M13R                774 .................................................
M13F                441 .................................................

pMEJ6T             1035 ggacttttcctcggcgtgctcgcgggccgcggcgacggcctgggtgcgcg
M13R                824 .................................................
M13F                391 .................................................

pMEJ6T             1085 cgagctccttctggtcgggcgagacggcgtcctcgaggtcgaagatatag
M13R                874 .................................................
M13F                341 .................................................

pMEJ6T             1135 gcgtcggccttgatcgtcttggccttctccagcgcgcgctggcgcgacga
M13R                924 .................................................
M13F                291 .................................................

pMEJ6T             1185 cggcatgtacagcacgctgcggcgcaccgacgccaccgaggcgtcctcgg
M13R                974 ......................----------------------------
M13F                241 .................................................
```

```
pMEJ6T          1235 cggcgagcccgggcccgcccggcgccgtcgagaggcccgcggcggccgcc
M13F             191 ..................................................

pMEJ6T          1285 acggtccgggcgggcgacatcgtcatatgaagggcgaattctgcagatat
M13F             141 ..................................................

pMEJ6T          1335 ccatcacactggcggccgctcgagcatgcatctagagggcccaattcgcc
M13F              91 ..................................................

pMEJ6T          1385 ctatagtgagtcgtattacaa
M13F              41 ....................
```

Sequence Alignment of Gene 8081

```
pMEJ7T           235 agctatttaggtgacactatagaatactcaagctatgcatcaagcttggt
M13R              23 ..................................................

pMEJ7T           285 accgagctcggatccactagtaacggccgccagtgtgctggaattcgccc
M13R              73 ..................................................

pMEJ7T           335 ttgctagcacaaacatgacgcccgaggaggaggcctccttgtcgtcctcc
M13R             123 ..................................................

pMEJ7T           385 ttgagcttgtcgagcgagagcttgagctccgcgttgaggcgctgcagcac
M13R             173 ..................................................

pMEJ7T           435 ctggtcgtgctcgaggttctcgttccagaagttgataaagtcttcctcgt
M13R             223 ..................................................

pMEJ7T           485 aaatgcggccgtggtcggcgcccgggatcttagcaaagagctcgagcatt
M13R             273 ..................................................

pMEJ7T           535 tcctcgggcgaggcgggctcgaagccgagaaccttggcggcgttctgcat
M13R             323 ..................................................

pMEJ7T           585 ctcgaactggtccacgaagccgctgccatcgtggtcgagcgagcggaaga
M13R             373 ..................................................

pMEJ7T           635 gcgcagacagatcggggagcggctgccgcggctcgtccgagaccgggtgc
M13R             423 ..................................................

pMEJ7T           685 gaaccgccaaactcgctgagaaagctcatggcgtcgtcgagctgcctttt
M13R             473 ..................................................

pMEJ7T           735 cttctcggcaccgtcccagccgagcaggtcggccataatgttggccacct
M13R             523 ..................................................
Seq             1059 --------------------------------..................

pMEJ7T           785 tcggcgcggcaaacttggccgcgtcagagttcaggtacgcgaggcgtgtg
M13R             573 ..................................................
Seq             1041 ..................................................

pMEJ7T           835 cgaatgctcagcatgtcgcgcaccgtgcgcgcgtattcgttgcgcacggc
M13R             623 ..................................................
Seq              991 ..................................................

pMEJ7T           885 gaactcgatctcggcctcgagatacgggtagttctccacgatgggaacgc
M13R             673 ..................................................
Seq              941 ..................................................
```

```
pMEJ7T          935 caaagcgcggccaacccttggaagtggggcgcacgagcttgcacacgtcc
M13R            723 ..................................................
Seq             891 ..................................................

pMEJ7T          985 caagcgcgcactccgtaggttcgtgcgaggtggcgcgccacattttcgga
M13R            773 ..................................................
Seq             841 ..................................................

pMEJ7T         1035 aatgctgtaagtctgaatgagctgcacaaagaggtcctcctcgtagccct
M13R            823 ..................................................
Seq             791 ..................................................

pMEJ7T         1085 cgccgccgacgaaggggatcttggtcgtgcggcacggaccggcgcgcgcc
M13R            873 ..................................................
Seq             741 ..................................................

pMEJ7T         1135 ttgagctccggagcttccttgatgacgcggtccacgacatcttctgccat
M13R            923 ..........................----------
Seq             691 ..................................................

pMEJ7T         1185 ctcacgcgaggtggtccacttgccgcccaccagaaagatgaccttggact
Seq             641 ..................................................

pMEJ7T         1235 ctgggttgaccgagatagtgtggtcgcggctggccggggcgccgggcgggg
Seq             591 ..................................................

pMEJ7T         1285 gcgtgcgggtccacggccagcgggcgccagccgcgccaggctgacaggac
Seq             541 ..................................................

pMEJ7T         1335 gtcggagcggcggacgcgaaggtgcgggctcaagtacttttttcacttcgt
Seq             491 ..................................................

pMEJ7T         1385 tcaggatccactggatctcttcctcgggggggtgccggcgtgctcgagggc
Seq             441 ..................................................

pMEJ7T         1435 tcgcccttgcggtccgtcgtgccgaccacggtgctgccctgccagggcag
Seq             391 ..................................................

pMEJ7T         1485 gaagaacaagaagcggccgtcgctggtgttgatgtccaagaggcccatgt
Seq             341 ..................................................

pMEJ7T         1535 cgggcgaggaaaagtacccaggcagcaccaggtgggtcccggctgcgccg
Seq             291 ..................................................

pMEJ7T         1585 cgcacggccggcggcgcgtcagggtcctcgagggcgcgcagctggtcggt
Seq             241 ..................................................

pMEJ7T         1635 aaagggccctcccgccaggacgatcttgttggcgcggatcgaaaaggtat
M13F           1048 --------------..................................
Seq             191 ..................................................

pMEJ7T         1685 cgcccgagatctcgtcgcgcgcgcgcacgcccgtgacgcgggcggcgtct
M13F           1012 ..................................................
Seq             141 ..................................................

pMEJ7T         1735 tgcgggtcctccttgagcatttccaccacgttaacgtggttggccacggt
M13F            962 ..................................................
Seq              91 ..................................................

pMEJ7T         1785 ggcgccctgctcggcagccgtcatagccatggccacgcaggttcgcgcgt
```

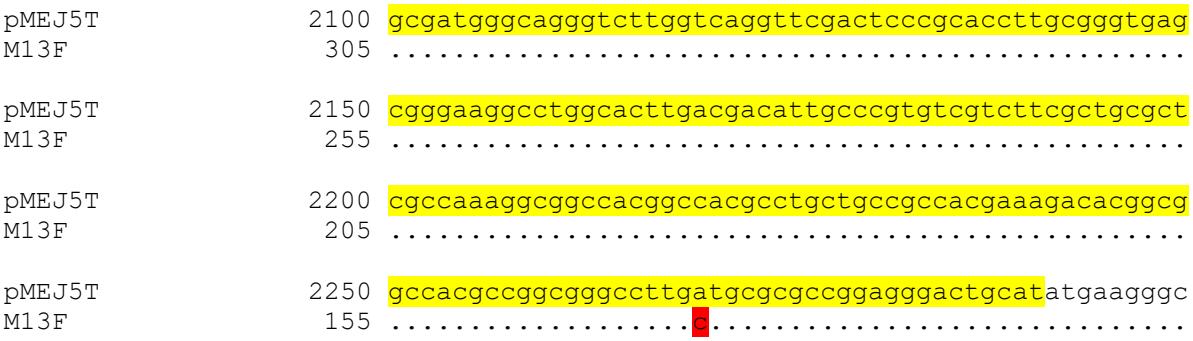
```
M13F             912 ..................................................
Seq               41 ..............................--------------------

pMEJ7T          1835 cgttgtgctggccttcataaaagacctgcgtatacttcatgcgcccgctg
M13F             862 ..................................................

pMEJ7T          1885 tcgagctgcgggaagcgctctcgcgcgccgcttggacatgacgtaact
M13F             812 ..................................................

pMEJ7T          1935 tggcgggcacgtaaagccgcagagcgaatcgtagaacttgaagatgcccg
M13F             762 ..................................................

pMEJ7T          1985 gcaggatcatgccggccatggcaaaaatcggatggcccatcggcggcggc
M13F             712 ..................................................

pMEJ7T          2035 caggagatccagcggtcgatgggcacggcgatggcgatccagttggtgag
M13F             662 ..................................................

pMEJ7T          2085 gtgcggctgggtctccagaaagtaccttcgctcgcgctgcgcgcccatga
M13F             612 ..................................................

pMEJ7T          2135 ccatcttgatctcgccgacaaagttcttgaccgactgcacggggctcagc
M13F             562 ..................................................

pMEJ7T          2185 agcgtctcgcgcgagagcagtgaggccacggccgtgcctaggtaccttac
M13F             512 ..................................................

pMEJ7T          2235 gccagcccagatgagcttggtgctgcgggagctggtctcgccggagaagt
M13F             462 ..................................................

pMEJ7T          2285 cctggcgctccacgagggccaccgagagccctcgcgacgccgcgtcgagc
M13F             412 ..................................................

pMEJ7T          2335 gccgtggcgcacccggtggcgcctccgccgatgaccaggacatcaaaggc
M13F             362 ..................................................

pMEJ7T          2385 ctcctcgcggcgcagcgctttcacctggtcctcgcgtgaggccacgccct
M13F             312 ..................................................

pMEJ7T          2435 tcttggcggccgtggacagtgtcctctccgccgccgacgtccctggcgcg
M13F             262 ..................................................

pMEJ7T          2485 gcggcggcggcggcggcgggcgtccgcgtggacaggagccgccgcgaggg
M13F             212 ..................................................

pMEJ7T          2535 cggcgttctcgcgcggcccaggttccgcagcgtcatcctcgccgccatat
M13F             162 ..................................................

pMEJ7T          2585 gaagggcgaattctgcagatatccatcacactggcggccgctcgagcatg
M13F             112 ..................................................

pMEJ7T          2635 catctagagggcccaattcgccctatagtgagtcgtattacaat
M13F              62 ..........................................
```

Sequence Alignment of Gene 6468

```
pMEJ5T           300 actagtaacggccgccagtgtgctggaattcgcccttgggcccgtagcat
M13R              88 ..................................................

pMEJ5T           350 ctcttcggggagcttcatggtctcaaggaagtcgagcgcgttcttggtct
```

```
M13R                138 ..................................................

pMEJ5T              400 cttcgctgcgcttggagcgcgaccagccgagcatggtgcccatgagattg
M13R                188 ..................................................
Seq                1001 ----------------------------------------.........

pMEJ5T              450 acgacctggggcacggcctcaacggcggccttgtggtcgacaaaggcgag
M13R                238 ..................................................
Seq                 992 ..................................................

pMEJ5T              500 acgggtcctccgggccaggacgtcgacagcagtgacggcgtactcttgct
M13R                288 ..................................................
Seq                 942 ..................................................

pMEJ5T              550 ggacggcaaaggcgacctcggcctcgaggaccgggtgcttggcgacgagg
M13R                338 ..................................................
Seq                 892 ..................................................

pMEJ5T              600 cgcttagggtggaggcccttggcgcgcgtggggtagccgtgcttgatgat
M13R                388 ..................................................
Seq                 842 ..................................................

pMEJ5T              650 ctcggcgatctgcagggcgcgcgtgccgtaggagctgacgaggtgctggg
M13R                438 ..................................................
Seq                 792 ..................................................

pMEJ5T              700 cgatgtcgcggtcaaagccgtagtcctcgcgaagagtgacagtgacgatg
M13R                488 ..................................................
Seq                 742 ..................................................

pMEJ5T              750 tcgtacttttggtcgcaaatgatgccggcgcggtcggcaccaatgagctt
M13R                538 ..................................................
Seq                 692 ..................................................

pMEJ5T              800 gagaccgagggtggcgcaagggccggccttagcaccgaggcgggggaact
M13R                588 ..................................................
Seq                 642 ..................................................

pMEJ5T              850 ccttgagcacgcggtcgacggcgtcctgggccatgcggcgctgggtggtc
M13R                638 ..................................................
Seq                 592 ..................................................

pMEJ5T              900 cacttgccgccgccgatggtgatgaggccgctgtccgagacctcgacgac
M13R                688 ..................................................
Seq                 542 ..................................................

pMEJ5T              950 gtgctcgcgggagatggccttggtgtccttggcgttgacgtcgcggacga
M13R                738 ..................................................
Seq                 492 ..................................................

pMEJ5T             1000 gggggcggatgccggaccaggcggcacggacgtcctcgtgggagacgcgg
M13R                788 ..................................................
Seq                 442 ..................................................

pMEJ5T             1050 cggttgaggtagcggccgcactcggcgaggacaaagtcgacctcgggaag
M13R                838 ..................................................
Seq                 392 ..................................................

pMEJ5T             1100 actgggcttggggcgcatagtgagcggggtggccgagtcggtagtgccga
M13R                888 ..................................................
Seq                 342 ..................................................
```

```
pMEJ5T          1150 cgatggtgccgccgctccaggggaggaagaagagtacgcggccgtcagag
M13R             938 ..................................................
Seq              292 ..................................................

pMEJ5T          1200 gtgttggggacgatgaggcccatctgggacgggctaaagtcgtcggagag
M13R             988 .........--------------------------------------------
Seq              242 ..................................................

pMEJ5T          1250 agtgacgtggatgccgccggccgggacgataatgtcctgctccttctttt
Seq              192 ..................................................

pMEJ5T          1300 ccgtagacgggccgttgtcctcgccgttggggtcgttggccatgtggcgc
Seq              142 ..................................................

pMEJ5T          1350 acggcgtcggaaaagggaccggtggcgttgacgacggtggtggcgcgcac
M13F            1032 ----------------------..........................
Seq               92 ..................................................

pMEJ5T          1400 gtcgatgaccttgccggtctgggtgtcgcgcacgcgggcgccgcagagct
M13F            1005 ..................................................
Seq               42 ..........----------------------------------------

pMEJ5T          1450 ttccttcctcgtccttgaggagcttgtcggcgttcatgtagttgacagcg
M13F             955 ..................................................

pMEJ5T          1500 gtagcgccggcctggatggccgtaaggaggatgagcaggccgtagcgcgc
M13F             905 ..................................................

pMEJ5T          1550 gtcgtccatctggccgtcgtagtacacaatagcaccggcgatcttggagg
M13F             855 ..................................................

pMEJ5T          1600 ggtcaagcatggggaagttgtagagcgcctgctgcttggtgaggaacatg
M13F             805 ..................................................

pMEJ5T          1650 gagctgggcacgcccgagttctggccagcaaccatgtcgtagaccttggt
M13F             755 ..................................................

pMEJ5T          1700 gccggcgtagtaataagggaccttccaccactcctcgatggggaccatga
M13F             705 ..................................................

pMEJ5T          1750 tggggagcggcttggcgatgaaggggcggcctggagcatgtgacttcgc
M13F             655 ..................................................

pMEJ5T          1800 tcttcgagcgcctccttgacgagcaggtactggccgtagtcgagcttgag
M13F             605 ..................................................

pMEJ5T          1850 gaaggcgttctcgaggtagcgcacgccgccgtggatgagcttggtggacc
M13F             555 ..................................................

pMEJ5T          1900 ggctagaagtgccggagctaaagtcctcgcgctcgacgacggcgacgcgg
M13F             505 ..................................................

pMEJ5T          1950 aggccgcgggttgcggcgtcgagggcgacgccggcgccgacgcagccgcc
M13F             455 ..................................................

pMEJ5T          2000 gccgatgaccaggaggtcgtacttttctgtctcgagctccttgagctgct
M13F             405 ..................................................

pMEJ5T          2050 cctcgcgcgagggcggggaccaggggaagaccatggtcccgtcgcggcac
M13F             355 ..................................................
```


pMEJ5T 2100 gcgatgggcagggtcttggtcaggttcgactcccgcaccttgcgggtgag
M13F 305 .....
pMEJ5T 2150 cgggaaggcctggcacttgacgacattgcccgtgtcgtcttcgctgcgct
M13F 255 .....
pMEJ5T 2200 cgccaaaggcggccacggccacgcctgctgccgccacgaaagacacggcg
M13F 205 .....
pMEJ5T 2250 gccacgccggcgggccttgatgcgcgccggagggactgcatatgaagggc
M13F 155 .....c.....

Appendix E – Abbreviations, Units and Symbols

˚C - celsius

3D – three dimensional

3XLB – 3 times concentrated Luria Broth

A – adenine

aa – amino acid

ACL – ATP-dependent citrate lyase

ACP – acyl carrier protein

ADP - adenosine diphosphate

AMP - adenosine monophosphate

amp - ampicillin

ATP - adenosine triphosphate

bla - beta lactamase (ampicillin resistance)

BLAST – Basic Local Alignment Search Tool

bp – base pair

BSA – bovine serum albumin

C – cytosine

CDW – cell dry weight

CoA – coenzyme A

CV – column volume

DH - dehydratase

DH/i - combined dehydratase and trans-cis isomerase

DHA – docosahexaenoic acid

DHAP – dihydroxy acetone phosphate

DNA – deoxy ribonucleic acid

dNTP – dinucleotide triphosphate

DTE - 1,4-dithioerythritol

DTT – 1,4-dithiothreitol

EDTA – Ethylene diamine tetraacetic acid

ELO – elongase

EPA - eicosapentaenoic acid

ER - enoyl reductase

$FAD/FADH_2$ – flavin adenine dinucleotide

FAS – fatty acid synthase

FPLC - fast protein liquid chromatography

G3P – glycerol 3-phosphate

G3PDH – glycerol 3-phosphate dehydrogenase

GMQE - Global Model Quality Estimate

GSQE - Quaternary Structure Quality Estimate

h – hour

HIVE-CUTs - High-performance Integrated Virtual Environment-Codon Usage Tables

ICDH - isocitrate dehydrogenase

IMAC - immobilized metal-affinity chromatography

IMP - inosine monophosphate

kan - kanamycin

kbp – kilo base pairs

kDa – kilo Dalton

KR - keto reductase

KS - ketoacyl transferase

LA – Luria broth agar

LB – Luria broth

M - molar

mAU – milli absorbance units

MDH – malate dehydrogenase

mG3PDH – mitochondial glycerol 3-phosphate dehydrogenase

min – minutes

MOPS - 3-morpholinopropane-1-sulfonic acid

M_w – molecular weight

NAD^+/NADH – nicotinamide adenine dinucleotide

$NADP^+$/NADPH – nicotinamide adenine dinucleotide phosphate

NCBI – National Center for Biotechnology Information

NEB – New England Biolabs

NTNU – Norwegian University of Science and Technology

OAA – oxaloacetate

OD_{λ} – optical density at wavelength λ

ON – over night

ORF – open reading frame

PAGE – poly acrylamide gel electrophoresis

PCR – polymerase chain reaction

PDB – Protein Data Base

P_i – inorganic phosphate

PKS – polyketide synthase

PUFA – poly unsaturated fatty acid

RNA - transfer ribonucleic acid

RO – reverse osmosis treated

rpm – rounds per minute

RT – room temperature

s – seconds

SDS – sodium dodecyl sulfate

SOC – Super Optimal broth with Catabolite repression

strep – streptomycin

TAE – Tris base, acetic acid, and EDTA

TBS - Tris-buffered saline

TBST – Tris-buffered saline Tween-20

TCA – Tricarboxylic acid cycle

TFB – Transformation buffer

TMS – Trace Mineral Mix

tRNA – transfer ribonucleic acid

UV – ultra violet

v/v – volume by volume

w/w – weight by weight

xg – times gravity

YPD – yeast extract peptone dextrose

Ω - omega