# Semi-rational engineering of Adh2 protein for better methanol utilization in yeast Komagataella phaffii

**Kalogjera, Lara**

UNIVERSITY OF ZAGREB
FACULTY OF FOOD TECHNOLOGY AND BIOTECHNOLOGY

# GRADUATE THESIS

Zagreb, March 31, 2023

Lara Kalogjera

# SEMI-RATIONAL ENGINEERING OF Adh2 PROTEIN FOR BETTER METHANOL UTILIZATION IN YEAST *Komagataella phaffii*

This study was carried out at the Institute of Microbiology and Microbial Biotechnology at the University of Natural Resources and Life Sciences, Vienna, under the supervision of Božidar Šantek, PhD, Full professor and Diethard Mattanovich, Univ. Prof. Dipl.-Ing. Dr. nat. techn., with the assistance of Charles Moritz, MSc.

This graduate thesis was carried out as a part of the ACIB Lighthouse Project with Univ. Prof. Dipl.-Ing. Dr. nat. techn. Diethard Mattanovich as a project leader.

I want to express my sincere gratitude to Univ. Prof. Dipl.-Ing. Dr. nat. techn. Diethard Mattanovich, who graciously welcomed me into his laboratory at the Institute of Microbiology and Microbial Biotechnology, University of Natural Resources and Life Sciences, Vienna. I am also deeply grateful to my supervisor, Charles Moritz, MSc, for his invaluable scientific guidance and assistance with my thesis. I want to express my appreciation to Lisa Lutz, MSc, and the rest of the lab technicians, as well as the entire research group, for answering all my questions with good will. Lastly, I would like to thank Božidar Šantek, PhD, Full professor for his mentorship at the Department of Biochemical Engineering, Faculty of Food Technology and Biotechnology, Zagreb.

**BASIC DOCUMENTATION CARD**

**Graduate Thesis**

**University of Zagreb**
**Faculty of Food Technology and Biotechnology**
**Department of Biochemical Engineering**
**Laboratory for Biochemical Engineering, Industrial Microbiology and Malting and Brewing Technology**
**Scientific area:** Biotechnical Sciences
**Scientific field:** Biotechnology

**Graduate university study programme:** Molecular Biotechnology

# SEMI-RATIONAL ENGINEERING OF Adh2 PROTEIN FOR BETTER METHANOL UTILIZATION IN YEAST *Komagataella phaffii*

*Lara Kalogjera, univ. bacc. biotechn. et pharm. inv.*
*0335003406*

**Abstract:** The methylotrophic yeast *Komagataella phaffii* is commonly used as an expression system for recombinant protein production. In a recent study, it was demonstrated that the promiscuous activity of alcohol dehydrogenase 2 (Adh2) is responsible for the majority of methanol oxidation in *K. phaffii* in the absence of *AOX* genes. Methanol oxidation by Adh increases the amount of cellular energy (NADH) produced via the dissimilatory pathway. However, in the absence of *AOX* genes Adh2 is unable to support a growth phenotype on methanol. This thesis aimed to employ semi-rational engineering of the Adh2 protein to produce mutant variants with improved *in vivo* activity towards methanol. Saturation mutagenesis was performed on target sites (A52, G272, L119, S249, T123, and Y297), and the activity of the mutated *ADH2* genes was determined through the *in vivo* production of eGfp under the *FLD1* promoter. The screening results were beneficial for confirming the efficiency of the biosensor strain.

**Keywords:** *Komagataella phaffii, Adh2, semi-rational enzyme engineering, site saturation mutagenesis, flow cytometry*
**Thesis contains:** 61 pages, 28 figures, 22 tables, 52 references, 6 supplements
**Original in**: English
**Graduate Thesis in printed and electronic (**pdf format**) form is deposited in:** The Library of the Faculty of Food Technology and Biotechnology, Kačićeva 23, Zagreb.
**Mentor:** Božidar Šantek, PhD, Full professor
**Co-mentor:** Diethard Mattanovich, PhD, Full professor
**Technical support and assistance**: Charles Moritz, MSc

**Reviewers:**
1. Blaženka Kos, PhD, Full professor (president)
2. Božidar Šantek, PhD, Full professor (mentor)
3. Diethard Mattanovich, PhD, Full professor (member)
4. Andreja Leboš Pavunc, PhD, Associate professor (substitute)

**Thesis defended:** March 31, 2023

**TEMELJNA DOKUMENTACIJSKA KARTICA**

**Diplomski rad**

**Sveučilište u Zagrebu**
**Prehrambeno-biotehnološki fakultet**
**Zavod za biokemijsko inženjerstvo**
**Laboratorij za biokemijsko inženjerstvo, industrijsku mikrobiologiju i tehnologiju slada i piva**
**Znanstveno područje:** Biotehničke znanosti
**Znanstveno polje:** Biotehnologija

**Diplomski sveučilišni studij:** Molekularna biotehnologija

# SEMI-RACIONALNO INŽENJERSTVO PROTEINA Adh2 U SVRHU BOLJE ISKORISTIVOSTI METANOLA U KVASCU *Komagataella phaffii*

*Lara Kalogjera, univ. bacc. biotechn. et pharm. inv.*
*0335003406*

**Sažetak:** Metilotrofni kvasac *Komagataella phaffii* učestalo se koristi kao ekspresijski sustav za proizvodnju rekombinantnih proteina. U nedavnom istraživanju, demonstrirano je da je promiskuitetna aktivnost alkohol dehidrogenaze 2 (Adh2) odgovorna za većinu oksidacije metanola u odsutnosti *AOX* gena. Oksidacija metanola putem alkohol dehidrogenaza povećava količinu stanične energije (NADH) proizvedenu tijekom disimilatornog puta. Unatoč tome, u odsutnosti *AOX* gena, Adh2 ne može podržati rast kvasca na metanolu. U ovom radu, cilj je bio upotrijebiti semi-racionalno inženjerstvo proteina Adh2 kako bi se proizvele mutirane varijante s poboljšanom *in vivo* aktivnošću prema metanolu. Usmjerena mutageneza bila je provedena na ciljnim mjestima u genu (A52, G272, L119, S249, T123, and Y297) i aktivnost mutiranih *ADH2* gena bila je određena *in vivo* produkcijom eGFP pod *FLD1* promotorom. Rezultati protočne citometrije bili su korisni kako bi se potvrdila efikasnost biosenzora.

**Ključne riječi:** *Komagataella phaffii, Adh2, semi-racionalno inženjerstvo, usmjerena mutageneza, protočna citometrija*
**Rad sadrži:** 61 stranicu, 28 slika, 22 tablice, 52 literaturnih navoda, 6 priloga
**Jezik izvornika**: engleski
**Rad je u tiskanom i elektroničkom (pdf format) obliku pohranjen u:** Knjižnica Prehrambeno-biotehnološkog fakulteta, Kačićeva 23, Zagreb
**Mentor:** prof. dr. sc. Božidar Šantek
**Komentor:** prof. dr. sc. Diethard Mattanovich, BOKU, Vienna
**Pomoć pri izradi**: Charles Moritz, MSc
**Stručno povjerenstvo za ocjenu i obranu:**
1. prof. dr. sc. Blaženka Kos (predsjednik)
2. prof. dr. sc. Božidar Šantek (mentor)
3. prof. dr. sc. Diethard Mattanovich (član)
4. izv. prof. dr. sc. Andreja Leboš Pavunc (zamjena)

**Datum obrane:** 31. ožujka 2023.

Contents

# 1. INTRODUCTION

Yeasts are commonly used as expression systems in molecular biology, as they are suitable for the large-scale production of recombinant eukaryotic proteins. The yeast *Komagataella phaffii* has emerged as a standard expression system for recombinant protein production due to the appropriate folding and secretion of recombinant proteins to the external environment of the cell (Karbalaei et al., 2020). *K. phaffii* displays methylotrophic properties distinctive for only a limited number of yeast genera (Wegat et al., 2022) and can assimilate methanol via the methanol utilization pathway. The *K. phaffii* expression system relies on the strong methanol-inducible promoters of the MUT pathway for the production of various protein and non-protein products.

The initial step of methanol utilization in *K. phaffii* is the oxidation of methanol to formaldehyde by alcohol oxidases 1 and 2. Until recently, the *AOX1* and *AOX2* knock-out mutants ($Mut^-$ phenotype) were considered unable to utilize methanol. In a recent study, it was demonstrated that the promiscuous activity of the alcohol dehydrogenases, mainly Adh2, allows for methanol metabolism in *K. phaffii* in the absence of alcohol oxidase genes (Zavec et al., 2021). This discovery represents a milestone that brought attention to the potential benefits of $Mut^-$ strain usage for recombinant protein production (Zavec, 2019). The theoretical benefits of methanol oxidation by alcohol dehydrogenases instead of alcohol oxidases include reduced oxygen demand, reduced heat generation, and increased energy efficiency through the production of an additional molecule of NADH for each molecule of methanol consumed. Despite the theoretical advantages, the limitations of using alcohol dehydrogenases for methanol oxidation in *K. phaffii* include low activity towards methanol and unfavourable thermodynamics at mesophilic temperatures. By overexpressing the *ADH2* gene in the *K. phaffii* $Mut^-$ strain, both the specific methanol uptake rate and the recombinant protein production were increased. However, the major limitation of utilizing $Mut^-$ Adh2 overexpression strains remains the lack of a growth phenotype.

To overcome the limitations of directed evolution, a semi-rational engineering approach was applied for the creation of a smaller functionally-rich library in this graduate thesis. Using the published crystal structure of the *K. phaffii* Adh2 protein (Zhang et al., 2018) and molecular dynamics simulations, the targets for site saturation mutagenesis were selected. Site saturation mutagenesis was performed using the overlap extension PCR with degenerate synthetic oligonucleotide primers to fully randomize the target sites, and the mutated *ADH2*

genes were cloned into the vector using the modular cloning system GoldenPiCS. Bacterial and yeast library were prepared and sequenced to confirm the site saturation mutagenesis and the successful transformation. Lastly, screenings of the strains were performed using *in vivo* flow cytometry to analyze the activity of the mutant Adh2 enzyme variants towards methanol.

Methylotrophic yeasts such as *K. phaffii* are poised to be useful production hosts on this sustainable feedstock and to expand the potential of a circular bioeconomy; however, the energetic inefficiency of the AOX-based MUT pathway represents a major limitation to the feasibility.

This graduate thesis aimed to employ semi-rational enzyme engineering to produce variants of the Adh2 enzyme with improved *in vivo* activity towards methanol. Coupled with strain engineering, this may lead to the realization of a more efficient Adh2-based MUT pathway in *K. phaffii*. Such strains could then be used as the basis for highly efficient conversion of methanol into biomass, proteins, and chemicals.

# 2. LITERATURE REVIEW

## 2.1. HISTORY OF *K. phaffii*

*Komagataella phaffii* (previously named *Pichia pastoris*) is a non-conventional methylotrophic yeast regarded as a highly renowned production host in modern biotechnology and an emerging yeast model organism in fundamental research (Bernauer et al., 2021). The discovery of the *P. pastoris* species is attributed to Alexandre Guilliermond, who isolated the yeast from chestnut trees in 1919 and originally named it *Zygosaccharomyces pastori* (Guilliermond, 1920). In the following years, the species was renamed *Saccharomyces pastori*. In his 1956 paper, Herman Phaff proposed the transfer of the *Saccharomyces pastori* and *Saccharomyces pini* species from the *Saccharomyces* genus to the *Pichia* genus (Phaff and Knapp, 1956). Both species possessed two unusual properties for the *Pichia* genus; neither formed a pellicle or a pseudomycelium. To accommodate these species, Phaff suggested modification of the *Pichia* genus definition (Phaff, 1956).

A defining discovery in the history of *K. phaffii* was the yeast's property of assimilating methanol as a sole source of carbon and energy in 1969. Prior to the discovery, the property of methanol utilization (MUT) in microorganisms was observed only in bacteria (Ogata et al., 1969). At the time, methanol was an inexpensive raw material, and it was easily accessible for the industrial cultivation of methylotrophic yeasts for biomass and single-cell protein production (Johnson and Echavarri-Erasun, 2011). The Phillips Petroleum Company developed such high-cell-density fermentation technology using *P. pastoris* as a production host. Due to the economic climate following the 1973 oil crisis, plants gained an advantage over methylotrophic yeasts for protein production, and the technology was repurposed for industrial yeast cultivation and bacterial fermentation. In the 1980s, the production of heterologous recombinant proteins started emerging as the primary application of the methylotrophic yeast *P. pastoris.* High-level expression systems were developed for the recombinant protein production in *P. pastoris*, most commonly using the highly expressed *AOX1* regulatory sequence as a promoter (Wegner, 1990). A study from 1989 indicated that the glycosylation patterns in *P. pastoris* resemble that of higher eukaryote to a higher degree than in *S. cerevisiae*, which accentuated the significance of *P. pastoris* as a model organism and a production host (Grinna and Tschopp, 1989).

It was only in 1995 that the *P. pastoris* strain was reclassified to the current genus, *Komagataella*, based on the phylogenetic analysis of the 18S and 26S ribosomal RNA

sequences (Yamada et al., 1995). The *Komagataella* genus was subdivided into the *Komagataella phaffii*, *Komagataella pastoris*, and *Komagataella pseudopastoris* species a decade later (Kurtzman, 2005). Thus far, seven methylotrophic species have joined the yeast genus *Komagataella*. The first sequenced *K. phaffii* genomes were published in 2009 for the GS115 and DSMZ70382 strains, and they consisted of 9.4 Mb distributed in 4 chromosomes with 5,450 possible protein-coding genes (De Schutter et al., 2009; Mattanovich et al., 2009). The published sequences contributed to a better understanding of the yeast's physiology and genetics, which consequently allowed for the creation of combinatorial optimization strategies for industrial protein production, including strain engineering and the construction of efficient expression cassettes (Fischer and Glieder, 2019; Peña et al., 2018; Prielhofer et al., 2017). In recent years, attempts to adapt the CRISPR/Cas9 system for precise and markerless genome editing in *K. phaffii* for metabolic engineering applications have shown success (Gassler et al., 2019; Weninger et al., 2018).

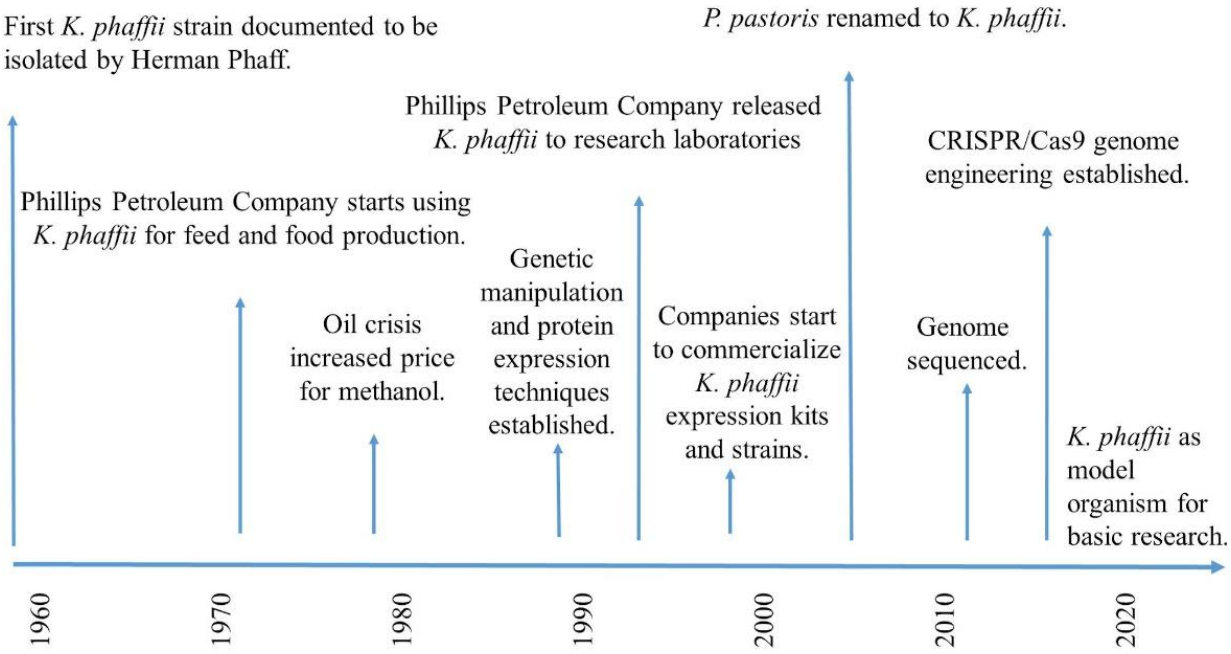



**Figure 1.** A timeline of the most important milestones in the history of *K. phaffii* research (adapted from Bernauer et al., 2021)

The advantageous characteristics of *K. phaffii* for industrial protein production, combined with years of research and numerous attempts to optimize the production processes, have enabled *K. phaffii* to become one of the most commonly used expression systems for recombinant protein production. Moreover, *K. phaffii* has also been used as a production host for various non-protein products using different carbon sources, and its non-conventional

properties brought attention to its appeal as a model organism (Ata et al., 2021). The most important milestones in the history of *K. phaffii* research are shown in Figure 1.

## 2.2. *K. phaffii* PROPERTIES

The yeast *K. phaffii* is commonly found in the phyllosphere, an environment predominantly inhabited by methylotrophic species. Due to the presence of methane-oxidizing bacteria and the pectin and lignin decomposition, the phyllosphere is rich in methanol which the methylotrophs use as a carbon and energy source. A study by Yurimoto et al. (2021) examined the methanol concentration in the phyllosphere using a methanol sensor based on the yeast *C. boidinii* expressing the yellow fluorescent protein Venus under the control of the *DAS1* promoter. The methanol concentration in the phyllosphere of growing young leaves in the dark period was 25–60 mM, in contrast to the methanol concentration greater than 250 mM (1 % (v/v) of methanol) on dead leaves. Yurimoto et al. (2021) have proposed that the methanol-induced gene expression in the methylotrophic yeasts results from the species' adaptation for growth in the methanol-abundant phyllosphere.

*K. phaffii* is a Crabtree-negative obligate aerobic heterotroph. Methylotrophy and the limited, non-fermentative glucose uptake under aerobic conditions are some of the prominent non-conventional features that distinguish *K. phaffii* from the conventional yeast *Saccharomyces cerevisiae*. *K. phaffii* is relatively easy to cultivate. As a result of the non-fermentative metabolism in the presence of glucose, *K. phaffii* can grow to high-cell densities over 100 g/L without accumulating toxic levels of ethanol (Cregg, 2007). However, in 2018, a group of scientists induced a respiro-fermentative phenotype in *K. phaffii* by overexpressing a single Gal4-like transcription factor, which demonstrated that a single mutation may have been a driving force for the evolution of the Crabtree effect (Ata et al., 2018). Besides methanol and glucose, *K. phaffii* can grow on various carbon sources, including glycerol, ethanol, sorbitol, L-rhamnose, mannitol, citric acid, and acetic acid (Ata et al., 2021). Recent engineering strategies to turn *K. phaffii* from a heterotroph to an autotroph have shown success and allowed the yeasts' growth on $CO_2$ as the only carbon and energy source (Gassler et al., 2020). Moreover, Gassler et al. (2022) explored the potential of improving the autotrophic growth of *K. phaffii* strains on $CO_2$ to produce value-added products.

*K. phaffii* expression systems most commonly rely on the strong methanol-inducible promoters of the MUT pathway for producing various bacterial and eukaryotic proteins.

Initially, such systems faced technical limitations due to the flammability, toxicity, and high metabolic heat release of methanol. Therefore, methanol-free expression systems were developed using constitutive promoters like the *GAP* dehydrogenase promoter. However, methanol-dependant systems based on the *AOX* promoter persisted as the preferred choice for industrial applications due to the high titer of recombinant proteins (Carneiro et al., 2022). In comparison with the well-studied conventional yeast *S. cerevisiae*, *K. phaffii* possesses certain advantages for heterologous protein production. The advantage of the *K. phaffii* expression system includes shorter and less immunogenic glycans and a higher yield of heterologous proteins than *S. cerevisiae* (Tran et al., 2017). Overall, *K. phaffii* expression systems ensure appropriate folding in the endoplasmatic reticulum and successful protein secretion to the extracellular environment, combined with the low-level secretion of endogenous proteins (Karbalaei et al., 2020).

## 2.3. BACTERIAL AND YEAST METHYLOTROPHY

Methylotrophic bacteria can proliferate on a wide range of reduced carbon substrates, such as methane, methanol, and methylamine. Their yeast counterparts, discovered much later, can only utilize methanol as a carbon source (Yurimoto et al., 2011). The ability of methylotrophic yeasts to use methanol as the sole carbon and energy source is based on the assimilation pathway that occurs in the peroxisomes. The methanol-utilizing pathway, named the xylulose-monophosphate (XuMP) cycle or the dihydroxyacetone (DHA) cycle, is common for all methylotrophic yeasts (Heistinger et al., 2020). In methylotrophic prokaryotes, three more aerobic methanol assimilating pathways are represented - the Calvin-Benson-Bassham (CBB) cycle, the Ribulose Monophosphate (RuMP) cycle, and the Serine cycle (Cotton et al., 2020) (Figure 2).

The XuMP cycle shares some similarities with the CBB pathway, except it is entirely localized in the peroxisomes. Upon methanol induction, enzymes employed in the MUT pathway in *K. phaffii* are synthesized in the peroxisomes. During the growth on methanol, the peroxisomes extensively proliferate, and can comprise 80 % of the cytoplasm (Hartner and Glieder, 2006). When switching the medium to glucose or ethanol, peroxisomal degradation occurs. In the natural habitat of *K. phaffii*, the fluctuation of the methanol concentration dictates peroxisomal proliferation and degradation (Yurimoto et al., 2021). For this reason, *K. phaffii* serves as an attractive model organism for investigating molecular mechanisms of peroxisome biogenesis and degradation (Heistinger et al., 2020).

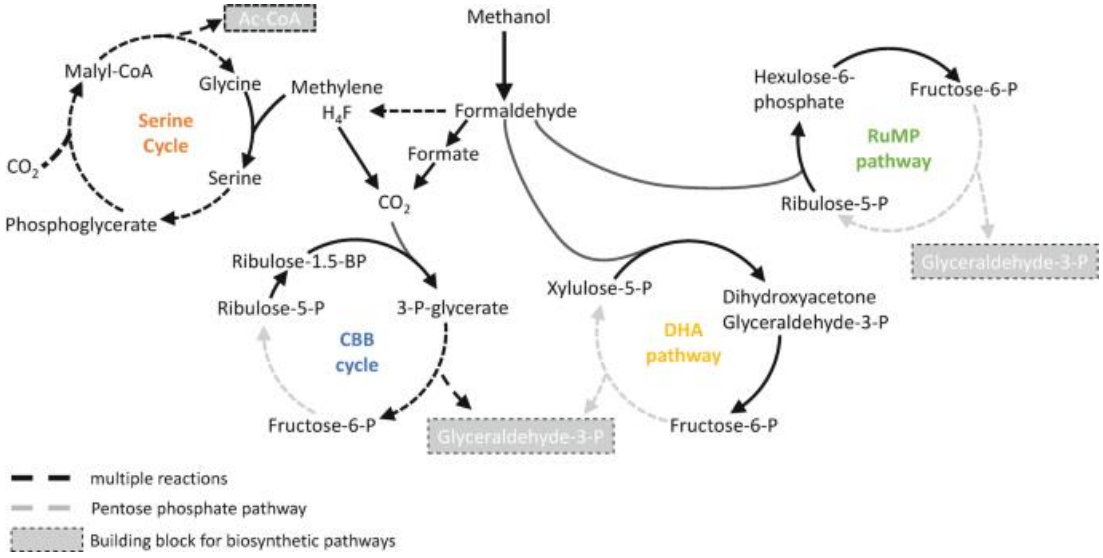



**Figure 2.** Common metabolic pathways for aerobic methanol utilization in aerobic methylotrophic yeasts and prokaryotes (adapted from Wendisch et al., 2021)

## 2.4. METHYLOTROPHY AND ITS ROLE IN CIRCULAR BIOECONOMY

Climate change and the increasing emission of greenhouse gases, in particular $CO_2$, seek for new solutions to reduce the harmful effects on the environment. One of the technologies for mitigating the carbon footprint is carbon capture and utilization, a powerful strategy toward a circular bioeconomy (Fabarius et al., 2021). Conversion of $CO_2$ into biomass or bio-chemicals is made possible by synthetic biology, and it eliminates the need for conventional organic feedstock and helps to establish $CO_2$-neutral or $CO_2$-negative industrial processes (Gassler et al., 2022; Fabarius et al., 2021). Methanol generated by hydrogenation or electrochemical reduction is regarded as sustainable feedstock (Cotton et al., 2020). With renewable energy, methanol can be derived from $CO_2$ and used as a carbon source for different production uses in *K. phaffii*. Methylotrophic yeasts such as *K. phaffii* can be engineered to achieve energetically-efficient production on sustainable feedstock and to exploit the full potential of a circular bioeconomy.

## 2.5. METHANOL UTILIZATION PATHWAY

In methylotrophic yeasts and bacteria, the prevalent first step of methanol metabolism is the formation of the central intermediate formaldehyde. The enzymes involved in methanol oxidation to formaldehyde differ depending on the methylotrophic species. Methylotrophic bacteria use alcohol dehydrogenases, and methylotrophic yeasts, including *K. phaffii*, utilize

alcohol oxidases. $H_2O_2$ is produced during the oxidation, which necessitates the localization of the alcohol oxidase to the peroxisome, where the $H_2O_2$ can be degraded (Zavec et al., 2021). Following methanol oxidation, formaldehyde is either assimilated for biomass synthesis or dissimilated for energy generation. If sufficient xylulose-5-phosphate is present within the peroxisome, formaldehyde assimilation occurs. Otherwise, formaldehyde diffuses to the cytosol, where the dissimilatory pathway occurs, and formaldehyde is further oxidized to $CO_2$ (van der Klei et al., 2006). In addition to the energy generation through NADH production, the dissimilatory pathway also performed the important physiological role of formaldehyde detoxification. In the assimilatory pathway, the compartmentalization of formaldehyde to the peroxisomes may also be a cellular strategy to protect the cell from the formaldehyde toxicity (Yurimoto et al., 2005). The methanol assimilation pathway proposed by Rußmayer et al. (2015) common for all methylotrophic yeasts is illustrated in Figure 3.

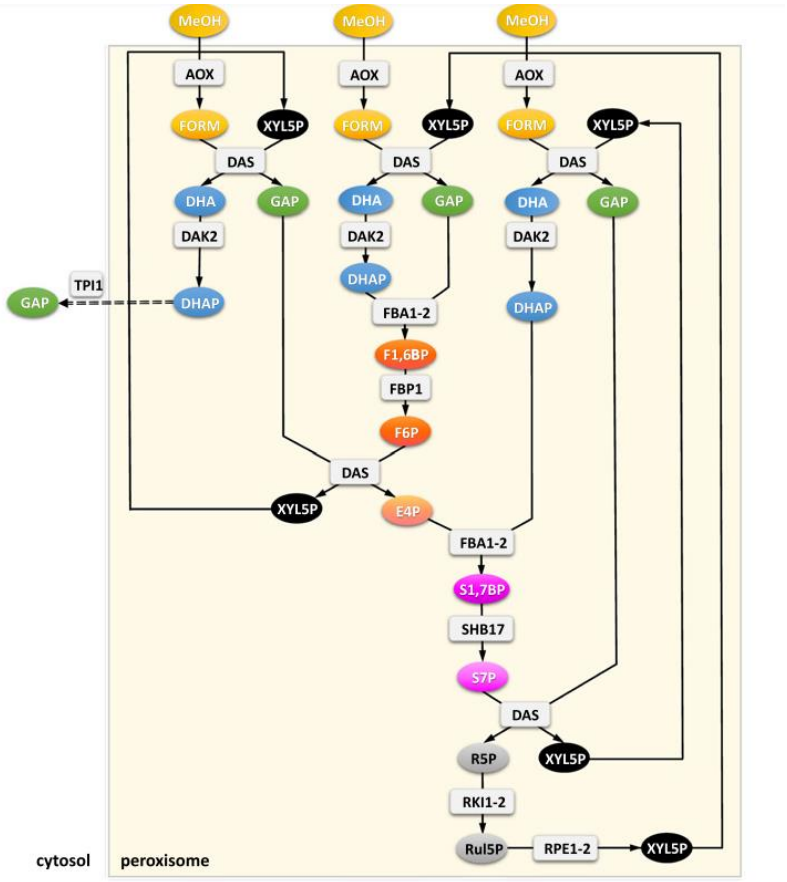



**Figure 3.** Methanol assimilation through the xylulose-monophosphate cycle localized to the peroxisomes.

The xylulose 5-phosphate regeneration occurs via sedoheptulose 1,7-bisphosphate formation. The net reaction of methanol assimilation is one mole DHAP or GAP out of peroxisomes per three moles of methanol (reprinted from Rußmayer et al., 2015)

The assimilatory pathway begins with a formaldehyde and xylulose-5-phosphate (Xu5P) reaction to form dihydroxyacetone (DHA) and glyceraldehyde-3-phosphate (GAP). The reaction is catalyzed by dihydroxyacetone synthase (Das) 1 and 2 in the peroxisomal matrix. DHA and GAP are further metabolized in the cytosol. DHA is phosphorylated by dihydroxyacetone kinase (Dak) into DHAP with the consumption of one ATP. Xu5P is regenerated over the intermediate sedoheptulose 1,7-bisphosphate. The net peroxisomal flux of carbon is one mole DHAP or GAP out of peroxisomes per three moles of methanol (Rußmayer et al., 2015).

In the dissimilatory pathway, formaldehyde diffuses out of the peroxisome into cytosol (Zavec, 2019). The reaction of formaldehyde with glutathione (GSH) forms S-hydroxymethylglutathione, which is reduced to $CO_2$ and NADH in two consecutive reactions (Gassler, 2019). The reactions are catalyzed by cytosolic formaldehyde dehydrogenase (Fld) and formate dehydrogenase (Fdh). The yield is two NADH molecules, which are used to produce energy for growth on methanol and the detoxification of formaldehyde and formate (Hartner and Glieder, 2006). The majority of methanol (50-80 %) is converted to $CO_2$ via the dissimilatory pathway, while the rest is assimilated to biomass and incorporated into the secreted recombinant proteins (Zavec et al., 2021).

## 2.6. MUT PHENOTYPES

Two alcohol oxidases are involved in the oxidation of methanol to formaldehyde in *K. phaffii* - alcohol oxidase 1 (Aox1) and alcohol oxidase 2 (Aox2). The terminal electron acceptor in the methanol oxidation reaction is $O_2$, which reduces the ATP yield. Different *K. phaffii* phenotypes are known based on the presence of *AOX1* and *AOX2* and the yeast's ability to utilize methanol (Zavec et al., 2021). $Mut^+$, the wild-type phenotype, is associated with active *AOX1* and *AOX2* genes that code for functional Aox proteins. The *AOX1* and *AOX2* genes are 97 % homologous, but the promoter regions of the genes bear differences, hence the variance in the expression of the two enzymes. The methanol-inducible *AOX1* promoter is repressed in the presence of excess glycerol, and it ensures extremely high recombinant protein expression levels (Stratton et al., 1998). The $P_{AOX2}$ produces far less enzyme than $P_{AOX1}$, whose product can make up more than 30 % of all soluble proteins in the cell. Therefore, *AOX1* knock-out mutants will grow on methanol at a significantly lower rate than the wild type or the *AOX2* knock-out mutants, displaying the methanol utilization slow ($Mut^S$) phenotype. The *AOX1* and

*AOX2* knock-out mutants display the Mut$^-$ phenotype, the methanol utilization negative phenotype.

The common *K. phaffii* strain for recombinant protein production is the Mut$^S$ strain, despite the disadvantages of such an expression system. They include the generation of excessive heat, high oxygen demand, and overall energetic inefficiency. To overcome these disadvantages, Zavec (2019) aimed to construct a *K. phaffii* Mut$^-$ strain capable of producing recombinant proteins on methanol induction, but the productivity was insufficient. However, Zavec (2019) demonstrated that, despite the limited cell growth, some form of methanol oxidation occurs in Mut$^-$ strains cultivated on methanol. This anomaly drew attention to the search for an alternative enzyme responsible for methanol oxidation in *K. phaffii* besides Aox1 and Aox2. In another study, Zavec et al. (2021) proved that the methanol metabolism in *K. phaffii* Mut$^-$ strains is possible due to the promiscuous activity of alcohol dehydrogenases (Adh).

## 2.7. ALCOHOL DEHYDROGENASE

Alcohol dehydrogenases are enzymes that catalyze the reverse oxidation of alcohols, using NAD or NADP as a cofactor (Zhang et al., 2018). The mechanism by which the alcohol dehydrogenase converts alcohol to acetaldehyde is shown in Figure 4.

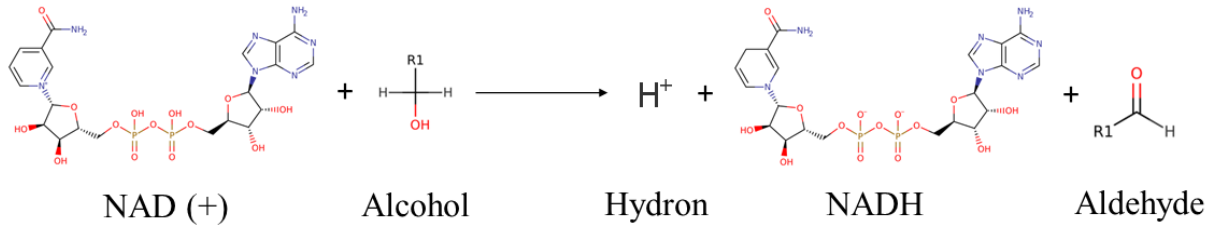



**Figure 4.** The conversion of alcohol to acetaldehyde by alcohol dehydrogenase in the cytoplasm by Holliday, G. L., & Sarkies, P. (n.d.). The alcohol dehydrogenase that catalyses the illustrated reaction belongs to the class I zinc-containing ADH superfamily (Retrieved December 22, 2022, from https://www.ebi.ac.uk/thornton-srv/m-csa/entry/256/)

In *K. phaffii*, six Adh genes are present, but the most highly transcribed ones are *ADH2* and *ADH900* (Zavec et al., 2021). In the 2016 study, Karaoğlan et al. characterized Adh2 (previously known as Adh3) as the only enzyme responsible for ethanol consumption in *P. pastoris*, but the genes responsible for ethanol production were unknown. In 2018, Zhang et al. published a crystal structure of the Adh2 enzyme, the first Adh structure from *K. phaffii*, solved

at 1.745 Å resolution (Zhang et al., 2018) (Figure 5). In the 2016 study, Karaoğlan et al. identified *ADH900* as the main gene responsible for ethanol production in *Pichia pastoris*. However, it was discovered that Adh2 also plays a minor role in ethanol production in the absence of the *ADH900* gene (Karaoğlan et al., 2020).

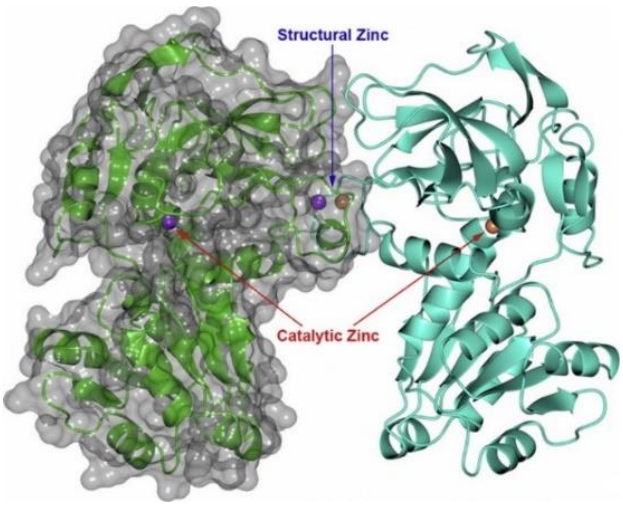



**Figure 5.** Dimeric arrangement and secondary structure assignment of *K. phaffii* GS115 Adh2 enzyme. The catalytic and structural zinc atoms are depicted with purple or coral in each monomer (adapted from Zhang et al., 2018)

In the new Zavec et al. (2021) study it is shown that the native Adh2 enzyme is responsible for the majority of methanol oxidation in the Mut$^-$ *K. phaffii* strains, and that the overexpression of the *ADH2* gene in such strains increased the specific methanol uptake rate and enhanced recombinant protein production. The theoretical benefits of methanol oxidation by Adh instead of Aox in *K. phaffii* include reduced oxygen demand, reduced heat generation, and increased energy efficiency through the production of an additional molecule of NADH for each molecule of methanol consumed. The theoretical benefits of methanol oxidation by Adh2 were supported by experimental data by Zavec et al. (2021). The specific productivity ($q_P$) of the Mut$^-$ strain is increased by the overexpression of the *ADH2* gene and is in the range of the $q_P$ of Mut$^S$ strains. The specific methanol uptake rate ($q_{MeOH}$) was increased compared to the Mut$^-$ Adh2 KO strain, which was a success but remained lower than the $q_{MeOH}$ of the Mut$^S$ strain. Although the productivity is comparable to the Mut$^S$ strain, Mut- strain possesses the additional benefits of reduced oxygen demand and heat output. Additionally, higher viability and purity of the produced protein have been noted in the Mut$^-$ strain (Zavec et al., 2021).

Despite the evident advantages of using Mut- *ADH2* overexpression *K. phaffii* strains for methanol utilization, the lack of growth phenotype in the absence of the AOX genes in such

strains remains a major limitation for applications involving the production of compounds from methanol. It is theorized that the inability to accumulate biomass stems from the cytosolic localization of the Adh2 enzyme. In Mut$^-$ *ADH2* overexpression strains, formaldehyde is almost exclusively present in the cytosol and needs to diffuse to the peroxisome to enter the assimilatory pathway. It is considered that the increased formaldehyde concentration in the cytosol changes the assimilation-to-dissimilation ratio in favor of dissimilation. The methanol dissimilation ratio ($qCO_2/q_{MeOH}$) was experimentally proven higher in the Mut$^-$ $P_{AOX1}$ HAS strain compared to the Mut$^S$ strain. In Mut$^-$ strains, 80 % of methanol is dissimilated to $CO_2$, compared to the 100 % dissimilation rate in the Mut$^-$ strain with the overexpressed *ADH2* gene (Zavec et al., 2021).

## 2.8. SEMI-RATIONAL ENGINEERING OF THE Adh2 PROTEIN

Directed evolution has become one of the most widespread tools for optimizing biocatalysts for practical applications, and it has transformed the field of protein engineering (Wang et al., 2021; Lutz, 2010). It is accomplished by coupling effective methods for generating diverse libraries of protein variants with high-throughput screening processes to detect the improved phenotypes. The molecular diversity is achieved by random mutagenesis and *in vitro* recombination methods. In the past few decades, scientists have turned to a combination of directed evolution with elements of rational enzyme modification to create smaller and functionally-rich libraries (Lutz, 2010). Such well-designed libraries aim to reduce the screening scale and overcome the limitations of directed evolution and rational design (Tang et al., 2012; Chica et al., 2005). Experimental data has shown that most of the mutated enzymes with beneficial properties have mutations in or near the residues involved in binding or catalysis (Chica et al., 2005). The availability of enzyme sequences and structural and functional information in combination with computational simulations allows scientists to apply a semi-rational engineering approach and select the target residues with appealing characteristics (Lutz, 2010).

In this graduate thesis, a semi-rational approach was adopted for the engineering of the *K. phaffii* Adh2 enzyme. The scheme of the semi-rational enzyme engineering approach is depicted in Figure 6. Specific residues were individually mutated using site saturation mutagenesis to create a small, well-designed mutant library with carefully chosen target sites. The targets for the site saturation mutagenesis were selected based on the published crystal structure of the *K. phaffii* Adh2 protein (Zhang et al., 2018) and the use of computational tools

for molecular dynamics simulations of the substrates and products of the enzyme. Site saturation mutagenesis was performed using the overlap extension PCR with degenerate synthetic oligonucleotide primers for the purpose of fully randomizing the target sites.

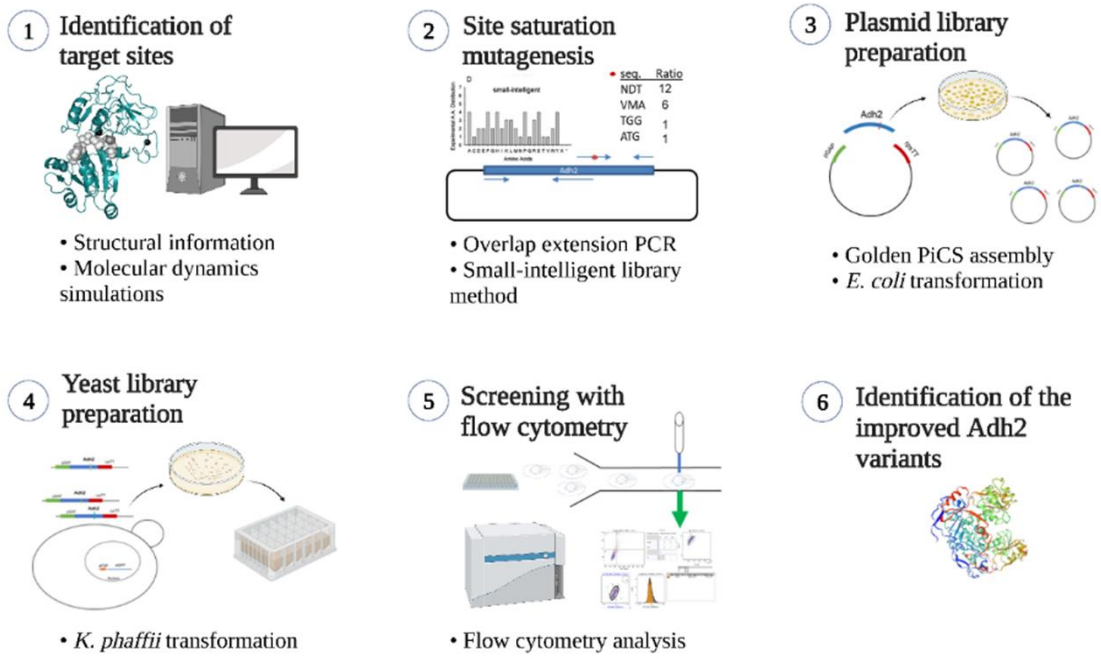



**Figure 6.** Scheme of the semi-rational approach for the evolution of the *K. phaffii* Adh2 enzyme (The figure was created with BioRender.com)

The mutated genes were cloned into the expression cassette using the state-of-the-art flexible modular cloning system for strain engineering in *K. phaffii*, Golden PiCS (Prielhofer et al., 2017). The plasmids containing the mutated *ADH2* gene were propagated in *E. coli* DH10B strain and expressed in *K. phaffii* CBS2612 biosensor screening strain. The plasmid isolation and sequencing was conducted on the bacterial and yeast level. The green fluorescent protein (eGfp) expressed by the biosensor strain enabled efficient screening of the strains carrying a mutation in the Adh2 enzyme. The screenings were performed using *in vivo* flow cytometry to analyze the performance of the enzyme variants in their native environment.

# 3. EXPERIMENTAL PART

## 3.1. WORKING STRAINS

### 3.1.1. *E. coli* strain

*Escherichia coli* K-12 strain DH10B, obtained from Invitrogen, was transformed with the constructed plasmids and used for plasmid propagation (Table 1).

**Table 1.** *Escherichia coli* strain used in this thesis

| Organism | Strain | Source |
|---|---|---|
| *Escherichia coli* | DH10B | Invitrogen |

### 3.1.2. *K. phaffii* strains

*Komagataella phaffii* strains used and generated in this thesis are listed in Table 2. The core strain used for the development of the biosensor strain was engineered and provided by Domen Zavec, Ph.D. The Mut$^-$ AdhKO strain was transformed with the biosensor plasmid, using the AOX integration site. The Mut$^-$ AdhKO biosensor strain was transformed with the plasmid containing the mutated *ADH2* gene, using the *RGI2* integration site. The Mut$^-$ AdhKO strain carries a resistance to hygromycin (Hyg) and kanamycin (G418), the Mut$^-$ AdhKO biosensor strain carries an additional resistance to Zeocin (Zeo), and the Mut- AdhKO biosensor Adh2 strain carries an additional resistance to Nourseothricin (NTC).

**Table 2.** *Komagataella phaffii* strains used and generated in this thesis

| Organism | Strain | Genotype | Phenotype | Designation | Source |
|---|---|---|---|---|---|
| *K. phaffii* | CBS2612 | Δaox1 Δaox2 Δadh2 Δadh900 | Mut$^-$ | AdhKO | Doctoral thesis of D. Zavec (Zavec, 2019) |
| *K. phaffii* | CBS2612 | Δaox1 Δaox2 Δadh2 Δadh900 pFld1_eGfp_CycTT | Mut$^-$ | AdhKO + biosensor | Generated in this thesis |
| *K. phaffii* | CBS2612 | Δaox1 Δaox2 Δadh2 Δadh900 pFld1_eGfp_CycTT pGAP_Adh2*_CycTT | Mut$^-$ | AdhKO + biosensor + Adh2 | Generated in this thesis |

*Single variant or mutant library

## 3.2. MATERIALS

### 3.2.1. Chemicals

The base chemicals used in this thesis are listed in Table 3.

**Table 3.** Base chemicals used in this thesis

| Chemical (abbreviation) | Manufacturer |
|---|---|
| 1 kb DNA Ladder | New England Biolabs (USA) |
| 1,4-Dithiothreitol (DTT) | Carl Roth (Germany) |
| Agar-Agar | Carl Roth (Germany) |
| CutSmart® Buffer | New England Biolabs (USA) |
| Gel Loading Dye Purple (6X) | New England Biolabs (USA) |
| Glacial acetic acid | Carl Roth (Germany) |
| Glucose | Agrana (Austria) |
| Glycerol | Carl Roth (Germany) |
| Hy-Soy Peptone | Quest (USA) |
| Lithium Chloride | Sigma-Aldrich (Germany) |
| Methanol ≥ 99.8 % (MeOH) | Carl Roth (Germany) |
| Midori Green Advance DNA Stain | Nippon Genetics (Germany) |
| Q5® High GC Enhancer | New England Biolabs (USA) |
| Q5® Reaction Buffer | New England Biolabs (USA) |
| Sodium hydroxide (NaOH) | Carl Roth (Germany) |
| D-Sorbitol | Carl Roth (Germany) |
| Sulfuric acid ($H_2SO_4$) | Merck (USA) |
| SYBR® Safe DNA Gel Stain | Invitrogen (USA) |
| Trisaminomethane (Tris) | Carl Roth (Germany) |
| Yeast Extract | Merck (USA) |

3.2.2. Buffers, media, and solutions

The buffers used in this thesis are listed in Table 4.

**Table 4.** Buffers used in this thesis

| Buffer | Components | pH |
|---|---|---|
| 1M Sorbitol | 182.2 g D-Sorbitol<br>Fill up with RO water to 1000.0 g | / |
| 1x PBS | 100.0 mL/L 10x PBS | 7.4 |
| 10x PBS | 2.4 g $KH_2PO_4$<br>18.0 g $Na_2HPO_4*2H_2O$<br>2.0 g KCl<br>80.0 g NaCl<br>Fill up with RO water to 1000.0 g | 6.8 - 6.9 |
| 50x TAE buffer | 242.0 g/L Tris Base<br>57.1 mL/L Glacial acetic acid<br>100.0 mL/L 0.5M EDTA solution (pH = 8)<br>Fill up with RO water to 1000.0 g | / |
| Running buffer | 1x TAE<br>5.0 µL/L Midori Green | / |

The media used in this thesis are listed in Table 5.

**Table 5.** Media used in this thesis

| Medium | Components | pH | Sterilization |
|---|---|---|---|
| 2x ASMØ.v6 | 12.60 g/L $(NH_4)_2HPO_4$<br>1.60 g/L $(NH_4)_2SO_4$<br>0.98 g/L $MgSO_4*7H_2O$<br>5.28 g/L KCl<br>0.11 g/L $CaCl_2*2H_2O$<br>44.0 g/L $C_6H_8O_7*H_2O$<br>2940.0 µl/L trace salts<br>40.0 mL/L $NH_4OH$ (25 %)<br>8.0 mL/L biotin (0.1 g/L) | Adjusted to 6.4-6.6 with solid KOH | Filtration (0.22 µm) |
| LB agar | 10.0 g/L hy soy peptone<br>5.0 g/L yeast extract<br>5.0 g/L NaCl<br>20.0 g/L Agar-Agar | Adjusted to 7.4-7.6 with 4N or 8N NaOH | Autoclaving for 20 minutes at 121 °C |
| LB medium | 10.0 g/L hy soy peptone<br>5.0 g/L yeast extract<br>5.0 g/L NaCl | Adjusted to 7.4-7.6 with 4N or 8N NaOH | Autoclaving for 20 minutes at 121 °C |
| YP agar | 20.0 g/L hy soy peptone<br>10.0 g/L yeast extract<br>20.0 g/L Agar-Agar<br>100.0 mL/L of 10x carbon source (glucose = YPD or glycerol = YPG) | Adjusted to 7.4-7.6 with 4N or 8N NaOH | Autoclaving for 20 minutes at 121 °C |

**Table 5.** Media used in this thesis (continuation)

| | | | |
|---|---|---|---|
| YP medium | 20.0 g/L hy soy peptone<br>10.0 g/L yeast extract<br>100.0 mL/L of 10x carbon source (glucose = YPD or glycerol = YPG) | Adjusted to 7.4-7.6 with 4N or 8N NaOH | Autoclaving for 20 minutes at 121 °C |

The solutions used in this thesis are listed in Table 6.

**Table 6.** Solutions used in this thesis

| Solution | Components | Sterilization |
|---|---|---|
| 10x Glucose | 220.0 g/L D-(+)-Glucose Monohydrate | Autoclaving for 20 minutes at 121 °C |
| 10x Glycerine | 126.0 g/L 100 % Glycerine | Autoclaving for 20 minutes at 121 °C |
| EnPump200 Polysaccharide Solution | 100.0 g/L EnPump 200 substrate (EnPresso) | Filtration (0.22 µm) |

3.2.3. Enzymes

The enzymes used in this thesis are listed in Table 7.

**Table 7.** Enzymes used in this thesis

| Enzyme | Manufacturer |
|---|---|
| AscI | New England Biolabs |
| BpiI | New England Biolabs |
| BsaI | New England Biolabs |
| EnPresso Reagent A | Sigma-Aldrich |
| T4 Ligase | New England Biolabs |
| Q5 Polymerase | New England Biolabs |

3.2.4. Antibiotics

The antibiotics used in this thesis are listed in Table 8.

**Table 8.** Antibiotics used in this thesis

| Antibiotic | Resistance | Stock conc. [mg/mL] | Working conc. *E. coli* [µg/mL] | Working conc. *K. phaffii* [µg/mL] |
|---|---|---|---|---|
| HYG | hphMX | 100 | 50 | 200 |

**Table 8.** Antibiotics used in this thesis (continuation)

| | | | | |
|---|---|---|---|---|
| NTC | natMX | 100 | 50 | 100 |
| ZEO | Sh ble | 100 | 25 | 25 |

### 3.3. SELECTION OF SITES FOR THE MUTAGENESIS

This thesis aimed to produce improved variants of the *ADH2* gene with enhanced activity toward methanol using a semi-rational engineering approach. The first step of the workflow was to identify and analyze the residues for site saturation mutagenesis. The structure of the Adh2 protein was initially analysed using PyMol, an open-source molecular visualization system. The European Bioinformatics Institute (EBI) Mechanism and Catalytic Site Atlas were used to examine the role of different catalytic residues in alcohol dehydrogenase I isolated from *Equus caballus*. The enzyme binds two zinc ions, but only one is active during the catalysis. Residues His51 and His47 are responsible for activating the catalytic function of Ser/Thr48, which is the deprotonation of the OH group of the alcohol substrate. It was concluded that the highly conserved residues directly involved in the catalysis and the zinc coordination weren't good initial target residues.

For an in-depth investigation of the potential target residues, molecular dynamics simulations were conducted using the GROningen MOlecular Simulation (GROMOS) software package and the published crystal structure of the Adh2 protein (Zhang et al., 2018). The molecular dynamics simulations were performed in collaboration with Enikö Hermann from the Institute of Molecular Modeling and Simulation (BOKU, Vienna). Firstly, a set of simulations was performed using the crystal structure of the Adh2 protein from the PDB database (5YAT) with the original orientation for dimerization. The first simulations with glycerol in the active site have shown the displaced orientation of the dimers. Based on the symmetry of the crystal structure, another possibility for dimerization was available. The alternative dimerization was closer to the crystal structures of Adh enzymes in other species (Figure 7).

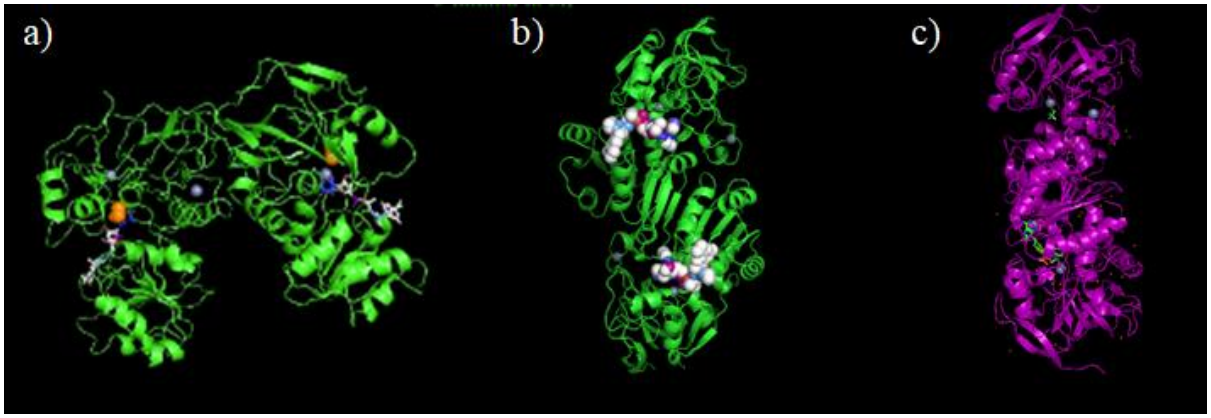



**Figure 7.** a) Dimerization in the original crystal structure of the *K. phaffii* Adh2 enzyme; b) Alternative dimerization model for *K. phaffii* Adh2 enzyme proposed by Enikö Hermann; c) Dimerization in the crystal structure of the *S. cerevisiae* Adh1 enzyme (all pictures were provided by Enikö Hermann from the Institute of Molecular Modeling and Simulation, BOKU Vienna)

For the $NAD^+$ binding residues, the alternative dimer orientation was used at 24 °C. Interactions between the substrate and the Adh2 residues were investigated. The trajectory analysis was performed using GROMOS, and the force field was 54a7. Simulations were performed 3x20 ns for MeOH and $NAD^+$ and 3x20 ns for formaldehyde and NADH. The enzymes are dimers, so the total for each state was 2x3x20 ns (120 ns). The electrostatic and van der Waals interactions indicated which amino acids were responsible for the binding in oxidized and reduced NAD states. A summary of the potential target residues given by the molecular dynamics simulations is depicted in Figure 8.

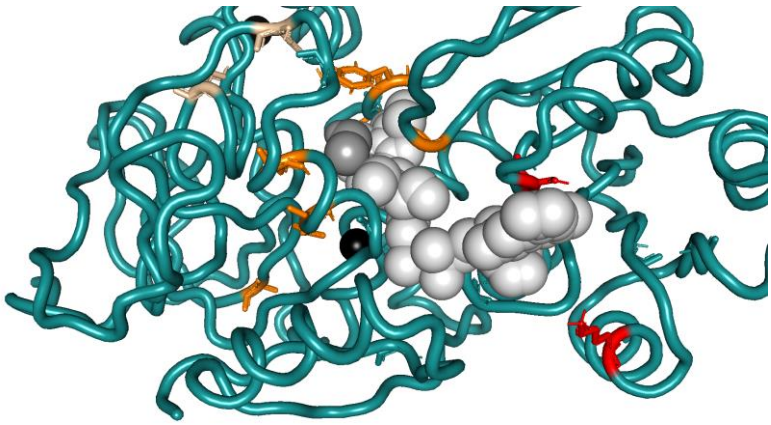


**Figure 8.** A summary of results by the molecular dynamics simulations performed using the GROMOS software package is depicted.

The residues involved in hydrogen bonding with formaldehyde and methanol are colored orange (A52, Y297, G272, D49, A72, G299); the residues that interact with $NAD^+$ and NADH are colored red (K209, S249); the residues located in other pocket targets are colored tan (T123, L59, L119)

Based on the performed visualizations and the previous research of the Adh2 structure, the residues selected as candidates for site saturation mutagenesis were Threonine 48, Alanine 52, Glycine 272, Leucine 59, Leucine 119, Serine 249, Threonine 123, and Tyrosine 297.

## 3.4. SITE SATURATION MUTAGENESIS

Site-saturation mutagenesis was employed in this thesis for the simple and effective generation of the novel Adh2 variants. Site-saturation mutagenesis was executed using the overlap extension PCR method with the help of degenerate synthetic oligonucleotide primers. The primers were synthesized based on the „small-intelligent“ library method (Tang et al., 2012), with the purpose of randomizing one target residue to allow for all 20 amino acids to be present in equal proportions with no stop codons.

The semi-rational approach was applied to design a mutant protein library of minimal size, with carefully chosen target residues, based on the molecular dynamics simulations generated by the previously described computational strategies. The simulations were made possible due to the availability of the high resolution of the crystal structure of the Adh2 protein, which allowed the exploration of the active site of the protein. The potential benefits of the semi-rational approach included a higher possibility of discovering notable mutants with increased activity towards methanol and reduced screening time. Overall, this approach was chosen to maximize the efficiency of the search for the optimized protein variants and to maximize the quality of the mutant library.

The site saturation mutagenesis screenings allowed us to thoroughly test the biosensor strain, which was used for the directed evolution of the Adh2 protein in the doctoral dissertation of Charles Moritz, Ph.D. student (Moritz, n.d.).

### 3.4.1. Primer construction (Small-Intelligent Library Method)

The applied method for the site saturation mutagenesis was the „small-intelligent“ focused library method developed by Tang et al. (2012). The method was chosen to avoid the usual problems associated with conventional mutagenesis methods, including amino acid biases and codon redundancy. The purpose of the small-intelligent method was to fully randomize one target site per mutagenesis cycle by using degenerate synthetic oligonucleotide primers.

Four primer solutions were needed for one mutagenesis site, including one universal forward primer, one newly synthesized reverse primer, a mixture of four newly synthesized

forward mutagenic primers, and one universal reverse primer. The mixture of four forward mutagenic primers included the NDT, VMA, ATG, and TAC codons in the target residue at a ratio of 12:6:1:1, respectively. Amino acid coverage of the used codons is shown in Table 9.

**Table 9.** Amino acid coverage of different codons used for the mutagenic primer design

| Codon | Amino acid | Ratio | Mixing volume (µL) |
|---|---|---|---|
| NDT | Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, Gly | 12 | 120 |
| VMA | Glu, Ala, Gln, Pro, Lys, Thr | 6 | 60 |
| ATG | Met | 1 | 10 |
| TAC | Trp | 1 | 10 |

The overlap region of the reverse and the forward primer for each target site was approximately 10 nucleotides long. When designing primers, the primer length was set to be between 18 and 25 nucleotides. The primers were designed to have melting temperatures within 5 °C of each other, and the GC content to be between 40-60 %. All the synthesized primers were diluted to a 10 µM working concentration. The design of the primers was done in the CLC Genomics Workbench program (Qiagen, Germany). All the primers that were used in the thesis are listed in Appendix, Supplementary Table 1. Primers MOC_36 and MOC_37 were used as end primers for the colony PCR method. Primers MOC_79 and MOC_80 are the forward and reverse universal end primers.

## 3.5. POLYMERASE CHAIN REACTION

Two types of polymerase chain reactions were used in the thesis: the overlap extension PCR for cloning purposes, and the colony PCR for the verification of the integration event. Both methods were conducted using the Bio-Rad C1000 Touch Thermal Cycler and the Q5® High-Fidelity DNA Polymerase (New England Biolabs, USA).

### 3.5.1. Overlap extension PCR

The overlap extension PCR cloning was used to generate DNA inserts that contain the *ADH2* gene with specific mutations at the site of interest. Four primer solutions were required, and three PCR reactions were performed while conducting the site saturation mutagenesis for each target residue.

The first PCR step included the synthesis of two overlapping DNA fragments in two separate reactions. For each PCR reaction, the master mix was prepared by combining the Q5 PCR Buffer, Q5 GC Enhancer, deoxynucleotide triphosphates (dNTPs), Q5 Polymerase, distilled water, and the DNA template (BB3rN pGAP_Adh2 DNA, concentration = 5 ng/ul) (Table 10). The universal forward Adh2 primer (MOC_79) and a newly synthesized non-mutagenic reverse primer, designed to target the upstream region of the specific site of interest, were added to the first PCR mixture tube. A mixture of four newly synthesized mutagenic degenerate forward primers, designed to target the residue of interest, was added to the second PCR mixture tube, along with the universal reverse Adh2 primer (MOC_80). The mixture of four mutagenic primer solutions was prepared using a ratio as previously described in Table 9.

The thermal cycling conditions used for the synthesis of the overlapping DNA fragments are shown in Table 11. At the end of the first PCR step, only the second of the two newly synthesized fragments contained an introduced mutation. The DNA fragments were separated using preparative gel electrophoresis to verify the correct length of the fragments. The fragments were then isolated from gels and purified using the standard “Mini elute” PCR clean-up protocol (innuPREP DOUBLEpure Kit, Analytik Jena). The concentrations of the purified fragments were determined (NanoDrop™ OneC, Thermo Scientific) and diluted with distilled water to reach the final concentration of 0.1 pmol/µl for the second PCR step.

**Table 10**. PCR mixture components for the first step of the overlap extension PCR

| **Reagent** | **Volume (µL)** |
|---|---|
| **Water** | 20.5 |
| Q5 PCR Buffer (5x) | 10.0 |
| Q5 GC Enhancer | 10.0 |
| dNTP (10 nM) | 1.0 |
| Q5 Polymerase (2U/µL) | 0.5 |
| Forward Primer | 2.5 |
| Reverse Primer | 2.5 |
| Template | 3 |
| | Final reaction volume = 50.0 |

**Table 11.** Thermal cycling conditions for the first step of the overlap extension PCR

| Step | Temperature (°C) | Duration | Number of cycles |
|---|---|---|---|
| Initial denaturation | 98 | 1 min | 1 |
| Denaturation | 98 | 10 s | 30 |
| Annealing | * | 30 s | |
| Elongation | 72 | * | |
| Final elongation | 72 | 2 min | 1 |

*Adjusted according to primers and template size

In the second PCR step, the purified DNA fragments were denatured and renatured, allowing the overlapping regions of the two fragments to bind complementary. The overlapping parts of the fragments acted as primers for the extension of the gene to the full length while introducing the mutations in the mutagenic forward primer. The master mix was prepared by combining the Q5 PCR Buffer, Q5 GC Enhancer, deoxynucleotide triphosphates (dNTPs), Q5 Polymerase, distilled water, and equimolar amounts of purified fragments to a total reaction volume of 45 µL (Table 12). The thermal cycling conditions for the second PCR step are given in Table 13. After the second PCR step, the universal forward Adh2 primer (MOC_79) and the universal reverse Adh2 primer (MOC_80) were added to the PCR tube. The purpose of the added outside primers was to amplify the whole mutated *ADH2* gene with the extensions for the Golden Gate assembly into the BB3rN backbone, which will be described in Section 3.8. The thermal cycling conditions for the third PCR step are listed in Table 14.

**Table 12.** PCR mixture components for the second and the third step of the overlap extension PCR

| Reagent | Volume (µL) | Note |
|---|---|---|
| Water | 20.5 | Second PCR step |
| Q5 PCR Buffer (5x) | 10.0 | |
| GC Enhancer | 10.0 | |
| dNTP (10 nM) | 1.0 | |
| Q5 Polymerase (2U/µL) | 0.5 | |
| Fragment 1 (1 pM) | 2.5* | |
| Fragment 2 (1 pM) | 2.5* | |
| Final reaction volume = 45.0 | | |
| Forward primer (MOC_79) | 2.5 | Third PCR step |
| Reverse primer (MOC_80) | 2.5 | |
| Final reaction volume = 50.0 | | |

*Purified PCR fragments were diluted with distilled water to reach a final concentration of 0.1 pmol/µl

**Table 13.** Thermal cycling conditions for the second step of the overlap extension PCR

| Step | Temperature (°C) | Duration | Number of cycles |
|---|---|---|---|
| Initial denaturation | 98 | 1 min | 1 |
| Denaturation | 98 | 10 s | 10 |
| Annealing | * | 30 s | |
| Elongation | 72 | 30 s | |
| Final elongation | 72 | 2 min | 1 |

*Adjusted according to the overlap region

**Table 14.** Thermal cycling conditions for the third step of the overlap extension PCR

| Step | Temperature (°C) | Duration | Number of cycles |
|---|---|---|---|
| **Initial denaturation** | 98 | 1 min | 1 |
| **Denaturation** | 98 | 10 s | 30 |
| **Annealing** | 66 | 30 s | |
| **Elongation** | 72 | 1 min | |
| **Final elongation** | 72 | 2 min | 1 |

3.5.2. Colony PCR

The colony PCR technique was used for the screening of yeast colonies that have grown on selective media to check the correct integration of the mutated *ADH2* gene in the *RGI2* site after the yeast transformation step. For each site saturation mutagenesis round, a minimum of 40 yeast colonies were picked for the colony PCR. Forty colonies were picked for sites A52, G272, L59, L119, and Y297, and 60 colonies were picked for sites T123 and S249. The isolated colonies were transferred into 30 µL of 0.2 M NaOH located in the PCR tubes using a sterile micropipette tip and briefly resuspended. The colonies were cooked for 10 minutes at 99 °C in the PCR thermal cycler and centrifuged afterward.

As templates for the second PCR reaction, DNA-containing supernatants were used. The master mix was prepared by combining all the other necessary reagents for a PCR reaction, including the Q5 DNA polymerase, deoxynucleotide triphosphates (dNTPs), forward and reverse primers, the Q5 PCR reaction buffer and distilled water (Table 15). The same primers (MOC_36 and MOC_37) were used to amplify the mutated *ADH2* gene for all the chosen mutagenesis sites. The thermal cycling conditions used for the colony PCR are shown in Table 16. The length of the amplified DNA fragments was expected to be around 1090 base pairs which is the length of the whole *ADH2* gene with the extensions for the Golden Gate assembly. The fragments were purified using the standard “Mini elute” PCR clean-up protocol

(innuPREP DOUBLEpure Kit, Analytik Jena) and sent for sequencing to analyze the incidence of different amino acids in the mutated site.

**Table 15.** Colony PCR mixture components for a total volume of 50 µL

| Reagent | Volume (µL) |
|---|---|
| Water | 30.5 |
| Q5 PCR Buffer (5x) | 10.0 |
| dNTP (10 nM) | 1.0 |
| Q5 Polymerase (2U/µL) | 0.5 |
| Forward Primer (MOC_36) | 2.5 |
| Reverse Primer (MOC_37) | 2.5 |
| Template | 3.0 |
| **Final reaction volume (µL)** | 50.0 |

**Table 16.** Colony PCR thermal cycling conditions

| Step | Temperature (°C) | Duration | Number of cycles |
|---|---|---|---|
| **Initial denaturation** | 98 | 1 min | 1 |
| **Denaturation** | 98 | 10 s | 30 |
| **Annealing** | 62 | 30 s | |
| **Elongation** | 72 | 40 s | |
| **Final elongation** | 72 | 2 min | 1 |

### 3.6. GEL ELECTROPHORESIS

Agarose gel electrophoresis was used to visualize and analyze the amplified PCR products. Components that are needed for the preparation of the 1.5 % agarose gels are shown in Table 17. Midori Green Advance DNA stain (Nippon Genetics Europe, Germany) was added to the analytical gels, and SYBR Safe DNA stain (Invitrogen, USA) was used for the preparative gels. The gel electrophoresis was performed at 130 V for 35 minutes. DNA fragments were visualized using the ChemiDoc Chemiluminescence Gel Imaging System (Bio-Rad Laboratories, USA). After the preparatory gel electrophoresis, DNA excision was done on a transilluminator.

**Table 17.** Agarose gel electrophoresis components

| Gel | Components |
|---|---|
| 1.5 % Agarose gel | 5.4 g Agarose<br>7.2 g 50x TAE<br>Fill up to 360.0 g with RO water<br>1.8 µL Midori Green (analytical gel) or 21 µL SYBR Safe (preparative gel) |

## 3.7. DNA EXTRACTION AND CONCENTRATION MEASUREMENT

DNA fragments generated by the overlap extension PCR method were purified after the preparative gel electrophoresis, using the standard protocol for DNA extraction from agarose gel slices (innuPREP DOUBLEpure Kit, Analytik Jena). Purification of the PCR products after the colony PCR was done using the "Mini Elute" protocol for purification and concentration of PCR products from PCR reactions (innuPREP DOUBLEpure Kit, Analytik Jena). The same protocol was used for the purification of plasmids after the linearization, to remove the enzyme from the solution. The plasmids in this thesis were isolated using the Mini Plasmid DNA Extraction Kit (HiYield® Plasmid Mini Kit, Süd-Laborbedarf, Germany). The concentrations of all the DNA molecules generated in this thesis were determined using the NanoDrop™ OneC (Thermo Scientific, USA).

## 3.8. GOLDEN PICS CLONING

The construction of the BB3rN pGap_Adh2*_rps25TT (Table 18) expression cassette was conducted using the highly efficient modular system for strain engineering of the yeast *K. phaffii*, Golden PiCS. It belongs to a universal cloning system GoldenMOCS (Golden Gate-derived Multiple Organism Cloning System). The cloning system consists of three hierarchical backbone (BB) levels. The lowest level, BB1, refers to the basic modules of the future expression cassette, and it includes promoters, coding sequences, and terminators. The basic modules are assembled into one transcription unit in the BB2 level. At the BB3 level, multiple genetic constructs are ligated into one expression vector (Prielhofer et al., 2017). Golden PiCS allows the simultaneous expression of up to eight expression units while using different standardized genetic parts (promoters, terminators, genome integration loci, and resistance marker cassettes). The plasmids used in this thesis and their standardized genetic building units (promoters, terminators, coding sequences) are listed in Table 18.

The Golden PiCS system was used for the simple and efficient *in vitro* assembly of our gene of interest directly into a specially designed BB3 vector, performing the digestion with the BsaI restriction enzyme. The gene of interest in this thesis was the mutated *ADH2* gene, and it was inserted into the vector using the extensions that contained the BsaI restriction sites needed for the Golden PiCS assembly (Figure 9).

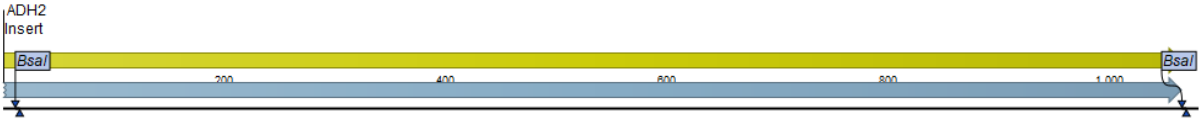



**Figure 9.** The *ADH2* gene with the extensions for the Golden Gate assembly into the BB3rN backbone

The backbone for the expression cassette creation was the BB3rN pGap_CloningSpacer_rps25TT plasmid with the NTC resistance marker (Figure 10). The BB3rN pGap_CloningSpacer_rps25TT plasmid contained the non-coding spacer DNA, which was removed during the Golden PiCS assembly, and replaced with the gene of interest (Figure 11). The integration site for the BB3rN pGap_Adh2*_rps25TT expression cassette in *K. phaffii* was the *RGI2* locus.

**Table 18.** Plasmids used in this thesis

| Plasmid | Backbone | Description |
|---|---|---|
| pFld1_eGfp_CycTT | BB3aZ (*AOX1*, ZEO) | Formaldehyde dehydrogenase 1 promoter, Green fluorescent protein gene, Cytochrome C terminator |
| pGap_CloningSpacer_rps25TT | BB3rN (*RGI2*, NTC) | Glyceraldehyde-3-phosphate dehydrogenase promoter, Cloning spacer DNA, Ribosomal Protein S25 terminator |
| pGap_Adh2*_rps25TT | BB3rN (*RGI2*, NTC) | Glyceraldehyde-3-phosphate dehydrogenase promoter, mutated Alcohol dehydrogenase 2 gene, Ribosomal Protein S25 terminator |

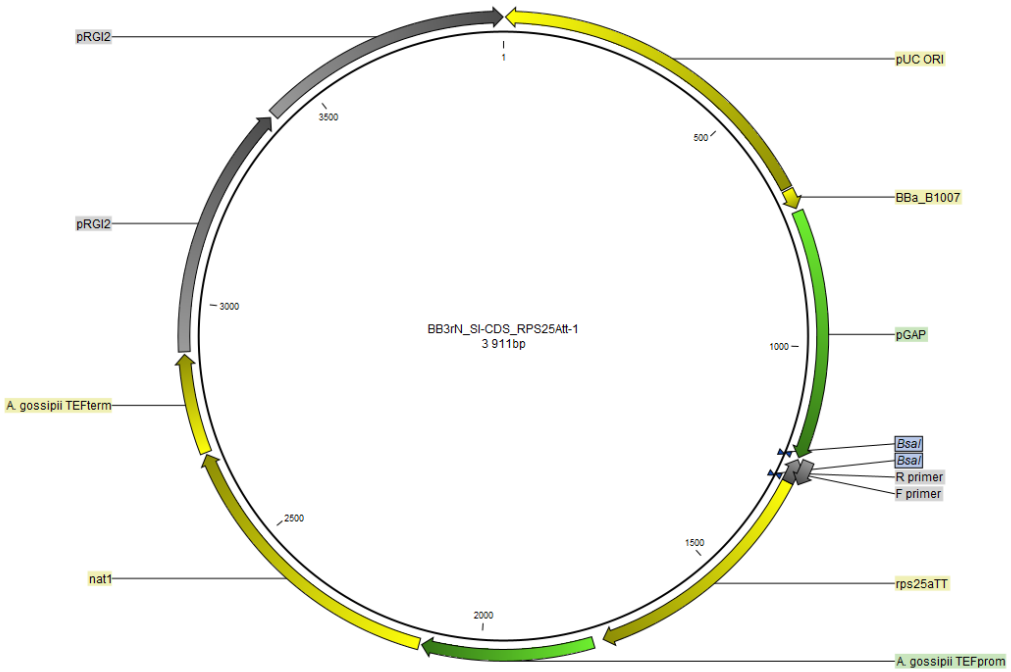



**Figure 10.** Illustration of BB3rN pGap_CloningSpacer_rps25TT plasmid (3911 bp).

The plasmid carries the following components: *GAP* promoter, non-coding spacer DNA, rps25 terminator, *A. gossipii TEF* promoter, *NAT1* gene, *A. gossipii TEF* terminator, *RGI2* promoter, and pUC origin of replication

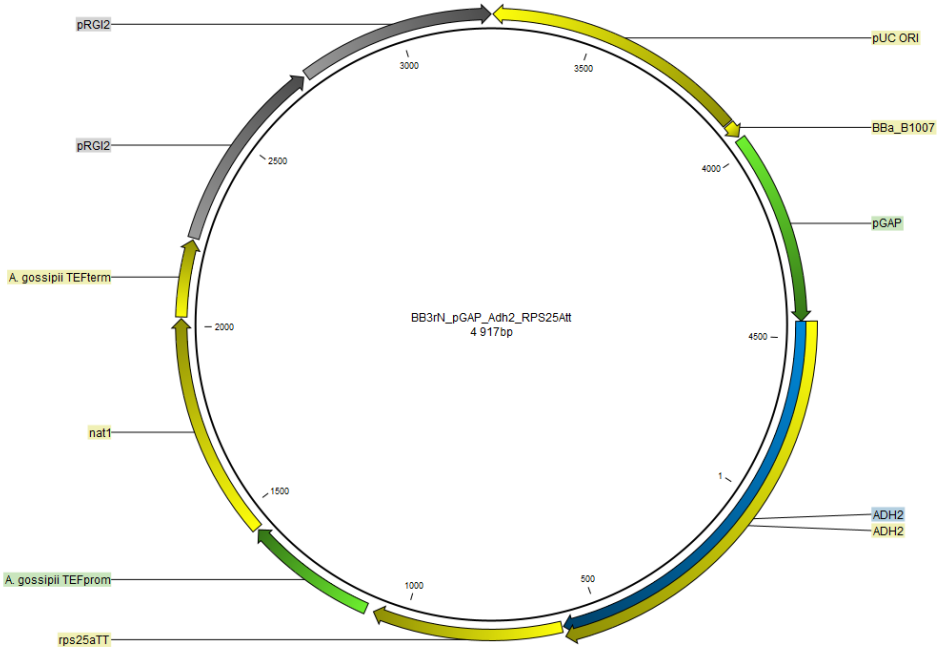



**Figure 11.** Illustration of BB3rN pGap_Adh2*_rps25TT plasmid (4917 bp).

The plasmid carries the following components: *GAP* promoter, mutated *ADH2* gene, rps25 terminator, *A. gossipii TEF* promoter, *NAT1* gene, *A. gossipii TEF* terminator, *RGI2* promoter, and pUC origin of replication

The components and the thermal cycling conditions for the Golden Gate assembly of the BB3rN pGap_Adh2*_rps25TT expression cassette are listed in Table 19 and Table 20, respectively. A larger volume of the mutated *ADH2* gene is added to the mixture in aspect to the BB3rN pGap_CloningSpacer_rps25TT plasmid to increase the probability of a successful insertion event.

**Table 19.** Components for the Golden Gate assembly of the BB3rN pGap_Adh2*_rps25TT expression cassette

| Component | Concentration | Volume (µL) |
|---|---|---|
| Backbone (BB3rN pGap_CDS_rps25TT) | 40 nM | 1 |
| Insert (mutated *ADH2* gene) | 40 nM | 3 |
| CutSmart™ Buffer | 10x | 2 |
| Aliquoted ATP | 10mM | 2 |
| T4 Ligase | 400,000 U/mL | 1 |
| BsaI | 20 U | 1 |
| Distilled water | / | 10 |

**Table 20.** Thermal cycling conditions for the Golden Gate assembly of the BB3rN pGap_Adh2*_rps25TT expression cassette

| Temperature (°C) | Duration | Number of cycles |
|---|---|---|
| 37 | 2 min | 50 |
| 16 | 5 min | |
| 37 | 10 min | 1 |
| 55 | 30 min | 1 |
| 80 | 10 min | 1 |

## 3.9. PLASMID LIBRARY PREPARATION

### 3.9.1. Chemically competent *E. coli* cell preparation

Competent *E. coli* DH10B cells were prepared by the lab technicians using the standard protocol for the preparation of chemically competent *E. coli* cells with RbCl, developed by Stefanie Müller in 2013 (Appendix, Supplementary Figure 1).

### 3.9.2. *E.coli* transformation

*E. coli* cells were transformed with the BB3rN pGap_Adh2*_rps25TT plasmid for plasmid propagation purposes. For each mutagenesis site, 100 µL of chemically competent *E.*

*coli* cells were thawed on ice. The total volume (20 µL) of the Golden Gate reaction mixture was quickly added to the competent cells, and the transformation mixture was incubated on ice for 30 minutes. Next, the transformation mixture was heat shocked for 90 seconds at 42 °C. The cells recovered on ice for 5 minutes and 1 mL of LB medium was added to the mixture, followed by a 1-hour incubation time in a shaker-incubator (Eppendorf, Thermomixer C) set to 37 °C and 450 rpm. The transformed cells were centrifuged for 1 minute at 11000 RCF. A part of the supernatant (600 µL) was removed and the transformed cells in the pellet were resuspended. For each mutagenesis site, four plates containing the selective LB medium (8.33 µg/mL NTC) were plated with 100 µL of the resuspended cells. The plated cells were incubated overnight at 37 °C or for two days at 25 °C.

3.9.3. Plasmid sequencing

The *E. coli* transformants were counted after the incubation period, and 10 colonies were selected for sequencing for each mutagenesis site. The precultures were started by transferring the marked colonies with a sterile toothpick into the glass culture tubes containing 4 mL of LB with 25 µg/mL NTC. The preculture glass tubes were incubated overnight at 37 °C and 180 rpm. The next day, 1.5 mL of the culture from each tube was taken for plasmid sequencing purposes, and 0.8 mL of the culture from each tube was taken for the cryopreservation of the transformed cells.

The amplified plasmids were purified from the *E. coli* cells using the Hi Yield® Plasmid Mini Kit (Süd-Laborbedarf, Germany). The concentration of the isolated plasmids was measured using the NanoDrop™ OneC (Thermo Scientific, USA). The plasmid DNA and the primers for sequencing reactions were dissolved in distilled water. For each sequencing reaction, 2 µl of the plasmid DNA and 3 µl of the primer (10 mM) were dissolved in 10 µl of distilled water. The final concentration of the plasmid DNA was 40-100 ng/µL. For each mutagenesis site, a primer that binds to the DNA region no less than 100 nucleotides away from the mutation site was chosen. The sequencing reaction tubes containing the plasmid DNA were sent for Sanger sequencing (Microsynth AG, Switzerland).

3.9.4. *E. coli* cryopreservation

*E. coli* transformants from the overnight cultures were cryopreserved in Thorbi® Kryo vials (National Lab, Germany). The vials were filled with 0.2 µl of 50 % glycerol solution, and

0.8 µl of the culture was pipetted into each solution. The vials were stored at a temperature of -80 °C.

3.9.5. Plasmid pooling and isolation

Plasmid pooling was performed to create a diverse plasmid library for yeast transformation. For each mutagenesis site, three plates that weren't used for sequencing were pooled. The number of *E. coli* colonies on the plates for pooling ranged from 100 to 680. Two mL of LB was pipetted onto a plate, and the colonies were stirred with a plastic cell spreader. The suspended cells were pipetted into the Eppendorf tubes and the plasmids were isolated using the Hi Yield® Plasmid Mini Kit (Süd-Laborbedarf, Germany). The concentration of the isolated plasmids was measured using the NanoDrop™ OneC (Thermo Scientific, USA). An appropriate amount of the isolated plasmids was used for the yeast transformation.

## 3.10. YEAST LIBRARY PREPARATION

The Mut$^-$ AdhKO biosensor strain (Δaox1 Δaox2 Δadh2 Δadh900 pFld1_eGfp_CycTT) was used for the yeast library preparation. The biosensor yeast cells were made competent for the successful integration of the BB3rN pGap_Adh2*_rps25TT plasmid.

3.10.1. Electrocompetent *K. phaffii* cell preparation

Electrocompetent biosensor yeast cells were prepared using the small-scale protocol for the preparation of electrocompetent *Pichia pastoris* cells, developed by Richard Zahrl in 2016 (Appendix, Supplementary Figure 2).

3.10.2. Plasmid linearization

The BB3rN pGap_Adh2*_rps25TT plasmid was linearized for the successful integration in the yeast genome via homologous recombination. A single or multiple integration events can occur in the yeast *K. phaffii*. Plasmids were linearized in a 46 µL reaction, with 40 µL of the plasmid DNA, 2 µL of the AscI enzyme, and 4 µL of CutSmart® Buffer, for 2 hours at 37 °C. The “Mini Elute” protocol for purification and concentration of PCR products from PCR reactions (innuPREP DOUBLEpure Kit, Analytik Jena) was performed to remove the enzyme from the reaction tube. The concentration of the linearized plasmid was determined using the NanoDrop™ OneC (Thermo Scientific, USA).

### 3.10.3. *K. phaffii* transformation

*K. phaffii* cells were transformed with the BB3rN pGap_Adh2*_rps25TT plasmid to build a yeast library. The plasmid linearization with the AscI enzyme encouraged the integration of the plasmid into the *RGI2* region of the *K. phaffii* genome. An 80 µL aliquot of electrocompetent *K. phaffii* cells was gently mixed with 20 µL of the linearized plasmid DNA. The wanted concentration of the plasmid DNA was 150 ng/µL. The yeast cells without the plasmid DNA were used as a negative control. The transformation mixture was transferred into a pre-chilled electroporation cuvette and incubated on ice for 5 minutes.

The electroporation was performed using the MicroPulser Electroporator (Bio-Rad Laboratories, USA) at 1.5 kV and four milliseconds. After the electroporation, the cells were resuspended in 1 mL of pre-chilled YPD media and transferred the mixture into a sterile microcentrifuge tube. The cells were regenerated for 2-2.5 hours at 28 °C and 180 rpm (Thermomixer C, Eppendorf) and plated on the selective agar plates in two different concentrations. The selective plates were prepared by melting 450 mL of YP agar and adding 50 mL of 10x glucose, 1 mL of HYG, 500 µL of NTC, and 125 µL of ZEO. For the first concentration, 100 µL of the transformation mixture was plated. The remaining volume was centrifuged for 1 minute at 11000 rpm, the supernatant was removed, and the remaining 100 µL of the transformation mixture was resuspended and plated. The plated cells were incubated for three days at 30 °C. After the first incubation period, a minimum of 40 large colonies were restreaked on the selective plates and incubated for three days at 30 °C.

### 3.10.4. Plasmid sequencing

The previously described colony PCR method was applied after the transformation to verify that the integration event was successful. The quality of a few PCR samples was checked using agarose gel electrophoresis. Twelve µL of the unpurified PCR amplicons was pipetted into the wells of a 96-well plate for sequencing. The sequencing plate was sent for PCR purification and Sanger sequencing (Microsynth AG, Switzerland). An appropriate amount of a primer (3 µL per sequencing reaction), specific for each mutagenesis site, was added to a sterile tube and sent in addition to the plate.

## 3.11. SCREENING WITH FLOW CYTOMETRY

### 3.11.1. Screening protocol

The purpose of the screening was to analyze the activity of the *ADH2* gene *in vivo*. The screening protocol was developed for the site saturation mutagenesis of the *ADH2* gene in the screening strain CBS2612 aZ_pFld1_eGfp_CycTT + rN_pGap_Adh2_rps25TT (Δaox1, Δaox2, Δadh2, Δadh9000). For each site saturation mutagenesis cycle, three biological replicates were screened for every identified amino acid at the mutation site. The first step of the screening process was identifying the isolates for the screening protocol and calculating the needed volume of the preculture media (Table 21). The untransformed Mut$^-$ AdhKO biosensor strain was used as the negative control. For positive control, the yeast clones from Sanger sequencing that displayed the wild-type amino acid sequence were used. The preculture medium (2.5 mL) was dispensed into the appropriate wells of the 24 deep-well cultivation plates and inoculated the medium with whole isolated colonies taken from freshly restreaked plates. The precultures were started by placing the 24 deep-well plates covered with VWR® Rayon Film (WVR, USA) on a shaker at 280 rpm and 25 °C for 24 hours.

**Table 21.** Preculture media preparation for the screening

| Sample | Number of samples | YPG (mL) | NTC (µL) | ZEO (µL) | HYG (µL) |
|---|---|---|---|---|---|
| **Preculture** | 40 | 100 | 50 | 25 | |
| **Negative control** | 3 | 7.5 | | 1.875 | 15 |

Before starting the cultivation of the main cultures, the precultures were cryopreserved by mixing 400 µL of the preculture with 100 µL of 50 % glycerol in Thorbi® Kryo vials (National Lab, Germany) and stored the vials at a temperature of -80 °C. It was assessed if the cultures showed growth and negative controls didn't grow and the starting time of the main culture cultivation was adjusted 24 hours apart from the beginning of the screening time. The media for the main cultures was prepared two hours before the start of washing. The EnPump200 kit (Enpresso, Germany) ensured glucose-limited cultivation of the yeast by implementing the enzymatic hydrolysis of the polysaccharide solution in 12.5 g/L polysaccharide and 0.4 % enzyme conditions. Glucose release rates were determined by the final concentration of the EnPresso Reagent A enzyme. The calculation for the main culture media preparation is shown in Table 22.

**Table 22.** Main culture media preparation for the screening

| | Stock conc. [g/L] | Final conc. [g/L] | Volume (mL) |
|---|---|---|---|
| **EnPump200 Polysaccharide Solution** | 100 | 25 | 15 |
| **EnPresso Reagent A** | 100 | 0.008 | 0.4 |
| **Distilled H2O** | | | 34.6 |
| **Total volume** | | | 50 |

*EnPresso Reagent A was added just before the main culture cultivation

The washing of the precultures started two hours before the start of the main culture cultivation. The weight of the preculture plates was balanced, and the plates were centrifuged at 2000 g for 5 minutes. The media was aspirated and resuspended in 1 mL of double concentrated (2x) ASMØ.v6, and the plate was put on a shaker set at 280 rpm for 2-5 minutes. The whole washing step was repeated once more. The optical density of the cells was measured at 600 nm in a 96-well microtiter plate. Distilled water was added into each well, and 1:11 and 1:121 dilutions were made by sequentially transferring 30 µL of the culture with a multichannel pipette into the microtiter plate and then into the adjacent well. The measured OD values were used to calculate the inoculation volumes of the washed cells into the main culture 24 deep-well plates. The media for the main culture plates was prepared by mixing 1 mL of 2x ASMØ.v6 and 1 mL of the main culture media with the added enzyme. The calculated volume was removed from the main culture plate and inoculated it with the same preculture volume. The main culture plates were put on a shaker set at 280 rpm. The cultures were fed 1 % methanol (20 µL) after 3 hours of growth at glucose-limited conditions and the second time 5 hours before the screening.

The main culture plates were taken off the shaker two hours before the screening with the flow cytometer. The washing was performed twice for the main cultures, using 1x PBS for resuspension. The optical density of a few randomly selected cultures was measured to estimate the dilution necessary to reach the optical density of 0.5-1 needed for screening. The optical density was measured by pipetting 10 µL of the culture into 990 µL of 1x PBS in the cuvette. All wells in the microtiter plate were filled with 1x PBS, and the calculated volume of the culture was serially transferred from the 24 deep-well plate into the first dilution and then the second dilution well. The screening of the yeast colonies was performed using the CytoFLEX S Flow Cytometer (Beckman Coulter, USA). The sample injection mode for microtiter plates

was chosen, and the cleaning panel was run before the experiment. A new experiment was created on the device, the plate was inserted, and the sample wells were defined. The flow cytometry protocol created by Domen Zavec (Zavec, 2019) was applied. All of the screening data were analysed in the same way by opening batch analysis with Kaluza Analysis Software (Beckman Coulter, USA). The FX values represented the dimensionless mean relative cell-associated product and were calculated based on the equation FX = (FL1/FS1,5) x 8000 on a cell-to-cell basis and then averaged for the whole measured population. The number of events measured was 20,000, the excitation wavelength was 488 nm, and the fluorescence was measured at 505-545 nm (Zavec, 2019).

# 4. RESULTS AND DISCUSSION

The purpose of this graduate thesis was to engineer and evolve the Adh2 protein in *K. phaffii* for its increased activity towards methanol, enabling its growth on methanol with improved energy efficiency. To achieve the set goal, a semi-rational approach was applied to fully randomize the chosen target residues (Alanine 52, Glycine 272, Leucine 59, Leucine 119, Serine 249, Threonine 123, Tyrosine 297) of the protein.

The mutated gene variants were incorporated into the BB3rN pGap_CloningSpacer_rps25TT vector using the Golden PiCS modular cloning system, and a plasmid library was created in the *E. coli* strain DH10B (Invitrogen, USA) for each mutagenesis site. The mutated variants of the *ADH2* gene were introduced into the *K. phaffii* biosensor screening strain, and the plasmids were sequenced on the bacterial and yeast level with the aim of analyzing the diversity of the constructed libraries. Finally, the *K. phaffii* mutants were screened *in vivo* using flow cytometry, and the screening data were analysed using the Kaluza Analysis Software.

## 4.1. SITE SATURATION MUTAGENESIS OF THE *ADH2* GENE

To generate the novel Adh2 variants, site saturation mutagenesis of the chosen target sites (T48, A52, G272, L59, L119, S249, T123, Y297) was employed. The site-directed mutagenesis was achieved using the overlap extension PCR method with the help of degenerate synthetic oligonucleotide primers, as described in Section 3.5.1. The amplification of the correctly-sized fragments was confirmed by preparatory gel electrophoresis on 1.5 % agarose gels. The DNA fragments were visualized using the ChemiDoc Chemiluminescence Gel Imaging System (Bio-Rad Laboratories, USA).

The primers used for the mutagenesis of each target site are listed in the Appendix, Supplementary Table 1. Two separate cycles of site saturation mutagenesis were performed on the first tested target site, G272, with the goal of optimizing the site saturation mutagenesis protocol. In the first mutagenesis cycle, different annealing temperatures and primer ratios were used to investigate the optimal annealing temperature and to examine if the NDT primer alone would give altered amino acid ratios. Figure 12 shows the agarose gel after the first PCR reaction. Two forward and reverse outside primer pairs were used in the first mutagenesis cycle to test the efficiency of different annealing temperatures. For the first PCR reaction with the *K. phaffii* Adh2 universal forward and reverse primers (MOC_79 and MOC_80), the annealing

temperature suggested by the NEB Tm Calculator (New England Biolabs, 2022) was 66 °C. The annealing temperature proposed by the NEB Tm Calculator (New England Biolabs, 2022) using the redesigned universal forward and reverse primers for the higher Tm (MOC_141 and MOC_142) was 70 °C. For the 66 °C annealing temperature PCR reaction, the G272 reverse primer designed for lower Tm (MOC_139) was added, and for the 70 °C annealing temperature, the G272 reverse primer designed for higher Tm (MOC_140) was added. Fragments 4 and 7 were synthesized using only the NDT forward primer to verify if the NDT and the fully randomized colonies display the same codon bias, which would rule out primer mixing as a source of error. All the bands were of the expected size (814 and 273 bp). The higher annealing temperature (70 °C) resulted in fainter bands for the second fragment than the lower annealing temperature (66 °C). We can conclude that the higher annealing temperature prevented the optimal binding of the primers to the DNA template. We used the universal forward and reverse primers (MOC_79 and MOC_80) for the rest of the site saturation mutagenesis cycles. Fragments 3, 4, 6, and 7 contained a newly introduced mutation at the target site (G272) and shared an overlapping region with the corresponding fragments 2 and 5.

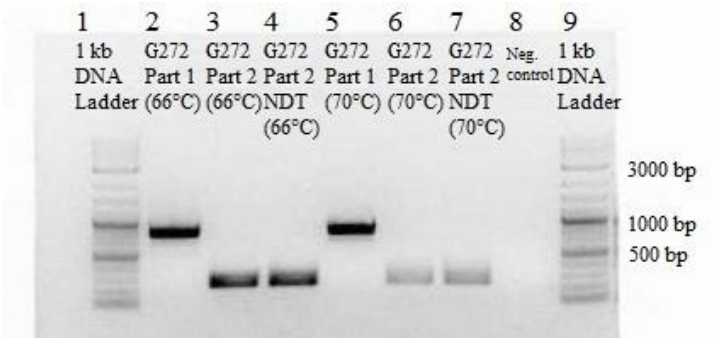



**Figure 12.** Agarose gel (1.5 %) electrophoresis results visualized after the first PCR reaction of the site saturation mutagenesis on target site G272.

Wells 1 and 9 are the 1 kb DNA Ladder (New England Biolabs, USA), wells 2, 3, 5 and 6 contain two pairs (2+3, 5+6) of overlapping DNA fragments, wells 4 and 7 contain the second fragment of the *ADH2* gene with the NDT codon in the mutation site, and well 8 is negative control (PCR master mix without the template DNA). The temperature indicated for each sample indicates the annealing temperature for the reaction

Figure 13 shows the agarose gel after the first PCR reaction of the second site saturation mutagenesis cycle on sites A52, G272, and Y297. For this cycle of site saturation mutagenesis, two different annealing temperatures were tested for target sites A52 and Y297. It was observed that the lower annealing temperature (62 °C) and the higher annealing temperature (66 °C) produced bands of similar brightness, but the lower annealing temperature gave slightly brighter bands. Fragments 2, 4, 6, 8, and 10 contained a newly introduced mutation at the target

site (G272, A52, or Y297) and shared an overlapping region with the corresponding fragments 1, 3, 5, 7, and 9.

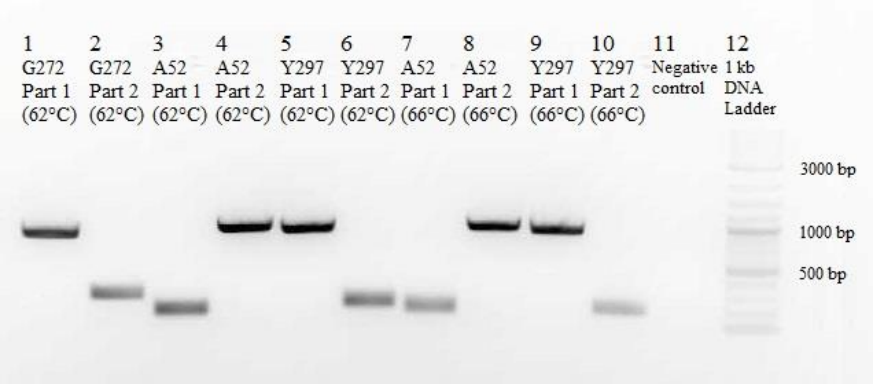



**Figure 13.** Agarose gel (1.5 %) electrophoresis results visualized after the first PCR reaction of the site saturation mutagenesis on target sites G272, A52, and Y297.

Wells 1-10 contain five pairs (1+2, 3+4, 5+6, 7+8, 9+10) of overlapping DNA fragments, well 11 is negative control (PCR master mix without the template DNA), and well 12 is the 1 kb DNA Ladder (New England Biolabs, USA). The temperature indicated for each sample indicates the annealing temperature for the reaction

The expected lengths of the fragments for G272, A52, and Y297 were 825 and 273 bp, 165 and 930 bp, and 900 and 196 bp, respectively, which correlated with the experimental length.

The agarose gel after the first PCR reaction of the third site saturation mutagenesis cycle on sites T123, S249, and T48 is shown in Figure 14. Given the slightly brighter bands produced by lower annealing temperature in the last mutagenesis cycle, approximately 4 °C lower annealing temperature than the annealing temperature proposed by the NEB Tm Calculator (New England Biolabs, 2022) was used. Fragments 3 and 6 contained a newly introduced mutation at the target site (T123 or S249) and shared an overlapping region with the corresponding fragments 2 and 4. The first fragment of the *ADH2* gene for the mutagenesis of site T48 didn't amplify. The second fragment for the mutagenesis of site T48 was synthesized using a forward primer with a codon for alanine. The mutagenesis was repeated in the fourth mutagenesis cycle to test if the fragment didn't amplify due to an issue while performing the PCR reaction or a poorly designed primer. The expected lengths of the fragments T123, S249, and T48 were 378 and 716 bp, 756 and 340 bp, and 153 and 939 bp, respectively.

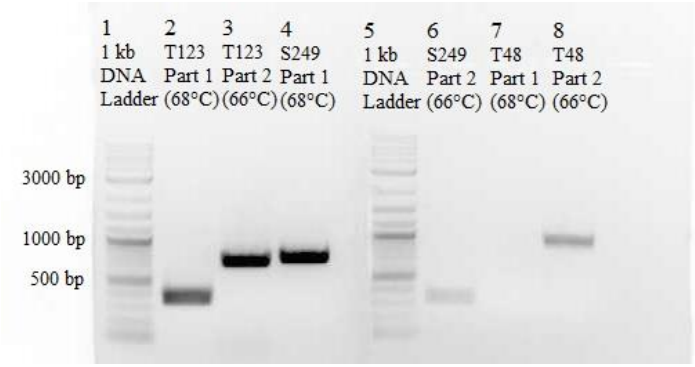



**Figure 14.** Agarose gel (1.5 %) electrophoresis results visualized after the first PCR reaction of the site saturation mutagenesis on target sites T123, S249, and T48.

Wells 1 and 5 are the 1 kb DNA Ladder (New England Biolabs, USA), and wells 2-4 and 6-8 contain three pairs of overlapping DNA fragments. All fragments were successfully synthesized and visualized, except for the first fragment of the *ADH2* gene for site T48. The temperature indicated for each sample indicates the annealing temperature for the reaction

Figure 15 shows the agarose gel after the first PCR reaction of the fourth site saturation mutagenesis cycle on sites L59, L119, and T48. The overlap extension PCR was repeated on site T48, and a lower annealing temperature was used in regard to the last mutagenesis cycle. The first fragment of the *ADH2* gene for the mutagenesis of site T48 didn't amplify, which indicated that the problem could be in the primer design, or that the fragment length was too short for the fragment to amplify successfully.

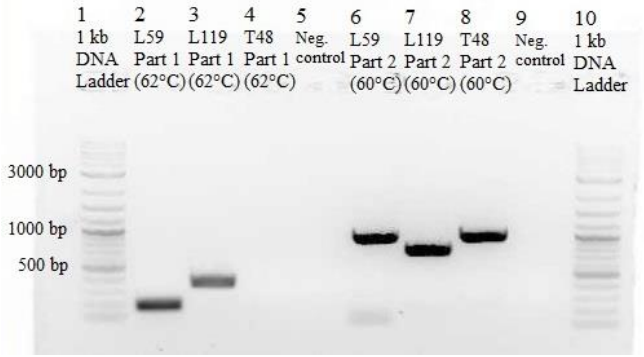



**Figure 15.** Agarose gel (1.5 %) electrophoresis results visualized after the first PCR reaction of the site saturation mutagenesis on target sites L59, L119, and T48.

Wells 1 and 10 are the 1 kb DNA Ladder (New England Biolabs, USA), and wells 2, 3, 6 and 7 contain two pairs of overlapping DNA fragments (2+6, 3+7). All fragments were successfully synthesized and visualized. The temperature indicated for each sample indicates the annealing temperature for the reaction

The purpose of mutating site T48 was to confirm its central role in the proton relay system and to use it as a control but decided to continue with other sites since the negative control was already available. Fragments 6 and 7 contained a newly introduced mutation at the

target site (L59 or L119) and shared an overlapping region with the corresponding fragments 2 and 3. The expected lengths of the fragments L59 and L119 were 186 and 904 bp, and 366 and 724 bp, respectively.

All the generated PCR fragments were excised from the gel and purified, the concentration was determined, and the fragments were diluted with distilled water to reach a 0.1 pmol/µl concentration for the second PCR step. The purpose of the second PCR reaction was to anneal the overlapping fragment pairs and to generate the whole mutated *ADH2* gene as a template. The third PCR reaction was performed with the universal forward (MOC_79) and the reverse primer (MOC_80) to amplify the mutated *ADH2* gene with the extensions for the Golden Gate assembly into the BB3rN backbone. The DNA fragments generated after the second and third PCR reactions were visualized on agarose gel. The expected length of the fragments was 1090 bp, the length of the whole *ADH2* gene with the extensions for the Golden Gate assembly.

Figure 16 shows the agarose gel after the second and third PCR reactions of the first site saturation mutagenesis cycle on target site G272. Fragments 3 and 5 contained the NDT codon in the G272 site, and fragments 2 and 4 contained the NDT, VMA, ATG, and TAC codons in the mutated site. In Figure 16, the characteristic U-shaped bands caused by the overloading of DNA into the wells can be observed.

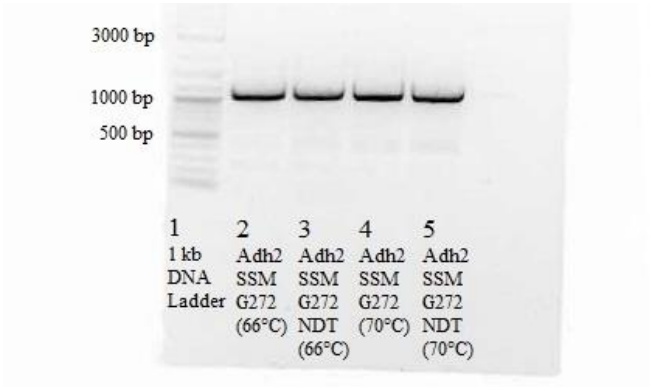



**Figure 16.** Agarose gel (1.5 %) electrophoresis results visualized after the second and third PCR reaction of the site saturation mutagenesis on target site G272.

Wells 1 is the 1 kb DNA Ladder (New England Biolabs, USA), and wells 2-5 are the *ADH2* genes with the mutation in the G272 site. The temperature indicated for each sample indicates the annealing temperature for the first PCR reaction

Figure 17 shows the agarose gel after the second and third PCR reactions of the second site saturation mutagenesis cycle on target sites A52, G272, and Y297. Well 1 contains the

*ADH2* gene with the mutation in the G272 site, wells 2 and 4 contain the *ADH2* gene with the mutated A52 site, and wells 3 and 5 contain the *ADH2* gene with the mutated Y297 site.

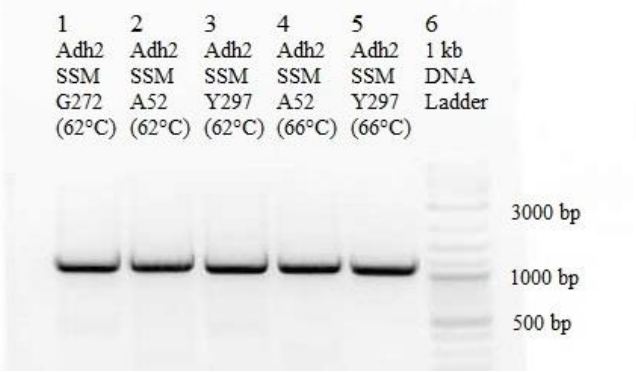



**Figure 17.** Agarose gel (1.5 %) electrophoresis results visualized after the second and third PCR reaction of the site saturation mutagenesis on target sites A52, G272, and Y297.

Well 6 is the 1 kb DNA Ladder (New England Biolabs, USA), wells 1-5 are the mutated *ADH2* genes. The temperature indicated for each sample indicates the annealing temperature for the first PCR reaction

Gel electrophoresis was used to verify the size of the generated DNA fragments after the second and third PCR reactions of the site saturation mutagenesis for the rest of the target sites (L59, L119, S249, and T123) (Figure 18). Well 2 (Figure 18) shows the unsuccessful amplification of the *ADH2* gene with the mutation in the T123 site, while well 5 contains the successfully amplified *ADH2* gene with the same mutated target site. The rest of the mutated *ADH2* genes (L59, L119, and S249) were successfully amplified. After successfully performing the site saturation mutagenesis, the novel Adh2 variants were used for the *in vitro* assembly of the BB3rN pGap_Adh2*_rps25tt plasmids using the Golden PiCS cloning system.

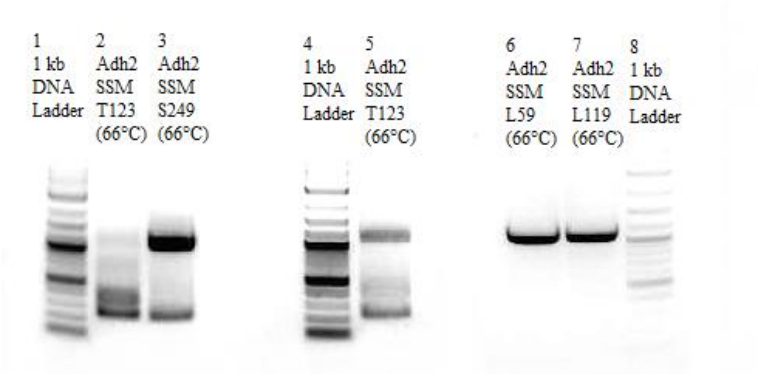



**Figure 18.** Agarose gel (1.5 %) electrophoresis results visualized after the second and third PCR reaction of the site saturation mutagenesis on target sites L59, L119, T123, and S249.

Wells 1, 4, and 8 are the 1 kb DNA Ladder (New England Biolabs, USA), well 2 shows unsuccessful amplification of the *ADH2* gene with mutated site T123, wells 3, 5, 6, and 7 are the mutated *ADH2* genes. The temperature indicated for each sample indicates the annealing temperature for the first PCR reaction

## 4.2. PLASMID SEQUENCE ANALYSIS ON BACTERIAL AND YEAST LEVEL

Transformation of *E. coli* strain DH10B was conducted with the constructed BB3rN pGap_Adh2*_rps25tt plasmids for propagation purposes. Ten *E. coli* transformants were selected for plasmid sequencing for each site saturation mutagenesis site except G272. For site G272, 40 colonies were screened in the first site saturation mutagenesis cycle and 12 in the second site saturation mutagenesis cycle. Different annealing temperatures and primer ratios were used in the first site saturation mutagenesis cycle to optimize the site saturation mutagenesis protocol. It was observed that the higher annealing temperature (70 °C) resulted in the abundance of the G272 wild-type strain. The experimental results of the first bacterial screening for site G272 were included in Figure 19. The number of G272 wild-type transformants for the annealing temperature of 70 °C indicated that the suggested annealing temperature was too high, which made the randomization of site G272 unsuccessful. The randomization was successful for the annealing temperature of 66 °C for both the full site saturation mutagenesis and the NDT-only mutagenesis. It can be noted that the lower annealing temperature produced appropriate amino acid codon distribution that correlated with the used primer mixes. Based on the presented screening results (Figure 19), lower annealing temperatures were used for the following site saturation mutagenesis cycles.

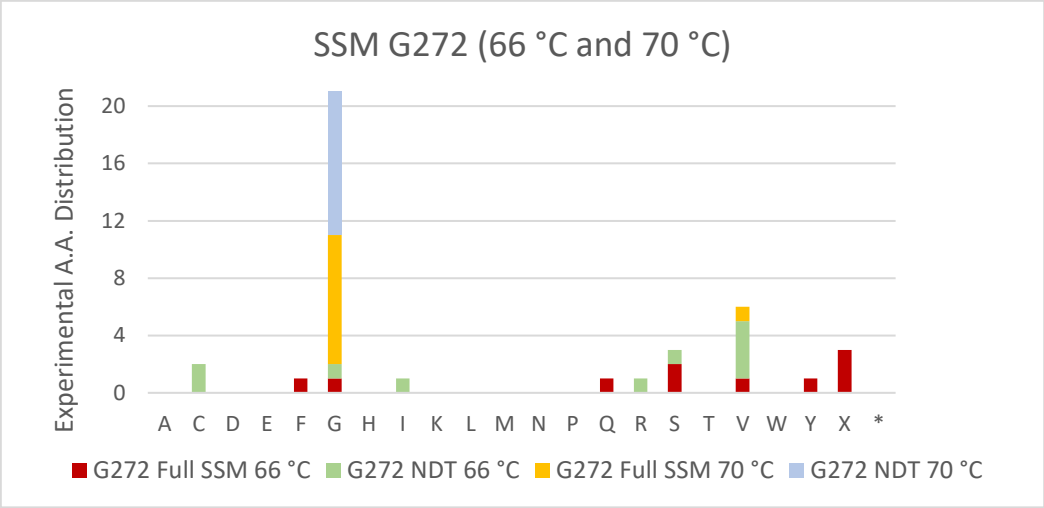



**Figure 19.** Experimental amino acid distribution of the small-intelligent randomization of site G272 in *K. phaffii ADH2* gene using different annealing temperatures (66 °C and 70 °C) and different primer mixes (full site saturation mutagenesis and NDT-only mutagenesis)

For the rest of the site saturation mutagenesis cycles, plasmid pooling was performed to create a diverse plasmid library to transform the competent Mut$^-$ AdhKO biosensor strain.

On the yeast level, a minimum of forty colonies were sent for sequencing for each site saturation mutagenesis cycle. Forty colonies were sequenced for sites A52, G272, L59, L119, and Y297, and sixty colonies were sequenced for sites T123 and S249. In the presented site saturation mutagenesis graphs, bacterial and yeast sequencing results are combined to compare the codon bias on both levels of sequencing (Figures 21-26). The graphs were made using the experimental and theoretical amino acid distributions of small-intelligent randomizations by Tang et al. (2012) as a reference (Figure 20).

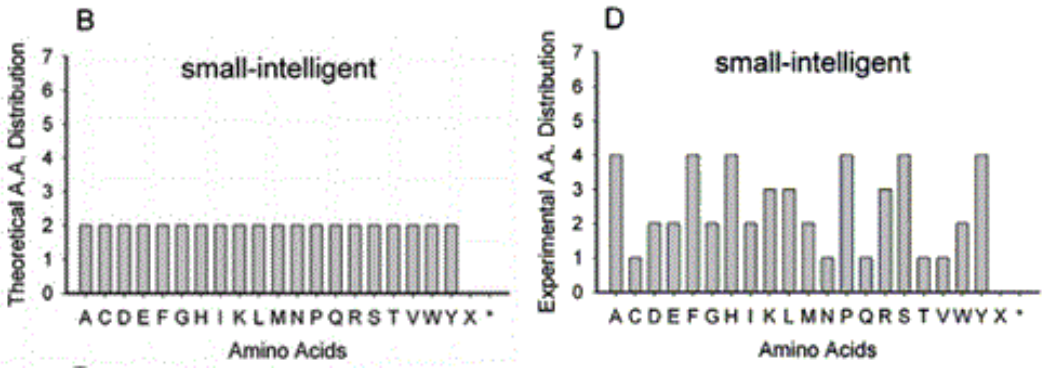
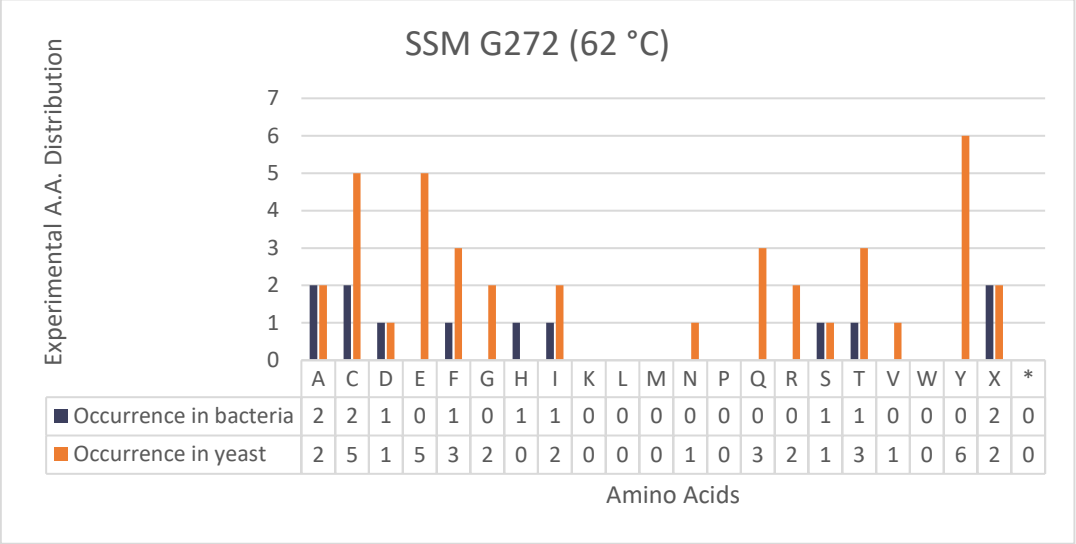



**Figure 20.** Experimental (B) and theoretical (D) amino acid distributions of the small-intelligent randomizations for one mutation site (adapted from Tang et al., 2012)

In the Tang et al. (2012) study, about 55 randomly picked *E. coli* TOP10 colonies were sequenced for each library. The experimental results (Figure 20.D) showed that all 20 amino acid codons were identified in the small-intelligent library, and no rare codons or stop codons were obtained (stop codons are marked by X, and rare codons marked with an asterisk in Figure 20). A correlation between the experimental coverage (Figure 20.D) and theoretical amino acid bias expectations (Figure 20.B) for the small-intelligent randomization was found, although the distribution wasn't perfect (Tang et al., 2012). In the same study, Tang et al. (2012) performed the randomization with the degenerate codon NNS. In the conventional NNS library, the coverage was 16 amino acids, and two stop codons and three rare codons of *E. coli* were obtained. In comparison with the NNS mutagenesis library, the small-intelligent library method provided a better coverage of amino acids and reduced the amino acid biases. The goal was to recreate the full coverage of amino acids or to get a coverage as high as possible.

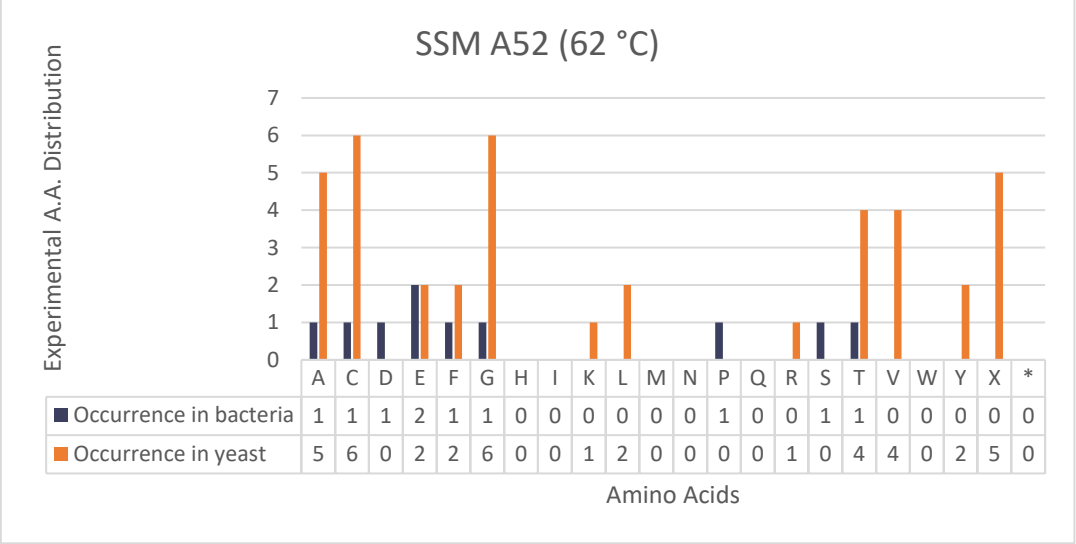
In Figures 21-26, the letter X is used to indicate uncertainty in the identification of the mutated amino acid, for example, if double bands are present in the sequence results. The double bands on the bacterial level indicate the possibility of a mixed colony, while the double

bands on the yeast level indicate the possible occurrence of multiple integration events. The experimental results of the small-intelligent randomization for site G272 (Figure 21) identified eight amino acids on the bacterial level and 14 amino acids on the yeast level, not counting the unknown amino acids (X). The experimental results of the small-intelligent randomization for site A52 (Figure 22) identified eight amino acids on the bacterial level and 11 amino acids on the yeast level.




| | A | C | D | E | F | G | H | I | K | L | M | N | P | Q | R | S | T | V | W | Y | X | * |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Occurrence in bacteria | 2 | 2 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2 | 0 |
| Occurrence in yeast | 2 | 5 | 1 | 5 | 3 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 3 | 2 | 1 | 3 | 1 | 0 | 6 | 2 | 0 |

**Figure 21.** Experimental amino acid distribution of the small-intelligent randomization of site G272 in *K. phaffii ADH2* gene




| | A | C | D | E | F | G | H | I | K | L | M | N | P | Q | R | S | T | V | W | Y | X | * |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Occurrence in bacteria | 1 | 1 | 1 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Occurrence in yeast | 5 | 6 | 0 | 2 | 2 | 6 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 4 | 0 | 2 | 5 | 0 |

**Figure 22.** Experimental amino acid distribution of the small-intelligent randomization for site A52 in *K. phaffii ADH2* gene

In the small-intelligent randomization of site Y297 (Figure 23), the results identified six amino acids on the bacterial level and 15 on the yeast level. It can be observed that the smaller coverage of amino acids is present on the bacterial level, which correlates to the number of screened colonies. Interestingly, it can be noted that certain amino acids are present on the bacterial level screening results but not on the yeast level screening results. This occurrence is unusual, given the higher number of screened colonies on the yeast level compared to the bacterial level, and could reflect that some protein variants could have an inhibitory effect on the yeast growth, but could also be due to the small library size. For the site saturation cycle on target sites A52, G272, and Y297, we can observe that the coverage of amino acids isn't complete, but it is relatively high.

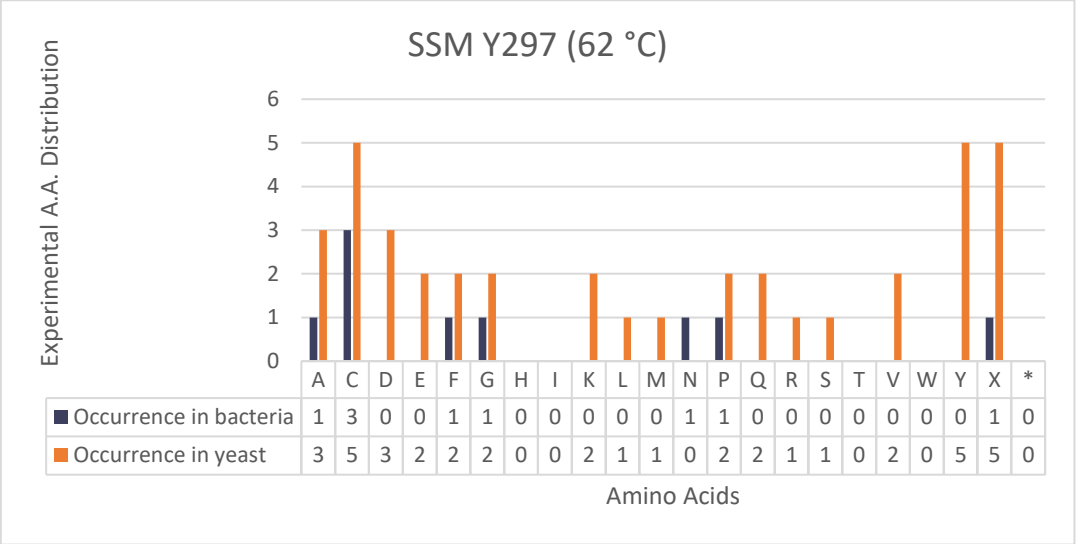



**Figure 23.** Experimental amino acid distribution of the small-intelligent randomization of the site Y297 in *K. phaffii ADH2* gene

In the next cycle of site saturation mutagenesis, target sites S249 and T123 were randomized (Figures 24 and 25). The results of the small-intelligent randomization for site S249 identified three amino acids on the bacterial level and eight amino acids on the yeast level. It can be observed that the amino acid coverage on both bacterial and yeast levels is low. The low coverage on the yeast level results from the predominant mutant strain with the GCA codon (A) in the mutated site. The results of the small-intelligent randomization for site T123 identified five amino acids on the bacterial level and ten amino acids on the yeast level.

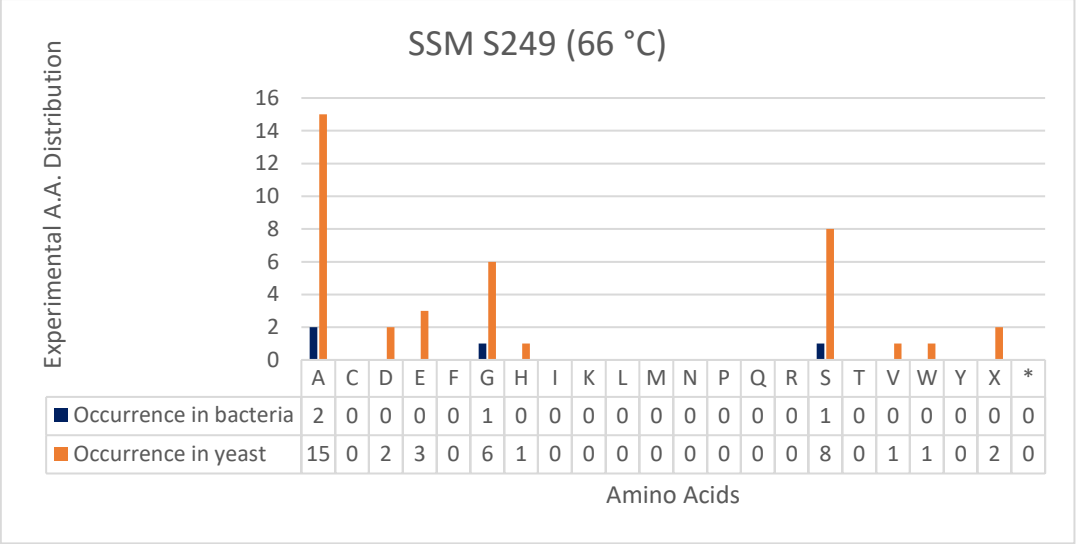



**Figure 24.** Experimental amino acid distribution of the small-intelligent randomization of the site S249 in *K. phaffii ADH2* gene

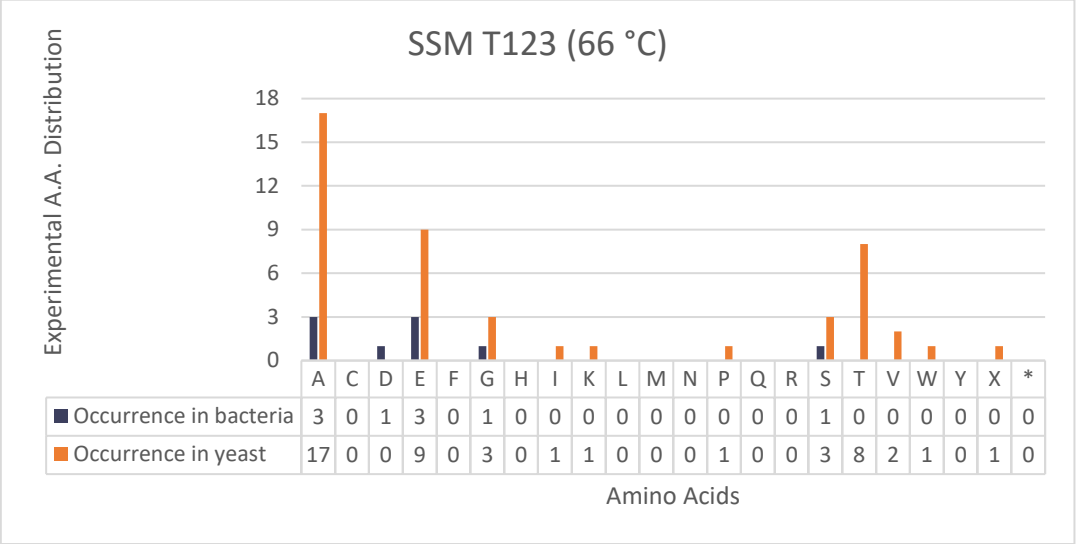
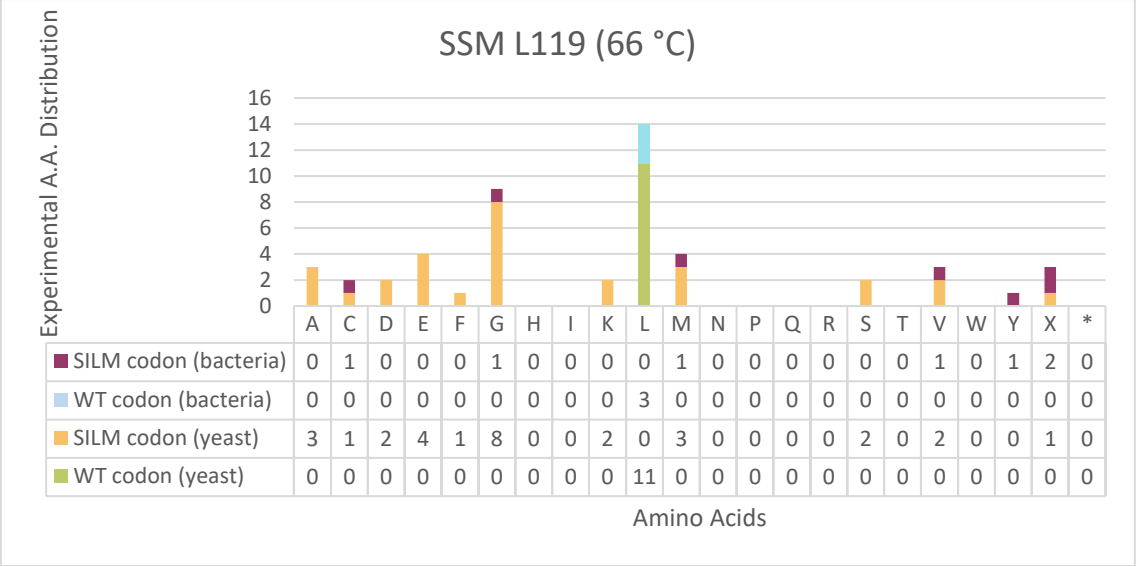



**Figure 25.** Experimental amino acid distribution of the small-intelligent randomization of the site T123 in *K. phaffii ADH2* gene

The sequencing results for site L59 on the bacterial level displayed only the wild-type genotype, so the yeast transformation wasn't performed. The cause of the unsuccessful site saturation mutagenesis is likely due to contamination with wild-type DNA. The results of the small-intelligent randomization for site L119 identified six amino acids on the bacterial level and 11 amino acids on the yeast level (Figure 26). The contamination with the wild-type strain can be observed on the bacterial and yeast level. L119 wild-type strains containing the CTG codon are colored differently than strains produced by mutagenesis to accentuate the absence of the CTT codon for leucine represented in the small-intelligent library method.




| | A | C | D | E | F | G | H | I | K | L | M | N | P | Q | R | S | T | V | W | Y | X | * |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SILM codon (bacteria) | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 0 |
| WT codon (bacteria) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SILM codon (yeast) | 3 | 1 | 2 | 4 | 1 | 8 | 0 | 0 | 2 | 0 | 3 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 1 | 0 |
| WT codon (yeast) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Figure 26.** Experimental amino acid distribution of the small-intelligent randomization of the site L119 in *K. phaffii ADH2* gene.

The bacterial and yeast strains that display the usual codons for the small-intelligent library method are colored different than the wild-type bacterial and yeast strains

Compared with the experimental amino acid bias expectations for the small-intelligent randomization by Tang et al. (2012), the full coverage of amino acids wasn't achieved. Codon distributions observed after the sequencing of plasmids isolated from *E. coli* and *K. phaffii* strains with the randomized *ADH2* gene are presented in Appendix (Supplementary Figures 3 and 4) and compared to the experimental codon distribution for the randomizations of one (D) and two contiguous mutation sites (F) by Tang et al. (2012). Possible causes of lower amino acid coverage include contamination with the wild type or a specific mutant strain, suboptimal annealing temperature, and imperfectly proportionate primer solutions. Additionally, a lower number of transformants was screened compared to the study by Tang et al. (2012), so the coverage could be improved by screening more strains. Contamination with the wild-type DNA can result in the occurrence of the wild-type genotype in all of the screened transformants. The contamination can originate from the impure reagents used for media or buffers or by non-sterile equipment handling. Too high of an annealing temperature can lead to non-specific primer binding and, subsequently, non-specific PCR products that reduce the reaction efficiency, while too low of an annealing temperature can prevent optimal binding of the primers to the templates. Another possible cause of lower amino acid coverage is imperfectly proportionate primer solutions made by the manufacturer or imperfect mixing of the primer solutions, in which case some codons are less represented, and some are represented in a higher degree.

It can be concluded that the best coverage was obtained in the second site saturation mutagenesis cycle on target sites A52, G272 and Y297 after the annealing temperature optimization in the first cycle. The sites with the best coverage all share the same annealing temperature of 62 °C, while the annealing temperature for the other sites was 66 °C. Contamination with the wild-type DNA has limited good coverage in the L59 bacterial screening results. Target sites S249 and T123 have low coverages due to the presence of one predominant mutant strain, while site L119 displays low coverage caused by the high number of the wild-type strain.

### 4.3. SCREENING WITH FLOW CYTOMETRY

The screening of the biosensor strains was performed using the CytoFLEX S Flow Cytometer (Beckman Coulter, USA) to analyze the activity of the mutant *ADH2* genes *in vivo*. The native *FLD1* promoter coupled to the *GFP* gene is induced by formaldehyde, and the activity of *ADH2* genes was determined through the *in vivo* production of eGfp under the *FLD1* promoter (Appendix, Supplementary Figure 5). The measured *in vivo* pFLD1 activities are presented in Figures 27 and 28 for different mutation sites. The y-axis in Figures 27 and 28 represents the promoter activity expressed as the dimensionless mean relative cell-associated product (FX values are calculated based on the equation listed in Section 3.11.1.). The horizontal gridline represents the FX value of the wild-type strain.

The order of amino acids on the x-axis of the graphs (Figures 27 and 28) is based on their side chain properties and size. The smaller amino acids are positioned left on the x-axis, and the bigger amino acids gravitate towards the right side of the x-axis. The amino acids with non-polar side chains are positioned left on the x-axis (yellow), followed by amino acids with polar side chains (purple), amino acids that are positively charged at pH = 7 (pink), and amino acids that are negatively charged at pH = 7 (blue). The negative control in Figures 27 and 28 is the untransformed $Mut^-$ AdhKO biosensor strain. The wild-type strains are colored purple, the mutant strains with lower Adh2 activity than the wild-type strain are colored blue, and the mutant strains with higher Adh2 activity than the wild-type strain are colored orange.

In the screening trial of target site G272 (Figure 27.A), only the strain with the serine residue out of the 13 mutant strains showed Adh2 activity in the same range as the wild-type strain. Mutant strains with alanine and cysteine residues displayed slightly lower Adh2 activity, while the others displayed far lower activity than the wild-type strain with an FX value similar to the control strain with no *ADH2* gene. Interestingly, serine and threonine mutants had very

different FX values despite their structural similarities. This could be due to the fact that threonine is larger in size than serine, and could affect the catalytic properties of the protein by improper folding. It can be observed that serine and cysteine are in the polar amino acid group and that alanine is structurally the most similar amino acid to glycine.

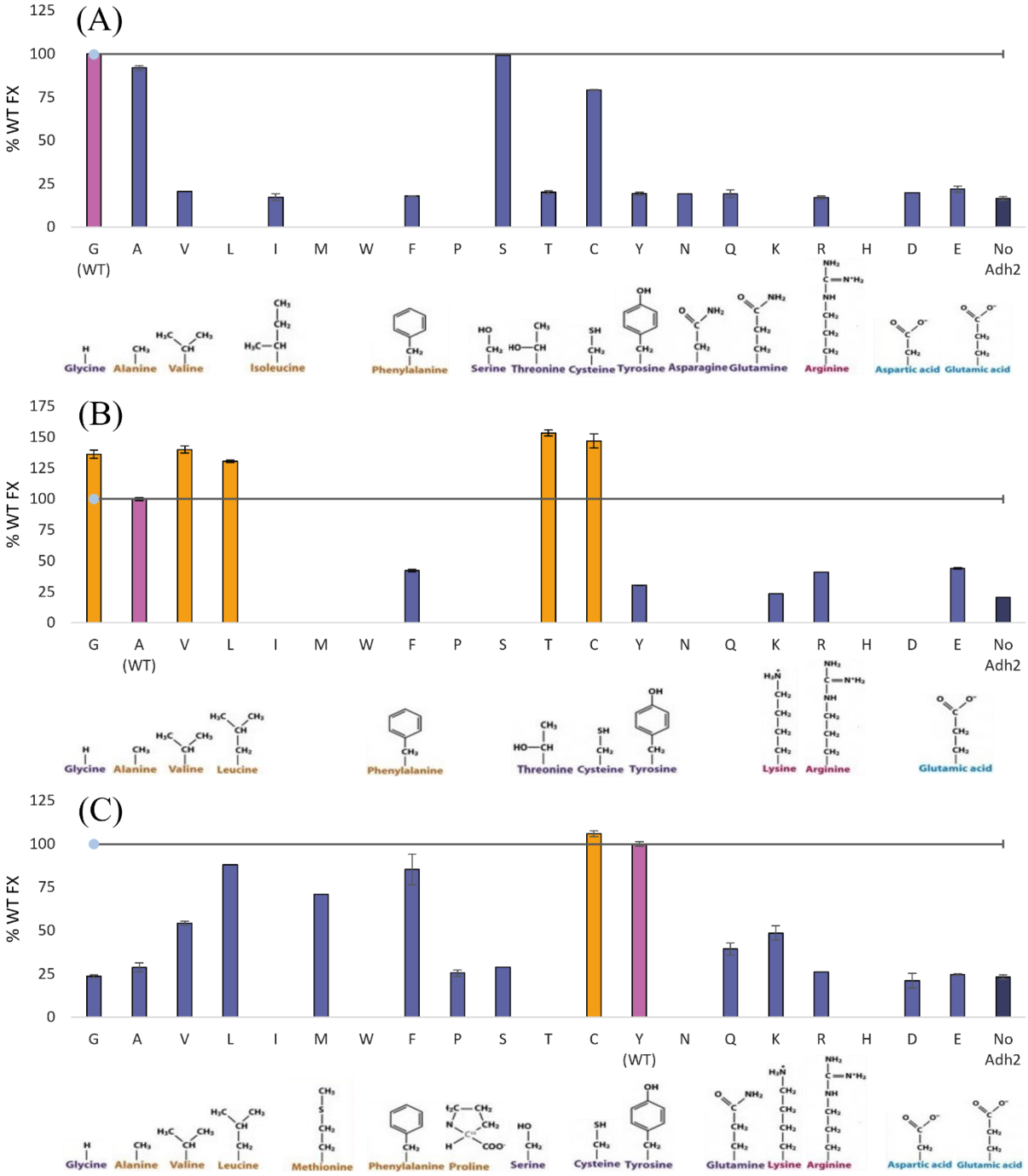



**Figure 27.** In vivo flow cytometry screening results for sites (A) G272, (B) A52, and (C) Y297 were analysed by Kaluza Analysis Software.

The gridlines represent the FX value of the strain with the wild-type. The structures of the represented amino acids are pictured below the graphs for size and structure reference

The screening trial of the G272 mutant strains didn't identify strains with improved methanol utilization phenotype compared to the wild type. The negative control strain had the lowest FX value, as expected.

In the second screening trial (Figure 27.B), five strains displayed higher Adh2 activity than the wild-type strain. Similar to the G272 screening trial, an amino acid with polar side chain, threonine, showed the highest Adh2 activity. Such results could be explained by the beneficial effect of the amino acid on methanol affinity of the Adh2 enzyme. Threonine was accompanied by cysteine, a weak polar amino acid, valine, glycine, and leucine, which all belong to the nonpolar group of amino acids and are structurally similar to adenine. The other mutant strains showed an Adh2 activity lower than 45 % WT FX. This screening trial produced mutant strain that possess up to 150 % WT FX.

The third screening trial (Figure 27.C) produced only one strain with higher Adh2 activity on methanol, with cysteine in the mutated residue. Again, it can be observed that cysteine is in the polar group of amino acids and possesses similar properties as the wild-type residue. Mutant strains with leucine and phenylalanine residues had the highest activity after the wild-type strain, followed by a methionine which wasn't obtained for the first two mutated sites.

In the screening trial of target site T123 (Figure 28.D), it was found that the highest Adh2 activity was present in the strain that codes for serine in the target site. It is intriguing that both amino acids have polar side chains and share similar properties. In this screening trial, plenty of mutant strains with Adh2 activity almost as high as the wild type were found, including glycine, alanine, valine, isoleucine, and lysine. This finding could indicate that site T123 doesn't strongly influence the catalytic site or that it doesn't significantly change the Adh2 enzyme structure. The negative control strain consistently had the lowest FX value out of all screened strains for the last four screenings. Strains with minor Adh2 activity improvements were produced in the last two screening trials.

In the fifth screening trial (Figure 28.E), low coverage of amino acids is present. It was observed that the amino acids glycine, alanine, and valine were continuously present in the screening trials. Despite the low coverage, three strains displayed Adh2 activity almost as high as the wild type, including glycine, alanine, and histidine.

The final screening trial (Figure 28.F) was performed on the strains with the randomized L119 codon. The consistency of the occurrence of amino acids glycine, alanine,

and valine were observed. The strain with phenylalanine in the target residue displays the highest Adh2 activity, although negligible compared to the wild type. After phenylalanine, highest Adh2 activity is displayed in the mutant strains with methionine and cysteine in the target residue. The negative control possesses the lowest FX value of all strains.

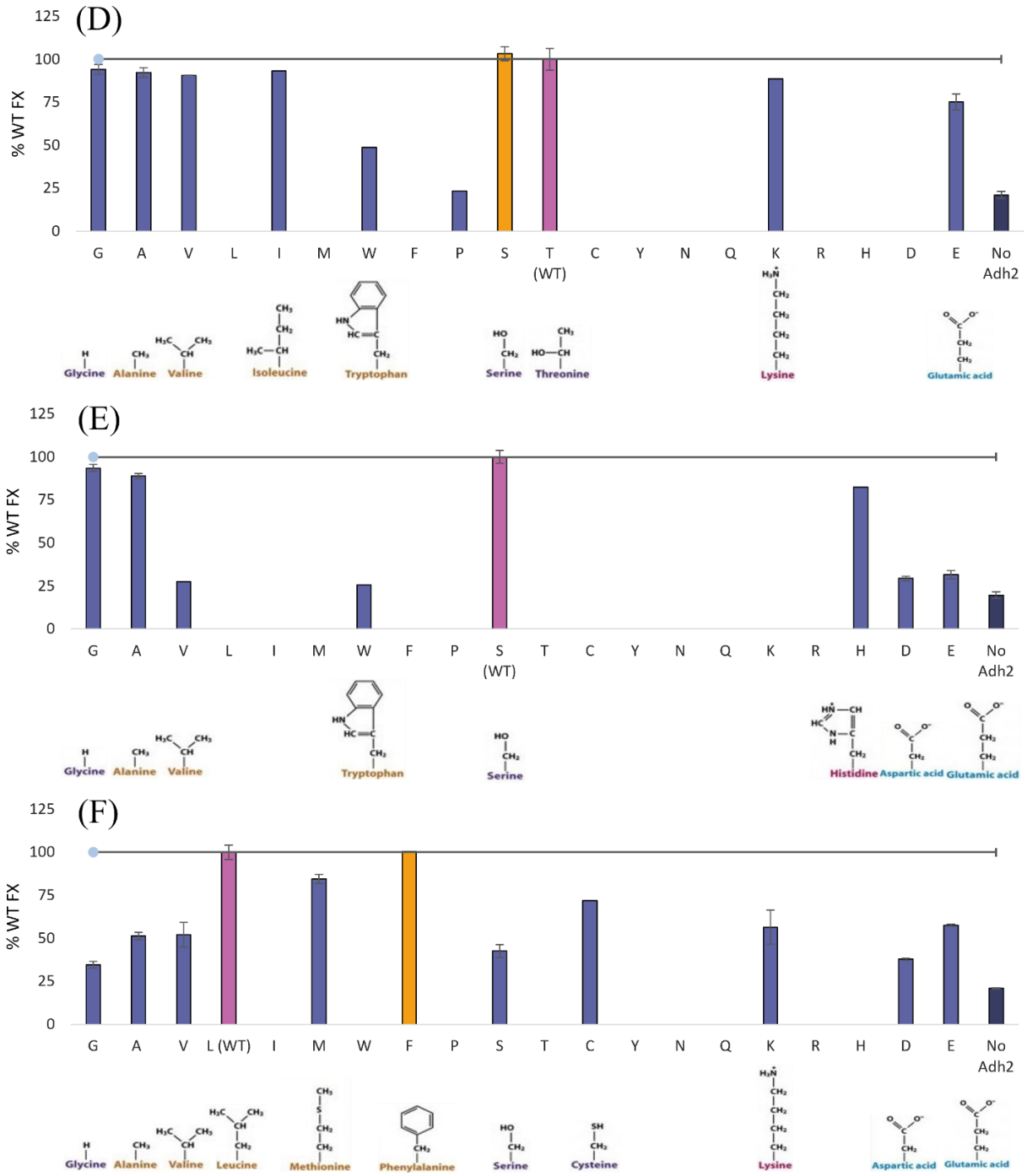



**Figure 28.** *In vivo* flow cytometry screening results for sites (D) T123, (E) S249, and (F) L119 were analysed by Kaluza Analysis Software.

The gridlines represent the FX value of the strain with the wild-type. The structures of the represented amino acids are pictured below the graphs for size and structure reference

In the final screening trial (Figure 28.F), A52 mutants from the second screening trial with higher FX values than the A52 wild-type strain were rescreened to verify the results (Appendix, Supplementary Figure 6). The rescreening results suggested that the A52 wild-type strain had a similar FX value to the other A52 mutant strains, only slightly better. The results indicate an issue with the cloning of the wild-type A52 strain that occurred during the second screening trial.

In this thesis, the focus was on a small subset of all the existing residues in Adh2. The screening trial in which the A52 codon was randomized produced the best results, but the rescreens suggested that a cloning error occurred in the A52 wild-type strain. In other trials, amino acids with polar uncharged side chains and amino acids with non-polar aliphatic side chains were the most represented amongst the highest recorded activities of the Adh2 enzyme compared to the wild type. These results may indicate that the aromatic and charged amino acids obtain properties that interfere with methanol utilization in the catalytic site of the Adh2 enzyme.

A small-intelligent library was developed with target residues that have a role in catalytic sites and are not highly conserved. In the 2015 study, Cahn et al. identified target residues that are not crucial in the catalytic sites but have a strong effect on the enzyme activity. Such target sites that improve enzyme activity through subtle structural changes cannot be predicted (Cahn et al., 2015). In the 2021 study, Gu et al. identified highly conserved target residues, and it was discovered that they have a strong effect on the catalytic function, even though the residues were distant from the active center of the enzyme. It is possible that some residues that are highly conserved or distant from the active site of the enzyme but would improve the enzyme's activity were overlooked. Such risks are unavoidable when designing a small library with a semi-rational approach.

In the 2016 study, Wu et al. performed site-saturation mutagenesis for the directed evolution of activator-independent methanol dehydrogenase from *Cupriavidus necator* N-1. In comparison with our small-scale mutant library, Wu et al. (2016) screened 4000 variants in two separate screening rounds and generated 15 possible positive variants. In this thesis, approximately 200 mutant strains were screened, including the negative controls and rescreening of the strains. The key advantage of our screening approach is that the clones were sequenced before the screening, so information was gained about the enzyme even for the clones that weren't improved.

Many studies have focused on the engineering of bacterial methanol dehydrogenases (Lee et al., 2020; Roth et al., 2019; Wu et al., 2016), but few research efforts on dehydrogenases paid attention to the alcohol dehydrogenases found in *K. phaffii*, with the accent being on its role on ethanol metabolism (Karaoğlan et al., 2020; Karaoğlan et al., 2016). A ground-breaking study by Zavec (2021) recently showed that *K. phaffii* utilizes methanol with its native alcohol dehydrogenase Adh2. That discovery opened the door for researchers to try to improve the methanol utilization process in *K. phaffii*. The screening data in this thesis examined the effect of site saturation mutagenesis of key residues in Adh2 on its activity towards methanol. Several mutated residues whose presence induced modest but seemingly beneficial improvements in the activity of Adh2 towards methanol were identified. Some of the mutated residues in site A52 produced FX values as high as 150 % of the wild-type FX value. The screening of site A52 was repeated to verify the consistency of the results, and the screening results portrayed significantly lower FX values. The rescreen indicated that the initial screening results may have shown higher FX values than realistic ones.

Overall, the screening data provided us with useful insight into the utility of the biosensor strain for *in vivo* flow cytometry screening. The FX values of the mutant strains were based on the properties of the side chains of the mutated residues. The screening results were crucial to confirm that the biosensor strain was an effective screening tool that could be used for high-throughput screenings of larger libraries for use in directed evolution. However, the limitation of the biosensor strain was the low transformation efficiency, which may have been reflected in the screening results. The biosensor has to be transferred into a yeast strain with higher transformation efficiency to generate larger libraries for directed evolution. The biosensor strain will be rebuilt in the X-33 *K. phaffii* strain for the doctoral dissertation of Charles Moritz (Moritz, n.d.) using CRISPR technology to be able to construct better yeast libraries, and the enzyme variants missing from the mutant library constructed in this graduate thesis will be filled in. To continue the search for the residues whose mutations would improve the enzyme's methanol utilization activity, further enzyme engineering of the X-33 biosensor strain will be implemented by Charles Moritz (Moritz, n.d.) using the directed evolution approach.

# 5. CONCLUSIONS

Based on the results presented in this graduate thesis, the following conclusions can be drawn:

1. Site-directed mutagenesis of the *ADH2* gene on target sites A52, G272, L119, S249, T123, and Y297 was successfully performed. Overlap extension PCR was unsuccessful on site T48, possibly due to faulty primer design or inadequate length of the first fragment of the *ADH2* gene for the fragment to generate successfully. The bacterial level sequencing results for target site L59 displayed only the wild-type genotype, presumably due to contamination with the wild-type strain.

2. *In vitro* assembly of the novel *ADH2* gene variants into the BB3rN pGap_CloningSpacer_rps25TT was successfully performed using Golden PiCS. The mutated *ADH2* genes with the extensions for the Golden PiCS assembly were cloned directly into a specially designed BB3rN vector.

3. The experimental amino acid coverage produced with the small-intelligent library ranged from 8 to 15 amino acids on the yeast level, compared to the full coverage of amino acids produced using the same methodology by Tang et al. (2012). Possible causes of lower coverage include contamination with the wild type or a specific mutant strain, suboptimal annealing temperature, and imperfectly proportionate primer solutions. A higher number of screened colonies could potentially improve the amino acid coverage.

4. The activity of the mutated *ADH2* genes was determined through the *in vivo* production of eGfp under the *FLD1* promoter. Out of six screening trials, four trials produced mutant strains with higher Adh2 activity than the wild-type strain. Out of the four screening trials, only one involved a mutant strain (A52) that exhibited FX value higher than 150 % of the wild-type activity.

5. It was noted that the *K. phaffii* mutant strains with amino acids that share similar properties to the wild-type amino acids produced the best results in the screening trial. Amino acids with polar uncharged side chains and amino acids with non-polar aliphatic side chains were the most represented amongst the highest recorded activities of Adh2 enzyme compared to the wild type.

6. In the last screening trial for site L119, the A52 mutant strains with the FX value higher than the A52 wild-type strain were rescreened, and the results showed that the FX values for those mutant strains were slightly lower than the FX value of the A52 wild-type strain. This anomaly

in screening can most likely be attributed to issues with the wild-type strain during the initial screening, as the mutants had similar relative activity to each other in the rescreening.

7. Further engineering of the Adh2 enzyme will be necessary to confirm if the *K. phaffii* A52 mutant strains display improved properties and to test more target sites to improve the methanol utilization process in *K. phaffii*. The enzyme variants missing from the mutant library constructed in this graduate thesis will be filled in the doctoral dissertation of Charles Moritz (Moritz, n.d.).

# 6. REFERENCES


Ata Ö, Rebnegger C, Tatto NE, Valli M, Mairinger T, Hann S, et al. (2018) A single Gal4-like transcription factor activates the Crabtree effect in Komagataella phaffii. *Nat Commun* **9**, 4911. https://doi.org/10.1038/s41467-018-07430-4

Ata Ö, Ergün BG, Fickers P, Heistinger L, Mattanovich D, Rebnegger C, et al. (2021) What makes Komagataella phaffii non-conventional? *FEMS Yeast Res* **21**, foab059. https://doi.org/10.1093/femsyr/foab059

Bernauer L, Radkohl A, Lehmayer LGK, Emmerstorfer-Augustin A (2021) Komagataella phaffii as Emerging Model Organism in Fundamental Research. *Front Microbiol* **11**, 607028. https://doi.org/10.3389/fmicb.2020.607028

Cahn JKB, Baumschlager A, Brinkmann-Chen S, Arnold FH (2015) Mutations in adenine-binding pockets enhance catalytic properties of NAD(P)H-dependent enzymes. *Protein Eng Des Sel* **29**, 31-38. https://doi.org/10.1093/protein/gzv057

Carneiro CVGC, Serra LA, Pacheco TF, Ferreira LMM, Brandão LTD, Freitas MN de M, et al. (2022) Advances in Komagataella phaffii Engineering for the Production of Renewable Chemicals and Proteins. *Fermentation* **8**, 575. https://doi.org/10.3390/fermentation8110575

Chica RA, Doucet N, Pelletier JN (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr opin biotechnol* **16**, 378–384. https://doi.org/10.1016/j.copbio.2005.06.004

Cotton CA, Claassens NJ, Benito-Vaquerizo S, Bar-Even A (2020) Renewable methanol and formate as microbial feedstocks. *Curr opin biotechnol* **62**, 168–180. https://doi.org/10.1016/j.copbio.2019.10.002

Cregg JM (2007) Pichia Protocols, 2nd ed., Humana Press, Totowa, NJ.

de Schutter K, Lin Y-C, Tiels P, van Hecke A, Glinka S, Weber-Lehmann J, et al. (2009) Genome sequence of the recombinant protein production host Pichia pastoris. *Nat Biotechnol* **27**, 561–566. https://doi.org/10.1038/nbt.1544

Fabarius JT, Wegat V, Roth A, Sieber V (2021) Synthetic Methylotrophy in Yeasts: Towards a Circular Bioeconomy. *Trends biotechnol* **39**, 348–358. https://doi.org/10.1016/j.tibtech.2020.08.008

Fischer JE, Glieder A (2019) Current advances in engineering tools for Pichia pastoris. *Curr opin biotechnol* **59**, 175–181. https://doi.org/10.1016/j.copbio.2019.06.002

Gassler T, Heistinger L, Mattanovich D, Gasser B, Prielhofer R (2019) CRISPR/Cas9-Mediated Homology-Directed Genome Editing in Pichia pastoris. *Methods mol biol* **1923**, 211–225. https://doi.org/10.1007/978-1-4939-9024-5_9

Gassler T (2019) Metabolic engineering of Pichia pastoris for utilization of carbon dioxide (Unpublished doctoral dissertation). University of Natural Resources and Life Sciences, Vienna

Gassler T, Sauer M, Gasser B, Egermeier M, Troyer C, Causon T, et al. (2020) The industrial yeast Pichia pastoris is converted from a heterotroph into an autotroph capable of growth on CO2. *Nat Biotechnol* **38**, 210–216. https://doi.org/10.1038/s41587-019-0363-0

Gassler T, Baumschabl M, Sallaberger J, Egermeier M, Mattanovich D (2022) Adaptive laboratory evolution and reverse engineering enhances autotrophic growth in Pichia pastoris. *Metab eng* **69**, 112–121. https://doi.org/10.1016/j.ymben.2021.11.007

Grinna LS, Tschopp JF (1989) Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast, Pichia pastoris. *Yeast* **5**, 107–115. https://doi.org/10.1002/yea.320050206

Gu J, Sim BR, Li J, Yu Y, Qin L, Wu L, et al. (2021) Evolutionary coupling-inspired engineering of alcohol dehydrogenase reveals the influence of distant sites on its catalytic efficiency for stereospecific synthesis of chiral alcohols. *Comput struct biotechnol* **19**, 5864–5873. https://doi.org/10.1016/j.csbj.2021.10.031

Guilliermond A (1920) Zygosaccharomyces Pastori, Nouvelle Espèce de Levures à Copulation Hétérogamique, Bulletin de la Société mycologique de France.

Hartner FS, Glieder A (2006) Regulation of methanol utilisation pathway genes in yeasts. *Microb Cell Fact* **5**, 39. https://doi.org/10.1186/1475-2859-5-39

Heistinger L, Gasser B, Mattanovich D (2020) Microbe Profile: *Komagataella phaffii*: a methanol devouring biotech yeast formerly known as *Pichia pastoris*. *Microbiology* **166**, 614–616. https://doi.org/10.1099/mic.0.000958

Holliday GL, Sarkies P (n.d.) Mechanism and Catalytic Site Atlas. https://www.ebi.ac.uk/thornton-srv/m-csa/entry/256/. Retrieved December 22, 2022.

Johnson EA, Echavarri-Erasun C (2011) Yeast Biotechnology. E: Kurtzmann CP, Fell JW, Boekhout T (edit.) The Yeasts: A Taxonomic Study, 5th ed., Elsevier Science, Amsterdam, Netherlands, pages 21-44.

Karaoğlan M, Karaoğlan FE, İnan M (2016) Functional analysis of alcohol dehydrogenase (ADH) genes in Pichia pastoris. *Biotechnol lett* **38**, 463–469. https://doi.org/10.1007/s10529-015-1993-z

Karaoğlan M, Karaoğlan FE, Yılmaz S, İnan M (2020) Identification of Major ADH Genes in Ethanol Metabolism of Pichia pastoris. *Yeast* **37**. https://doi.org/10.1002/yea.3443

Karbalaei M, Rezaee SA, Farsiani H (2020) Pichia pastoris: A highly successful expression system for optimal synthesis of heterologous proteins. *J Cell Physiol* **23**, 5867–5881. https://doi.org/10.1002/jcp.29583

Kurtzman CP (2005) Description of Komagataella phaffii sp. nov. and the transfer of Pichia pseudopastoris to the methylotrophic yeast genus Komagataella. *Int j syst evol micr* **55**, 973–976. https://doi.org/10.1099/ijs.0.63491-0

Lee JY, Park SH, Oh SH, Lee JJ, Kwon KK, Kim SJ, et al. (2020) Discovery and Biochemical Characterization of a Methanol Dehydrogenase From *Lysinibacillus xylanilyticus*. *Front bioeng biotechnol* **8**, 67. https://doi.org/10.3389/fbioe.2020.00067

Lutz S (2010) Beyond directed evolution - semi-rational protein engineering and design. *Curr opin biotechnol* **21**, 734–743. https://doi.org/10.1016/j.copbio.2010.08.011

Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, et al. (2009) Genome, secretome and glucose transport highlight unique features of the protein production host Pichia pastoris. *Microb Cell Fact* **8**, 29. https://doi.org/10.1186/1475-2859-8-29

Moritz C (n.d.) Unpublished doctoral dissertation. University of Natural Resources and Life Sciences, Vienna.

New England Biolabs (2022) Tm Calculator v1.15.0. https://tmcalculator.neb.com/#!/main. Retrieved December 16, 2022.

Ogata K, Nishikawa H, Ohsugi M (1969) A Yeast Capable of Utilizing Methanol. *Agric Biol Chem* **33**, 1519–1520. https://doi.org/10.1080/00021369.1969.10859497

Peña DA, Gasser B, Zanghellini J, Steiger MG, Mattanovich D (2018) Metabolic engineering of Pichia pastoris. *Metab Eng* **50**, 2–15. https://doi.org/10.1016/j.ymben.2018.04.017

Phaff HJ (1956) A proposal for amendment of the diagnosis of the genus Pichia hansen. *Antonie van Leeuwenhoek* **22**, 113–116. https://doi.org/10.1007/BF02538318

Phaff HJ, Knapp EP (1956) The taxonomy of yeasts found in exudates of certain trees and other natural breeding sites of some species of Drosophila. *Antonie van Leeuwenhoek* **22**, 117–130. https://doi.org/10.1007/BF02538319

Prielhofer R, Barrero JJ, Steuer S, Gassler T, Zahrl R, Baumann K, et al. (2017) GoldenPiCS: a Golden Gate-derived modular cloning system for applied synthetic biology in the yeast Pichia pastoris. *BMC Syst Biol* **11**, 123. https://doi.org/10.1186/s12918-017-0492-3

Roth TB, Woolston BM, Stephanopoulos G, Liu DR (2019) Phage-Assisted Evolution of Bacillus methanolicus Methanol Dehydrogenase 2. *ACS Synt Biol* **8**, 796–806. https://doi.org/10.1021/acssynbio.8b00481

Rußmayer H, Buchetics M, Gruber C, Valli M, Grillitsch K, Modarres G, et al. (2015) Systems-level organization of yeast methylotrophic lifestyle. *BMC biol* **13**, 80. https://doi.org/10.1186/s12915-015-0186-5

Stratton J, Chiruvolu V, Meagher M (1998) High cell-density fermentation. *Methods mol bio* **103**, 107–120. https://doi.org/10.1385/0-89603-421-6:107

Tang L, Gao H, Zhu X, Wang X, Zhou M, Jiang R (2012) Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers. *BioTechniques* **52**, 149–158. https://doi.org/10.2144/000113820

Tran A-M, Nguyen T-T, Nguyen C-T, Huynh-Thi X-M, Nguyen C-T, Trinh M-T, et al. (2017) Pichia pastoris versus Saccharomyces cerevisiae: a case study on the recombinant production of human granulocyte-macrophage colony-stimulating factor. *BMC Res Notes* **10**, 148. https://doi.org/10.1186/s13104-017-2471-6

van der Klei IJ, Yurimoto H, Sakai Y, Veenhuis M (2006) The significance of peroxisomes in methanol metabolism in methylotrophic yeast. *Biochim Biophys Acta* **1763**, 1453-62. https://doi.org/10.1016/j.bbamcr.2006.07.016

Wang Y, Xue P, Cao M, Yu T, Lane ST, Zhao H (2021) Directed Evolution: Methodologies and Applications. *Chem rev* **121**, 12384–12444. https://doi.org/10.1021/acs.chemrev.1c00260

Wegat V, Fabarius JT, Sieber V (2022) Synthetic methylotrophic yeasts for the sustainable fuel and chemical production. *Biotechnol biofuels bioprod* **15**, 113. https://doi.org/10.1186/s13068-022-02210-1

Wegner GH (1990) Emerging applications of the methylotrophic yeasts. *FEMS Microbiol Rev* **7**, 279–283. https://doi.org/10.1111/j.1574-6968.1990.tb04925.x

Weninger A, Fischer JE, Raschmanová H, Kniely C, Vogl T, Glieder A (2018) Expanding the CRISPR/Cas9 toolkit for Pichia pastoris with efficient donor integration and alternative resistance markers. *J cell biochem* **119**, 3183–3198. https://doi.org/10.1002/jcb.26474

Wu TY, Chen CT, Liu JT, Bogorad IW, Damoiseaux R, Liao JC (2016) Characterization and evolution of an activator-independent methanol dehydrogenase from Cupriavidus necator N-1. *Appl microbiol biotechnol* **100**, 4969–4983. https://doi.org/10.1007/s00253-016-7320-3

Yamada Y, Matsuda M, Maeda K, Mikata K (1995) The phylogenetic relationships of methanol-assimilating yeasts based on the partial sequences of 18S and 26S ribosomal RNAs: the proposal of Komagataella gen. nov. (Saccharomycetaceae). *Biosci biotechnol biochem* **59**, 439–444. https://doi.org/10.1271/bbb.59.439

Yurimoto H, Kato N, Sakai Y (2005) Assimilation, dissimilation, and detoxification of formaldehyde, a central metabolic intermediate of methylotrophic metabolism. *Chem rec* **5**, 367–375. https://doi.org/10.1002/tcr.20056

Yurimoto H, Oku M, Sakai Y (2011) Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis. *Int j microbiol* **2011**, 101298. https://doi.org/10.1155/2011/101298

Yurimoto H, Shiraishi K, Sakai Y (2021) Physiology of Methylotrophs Living in the Phyllosphere. *Microorganisms* **9**, 809. https://doi.org/10.3390/microorganisms9040809

Zavec D (2019) Engineering Pichia pastoris for improved methanol utilization (Unpublished doctoral dissertation). University of Natural Resources and Life Sciences, Vienna

Zavec D, Troyer C, Maresch D, Altmann F, Hann S, Gasser B, et al. (2021) Beyond alcohol oxidase: the methylotrophic yeast Komagataella phaffii utilizes methanol also with its native alcohol dehydrogenase Adh2. *FEMS yeast res* **21**, foab009. https://doi.org/10.1093/femsyr/foab009

Zhang H, Li Q, Wang L, Chen Y (2018) Investigation of structure and function of mitochondrial alcohol dehydrogenase isozyme III from Komagataella phaffii GS115. *Biochim Biophys Acta Gen Subj* **1862**, 1199–1208. https://doi.org/10.1016/j.bbagen.2018.02.012

# 7. APPENDIX

**Supplementary Table 1**: List of primers used in this graduate thesis. All primers were designed and synthesized in this thesis except for MOC_36, MOC_37, MOC_79, and MOC_80

| Primer name | Sequence | Primer purpose | Tm for Q5 polymerase (°C) |
|---|---|---|---|
| MOC_36 | ATGTCTCCAACTATCCCAAC | *K. phaffii* Adh2 F universal | 61 |
| MOC_37 | TTATTTGGAAGTGTCCACAACG | *K. phaffii* Adh2 R universal | 63 |
| MOC_79 | CATGTCTCCAACTATCCCAACTACAC | *K. phaffii* Adh2 F universal high Tm | 67 |
| MOC_80 | AAGCTTATTTGGAAGTGTCCACAAC | *K. phaffii* Adh2 R universal high Tm | 65 |
| MOC_114 | CGGTGTTATTAACGTC**NDT**GTTTCCCCAC | **Ser249** F Primer NDT | 69-73 |
| MOC_115 | CGGTGTTATTAACGTC**VMA**GTTTCCCCAC | **Ser249** F Primer VMA | 69-73 |
| MOC_116 | CGGTGTTATTAACGTC**ATG**GTTTCCCCAC | **Ser249** F Primer ATG | 71 |
| MOC_117 | CGGTGTTATTAACGTC**TGG**GTTTCCCCAC | **Ser249** F Primer TGG | 73 |
| MOC_118 | GACGTTAATAACACCGTGTGGACC | **Ser249** R Primer | 68 |
| MOC_119 | ACCTGTCTGGTTTC**NDT**CACGACGGA | **Thr123** F primer NDT | 71-76 |
| MOC_120 | ACCTGTCTGGTTTC**VMA**CACGACGGA | **Thr123** F primer VMA | 72-76 |
| MOC_121 | ACCTGTCTGGTTTC**ATG**CACGACGGA | **Thr123** F primer ATG | 74 |
| MOC_122 | ACCTGTCTGGTTTC**TGG**CACGACGGA | **Thr123** F primer TGG | 76 |
| MOC_123 | GAAACCAGACAGGTCAGCC | **Thr123** R primer | 66 |
| MOC_124 | CGAGATCAAGGGATCT**NDT**GTTGGAAACAGAG | **Y297** F Primer NDT | 70-73 |
| MOC_125 | CGAGATCAAGGGATCT**VMA**GTTGGAAACAGAG | **Y297** F Primer VMA | 70-73 |
| MOC_126 | CGAGATCAAGGGATCT**ATG**GTTGGAAACAGAG | **Y297** F Primer ATG | 71 |
| MOC_127 | CGAGATCAAGGGATCT**TAC**GTTGGAAACAGAG | **Y297** F Primer TGG | 72 |
| MOC_128 | AGATCCCTTGATCTCGATGGAC | **Y297** R Primer | 66 |
| MOC_129 | CACACTGATTTGCAC**NDT**TGGAAGGGTG | **A52** F Primer NDT | 70-74 |
| MOC_130 | CACACTGATTTGCAC**VMA**TGGAAGGGTG | **A52** F Primer VMA | 70-74 |

| | | | |
|---|---|---|---|
| MOC_131 | CACACTGATTTGCAC**ATG**TGGA AGGGTG | **A52** F Primer ATG | 72 |
| MOC_132 | CACACTGATTTGCAC**TAC**TGGA AGGGTG | **A52** F Primer TGG | 71 |
| MOC_133 | GTGCAAATCAGTGTGACAGACA | **A52** R Primer | 65 |
| MOC_134 | GGTAAGGTTGTTTTGGTT**NDT**CT GCCATCTGG | **G272** F Primer NDT | 71 - 74 |
| MOC_135 | GGTAAGGTTGTTTTGGTT**VMA**C TGCCATCTGG | **G272** F Primer VMA | 71 - 75 |
| MOC_136 | GGTAAGGTTGTTTTGGTT**ATG**CT GCCATCTGG | **G272** F Primer ATG | 73 |
| MOC_137 | GGTAAGGTTGTTTTGGTT**TAC**CT GCCATCTGG | **G272** F Primer TGG | 73 |
| MOC_138 | AACCAAAACAACCTTACCCAAA GTTCT | **G272** R Primer | 67 |
| MOC_139 | AACCAAAACAACCTTACCCAAA GTT | **G272** R Primer (lower Tm) | 65 |
| MOC_140 | AACCAAAACAACCTTACCCAAA GTTCTAACG | **G272** R Primer (higher Tm) | 69 |
| MOC_141 | GATAGGTCTCACATGTCTCCAAC TATCCCAACTACACAAAAGG | Redesign of MOC_79 F primer for higher Tm | 70 |
| MOC_142 | GATCGGTCTCTAAGCTTATTTGG AAGTGTCCACAACGT | Redesign of MOC_80 R primer for higher Tm | 69 |
| MOC_146 | TGACTGGCCA**NDT**GACAACAAG CTT | **L59** F Primer NDT | 69 - 73 |
| MOC_147 | TGACTGGCCA**VMA**GACAACAAG CTT | **L59** F Primer VMA | 69 - 73 |
| MOC_148 | TGACTGGCCA**ATG**GACAACAAG CTT | **L59** F Primer ATG | 71 |
| MOC_149 | TGACTGGCCA**TAC**GACAACAAG CTT | **L59** F Primer TAC | 71 |
| MOC_150 | TGGCCAGTCACCCTTCCAG | **L59** R Primer | 70 |
| MOC_151 | CAAGGCTGAC**NDT**TCTGGTTTCA CC | **L119** F Primer NDT | 68 - 72 |
| MOC_152 | CAAGGCTGAC**VMA**TCTGGTTTC ACC | **L119** F Primer VMA | 68 - 73 |
| MOC_153 | CAAGGCTGAC**ATG**TCTGGTTTC ACC | **L119** F Primer ATG | 70 |
| MOC_154 | CAAGGCTGAC**TAC**TCTGGTTTCA CC | **L119** F Primer TAC | 69 |
| MOC_155 | GTCAGCCTTGGCACAACTGGATT | **L119** R Primer | 70 |

**Supplementary Figure 1:** Protocol for Preparation of chemically competent *E. coli* with RbCl and transformation created by Stefanie Müller in 2013



# Preparation of chemically competent *E. coli* with RbCl and transformation

| Effective Date: | Created by: |
|---|---|
| 16.10.2013 | Stefanie Müller |
| Checked by (Sign/Date): | Michael Sauer |
| Major changes to previous version: | |

## 1. Purpose and scope

Preparation of chemically competent *E. coli* cells and transformation.

## 2. Materials and equipment

**Solution TFB1:** 30 mM potassium acetate, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM RbCl, and 15% glycerol; adjust to pH 5.8 (with 1 M acetic acid), filter-sterilize, and store at 4°C (ready to use) or at room temperature (cool down before use).

**Solution TFB2:** 100 mM MOPS (or PIPES), 75 mM $CaCl_2$,10 mM RbCl, and 15% glycerol; adjust to pH 6.5 (with 1 M KOH), filter-sterilize, and store at 4°C (ready to use) or at room temperature (cool down before use).

## 3. Procedure

**Preparation of chemically competent cells**

- *E. coli* strain DH10B (or another appropriate strain) is inoculated from a glycerol stock onto an LB plate; the inoculum is streaked on the plate using a loop so as to obtain individual colonies. The plate is incubated overnight at 37°C.
- Inoculate 10 mL of LB medium with a single colony and incubate the flask overnight in a shaker-incubator (37°C, shaking 180 rpm).
- The following day, transfer 2 mL of this culture to a flask containing 200 mL LB medium and incubate for about 2 h until $OD_{600}$ reaches 0.6.

[This protocol was adapted/developed by the Department of Biotechnology, BOKU University of Natural Resources and Life Sciences Vienna. Please acknowledge this.]

- Cool down the cells on ice for 10 min. The cells are pelleted in a centrifuge for 5 min at 4,500 rpm (4,000 × g) at 4°C. The cells are resuspended in 0.4 volume of ice-cold TFB1.
- Repeat the centrifugation step. Resuspend the pellet in 1/25 volume of ice-cold TFB2.
- The cells are aliquoted 100 µL per tube and stored at −80°C.

**Transformation**

- Frozen chemically competent cells (100 µL per tube) are thawed on ice.
- The entire ligation (or other DNA sample to be transformed) is added to the cells, and the mix incubated on ice for 30 min.
- The cells and DNA mix is heat shocked for 90 s at 42°C.
- Allow the cells to recover on ice for 5 min.
- Add 1 mL of LB medium to the cells and incubate the tube at 37°C in a shaker-incubator (150 rpm) for 45 min to 1 h.
- Plate the appropriate amount of cells on selective LB agar.

## 4. Data analysis and documentation

Colonies on selective plate.

**Supplementary Figure 2:** Protocol for Preparation of electro-competent *Pichia pastoris* cells – small scale created by Richard Zahrl in 2016



# Preparation of electro-competent Pichia pastoris cells – small scale protocol

| Effective Date: | Created by: |
|---|---|
| 18.04.2016 | Richard Zahrl |
| Checked by (Sign/Date): | |
| Major changes to previous version: | adaption of our protocol to resemble www.ncbi.nlm.nih.gov/pubmed/14740498 |

## 1. Purpose and scope

Generation of Pichia pastoris cells, which are competent for transformation by electroporation.

## 2. Health and safety considerations

Irritating substances are used. Wear suitable protective clothing and gloves.

## 3. Materials and equipment

- Centrifuge for Falcon-Tubes (50 mL), cooled to 4 °C
- Shaker (180 rpm; 25 °C)
- Eppis (approx. 8; ice-cold)
- YPD medium (+ appropriate antibiotics)
- DTT (1 M)
- Tris-HCl (1 M, pH 7.5)
- Lithium **chloride** (10 M; *may need a drop of acid for complete solubility*)
- Sorbitol (1 M)

- pre-treating solution (20 mL are needed for one 100 mL cell batch)
  - 7.4 mL AD
  - 12 mL of 1 M Sorbitol (for a final concentration of 0.6 M)
  - 200 µL of 1 M Tris-HCl, pH 7.5 (for a final concentration of 10 mM)
  - 200 µL of 1 M DTT (for a final concentration of 10 mM)
  - 200 µL of 10 M Lithium chloride (for a final concentration of 100 mM)

## 4. Procedure

*Note: Please make sure that the sorbitol (1 M) for usage after pre-treatment is **ice-cold before you start**. Also, take care that you cells are kept ice-cold throughout the whole procedure, from pre-treatment until freezing the final aliquots at -70 °C.*

Day 1 (evening):

- Prepare the pre-culture: inoculate 5-10 mL selective YPD medium with a fresh colony (no older than 3 weeks) and incubate overnight (o/n; ~ 16 h) (shaker; 180 rpm; 25 °C)

Day 2 (evening):

- Main culture: Inoculate 100 mL non-selective YPD medium with the preculture and incubate o/n ( ~ 16 h) (shaker; 180 rpm; 25 °C); the end-OD should be between 1.2 and 2.5. Calculate the volume of pre-culture needed for inoculation of the main culture according to the following equation:

$$V_{inoc}[\mu L] = \frac{OD_m * V_m}{e^{\mu * t}} * \frac{1000}{OD_{pre}}$$

$OD_m$ ….$OD_{600}$ main culture after time t (use $OD_{600}$ 1.5 for calculation)
$V_m$ = volume main culture [mL]
$t_m$ = incubation time of the main culture [h] (at least 15 h)
µ = 0.3 $h^{-1}$ for *P. pastoris* wild type in YPD at 25 °C
$OD_{pre}$ = $OD_{600}$ pre-culture

Day 3 (morning):

- Measure OD and split 90 mL of the culture into **2 Falcon-tubes** (2x45 mL)
- Harvest the cells by centrifuging (5 min; 1500 x g and 4 °C); discard supernatant

- Add 10 mL of **pre-treating solution** (25 °C) to each of the pellets and resuspend the cells
- Incubate the cells for **30 min** (Shaker; 180 rpm; 25 °C)
- Add 40 mL **ice-cold sorbitol (1 M)** to each Falcon-tube and harvest cells (5 min; 1500 x g; 4 °C), discard supernatant
- **Combine pellets** in 1 Falcon and resuspend cells in 45 mL **ice-cold sorbitol (1 M)**
- Harvest the cells (5 min; 1500 x g; 4 °C), discard supernatant
- Resuspend cells in 45 mL **ice-cold sorbitol (1 M)**
- Harvest them (5 min; 1500 x g; 4 °C), discard supernatant
- Resuspend cells in **200 µL ice-cold sorbitol (1 M)**; aliquot cells into pre-chilled Eppendorf tubes **(à 80 µL)** and keep them on ice until transformation (long term-storage: -80 °C)

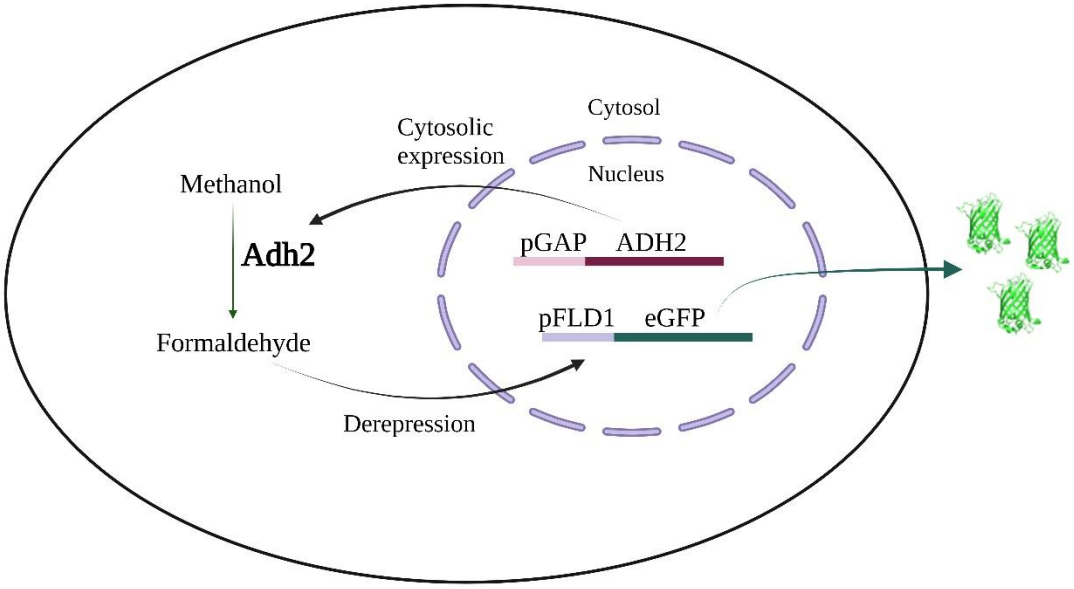
**Supplementary Figure 3:** Codon distribution observed after the sequencing of the *E. coli* strains with the randomized *ADH2* gene compared to the experimental codon distribution for the randomizations of one (D) and two contiguous mutation sites (F) by Tang et al. (2012)

| Amino Acid | SSM G272 B | SSM A52 B | SSM Y297 B | SSM S249 B | SSM T123 B | SSM L119 B | Tang et al.* D | Tang et al.* F | Codon* |
|---|---|---|---|---|---|---|---|---|---|
| F | 0.1 | 0.1 | 0.1 | 0 | 0 | 0 | 0.08 | 0.05 | TTT |
| I | 0.1 | 0 | 0 | 0 | 0 | 0 | 0.04 | 0.03 | ATT |
| K | 0 | 0 | 0 | 0 | 0 | 0 | 0.06 | 0.03 | AAA |
| Y | 0 | 0 | 0 | 0 | 0 | 0.1 | 0.08 | 0.02 | TAT |
| S | 0.1 | 0.1 | 0 | 0.1 | 0.1 | 0 | 0.08 | 0.07 | AGT |
| C | 0.2 | 0.1 | 0.3 | 0 | 0 | 0.1 | 0.02 | 0.05 | TGT |
| D | 0.1 | 0.1 | 0 | 0 | 0.1 | 0 | 0.04 | 0.09 | GAT |
| E | 0 | 0.2 | 0 | 0 | 0.3 | 0 | 0.04 | 0.05 | GAA |
| H | 0.1 | 0 | 0 | 0 | 0 | 0 | 0.08 | 0.02 | CAT |
| L | 0 | 0 | 0 | 0 | 0 | 0.3 | 0.06 | 0.04 | CTT |
| N | 0 | 0 | 0.1 | 0 | 0 | 0 | 0.02 | 0.05 | GAU |
| M | 0 | 0 | 0 | 0 | 0 | 0.1 | 0.04 | 0.04 | ATG |
| Q | 0 | 0 | 0 | 0 | 0 | 0 | 0.02 | 0.02 | CAA |
| T | 0.1 | 0.1 | 0 | 0 | 0 | 0 | 0.02 | 0.07 | ACA |
| V | 0 | 0 | 0 | 0 | 0 | 0.1 | 0.02 | 0.08 | GTT |
| G | 0 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.04 | 0.12 | GGT |
| P | 0 | 0.1 | 0.1 | 0 | 0 | 0 | 0.08 | 0.07 | CCA |
| R | 0 | 0 | 0 | 0 | 0 | 0 | 0.06 | 0.04 | CGT |
| W | 0 | 0 | 0 | 0 | 0 | 0 | 0.04 | 0.02 | TGG |
| A | 0.2 | 0.1 | 0.1 | 0.2 | 0.3 | 0 | 0.08 | 0.04 | GCA |
| X | 0.2 | 0 | 0.1 | 0 | 0 | 0.2 | 0 | 0 | / |
| * | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | / |

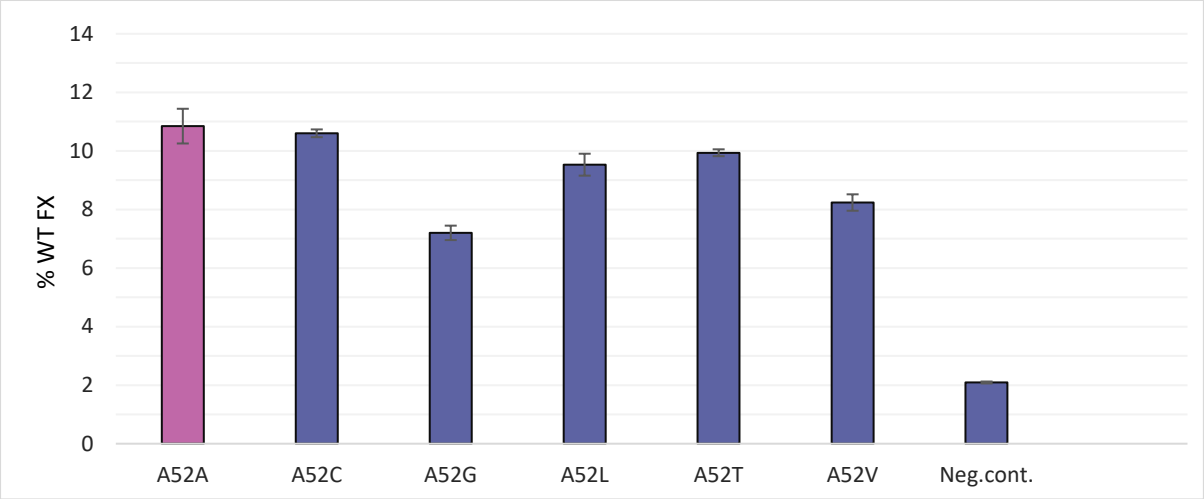
**Supplementary Figure 4:** Codon distribution observed after the sequencing of *K. phaffii* strains with the randomized *ADH2* gene compared to the experimental codon distribution for the randomizations of one (D) and two contiguous mutation sites (F) by Tang et al. (2012)

| Amino Acid | SSM G272 Y | SSM A52 Y | SSM Y297 Y | SSM S249 Y | SSM T123 Y | SSM L119 Y | Tang et al.* D | Tang et al.* F | Codon* |
|---|---|---|---|---|---|---|---|---|---|
| F | 0.075 | 0.05 | 0.05 | 0 | 0 | 0.025 | 0.08 | 0.05 | TTT |
| I | 0.05 | 0 | 0 | 0 | 0.016 | 0 | 0.04 | 0.03 | ATT |
| K | 0 | 0.025 | 0.05 | 0 | 0.016 | 0.05 | 0.06 | 0.03 | AAA |
| Y | 0.15 | 0.05 | 0.125 | 0 | 0 | 0 | 0.08 | 0.02 | TAT |
| S | 0.025 | 0 | 0.025 | 0.133 | 0.05 | 0.05 | 0.08 | 0.07 | AGT |
| C | 0.125 | 0.15 | 0.125 | 0 | 0 | 0.025 | 0.02 | 0.05 | TGT |
| D | 0.025 | 0 | 0.075 | 0.033 | 0 | 0.05 | 0.04 | 0.09 | GAT |
| E | 0.125 | 0.05 | 0.05 | 0.05 | 0.15 | 0.1 | 0.04 | 0.05 | GAA |
| H | 0 | 0 | 0 | 0.016 | 0 | 0 | 0.08 | 0.02 | CAT |
| L | 0 | 0.05 | 0.025 | 0 | 0 | 0.275 | 0.06 | 0.04 | CTT |
| N | 0.025 | 0 | 0 | 0 | 0 | 0 | 0.02 | 0.05 | GAU |
| M | 0 | 0 | 0.025 | 0 | 0 | 0.075 | 0.04 | 0.04 | ATG |
| Q | 0.075 | 0 | 0.05 | 0 | 0 | 0 | 0.02 | 0.02 | CAA |
| T | 0.075 | 0.1 | 0 | 0 | 0.133 | 0 | 0.02 | 0.07 | ACA |
| V | 0.025 | 0.1 | 0.05 | 0.016 | 0.033 | 0.05 | 0.02 | 0.08 | GTT |
| G | 0.05 | 0.15 | 0.05 | 0.1 | 0.05 | 0.2 | 0.04 | 0.12 | GGT |
| P | 0 | 0 | 0.05 | 0 | 0.016 | 0 | 0.08 | 0.07 | CCA |
| R | 0.05 | 0.025 | 0.025 | 0 | 0 | 0 | 0.06 | 0.04 | CGT |
| W | 0 | 0 | 0 | 0.016 | 0.016 | 0 | 0.04 | 0.02 | TGG |
| A | 0.05 | 0.125 | 0.075 | 0.25 | 0.283 | 0.075 | 0.08 | 0.04 | GCA |
| X | 0.05 | 0.125 | 0.125 | 0.033 | 0.016 | 0.025 | 0 | 0 | / |
| * | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | / |

**Supplementary Figure 5:** A schematic representation of the *K. phaffii* CBS2612 biosensor strain mechanism of action is illustrated (Δaox1 Δaox2 Δadh2 Δadh900). The *in vivo* production of eGfp under the *FLD1* promoter determines the promoter activity expressed as the dimensionless mean relative cell-associated product (FX)



**Supplementary Figure 6:** Rescreening data from the final screening trial on site L119 (Figure 28.F) for A52 mutants from the second screening trial with higher FX values than the A52 wild-type strain. The rescreening results showed lower FX values for the A52 mutant strains, which indicates an issue in the screening process



DECLARATION OF ORIGINALITY

I LARA KALOGJERA declare that this master's thesis is an original result of my own work and it has been generated by me using no other resources than the ones listed in it.

Lara Kalogjera

Signature