Master Thesis

New Pichia pastoris platform strains for protein expression

Student: *Viktorija Vidimče (Викторија Видимче)*

Supervisor:

Glieder, Anton, Ao.Univ.-Prof. Mag.rer.nat.Dr.rer.nat.

Institute of Molecular Biotechnology

Graz University of Technology

Graz, June 2012

Danksagung:

(Diese Seite wird händisch geschrieben)

Благодарност

(Diese Seite wird händisch geschrieben)

Abstract

Pichia pastoris is a methylotrophic yeast widely used for industrial production of heterologous proteins. Proper protein folding which often depends on post translational modifications is of crucial importance for producing active recombinant protein. This master thesis was focused on a genes coding for enzymes involved in protein posttranslational modification.

First of all, DNA sequence analysis of all genes amplified from Pichia pastoris CBS7435 was performed, and the results were compared with the existing sequences in GenDB. This comparison did not show any unexpected results, but showed some sequence mismatches revealed by the much more accurate new generation sequencing methods.

Four FLP (recombinase) PpPDI promoter specific P_{AOX1}-replacement cassettes were designed. In principle a previously generated FLP-recombinase cassette was used. It was redesigned and modified for gene replacement use, instead of its original knock-out function. These resulted in four P_{AOX1}-replacement cassettes all carrying the recombinase gene, a gene for zeocin resistance, the AOX promoter (P_{AOX1}), and a 500 bp long sequence which was homologous to the PDI genes. In addition, new Pichia pastoris ΔKU70 strains carrying genes for model proteins were designed. As model proteins CalB (Candida antartica lipase B), and HRP (horseradish peroxidase) were used. Out of over 250 clones screened for each construct, one single copy and one multi copy clone was chosen for each protein.

Furthermore two knock-out cassettes were designed, both of them of ends-out type. The first one carried homologous sequences of the upstream and downstream region of PpOCH1, and the second one of PpHOC1. In addition, two new Pichia pastoris strains (CSB7435 and ΔKU70) carrying an HNL5 gene were designed. HNL5 codes for hydroxynitrile lyase, an isoenzyme from Prunus dulcis[*]. Out of over 250 clones screened for each strain, two multi copy clones were chosen for knockout cassette transformation. However, no transformation resulted in a knockout clone, which opened a broad discussion, leading to new strategies for the generation of an OCH1 or HOC1 knockout.

[*] *Prunus dulcis* is also known as *Prunus amygdalus.*

Index

1. *Introduction to Pichia pastoris*

1.1 Brief overview of Pichia pastoris phylogenetic determination

1.2 Historical overview of Pichia pastoris becoming industry-relevant yeast

1.3 Methanol utilization (MUT) pathway in Pichia pastoris

1.3.1 Pichia pastoris alcohol oxidases

1.4 Pichia pastoris as industrial protein production host

1.5 Pichia pastoris related scientific research areas

1.1 Brief overview of *Pichia pastoris* phylogenetic determination

Almost 30 years ago, Koichi Ogata discovered yeast species able to metabolize methanol (MeOH) and utilize it as sole carbon and energy source, and he defined them as methylotrophic yeasts[1]. At this time the determination of yeasts was based upon their phenotypic differences, such as the ability to grow on certain carbon and nitrogen compounds.

In the 1980's, Kurtzman introduced a new way to determine phylogenetic relations between organisms, based on their nuclear DNA (nDNA) relatedness. Later, numerous studies have reported about other phylogenic determination methods such as rRNA and rDNA sequence analysis. According to rRNA/rDNA phylogenetic sequence analysis of D1/D2 domain of 5S, 18S and 26S, all ascomycetous yeasts are divided in three major lineages: the hamiascomycetes, the euacsomycetes and the archiascomycetes[2]. After all known methylotrophic yeasts have been analysed for divergence in partial sequences of nuclear large subunit (LSU) and nuclear small subunit (SSU) of ribosomal RNA (rRNA), the following three genera have been proposed and described: *Komagataella*, *Ogataea* and *Pichia*[2,3]. Although genetic determination methods have been used in all these studies, their results did not make a stable phylogenetic basis. However, they have indicated the beginning of a new era in taxonomy, where gene sequence analysis presents the most exact way to phylogenetic determination of living organisms.

Recently, two new strains, *Komagataella phaffii* and *Komagataella pseudopastoris* have been described within the genus *Pichia*. Their phenotypic similarity with *Pichia pastoris*[*], strain which is widely used in biotechnology, initiated additional sequence analysis, after which *Pichia pastoris* was defined as legitimate synonym of *Komagataella phaffii*[* *] [4,3].

Pichia pastoris is homothallic ascomycetous yeast and belongs to the Saccharomycetaceae family (http://www.ncbi.nlm.nih.gov/Taxonomy/)[1]. Usually this yeast exists as haploid, but diploid generations can be achieved by growing the cells on nitrogen-limited media[1].

All known *Pichia pastoris* strains had been stored in the ARS Culture Collection, also known as Northern Regional Research Laboratories Collection. They are all derivatives of the wild-type strain, registered as NRRL-Y 11430 (http://nrrl.ncaur.usda.gov/)[5]. In the Centraalbureau voor Schimmelcultures this strain has been deposited as CBS7435[6], as well as ATCC 76273, and CECT 11047 (http://www.uniprot.org/taxonomy/4922).

Recently performed genome sequencing projects of *Pichia pastoris* CBS7435 strain, displayed a leap forward in the understanding of the biology of this methylotroph. Within the 9,35Mbp high-quality genome sequence, 5007 coding domain sequences (CDS) located on four chromosomes have been automatically annotated[6]. Of all the predicted gene sequences, 83,62% are single exon genes, 12,42% contain a single intron, 2,86% contain 2 introns, 0,84%

[*] This yeast was also used as experimental organism within this master thesis.

[* *] The name *Pichia pastoris* is more common used synonym of *Komagataella phaffii*.

contain 3 introns, 0,18% contain 4 introns, and 0,008% of all contain up to 5 introns[6]. Within this project also the *P. pastoris* mitochondrial genome was sequenced, and manually annotated; it was the first mitochondrial genome of a methylotrophic yeast to be sequenced[6].

1.2 Historical overview of *Pichia pastoris* becoming industry-relevant yeast

The demand for high amounts of different proteins in the world market continuously increases. However, only a limited number of these proteins can be produced at high yields by their natural host organisms. Therefore, different microorganisms, including *P. pastoris*, are being used as production hosts for heterologous proteins.

The *Pichia* story has started in the 1970s in the Phillips Petroleum Company. This company developed all media and protocols for growing of *P. pastoris* as methylotrophic yeast, with the main focus on production of single cell proteins (SCP) meant to be used as high-protein animal feed[1]. Unfortunately, very soon after the initial research work with *P. pastoris* has been done, methanol costs have been dramatically increased by oil crises. As a result of that, further development of *P. pastoris* related processes had to be suspended for an indefinite period of time. However, years later the company signed an agreement with the Salk Institute Biotechnology/Industrial Associates, Inc. (SIBIA, La Jolla, CA) and the research work on this yeast continued[1]. It led to the development of vectors, strains and protocols for molecular manipulation of *P. pastoris*.

At that time the scientific attention was attracted by the strong alcohol oxidase 1 promoter (P_{AOX1}). Besides its strength, there is one other desired feature, namely its ability to be tightly regulated, a feature that made P_{AOX} the most implemented promoter for driving the industrial heterologous protein production in *P. pastoris*. In addition, this methylotroph prefers a respiratory rather than a fermentative growth, which enables high cell density production environment (>130g/l dry cell weight)[7,1].

A basic *P. pastoris* expression kit is available from Life Technologies, licensed from Research Corporation Technologies RCT (Tuscon, AZ, USA) who commercialized the products of Phillips Petroleum[8]. A new FTO (freedom to operate) system was established by TU Graz together with VTU Technology and ACIB GmbH and is available from TU Graz. This system is based on the wild type strain CBS 7435.

1.3 Methanol utilization (MUT) pathway in *Pichia pastoris*

Like in all presently known methylotrophic yeasts, **m**ethanol **ut**ilization (MUT) in *P. pastoris* takes place in two different cell compartments: in the peroxisomes and in cytosol.

Based on the recent *P. pastoris* genome sequencing study, a more detailed metabolic network of the MUT pathway has been published[6]. Altogether there are nine genes coding for enzymes directly involved in this pathway, and they belong to seven different enzyme families[6]. Within this group of nine enzymes, two pairs of isoenzymes have been identified, namely AOX1 and AOX2, and DAS1 and DAS2. Having two *DAS* genes involved in MUT pathway makes *P. pastoris* unique among all methylotrophs[6]. It is also important to be mentioned that all nine genes are differently controlled. Although all of their promoters are methanol inducible, those promoting genes included in assimilation pathways are repressed by glucose, while those included in dissimilation pathways are not[9].

The promoter of the *AOX1* gene is the most used promoter for *P. pastoris* protein expression and also preferred for my studies due to its properties and FTO situation.

1.3.1 *Pichia pastoris* alcohol oxidases

During the past decades scientists working with *P. pastoris* have designed many different mutated strains relevant for specific scientific and industrial tasks. As a result of that, nowadays three methanol utilization *P. pastoris* phenotypes are widely used. Mut^+ (**m**ethanol **ut**ilization **plus**) with intact *AOX1* and *AOX2* genes, Mut^s (**m**ethanol **ut**ilization **s**low) with disruption in *AOX1* gene, and Mut^- (**m**ethanol **ut**ilization **minus**) where both, *AOX1* and *AOX2* genes are disrupted[8]. As a consequence of above mentioned gene disruptions these three different strain phenotypes show different growing rates. Namely, Mut^+ grows at wild type rate, Mut^s grows slower and Mut^- is unable to grow on methanol medium[8].

1.4 *Pichia pastoris* as industrial protein production host

P. pastoris exhibits a propensity for homologous recombination between its genomic and foreign or synthetic DNA sequence[1]. Inserted vectors show high stability in the genome[8]. These particular features allowed *P. pastoris* to become an easy manipulated host for heterologous protein production.

In general, the approach of inserting heterologous gene sequences into *P. pastoris* genome can be briefly described with following steps:

1. Design of *P. pastoris* codon optimized shuttle vector which contains at least one copy of the heterologous gene and the marker gene (usually gene for antibiotic resistance),
2. Transformation of the transition host cell (usually *Escherichia coli*) with circular shuttle vector,
3. Transformation of *P. pastoris* host strain with linearized shuttle vector,
4. Screening for desired protein activity (usually with spectrophotometric screening assays, capillary electrophoresis or dot blot technologies).

The heterologous genes inserted in *P. pastoris* can be under the control of P_{AOX1}, constitutive promoter, or synthetic promoter (Table 1). The following reasons explain the need of different promoter variants:

- Depending on the nature of the protein to be expressed, weak expression is in some cases preferred in order to ensure proper protein folding[1] and to avoid possible cytotoxicity,
- Protein overexpression can in some cases induce new bottlenecks in the cell metabolic pathways[2],
- MeOH inducible promoters are undesired in the biotechnological production of food, since this compound is toxic to the humans[1].

Promoter designation	P_{GAP}	P_{FLD1}	P_{PEX8}	P_{YPT1}
Gene description	glyceraldehyde 3-phosphate dehydrogenase	gluthathione-dependent formaldehyde dehydrogenase	peroxisomal matrix protein	guanosine triphosphatase (GTP-ase)
Promoter type	strong constitutive	MeOH and methylamine inducible	MeOH inducible	moderate constitutive
Important features	no MeOH demand, not recommended for driving expression of cytotoxic protein	methylamin = nontoxic and inexpensive nitrogen source	recommended for driving expression of cytotoxic proteins	recommended for driving expression of cytotoxic proteins

Table 1: Brief depiction of four (P_{GAP}, P_{FLD1}, P_{PEX8}, and P_{YPT1}) *P. pastoris* promoters used for driving heterologous gene expression[11,1,8].

The copy number of heterologous gene sequences inserted in *P. pastoris* can vary. In general no standard correlation between gene copy number and protein yield has been

defined[12]. Depending on the protein to be expressed and also depending on the promoter which is being used for driving the expression, drastic variation in the expressed protein level has been observed. Therefore screening studies for strains carrying different gene copy number are necessary to be done in order to find the optimal strain for industrial use.

The capability of *P. pastoris* for expression of intracellular and extracellular (secreted) proteins is generally known. The intracellular expressed proteins can also be targeted into the peroxisomes (cube-shaped cell organelles also called microbodies)[13]. Herewith possible cytotoxicity by the heterologous protein can be avoided. Only few proteins are naturally secreted by *P. pastoris*, thus the majority of the secreted proteins (sometimes over 80%) in the fermentation medium is the heterologous one[14,15].

Unfortunately, there are also cellular and environmental factors negatively affecting industrial heterologous protein production with *P. pastoris*. For instance, secreted heterologous proteins in high cell density fermentation medium are easily degraded by *P. pastoris* proteases. One way to overcome this problem is the use of protease-deficient strains such as SMD-1163 (*his4 pep4 prb1*), SMD-1165 (*his 4 prb1*), and SMD-1168 (*his4 pep4*)[1, 8,16]. However the lower viability and growth rate of this strains make them useful only in cases where no other solution is available[1].

In addition it has to be mentioned that there are no secreted proteins by *P. pastoris*. Thus protease activity in the culture medium seems to come from lysed cells growing under physiological and physical stress.

A viable solution to increase product yield consists of the adjustment of fermentation parameters, such as pH and temperature. *P. pastoris* shows efficient growth rate in the pH range from 3.0 to 7.0, an interval wide enough to allow significant process adjustments[14,8]. Increase of the product yield has been shown also by lowering the temperature in the fermentation from 30°C down to 20°C[8].

The cultivation process of *P. pastoris* (using P_{AOX1}) in bioreactors can in general be divided in three phases. The first one is the growth phase (or batch phase) where the culture grows in salt medium on non-fermentable carbon sources (C-sources). The second phase (transition phase) is characterized by addition of the same C-substrate on the cell growth-limiting rate in order to prepare the culture for the next phase. In the third phase (induction phase), an appropriate amount of methanol is added, and the methanol utilization pathway and the expression of foreign gene are induced[17]. An important fact is that some C-sources such as glycerol, glucose, and ethanol repress the P_{AOX1}, which decreases the heterologous protein yield[17,9]. However, for other C-sources such as sorbitol, mannitol, trehalose, and alanine no repression to the P_{AOX1} has been observed[9]. In general, *P. pastoris* requirements for the growth and production medium include glycerol and methanol (if a methanol inducible promoter is used) as main substrates, and minerals, biotin, salt, and trace elements[7]. All these compounds are inexpensive which makes the industrial use of this yeast interesting. The medium is also free of toxins and pyrogens[18]. During the fermentation, adequate monitoring and control of the fermentation parameters are also necessary in order to achieve maximal product yields. Directed by this need, bioprocess development continuously aims to improve of the existing and development of new monitoring methods.

Another process – based method to increase the yield of intact heterologous proteins consists of the addition of amino acid-rich supplements to the fermentation medium. This is an effect of partial saturation of the existing proteases in the fermentation medium by the amino acid suplements[14].

All above mentioned features contributed to turn this methylotrophic yeast into a more and more attractive tool for biotech and pharmaceutical companies. It is already a host strain for industrial production of over 400 different heterologous proteins[14,7].

1.5 *Pichia pastoris* related scientific research areas

Nowadays, the growth of all biotech companies is characterized by high demand for production of new proteins with customized features. And even thought that the technology for the use of *P. pastoris* is already mature, there are still very important topics waiting to be more precisely described.

In the year 1995 only 4% of total recombinant proteins reported by PubMed* have been produced in *P. pastoris*. Up to year 2009 this number increased to 17%[19].

Many publications regarding *P. pastoris* state that this yeast is also important for basic molecular biology studies. To avoid any confusion, strict distinction has to be made between two *P. pastoris* oriented research areas. In the first one, *P. pastoris* represents model organism for studying all gene- and protein-linked processes inside the cell. Only proper understanding of the molecular basis of this organism can facilitate its fast and effective manipulation. In the second area *P. pastoris* is host cell for expression of high yields of desired protein, but the research interest is focused on protein itself. Both areas are connected since the most of hypothesis made in the first area arise from case studies already done in the second area and vice versa. However, incomparability of the case studies due to varying cultivation conditions is being observed. Thus comparison of case studies results is not always of scientific relevance.

One of the most important research topics concerning *P. pastoris* as model organism is the biogenesis of the peroxisomes. After methanol induction, the number of peroxisomes dramatically increases and they can take up to 80% of the cell volume[20]. Peroxisomes play a role in storing the heterologously expressed proteins, hence protecting the cell from possible toxicity produced by the same. Additional, being stored in the peroxisomes, expressed proteins are unavailable for the proteases, and they are protected from proteolytic degradation[21]. Short amino acid sequences located either at the C-terminus or at the N-terminus of an expressed protein, are responsible for the peroxisomal protein targeting. They are called peroxisome targeting sequences (PTS)[21]. The fact that different PTS exist for the peroxisome matrix proteins, makes an assumption of existing different mechanisms for protein targeting to the peroxisomes[21].

As already said, differently regulated promoter variants are of great importance for the industrial use of this yeast (Paragraph 1.4). This opens a new research topic, identification of regulatory regions within promoter sequences, understanding of their regulation on molecular level, and design of new synthetic promoters[9].

Also other topics which influence the expression efficiency of *P. pastoris* have been defined. For instance, an *AOX1*- case study has shown that the composition and the length of the 5'- untranslated region (5'-UTR) influences the heterologous protein expression[12]. For that reason, and in order to maintain optimal protein expression, it has been suggested to keep the 5'-UTR sequence as similar as possible to that of *AOX1* mRNA[12]. It is also important to avoid AUG triplets in the 5'-UTR in order to ensure exact start codon (AUG) recognition by RNA polymerase[12]. Another important parameter is the A+T composition of cDNA. Since AT-rich regions characterize termination sequence regions, their occurrence in a gene coding domain should be therefore avoided if possible, because it can induce premature transcription termination[12].

* PubMed is a free database offering primarily abstracts and references from life sciences. (http://www.ncbi.nlm.nih.gov/pubmed/)

With protein secretion, the cell lysis step is avoided and it makes downstream processes quite easier and cheaper, and also increases the yield of the expressed protein. The *Saccharomyces cerevisiae* pre-pro α-mating factor (AMF) signal sequence, is standardly used secretion signals in this methylotrophic yeast[8,12]. However not all proteins can be secreted in *P. pastoris*. There is a supposition that using a protein's natural secretion signal, promotes its efficient secretion. Because of that, different heterologous sequences are being evaluated in *P. pastoris*, and it has been shown that additional optimization is usually needed depending on the protein to be secreted. For this reason, detection of new *P. pastoris* secretion signal sequences, description of their transport mechanisms, and definition of their sequence optimization possibilities, are other currently popular research topics.

Another popular research topic is the recombination in the genome of *P. pastoris* (Chapter 1.4). Also protein folding and posttranslational modifications such as glycosylation are being studied intensively (Chapter 1.2 and 1.3). These two topics are of enormous importance for both, producing of correctly folded proteins and increasing the yield of the product.

For all the above mentioned problems (limitations), strategies in order to enhance the production are being improved by the research groups working with *P. pastoris* all over the world. And all of these research studies should help to make this yeast more predictable production host, thereby saving resources and time spent for trial and error experiments.

2. *Recombination in Pichia pastoris*

2.1 Homologous recombination in general

2.2 Single site homologous recombination in Pichia pastoris

2.3 Double site homologous recombination in Pichia pastoris

2.4 FLP site-specific recombination system derived from Saccharomyces cerevisiae 2µm plasmid

2.5 Non-homologous recombination in general

2.1 Homologous recombination in general

Different environmental conditions can induce DNA double-strand breaks (DSBs) in the cell genome. The cell possesses evolutionary developed DNA repair mechanisms, such as non-homologous DNA end-joining (NHEJ), as well as enzymes called telomerases[52] (paragraph 2.5). But, homologous recombination (HR) is able to induce DSBs repair in the cell as well, by that ensuring genome stability[52]. HR happens between DNA regions shearing high percentage of sequence similarity. During this process DNA sequences are integrated or replaced[53] (Figure 2). Thereby, several recombinases catalyze HR in meiosis or in mitosis, but there are also helicases which regulate the HR negatively[52]. In the eukaryotic cells, all proteins that play role in HR are products of *RAD52* group of genes, and are highly conserved[52,54].

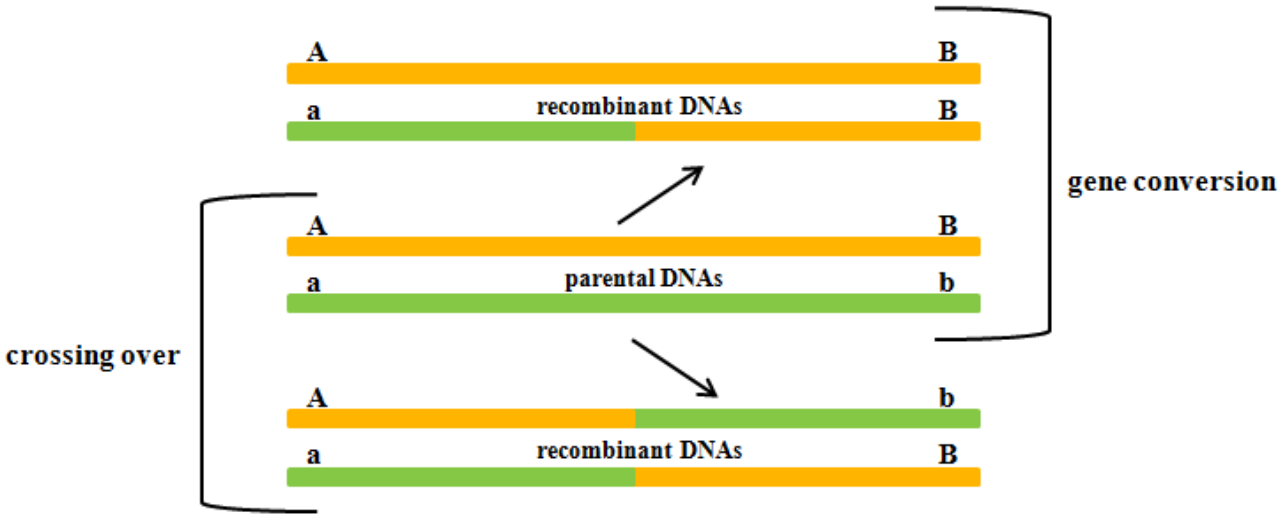

Figure 1:[53] **DNA homologous recombination process.**
Gene replacement ("gene conversion"), and gene exchange ("crossing over") processes are depicted in the upper and down part of the picture, respectively. The orange colored AB-line and the green colored ab-line present two homologous DNA sequences.

After entering the cell and the cell nucleus, homologous DNA sequences initiate a HR cell response[52]. On the one of the DSB sites, DNA is cut by 5'-3' exonucleases and the results are two single-strand 3'-OH DNA overhangs[52,54]. These overhangs are intruded into the homologous DNA and form D-loop[52]. Starting from the 3'-OH end, complement DNA strand is being synthesized, and finally annealed to the D-loop and connected to the second DSB[52]. This is how two Holliday junctions (HJs) are formed, after what both, crossover as well as non-crossover products occure[52]. This process is called double-strand break repair (DSBR)[52] (Figure 2).

Two other HR processes, called synthesis-dependant strand annealing (SDSA) and single-strand annealing (SSA) have also been described[52]. In the SDSA, instead of D-loop, the initial homologous DNA strand anneals to the second DSB, and thereby no HJs are formed[52].

The product of SDSA are always non-crossover sequences[52]. In SSA mechanism, repeated DNA-sequences, located upstream and downstream of DSB join together, by that deleting the sequence from between[52].

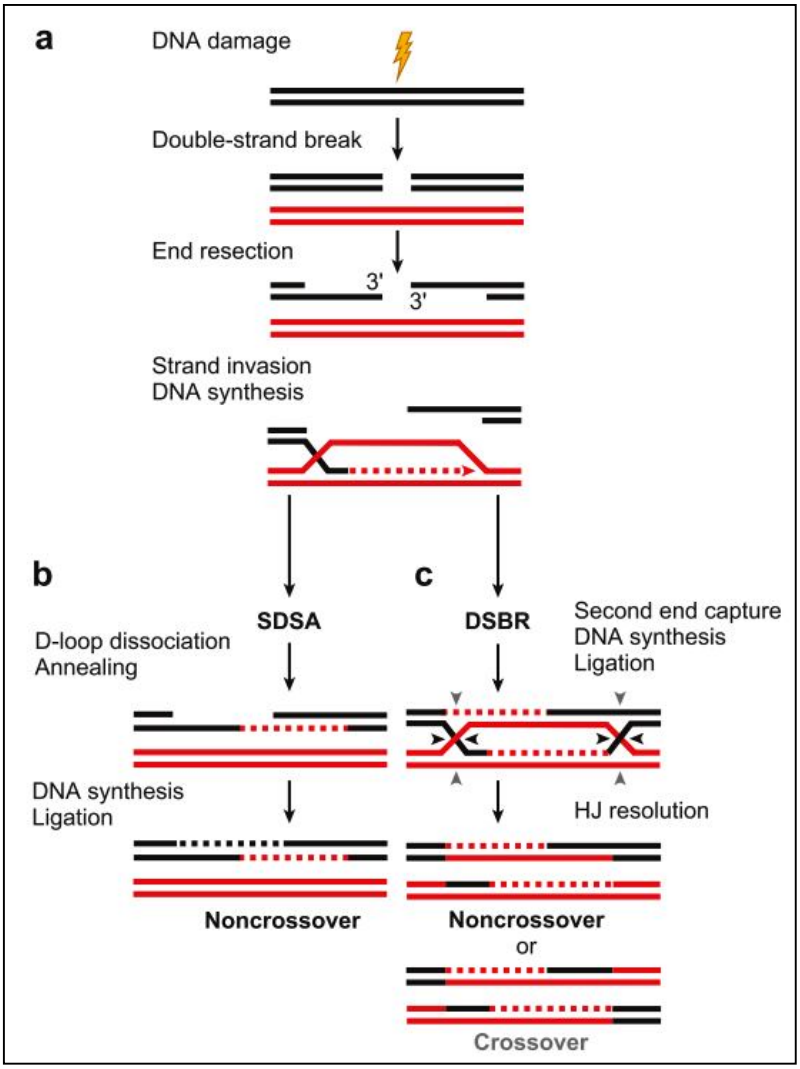

Figure 2:[52] **Pathways of DNA double-strand break repair (DSBR) by homologous recombination (HR).**
a) After double-strand break (DSB) formation, 3' single-strand DNA (ssDNA) overhangs, needed for the HR machinery are formed. Strand invasion followed by DNA synthesis occur subsequent. **b)** Synthesis-dependent strand annealing (SDSA) pathway, where non-crossover reaction product is synthesized. **c)** Double-strand break repair (DSBR) pathway, where crossover or non-crossover reaction product is synthesized.

Nowadays, HR induced by transformation of linear homologous DNA sequence into the cell, is widely used method for *in vivo* gene manipulations, and is also known as targeted insertion mutagenesis[54]. Thereby, linear DNA integration sequences (linear DNA integration cassettes) of two different types exist, and they are termed as "ends-in vector" and "ends-out vector"[54,55] (Figure 3).

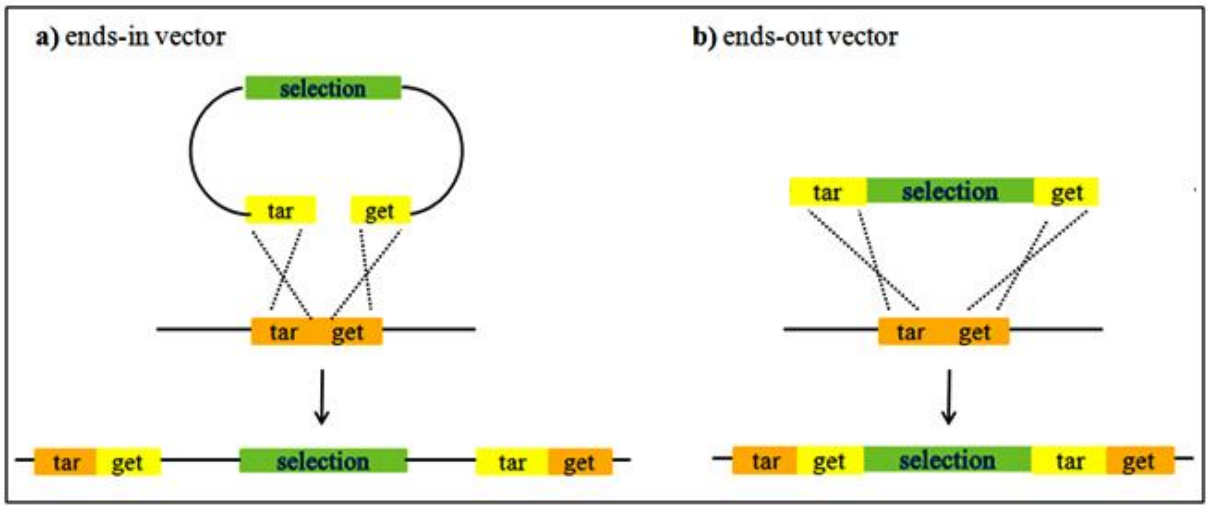

Figure 3:[54] **Two types of linear integration cassettes.**
The green colored part termed as "selection" and yellow colored "target" part are both located on the vector and represent selection gene and homologous sequence, respectively. The orange colored "target" part represents the genome target sequence, which is homologous with the yellow "target". The x-pointed lines depict crossing over event.

During the integration of an ends-in vector, the gap formed in the middle is filled in using the host DNA sequence as a template[54]. Thereby, two crossing over events between the genomic target sequence and respective homologous flanking regions on the vector, result in duplication of the target sequence in the genome[54]. Duplication occurs also during the integration of ends-out vector, but in this case no gap is formed[54]. Designed truncations or mismatches in the flanking ends of both, ends-in and ends-out vectors, could be used for targeted gene disruptions. In general, the ends-out vectors are preferred, because they show higher mitotic stability in comparison with ends-in vectors, and circular plasmids[54].

HR has been first analyzed and defined in the yeast *S. cerevisiae*[52]. For this experiment, linear plasmid containing *S. cerevisiae* homologous DNA sequence has been transformed in the cell, and successfully integrated in yeast genome[52]. Little knowledge about the homologous recombination processes in *P. pastoris* has been obtained from basic molecular studies. Since the classical yeast system *S. cerevisiae* has been analyzed and described in detail, most of the recombination strategies used for its genetic manipulation have been simply adjusted for other yeast, including *P. pastoris*[54].

2.2 Single site homologous recombination in *Pichia pastoris*

A common method of targeted mutagenesis in *P. pastoris* is the gene insertion through single site homologous recombination. With this method, circular and linearized plasmids carrying *P. pastoris* homologous sequences are integrated in the genome. Thereby the resulted transformants (clones) show high mitotic stability, even without selection pressure[56].

Observed at the molecular level, single site homologous recombination includes a crossing over event occurring between plasmid or expression cassette sequence complement to some part of *P. pastoris* genome. In this way, the whole plasmid or expression cassette is integrated in the *P. pastoris* genome at the desired locus (usually *AOX1* locus) (Figure 4). Thereby, the probability of spontaneous achievement of multi copy clones is 1-10%[56]. Besides other factors, the copy number of genome inserted sequences is also defined by DNA sequence size (bp) and amount (µg) used for the transformation. However, although higher DNA amount results usually in higher copy number, it influences the stability and viability of the cell, and therefore the possible copy number is not unlimited.

It is important to be mentioned that in *P. pastoris* single site homologous recombination occurs with much higher probability in comparison with double site homologous recombination and no part of the genome sequence is lost[56].

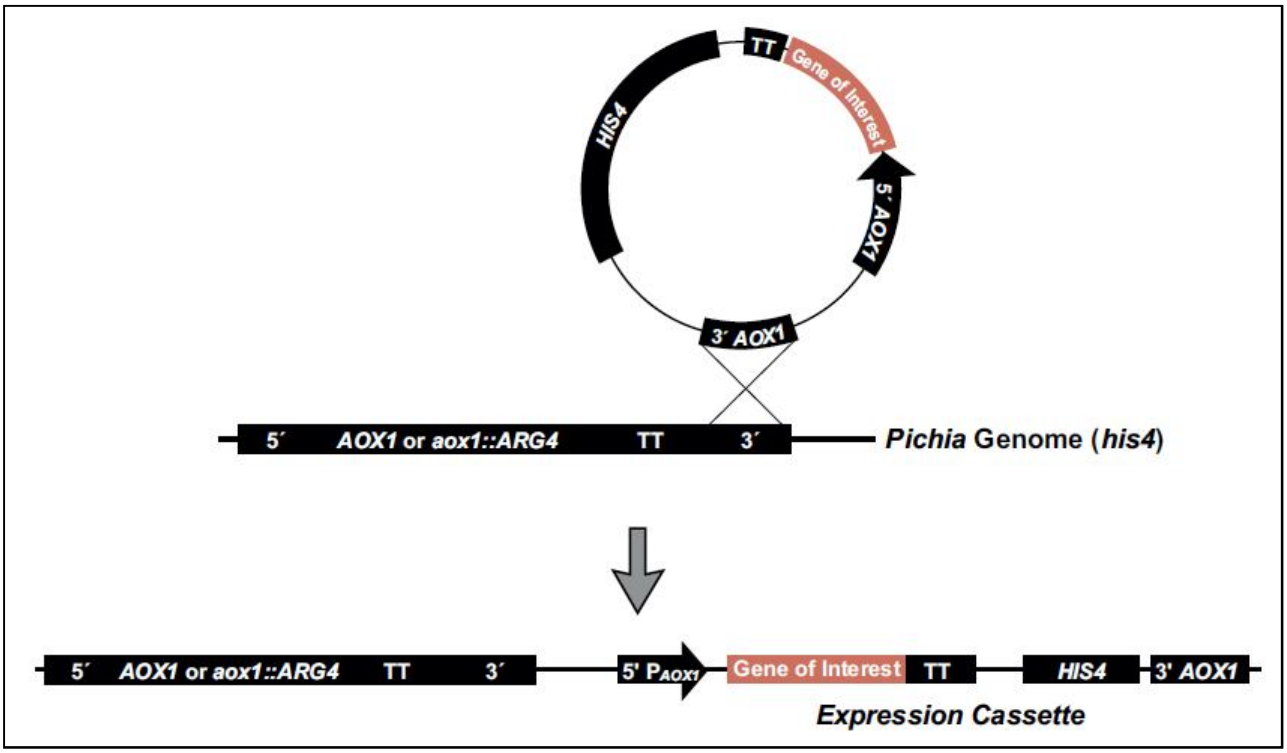

Figure 4:[56] **Single site homologous recombination of linearized plasmid in *P. pastoris*.**
P. pastoris was transformed with circular plasmid, or plasmid linearized in the 3` *AOX1* region, and after a single crossing over event it becomes integrated in the genome at the *AOX1* locus.
(*HIS4* represents a marker gene; crossing over events are marked with ×)

2.3 Double site homologous recombination in *Pichia pastoris*

Double site homologous recombination is another way for genetic manipulation of *P. pastoris*. In this method, *P. pastoris* was transformed with a linear expression cassette carrying two homologous flanking ends. Observed at the molecular level, two crossing over events happen and they result in gene replacement[56] (Paragraph 2.1). Thereby the part of the genome

which ends with sequences complementary to the expression cassette flanking ends, is deleted (Figure 5).

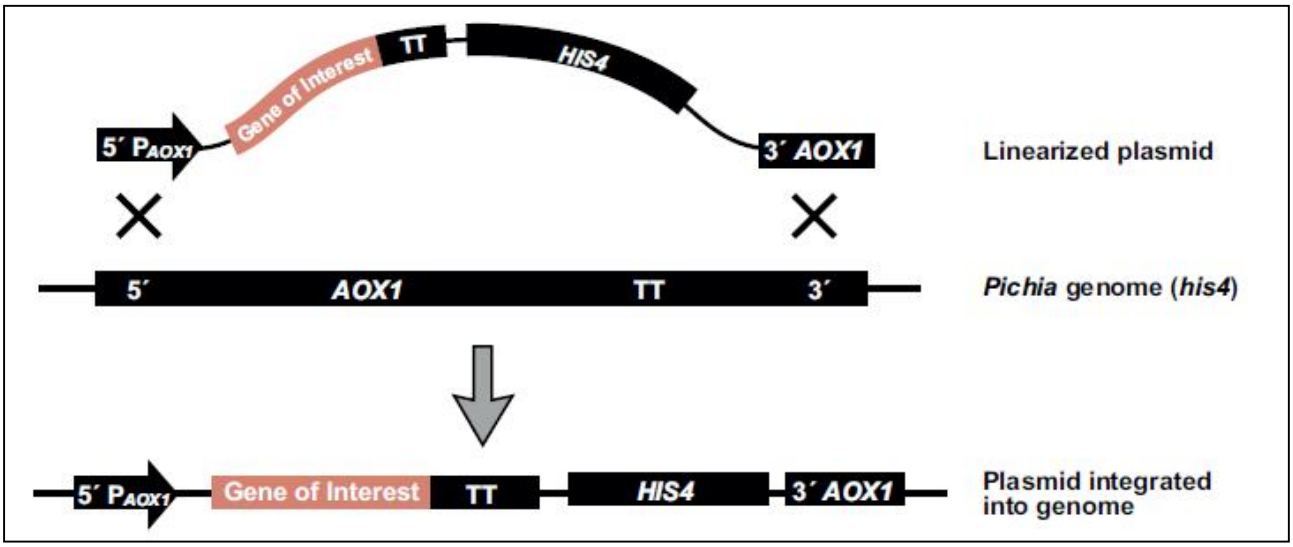

Figure 5:[56] **Double site homologous recombination of linearized plasmid or linear expression cassette in *P. pastoris*.**[*]
After transformation, two crossing over events happen, and the part of the genome located between P_{AOX1} and 3`AOX1 gets deleted and replaced with the gene of interest. This figure depicts the deletion of *PpAOX1* gene which results in a new phenotype, the *P. pastoris* Mut^{s} strain.
(*HIS4* represents a marker gene; crossing over events are marked with ×)

Double site homologous recombination method is widely used for design of gene *knock-out* and *knock-in* strains. Different transformation experiments have shown positive correlation between the length of flanking ends and the transformation efficiency in *P. pastoris*[57] (Chart 1).

[*] This process is also called omega insertion, simply because its' depiction has the shape of the Greek letter omega (Ω).

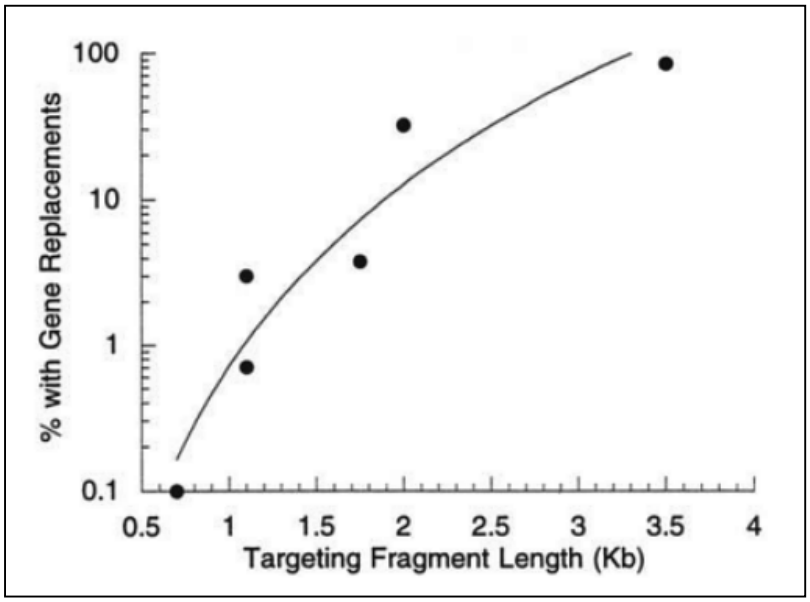

Chart 1:[57] **Correlation between targeting fragment length (flanking ends length) and transformation efficiency in %, in *P. pastoris*.**

2.4 FLP site-specific recombination system derived from *Saccharomyces cerevisiae* 2µm plasmid

Site specific recombination is a process well known from both prokaryotes and eukaryotic organisms[58,59]. It is also called "conservative recombination", because no DNA is synthesized or degraded during this process[60]. Site-specific recombination can be seen as a special type of homologous recombination, because it is independent of the endogenous recombination machinery of the host cell[61].

Different recombinase systems are common in circular, and double stranded yeast plasmids, since they provide the plasmid amplification mechanism[59]. The yeast *Saccharomyces cerevisiae* for instance, synthesizes recombinase termed FLP ("flip"), which is located on the 2µm plasmid in the cell, and between two identical DNA target sequences, termed FRT[58,60,61]. FRT target sequences are each 34 bp long, built of two identical 13 bp regions, which are separated by one 8 bp long region (so called "core region")[60] (Figure 6). This core region is the site of crossover recombination event[60]. The 13 bp flanking regions are rich on AT nucleotides[60]. They can be oriented in the same or in the opposite direction (inverted)[60]. When FRT target sequences lay in the same direction, recombinase cut out the DNA sequence in between, and the so called "reaction of excision" happens[58,61]. In contrast, positioning of the FRT target sequences in opposite direction results in DNA inversion[61].

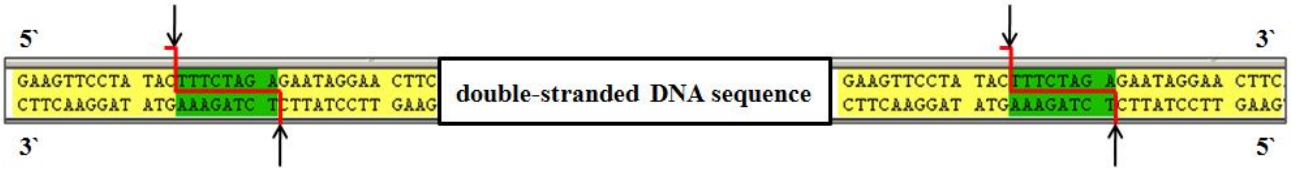

Figure 6: FLP target DNA sequence called FRT.
The two 13 bp long identical flanking regions of each FRT, oriented in the same direction, are colored yellow, and the 8 bp long core regions are colored green. The black arrows show the recombinase restriction sites. The DNA sequence between the red lines is deleted during the recombination.[62]

The catalytic mechanism of FLP protein has been reported almost 25 years ago. During the recombination event FLP brakes the junction on the last nucleotide at the 5'-terminus of the core region[59]. In this way two 8 bp long single stranded DNA nicks are produced[58,62]. At the next reaction step, the recombinase binds covalently to the free 3'-terminus[58,63]. It has been reported that this is a 3'-phosphothyrosine linkage, connecting the DNA and the recombinase[58]. For that reason, FLP recombinase has been classified to the topoisomerase enzyme-family, which includes enzymes able to use reversible DNA-tyrosine bonds as reaction intermediates[58,62]. For its catalytic activity, no co-factors or other proteins are required[61]. It has been assumed that at least two or more FLP molecules bind symmetrically to the both FRT sites and support the cleavage and religation of the DNA strands.[62] The FLP protein is also responsible for the proper DNA strand motions in order to ensure optimal positions for the formation of all four new phosphodiester bonds.[63]

In comparison to the site-specific recombinases, which usually bind to specific DNA fold, FLP is able to bind to supercoiled, relaxed, circular and linear DNA sequences[60]. FLP is also able to catalyze the recombination between intra- as well as intermolecular sequences[60]. Mutational studies provided in order to determine possible hot-spots within the FRT sequence, have shown decrease of FLP activity in all clones caring a mutation in the core region[59]. However one mutation within the flanking ends increased the FLP activity[59].

The FLP recombinase system is well described, and commonly used for insertion, deletion, inversion, as well as translocation of DNA sequences in different organisms, such as bacteria and yeasts, and also in animal cells[59,64,65]. In *P. pastoris*, FLP is a commonly used mechanism for the elimination of prokaryotic sequences and genes of resistance from its genome, and also for gene deletion studies[65].

Another site specific recombination system which was successfully applied in *P. pastoris* was the Cre-lox system.[66] This is a recombinase system of the bacteriophage P1 and its mechanism is quite similar to that of FRT.[63]

2.5 Non-homologous recombination in general

Non-homologous recombination (NHR), also called illegitimate integration, is a process where DNA sequences which share no or very low sequence similarity are recombined[54]. NHR possesses its own protein machinery, and shows a quite different reaction mechanism from those of HR[54].

Additional to NHR, the non-homologous end-joining (HNEJ) represents a cell mechanism for DSBs repair in the genome[54]. One of the proteins detected to contribute to this mechanism is termed as Ku-protein in mammalian cells, or HDF (high affinity DNA-binding

factor) protein in yeast *S. cerevisiae*[67]. Ku is a heterodimer, which binds with high affinity to the linear DNA duplex, but it does not bind to circular DNA[68]. It is built of two subunits called Ku70 and Ku80, and bound to the telomeric DNA[68]. Both monomers make a ring around the DNA, and so Ku participates in HNEJ with its capability to orient the DNA strand in the right direction for further processing and ligation[68]. Before the ligation can start, polymerases fill in the gaps and nucleases trim the excess strands[68].

NHR and HR compete for every introduced foreign DNA sequence in the cell[54]. For this reason, researchers designed strains which are not able to do NHR, in order to ensure higher HR transformation rate. These are usually knock-out strains, such as *P. pastoris* ΔKU70 strain (Näätsaari et al., manuscript submitted).

3. *Introduction to protein disulfide isomerases (PDIs)*

3.1 Protein folding-disulfide bonds (-S-S-) formation

3.2 General description of protein disulfide isomerases

3.3 Protein disulfide isomerases reaction mechanism

3.4 Brief comparison of protein disulfide isomerases from different organisms

3.4.1 Pichia pastoris protein disulfide isomerases

3.1 Protein folding - disulfide bonds (-S-S-) formation

Proteins' native configuration is determined by their amino acid (AA) sequence. Folding processes performed *in vitro* have shown that proteins can fold spontaneously in water solution[22,23]. However, inside the cell these processes are assisted by folding enzymes and helping proteins called chaperones[22]. Chaperons are helping-proteins, which possess a capability to vary the protein folding velocity, and also to control the exactness of disulfide bonds formation[24].

Except for secreted proteins, which are mainly folded in the endoplasmic reticulum (ER) lumen, for all other proteins it is usual to be folded inside the organelle of their residence[22].

The folding process begins on the nascent protein chain, since elongation is slower than folding[22]. This phase is called "co-translational folding" and takes place in the cytosol[22].

For secreted proteins, the folding process continues in ER lumen, since their folding pattern usually includes disulfide bonds between cysteine (Cys) residues. Namely, ER lumen provides a more oxidizing environment and different reaction conditions than the cytosol. It shows approximately neutral pH values, and high concentration of Ca^{2+} and, these conditions favor the disulfide bonds formation[22]. At the beginning of a protein folding process within ER, some of the Cys residues are mispaired[24]. Thus, structures without correct tertiary and quaternary structure, and with high hydrophobic surface are formed[23]. These intermediates form rapidly, possess significant secondary structure, and is called "molten globule" or "compact intermediate"[23,22]. The later reactions in the folding process are slower, and include disulfide bonds formation and isomerisation, reactions catalyzed by protein disulfide isomerases (PDIs)[24,22]. The disulfide bonds are of high importance for secreted proteins, because they provide higher protein stability in extracellular environment[25]. It is important to be mentioned, that also enzymes different than PDIs, are included in protein folding process in ER[22,23]. For instance, peptidyl-prolyl isomerases (PPIs), detected in both, prokaryotic and eukaryotic cells, catalyze *cis-trans* isomerisation of peptide bond N-terminal to proline (Pro) residues, and also play a role of chaperones in the cell[26,23]. Also sugar moiety transferases, which predominantly play a role in protein *N*-glycosylation, are important for correct protein folding[22]. In particular, they support proper positioning of defined protein domains, and favor or disfavor certain hydrophilic and hydrophobic interactions in ER[22] (Chapter 1.3).

Protein folding is finished by protein subunits rearrangement and assembling of monomers in homo- or heterooligomers, which is a concentration-dependent process[22].

Different steps of the secretory pathway in *P. pastoris* have been supposed to become bottlenecks of overexpression of heterologous proteins. One speculation is that in some cases ER is overburdened by the huge amount of proteins inside, which results in production of misfolded proteins, and induces cell stress response[24].

In general, there are two non-natural cell states, the so called metabolic and environmental stress[27]. The metabolic stress is caused by cell abuse during heterologous protein expression[27]. In cases where no natural protein conformation is reached inside ER, the cell activates endoplasmic reticulum associated protein degradation (ERAD) mechanism,

transporting the protein back to the cytosol and degrading it[27]. Thus ERAD can be seen as typical metabolic stress response of the cell[27]. The environmental stress is caused by fermentation conditions, which can vary in comparison to the natural ones[27].

The protein quality control (also called ‘’architectural editing’’) provided by protein PDIs, as well as ERAD, is as well important to be mentioned since it guaranties that only proteins with proper configuration will enter the secretion downstream steps[22].

3.2 General description of protein disulfide isomerases

Almost 40 years after their discovery, PDIs still occupy researcher’s attention[28]. Together with glutaredoxin (Grx), they belong to the thioredoxin (Trx) superfamily of thiol/disulfide exchange enzymes, which are characterized by a thioredoxin fold and -CXXC/S- active site motif[29,30]. This thioredoxin motif contains two Cys residues, which shuttle between dithiol and disulfide form, during catalysis of the redox reaction[30,31]. The main difference between Grx and PDI is their localization inside the cell. Grx is predominantly localized in the cytoplasm, but it has also been found in mitochondria and nucleus[29]. In contrast, PDIs are soluble proteins found in ER lumen[29,28].

A cell targets its ER resident soluble proteins to the right cell compartment using different targeting sequences. In this way it determines which of the ER resident proteins retain in the lumen and which have to be retrieved from the downstream secretory organelles[32]. For instance, the C-terminally located tetrapeptide (-K/HDEL) causes protein retrieval from Golgi apparatus (Golgi)[32]. These sequences are recognized by membrane receptor proteins of the organelles[32]. It is known that also Ca^{2+} concentration plays a role in the ER protein retention mechanism[32].

PDIs contained in the ER catalyze the oxidation of thiol groups between Cys residues in proteins, but they also play a role of chaperones[28]. Some disulfide bonds form spontaniously, and mispaired Cys residues, as well as early misfolded proteins are produced[24]. Depending on the redox conditions, PDIs can reduce already formed disulfide bonds in proteins, thus performing folding pattern controlling processes, and protein isomerisation[24]. This capability of PDIs comes from the misfolded proteins´ hydrophobic surface[28] (Figure 7). Namely, early misfolded proteins have mainly hydrophobic surfaces[28]. Since PDIs surface is also mostly hydrophobic, driven by the favorable hydrophobic interactions, reduction and isomerisation can easily happen[28]. In addition, quite flexible PDI conformation enables proteins with different size to be properly refolded[28]. Recent studies have shown that different PDIs in the cell posses different product specificity, which can be seen as an explanation of the cell need of synthesizing different PDIs[31].

Crystallographic studies of yeast PDIs have shown that they contain four thioredoxin domains (**a**,**a’**,**b**, and **b’**) and an acidic C-terminus, which carries an ER-retention signal (-K/HDEL)[28,32]. In particular, two thioredoxin domains (**a** and **a’**) show catalytic activity, but two of them (**b** and **b’**) are catalytically inactive[28]. Both active sites are placed opposite to each other, separated by 28Å long hydrophobic area[28] (Figure 7). Even structurally very similar, these two domains (**a** and **a’**) differ from each other by their redox state[28]. Namely, cysteine residues inside the active center of the **a** domain are connected with each other by disulfide bond, thus they are oxidised[28]. In contrast, cysteine residues inside the active center of the **a’** domain are in reduced state[28]. This feature is crucial for the redox activity of PDIs.

Another crucial feature is the different orientation of both Cys residues within one thioredoxin active center of PDI[29]. The Cys residue nearer to the PDI N-terminus (-**C**XXC-) is oriented toward enzyme surface, and reacts as nucleophile[29]. The second Cys residue (-CXX**C**-) is located deeper inside the active center, thus being able to react with the N-terminus Cys (-**C**XXC-) only[29]. The nature of two amino acids staying between the active Cys residues (-C**XX**C-), and the amino acid sequence surrounding the active center are also factors having an influence on thioredoxin redox potential[29]; enzymes containing thioredoxin are able to catalyze protein folding by extremely different redox potential[29].

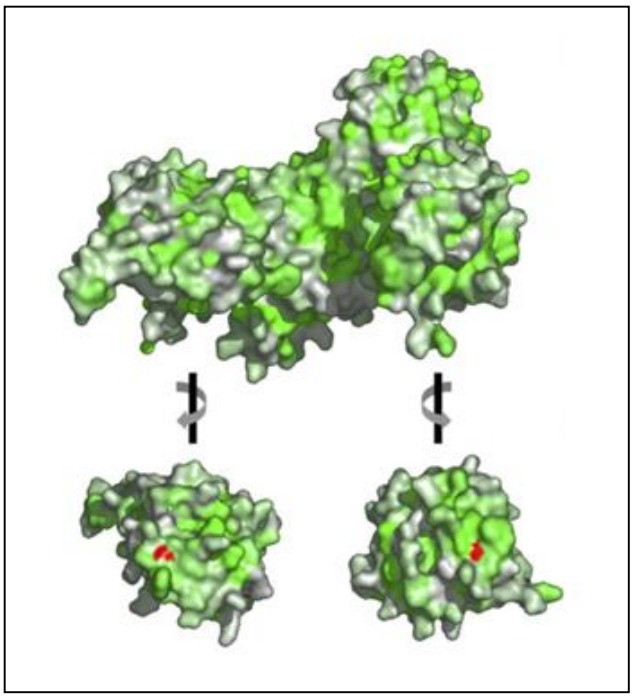

Figure 7: [28] **Depiction of hydrophobicity and hydrophilicity distribution on the PDI surface.** According to the normalized consensus hydrophobicity scale of the exposed residues, the green marked regions are hydrophobic and gray regions are hydrophilic. Down left and right show the **a-** and **a'-** domain, respectively. Their active centers are colored in red.

Expression of heterologous proteins with a high number of Cys residues can be hindered by a folding bottleneck in the ER: thus, the more dithiols to be oxidized, the higher PDI concentrations are required[24].

3.3 Protein disulfide isomerases reaction mechanism

In the ER lumen, glutathione (GSH) contributes to reaching the optimal redox conditions by counterbalancing the flow of oxidizing equivalents through ER membrane-associated oxidoreductases[33] (Figure 8). The oxidoreductases are included in direct disulfide bonds transfer to ER PDIs[33]. Further, PDIs are responsible for the transfer of oxidizing equivalents derived from the ER membrane proteins, to the secretory protein (substrate protein)[29].

PDI catalyses oxidation by performing a disulfide exchange reaction between thioredoxin Cys residues from the PDI active center, and Cys residues located inside the substrate protein[31] (Figure 9).

Following formula explains the chemical reaction steps of thiol-disulfide exchange reaction[34]:

$$\mathbf{R_1S^-} + \mathbf{R_2SSR_3} \rightarrow \mathbf{R_2S^-} + \mathbf{R_1SSR_3}$$

$\mathbf{R_xS^-}$ represents an thiolate anion, and $\mathbf{R_xSSR_y}$ disulfide bound amino acid residues[34]. Thiol-disulfide exchange reaction can happen between Cys residues of different proteins, but also within the same protein, which is then called disulfide reshuffling[34]. For every disulfide bond formation, two thiol-disulfide reactions happen, where after the first one, a mixed disulfide intermediate is formed, between the protein and the redox equivalent[34], thus a distance between the Cys residues influence the rate of this reaction[34]. This reaction depends furthermore on pH, electrostatic environmental conditions, and p*Ka* of the thiol groups[34]. Disulfide bonds destabilize the nascent protein chain by imposing distance and angle between C^{β} and S^{γ} atoms of the joined cysteine residues, and by that reduce their entropy[34]. They also support stabilization of folded proteins by enhancement of interactions between hydrophobic protein domains[34].

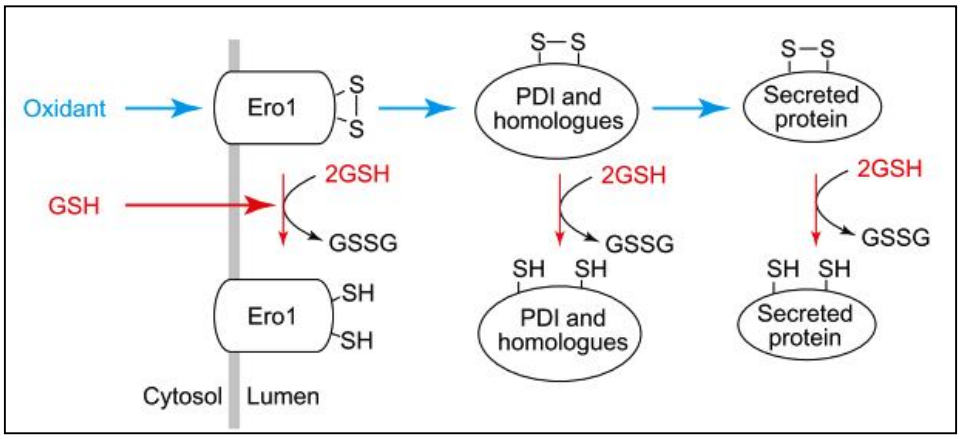

Figure 8:[33] **Redox balance in the lumen of endoplasmic reticulum.**
Ero1 is an ER membrane bound protein responsible for the transfer of oxidizing equivalents from the cytosol to PDI in the ER lumen. GSH = reduced glutathione; GSSG = oxidized glutathione;

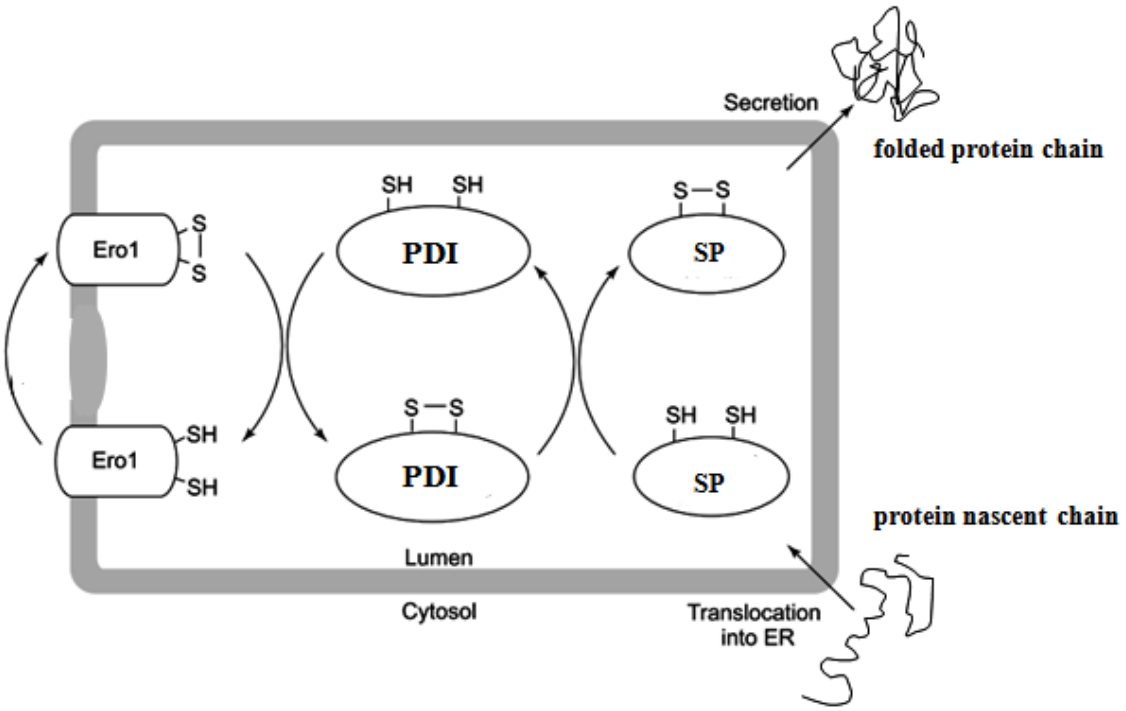

Figure 9:[33] **Depiction of disulfide bond formation in the ER lumen.**
Ero1 is ER membrane bound protein responsible for transfer of oxidizing equivalents from cytosol to PDI in ER lumen. The figure shows only the thioredoxin domains of PDI. Disulfide bond forms by thiol-disulfide exchange reaction between PDI and the substrate protein (SP).

3.4 Brief comparison of protein disulfide isomerases from different organisms

Although enzymes responsible for disulfide bonds formation occur in all organisms, depending on the cell type their main features vary drastically. For instance, these enzymes in bacterial cells are located in the cytosol. Two bacterial enzyme-systems, DsbA-DsbB and DcbC-DsbD are responsible for secreted protein folding and isomerisation, respectively[25]. In contrast, in yeast cells, both, oxidation and isomerisation are a performed by PDIs[25].

In the yeast *Saccharomyces cerevisiae*, besides *PDI1*, four other homologous genes coding for ER PDIs have been detected: *MPD1*, *MPD2*, *EUG1*, and *EPS1*[35,25]. Since there is an evidence that *PDI1* deletion is lethal for *S. cerevisiae*, further deletion studies have been provided in order to define the functions of *Sc*PDI1homologs[35,36]. The study by Nørgaard, P. et al. has shown that only *MPD1*, has the same capability as *PDI1*, being able to catalyze all folding reactions essential for the cell[35]. In *S. cerevisiae*, an active site motif containing only one Cys residue (monothiol; -**C**GFS-) has also been identified[30].

The short comparison of mammalian and yeast PDIs shows a difference in the location of the two non-active Cys residues: in mammalian cells they are located in the **b'** domain, while in yeast cells in the **a** domain[25].

3.4.1 *Pichia pastoris* protein disulfide isomerases

The first *P. pastoris* PDI (*[*]*Pp*PDI), analyzed and reported in 2001, includes two double-cysteine redox-active sites, and belongs to the thioredoxin superfamily of protein-thiol oxidoreductases[36]. Within 517 AA long *Pp*PDI, two redox-active domains (-CGHC-) have been detected, separated from each other by 338 AA long domain[36]. Also two potential glycosylation sites, one N-terminal putative signal sequence, and C-terminal ER-retrieval sequence have been detected[36] (Figure 10).

Figure 10:[36] **DNA and AA sequences showing the most important domains within *Pp*PDI1.**

```
 661 CTTCTTCTTCTTCCTATCTATAAGAatgcaattcaactgggatattaaaactgtggcaag 720
   1                          M  Q  F  N  W  D  I  K  T  V  A  S  12
 721 tattttgtccgctctcacactagcacaagcaagtgatcaggaggctattgctccagagga 780
  13  I  L  S  A  L  T  L  A  Q  A  S  D  Q  E  A  I  A  P  E  D  32
 781 ctctcatgtcgtcaaattgactgaagccacttttgagtctttcatcaccagtaatcctca 840
  33  S  H  V  V  K  L  T  E  A  T  F  E  S  F  I  T  S  N  P  H  52
```

a) Potential signal sequence (underlined domain).

```
 841 cgttttggcagagtttttgccccttggtgtggtcactgtaagaagttgggccctgaact 900
  53  V  L  A  E  F  F  A  P  W  C  G  H  C  K  K  L  G  P  E  L  72
```

b) The first active center domain (-APWCGHCH-), containig the conserved thioredoxin motif (-CGHC-) is located between 59 and 66 AA.

```
1081 tgtcagctatatgctaaagcagagtttacccccctgtcagtgaaatcaatgcaaccaaaga 1140
 133  V  S  Y  M  L  K  Q  S  L  P  P  V  S  E  I  N  A  T  K  D  152
1141 tttagacgacacaatcgccgaggcaaaagagcccgtgattgtgcaagtactaccggaaga 1200
 153  L  D  D  T  I  A  E  A  K  E  P  V  I  V  Q  V  L  P  E  D  172
1201 tgcatccaacttggaatctaacaccacattttacggagttgccggtactctcagagagaa 1260
 173  A  S  N  L  E  S  N  T  T  F  Y  G  V  A  G  T  L  R  E  K  192
1261 attcacttttgtctccactaagtctactgattatgccaaaaaatacactagcgactcgac 1320
 193  F  T  F  V  S  T  K  S  T  D  Y  A  K  K  Y  T  S  D  S  T  212
```

c) Two potential glycosilation sites (*N*-148 and *N*-179).

```
1861 taaagatgttctagtcaagtactacgccccttggtgtggtcactgtaagagaatggctcc 1920
 393  K  D  V  L  V  K  Y  Y  A  P  W  C  G  H  C  K  R  M  A  P  412
```

d) The second active center domain (-APWCGHCH-) containig the conserved thioredoxin motif (-CGHC-) is located between 401 and 408 AA.

```
2161 agcactcagaccagtcgaggaagaaaaggaagctgaagaagaagctgaaagtgaggcaga 2220
 493  A  L  R  P  V  E  E  E  K  E  A  E  E  E  A  E  S  E  A  D  512
2221 cgctcacgacgagctttaaGAATGTCCGGGGTTTCTATATTTACTGTAACTAGGTTTTAT 2280
 513  A  H  D  E  L  *                                           517
```

e) Yeast tipical ER retriveal sequence (-HDEL).

[*] An abbreviation *Pp* will be used for *Pichia pastoris* from now on.

Since an absolutely correct folding of heterologous proteins is of crucial relevance for all therapeutic proteins (biopharmaceuticals and biosimilars) and for activity of most disulfide bridge containing enzymes, many case studies have been enquiring how PDI influences the overexpression of heterologous proteins in *P. pastoris.*

For instance, ancrod (a snake venom serine protease) has been expressed in *P. pastoris* by co-overexpression of *Pp*PDI1, and the results have shown a twofold increase of the secretion during the bioreactor cultivation[37]. In this case, *Pp*PDI1 in *P. pastoris* genome was under the control of the constitutive *Pp* P_{GAP}[37]. In another case study, a different strategy for co-overexpression of *PpPDI* gene has been used. Heterologous expression of the *Necator americanus* secretory protein (Na-ASP1) in *P. pastoris* has been tested, simultaneously co-overexpressing the *PpPDI* gene which was integrated with varying copy numbers[24]. Within this study, a positive correlation between PDI copy number and secreted Na-ASP1 has been observed[24].

4. Introduction to protein glycosylation

4.1 Posttranslational protein glycosylation

4.2 Brief overview of Pichia pastoris N-glycosylation pathway

4.3 Pichia pastoris strain containing human-like N-glycosylation pattern

4.1 Posttranslational protein glycosylation

Protein glycosylation is the most common protein modification[38]. It has been found in both prokaryotes and eukaryotes. Within the eukaryotic cells it is a highly conserved pathway[39]. All carbohydrates are covalently bound to the proteins, and different proteins in the cell are characterized by certain glycosylation pattern[38]. Protein glycosylation is of high relevance for the cell: it enables proper protein folding, which is necessary for achieving proper protein activity[15], and all properties of the proteins, such as size, conformation, surface solubility and charge, are influenced by the oligosaccharides contained in their glycosylation pattern[38].

There are two types of protein glycosylation in the cell, *O*- and *N*-glycosylation. *O*-glycosylation is characterized by a covalent bond between *N*-acetylgalactosamine (GalNAc) and threonine (Thr; T), or serine (Ser; S) hydroxyl (-OH) group[40,38]. *N*-glycosylation, in contrast, is characterized by a covalent bond between *N*-acetylglucosamine (GlcNAc) and specific asparagine (Asn; N) amino acids in the protein chain[15,38]. Asn-Xaa*-Thr/Ser (-N-X-T/S-) is the amino acid consensus sequence for *N*-glycosylation in yeast[40], with the carbohydrate binding to the amide nitrogen of Asn[40].

There are three main classes of *N*-glycosylated proteins in eukaryotes: complex, high mannosylated, and hybrid. They all share the same pentasaccharide structure, with two GlcNAc and three mannose (Man) moieties, which is called "core part"[38]. The rest of the polysaccharide chain is called "antenna", and it differs among all three groups of *N*-glycoproteins[38] (Figure 11). Complex *N*-glycoproteins can contain more antennas, and additional GlcNAc, galactose and sialic acid, but they do not contain more than those three Man moieties, located in the core part[38]. Highly mannosylated as hyper mannosylated *N*-glycoproteins, contain a high number of Man moieties, but they do not contain any other carbohydrate in their antenna part[38]. In contrast, hybrid *N*-glycoproteins can contain more Man residues, but also other carbohydrate moieties within their antenna[38].

Although the glycosylation patterns differ between different cell types, the *N*-glycosylation pathway always takes place in the same cell organelles, namely in ER and in Golgi. More precisely, it takes place on the both sides of ER, and is catalyzed by cytoplasm-oriented and lumen-oriented glycosyltransferases[39].

The core part synthesis starts on the cytosolic site of the ER membrane (ER surface)[39]. Its initial reaction is the assembly of N-acetyl-glucosamine-phosphate and a membrane-bound lipid career molecule, called dolichyl monophosphate (Dol**-P)[39]. The resulted product, *N*-acetylglucosamine- pyrophosphatidyldolichol (GlcNAc-PP-Dol) is further changed by the addition of one N-acetylglucosamine (GlcNAc) and five Man residues, and a heptasaccharide intermediate, $Man_5GlcNAc_2$-PP-Dol is formed[39]. Before further reaction steps occur, $Man_5GlcNAc_2$-PP-Dol is transferred into the ER lumen, where progressive addition of Man

* (Xaa; X) – means that theoretically any amino acid can stay in the middle of the consensus sequence, but frequently found residues are proline and aspartate[38].

** Dolichol (Dol) is an polyisoprenol, in yeast cells commonly built of 14 isoprene units[39].

and glucose (Glc) residues results in $Glc_3Man_9GlcNAc_2$-PP-Dol[39]. In the final step of the core synthesis, $Glc_3Man_9GlcNAc_2$ is covalently bound to a specific *N*-residue of nascent protein chain, by catalytic support of oligosaccharide transferase comlex[38,39]. Soon after that, three Glc and one Man residues are trimmed, and the resulting $Man_8GlcNAc_2$ polysaccharide is transported to the Golgi[41]. Up to the point where $Man_8GlcNAc_2$ enters the Golgi, all glycosylation steps are the same in yeast and mammalian cells. However, further *N*-glycosylation steps, within the Golgi, differ drastically between yeast and mammalians.

Mammalian glycoproteins are mainly of complex type. After entering the Golgi, $Man_8GlcNAc_2$ is trimmed by several α-1,6 mannosidases to $Man_5GlcNAc_2$, whose core enables addition of different carbohydrates and further branching[42,43]. In the next step, the chain is extended by addition of ß-1,2-linked GlcNAc, a reaction catalyzed by *N*-acetylglucosaminyl transferase I (GnT-I)[43,42]. After that, two Man residues are eliminated by mannosidase II (Man-II), and second ß-1,2-linked GlcNAc is added by GnT-II[42,43]. The further chain extension includes addition of galactose and sialic acid, reactions which are catalyzed by galactosyltransferases and sialyltransferases, respectively[42,43].

Yeast cells, in contrast, do not terminate the glycosylation with the sialic acid[44]. Instead of that, a negative charge of the N-glycan is achieved by addition of phosphate, glucuronic acid or pyruvate[44]. They synthesize mainly high mannosylated (hypermannosylated) proteins, which can also contain Gal residues[44]. One yeast species, known to produce hybrid-proteins, by addition of extra GlcNAc moieties is *Kluyveromyces lactis*[44].

Because of that, metabolic engineering is necessary to be performed in order to adapt yeast cells for production of proteins with humanized glycosylation for therapeutic application (Paragraph 4.3).

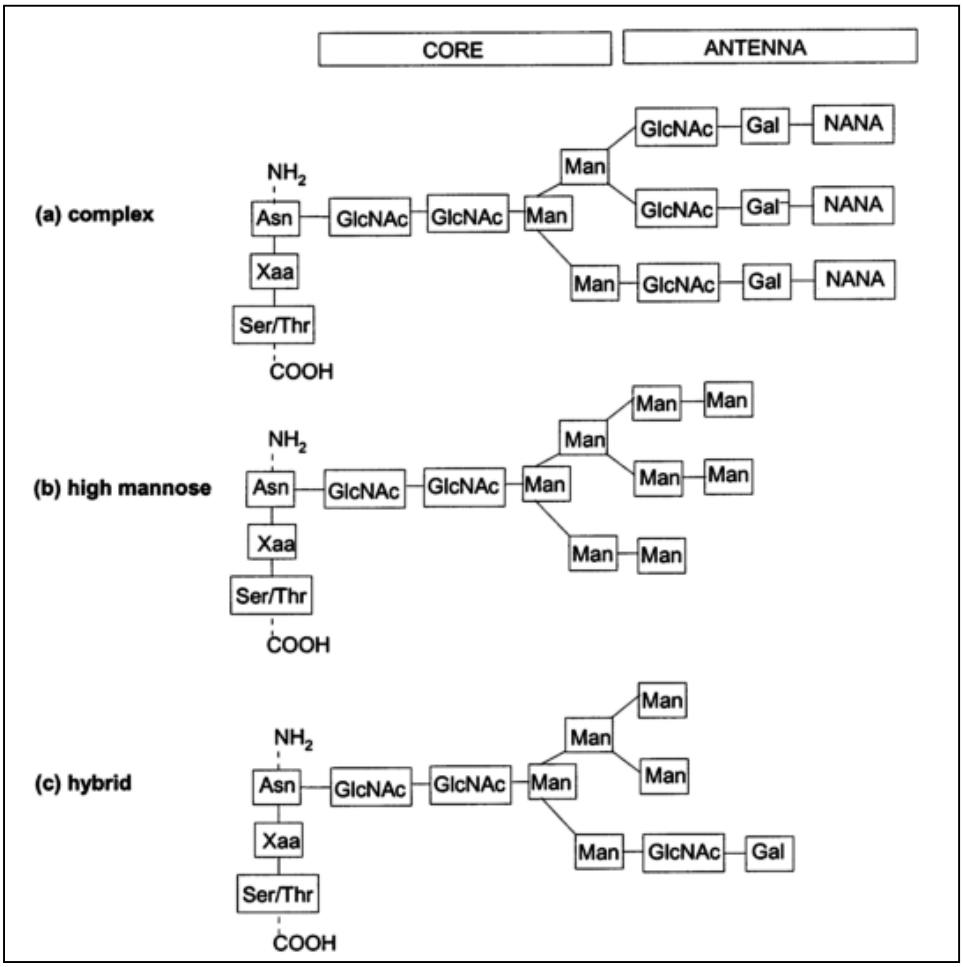

Figure 11:[38] **Main glycosylation classes.**
The common part for all glycosylated proteins is termed CORE, and the protein specific patterns are termed ANTENNA. **a) Complex –** contains only Man moieties. **b) High mannose –** contains high number of Man moieties, but no other carbohydrates. **c) Hybrid –** can contain more than three Man moieties, and also other carbohydrates.

4.2 Brief overview of *Pichia pastoris* N-glycosylation pattern

As already mentioned in this thesis, the capability of *P. pastoris* to secret heterologous proteins is of high relevance for its industrial use. It makes downstream processing of expressed heterologous proteins easier and cheaper, and it ensures protein glycosylation, since ER and Golgi, where protein glycosylation generally happens, are parts of the secretory pathway[15].

As well as many other pathways in yeasts, also *N*-glycosylation has been first described for *S. cerevisiae*. This knowledge has been used later as comparison basis in studies analyzing the common glycosylation patterns in *P. pastoris*.

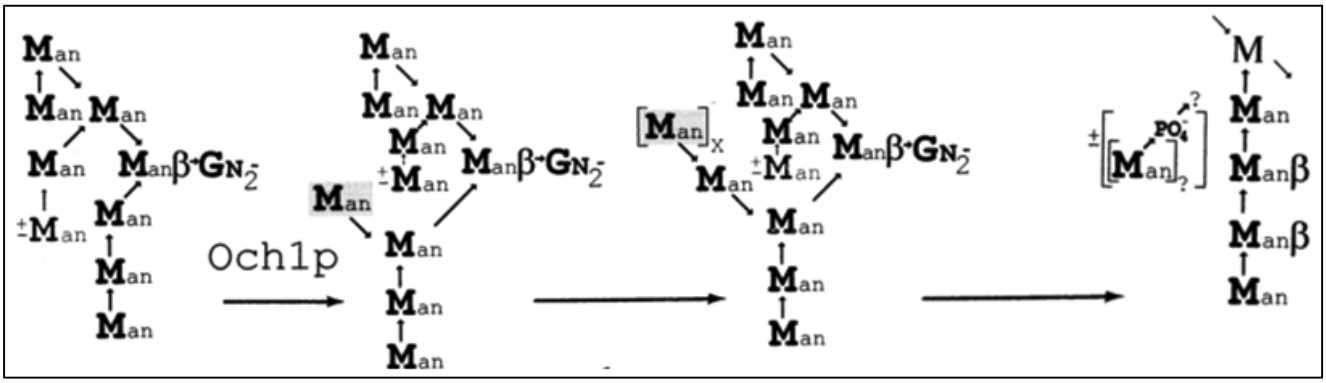

Figure 12:[44] ***N*-linked outer chain extension in *P. pastoris*.**

As already said in the last paragraph, when secreted proteins leave ER, they are "half glycosylated" carrying $Man_8GlcNAc_2$ polysaccharide, which is a precursor for their further glycol-modifications.

The glycosylation pathway continues within Golgi in *P. pastoris* by α-1,6 mannose extension[41]. Thereby, the initial step for the hypermannosylation is catalyzed by an α-1,6 mannosyltransferase, encoded by *PpOCH1* gene[41,42] (Figure 12). It is a transferring reaction of Man residue from GDP-Man to the α-1,3-linked core Man residue on the lower arm of the $Man_8GlcNAc_2$[41,44]. α-1,6 mannose extension formed by this reaction represents a substrate for further Man residues addition, catalyzed by different transferases in *P. pastoris*[41].

Interesting to be mentioned is that phosphate residues have been detected in many of the *P. pastoris N*-glycoproteins[44]. It has been found that the *PNO1*gene is responsible for phosphomannosylation of *N*-glycoproteins and that it promotes the glycosylation of core residues, but not of the outer sugar chain[8].

4.3 *Pichia pastoris* strain containing human-like glycosylation pattern

Glycosylation plays an important role in cell and tissue distribution of proteins as well as in blood clearance in humans[42]. As mentioned above, the human *N*-glycosylation pattern is characterized by a terminal sialic acid residue[42]. Pharmaceutical kinetic studies have shown that glycoproteins having galactose, *N*-acetylglucosamine or Man terminal residues persist shorter in blood then glycoproteins with terminal sialic acids[42]. On the other hand, yeast glycoproteins can contain phosphate residues, ß-mannose or α-1,3 mannose, which activate immunogenic reactions in human cells[42]. For instance, there are three genes (*MNN1*, *MNT2* and *MNT3*) coding for α-1,3-mannosyltransferases in *S. cerevisiae*.[44,45,46] *ScMNN1* is a Golgi located protein, and takes part in both *N*-glycosylation and *O*-glycosylation, while *ScMNT2* and *ScMNT3* are included only in *O*-glycosylation.* (http://www.ncbi.nlm.nih.gov/) The *N*-linked oligosaccharides added to *P. pastoris* proteins vary in size, having structures from $Man_8GlcNAc_2$ to $Man_{14}GlcNAc_2$, and proteins with more than 30 Man residues have been

* Information source: http://www.ncbi.nlm.nih.gov/.

detected[47,15]. However there is no evidence of alpha-1,3 terminal linked mannose moieties in *P. pastoris* proteins, and this feature favored the use of *P. pastoris* over *S. cerevisiae* for the industrial production of pharmaceutical proteins.[48,40,49] More precisely, *P. pastoris* contains only one gene coding for α-1,3-mannosyltransferase, which is called *RHK1* and is thought to be included in the core part synthesis in the endoplasmic reticulum.[**]

For all these reasons Chinese hamster ovary (CHO) cells, have been proven to be more suitable for production of therapeutic glycoproteins[41]. On the other hand, the fact that yeast cells are easier for handling in bioreactors indicates the importance and necessity of heterologous expressed therapeutic proteins to be human-like glycosylated. For that reason, in parallel to the successfully increased yields of heterologous proteins produced in *P. pastoris*, scientists have also concentrated on quality improvements of synthesized proteins. They have opened a new applied research topic, trying to design yeast cells with human-type *N*-glycosylation pathway. In the last decade a research group of Gerngross et al. designed a *P. pastoris* strain containing a fully humanized *N*-glycosylation pathway.

Since the *P. pastoris* glycosylation machinery produces hypermannosylated proteins containing up to 40 Man residues, the first engineering step consisted in trying to stop further addition of Man residues[41, 42]. With this aim, inactivation of *PpOCH1* gene by gene disruption has been done similar to the strategy applied for *S. cerevisiae*.[42,50]. In the second step, an HDEL-tagged α-1,2 mannosidase from the filamentous fungi *T. reesei*, has been overexpressed[43,42]. This enzyme is able to eliminate terminal α-1,2 mannose residues, and the reaction results in formation of $Man_5GlcNAc_2$[42,43]. Further glycosylation of $Man_5GlcNAc_2$ seems to be more difficult task for the cell, because the most *P. pastoris* endogenous glycosyltransferases are not able to use this *N*-glycoprotein as a substrate[43]. For this reason, in the next step a gene for uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) transporter has been integrated in *P. pastoris* genome, thus producing for the first time hybrid glycoproteins ($GlcNAcMan_5GlcNAc_2$) in this yeast[41]. Within the further steps, mannosidase II and *N*-acetylglucosaminyl transferase II (GnT-II) genes have been introduced, and the endogenous gene *ALG3* has been deleted, which resulted in the formation of first complex glycoproteins ($GlcNAc_2Man_3GlcNAc_2$) in *P. pastoris*[41].

The most difficult engineering step has been the introduction of all genes responsible for synthesis and addition of sialic acid in *P. pastoris*[51]. Generally, the following features make yeasts unable for sialylation:

1. they are not able to synthesize ß-1,4 galactose terminated substrates,
2. they are not able to synthesize activated sugar cystidine monophosphate (CMP)-sialic acid,
3. they are not able to synthesize transport molecule for CMP-sialic acid,
4. they are not able to synthesize sialyltransferase, needed for the final step where CMP-sialic acid is added to galactose[51].

However, based on earlier research work, Hamilton et al. have designed the first *P. pastoris* strain capable of synthesizing complex sialylated glycoproteins[41]. Engineered *P. pastoris* strains and plasmids for heterologous protein production with fully adjusted human *N*-glycosylation pattern are already available made by the new patented technology.

[**] Information source: https://gendb.cebitec.uni-bielefeld.de/

5 Objectives

The industrial use of *P. pastoris* for production of pharmaceutical proteins shows a continuously increasing course in the last two decades. In parallel, the number of research groups working with this methylotroph grows as well. Thereby almost every improvement is patented, and thus not of common usage.

In contrast, the results of this master thesis were planed as a contribution to the *P. pastoris* strain collection (*P. pastoris* pool), established at the Institute of Molecular Biotechnology (IMBT) at Graz University of Technology (TUG). The basic idea of the *P. pastoris* pool was to make collection of patent free strains with adjusted industrially relevant features. With this aim, both research topics within this master thesis were chosen according to actual lacks in the biotechnological protein production using *P. pastoris* as production host.

The first research project was focused on *P. pastoris* protein disulfide isomerases (*Pp*PDIs). *PpPDI1* gene has been patented since more than fifteen years. However, the recent *P. pastoris* sequencing study, done at our institute, indicated the existence of three additional genes coding for enzymes with potential protein disulfide isomerase catalytic activity.

PDIs play a role in protein folding processes in the cell (Chapter 1.2). Since proper protein folding is of crucial relevance for the protein activity, one of the possibilities to avoid misfolding of heterologous proteins is the co-overexpression of *PDIs* in the cell. For that reason, the aim of this thesis was the design of *PpPDI*-specific P_{AOX1} -replacement cassettes. Also single and multi copy *P. pastoris* strains carrying *CalB* and *HRP* genes should be designed as reporter strains within this master thesis. The four newly designed recombination cassettes, as well as *P. pastoris* strains, should be used in the future for further research work. My final aim was to replace the *PpPDI* natural promoters with the strong *AOX1* promoter (P_{AOX1}) in order to overexpress the same. Disulfide bonds contained in CalB (*Candida antartica* lipase B) and HRP (horseradish peroxidase) make them suitable model proteins, for testing the influence certain *PDI* co-overexpression has on their folding, since it was known from previous experiments that PDI over-expression enhances active enzyme production.

Within the second research project, my interest was focused on *P. pastoris* genes coding for α-1,6 mannosyltransferase. Namely, the first known gene coding for this enzyme in *P. pastoris* was *OCH1*. Disruptions of this gene resulted in non-hypermannosylated secreted proteins, suitable for additional glycosylation manipulations (Chapter 1.3). However, there is one more gene supposed to code for an enzyme with α-1,6 mannosyltransferase activity in *P. pastoris*. Its sequence was automatically annotated as *HOC1* within the recent *P. pastoris* sequencing study, but nothing is yet known about its structure and activity. Therefore, the aim of the second project was the design of *P. pastoris OCH1* and *HOC1* knock-out strains. For this reason, *OCH1* and *HOC1* specific knock-out cassettes should be designed.

Within this project a new *P. pastoris* strain carrying *HNL* gene should be also designed as a reporter strain to verify possible glycosylation effects. *HNL5* codes for hydroxynitrile lyase isoenzyme 5 from *Prunus amygdalus* and is known as high mannosylated protein, and is thereby an adequate model protein for this project.

My first aim was to design a *P. pastoris* strain for secretion of low mannosylated proteins. However, this aim was based on our hypothesis that *OCH1* deletion has the same effect as *OCH1* disruption. One additional aim was to gain information how *HOC1* deletion reflects on the protein glycosylation.

6 Instruments and associated equipment

6.1 Centrifuges and associated materials

- Centrifuge 5810 R/Eppendorf AG, Hamburg, Germany
- Centrifuge 5415 R/Eppendorf AG, Hamburg, Germany
- Avanti™ centrifuge J-20 XP/Beckman Coulter™, Inc, Vienna Austria
- JA-10 Rotor, fixed angle/Beckman Coulter™, Inc, Vienna Austria
- Nalgene® Labware 500 ml PPCO Centrifuge Bottles/Thermo Fisher Scientific, Rochester, NY, USA

6.2 Shakers and incubators

- Titramax 1000 (1,5 mm)/Heidolph Instruments, Schwabach, Germany
- Thermomixer comfort (1.5 ml)/Eppendorf AG, Hamburg, Germany
- Multitron II incubator shaker (25 mm)/Infors AG, Bottmingen-Basel, Switzerland
- RS 306 rotary shaker (50 mm)/Infors AG, Bottmingen-Basel, Switzerland
- Binder drying oven/Binder GmbH, Tuttlingen, Germany

6.3 Polymerase chain reaction (PCR) thermocyclers

- GeneAmp®PCR System 2700/Applied Biosystems, Foster City, CA, USA
- Mastercycler® personal/Eppendorf AG, Hamburg, Germany

6.4 Photometers, plate readers and associated materials

- DU 800 Spectrophotometer/Beckman coulter Inc, Fullerton, CA, USA
- Spectramax Plus 384/Molecular Devices, Ismaning/München, Germany
- Semi-micro cuvette 10x4x45 mm, Polystyrene/Sarstedt Aktiengesellschaft & Co., Nümbrecht, Germany
- NanoDrop 2000c Spectrophotometer/peqlab Biotechnologie GmbH, Polling, Austria

6.5 Gel electrophoresis and associated materials and devices

- Biozym LE Agarose/Biozym Biotech Trading GmbH, Vienna, Austria
- 6x Orange DNA Loading Dye/Fermentas GmbH, St. Leon-Rot, Germany
- GeneRuler™ 1kb DNA-Ladder/Fermentas GmbH, St. Leon-Rot, Germany
- Power Pac™ Basic (power supply)/Bio-Rad Laboratories, Vienna, Austria
- Sub-cell GT/Bio-Rad Laboratories, Vienna, Austria
- GelDoc-It™ Imaging Systems/UVP®, Cambridge, UK

- BioImaging Systems Gel HR Camera 6100 Series/UVP®, Cambridge, UK
- Chroma 43 mittelwellig 302 nm/Laborgeräte Vetter GmbH, Wiesloch, Germany

6.6 Electroporation materials

- Gene Pulser™/Bio-Rad Laboratories, Vienna, Austria
- Electroporation cuvette 2 mm/Bridge Bioscience™, Rochester, NY, USA

6.7 Reaction tubes

- Micro-centrifuge tubes, 1.5 ml with lid/Greiner Bio GmbH, Frickenhausen, Germany
- PP-Tube, sterile, cap, 12 ml/ Greiner Bio GmbH, Frickenhausen, Germany
- PP-Tube, sterile, 15 ml/Greiner Bio GmbH, Frickenhausen, Germany
- PP-Tube, sterile, +/- support skirt, 50 ml/Greiner Bio GmbH, Frickenhausen, Germany

6.8 Pipettes and pipette tips

- Denville XL 3000i™ (10 µl, 20 µl, 200 µl, 1000 µl)/Denville Scientific (Europe), P.O. Box 2108, Westbourne, UK
- Eppendorf Research® pipette (adjustable volume), 0.1-2.5 µl/ Eppendorf AG, Hamburg, Germany
- Biohit Proline® multichannel pipettor, 8 channels, 5-50 µl/Biohit Plc., Helsinki, Finland
- Biohit Proline® multichannel electronic pipettor, 8 channels 50-1200 µl/Biohit Plc., Helsinki, Finland
- Pipette tips, micro P10/Greiner Bio-One GmbH, Frickenhausen, Germany
- Pipette tips 200/Greiner Bio-One GmbH, Frickenhausen, Germany
- Pipette tips 1000/Greiner Bio-One GmbH, Frickenhausen, Germany
- Filter pipette tips 10 µl/Greiner Bio-One GmbH, Frickenhausen, Germany
- Filter pipette tips 200 µl/Greiner Bio-One GmbH, Frickenhausen, Germany
- Filter pipette tips P1000/Greiner Bio-One GmbH, Frickenhausen, Germany

6.9 Microplates

- PS-Microplate 96-well, flat bottom/Greiner Bio-One GmbH, Frickenhausen, Germany
- MicroAmp® Optical 96-Well Reaction Plate/Applied Biosystems, Foster City, CA, USA
- MicroAmp® Optical Adhesive Covers/Applied Biosystems, Foster City, CA, USA
- 96-well footprint deep well plate, PP/Bel-Art Products, Pequannock, NJ, USA
- Cover for deep well plate/Bel-Art Products, Pequannock, NJ, USA

6.10 Other materials and devices

- Rotilabo® syringe filters, CME, sterile, Ø 33 mm, 0.22 µm pore size /Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- Vortex-Genie 2/Scientific Industries Inc., Bohemia, NY, USA
- Sartorius BL 120S scale/ Sartorius Stedim Biotech GmbH, Göttingen, Germany
- inoLab® pH720 pH meter/WTW, Weilheim, Germany
- SenTix 21 PLUS® pH electrode/WTW, Weilheim, Germany
- MR 3000 magnetic stirrer/Heidolph, Schwabach, Germany
- Arium® basic ultrapure water system/Sartorius Stedim Biotech GmbH, Göttingen, Germany
- Petri plates/Greiner Bio GmbH, Frickenhausen, Germany
- Desalting membranes/Millipore, USA
- Pipettes 5ml, 10ml, 25ml/Greiner Bio GmbH, Frickenhausen, Germany
- Adhesive foils/Greiner Bio GmbH, Frickenhausen, Germany

7 Materials

7.1 Media and buffers

7.1.1 *Escherichia coli* media

- Low Salt Luria Bertani (LB) (agar): 10 g/l Bacto™ tryptone, 5 g/l Bacto™ yeast extract, 5 g/l NaCl, (15 g/l agar), autoclaved. If needed, antibiotics are added to a final concentration of 100 mg/l ampicillin, 25 mg/l zeocin™, 50 mg/l kanamycin.
- SOC medium: 20 g/l Bacto™ tryptone, 0.58 g/l NaCl, 5 g/l Bacto™ yeast extract, 2 g/l $MgCl_2$, 0.18 g/l KCl, 2.46 g/l $MgSO_4$, 3.81 g/l α-D(+)-glucose monohydrate, autoclaved

7.1.2 *Pichia pastoris* media

7.1.2.1 Stocks

- 500x B: 10 mg / 50 ml d-Biotin, filter sterilized
- 10x D: 220 g/l α-D(+)-glucose monohydrate, autoclaved
- 10x YNB: 134 g/l Difco™ yeast nitrogen base w/o amino acids, autoclaved
- 10x S: 200 g/l D-sorbitol, autoclaved

7.1.2.2 Media for cultivation in 96-deep well plates

- Yeast extract, peptone, dextrose (YPD) (agar): 10 g/l Bacto™ yeast extract, 20 g/l Bacto™ peptone, 900 ml dH_2O, (15 g/l agar), autoclaved. After autoclaving, adding 100 ml 10x D. If needed, antibiotics are added to a final concentration of 100 mg/l zeocin™, 300 mg/l geneticin.
- BMD1%, 1 l: 50 ml 10X D, 200 ml 10x PPB, 100 ml 10x YNB, 2 ml 500x B, 650 ml dH_2O
- BMM2, 1 l: 10 ml MeOH, 200 ml 10x PPB, 100 ml 10x YNB, 2 ml 500x B, 690 ml dH_2O
- BMM10, 1 l: 50 ml MeOH, 200 ml 10x PPB, 100 ml 10x YNB, 2 ml 500x B, 650 ml dH_2O

7.1.2.3 Media for genomic DNA isolation from yeasts

- Yeast lysis buffer: 4 ml Triton X-100, 20 ml 10% SDS, 4 ml 0.5 M NaCl, 0.4 ml 0.5 M EDTA, 2.0 ml 1.0 M Tris, pH 8.0, 169.6 ml dH_2O
- TE buffer: 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, pH 8.0

7.1.3 Buffers

7.1.3.1 Buffer solutions for enzyme characterization

- 50 mM NaOAc (sodium acetate): 4.1 g/l $NaC_2H_3O_2$, pH 4.5 adjusted with 100 % acetic acid
- 20x ABTS stock: 440 mg 2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) di-ammonium salt, add 50 mM NaOAc to 50 ml
- p-nitrophenol butyrate stock: 42µl pNPB 98% were added to 458µl DMSO
- 300mM Tris/HCl buffer: pH 7.0
- 1M citrate phosphate buffer: pH 5.0 (210.13g citric acid mono-hydrate in 1l dH_2O, 228.23g K_2HPO_4. 3 H_2O in 1l dH_2O)
- 0.1M phosphate-citrate buffer: pH 3.0 (21.013g citric acid mono-hydrate in 1l dH_2O, 22.823g K_2HPO_4. 3 H_2O in 1l dH_2O)
- 0.06M mandelonitrile solution: 80mg mandelic acid nirile in 10ml 0.1M phosphate-citrate buffer

7.2 Other chemicals and solutions

- 10x dNTP mix: 20 mM each: dATP, dTTP, dCTP, dGTP
- dATP / dCTP / dTTP / dGTP, 100 mM/Fermentas GmbH, St. Leon-Rot, Germany
- 50x TAE buffer: 242 g Tris base, 57.1 ml glacial acetic acid, 100 ml 0.5M EDTA, add dH_2O to a final volume of 1l
- BEDS: 10 mM bicine-NaOH, pH 8.3, 3 % (v/v) ethylene glycol, 5 % (v/v) DMSO, 1 M sorbitol, filter sterilized
- 1M DTT: 154,25g in 1l dH_2O

7.3 Enzymes

7.3.1 Restriction enzymes

- BamHI 10u/µl, 10xUnique buffer, 5'G↓*GATCC3'/Fermentas GmbH, St. Leon-Rot, Germany
- BamHI 1FDU/µl, 10xFastDigest® Buffer, 10xFastDigest® Green Buffer, 5'G↓*GATCC3'/Fermentas GmbH, St. Leon-Rot, Germany
- EcoRI 10u/µl, 10xUnique buffer, 5'G↓AATTC3'/Fermentas GmbH, St. Leon-Rot, Germany
- EcoRI 1FDU/µl, 10xFastDigest® Buffer, 10xFastDigest® green Buffer, 5'G↓AATTC3'/Fermentas GmbH, St. Leon-Rot, Germany

* ↓represents the enzyme restriction site.

- NotI 10u/µl, 10xOrange buffer, 5'GC ↓ GGCCGC3'/Fermentas GmbH, St. Leon-Rot, Germany
- NotI 1FDU/µl, 10xFastDigest® Buffer, 10xFastDigest® Green Buffer, 5'G ↓ AATTC3'/Fermentas GmbH, St. Leon-Rot, Germany
- NcoI 10u/µl, 10xTango buffer, 5'C ↓ CATGG3'/Fermentas GmbH, St. Leon-Rot, Germany
- NcoI 1FDU/µl, 10xFastDigest® Buffer, 10xFastDigest® Green Buffer, 5'C ↓ CATGG3'/Fermentas GmbH, St. Leon-Rot, Germany
- SmiI(SwaI) 10u/µl, 10xOrange buffer, 5'ATT ↓ AAAT3'/Fermentas GmbH, St. Leon-Rot, Germany
- SwaI(SmiI) 1FDU/µl, 10xFastDigest® Buffer, 10xFastDigest® green Buffer, 5'G ↓ AATTC3'/Fermentas GmbH, St. Leon-Rot, Germany

7.3.2 DNA polymerases

- Phusion® High-Fidelity DNA-Polymerase 2 u/µl, 5x HF reaction buffer or 5x GC buffer/Finnzymes Oy, Espoo, Finland
- GoTaq® Polymerase 5 u/µl, 5x Green GoTaq® Buffer/Promega GmbH, Madison, WI, USA
- PfuUltra™ High-Fidelity DNA-Polymerase 2.5u/µl, 10× PfuUltra™ HF reaction buffer/Stratagene, La Jolla, CA

7.4 Chemicals

- ABTS (2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) diammonium), from Sigma-Aldrich Handels GmbH, Vienna, Austria
- α-D(+)-glucose monohydrate, product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- Ampicillin, product of Sigma-Aldrich Handels GmbH, Vienna, Austria
- Aqua bidest. "Fresenius", product of Fresenius Kabi Austria GmbH, Graz, Austria
- Bacto™ peptone, product of Becton Dickinson and Company, Sparks, MD, USA
- Bacto™ tryptone, product of Becton Dickinson and Company, Sparks, MD, USA
- Bacto™ yeast extract, product of Becton Dickinson and Company, Sparks, MD, USA
- Benzaldehyde cyanohydrin (mandelonitrile), from Sigma-Aldrich Handels GmbH, Vienna, Austria
- Citronensäure, product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- D-biotin, product of Fluka Chemia AG, Basel, Switzerland
- Difco™ yeast nitrogen base w/o amino acids, product of Becton Dickinson and Company, Sparks, MD, USA
- Dimethyl sulfoxide (DMSO), product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- Dithiothreitol (DTT), product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- D-sorbitol, product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- Ethanol abs., product of Merck KGaA, Darmstadt, Germany
- Ethidium bromide (≥ 98 %), product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- Geneticin disulfate (G418 sulfate), product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- HCl (37%), product of Merck KGaA, Darmstadt, Germany

- K_2HPO_4, product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- Kanamycin, product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- KH_2PO_4, product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- LB agar, product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- Methanol (≥ 99.8 %), product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- Na_2HPO_4, product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- NaOAc ($NaC_2H_3O_2$, sodium acetate), product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- NaOH, product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- Phenol:chloroform:isoamylalcohol (25.24:1), product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- p-nitrophenol butyrate, product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- Tris, product of Carl Roth GmbH + Co. KG, Karlsruhe, Germany
- Zeocin™, product of InvivoGen-Eubio, Vienna, Austria

7.5 Existing strains

- Name: *Escherichia coli* Top 10F' (IMBT-TUG* strain collection number: 1482) / Source: Invitrogen Corporation, San Diego, USA
- Name: *Pichia pastoris* CBS7435 (IMBT-TUG strain collection number: 3132) / Source: *Pichia pastoris* wild type
- Name: *Pichia pastoris* Mut^S (IMBT-TUG strain collection number: 3445) / Source: Based on *P. pastoris* CBS7435, engineered at Graz University of Technology (Näätsaari et al., manuscript in preparation)
- Name: *Pichia pastoris* ΔKU70 (IMBT-TUG strain collection number: 3499) / Source: Based on *P. pastoris* CBS7435, engineered at Graz University of Technology (Näätsaari et al., manuscript in preparation)

7.6 Existing plasmids

Following six existing plasmids were used within this thesis: pPpT4_SmiI, pPpKan_SmiI, pPpT4_SmiI_Cycle3GFP, pPpT4_SmiI_HRPC1ASayn #0, pPpT4_SmiI_CalB #1, pGAPZ A-PaHNL5_alpha, and pJet1.2blunt_ends. Two existing recombination cassettes were also used: FLIPPER_pAOX1 and HIS_Flipper_old_FRT. Both cassettes exist in pJet1.2 cloning vector.

* IMBT-TUG is an abbreviation for the Institute of Molecular Biotechnology at Graz University of Technology.

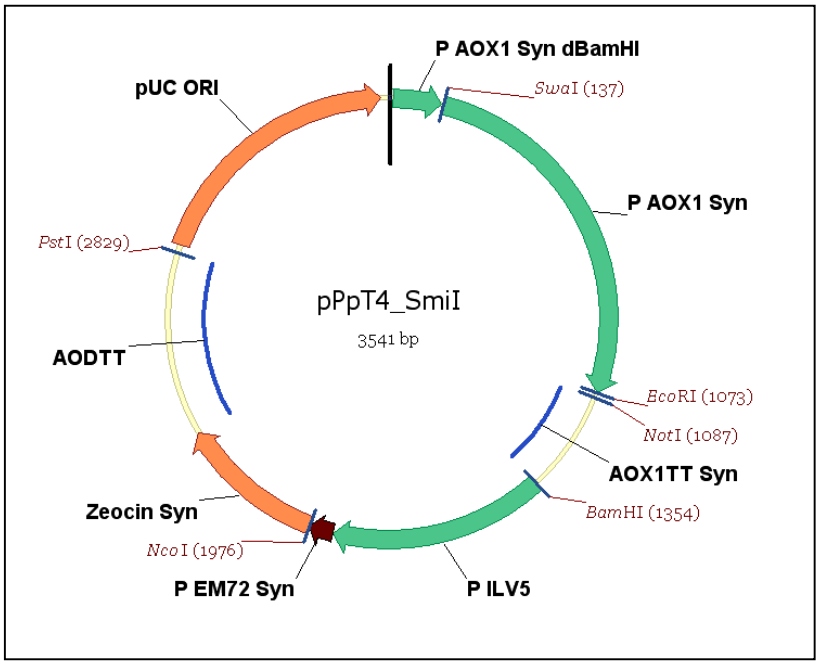

Figure 13: pPpT4_SmiI plasmid. pUC_ORI (pUC origin of replication); PAOX1_Syn_dBamHI (synthetic *AOX*1 promoter upstream region); PAOX1_Syn (synthetic *AOX*1 promoter region); AOX1TT_Syn (synthetic terminator); pILV5 (eukaryotic promoter); pEM72Syn (synthetic prokaryotic promoter); Zeocin_Syn (gene coding for zeocin resistance); AODTT (*AOD* terminator); SwaI (SmiI), EcoRI, NotI, BamHI, NcoI, and PstI are the restriction sites for the respective restriction enzymes.

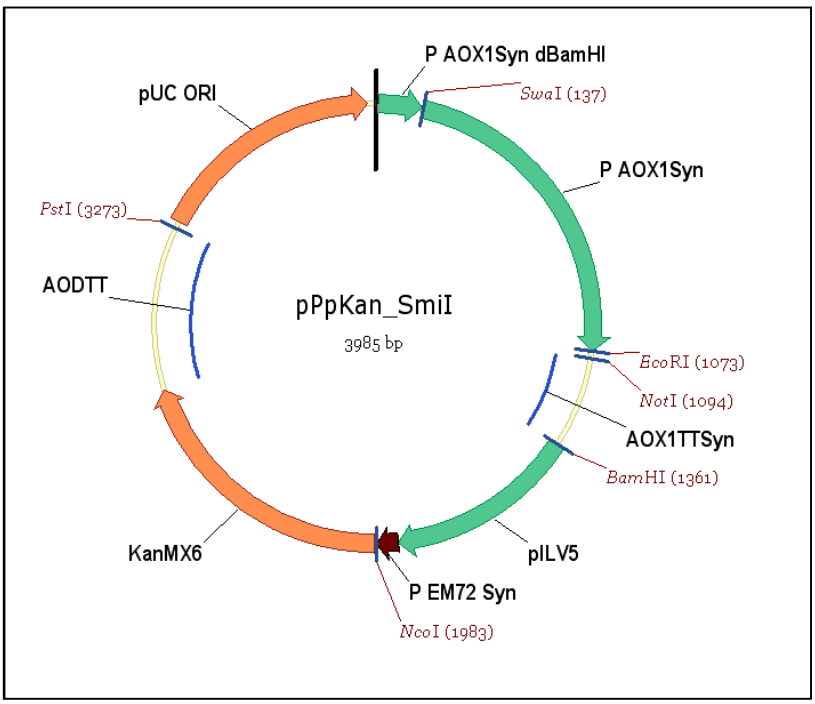

Figure 14: pPpKan_SmiI plasmid. pUC_ORI (pUC origin of replication); PAOX1_Syn_dBamHI (synthetic *AOX*1 promoter upstream region); PAOX1_Syn (synthetic *AOX*1 promoter region); AOX1TT_Syn (synthetic terminator); pILV5 (eukaryotic promoter); pEM72Syn (synthetic prokaryotic promoter); KanMX6 (gene coding for kanamycin/geneticin resistance); AODTT (*AOD* terminator); SwaI (SmiI), EcoRI, NotI, BamHI, NcoI, and PstI are the restriction sites for the respectively restriction enzymes.

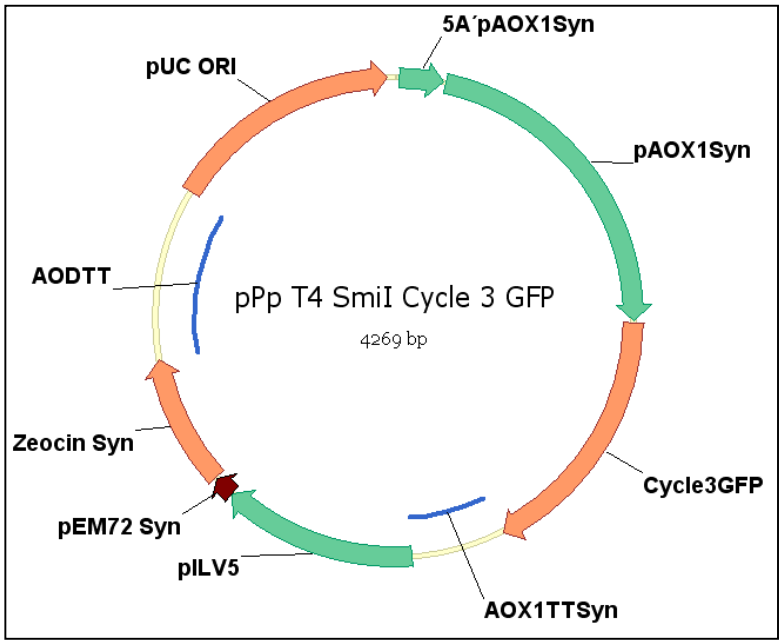

Figure 15: pPpT4_SmiI_Cycle3GFP plasmid. pUC_ORI (pUC origin of replication); PAOX1_Syn_dBamHI (synthetic *AOX*1 promoter upstream region); PAOX1_Syn (synthetic *AOX*1 promoter region); Cycle3GFP (synthetic gene coding for green fluorescence protein); AOX1TT_Syn (synthetic terminator); pILV5 (eukaryotic promoter); pEM72Syn (synthetic prokaryotic promoter); Zeocin_Syn (gene for zeocin resistance); AODTT (*AOD* terminator);

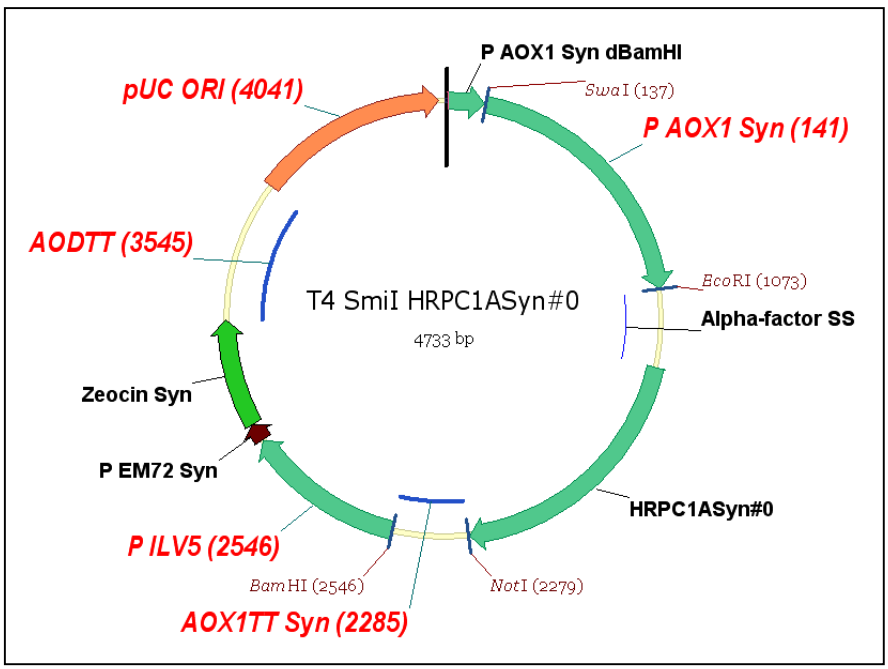

Figure 16: pPp-T4_SmiI HRPC1ASayn #0 plasmid. pUC_ORI (pUC origin of replication); PAOX1_Syn_dBamHI (synthetic *AOX*1 promoter upstream region); PAOX1_Syn (synthetic *AOX*1 promoter region); Alpha-factor_SS (*S. cerevisiae* secretion signal sequence); HRPC1ASyn#0 (synthetic gene coding for Horseradish peroxidase); AOX1TT_Syn (synthetic terminator); pILV5 (eukaryotic promoter); pEM72Syn (synthetic prokaryotic promoter); Zeocin_Syn (gene for zeocin resistance); AODTT (*AOD* terminator); The restriction enzymes (sites) SwaI (SmiI) and BamHI were used for plasmid linearization, and EcoRI and NotI for cloning of *HRPC1ASyn#0*.

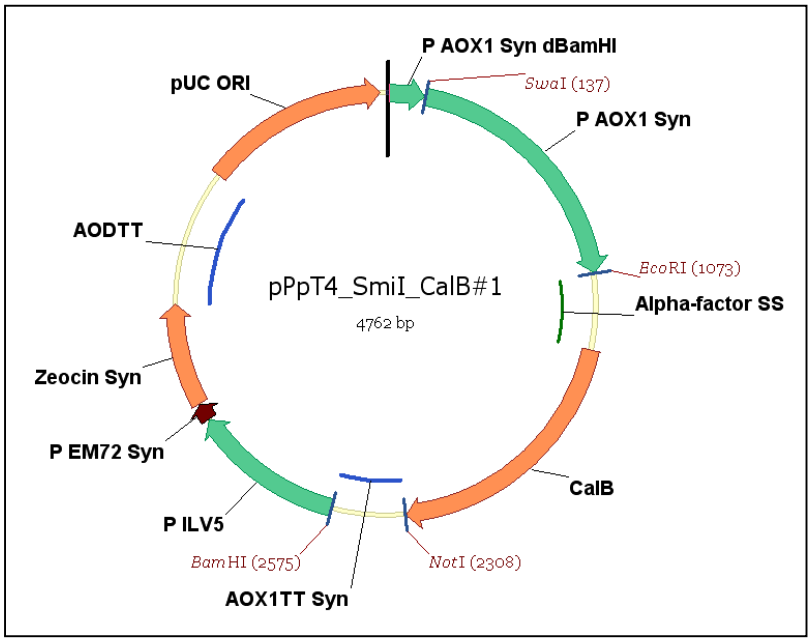

Figure 17: pPp-T4_SmiI CalB #1 plasmid. pUC_ORI (pUC origin of replication); PAOX1_Syn_dBamHI (synthetic *AOX*1 promoter upstream region); PAOX1_Syn (synthetic *AOX*1 promoter region); Alpha-factor_SS (*S. cerevisiae* secretion signal sequence); CalB (gene coding for *Candida antartica* lipase); AOX1TT_Syn (synthetic terminator); pILV5 (eukaryotic promoter); pEM72Syn (synthetic prokaryotic promoter); Zeocin_Syn (gene for zeocin resistance); AODTT (*AOD* terminator); The restriction enzymes (sites) SwaI (SmiI) and BamHI were used for plasmid linearization, and EcoRI and NotI for cloning of *CalB*.

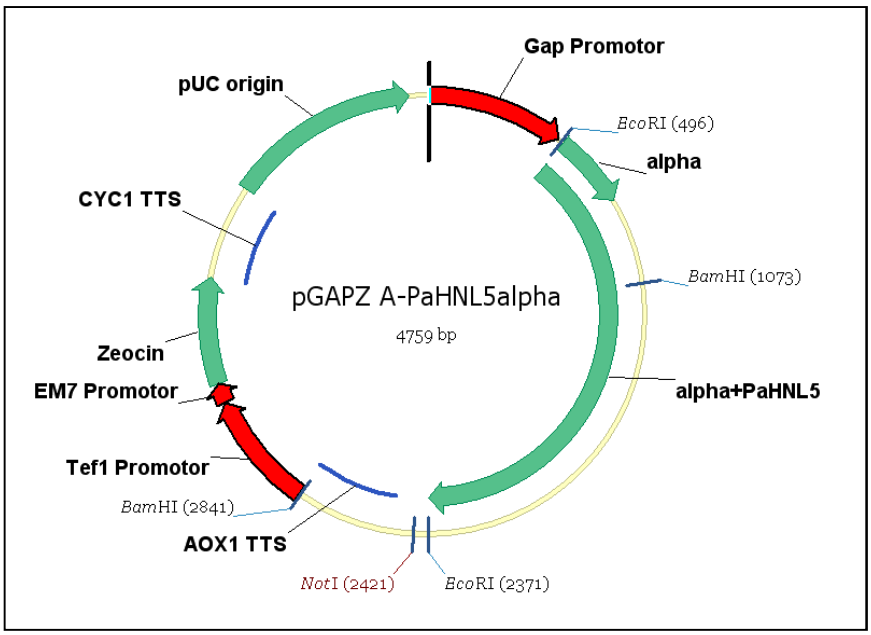

Figure 18: pGAPZ A-PaHNL5 alpha plasmid. pUC_origin (pUC origin of replication); Gap_promoter (constitutive promoter of *P. pastoris* gene coding for glyceraldehyde 3-phosphate); alpha factor (secretion signal sequence); PaHNL5 (*Prunus amygdalus* gene coding for hydrolynitrile lyase); AOX1_TTS (synthetic terminator); Tef1_promoter (eukaryotic promoter); EM7_promoter (synthetic prokaryotic promoter); Zeocin (gene coding for zeocin resistance); CYC1_TT (*S. cerevisiae* iso-1-cytochrome c terminator);

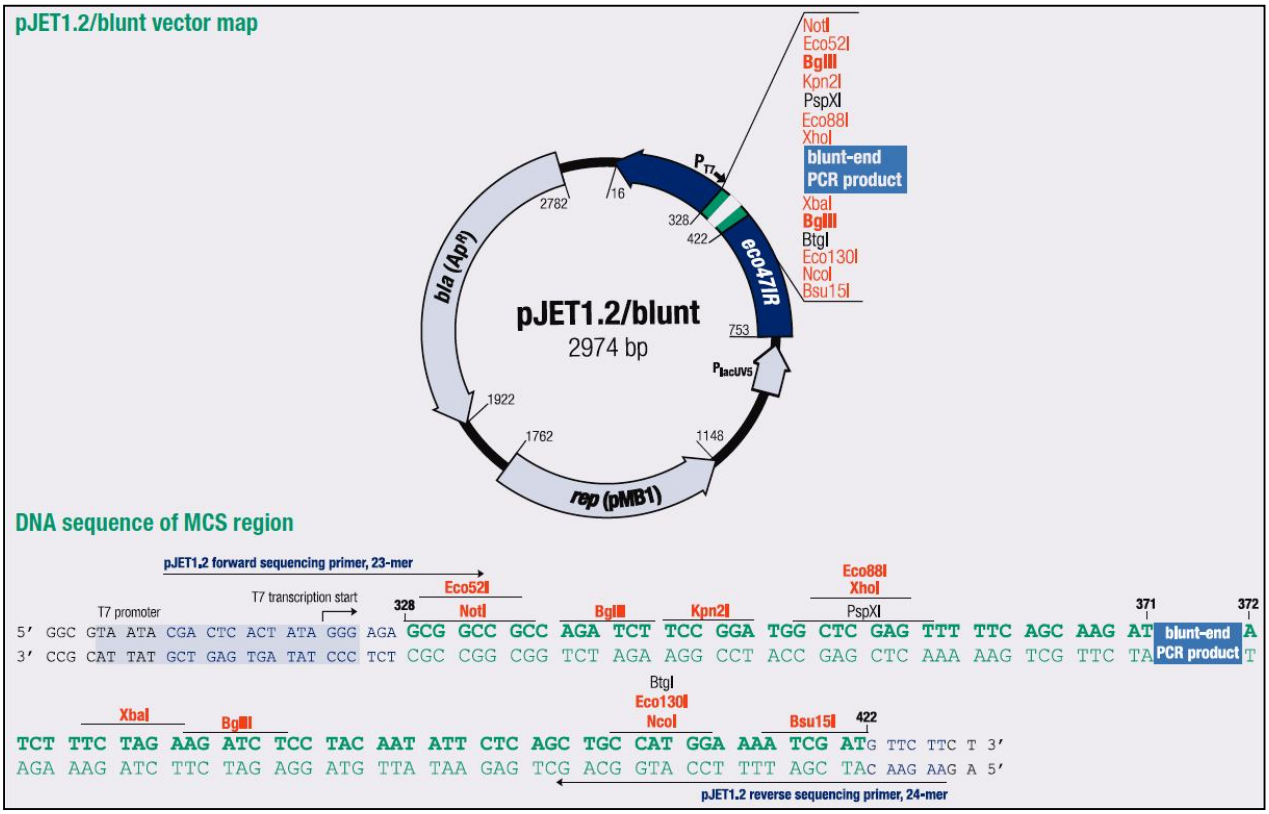

Figure 19: pJet1.2blunt ends cloning vector. Rep pMB1 (replicon from pMBI plasmid); bla_AmpR (gene for ampicillin resistance); eco47IR (lethal gene for positive selection of recombinants which is disrupted by the inserted sequence); P*lac*UV5 (modified P_{lac} promoter); P_T7 (T7_RNA polymerase promoter); and MCS (the space between 328bp and 422bp is the multiple cloning site);

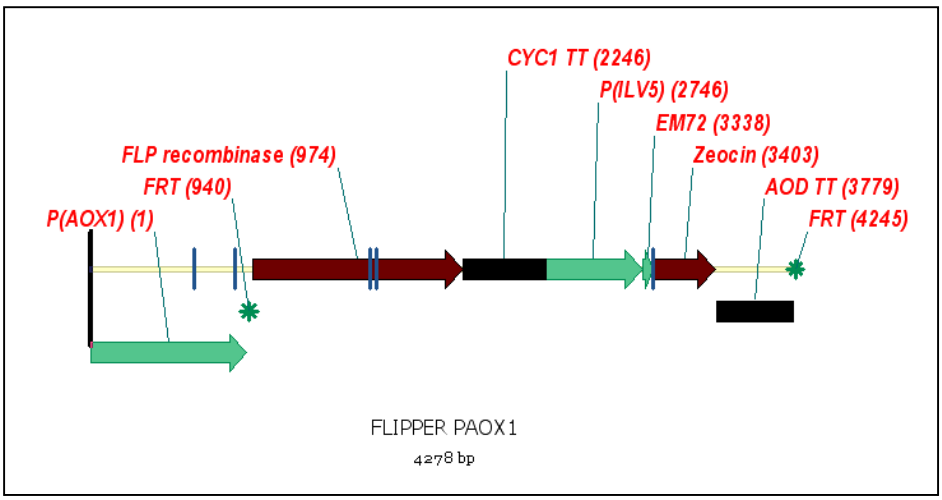

Figure 20: Existing FRT recombination cassette (FLIPPER_PAOX1) has been cloned in pJet1.2blunt_ends cloning vector. The cassette contains following parts: P(AOX)(1) (*P. pastoris* alcohol oxidase 1 promoter), FRT (recombinase target sequences), FLP_recombinase (gene coding for *S. cerevisiae* 2µm plasmid recombinase), CYC_TT (*S. cerevisiae* iso-1-cytochrome c terminator),P(ILV5) (eukaryotic promoter), EM72 (synthetic prokaryotic promoter), Zeocin (gene coding for zeocin resistance), AOD_TT (*AOD* terminator).

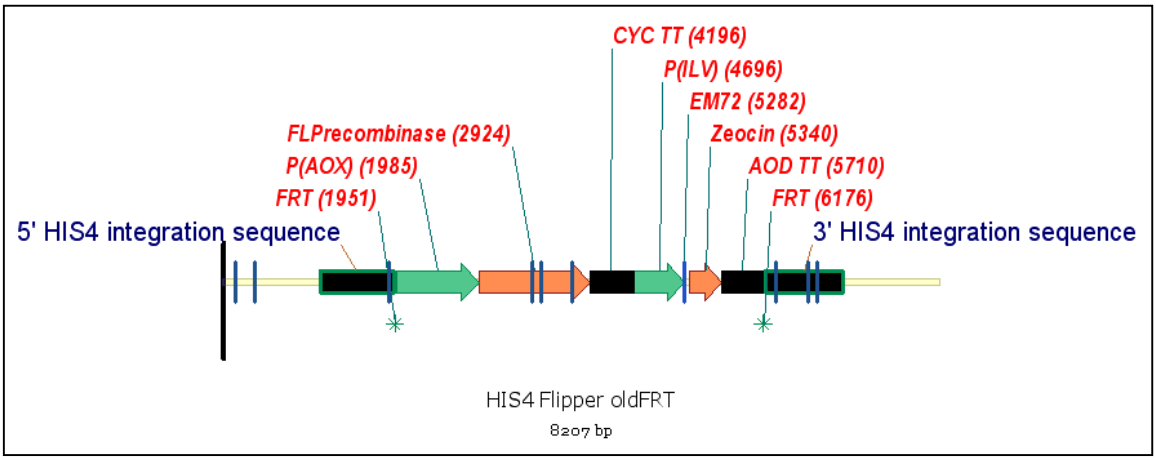

Figure 21: Existing FRT recombination cassette (HIS4_Flipper_old_FRT). The cassette contains following parts: 5`- and 3` HIS4 integration sequence (homologous sequences to *P. pastoris* HIS locus), FRT (recombinase target sequences), P(AOX)(1) (*P. pastoris* alcohol oxidase 1 promoter), FLP_recombinase (gene coding for *S. cerevisiae* 2µm plasmid recombinase), CYC_TT (*S. cerevisiae* iso-1-cytochrome c terminator),P(ILV5) (eukaryotic promoter), EM72 (synthetic prokaryotic promoter), Zeocin (gene coding for zeocin resistance), AOD_TT (*AOD* terminator).

7.7 Primers

From all used primers within this master thesis, 37 were primers existing at the Institute of Molecular Biotechnology at Graz University of Technology, and 146 were newly designed. All the primers were designed with Vector NTI® (Bioinformatics software package from Invitrogen, USA) and purchased from Integrated DNA Technologies, BVBA, Leuven, Belgium.

7.7.1 Existing primers

Primer designation	Primer name	Primer sequence (5`-3`)
p-07-352	AOXSyn1fw	GGATCCAATTCCCGCTTTGACTGCCTGAAATCTCC
p-07-353	AOXSyn2fw	GGAACCTAATATGACAAAAGCGTGATCTCATCC
p-07-604	EM72_Nco_rv	ACCCATGGTTTAGTCCTCCTTACACC
p-07-628	1FRTFseq	GGTGCACCTGTGCCGAAACG
p-07-629	2FRTRseq	GTTCCGTTATGTGTAATCATCCAAC
p-07-636	2RnewCYC	CAAAGGAAAAGGGGCCTGTTTATATGCGTCTATTTATGTAG
p-07-637	CYC1F	ACATAAATAGACGCATATAAACAGGCCCCTTTTCCTTTGTCGATATC
p-07-638	CYC1R	GAAACAAGACATTACTGAAGTCGACAACTAAACTGGAATGTGAGG
p-08-134	P(AOX)F	GAAGTTCCTATACTTTCTAGAGAATAGGAACTTCAGATCTAACATCC AAAGACGAAAGG
p-08-136	FLPF	CAACTAATTATTGAAACGATGCCACAATTTGATATATTATGTAAAAC ACCACCTAAGG
p-08-151	pucPstfw	AAACTGCAGGTACCACTGAGCGTCAGACCCC
p-08-483	pJET1.2Fw	CGACTCACTATAGGGAGAGCGGC
p-08-484	pJET1.2Rv	AAGAACATCGATTTTCCATGGCAG
p-08-690	ILV5_fw	GCGAACAAATTAAGAGATAATCTCG
p-08-691	AODTT_fw	CGGATGTTTTATTATCTATTTATGC
p-08-692	PUC_rv	GCAGAGCGAGGTATGTAGGC
p-08-694	Kan_rv	CTCATCTGTAACATCATTGGCAAC
p-08-849	AOXSyn_end_fw	CAGAAGGAAGCTGCCCTGTC
p-09-080	KanMX6_fw	GCTCGATGAGTTTTTCTAACC
p-09-132	GFP_AscI_rv	TTTGGCGCGCCTTACTTGTACAATTCATCCATGCC
p-09-155	Kan_fw	GCCAATGATGTTACAGATGAGATGG
p-09-156	Kan_Mitte_fw	GCAATCACGAATGAATAACGG
p-09-194	AOXTTR	TGAAGCCTGCATCTCTCAG
p-09-304	KanII_rev	TATCGGTCTGCGATTCCGACTCGTC
p-09-337	Zeocin_rv	CTGCGGAGATGAACAGGGTAA
p-09-511	PaHNL_V317ASynI_fw	GGAACTACCATCATTAACGCAGG
p-09-512	PAHNL_V317ASynII_fw	CACCTAATCCTATTGAAGCATCTGTTG
p-10-632	KanMXintfw	gatcgcgtatttcgtctcg
p-10-633	KanMXintrv	cgactgaatccggtgagaatg
p-10-595	pILV5rv	GAAAATTTTTTCGGATAATTTTTTAAAGCGAG

Table 2: List of existing primers, used in this thesis.

7.7.2 Primers designed in this thesis

Primer designation	Primer name	Primer sequence (5`-3`)
p-09-514	pdi_1_upstr_fw_	gtggttgtgttacttcctgatagc
p-09-515	pdi_1_term_rv_	cgtattgcgcgaagtgatgc
p-09-516	pdi_2_upstr_fw_	gatgaggagctttatagacg
p-09-517	pdi_2_gen_rv_	cttcaacgagcttagcaacctg
p-09-518	pdi_3_upstr_fw_	ccgccaaacagggctgctgactc
p-09-519	pdi_3_term_rv_	gatagaccaggcggagtgcaagttc
p-09-520	pdi_4_upstr_fw_	gcgatatgtaacctttgacgtagcc
p-09-521	pdi_4_term_rv_	cctgctacacgagacaattccg
p-10-052	pdi_1_vor_prom_rv_	gcctgtggaaagaactgtagaga
p-10-053	pdi_1_prom_fw_	cctttctctctgtgctgtcaactc
p-10-054	pdi_2_vorn_rv_	gtattatatctcccagcaacg
p-10-055	pdi_2_prom_fw_	ggtgcaaagagctggagatgg
p-10-056	pdi_3_prom_rv_	caggctcccaagggacttggc
p-10-057	pdi_3_prom_fw_	cttggatggttacttgagtaacgtcg
p-10-058	pdi_4_upstr_rv_	gcctcttggtggtgtacttggg
p-10-059	pdi_4_upstr_fw_	cgtctgtacttcaacctcatctcc
p-10-127	pdi_4_seq_3_fw	gcccacaaacccataaaag
p-10-128	pdi_4_seq_2_fw	gactttgaggataacag
p-10-129	pdi_4_seq_1_fw	gactttctgtaatgagc
p-10-130	pdi_3_seq_4_rv	ctcaccttccacaac
p-10-131	pdi_3_seq_3_fw	ggttggaaggtgaagtgc
p-10-132	pdi_3_seq_2_fw	ggaagggaatggactcg
p-10-133	pdi_3_seq_1_fw	gttgagttgatagccgc
p-10-134	pdi_2_seq_2_fw	cttctgtttccttccatc
p-10-135	pdi_2_seq_1_fw	cgtattgccacttccatc
p-10-136	pdi_1_seq_4_fw	ggtgaggttgaggtcc
p-10-137	pdi_1_seq_3_fw	gaatagaacacgaacactg
p-10-138	pdi_1_seq_2_fw	gcaccaccttacaccac
p-10-139	pdi_1_seq_1_fw	ggtggacgaagaagtctcc
p-10-245	a_rv_pdi_1_	GCTAAACTGCTCCCACCATTG
p-10-246	b_rv_pdi_1_	TACCTGCGGGACTTTCTAACACG
p-10-247	c_fw_pdi_1_	GTTCTACGACCTCTGATGACCTG
p-10-248	d_rv_pdi_1_	CAGTGAGAAAGTGGTGTAAGGTGG
p-10-249	e_rv_pdi_1_	GTTCGTTCATCGGGCAGTTCTTTC
p-10-250	f_rv_pdi_1_	AATGCTTTGGCTCTGTCTTTG
p-10-251	g_rv_pdi_1_	TCTGCCTCTCCTGCGATGAATTTC

Table 3-7: List of new primers, designed in this thesis.

Primer designation	Primer name	Primer sequence (5`-3`)
p-10-252	h_fw_pdi_1_	GCTGCTGCTGCCGATATTATTAAACC
p-10-253	a_rv_pdi_2_	ATCAGTGGCGAATGGCGGTGC
p-10-254	b_fw_pdi_2_	TAGTAGTAGTAGCACCGCCATTCG
p-10-255	c_rv_pdi_2_	TTTGGAGAAGAGATGGAAGTGGC
p-10-256	d_rv_pdi_2_	GCTCAACTGGTCCTGTCCTTTCTC
p-10-257	e_rv_pdi_2_	TCTTCTATTGGTGCTTGATGTGGG
p-10-258	f_fw_pdi_2_	ATCCCACATCAAGCACCAATAG
p-10-259	a_rv_pdi_3_	TCTATTTTTGGTTTGGTCTGCTTC
p-10-260	b_rv_pdi_3_	CAACACCAAATCACACAAGTCAAC
p-10-261	c_rv_pdi_3_	ACGTAAAAAGAAGAAGCAATGCAC
p-10-262	7_fw_pdi_3_	AATGTGGTTGGAAGGTGAAGTGC
p-10-263	d_fw_pdi_3_	CGAATGAATCCAGAGTATGAAAAGC
p-10-264	e_fw_pdi_3_	GTTCAATCTTTCCCGACTATCCTG
p-10-265	f_rv_pdi_3_	TAATGACAAAACCAGAACAGCAGG
p-10-266	g_rv_pdi_3_	GAAACTCCTCCATAAATCTACTTG
p-10-267	a_rv_pdi_4_	CGAAAAGTCCCCAAATCTCACAG
p-10-268	b_rv_pdi_4_	GAACAACATCCCAACACCGTC
p-10-269	c_rv_pdi_4_	ATCCGTTTTAGTTCAGTCAATGG
p-10-270	d_rv_pdi_4_	CGAAAAAGCAATCCAAGAAGAAGG
p-10-271	e_rv_pdi_4_	TCAAAACAGGAAACCCGACTACC
p-10-272	f_fw_pdi_4_	GGGTATGCCAACAAGAGGGTC
p-10-273	g_fw_pdi_4_	GGAAATGACGGTGTTGGGATGTTG
p-10-309	pdi_1_term_fw	tgaagaagaagctgaaagtgaggc
p-10-310	pdi_1_term_rv	ccgcttagttctcttcttcaccttg
p-10-311	pdi_2_term_fw	gaaaagcgaagaagggcaagaagg
p-10-312	pdi_2_term_rv	tcctcatcatccagtaacgaagcc
p-10-313	pdi_3_term_fw	aatctcacatgctccggccacac
p-10-314	pdi_3_term_rv	tgtgtatggaggatgttgatgccg
p-10-315	pdi_4_term_fw	gatgttgaggagtctggggaagc
p-10-316	pdi_4_term_rv	gaagctccacaaaatctccacaatac
p-10-533	pAox_FRT_fw	atagtcctgtcgggtttcgccac
p-10-534	pAox_FRT_rv	gttcctattctctagaaagtataggaacttccgtttcaataattagttg
p-10-535	GFP_FRT_fw	atactttctagagaataggaacttcaaacgatggctagcaaaggagaag
p-10-536	GFP_FRT_rv	cacgaagtcaaaatggctcaccactgc
p-10-561	GFP_FRT_mut_rv	gttcttctcctttgctagccatcgtttgaagttcctattctctagaaagtataggaacttccgtttcaataat tagttgttttttg

Table 3-7: List of new primers, designed in this thesis.

Primer designation	Primer name	Primer sequence (5`-3`)
p-10-562	GFP_FRT_mut_fw	caaaaaacaactaattattgaaacggaagttcctatactttctagagaatagg aacttcaaacga tggctagcaaaggagaagaac
p-10-575	GFP_FRT_mut_1-Kozak_rv	gttcttctcctttgctagccatgaagttcctattctctagaaagtataggaacttc cgtttcaataattagt tgtttttg
p-10-576	GFP_FRT_mut_1-Kozak_fw	caaaaaacaactaattattgaaacggaagttcctatactttctagagaatagg aacttcatggct agcaaaggagaagaac
p-10-581	pdi_4_pAOX_long_rv_	GTCATCAAGCTCTTCAGCATAATATCTTGCCAGAG
p-10-582	pdi_4_pAOX_long_fw_	caaaaaacaactaattattgaaacgatgaagtcgttactgctacttc
p-10-583	pdi_3_pAOX_long_rv_	GCAATCTTTTGCAATATCCACACCACGAAGCAG
p-10-584	pdi_3_pAOX_long_fw_	gatcaaaaaacaactaattattgaaacgatgaaaatattaagtgcattgc
p-10-585	pdi_2_pAOX_long_rv_	AGCTGGGTCTTCTCAGAATTTCCCAGGG
p-10-586	pdi_2_pAOX_long_fw_	caaaaaacaactaattattgaaacgatgaagttactatccttgg
p-10-587	pdi_1_pAOX_long_rv_	GCTTCAGCATATGATTTGAAGGTGGATCCG
p-10-588	pdi_1_pAOX_long_fw_	caaaaaacaactaattattgaaacgatgcaattcaactggaatattaaaac
p-10-589	pAOX_FRT_long_rv_	CGTTTCAATAATTAGTTGTTTTTTGATCTTCTCAA GTTGTCG
p-10-590	pAOX_FRT_long_fw_	ctagaagaataggaacttcagatctaacatccaaagacgaaag
p-10-591	EM72_elim_rv_	CCTTTCGTCTTTGGATGTTAGATCTGAAGTTCCTAT TCTTCTAGAAAG
p-10-592	EM72_elim_fw_	ccaaatatcgtctccacaaaaaaataaactcacctctgctgttcc
p-10-593	pILV5_long_rv_	GTGAGTTTATTTTTTTGTGGAGACGATATTTGGAA GAG
p-10-594	pILV5_long_fw_	gaagttcctatactttctagagaataggaacttcagatctaacatcc
p-10-670	pILV5_kurz_rv_	TTTGTGGAGACGATATTTGGAAGAG
p-10-671	pILV5_kurz_fw_	gaagttcctatactttctagag
p-10-734	OCH_2_gene_rv	TCTCAAAATACCCATTCGAGTTGG
p-10-735	OCH_2_gene_fw	cttacaggcaatgcaacgttcttc
p-10-736	OCH1_gene_rv	ATATTGGAACGTCCCCCAGTTG
p-10-737	OCH1_gene_fw	gtttacattgaccagatgatttggc
p-10-800	OCH1_249_upstr_fw	GTTTACATTGACCAGATGATTTGGCTTTTTCTCTGT TCAATTCAC
p-10-801	OCH1_249_upstr_rv	GTCCTCGTTTCGAAGTACCTAGGGCTGATGATATTT GCTACGAAC
p-10-802	OCH1_249_downstr_fw	GCGAGGTCACTCAGTCAGAAAGCTAGAGTAAAAT AGATATAGCGAG
p-10-803	OCH1_249_downstr_rv	ATATTGGAACGTCCCCCAGTTGCTTTAGGGTATCTT CATC
p-10-804	OCH1_636_upstr_fw	CTTACAGGCAATGCAACGTTCTTCTCCGGATGG

Table 3-7: List of new primers, designed in this thesis.

Primer designation	Primer name	Primer sequence (5`-3`)
p-10-805	OCH1_636_upstr_rv	TCGTTTCGAAGTACCTAGCTTGGACAGTAAGAGAGAATCG
p-10-806	OCH1_636_downstr_fw	CAGATAGCGAGGTCACTCAGTCGAGAATTAATCGCC
p-10-807	OCH1_636_downstr_rv	TCTCAAAATACCCATTCGAGTTGGTTGTCAGAGGGG
p-10-812	AOX_Linker_rv_	CGTGAAGTCCTCGTTTCGAAGTACCTAGCGTTTCAATAATTAG
p-10-813	FRT_Linker_fw_	gtcagatagcgaggtcactcagtcgaagttcctatactttctag
p-10-824	HNLwt_end_fw	ATTGTGTTAGCGGCATGAAGAAGC
p-10-825	HNLwt_middle_fw	CTTTTGACAATAATGGAACGCGAC
p-10-827	alpha_new_end_fw	tactactattgccagcattgctgc
p-10-840	KanMX_fw	ctaggtacttcgaaacgaggacttcac
p-10-841	KanMX_rv	GACTGAGTGACCTCGCTATCTGACTG
p-10-842	Rec_mit_fw	cggaagaagcagataagggaaatagccac
p-10-843	Rec_mit_rv	GATCTCCCAGATGCTTTCACCCTCAC
p-10-844	Linker_FRT_rv_	CGTGAAGTCCTCGTTTCGAAGTACC
p-10-845	Linker_FRT_fw_	gtcagatagcgaggtcactcagtcg
p-10-866	CycTT_ILV5_long	attccagtttagttgtcgacttcagtaatgtcttgtttcttttg
p-10-867	FRT_Zeo_rv_long	CCTTTCGTCTTTGGATGTTAGATCTGAAGTTCCTATTCTCTAGAAAG
p-10-868	OCH1_249_colony_fw	cgcctggtagggatgcaatacaagg
p-10-869	OCH1_636_colony	TCTCCCAGTGAAGACAATAATATATGGG
p-10-870	pILV_nat_rv	CCCATTTTGTGGAGACGATATTTGGAAGAGAAGGG
p-10-871	pILV_KanMX_nat_fw	ccttctcttccaaatatcgtctccacaaaatgggtaaggaaaagac
p-10-888	PDI_4.4._rv_seq	GTCCGTGAGTTCCCATATCGTACC
p-10-889	PDI_4.2._rv_seq	CTGGATTACATCATCCTTTGGAGTG
p-10-890	PDI_4.1._rv_seq	GCTTGTGCTAGTGTGAGAGCGGAC
p-10-891	AOD_TT_end_fw	tggggaaacttggatctgattacc
p-11-021	249_up_2500_long_fw	ctgattggggatattctagttacagcactaaacaactggcgatac
p-11-022	249_down_2500_long_rv	ATTGAGTATATGGGCACTTGCCCCTTTATTTATGACTTGGG
p-11-023	636_up_2500_long_fw	gctatgtcagtctctgcttctgaatctgaggagtttatccc
p-11-024	636_down_2500_long_rv	TAAGTTCAGCTCATAATTCCTCTGTTATCCATTGCCAATTAGACC
p-11-025	249_KanMX_direkt_ovlap_rv	GTGAGTCTTTTCCTTACCCATGGCTGATGATATTTGCTACG

Table 3-7: List of new primers, designed in this thesis.

Primer designation	Primer name	Primer sequence (5`-3`)
p-11-026	249_KanMX_direkt_ovlap_fw	gcagtttcatttgatgctcgatgagtttttctaaagaaagctagag
p-11-027	636_KanMX_direkt_ovlap_rv	CGTGAGTCTTTTCCTTACCCATCTTGGACAGTAAG AGAGAATC
p-11-028	636_KanMX_direkt_ovlap_fw	gatgctcgatgagtttttctaagagaattaatcgccaaaatc
p-11-061	249_up_seq_2500_fw	gtcaaaccaacgtaaaaccctctgg
p-11-060	249_up_seq2_2500_fw	cccatagatctgcctgacagcc
p-11-059	249_down_seq_2500_rv	GATGAGCAAATCCCAGCTCAGTG
p-11-058	249_down_seq2_2500_rv	CATCTTGGGCAGAATAATGACCACG
p-11-057	636_up_seq_2500_fw	ggcaacattgatcacaacgtcg
p-11-056	636_up_seq2_2500_fw	gtagatagtgggaaacccttcccac
p-11-055	636_down_seq_2500_rv	GAGCGGGTCCAGTTGCCTCTAC
p-11-054	636_down_seq2_2500_rv	CAAACTTCACTGTGGACCGGATAGAC
	249_lok_fw_	gctcctgacgtaggcctagaacagg
	249_lok_rv_	CGTCTGGTTTAGAAGACAGGCCG
	249_gen_fw_	atggcgaaggcagatggcag
	249_gen_rv_	TTAGTCCTTCCAACTTCCTTCAAATG
	636_lok_fw_	cctgttcccatttgtcaccattagc
	636_lok_rv_	GCTCAGAAACCACTCCCAGAACCTAC
	636_gen_fw_	gaaagatgtgaagtttctttgttggcg
	636_gen_rv_	TCAAGATTGGGTGACCAGGCTTC

Table 3-7: List of new primers, designed in this thesis.

8 Kits, protocols and methods

8.1 DNA isolation

8.1.1 Genomic DNA isolation and determination of its quality and quantity

In order to isolate genomic DNA (gDNA) from *P. pastoris*, following protocol was applied (cf. Hoffman et al.). A single colony of the desired strain was inoculated to 7-10ml sterile YPD medium and incubated over night at 28 °C, 110rpm. The grown culture (OD_{600} < 10) was centrifuged for 5 min at 500 x g, the pellet was re-suspended in 500µl sterile dH_2O and transferred to a sterile micro-centrifuge tube. The cells were spun down and the pellet was re-suspended in 200µl yeast lysis buffer. 200µl phenol:chloroform:isoamyl alcohol (25:24:1) plus approximately 300mg acid-washed glass beads were added and the suspension was vortexed for 4 min. After addition of 200µl TE buffer, the suspension was centrifuged for 5 min at 16,100 x g and the aqueous phase was transferred to a new micro centrifuge tube. The DNA was precipitated by addition of 1ml ice cold absolute EtOH and incubation for 10 min at -20 °C, spun down and air dried. The DNA pellet was re-suspended in 400µl TE plus 5 µl 2 mg/mL RNaseA and incubated for > 4 h at 37 °C. 10 µl 4 M ammonium acetate and 1ml absolute EtOH were added and the DNA was precipitated as described above. The DNA pellet was washed with 70 % ice cold EtOH, spun down and air dried. The DNA was dissolved in a final volume of 50µl dH_2O.

8.1.2 Plasmid isolation

For plasmid isolation, GeneJET™ Plasmid Miniprep Kit (Fermentas GmbH, St. Leon-Rot, Germany) was used according to the manufacturer's recommendations, except for the following deviations: cells were abraded from 1/2 or 1/4 an agar plate with a sterile toothpick, and plasmid elution was done with 50µl deionized H_2O (dH_2O).

8.2 DNA purification

DNA purification was done with the Wizard® SV Gel and PCR Clean-Up System (Promega GmbH, Mannheim, Germany). Applied deviation from the manufacturer's recommendations was the elution of DNA with 18-20µl dH_2O.

8.3 Polymerase chain reaction (PCR)

8.3.1 Standard PCR and overlap extension PCR (OE-PCR)

Standard PCRs were performed with Phusion® High-Fidelity DNA-Polymerase, according to the manufacture's recommendation. The annealing temperature was by default chosen to be 4-5 °C below the primer DNA melting temperature (T_m). T_m for the primers shorter than 24bp was calculated with Vector NTI® (Bioinformatics software package from Invitrogen, USA), and for the primers longer than 24bp, *Tm* from the manufacturer's certificate was taken. All primers were used in a final concentration of 400nM, dNTPs in a final concentration of 200 µM. The elongation time was calculated in dependence of the expected PCR product length, considering a processivity of 1kb per 10-30s of the Phusion® High-Fidelity DNA-Polymerase. For the OE-PCRs, in the first 17 cycles, template overlapping regions played a role of primers connecting the fragments in appropriate order. Primers in a final concentration of 800nM were added after the 17th cycle, for the next 18 cycles. OE-PCRs were performed with 2-3 templates, and their ratio was determined using the following formula:

ng (shorter fragment) = kb (shorter fragment) x ng (longer fragment) / kb (longer fragment)

ng = nano gram
kb = kilo base

Every PCR included at least one negative control, where no template was added. 35 cycle of following PCR-program were preformed for all PCRs: 98°C (30``)- 98°C (10``)- 51-58°C (20``)- 72°C (10-106``)- 72°C (2`)- 4°C (storage).

8.3.2 Colony PCR

Single colonies from YPD (antibiotic)-agar plates were picked and suspended in 20µl dH_2O, and heated at 95°C for 10min. After that the samples were centrifuged for 4-8sec at maximal speed (13.200rpm) and 5µl from the supernatant were used as a template. All primers were used in a final concentration of 400nM, dNTPs in a final concentration of 200µM. The elongation time was calculated in dependence of the expected PCR product length, considering a processivity of 1kb per 1min of the GoTaq® polymerase. Every PCR included at least one negative control, where no template was added, and one positive control using genomic DNA from *P. pastoris* CBS7435 as a template. 35 cycles of following PCR-program were preformed for all PCRs: 95°C (30``)- 95°C (10``)- 51-58°C (20``)- 72°C (1-3`)- 72°C (7`)- 4°C (storage).

8.4 Two steps site directed mutagenesis method

Two steps site directed mutagenesis was performed with PfuUltra™ High-Fidelity DNA-Polymerase, according to the manufacture's recommendation. All primers were used in a final concentration of 400nM, and dNTPs in a final concentration of 200µM. The circular plasmid to be mutated was used as template in concentration of 50ng/µl. The elongation time was calculated in dependence of the expected PCR product length (plasmid size), considering a processivity of 1kb per 1min of the PfuUltra™ High-Fidelity DNA-Polymerase. 18 cycles of following PCR-program were preformed: 95°C (30``)- 95°C (10``)- 52°C (1`)- 76°C (4,3`)- 72°C (7`)- 4°C (storage). T_m for the mutagenesis primers used in this experiment was calculated using following formula:

$T_m = 81.5 + 0.41(\%GC) - 675/N - \%$ mismatches

N = primer length in bp, without the bases to be inserted

The PCR products were digested with DpnI in order template plasmid to be eliminated and additional *E. coli* TOP10F` electro-competent cells were transformed with these PCR products.

8.5 Agarose gel electrophoresis

1 %-agarose gels were made with agarose and 1xTAE buffer and approximately 2µl EtBr (≥ 98 % ethidium bromide). For preparative-gels, DNA samples were mixed with 6x Loading Dye, and for control-gels they were mixed with 1x Loading Dye. Electrophoresis for preparative-gels was run for 80-100min at a voltage of 80-90V and for control-gels for 35-50min at a voltage of 110-120V. In general, small DNA fragments need shorter running time and large DNA fragments have to be run longer in order to get nicely separated bands. After every run of preparative-gel, the samples were cut out of the gel at the Chroma 43. All samples on the gels were analyzed with the GelDoc-It™ Imaging System. As standard, 5µl of 0.1µg/µl GeneRuler™1kb DNA Ladder were used.

8.6 Standard ligations

Ligation of an insert to a vector was done at 16 °C over night or at 22°C for 2-4 hours, with T4 DNA Ligase (400u/µl, 10xT4 DNA Ligase Reaction Buffer/New England BioLabs® Inc., Ipswich, MA, USA) according to the manufacturer's recommendations. The ratio of the fragments to be ligated was calculated using following formula:

ng (insert) = kb (insert) x ng (vector) / kb (vector)

ng = nano gram
kb = kilo base

8.7 Cloning to the pJET1.2/blunt ends vector

The CloneJET™ PCR cloning Kit (Fermentas GmbH, St. Leon-Rot, Germany) was used for cloning PCR products to the pJET1.2/blunt ends vector according to the manufacturer's recommendations. The incubation time for ligation was 20-25min at room temperature (RT) with a molar vector: insert ratio of 1:3. Desalting was done for 45-60min against dH_2O at room temperature prior to electroporation.

8.8 Transformations

8.8.1 Transformation to *Escherichia coli*

DNA (approximately 100ng, maximum volume of 10-20µl) and electro-competent cells (80 µl *E. coli* Top 10F') were mixed and incubated for 15min on ice. After that the cell suspension was transferred to electro cuvettes and pulsed with 2.5kV, 200ω, and 25µF for 4,1-4,6sec. 500µl SOC medium were added immediately and the cell suspension was mixed briefly by slowly pipetting up and down. The cell suspension was incubated for 45-60min at 37 °C and 500rpm. 20µl, 120µl and the rest of the cell suspension (cells concentrated through brief centrifugation) were plated on LB-agar plates containing the appropriate antibiotic. Incubation of the plates was done at 37 °C for approximately 24h (over night).

8.8.2 Transformation to *Pichia pastoris*

Transformation of *P. pastoris* was done as described by Lin-Cereghino et al. Competent cells were prepared by inoculating a single colony of the respective strain to 7-10ml YPD medium and grown over night at 28 °C, 110rpm. From the over night-culture, a main culture was grown from OD_{600} = 0.3 to a final OD_{600} of 0.8-1.0 (again in YPD medium). The cells were harvested by centrifugation at 4000rpm for 5min. After that, 9ml ice cold BEDS and 1ml 1M DTT were added to the cell pellet and the suspension was mixed briefly by hand for 5min, and centrifuged for 5min at 4000rpm. After centrifugation, the cells from the cell pellet were re-suspended in 1ml ice cold BEDS. 80µl of these electro-competent cells were mixed with 500ng to 2µg DNA (maximum volume of 20µl) and incubated for 15min on ice. After that, the cell suspension was transferred to electro cuvettes and pulsed with 1.5kV, 200ω, 25µF, and 1000µl 1M sorbitol were added immediately after the pulse. The cell suspension was mixed briefly by slowly pipetting up and down. The cells were incubated for 2h and 28 °C for regeneration, and then plated on YPD-agar plates containing the appropriate antibiotic(s). Aliquots of 50µl, 150 µl, 300µl and the rest of the cell suspension (cells concentrated through brief centrifugation) were plated and incubated at 28 °C for approximately 48-72h (2 - 3 days).

8.9 Micro-scale cultivation of *Pichia pastoris* in 96-deep well plates

In order to screen for positive *P. pastoris*, micro-scale cultivation in 96-deep well plates (DWP) was done similar to the protocol described by Weis et al. All DWPs were filled up with 300µl BMD1% per well, and each well was inoculated with cells from a single colony.

Three wells per DWP were inoculated with single colonies of strains used as a negative control, and 3-6 wells per DWP were inoculated with single colonies from strains used as a positive control. The DWPs were incubated at 28 °C, 320rpm, 80% humidity for 5 days (for the short protocol) or 7 days (for the long protocol). The induction for the short protocol was done after approximately 48h by adding 250µl BMM2 per well and after approximately 60h and 72h by adding 50µl BMM10 per well each time. The induction for the long (standard) protocol was done after approximately 60h by adding 250µl BMM2 per well and after approximately 72h, 84h and 96h by adding 50µl BMM10 per well each time. The cultivation was stopped 48h after the start of induction, and the cells were harvested by centrifugation at 4000rpm for 5min. For the photometric screening assays used in this thesis, samples were taken from the supernatant.

8.10 Photometric enzyme activity assays

8.10.1 HRP enzyme activity assay

After harvesting the cells from the micro-scale cultivation, 15µl supernatant from each well of the DWPs were transferred to microtiter plates mixed with 140µl ABTS (2,2'-Azino-bis-(3-ethylbenzthiazoline-6-sulfonic acid)) assay solution (1ml 20xABTS stock: 19ml 50mM NaOAc pH 4.5, and 1.75 µl 30 % (v/v) H_2O_2) and the absorbance at 405nm and 25°C was followed over 5min using the plate-reader (Spectramax Plus384, Molecular Devices, Sunnyvale, CA, USA).

8.10.2 CalB enzyme activity assay

After harvesting the cells from the micro-scale cultivation, 20µl supernatant from each well of the DWPs were transferred to microtiter plates. After that, 100µl substrate-solution (p-nitrophenol butyrate stock: 42µl pNPB 98% added to 458µl DMSO) were mixed with 10ml of 300mM Tris/HCl pH 7.0 buffer. The substrate-solution autolyzes and for that reason aliquots of the same were stored at -20°C. 180µl of the substrate-mixture were added to the 20µl supernatant and the absorbance at 405nm and 25°C was followed over 5min using plate-reader (Spectramax Plus384, Molecular Devices, Sunnyvale, CA, USA).

8.10.3 HNL enzyme activity assay

This assay for analyzing the enzymatic activity of HNL was done according to Weis, R. et al., FEMS Yeast Research, 2004. After harvesting the cells from the micro-scale cultivation, 20µl supernatant from each well of the DWPs were transferred to microtiter plates and mixed with 130µl of 1M citrate phosphate buffer pH 5.0. Shortly before the measurement was started, 50µl racemic 0,06M mandelonitrile solution (80mg racemic mandelonitrile, dissolved in 0.1M citrate phosphate buffer pH 3.0, 15mM final concentration), the absorbance at 280nm and 25°C was followed over 5min using plate-reader (Spectramax Plus384, Molecular Devices, Sunnyvale, CA, USA).

8.11 Green fluorescence protein (GFP) fluorescence measurement

GFP was produced intracellular in *P. pastoris* Mut^s strain. After micro scale cultivation was finished, GFP expression was measured with a plate reader (Spectramax Plus384, Molecular Devices, Sunnyvale, CA, USA), as follows:
- 200µl of undiluted suspended cell culture from each plate well were transferred to microtiter plates for measuring the optical density (OD_{600}) of the cells;
- after that, 200µl of suspended culture from each well were transferred to microtiter plates and mixed briefly;
- endpoint fluorescence measurement of the cell suspensions was provided by excitation maximum at 395nm, emission maximum at 507nm, and cutoff at 495nm;

9 Bioinformatics' programs and web tools

9.1 Programs

- The results of photometric screening assays were analyzed using SoftMax Pro 4.8/Molecular Devices, Ismaning, Germany.
- The DNA sequencing results were analyzed using DNASTAR Lasergene®, USA (SeqMan and EditSeq).
- The plasmid and cassettes maps, as well as the primers, were designed using Vector NTI® Bioinformatics software package from Invitrogen, USA.

9.2 Web tools

- Multiple sequence alignments were provided by ClustalW2: http://www.ebi.ac.uk/Tools/msa/clustalw2/
- Sequences searches were provided by BLAST (Basic Local Alignment Search Tool) - NCBI: http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome
- *Pichia pastoris* DNA sequences were taken from GenDB: http://www.cebitec.uni-bielefeld.de/groups/brf/software/gendb_info/
- DNA sequencing results: https://shop.lgcgenomics.com/

10. Pichia pastoris PDI-specific P_{AOX1}-replacement cassettes

10.1 Project aim

The aim of this project was to design four *PpPDI*-specific P_{AOX1}-replacement cassettes. The replacement of the natural *PpPDI* promoter with PpP_{AOX1} should result in *Pp*PDI overexpression. Additionally, four *P. pastoris* ΔKU70 strains carrying genes coding for the proteins CalB and HRP were designed to evaluate the effect of *Pp*PDI overexpression.

10.2 Project strategy

10.2.1 *Pichia pastoris* protein disulfide isomerase genes (*PpPDIs*)

In the recently performed genome sequencing project of *P. pastoris* CBS7435, four genes coding for protein disulfide isomerase have been detected and annotated as it follows: PP7435_Chr4-0183 (CCA40358.1), PP7435_Chr1-0128 (CCA36295.1), PP7435_Chr1-0470 (CCA36622.1), and PP7435_Chr4-0107 (CCA40283.1). They vary in the length. PP7435_Chr4-0183 is 876bp, PP7435_Chr1-0128 897bp, PP7435_Chr1-0470 1110bp, and PP7435_Chr4-0107 1554bp long.

Unlike the length, the GC content does not differ drastically between all four gene sequences, with a minimum of 39% in PP7435_Chr1-0470 and a maximum of 43% in PP7435_Chr4-0107. All four genes do not contain introns, and all of them contain two thioredoxin domains, with the exception of PP7435_Chr4-0183, which contains three. All four expressed PDIs are supposed to be located in the lumen of the endoplasmic reticulum and to catalyze the rearrangement of -S-S- bonds in proteins (Table 8).

Gene designation (GenDB)	PP7435_Chr4-0107	PP7435_Chr1-0128	PP7435_Chr1-0470	PP7435_Chr4-0183
NCBI Accession number	CCA40283.1	CCA36295.1	CCA36622.1	CCA40358.1
Start/Stop codon	ATG/TAA	ATG/TAG	ATG/TAA	ATG/TAG
Length [bp]	1554	897	1110	876
GC content [%]	43,308	42,029	39,189	41,438
Introns	no	no	no	no
AA sequence length	517	298	369	291
Conserved domains	2 thioredoxin domains	2 thioredoxin domains	2 thioredoxin domains	3 thioredoxin domains
Subcellular location	lumen of endoplasmic reticulum	lumen of endoplasmic reticulum	lumen of endoplasmic reticulum	lumen of endoplasmic reticulum
Gene product	prolyl 4-hydroxylase	not described	protein disulfide isomerase family A	protein disulfide isomerase family A
EC number	5.3.4.1	not described	5.3.4.1	5.3.4.1
Gene description	protein disulfide isomerase	not described	protein disulfide isomerase 1	protein disulfide-isomerase A4

Table 8: Main observations about *P. pastoris PDIs* made in the genome sequencing project.

A multiple sequence alignment of all four *PDI* coding domains was provided using ClustalW2. It revealed DNA sequence similarities of 55% to 67% (Table 9; Appendix 3). The multiple sequence alignment of the 3000bp long upstream and 1000bp long downstream region, showed sequence similarities between 48% and 52% (Table 10 and 11; Appendix 5). Also a multiple sequence alignment of the AA sequences of all four PDIs was performed. It showed low sequence similarity of 17% to 22% (Table 12; Appendix 6).

SeqA	Name	Length	SeqB	Name	Length	Score
1	CCA40283.1	1554	2	CCA36295.1	897	67.0
1	CCA40283.1	1554	3	CCA36622.1	1110	60.0
1	CCA40283.1	1554	4	CCA40358.1	876	64.0
2	CCA36295.1	897	3	CCA36622.1	1110	60.0
2	CCA36295.1	897	4	CCA40358.1	876	55.0
3	CCA36622.1	1110	4	CCA40358.1	876	57.0

Table 9: Summary of DNA multiple sequence alignment of all four *PpPDIs*.

SeqA	Name	Length	SeqB	Name	Length	Score
1	CCA40283.1	3000	2	CCA36295.1	3000	49.0
1	CCA40283.1	3000	3	CCA36622.1	3000	50.0
1	CCA40283.1	3000	4	CCA40358.1	3000	50.0
2	CCA36295.1	3000	3	CCA36622.1	3000	51.0
2	CCA36295.1	3000	4	CCA40358.1	3000	50.0
3	CCA36622.1	3000	4	CCA40358.1	3000	51.0

Table 10: Summary of DNA multiple sequence alignment of 3000 bp upstream region of all four *PpPDIs.*

SeqA	Name	Length	SeqB	Name	Length	Score
1	CCA40283.1	1000	2	CCA36295.1	1000	49.0
1	CCA40283.1	1000	3	CCA36622.1	1000	52.0
1	CCA40283.1	1000	4	CCA40358.1	1000	50.0
2	CCA36295.1	1000	3	CCA36622.1	1000	50.0
2	CCA36295.1	1000	4	CCA40358.1	1000	48.0
3	CCA36622.1	1000	4	CCA40358.1	1000	50.0

Table 11: Summary of DNA multiple sequence alignment of 1000 bp downstream region of all four *PpPDIs.*

SeqA	Name	Length	SeqB	Name	Length	Score
1	CCA40283.1	517	2	CCA36295.1	298	22.0
1	CCA40283.1	517	3	CCA36622.1	369	18.0
1	CCA40283.1	517	4	CCA40358.1	291	14.0
2	CCA36295.1	298	3	CCA36622.1	369	19.0
2	CCA36295.1	298	4	CCA40358.1	291	15.0
3	CCA36622.1	369	4	CCA40358.1	291	17.0

Table 12: Summary of DNA multiple sequence alignment of all four *PpPDIs.*

10.2.2 Amplification, cloning and sequencing of *PpPDI*s

All four *PDI* genes from *Pichia pastoris* CBS7435 strain (PP7435_Chr4-0107, PP7435-Chr1-0128, PP7435_Chr1-0470, and PP7435_Chr4-0183) were amplified together with their natural promoter and terminator. Thereby, the promoter and terminator region were approximately 3000bp and 500bp long, respectively. For that aim, genomic DNA (gDNA) from *P. pastoris* CBS7435 strain was isolated according to the protocol of Hoffman *et al.* Since each *PpPDI*-sequence to be amplified was longer than 4000bp, they were amplified as three shorter pieces by using three different primer pairs. After that, the three pieces were joined together by OE-PCR. (Table 13; Figure 22-25)

Cassettes located gene	PP7435_Chr4-0107	PP7435_Chr1-0128	PP7435_Chr1-0470	PP7435_Chr4-0183
Cassette length [bp]	5365	4754	4880	4763
Primer used for whole cassette amplification	p-09-514 / p-10-310	p-09-516 / p-10-312	p-09-518 / p-10-056	p-09-520 / p-10-058
Length of fragment part 1 [bp]	1884	1516	1723	1731
Primer used for fragment part 1	p-09-514 / p-10-052	p-09-516 / p-10-054	p-09-520 / p-10-057	p-09-521 / p-10-059
Length of fragment part 2 [bp]	3235	2888	3266	2730
Primer used for fragment part 2	p-09-515 / p-10-053	p-09-517 / p-10-055	p-10-313 / p-10-314	p-10-315 / p-10-316
Length of fragment part 3 [bp]	588	633	457	680
Primer used for fragment part 3	p-10-309 / p-10-310	p-10-311 / p-10-312	p-09-518 / p-10-314	p-09-520 / p-10-316
Length of OE-region 1 [bp]	202	66	375	190
Length of OE-region 2 [bp]	140	217	191	188

Table 13: Primer pairs used for amplification of the *Pp*PDI cassettes and cassette fragments including the their lengths and the lengths of the overlap extension (OE) regions.

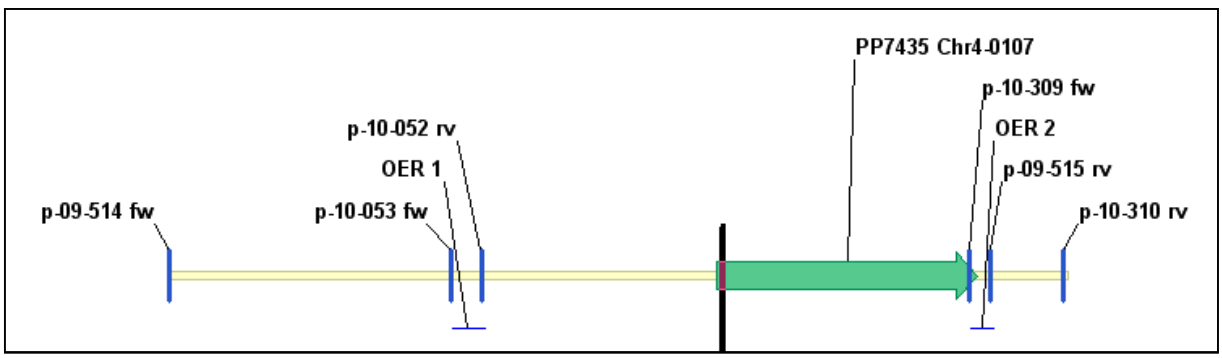

Figure 22: Whole amplified PP7435_Chr4-0107 sequence, including the overlap extension regions 1 and 2 (OER1 and OER2) and the three primer pairs which were used.

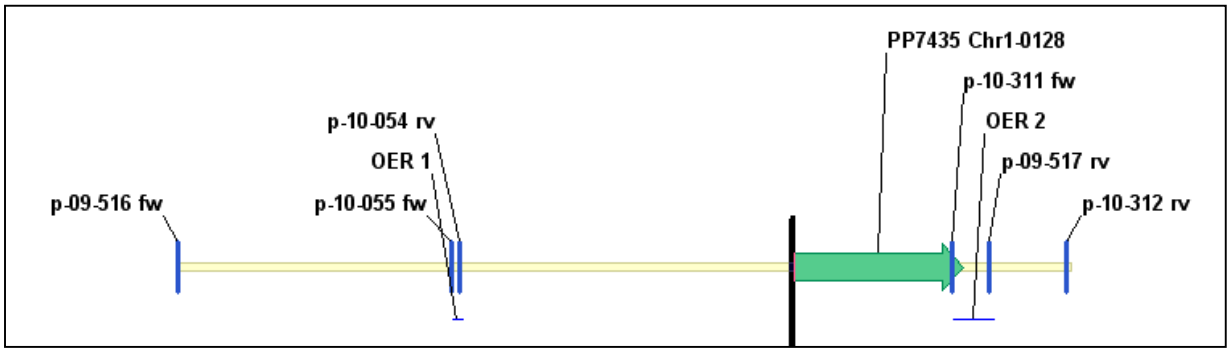

Figure 23: Whole amplified PP7435_Chr1-0128 sequence, including overlap extension regions 1 and 2 (OER1 and OER2) and the three primer pairs which were used.

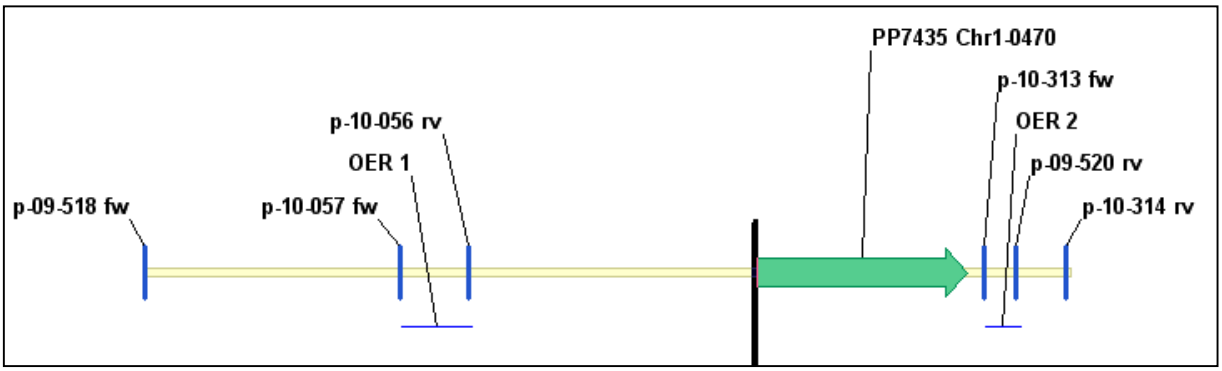

Figure 24: Whole amplified PP7435_Chr1-0470 sequence is shown, including overlap extension regions 1 and 2 (OER1 and OER2) and the three primer pairs which were used.

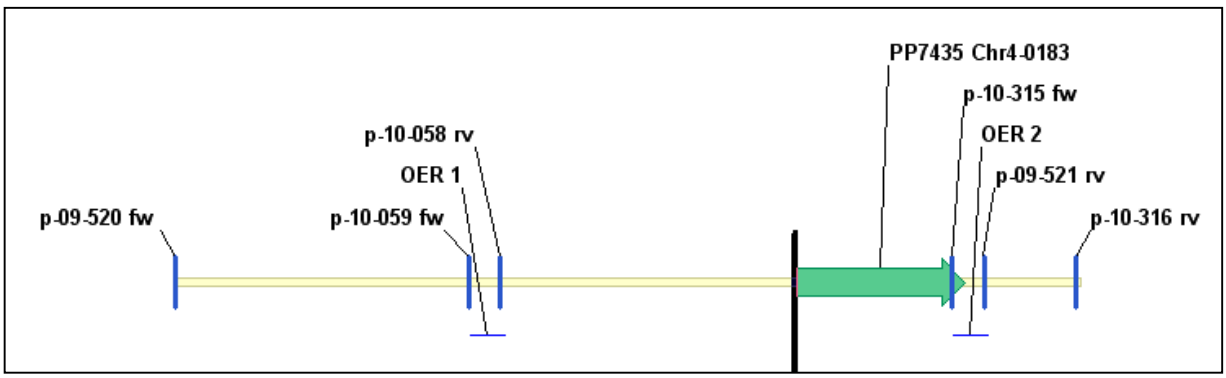

Figure 25: Whole amplified PP7435_Chr4-0183 sequence is shown, including overlap extension regions 1 and 2 (OER1 and OER2) and the three primer pairs which were used.

After PCR amplification of all sequences and their rejoining, they were ligated into pJet1.2/ blunt ends cloning vector using T4-DNA ligase. *E. coli* TOP10F` electro-competent cells were transformed with this plasmid, and plated on LB-ampicillin plates. After that, pJet plasmids from each strain were isolated using GeneJET™ Plasmid Miniprep Kit. The ligated *PpPDI* sequences were amplified from the plasmids and sent for sequencing. Using respective forward and reverse primer (Table 14), both DNA strands of all *PDI* sequences were fully sequenced. Terminator regions were sequenced using standard pJet sequencing primers: p-08-483_fw and p-08-484_rv. All samples were sequenced by LGC Genomics Company (Berlin, Germany) using Sanger Technology.

PP7435_Chr4-0107	PP7435_Chr1-0128	PP7435_Chr1-0470	PP7435_Chr4-0183
p-09-514 fw	p-09-516 fw	p-09-518 fw	p-09-520 fw
p-09-515 rv	p-09-517 rv	p-09-519 rv	p-09-521 rv
p-10-052 rv	p-10-054 rv	p-10-056 rv	p-10-058 rv
p-10-053 fw	p-10-055 fw	p-10-057 fw	p-10-059 fw
p-10-136 fw	p-10-134 fw	p-10-130 rv	p-10-127 fw
p-10-137 fw	p-10-135 fw	p-10-131 fw	p-10-128 fw
p-10-138 fw	p-10-253 rv	p-10-132 fw	p-10-129 fw
p-10-139 fw	p-10-254 fw	p-10-133 fw	p-10-267 rv
p-10-245 rv	p-10-255 rv	p-10-259 rv	p-10-268 rv
p-10-246 rv	p-10-256 rv	p-10-260 rv	p-10-269 rv
p-10-247 fw	p-10-257 rv	p-10-261 rv	p-10-270 rv
p-10-248 rv	p-10-258 fw	p-10-262 fw	p-10-271 rv
p-10-249 rv	p-10-311 fw	p-10-263 fw	p-10-272 fw
p-10-250 rv	p-10-312 rv	p-10-264 fw	p-10-273 fw
p-10-251 rv	/	p-10-265 rv	p-10-315 fw
p-10-252 fw	/	p-10-266 rv	p-10-316 rv
p-10-309 fw	/	p-10-313 fw	/
p-10-310 rv	/	p-10-314 rv	/

Table 14: Primers used for sequencing of four *PpPDI* genes including their promoter and terminator regions.

The obtained sequences were compared with the corresponding NCBI sequences. Thereby, several mismatches have been detected:
- for the gene PP7435_Chr1-0128 (CCA36295.1): 19bp downstream of the gene stop codon, the sequence contains 13T instead of 11T as recorded in NCBI;
- for the gene PP7435_Chr1-0470 (CCA36622.1): 509bp downstream of the gene start codon, the sequence contains 9T instead of 8T as recorded in NCBI;

Interestingly, all *P. pastoris* CBS7435 sequences stored in NCBI have been obtained using new generation sequencing method (454 Life Sciences, Roche). Küberl *et al.* have already reported mismatches found in homo-polymer stretches by comparison of the sequences obtained using 454 Illumina and Sanger (LGC Genomics, Germany)[6] technology. This can also explain the sequence mismatches detected within the *PpPDIs*, since all mismatches were located in poly-T regions.

10.2.3 Design of *PpPDI*-specific P_{AOX1}-replacement cassettes

Based on existing FLP-recombinase knockout cassettes (Figure 20 and 21), two variants (variant A and B) of *PpPDI*-specific P_{AOX1}-replacement cassettes were initially designed *in silico* (Figure 26 and 27). The differences between these two cassette variants could theoretically have an influence on the transformation rate, the strength of the promoter and also the expression of the desired *PpPDI*. Namely, the cassette variant A contains one P_{AOX1} and is approximately 1kb shorter than the cassette variant B. This should theoretically be an advantage for the transformation. On the other hand, in the cassette variant A one FRT sequence stays between the PpP_{AOX1} and the *PpPDI* after recombination, and could negatively influence the promoter and also the *PDI* expression. For that reason cassette variant B was designed. It contains two P_{AOX1} and after recombination one of the two P_{AOX1} stayed in the genome and is connected to the *PDI* directly, without any additional sequences in between.

Variant A was built out of four parts (part I-IV) (Figure 26). Part I contained only the PpP_{AOX1}. Part II contained the FRT-target sequence, recombinase (*FLP*) gene and Cyc terminator. Part III contained the *P. pastoris/E. coli* codon optimized synthetic zeocin resistance gene, placed under the control of *PpILV5* promoter and *Pp*Aod terminator. Part IV contained again a FRT-target sequence and 500bp long *PpPDI*-specific homologous sequence (start codon plus 500bp downstream region).

Variant B was built out of three parts (part I-III) (Figure 27). Part I contained the FRT target sequence and recombinase (*FLP*) gene, placed under the control of PpP_{AOX1} and CycTT. Part II contained the *P. pastoris/E. coli* codon optimized synthetic zeocin resistance gene, placed under the control of *PpILV5* promoter and *Pp*Aod terminator. Part III contained the FRT target sequence, PpP_{AOX1}, and 500 bp long *PpPDI*-specific homologous sequence, as described above.

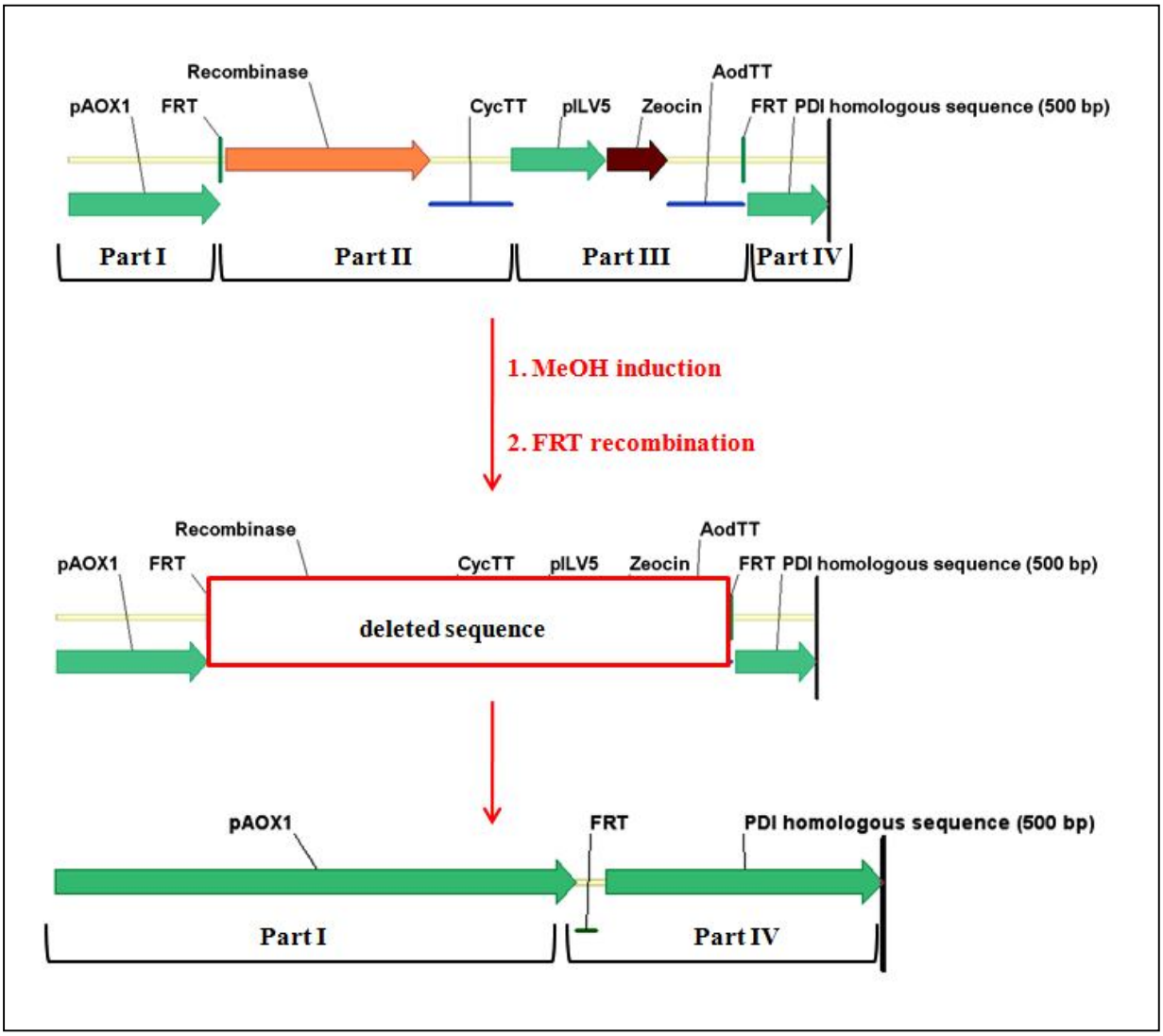

Figure 26: *PpPDI*-specific PpP_{AOX1}-replacement cassette / Variant A. The upper part of the figure shows the recombination cassette including all four parts (Part I-IV). The middle part shows the part of the cassette which should be deleted after MeOH induction and recombinase expression. The bottom part of the figure shows the parts of the cassette (Part I and Part II) staying in the genome after FRT recombination process. pAOX1 (PpP_{AOX1}); FRT (FRT target sequence); Recombinase (gene coding for FLP* Recombinase); CycTT** (*S. cerevisiae* iso-1-cytochrome c terminator); pILV5 (eukaryotic promoter); Zeocin (gene coding for zeocin resistance); AodTT (AOD terminator); PDI homologous sequence (500bp long sequence homologous to the 500bp upstream region of each *PpPDI*);

* NCBI accession number: J01347.1 / FLP and FRT present *S. cerevisiae* 2-micron plasmid derived recombinase system.

** NCBI accession number: L26347.1 / Presents *S. cerevisiae* iso-1-cytochrome c terminator.

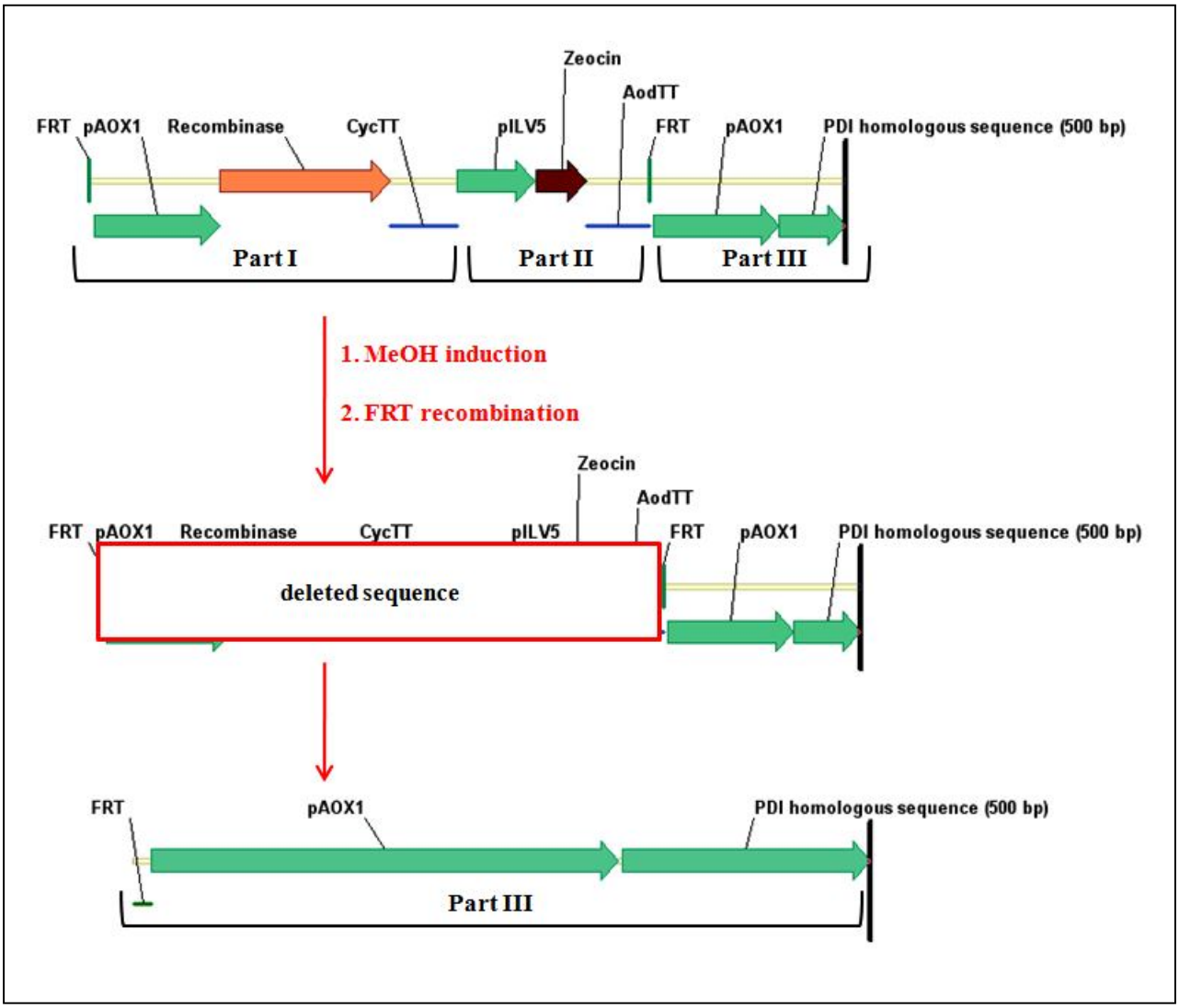

Figure 27: *PpPDI*-specific *PpP*$_{AOX1}$-replacement cassette / Variant B. The upper part of the figure shows the recombination cassette including all four parts (Part I-IV). The middle part shows the part of the cassette deleted after MeOH induction and recombinase expression. The lowest part of the figure shows the parts of the cassette (Part I and Part II) staying in the genome after FRT recombination process. FRT (FRT target sequence); pAOX1 (*PpP*$_{AOX1}$); Recombinase (gene coding for FLP Recombinase); CycTT (*S. cerevisiae* iso-1-cytochrome c terminator); pILV5 (eukaryotic promoter); Zeocin (gene coding for zeocin resistance); AodTT (AOD terminator); PDI homologous sequence (500bp long sequence homologous to the 500bp upstream region of each *PpPDI*);

After transformation the cassette should integrate into the *P. pastoris* genome according to the homologous sequence used (PDI locus). The obtained clones should be cultivated and induced with MeOH for the FLP-recombinase to be expressed. FLP-recombinase catalyzes the recombination of both FRT target sequences (Paragraph 2.4). After recombination, part I and part II should be excised from the *P. pastoris* genome, and the *PpP*$_{AOX1}$ should be integrated upstream of the start codon of each *PpPDI* (Figure 26 and 27).

Since it is known that shorter recombination cassettes integrate easier in genome, the cassette variant A which is shorter, was favored. However, by using cassette variant A, one FRT target sequence stays between *PpP*$_{AOX1}$ and the 500 bp long *PpPDI*-specific homologous sequence after FRT recombination. Since it was not known if this FRT will influence the strength of the promoter or not, a cassette validation experiment was done first.

10.2.3.1 Evaluation of efficiency of the cassette variant A

For the evaluation of the efficiency of cassette variant A, green fluorescent protein (GFP) was used as reporter protein. The existing plasmid pPpT4_SmiI_Cycle3GFP was mutated according to the "Two Steps Site Directed Mutagenesis" protocol from Stratagene.

The first mutagenesis approach included insertion of a FRT target sequence between the Kozak consensus sequence of $P_{AOX1syn}$ and the start codon of *GFP* (ATG) (Figure 28), similar to the existing FRT-recombinase cassette (FLIPPER_pAOX1).

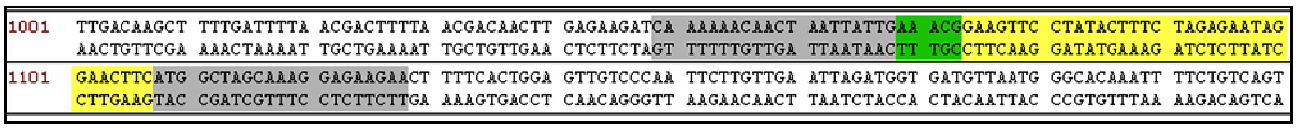

Figure 28: Region of mutagenesis (the gray-green-yellow-gray colored mutagenesis primer) including one Kozak sequence. FRT target sequence is colored yellow, Kozak sequence is colored green, and the rests of the mutagenesis primer are colored gray.

The second mutagenesis approach included insertion of two Kozak consensus sequences located upstream and downstream of the FRT target sequence of $P_{AOX1syn}$ (Figure 29).

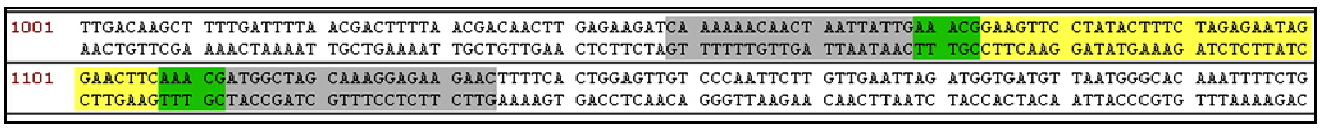

Figure 29: Region of mutagenesis (the gray-green-yellow-gray colored mutagenesis primer) including two Kozak sequences. FRT target sequences are colored yellow, Kozak sequence is colored green, and the rests of the mutagenesis primer are colored gray.

Mutagenesis primer containing one FRT (p-10-561 and p-10-562), or two FRTs (p-10-575 and p-10-576) were used. *E. coli* TOP10F` electro-competent cells were transformed wuth the mutated plasmids (Figure 30). The cells were plated on LB-zeocin plates and incubated according to the standard protocol. After that, 3 randomly chosen clones were streaked out, and incubated for another 24h. From these 3 clones, plasmids were isolated using GeneJET™ Plasmid Miniprep Kit. As the insertion of only 34 or 68bp, which is the length of one or two FRTs, is too small to be detected on an agarose gel, plasmids were sent for sequencing without prior control restriction step. One positive (sequence verified) clone of each transformation was then linearised with *Smi*I over night, and *P. pastoris* Mut^S electro-competent cells were transformed with it. In parallel, pPpT4_SmiI_Cycle3GFP unmutated plasmid was also linearised over night with SmiI, and *P. pastoris* Mut^S electro-competent cells were transformed, and plated on YPD-zeocin plates. After that, 90 clones from the first transformation (carrying plasmids with one FRT), 45 clones from the second transformation

(carrying plasmids with two FRTs), and 144 clones from the third transformation (carrying unmutated plasmid) were cultivated in deep well plates (DWPs). Thereby one half of each DWP was inoculated with clones carrying mutated and the other half with the clones carrying unmutated plasmid.

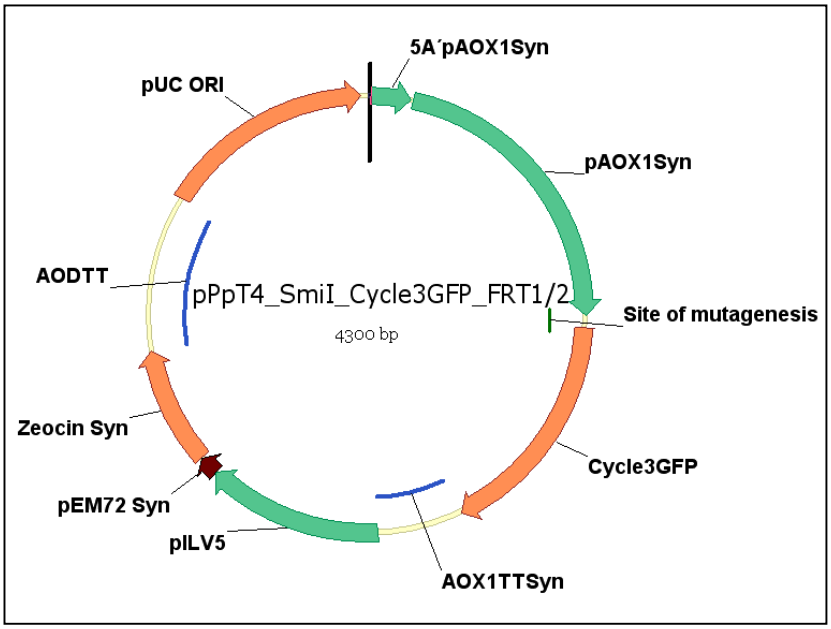

Figure 30: pPpT4_SmiI_Cycle3GFP_FRT1/2 represents both, plasmid carrying one and two FRT sequences. pUC_ORI (pUC origin of replication); PAOX1_Syn_dBamHI (synthetic *AOX*1 promoter upstream region); PAOX1_Syn (synthetic *AOX*1 promoter region); Site of mutagenesis (sequence containing one or two FRT target sequences); Cycle3GFP (synthetic gene coding for green fluorescence protein); AOX1TT_Syn (synthetic terminator); pILV5 (eukaryotic promoter); pEM72Syn (synthetic prokaryotic promoter); Zeocin_Syn (gene for zeocin resistance); AODTT (*AOD* terminator);

10.2.3.2 Results and discussion

For calculation of the final values, the measured fluorescence values were normalized by OD_{600} values. The average negative control was subtracted in addition. The final values for clones carrying plasmids with integrated FRT target sequences were than compared with that for clones carrying pPpT4_SmiI_Cycle3GFP empty plasmid (Figure 31, 32 and 33).

All clones carrying 1 or 2 FRTs showed drastically lower fluorescence values than clones carrying an unmutated reporter plasmid. More precisely, in DWP1, the highest measured value for clones carrying unmutated plasmid was 5149,34 RFU (relative fluorescence units), while clones carrying plasmid with one FRT was 316,80 RFU only. Approximately the same ratio was seen for the DWP2 and DWP3.Comparing low producer clones for all three DWPs, the control plasmid based strains reached the level of 500 RFU, while for the clones carrying one or two FRTs, only 50 and 30RFU were measured, respectively. The average ratios of the activity of clones carrying unmutated plasmid and clones carrying plasmids with FRT vary between 10,5 and 8,5 for clones with one and two FR-target sequences, respectively.

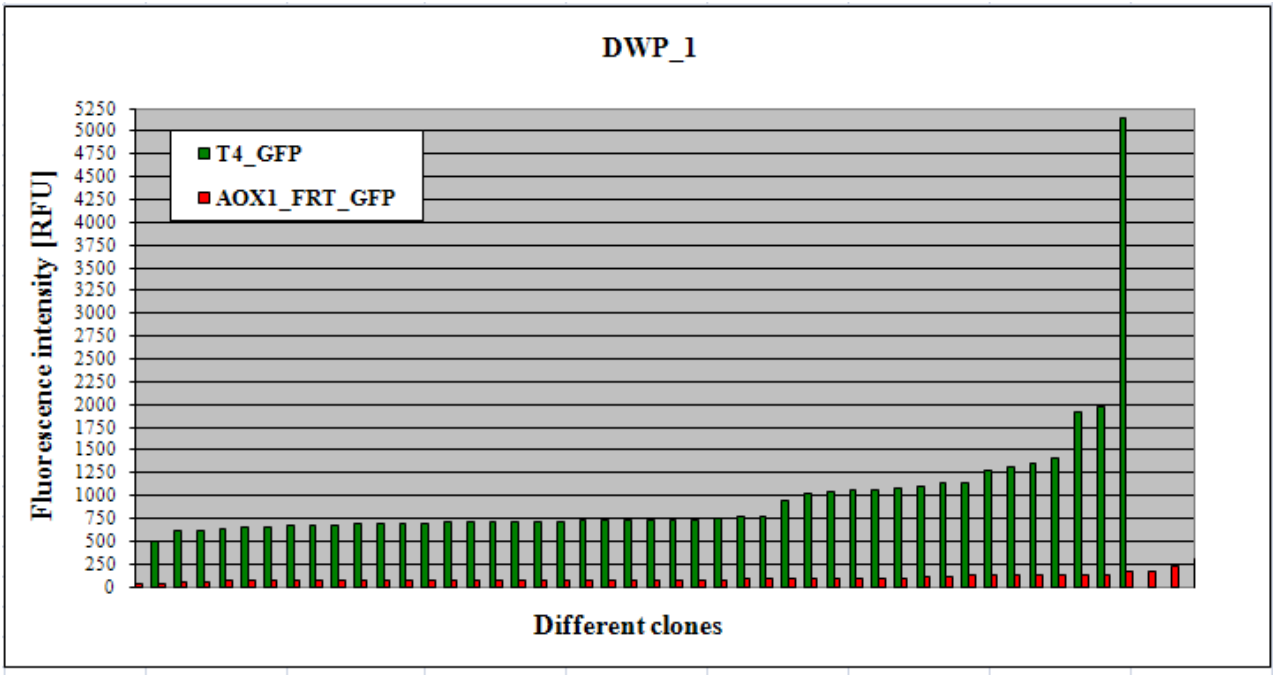

Figure 31: GFP fluorescence measurement screening results of DWP No.1. The green colored clones carry PpT4-GFP-AOX1Syn (unmutated plasmid), and the red colored clones carry PpT4-GFP-AOX1Syn with integrated one FRT target sequence (mutated plasmid). From 93 clones screened in total, only one was negative, showing activity in the range of the negative control.

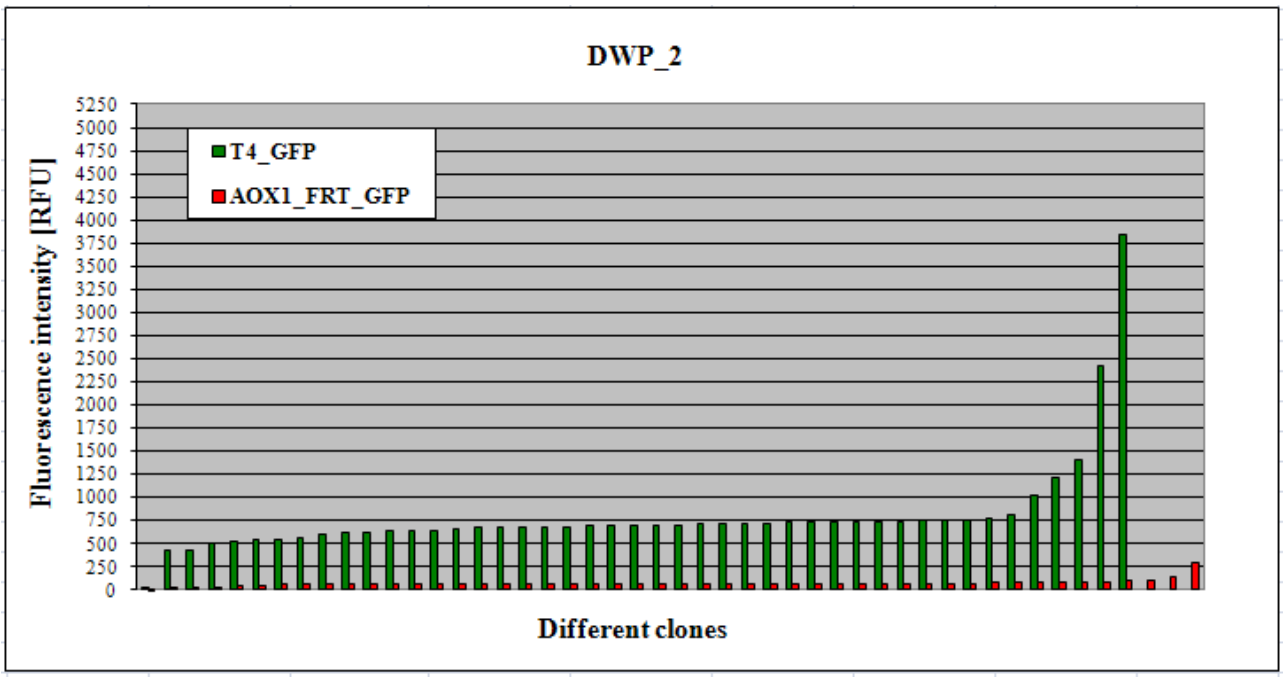

Figure 32: GFP fluorescence measurement screening results of DWP No.2. The green colored clones carry PpT4-GFP-AOX1Syn (unmutated plasmid), and the red colored clones carry PpT4-GFP-AOX1Syn with integrated one FRT target sequence (mutated plasmid). From 93 clones screened in total, three clones were negative, showing activity in the range of the negative control.

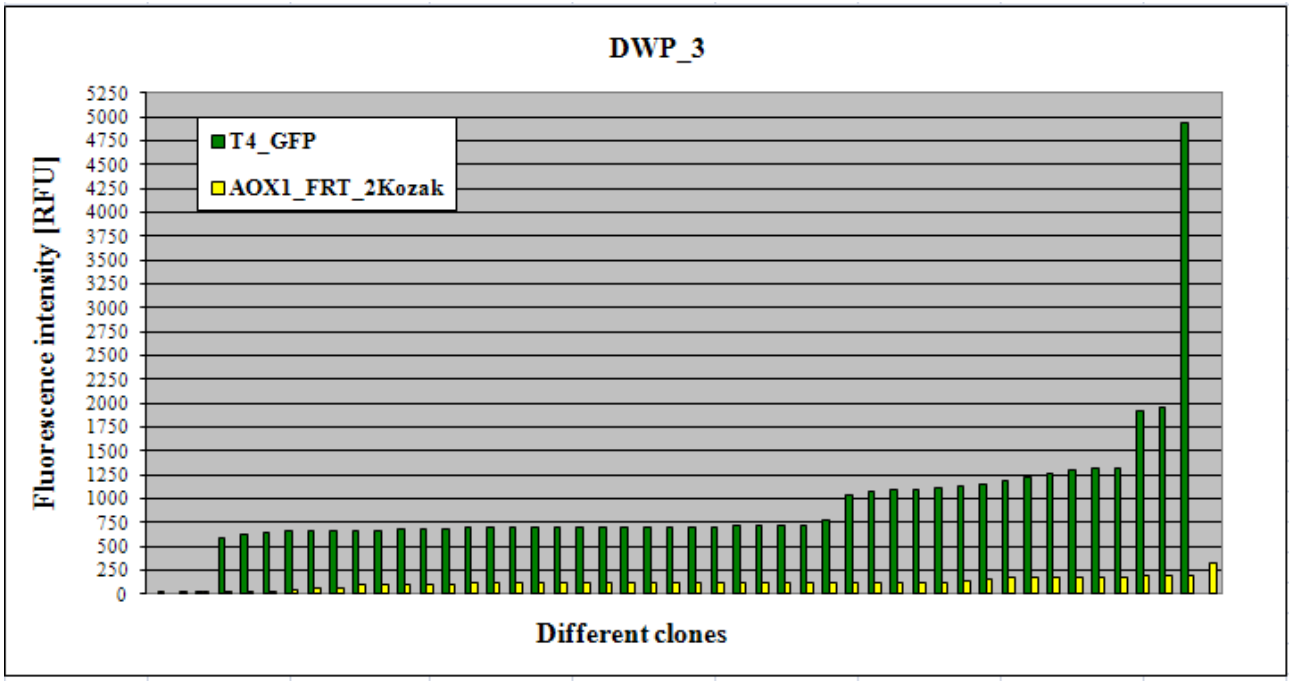

Figure 33: GFP fluorescence measurement screening results of DWP No.3. The green colored clones carry PpT4-GFP-AOX1Syn (unmutated plasmid), and the red colored clones carry PpT4-GFP-AOX1Syn with integrated two FRT target sequences (mutated plasmid). From 93 clones screened in total, seven clones were negative, showing activity in the range of the negative control.

From these results, a clear conclusion could be made, namely that a 34bp long FRT target sequence placed upstream of the target gene has a dramatically negative influence on expression level, (8 times less fluorescence). Also an insertion of a suitable Kozak consensus sequence right before the start codon did not improve the situation. A possible explanation for this can be found in the nature of the FRT target sequence. The two complement flanking ends of the FRT can join together, thereby making 13 bp long hairpin, and a loop of 8 bp. This intramolecular DNA base pairing might result in diverging of DNA and Polymerase.

Since the final aim of using this cassette is overexpression of the desired *PpPDI*, the cassette variant A as shown in this approach, is not convenient for this aim. For that reason, further work was focused on developing the cassette variant B.

10.2.3.3 Design of cassette variant B

The task within this part of the project was the redesign of the existing FRT-recombinase cassette (HIS4_Flipper_old_FRT) into *PpPDI*-specific P_{AOX1}-replacement cassette variant B (Figure 34; see also Figure 27). For that aim, the eight different cassettes' parts were separately amplified using PCR, and joined together using OE-PCR.

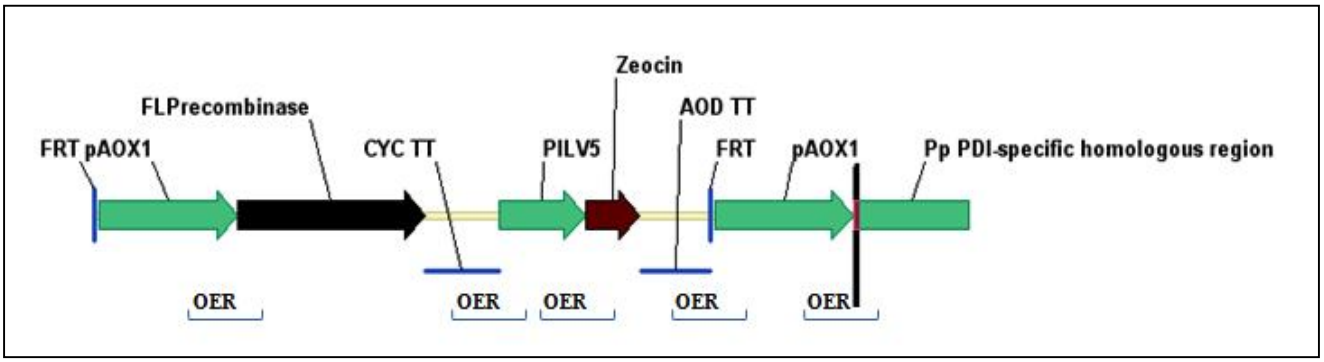

Figure 34: Construction of *PpPDI*-specific PpP_{AOX1}-replacement cassette / Variant B. PpPDI-specific homologous region, presents the homologous 3'-flanking end of the cassette. It is specific for each PDI and it varies in size from 528 to 813bp. OER is the abbreviation for overlap extension region, and shows the site of connection of the separately amplified cassettes' parts.

The first part, PpP_{AOX1} (FRT-pAOX1), was amplified using the primer p-10-594(fw) and p-10-589(rv), and pPpKan_SmiI plasmid as a template. It was 973bp long, and contained one FRT target sequence on the 5'-terminus.

The second part, FLPrecombinase-CYCTT, was amplified using the overlapping primer p-08-136(fw) and p-07-638(rv). This part was 1809bp long, and it was amplified from a pJet plasmid carrying the FLIPPER_PAOX1 cassette.

The third part, PILV5 (PpP_{ILV5} promoter), was amplified from *P. pastoris* CBS7435 genomic DNA, using p-10-866(fw) and p-10-593(rv) primer, and it was 619bp long.

The fourth part, Zeocin-AODTT, was amplified from the pJet1.2 plasmid carrying the FLIPPER_PAOX1 cassette. To do so, two overlapping primer, p-10-592(fw) and p-10-867(rv) were used. The p-10-592(fw) primer was designed to contain homologous sequences from both, PILV5 and Zeocin in order to ensure overlapping regions for joining PILV5 and Zeocin and deleting the prokaryotic promoter EM72 from the existing HIS4_Flipper_old_FRT cassette (Figure 35). EM72 is an unnecessary prokaryotic promoter sequence integrated into the ILV5 promoter and was therefore deleted. The total length of the fourth part was 919 bp.

```
2               CCGAAAAAATTTTCTTCCCTTCTCTTCCAAATATCGTCTCCACAAA---------- 586
3               -------------------------CCAAATATCGTCTCCACAAA---------- 20
1               CCGAAAAAATTTTCTTCCCTTCTCTTCCAAATATCGTCTCCACAAATCTAGAGTGTTGAC 600
                                         ********************

2               --------------------AAAATA---------------------------------A 593
3               --------------------AAAATA---------------------------------A 27
1               ACTTTATACTTCCGGCTCGTATAATACGACAAGGTGTAAGGAGGACTAAACCATGGCTAA 660
                                   * ****                                   *

2               ACTCACCTCTGCTGTTCCAGTCCTGACTGCTCGTGATGTTGCTGGTGCTGTTGAGTTCTG 653
3               ACTCACCTCTGCTGTTCC------------------------------------------ 45
1               ACTCACCTCTGCTGTTCCAGTCCTGACTGCTCGTGATGTTGCTGGTGCTGTTGAGTTCTG 720
                ******************
```

Figure 35: Primer p-10-592 (sequence number 3 in this figure) was used for deletion of EM72 prokaryotic promoter from the existing cassette. The sequence number 1 depicts the end of the PpP_{ILV5} and the beginning of the *Zeocin* coding domain (where EM72 is already deleted). The sequence number 2 depicts the preexisting part of the cassette including EM72 prokaryotic promoter.

PpPDI-specific homologous region were different for all four cassettes. They were amplified from *P. pastoris* CBS7435 genomic DNA (Table 15). After all parts of the cassettes were separately amplified, they were connected by several sequential steps of OE-PCR. In a first step the PpP_{AOX1} and *Recombinase*-CycTT parts were connected using the primer p-10-813(fw) and p-07-638(rv). Since PpP_{AOX1} was two times in each cassette, primer p-10-813(fw) ends with a linker sequence at its 5'-terminus, in order to avoid a double primer binding site in the final OE-PCR step. In the second OE-PCR step, the PpP_{ILV5} and *Zeocin*-AODTT parts were connected using the primer p-10-866(fw) and p-10-867(rv). Within the third OE-PCR step, PpP_{AOX1} part was connected with each *PpPDI*-specific part. During the final OE-PCR step PpP_{AOX1}-*Recombinase*-CycTT, PpP_{ILV5} -*Zeocin*-AODTT, and PpP_{AOX1} -*PpPDI*-specific parts were connected.

All amplified parts (fragments) were purified using preparative agarose gel electrophoresis, and Wizard® SV Gel and PCR Clean-Up System from Promega. Their sizes were controlled using 1kbDNA-ladder as standard, and their concentrations were measured by NanoDrop spectrophotometer. The whole cassettes were ligated with pJet1.2/blunt ends cloning vector. After that *E. coli* TOP10 F` electro-competent cells were transformed with the plasmid, and plated on LB-ampicillin plates.

Cassettes located gene	PP7435_Chr4-0107	PP7435_Chr1-0128	PP7435_Chr1-0470	PP7435_Chr4-0183
Length of the *PDI* specific homologous region	813	528	548	589
Primer used for the PDI specific homologous region amplification	p-10-588 / p-10-587	p-10-586 / p-10-585	p-10-584 / p-10-583	p-10-582 / p-10-581
Primer used for the third OE-PCR step of cassette construction	p-10-594 / p-10-587	p-10-594 / p-10-585	p-10-594 / p-10-583	p-10-594 / p-10-581
Primer used for the final OE-PCR step (whole cassette amplification)	p-10-813 / p-10-587	p-10-813 / p-10-585	p-10-813 / p-10-583	p-10-813 / p-10-581

Table 15: Overview of OE-PCR details connected to the *PpPDI*-specific promoter replacement cassette construction.

Regarding the construction of *PpPDI*-specific P_{AOX1}-replacement cassettes, difficulties occurred during the final OE-PCR step. Namely, as P_{AOX1} was used two times in one cassette, the forward primer (p-10-594) was binding more strongly to the second P_{AOX1}, resulting in high concentration of a short incorrect band instead of the band of full length cassette. However, this problem could be solved by T_m gradient OE-PCR optimization. The second and less expected difficulty was that after sequencing the full length cassettes point mutations were found in almost each of them. For this reason, additional PCRs and OE-PCRs, as well as restrictions and ligations were done in order to eliminate these.

10.3 Design of *Pichia pastoris* reporter strains carrying genes coding for model proteins CalB and HRP

To evaluate the effect of *PpPDI* up-regulation *Candida antartica* lipase (CalB) and *Horseradish* peroxidase (HRP) were used as model proteins. For that reason, within this thesis four new *P. pastoris* ΔKU70 strains were designed, two of them carrying a *CalB*, and the other two carrying an *HRP* gene. For each model protein one single copy and one multi copy strain were chosen from the obtained screening landscapes.

Since *PpPDI*-specific P_{AOX1}-replacement cassettes contain a *Zeocin* resistance gene, in order to achieve a second selection basis, the model proteins were inserted into pPpKan_SmiI plasmid which carries a kanamycin resistance gene.

10.3.1 Design of expression vectors carrying *CalB* and *HRP* gene

pPpT4_SmiI_CalB and pPpT4_SmiI_HRPC1A plasmids were isolated from *E. coli* using GeneJET™ Plasmid Miniprep Kit from Fermentas. *CalB* and *HRP* genes together with α-factor signal sequence for secretion, were cut from the plasmids using restriction enzymes *EcoR*I and *Not*I. *EcoR*I cuts both plasmids one time, at the α-factor 5'-terminus. *Not*I cuts also one time, at *CalB* and *HRP* 3'-terminus, respectively. In addition, an empty pPpKan plasmid was cut with the same restriction enzymes, *EcoR* I and *Not* I. The DNA sequence for the α-factor signal sequence for secretion together with *CalB* was a 1235 bp long fragment, and the fragment containing α-factor together with HRP was 1206 bp long. The empty pPpKan plasmid cut with EcoR I and Not I was 3964 bp long. After ligation of the fragments, six possibly positive clones carrying CalB and two carrying HRP were cut for size control with Fast Digest restriction enzymes EcoRI and BamHI. Subsequently, two plasmids of correct size were sent for sequencing. Following primers were used for sequencing of both, pPpKan_SmiI_CalB and pPpKan_SmiI_HRPC1A p-08-849 and p-09-194 (Figure 36 and 37). For both constructs a positive clone was found.

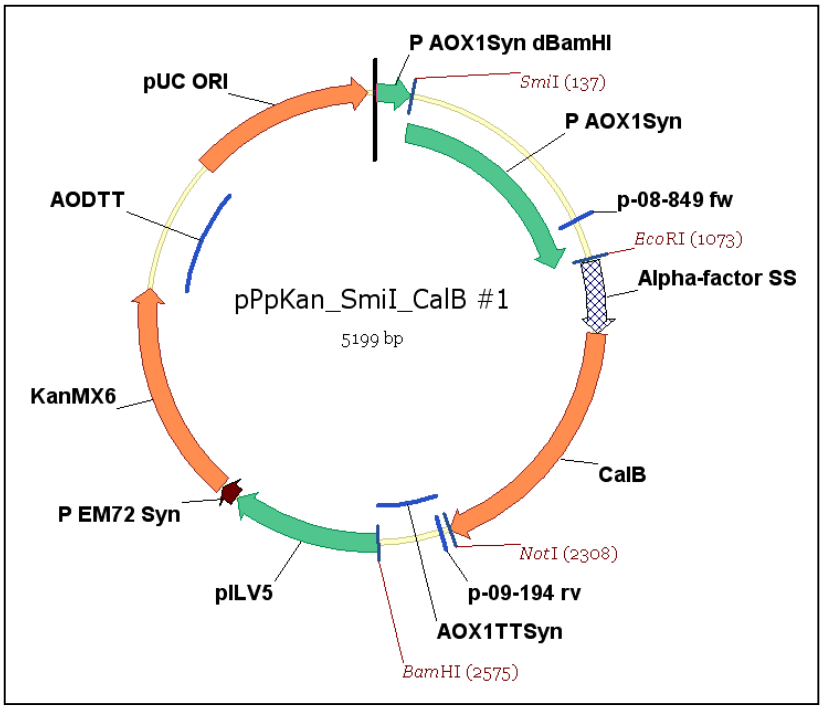

Figure 36: pPpKan_SmiI_CalB #1 is derived from pPpKan_SmiI by insertion of α-factor (alpha-factor secretion signal sequence) and *CalB* gene, through restriction with EcoR I and Not I and ligation using T4DNA ligase.

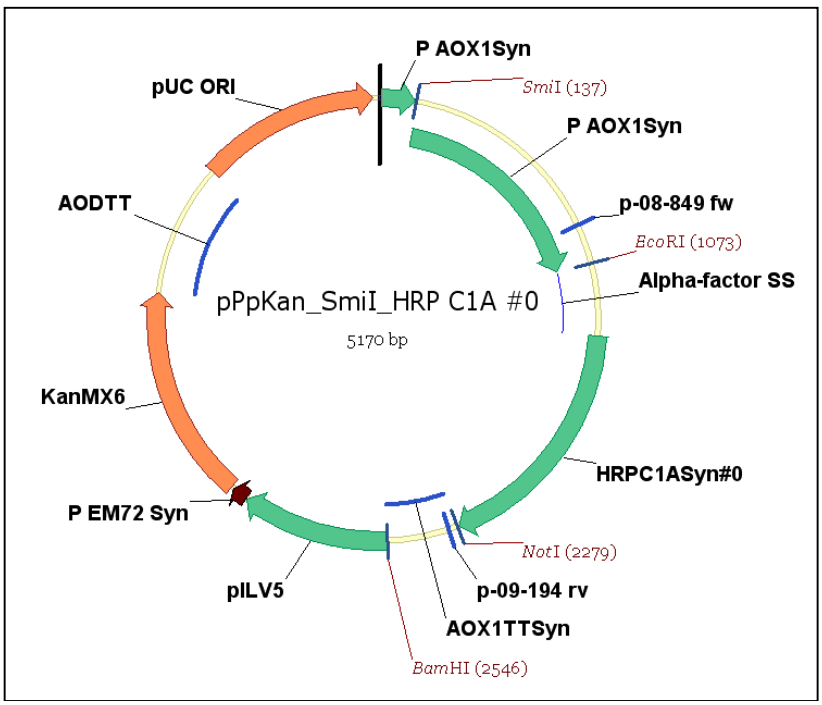

Figure 37: pPpKan_SmiI_HRPC1A #0 is derived from pPpKan_SmiI by insertion of α-factor (alpha secretion signal sequence) and *HRP* gene, through restriction with EcoR I and Not I and ligation using T4DNA ligase.

10.3.2 Transformation of *CalB* and *HRP* expression vectors to *P. pastoris* ΔKU70

Plasmids pPpKan_SmiI_CalB and pPpKan_SmiI_HRPC1A were linearised with *Smi*I over night. After that, *P. pastoris* ΔKU70 electro-competent cells were transformed with 1-2µg by electroporation, and plated on YPD-geneticin plates. 6 DWPs were inoculated with transformants. *P. pastoris* ΔKU70 strain was used as negative control on each plate. As control strains for CalB screening *P. pastoris* Mut^{s} pPpT4_SmiI_CalB clones A9 (low activity level) and E9 (high activity level) were used. For the HRP screening *P. pastoris* Mut^{s} pPpT4_SmiI_C1AHRP clones H12 (low activity level) and H5 (high activity level) were used.

10.3.3 Screening results of *P. pastoris* ΔKU70 *CalB* and *HRP* reporter strains

After cultivation, all DWPs were screened for positive clones using CalB and HRP photometric activity assays. Thereby clones from both the lowest and the highest activity level were chosen and rescreened. This was done in order to prove if the overexpression of the desired *PDI* has the same effect on the expression level of the model protein in the single copy and the multi copy clones.

For the screening for clones with CalB activity, from 87 wells inoculated with single colonies in DWP_A, DWP_B, DWP_1, and DWP_3, enzyme activity was measured for 72, 57, 46, and 86 clones, respectively. For the DWP_4 and DWP_8, 41 of inoculated 59 wells, and 35 of inoculated 43 wells contained clones secreting CalB. The final values were calculated by reducing each value with the average value of the negative control of each DWP. (Landscapes in Appendix 1) Seven clones showing highest activity were chosen and rescreened (Figure 38). Also five randomly chosen clones from the lowest activity level were rescreened (Figure 39). The aim of the rescreen was to define the most stable clones with high and low activity. Thereby, the rescreen was performed with eight biological replicates for each clone, inoculated in vertical lines of the DWPs. This inoculation order seemed to be necessary in order to avoid measuring errors which could occur due to different oxygen distribution. The cultivation media and protocol were the same as that used for the screening.

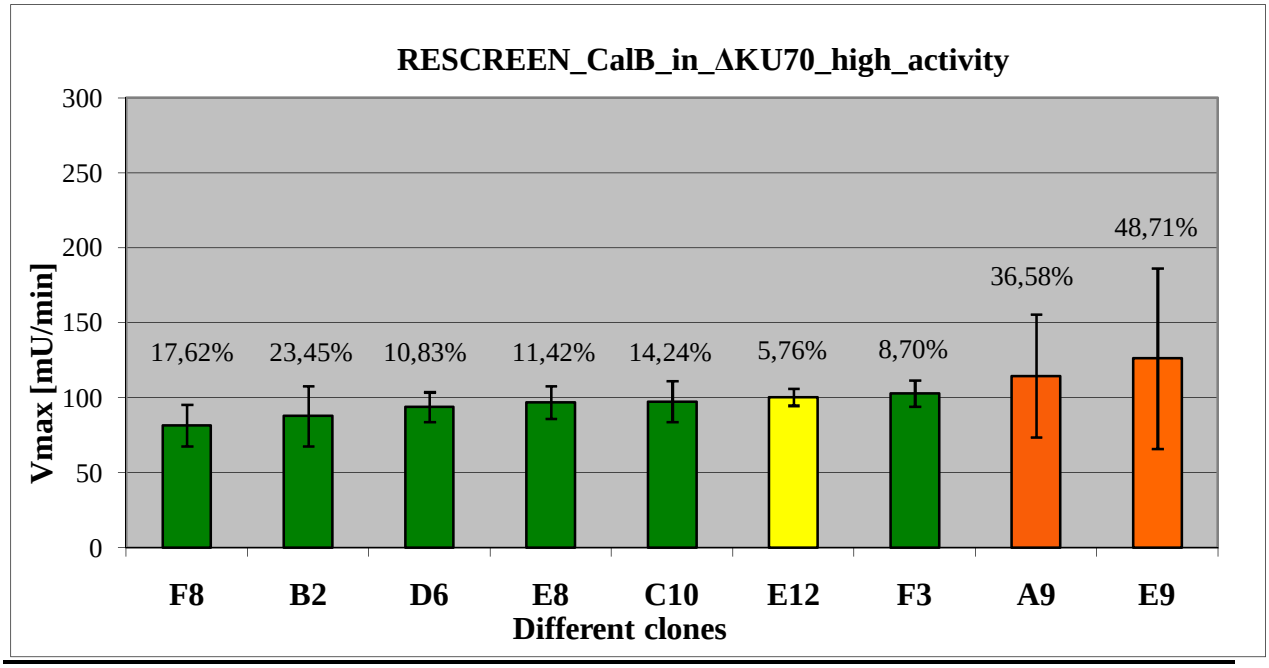

Figure 38: Rescreen of seven clones showing highest activity (F8, B2, D6, E8, E12 and F3). The orange colored bars present the clones used as positive control: *Pp*Mut^{s}_CalB_E9 as high activity and *Pp*Mut^{s}_CalB_A9 as low activity clone. The yellow colored bar presents the most stable clone showing high activity (*Pp*ΔKU70_CalB_E12), which were chosen to be used for the further experiments.

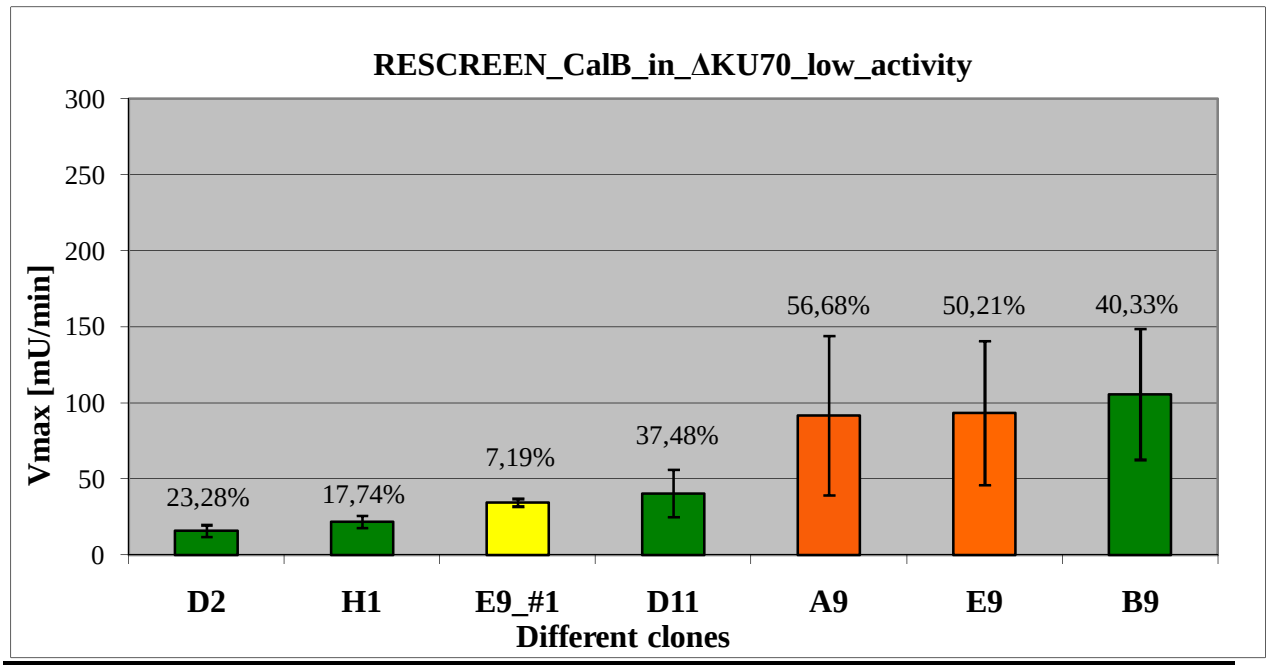

Figure 39: Rescreen of five clones showing low activity (D2, H1, E9_#1, D11 and B9). The orange colored bars present the clones used as positive control: *Pp*Mut^{s}_CalB_E9 as high activity and *Pp*Mut^{s}_CalB_A9 as low activity clone. The yellow colored bar presents the most stable clone showing low activity (*Pp*ΔKU70_CalB_E9_#1), which were chosen to be used for the further experiment.

For the screening for clones with HRP-activity, from 67 wells inoculated with single colonies in DWP_D, DWP_2, and DWP_5, enzyme activity was measured by 51, 25, and 27 clones, respectively. For the DWP_C and DWP_6, 78 of inoculated 87 wells, and 51 of inoculated 73 wells contained clones secreting HRP. The final values were calculated by reducing each value with the average value of the negative control of each DWP. (Landscapes in Appendix 2) Six clones were chosen for rescreening, three with the highest activity and three randomly chosen low activity clones (Figure 40). Again, the aim of the rescreen was to define the most stable clones with high and low activity. The rescreen was done in the same way as described for rescreen of clones with CalB activity.

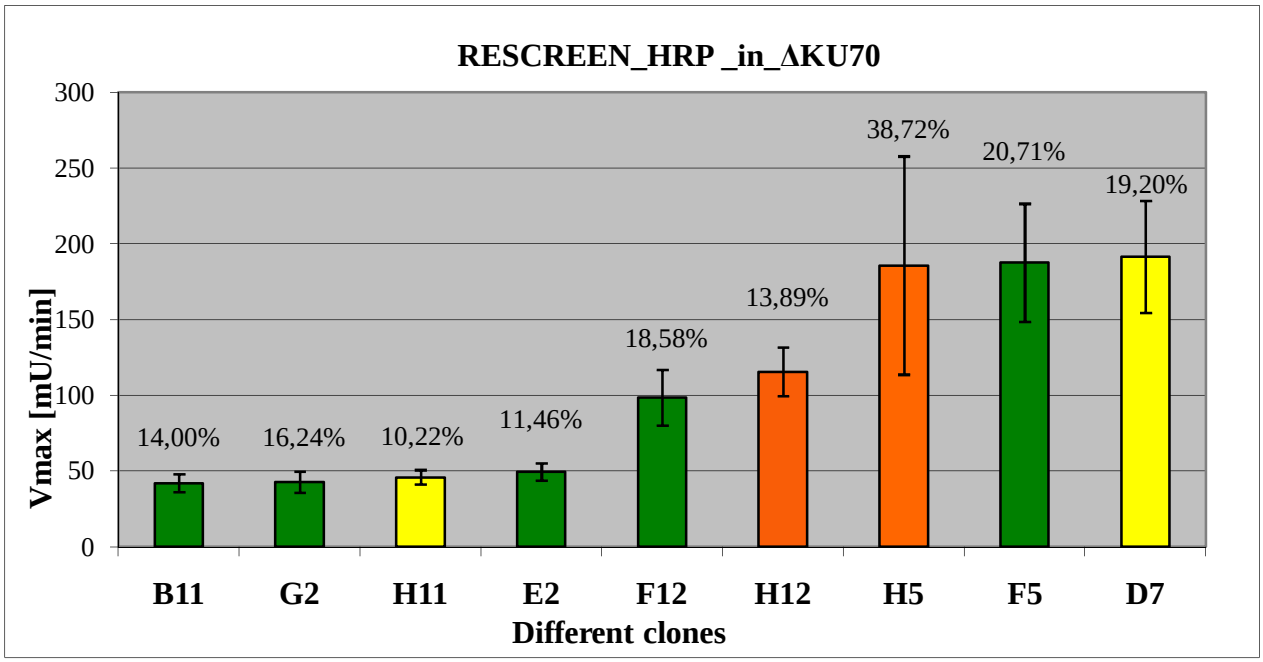

Figure 40: Landscape of photometric HRP rescreen of the most stable single and multi copy clones. The orange colored bars present the clones used as positive control: *Pp*Mut^{s}_HRP_H5 as high activity and *Pp*Mut^{s}_HRP_H12 as low activity clone. The yellow colored bars present the most stable clones showing low (*Pp*ΔKU70_HRP_H11) and high activity (*Pp*ΔKU70_HRP_D7), which were chosen to be used for the further experiments.

Cloning of *CalB* and *HRP* expression vectors to *P. pastoris* ΔKU70 gave expected results. The *Pp*ΔKU70 strain grows slower than the wild type (CBS7435). For that reason 48h after the transformation the *Pp*ΔKU70 colonies were still too small to be picked. On the other hand, approximately 72h after the transformation it was difficult to distinguish between the positive clones and the background. The background consisted of the so called satellite colonies, which are not positive but are able to grow in the certain radius of each positive clone, where the antibiotic in the medium got degraded. For that reason 48h after the transformation the coming colonies were labeled on the outside of the plate bottom. Thus the positive colonies were easily separated from their satellite colonies. After the first rescreen, standard deviations of more than 20% were seen for all rescreened clones, which implied a possible cultivation or measurement mistake (the results of the first rescreen are not shown in the text). For that reason, another rescreen experiment was done. This time, the eight biological replicates of each colony were inoculated in vertical lines of the DWPs. This sample order

ensured equal aeration of all biological replicates and identical activity measurement using an 8-channel pipette. After these optimizations, activity standard deviations lower than 20% were detected, which is in general acceptable percentage for micro scale cultivation of *P. pastoris* strains. Three of four clones used as positive control (*Pp*Mut^s_CalB_A9, *Pp*Mut^s_CalB_E9 and *Pp*Mut^s_HRP_H5) showed standard deviation higher than 20%. However, this fact did not influence the screening and rescreening experiments, since the positive control strains are *Pp*Mut^s and they could not be directly compared with *Pp*ΔKU70 strains. They were used in this thesis for approximately comparison of the single copy and multi copy expression level of the selected clones, and also in order to confirm that the assays work properly. After the rescreen, four clones shoving best reproducibility were chosen for further experimental work. *Pp*ΔKU70_CalB_E12 clone was chosen as high activity CalB clone with standard deviation of only 5,76%. This clone is very possible a single copy clone, as it is known that single copy CalB strains show highest enzyme activity[69]. However, without copy number determination this is only speculation. From the lowest activity level on the landscape, clone *Pp*ΔKU70_CalB_E9_#1 with standard deviation of 7,19% was chosen. For HRP, since the enzyme activity to a certain extent correlates to the gene copy number, the chosen *Pp*ΔKU70_HRP_H11 and *Pp*ΔKU70_HRP_D7 are assumed to be single and multi copy clones, showing 10,22% and 19,20% standard deviation of their activities.

10.4 Conclusion

Since the construction of the *PpPDI*-specific P_{AOX1}-replacement cassettes was more time-consuming than it was though at the beginning, the transformation of the cassettes to the *Pp*ΔKU70 reporter strains (*Pp*ΔKU70_CalB_E12, *Pp*ΔKU70_CalB_E9_#1, *Pp*ΔKU70_HRP_H11 and *Pp*ΔKU70_HRP_D7) will be a part of another project. Actually, after cassette integration in a proper *PpPDI* locus, and AOX1 promoter insertion, it will be possible to analyze the influence of the *PpPDI* overexpression on the expressed heterologous proteins (CalB and HRP) in the four reporter strains. As it has been shown in many case-studies, co-expression of *PDI* can have positive, negative, and no impact on heterologous protein expression. One explanation for that could be the different substrate specificity among the PDIs. It would be also interesting to know if the impact of *PpPDI* co-expression on heterologous protein expression varies depending on the gene copy number of the reporter strains. For these reasons, the reporter strains designed within this thesis were chosen from the lowest and highest activity level, and carry different reporter genes (CalB and HRP). Further experiments should direct us to the final aim of this project, which was the construction of *P. pastoris* patent free platform strains for production of higher yield of correctly folded heterologous proteins.

11. Design of 1,6-manosyltrasferase knockout strains of Pichia pastoris

11.1 Project aim and strategy

11.2 Sequence analysis of PpOCH1 and PpHOC1 genes

11.3 Design of PpOCH1 and PpHOC1 knockout cassettes

11.4 Design of expression vector carrying gene coding for model protein HNL

11.4.1 Design of Pichia pastoris reporter strain expressing HNL

11.4.2 Screening results of HNL reporter strain

11.5 Transformation of PpOCH1 and PpHOC1 knockout cassettes to P. pastoris HNL reporter strain

11.5.1 Results and discussion

11.5.2 Design and screening of P. pastoris ΔKU70 HNL reporter strain

11.6 Conclusion and further perspectives

11.1 Project aim and strategy

The aim of this project was to design *P. pastoris OCH1 and P. pastoris HOC1* knockout strains. *OCH1* and *HOC1* are genes coding for α-1,6 mannosyltransferases in *P. pastoris.* As model protein for analyzing the protein glycosylation in the knockout strains, *P. pastoris* CBS7435 and *P. pastoris* ΔKU70 strains carrying the *PaHNL5 (Prunus amygdalus* hydroxynitrile lyase 5) gene were first designed. The knockout cassettes contain *OCH1* and *HOC1* locus homologous flanking ends on the 3` and 5` end, so the gene deletion should occur through double site homologous recombination.

11.2 Sequence analysis of *PpOCH1* and *PpHOC1* genes

According to GenDB and the NCBI database, two genes code for α-1,6 mannosyltransferase in *P. pastoris* CBS7435 (wild type) strain. The first one, called *OCH1*, is located on chromosome are (PP7435_Chr1-0258) and codes for a 404 amino acids long α-1,6 mannosyltransferase. The second one is called *HOC1* (PP7435_Chr3-0582). It is located on chromosome three and codes for a 286 amino acids long α-1,6 mannosyltransferase. The %GC content vary between 41,1% for *PpHOC1* and 43,7% for *PpOCH1*. *PpOCH1* contains no introns, in contrast to *PpHOC1* which has one 130bp long intron. Both, OCH1 and HOC1 belong to the enzyme class EC 2.4.1.

According to the analysis and comments made within the recent *P. pastoris* CBS7435 genome sequencing project, inside the cell the gene product of *OCH1* is anchored to the endoplasmic or Golgi membrane and the product of *HOC1* is a Golgi membrane protein.*

Providing DNA multiple sequence alignment of *PpOCH1* and *PpHOC1* coding domains, sequence similarity of 57% has been observed (Figure 41). The multiple sequence alignment of the 1000bp long upstream and downstream regions, showed a sequence similarity of 49% and 50%, respectively (Figure 41; Appendix 5). The multiple sequence alignment of the amino acid sequences of *Pp*OCH1 and *Pp*HOC1 showed a low sequence similarity of 25% (Figure 41; Appendix 6).

Alignment 1: Multiple sequence alignment of the coding domain sequences of *PpOCH1* and *PpHOC1*.

```
CLUSTAL 2.1 multiple sequence alignment


CCA36419.1      ATG-GCGAAGGCAGATG-GCAGTTTGCTCTACTA-TAATCCTC--ACAATCCA--CCCAG 53
CCA39540.1      ATGTACGAA---AGATGTGAAGTTT-CTTTGTTGGCGATTTTCTTGCAGCGCGATCGCGG 56
                ***  ****   ***** * ***** ** *  *    **  **   **   *   * * *
```

* Information source: https://gendb.cebitec.uni-bielefeld.de/

```
CCA36419.1      AAGGTATTACTTC--TACATGGCTATATTCGCCGTTTCTGTCATTTGCGTTTTGTACGGA 111
CCA39540.1      TCGGTATGATTTCCCTTCA-GACCAAAA--ACAGATATAGATATT--CAATTGAAACTCG 111
                  ***** * ***  * ** * * * *    * * *   *  ***  *  **   **

CCA36419.1      CCCTCACAACAATTATCATCTCCAAAAATAGACTATGATCCATTGACGCTCCGATCACTT 171
CCA39540.1      CCCTTCCAACA--------CCCTGAA-------CACCACCCCCTAATGCAACAA-CAGCT 155
                ****  *****        * *  **        *  * **  * * **  * * **  *

CCA36419.1      GATTTGAAGACTTTGGAAGCTCCTTCACAGTTGAGTCCAGGCACCGTAGAAGATAATCTT 231
CCA39540.1      GTTTAGAAAAGT----AATTTGGCTCACGGTTGGACTCA-------TA-ACGGTGATATT 203
                * ** *** * *    **  *   **** ****    **       ** * * * ** **

CCA36419.1      CGAAGACAATTGGAGTTTCATTTTCCTTA-CCGCAGTTACGAACCTTTTCCCCAACATAT 290
CCA39540.1      -GGTAATAATAAAGATTTCATCCAGCAAGTCCACCGCCACAGACCTT---CAGAAAGTAT 259
                 *   * ***     ******    *    ** * *  **  *****   *  **  ***

CCA36419.1      TTGGCAAACGTGGAAAGTTTCTCCC----TCTGAT-AGTTCCTTT--CCGAAAAACTTCA 343
CCA39540.1      T-GAAGAATGCCAATATTTTACCGCAGGACGTGATCAATTACAATAGCAGAAAAGTAACG 318
                * *   ** *   * * ***  * *      **** * ** *  *  * *****    *

CCA36419.1      AA--GACTTAGGTGAAAGTTGGCTGCAAAGGTCCCCAAATTATGATCATTTTGTGATACC 401
CCA39540.1      GATGAACTCGCTTCAAAGCTGGACG---AGATT--CAAA--AGAAGTATCTCTCGAAGCA 371
                 *   ***    * **** ***  *   ** *   ****  *  *  ** *   **  *

CCA36419.1      CGATGATG----CAGCATG---GGAACTTATTCACCATGAATAC--GAACGTGTACCAGA 452
CCA39540.1      AGATGATAGAATTAGCAAGCTCGAAGCTGAACGGGCAGATCTACTGGAACAGGT--TAGA 429
                 ******      **** *   * * ** *     **    ***  ****  **   ***

CCA36419.1      AGTCTTGGAAGCTTTCCACCTGCTACCAGAGCCCATTCTAAAGGCCGATTTTTTCAGGTA 512
CCA39540.1      TTTCTAAGGAAC---CCCCCTGC----AGGATCAAGTTTAAGAGAAAAATTGGCT---TA 479
                  ***  * * *   ** *****    **   * * * ***  *   * **       **

CCA36419.1      TTTGATTCTTTTTGCCCGTGGAGGACTGTATGCTGACATGGACACTATGTTATTAAAACC 572
CCA39540.1      TCTGTTTCCTTATAAT----GAAAACGGCAAATTC-CCTGCTTA-TATATGGC--AAAC- 530
                * ** *** ** *       **  ** * *   *  * **   * *** *     ****

CCA36419.1      AATAGAATCGTGGCTGACTTTCAATGAAACTATTGGT--GGAGTAAAAAACAATGCTGG- 629
CCA39540.1      -ATGGAAGTATGGCT-----TGAATG--ACGATCGGTTTGGAGAAAAGTTCAAAGAAGGC 582
                 ** ***   *****     * ****  ** ** ***  **** ***   *** *  **

CCA36419.1      GTTGGTCATTGGTATT-GAGGCTGATCCTGATAGACCTGATTGGCACGACTGGTATGCTA 688
CCA39540.1      GAAACTCAGTGGGCTTCGAAG--AATCCTGGT----TTCGTT--CATGAGTTGTTT---- 630
                *    *** ***  ** ** *   ****** *     *  **  ** ** * ** *

CCA36419.1      GAAGGATACAATTTTGCCAATGGGCAATTCAGTCCAAACGAGGACACCCAGCACTGCGTG 748
CCA39540.1      -AACGATG--ATACTTCCGGTGTG---TTTA-TTCA----------CCATCTGTATATC 672
                 ** ***   **  * **  ** *   ** * * **          *** *  *   *

CCA36419.1      AACTGATTGTAAGAGTTGTCAGCACGACTTTACGGAAAGAGAAAAGCGGTTACTTGAACA 808
CCA39540.1      AAT--GTTCCAGAAGTGATCAAAGCA----TACG----------AGCTGCTTCCC-AACA 715
                **    **  *  ***  ***   *     ****          *** * * *   ****

CCA36419.1      TGGTGGAAGGAAAGGATCGTGGAAGTGATGTGATGGACTGGACGGGTCCAGGAATATTTA 868
CCA39540.1      TA-------------ATCTTGAAA--------ATGGACT-----------------TCTT 737
                *              *** ** **        *******                 * *

CCA36419.1      CAGACACTCTATTTGATTATATGACTAATGTCAATACAACAGGCCACTCAGGCCAAGGAA 928
CCA39540.1      CAGA-----TATTTGGTTTTAT--------------------------ACGCCAAG--- 762
                ****     ****** ** ***                          * ******

CCA36419.1      TTGGAGCTGGCTCAGCGTATTACAATGCCTTATCGTTGGAAGAACGTGATGCCCTCTCTG 988
CCA39540.1      --GGAGGTGTCTATGCAGA-------------CGTTG-----ATACTATGCCTCTTCAG 801
                  **** ** **  **  *              *****     *    *****   ** *

CCA36419.1      CCCGC-CCGAACGGAGAGATGTTAAAAGAGAAAGTCCCAGGTAAATATGCACAGCAGGTT 1047
CCA39540.1      CCTGTACCAAACTG---GATTCCTGAAAATGTCTCCCCAAA-AAGCAT------CGGGAT 851
                ** *  ** *** *   ***     ** *      ****   **  **      * ** *
```

```
CCA36419.1      GTTTTATGGGAACAATTTACCAACCTGCGCTCCC---CCAAATTAATCGACGATATTCTT 1104
CCA39540.1      GATC-ATTGGAA-----TACAAAACGATGCTAACAATCCAGATTGGAAAAAGAT-TACGT 904
                * *  ** ****     *** ** *   ***  *   *** ***     * *** * * *

CCA36419.1      ATTCTTCCGATCACCAGCTTCAGTCCAGGGATTGGCCACAGTGGAGCTGGAGATTTGAAC 1164
CCA39540.1      A--CATC--GTCTACAATTT---TCCA---ATTGGT-GTATTCAAGC--GAAGCCTGG-- 949
                *  * **   **  **  **   ****   *****    * *  ***  **    **

CCA36419.1      CATCACCTTGCATATATTAGGCATACATTTGAAGGAAGTTGGAAGGACTAA 1215
CCA39540.1      --TCACC---CA-AT-----------CTTGA------------------- 963
                  *****   ** **            ****
```

Alignment 2: Multiple amino acid sequence alignment of *Pp*OCH1 and *Pp*HOC1.

```
CLUSTAL 2.1 multiple sequence alignment


CCA36419.1      -MAKADGSLLYYNPHNPPRRYYFYMAIFAVSVICVLYGPSQQ---------------LS 43
CCA39540.1      MYERCEVSLLAIFLQRDRGRKVIWLTVGLITVILVIIKISSSKSTATDLQKVLKNANILP 60
                   :.: ***    :.   *  :::::  ::** *:   *..                *.

CCA36419.1      SPKIDYDP-------------------------LTLRSLDLKTLEAPSQLSPGTVEDNL 77
CCA39540.1      QDVINYNSRKVTDELASKLDEIQKKYLSKQDDRISKLEAERADLLEQVRFLRNPPAGSSL 120
                .  *:*:.                            .*.:   . **    *   .. ..*

CCA36419.1      RRQLEFHFPYRSYEPFPQHIWQTWKVSPSDSSFPKNFKDLGESWLQRSPNYDHFVIPDDA 137
CCA39540.1      REKLAYLFPYNENGKFPAYIWQTWKYGLNDDRFGEKFKEGETQWASKNPGFVHELFNDDT 180
                *.:* : ***..   ** :****** . .*. * ::**:   .* .:.*.: * :: **:

CCA36419.1      AWELIHHEYERVPEVLEAFHLLPEPILKADFFRYLILFARGGLYADMDTMLLKPIESWLT 197
CCA39540.1      SGVFIHHLYINVPEVIKAYELLPNIILKMDFFRYLVLYAKGGVYADVDTMPLQPVPNWIP 240
                :  :*** * .****::*:.***: *** ******:*:*:**:***:*** *:*: .*:.

CCA36419.1      FNETIGGVKNNAGLVIGIEADPDRPDWHDWYARRIQFCQWAIQSKRGHPALRELIVRVVS 257
CCA39540.1      ENVSP----KSIGMIIGIQNDANNPDWK-----KITYIVYNFP------------IGVFK 279
                 * :     :. *::***: *.:.***:     :* :  : :             : *..

CCA36419.1      TTLRKEKSGYLNMVEGKDRGSDVMDWTGPGIFTDTLFDYMTNVNTTGHSGQGIGAGSAYY 317
CCA39540.1      RSLVTQS---------------------------------------------------- 286
                 :* .:.

CCA36419.1      NALSLEERDALSARPNGEMLKEKVPGKYAQQVVLWEQFTNLRSPKLIDDILILPITSFSP 377
CCA39540.1      ------------------------------------------------------------


CCA36419.1      GIGHSGAGDLNHHLAYIRHTFEGSWKD 404
CCA39540.1      ---------------------------
```

Figure 41: Summary of DNA and amino acids multiple sequence alignment of *PpOCH1* (PP7435_Chr1-0258) and *PpHOC1* (PP7435_Chr3-0582).

a) Multiple sequence alignment of the coding domain sequences of *PpOCH1* and *PpHOC1*.

SeqA	Name	Length	SeqB	Name	Length	Score
1	CCA36419.1	1215	2	CCA39540.1	963	57.0

b) Multiple sequence alignment of the 1000bp upstream region of *PpOCH1* and *PpHOC1*.

SeqA	Name	Length	SeqB	Name	Length	Score
1	CCA36419.1	1000	2	CCA39540.1	1000	49.0

c) Multiple sequence alignment of the 1000bp downstream region of *PpOCH1* and *PpHOC1*.

SeqA	Name	Length	SeqB	Name	Length	Score
1	CCA36419.1	1000	2	CCA39540.1	1000	48.0

d) Multiple sequence alignment of the amino acid sequences of *PpOCH1* and *PpHOC1*.

SeqA	Name	Length	SeqB	Name	Length	Score
1	CCA36419.1	404	2	CCA39540.1	286	25.0

The coding domain sequences (CDS) of the genes *OCH1* (PP7435_Chr1-0258) and *HOC1* (PP7435_Chr3-0582) together with 1000bp upstream and downstream regions were amplified from genomic DNA (gDNA) of *P. pastoris* CBS7435 (wild type strain) with the primer pairs p-10-737(fw) and p-10-736(rv), and p-10-735(fw) and p-10-734(rv), respectively. The 2258bp and 1994bp long sequences were purified by preparative 1%-agarose gel electrophoresis, and Wizard® SV Gel and PCR Clean-Up System from Promega. After that they were ligated with a pJet1.2/blunt ends cloning vector using T4-DNA ligase. *E. coli* TOP10F` electro-competent cells were transformed with the plasmid by electroporation. Aliquots of 20µl and 120µl of each transformation were plated on LB-ampicillin plates, and incubated for 24h at 37°C. After a incubation period, four positive clones were chosen, streaked out on LB-ampicillin plates, and incubated for another 24h at 37°C. Plasmids were isolated and sent for sequencing. All samples were sequenced using Sanger Technology by the LGC Genomics Company (Berlin, Germany).

The sequencing results were compared with the sequences contained in NCBI. Thereby following mismatches were detected:

- for the gene PP7435_Chr1-0258 (CCA36419.1), 153bp downstream to the gene start codon, 1G is contained instead of 1A as recorded in NCBI. However, this point mutation is silent one and the amino acid proline is not changed (both codes CC**A** and CC**G** code for Pro). Both, *OCH1* and *HOC1* DNA sequences from CBS7435 and GS115 contained in NCBI show 100% identity.

-for the gene PP7435_Chr1-0258 (CCA36419.1), 244bp downstream to the gene stop codon, the sequence contains 7C instead of 8C as recorded in GenDB;

11.3 Design of *PpOCH1* and *PpHOC1* knockout cassettes

Plasmids of two clones showing correct sequences were used in the further work for the construction of the knockout cassettes. Both, *PpOCH1* and *PpHOC1* knockout cassettes contain the following three parts: homologous upstream region of *OCH1* or *HOC1* gene, the gene for ampicillin resistance together with the regulatory elements (*Pp* pILV5 promoter and AODTT terminator), and homologous downstream region of *OCH1* or *HOC1* gene. All parts were amplified using PCR. For the *PpOCH1* and *PpHOC1* upstream regions the primer pairs: p-10-800(fw) and p-10-801(rv), and p-10-804(fw) and p-10-805(rv) were used, respectively. For the *PpOCH1* and *PpHOC1* downstream regions the primer pairs: p-10-802(fw) and p-10-803(rv), and p-10-806(fw) and p-10-807(rv) were used, respectively. For the amplification of the whole core of the cassette, primers containing junction sequences (Linker 1 and Linker 2) were used (Figure 42).

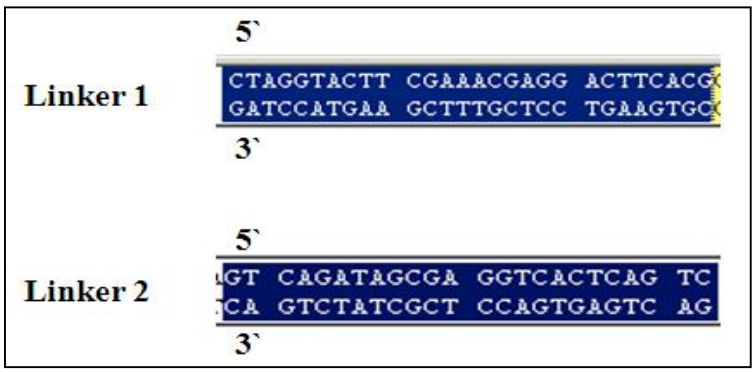

Figure 42: Linker 1 (28bp) and Linker 2 (24bp). Two junction sequences used as overlapping regions for the connection of the cassette core region with the homologous upstream and downstream regions of *PpOCH1* and *PpHOC1*.

All cassettes' parts were joined together by OE-PCR, and the cassettes were ligated with pJet1.2/blunt ends cloning vector. *E. coli* TOP F` cells were transformed with these cassettes. The homologous flanking ends were connected with the core of the cassette with OE-PCR, where the linker regions were used as overlapping sequences (Figure 43). The cells were plated on LB-ampicillin plates, as aliquots of 20µl and 120µl for each transformation. After an incubation period of approximately 24h at 37°C, positive clones were chosen, streaked out on LB-ampicillin plates, and incubated for additional 24h at 37°C. Plasmids of four clones of each transformation were isolated using the GeneJET™ Plasmid Miniprep Kit and sent for sequencing. Two clones carrying knockout cassettes for *PpOCH1* and *PpHOC1* which showed correct sequences, were used for the further *P. pastoris* transformation.

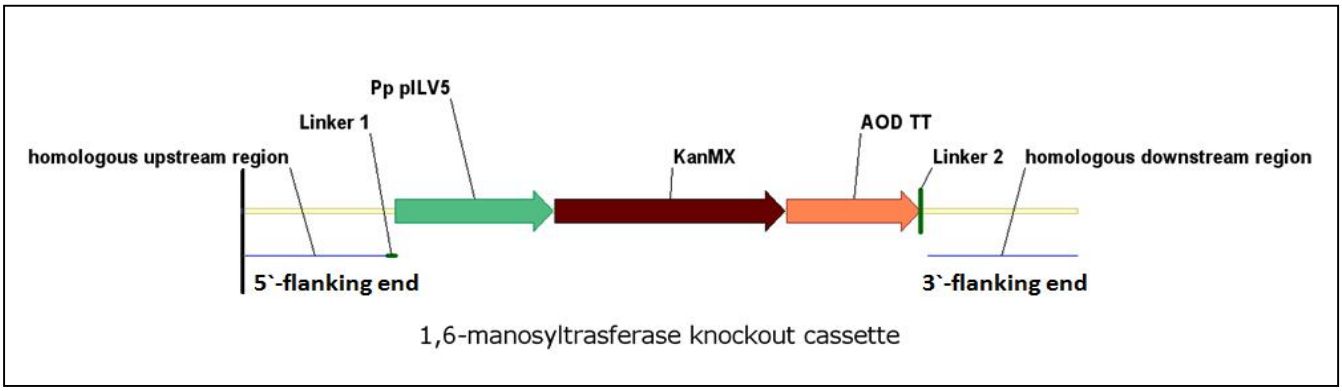

Figure 43: 1,6-manosyltransferase knockout cassette. Green colored is the *Pichia pastoris* ILV5 promoter, red colored is the synthetic gene for kanamycin resistance (KanMX), and orange colored is the AOD terminator. The 5`- and 3`-flanking ends are 500bp long, and they are homologous to the upstream and downstream regions of the CDS of *PpOCH1* and *PpHOC1*, respectively. Linker1 (28bp) and Linker2 (24bp) are short DNA junction sequences.

11.4 Design of expression vector carrying gene coding for model protein HNL

The plant isoenzyme 5 of hydroxynitrile lyase from *Prunus amygdalus* (*Pa*HNL5) contains 13 potential N-glycosylation motifs.[70] Previous studies have shown that heterologously expressed secreted *Pa*HNL5 in *P. pastoris* is highly glycosylated.[70] (Figure 44) [70] For that reason, *Pa*HNL5 was chosen to be a model protein within this project since the main goal was to prove the effect of the *OCH1* and *HOC1* deletion on the protein *N*-glycosylation in *P. pastoris*.

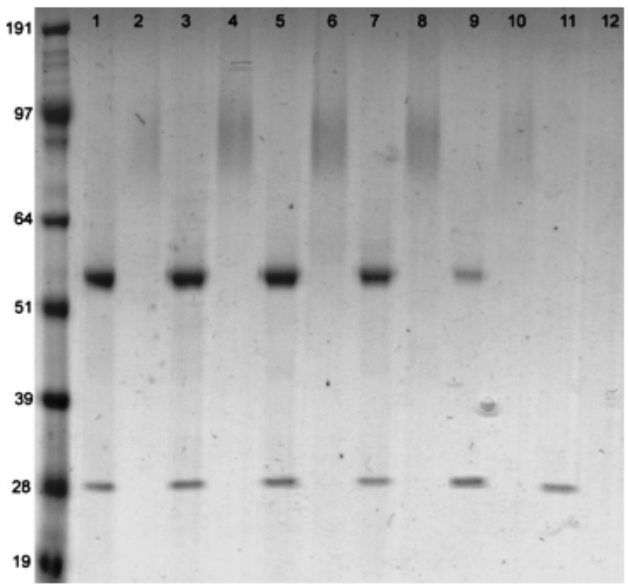

Figure 44:[70] **Protein gel of different samples of glycosylated and deglycosylated *Pa*HNL5 secreted in *P. pastoris*.** At the positions 2, 4, 6 and 8 are different samples of the native form of *Pa*HNL5 secreted in *P. pastoris*. At the positions 1, 3, 5 and 7 are different samples of *Pa*HNL5 secreted in *P. pastoris* and additionally EndoH-deglycosylated. The high glycosylated native form of the enzyme shows a smear, while its deglycosylated form shows a clear band in the range of approximately 60 kDa.[70]

Plasmid carrying *HNL* gene was isolated (pGAPZA_α_PaHNL5alpha) using GeneJET™ Plasmid Miniprep Kit. It was cut with BamHI in two fragments, the shorter one (1768bp), and the larger one (2991bp). The fragments were purified using preparative 1%-agarose gel electrophoresis, and Wizard® SV Gel and PCR Clean-Up System from Promega. Additionally, the shorter fragment was cut with the restriction enzyme NotI and the larger fragment with EcoRI. The pieces of 577bp and 1348bp respectively, were purified in the same way, and their sizes were controlled by 1%-agarose gel electrophoresis, using 1kb DNA ladder as standard. The concentrations were measured with NanoDrop spectrophotometer. In parallel pPpT4_SmiI empty plasmid was isolated and cut with the same restriction enzymes as the fragments, NotI and EcoRI. The restricted plasmid was purified, its size was controlled and its concentration was measured, in the same way as for all other fragments. After that the vector was ligated with both fragments using T4DNA-ligase, for 1h 30min at 22°C, and after ligation the samples were desalted for 60min against dH_2O at room temperature prior to electroporation. *E. coli* TOP10F` electro-competent cells were used for transformation and aliquots of 50µl and 150µl were plated on LB-zeocin plates. The plates were incubated for approximately 24h at 37°C, and after that three clones were streaked out for additional 24h at 37°C. From these three clones plasmids were isolated and a control cut with BamHI (Fast Digest) for 15min at 37°C was done. The size of the plasmid was controlled using 1%-agarose gel and the concentration was measured with NanoDrop spectrophotometer. Plasmids from all three clones were sent for sequencing.

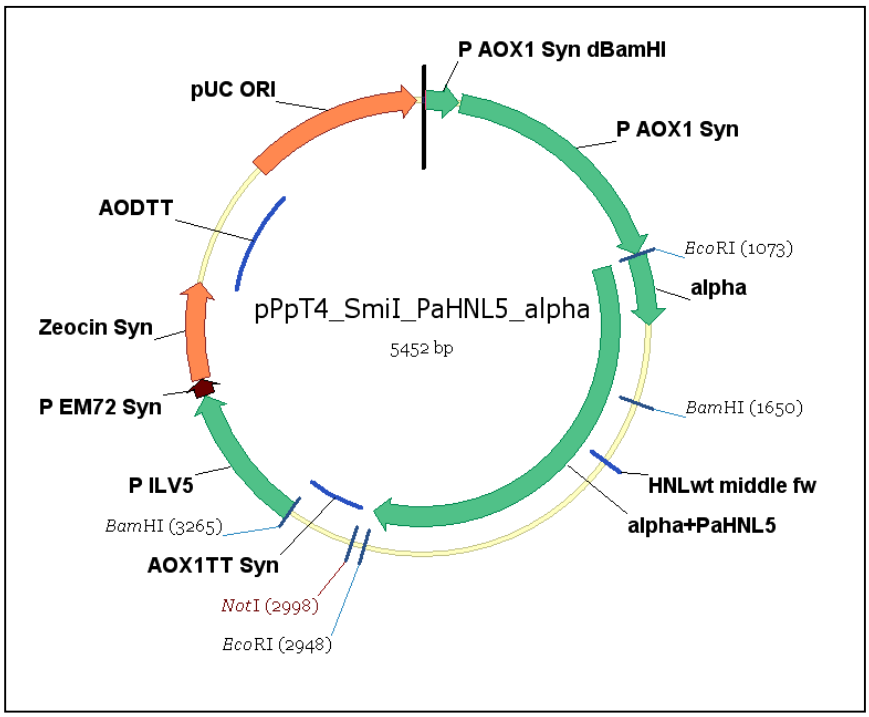

Figure 45: pPpT4_SmiI_PaHNL5_alpha plasmid. pUC_ORI (pUC origin of replication); PAOX1_Syn_dBamHI (synthetic *AOX*1 promoter upstream region); PAOX1_Syn (synthetic *AOX*1 promoter region); alpha factor (secretion signal sequence); PaHNL5 (gene coding for *Prunus amygdalus* hydroxynitrile lyase 5); AOX1TT_Syn (synthetic terminator); pILV5 (eukaryotic promoter); pEM72Syn (synthetic prokaryotic promoter); Zeocin_Syn (gene for zeocin resistance); AODTT (*AOD* terminator);

11.4.1 Design of a *Pichia pastoris* HNL reporter strain

The plasmid showing correct sequencing results was linearized, over night and at 30°C with SmiI. The linearized plasmid was purified using preparative 1%-agarose gel electrophoresis, and Wizard® SV Gel and PCR Clean-Up System from Promega and *P. pastoris* CBS7435 was transformed using 1-2µg. Aliquots of 150µl, 300µl and the rest of the cells were plated on YPD-zeocin plates and incubated for approximately 48h and 72h, respectively. After the incubation period, 4 deep-well plates (DWPs) were filled with 300µl sterile BMD1% medium per well and each well was inoculated with a single colony picked from the transformation plates. Thereby, three wells per DWP were inoculated with a single colony of the *P.p.* CBS7435 strain, used as negative control. In order to ensure positive control for approximate correlation of the expression level, three wells per plate were inoculated with single colonies of NHL5_V317α strains. This is a *P. pastoris* Mut^{s} strain carrying a single copy of a synthetic *HNL5* gene, variant cloned with a pPpT4_SmiI cloning vector. All DWPs were cultivated at 28°C, 110 rpm and 80% humidity, according to the *P. pastoris* long DWPs-cultivation protocol.

11.4.2 Screening results of *P. pastoris* CBS7435 *HNL* reporter strain

After finishing the micro scale cultivation, all DWPs were screened for positive clones using a photometric HNL activity assay.

Thereby from 90 wells inoculated with single colonies in DWP_1 (Figure 46), 61 in DWP_2 (Figure 47), 90 in DWP_3 (Figure 48), and 90 in DWP_4 (Figure 49), enzyme activity was measured for 90, 50, 77, and 82 clones, respectively. The final measure values were calculated by reducing each value with the average value for the negative control for each DWP.

Six clones showing highest activity were chosen and rescreened (Figure 50). The aim of the rescreen was to define the most stable clones with high activity. Thereby, the rescreen was performed with eight replicates for each clone, inoculated in vertical lines of the DWPs. This inoculation order seemed to be necessary in order to avoid measuring errors which could occur due to different oxygen distribution. The cultivation media and protocol were the same as that used for the screening.

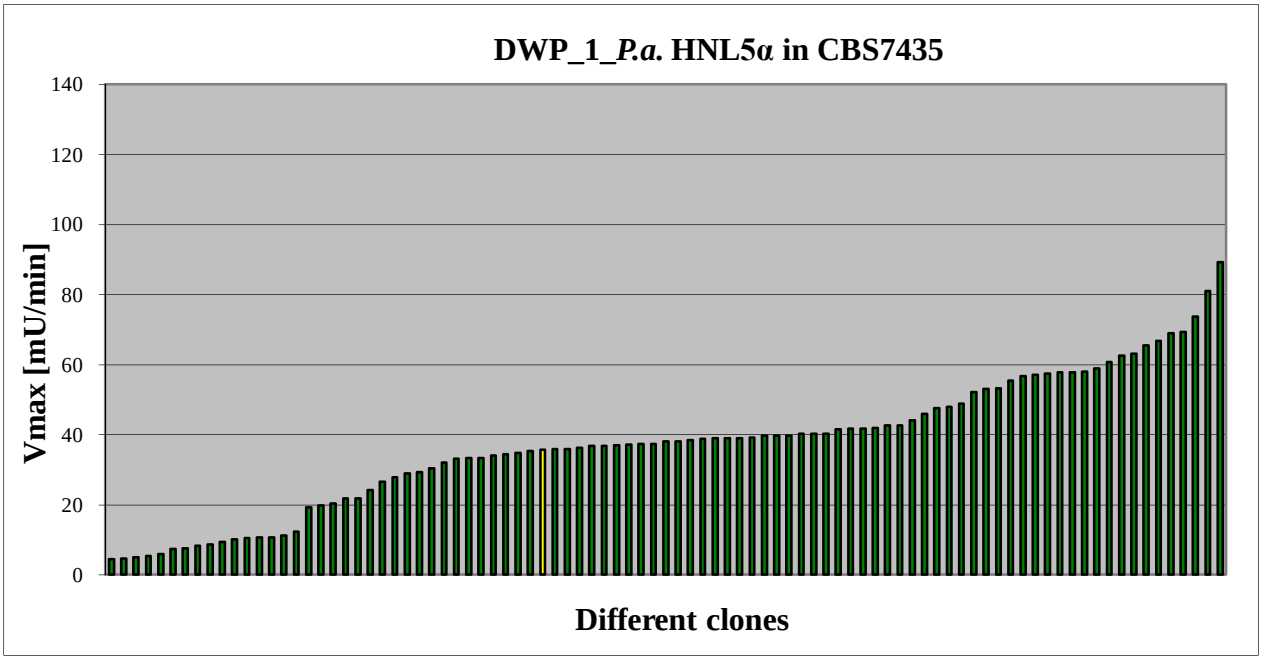

Figure 46: Landscape of photometric HNL activity screening of DWP_1 (*HNL5a* in *P.p.*CBS7435). The yellow colored bar presents the clone used as positive control (HNL5_V317α).

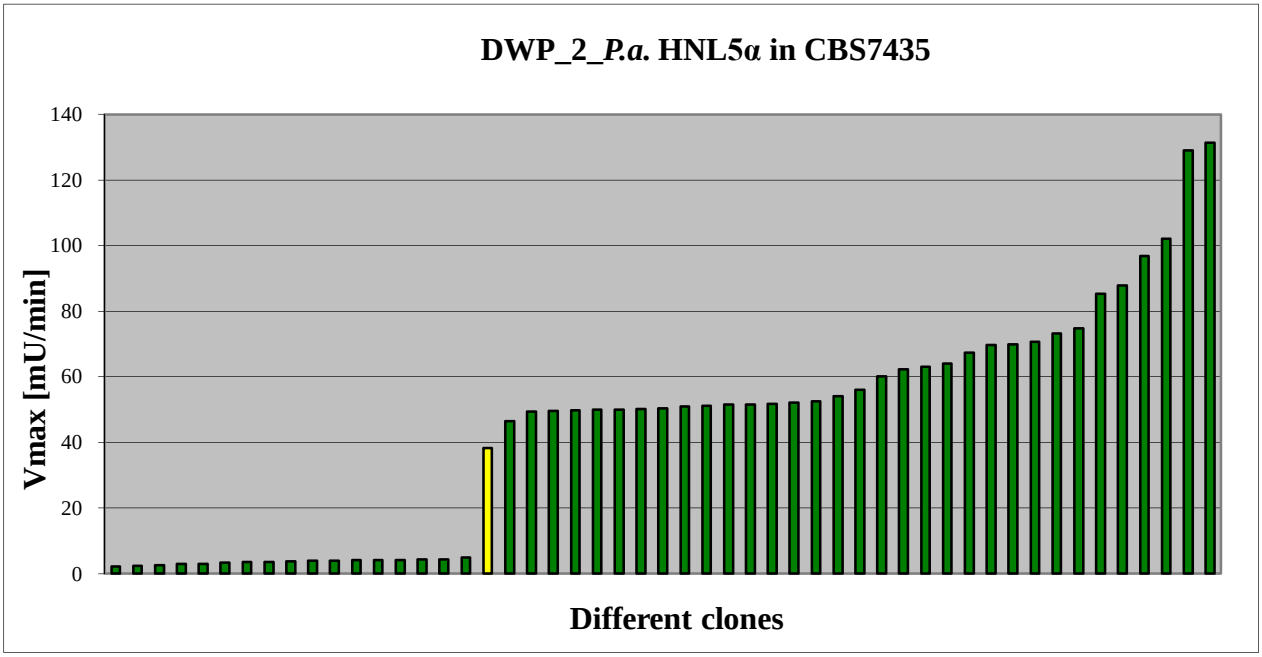

Figure 47: Landscape of photometric HNL activity screening of DWP_2 (*HNL5a* in *P.p.*CBS7435). The yellow colored bar presents the clone used as positive control (HNL5_V317α).

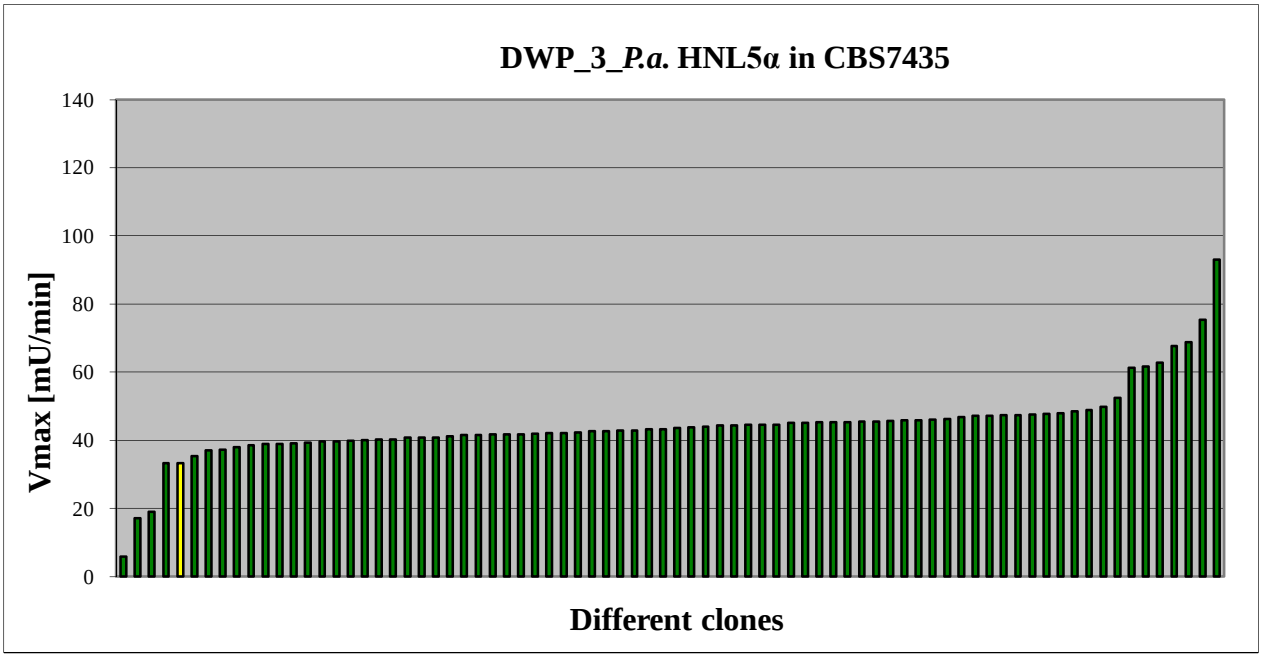

Figure 48: Landscape of photometric HNL activity screening of DWP_3 (*HNL5a* in *P.p.*CBS7435). The yellow colored bar presents the clone used as positive control (HNL5_V317α).

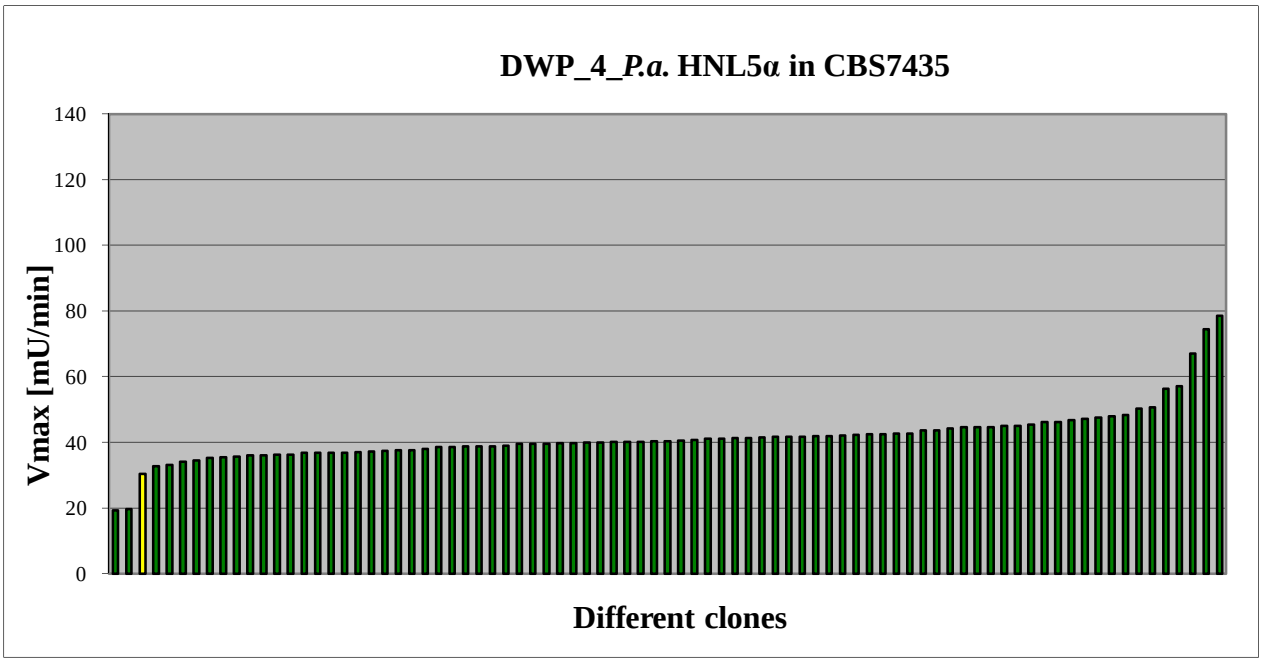

Figure 49: Landscape of photometric HNL activity screening of DWP_4 (*HNL5a* in *P.p.*CBS7435). The yellow colored bar presents the clone used as positive control (HNL5_V317α).

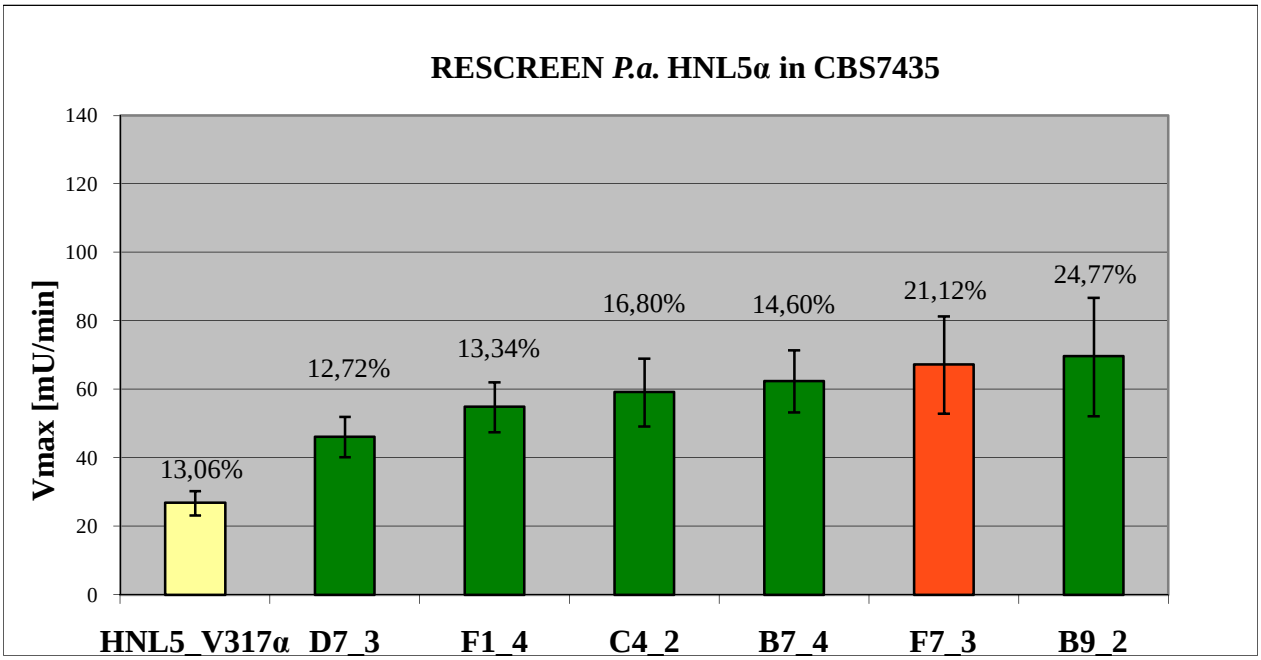

Figure 50: Landscape of photometric HNL rescreen of the most stable multi copy clones (*HNL5α* in *P.p.*CBS7435). The yellow colored bar presents the clone used as positive control (HNL5_V317α). The orange colored bar represents the clone chosen as reporter strain for the further work (F7_3).

11.5 Transformation of the *P. pastoris* HNL reporter strain with *PpOCH1* and *PpHOC1* knockout cassettes

The *PpOCH1* and *PpHOC1* knockout cassettes were amplified from the corresponding plasmids using PCR and following primer pairs: p-10-737 and p-10-736 and p-10-735 and p-10-734, respectively. The PCR products were purified using preparative 1%-agarose gel electrophoresis, and Wizard® SV Gel and PCR Clean-Up System from Promega. Their concentrations were measured using a NanoDrop spectrophotometer. Different amounts of the purified knockout cassettes (500ng to 1,5µg) were used for transformation of electro competent cells of *P. pastoris* CBS7435 carrying the *Pa*HNL5_alpha gene (clone F7_3). *Pa*HNL5_alpha is a HNL5 gene where the wild signal sequence was replaced by the alpha factor signal sequence. After transformation, aliquots of 50µl, 150µl and the rest of the cells (cells concentrated by brief centrifugation) were plated on YPD-zeocin/geneticin plates. Also aliquots of 330µl concentrated by brief centrifugation were plated on YPD-zeocin/geneticin plates. All transformation plates were incubated for approximately 48h at 28°C and 80% humidity. After the incubation period deep-well plates (DWP) were filled with 300µl sterile BMD1% medium per well and inoculated with single colony picked from the transformation plates. This time not every, but every second well of the DWPs was inoculated. Namely, strong background of satellite colonies was seen on the transformation plates which made it difficult to pick the real positive clones. However, the inoculation of every second well increased the distance between each two colonies on the stamp plate and thereby disabled the growth of the

satellite colonies. Three wells per DWP were inoculated with a single colony of *P. pastoris* CSB7435 strain, used as negative control. After approximately 48h micro scale cultivation stamp plates of all DWPs were made and incubated for additional 48h at 28°C and 80% humidity. After that colony PCRs were made. One of the primers used for the colony PCRs was binding to the cassettes and the other to the *PpOCH1* or *PPHOC1* locus, respectively. These colony PCRs resulted with no products. Colony PCRs were also done using primers which bind inside the knockout cassettes. These resulted in PCR product with right size (the size was controlled on 1%-agarose gel).

11.5.1 Results and discussion

All together, nine transformations were done and their transformation rate varied between 20 and 30 clones per transformation plate. However, after colony PCR of over 150 clones per gene to be knocked out, no knockout strain was detected. All colony PCR products showed bands characteristic for *PpOCH1* and *PpHOC1* locus, indicating with that the knockout cassettes integrated in the genome by non-homologous recombination, and not at the desired OCH or HOC locus. For that reason the project strategy was slightly changed and a new HNL reporter strain was designed. This time the *HNL* reporter gene was cloned to *P. pastoris* **Δ**KU70 strain which is deletion strain where no non-homologous recombination can happen or only rarely happens (Paragraph 2.5). Additionally, *PpOCH1* and *PpHOC1* knockout cassettes containing 2500bp upstream and downstream homologous regions were also designed and *P. pastoris* CBS7435_HNL as well as *P. pastoris* **Δ**KU70_HNL electro-competent cells were transformed with the cassettes. Including 2500bp long flanking ends in the knockout cassettes drastically increased their size, and could negatively influence the rate of cassette integration in the genome. However, it is known from different experimental data that the size of the homologous sequences and the efficiency of their integration in *P. pastoris* genome follow logarithmic function (Paragraph 2.3). For that reason also longer knockout cassettes were designed and *P. pastoris* CBS7435 HNL as well as *P. pastoris* **Δ**KU70 HNL were transformed with the same.

11.5.2 Design and screening of *P. pastoris* ΔKU70 HNL reporter strain

The same plasmid carrying the PaHNL5_alpha gene used for design of *P. pastoris* CBS7435 HNL reporter strain was used also for design of *P. pastoris* ΔKU70 HNL reporter strains. It was linearized and *P. pastoris* ΔKU70 electro competent cells were transformed and plated out on LB-zeocin/geneticin plates. The possibly positive clones were picked and cultivated in DWPs and after that they were screened and rescreened using HNL enzyme activity assay. All experimental steps were done in the same way as for the design, screening and rescreen of *P. pastoris* CBS7435 HNL (Paragraph 11.4.1 and 11.4.2). For the amplification

of the 2500bp long *PpOCH1* and *PpHOC1* upstream and downstream homologous regions, genomic DNA of *P.p*CBS7435 was used as a template.

For the micro scale (DWPs) cultivation, 90 wells were inoculated with single colonies in DWP_1 (Figure 51), DWP_2 (Figure 52), DWP_3 (Figure 53), DWP_4 (Figure 54). Enzyme activity was measured for 90, 89, 90, and 90 clones, respectively. The final measure values were calculated by reducing each value with the average value for the negative control for each DWP. From all DWPs, eight clones showing highest activity were chosen and rescreened (Figure 55).

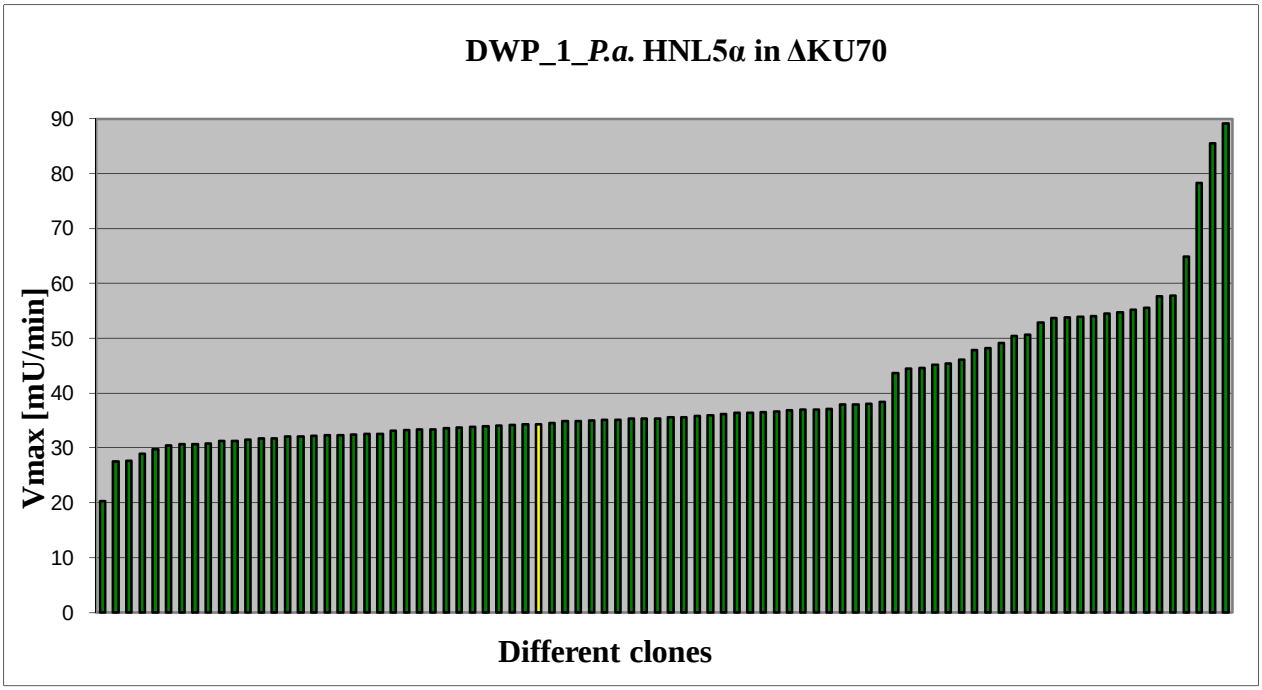

Figure 51: Landscape of photometric HNL activity screening of DWP_1 (*HNL5a* in *P.p.* ΔKU70). The yellow colored bar presents the clone used as positive control (HNL5_V317α).

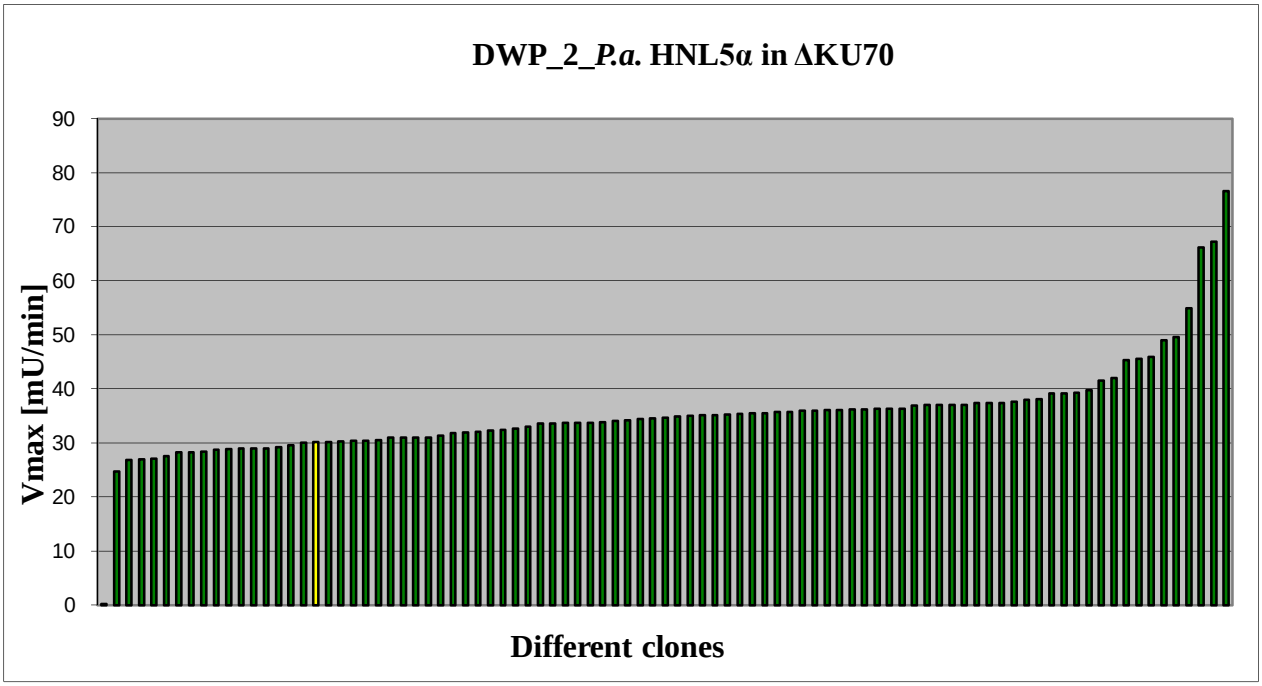

Figure 52: Landscape of photometric HNL activity screening of DWP_2 (*HNL5α* in *P.p.* ΔKU70). The yellow colored bar presents the clone used as positive control (HNL5_V317α).

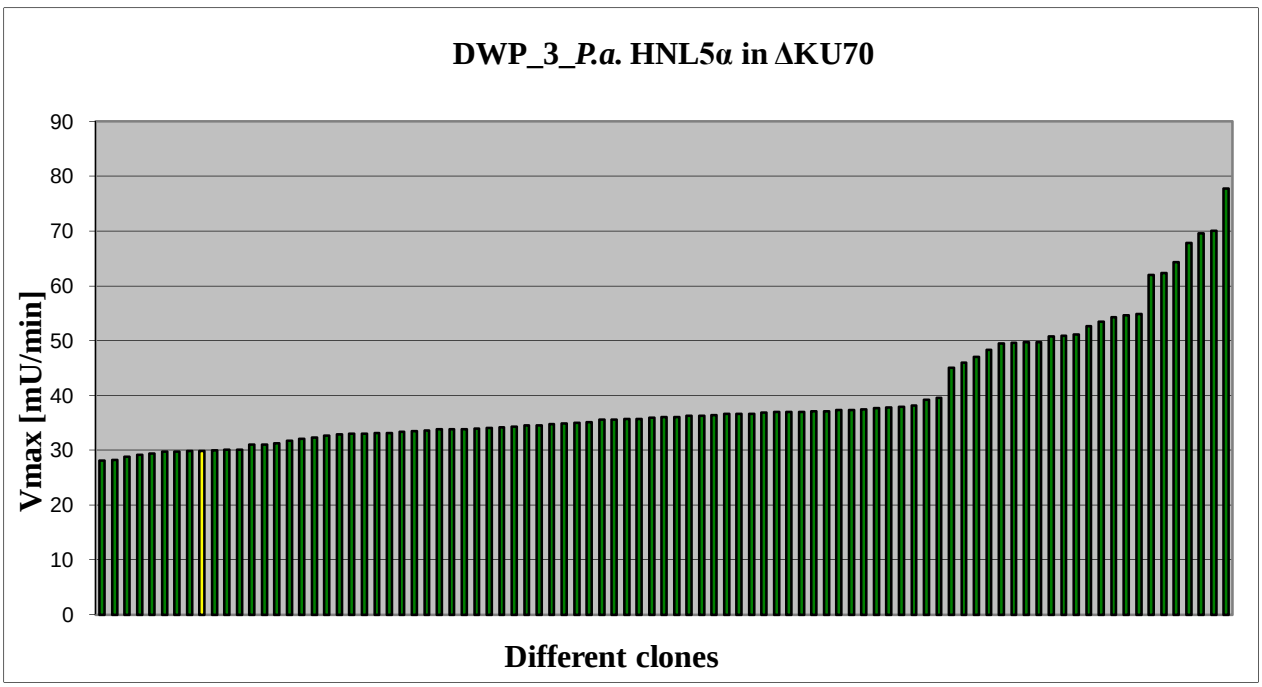

Figure 53: Landscape of photometric HNL activity screening of DWP_3 (*HNL5α* in *P.p.* ΔKU70). The yellow colored bar presents the clone used as positive control (HNL5_V317α).

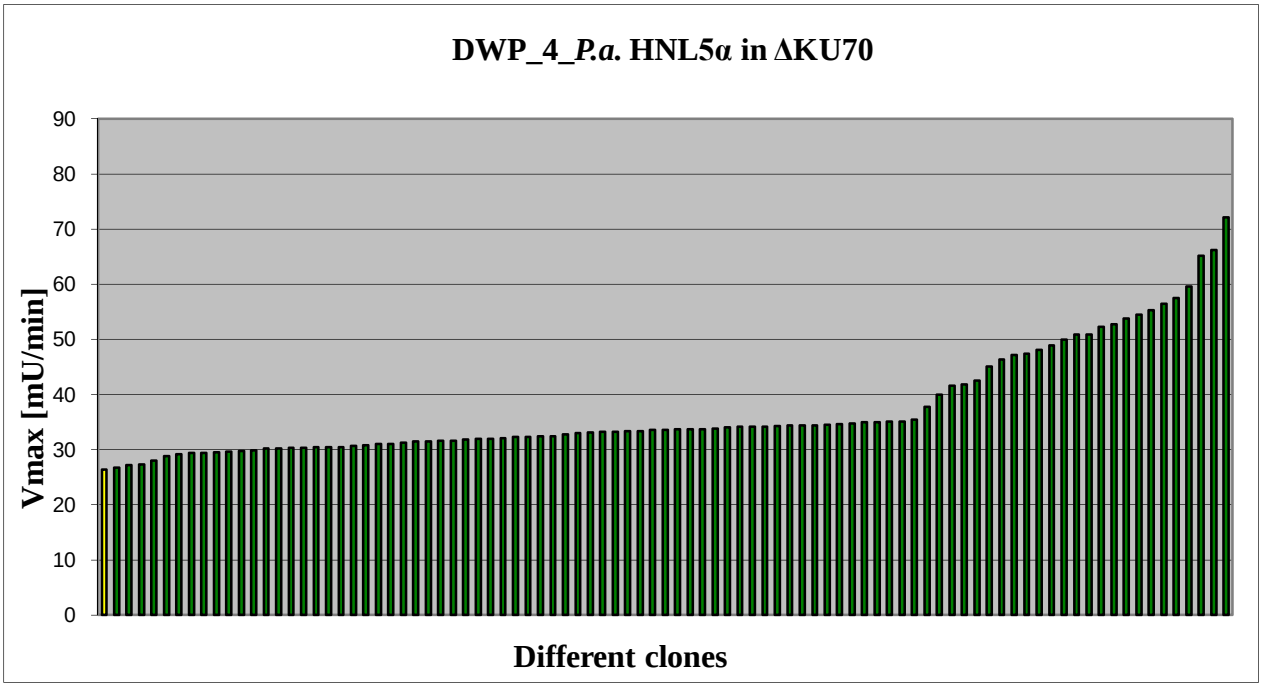

Figure 54: Landscape of photometric HNL activity screening of DWP_4 (*HNL5a* in *P.p.* ΔKU70). The yellow colored bar presents the clone used as positive control (HNL5_V317α).

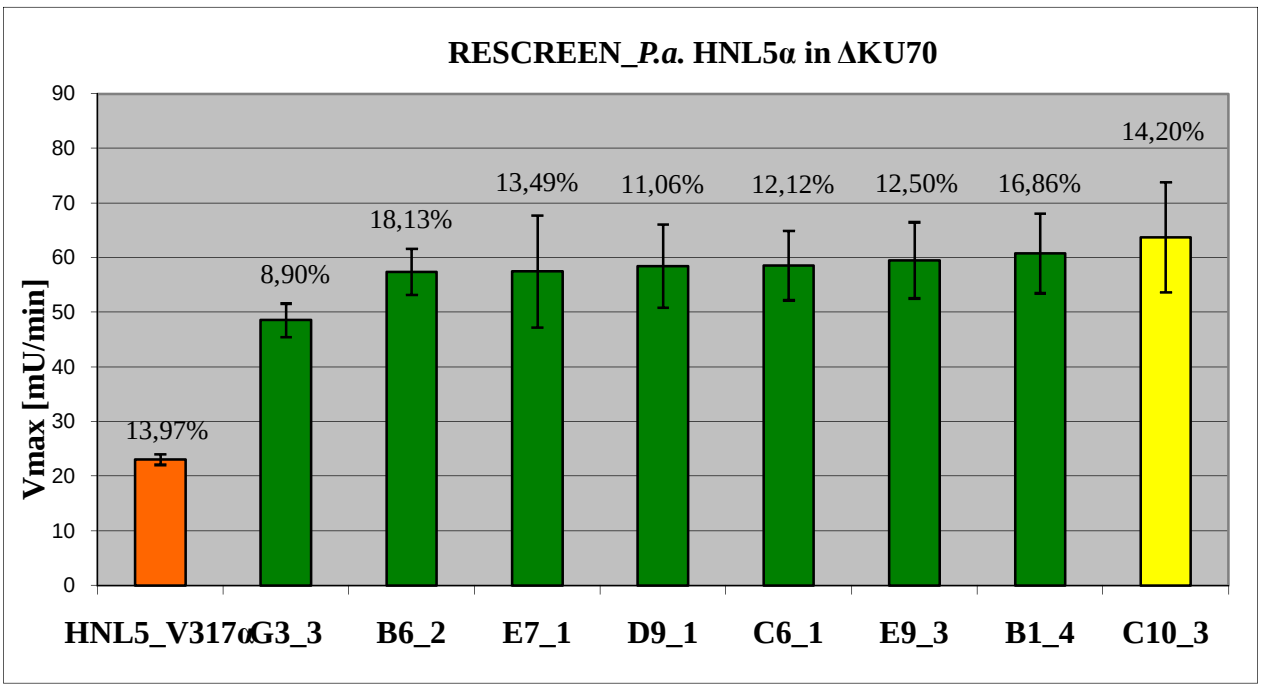

Figure 55: Landscape of photometric HNL rescreen of the most stable multi copy clones (*HNL5a* in ΔKU70). The orange colored bar presents the clone used as positive control (HNL5_V317α). The yellow colored bar represents the clone chosen as reporter strain for the further work (C10_3).

11.6 Results, discussion and further perspectives

Three transformations of *PpOCH1* and *PpHOC1* knockout cassettes (cassettes containing 2500bp long homologous regions) to *P. pastoris* CBS7435_HNL5 reporter strain (clone F7_3) were done and no knockout clones were detected.

For the transformation of the *P. pastoris* ΔKU70_HNL reporter strain (clone C10_3), eight transformations were done which resulted in all together 10 clones only. And, again after colony PCR none of these clones displayed on of the desired knockouts (Δ*OCH1* or Δ*HOC1*). It is also important to be mentioned that all 10 clones were found in the transformation with the long cassettes and with the highest DNA amount of approximately 1,5µg.

As already said (Paragraph 11.2), *PpOCH1* and *PpHOC1* gene products are ER and Golgi membrane located proteins, and theoretically their inactivation or deletion could disturb the function of these two crucial cell organelles. Also absence of proper glycosylation pattern in the other membrane proteins, caused by *OCH1* or *HOC1* inactivation or deletion could be lethal to the cell. Since *P. pastoris* ΔKU70 strain has already been used for design of knockout strains in our research group and it worked with quite high efficiency, the first logical final conclusion can be that the deletions of *PpOCH1* as well as *PpHOC1* are lethal for the cells or that the genomic locus is hardly accessible and even more transformants need to be analyzed.

However, Jacobs et al. has already reported about *P. pastoris* strain with disrupted *PpOCH1* gene and about the influence of *PpOCH1* disruption on the *N*-glycosylation pathway in the cell. Also report about *PpOCH1* deletion has been recently published by Zhang et al.. Unfortunately, this publication is in Chinese language, and thus not accessible for everyone. Taking in consideration these two publications, it seems that the inactivation or deletion of at least *PpOCH1* is not lethal to the cell. No publications with similar topic were found about the *PpHOC1*.

After everything that is mentioned above, it seems that the strategy used in this thesis project was not optimal for deletion of the desired genes.

However, because of the biotechnological importance of design of *P. pastoris* strains with engineered *N*-glycosylation pathway, further perspectives of this project were discussed. One of the ideas was the optimization of the cloning method including the preparation of *P. pastoris* competent cells. Other strategy changes would include the redesign of the knockout cassettes. Namely, elimination of other *P. pastoris* homologous parts from the cassettes, as PpP_{ILV5} should decrease the possibility of cassette integration at non-desired loci. One applicable improvement of the strategy was seen in the use of *PpOCH1* and *PpHOC1* natural promoter and terminator regions instead of PpP_{ILV5} and AODTT, respectively. In this way, the cassettes would be shorter, since the promoter and terminator would at the same time play a role of homologous flanking ends. This could increase the transformation rate and the possibility of cassettes proper integration in the genome. However, for that aim the promoters of *PpOCH1* and *PpHOC1* must be first validated.

Also totally different strategies, like using the recombinase systems could be applied, but this was not a topic of our discussion within this thesis.

12. References:

1. Cereghino, J.L. & Cregg, J.M. Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiology Reviews* **24**, 45-66 (2000).

2. Kurtzman, C.P. & Robnett, C.J. Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie van Leeuwenhoek* **73**, 331-371 (1998).

3. Kurtzman, C.P. Biotechnological strains of Komagataella (Pichia) pastoris are Komagataella phaffii as determined from multigene sequence analysis. *Journal of industrial microbiology & biotechnology* **36**, 1435-8 (2009).

4. Kurtzman, C.P. Description of Komagataella phaffii sp. nov. and the transfer of Pichia pseudopastoris to the methylotrophic yeast genus Komagataella. *International journal of systematic and evolutionary microbiology* **55**, 973-6 (2005).

5. Higgins, D.R. & Cregg, J.M. Introduction to Pichia pastoris. *Methods in molecular biology (Clifton, N.J.)* **103**, 1-15 (1998).

6. Küberl, A. *et al*. High-quality genome sequence of Pichia pastoris CBS7435. *Journal of biotechnology* **154**, 312-20 (2011).

7. Cereghino, G.P.L., Cereghino, J.L., Ilgen, C. & Cregg, J.M. Production of recombinant proteins in fermenter cultures of the yeast Pichia pastoris. *Current opinion in biotechnology* **13**, 329-32 (2002).

8. Macauley-Patrick, S., Fazenda, M.L., McNeil, B. & Harvey, L.M. Heterologous protein production using the Pichia pastoris expression system. *Yeast (Chichester, England)* **22**, 249-70 (2005).

9. Hartner, F.S. & Glieder, A. Regulation of methanol utilisation pathway genes in yeasts. *Microbial cell factories* **5**, 39 (2006).

10. Ruth, C. & Glieder, A. Perspectives on synthetic promoters for biocatalysis and biotransformation. *Chembiochem□: a European journal of chemical biology* **11**, 761-5 (2010).

11. Sears, I.B., O'Connor, J., Rossanese, O.W. & Glick, B.S. A versatile set of vectors for constitutive and regulated gene expression in Pichia pastoris. *Yeast (Chichester, England)* **14**, 783-90 (1998).

12. SREEKRISHNA, K. Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast Pichia pastoris*1. *Gene* **190**, 55-62 (1997).

13. Cregg, J.M., Madden, K.R., Barringer, K.J., Thill, G.P. & Stillman, C.A. Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. *Molecular and cellular biology* **9**, 1316-23 (1989).

14. Cregg, J.M., Cereghino, J.L., Shi, J. & Higgins, D.R. Recombinant protein expression in Pichia pastoris. *Molecular biotechnology* **16**, 23-52 (2000).

15. Montesino, R., García, R., Quintero, O. & Cremata, J. a Variation in N-linked oligosaccharide structures on heterologous proteins secreted by the methylotrophic yeast Pichia pastoris. *Protein expression and purification* **14**, 197-207 (1998).

16. Cregg, J.M. *et al.* Expression in the yeast Pichia pastoris. *Methods in enzymology* **463**, 169-89 (2009).

17. Meagher, M.M. & Inan, M. Non-Repressing Carbon Sources for Alcohol Oxidase (AOXI) Promoter of Pichia pastoris Non-Repressing Carbon Sources for Alcohol Oxidase (AOXI) Promoter of Pichia pastoris.

18. © 1993 Nature Publishing Group http://www.nature.com/naturebiotechnology. *Group* **11**, 6 (1993).

19. Sørensen, H.P. Towards universal systems for recombinant gene expression. *Microbial cell factories* **9**, 27 (2010).

20. Gellissen, G. Heterologous protein production in methylotrophic yeasts. *Applied microbiology and biotechnology* **54**, 741-50 (2000).

21. Faber, K.N., Harder, W., Ab, G. & Veenhuis, M. Review: methylotrophic yeasts as factories for the production of foreign proteins. *Yeast (Chichester, England)* **11**, 1331-44 (1995).

22. Ews, R. The endoplasmic reticulum as a. *Trends in Cell Biology* 227-231 (1992).

23. © 1992 Nature Publishing Group. *Group* (1992).

24. Inan, M., Aryasomayajula, D., Sinha, J. & Meagher, M.M. Enhancement of protein secretion in Pichia pastoris by overexpression of protein disulfide isomerase. *Biotechnology and bioengineering* **93**, 771-8 (2006).

25. Xiao, R. *et al.* The contributions of protein disulfide isomerase and its homologues to oxidative protein folding in the yeast endoplasmic reticulum. *The Journal of biological chemistry* **279**, 49780-6 (2004).

26. Shaw, P.E. Peptidyl-prolyl isomerases: a new twist to transcription. *EMBO reports* **3**, 521-6 (2002).

27. Mattanovich, D., Gasser, B., Hohenblum, H. & Sauer, M. Stress in recombinant protein producing yeasts. *Journal of biotechnology* **113**, 121-35 (2004).

28. Tian, G., Xiang, S., Noiva, R., Lennarz, W.J. & Schindelin, H. The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. *Cell* **124**, 61-73 (2006).

29. Xiao, R., Lundström-Ljung, J., Holmgren, A. & Gilbert, H.F. Catalysis of thiol/disulfide exchange. Glutaredoxin 1 and protein-disulfide isomerase use different mechanisms to enhance oxidase and reductase activities. *The Journal of biological chemistry* **280**, 21099-106 (2005).

30. Gallogly, M.M., Starke, D.W. & Mieyal, J.J. Mechanistic and kinetic details of catalysis of thiol-disulfide exchange by glutaredoxins and potential mechanisms of regulation. *Antioxidants & redox signaling* **11**, 1059-81 (2009).

31. Jessop, C.E., Watkins, R.H., Simmons, J.J., Tasab, M. & Bulleid, N.J. Protein disulphide isomerase family members show distinct substrate specificity: P5 is targeted to BiP client proteins. *Journal of cell science* **122**, 4287-95 (2009).

32. van Vliet, C., Thomas, E.C., Merino-Trigo, A., Teasdale, R.D. & Gleeson, P. a Intracellular sorting and transport of proteins. *Progress in Biophysics and Molecular Biology* **83**, 1-45 (2003).

33. Frand, a R., Cuozzo, J.W. & Kaiser, C. a Pathways for protein disulphide bond formation. *Trends in cell biology* **10**, 203-10 (2000).

34. Wedemeyer, W.J., Welker, E., Narayan, M. & Scheraga, H.A. Current Topics Disulfide Bonds and Protein Folding †. *Current* **39**, (2000).

35. Nørgaard, P. *et al.* Functional differences in yeast protein disulfide isomerases. *The Journal of cell biology* **152**, 553-62 (2001).

36. Warsame, A., Vad, R., Kristensen, T. & Oyen, T.B. Characterization of a gene encoding a Pichia pastoris protein disulfide isomerase. *Biochemical and biophysical research communications* **281**, 1176-82 (2001).

37. Zhang, S.-T. *et al.* Co-overexpression of PpPDI Enhances Secretion of Ancrod in Pichia pastoris. *Applied biochemistry and biotechnology* (2011).doi:10.1007/s12010-011-9192-4

38. Charlwood, J., Bryant, D., Skehel, J.M. & Camilleri, P. Analysis of N-linked oligosaccharides: progress towards the characterisation of glycoprotein-linked carbohydrates. *Biomolecular engineering* **18**, 229-40 (2001).

39. Helenius, J. & Aebi, M. Transmembrane movement of dolichol linked carbohydrates during N-glycoprotein biosynthesis in the endoplasmic reticulum. *Online* **13**, 171-178 (2002).

40. Bretthauer, R.K. & Castellino, F.J. Glycosylation of Pichia pastoris-derived proteins. *Biotechnology and applied biochemistry* **30 (Pt 3)**, 193-200 (1999).

41. Hamilton, S.R. & Gerngross, T.U. Glycosylation engineering in yeast: the advent of fully humanized yeast. *Current opinion in biotechnology* **18**, 387-92 (2007).

42. Vervecken, W. *et al.* In Vivo Synthesis of Mammalian-Like , Hybrid-Type N-Glycans in Pichia pastoris. *Society* **70**, 2639-2646 (2004).

43. Jacobs, P.P., Geysens, S., Vervecken, W., Contreras, R. & Callewaert, N. Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch technology. *Nature protocols* **4**, 58-70 (2009).

44. Gemmill, T.R. & Trimble, R.B. Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochimica et biophysica acta* **1426**, 227-37 (1999).

45. Yip, C.L.A.M. *et al.* Cloning and analysis of the Saccharomyces cerevisiae MNN9 and MNNI genes required for complex glycosylation of secreted proteins. *Proteins* **91**, 2723-2727 (1994).

46. Romero, P.A. *et al.* Mnt2p and Mnt3p of Saccharomyces cerevisiae are members of the Mnn1p family of α-1 , 3-mannosyltransferases responsible for adding the terminal mannose residues of O-linked oligosaccharides. *Glycobiology* **9**, 1045-1051 (1999).

47. Grinna, L.S. & Tschopp, J.F. Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast, Pichia pastoris. *Yeast (Chichester, England)* **5**, 107-15 (1989).

48. Romanos, M. Advances in the use of Pichia pastoris for high-level gene expression. *Current Opinion in Biotechnology* **6**, 527-533 (1995).

49. Davidson, R.C. *et al.* Functional analysis of the ALG3 gene encoding the Dol-P-Man: Man5GlcNAc2-PP-Dol mannosyltransferase enzyme of P. pastoris. *Glycobiology* **14**, 399-407 (2004).

50. Nakayama, K., Nagasu, T., Shimma, Y., Kuromitsu, J. & Jigami, Y. OCH1 encodes a novel membrane bound mannosyltransferase: outer chain elongation of asparagine-linked oligosaccharides. *The EMBO journal* **11**, 2511-9 (1992).

51. Hamilton, S.R. *et al.* Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science (New York, N.Y.)* **313**, 1441-3 (2006).

52. San Filippo, J., Sung, P. & Klein, H. Mechanism of eukaryotic homologous recombination. *Annual review of biochemistry* **77**, 229-57 (2008).

53. Shibata, T. Functions of homologous DNA recombination. *Science And Technology* **41**, 21-23 (2001).

54. Klinner, U. & Schäfer, B. Genetic aspects of targeted insertion mutagenesis in yeasts. *FEMS microbiology reviews* **28**, 201-23 (2004).

55. Hastings, P.J., Mcgill, C., Shafert, B. & Strathem, J.N. Ends-Out Recombination in Yeast. **558**, (1993).

56. Kit, P.E. Expression Kit.

57. *Pichia protocols (Google eBook)*. 270 (Humana Press: 1998).at <http://books.google.com/books?id=0qwykqk-lUgC&pgis=1>

58. Sadowski, P.D. Recombinase of. *Microbiology* **5**, 3274-3279 (1985).

59. McLeod, M., Craft, S. & Broach, J.R. Identification of the crossover site during FLP-mediated recombination in the Saccharomyces cerevisiae plasmid 2 microns circle. *Molecular and cellular biology* **6**, 3357-67 (1986).

60. Sadowski, P. Site-specific recombinases: changing partners and doing the twist. *Journal of bacteriology* **165**, 341-7 (1986).

61. Sauer, B. Site-specific recombination: developments and applications. *Current opinion in biotechnology* **5**, 521-7 (1994).

62. Gronostajskit, R.M. & Sadowski, P.D. The FLP Recombinase of the Saccharomyces cerevisiae 2pLm Plasmid Attaches Covalently to DNA via a Phosphotyrosyl Linkage. (1985).doi:10.1128/MCB.5.11.3274.Updated

63. Inman, R.B. & Coxg, M.M. Protein-based Asymmetry and Protein-Protein Interactions Site-specific Recombination * in FLP. **265**, (1990).

64. Andrews, B.J., McLeod, M., Broach, J. & Sadowski, P.D. Interaction of the FLP recombinase of the Saccharomyces cerevisiae 2 micron plasmid with mutated target sequences. *Molecular and cellular biology* **6**, 2482-9 (1986).

65. Kii, N.J., Snaith, I.I.Y.M.R. & Murray, J.A.H. tools for gen0me engineering. *Cloning* (1993).

66. Marx, H., Mattanovich, D. & Sauer, M. Overexpression of the riboflavin biosynthetic pathway in Pichia pastoris. *Microbial cell factories* **7**, 23 (2008).

67. Gravel, S. Yeast Ku as a Regulator of Chromosomal DNA End Structure. *Science* **280**, 741-744 (1998).

68. Walker, J.R., Corpina, R. a & Goldberg, J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**, 607-14 (2001).

69. Abad, S. *et al.* Real-time PCR-based determination of gene copy numbers in Pichia pastoris. *Biotechnology journal* **5**, 413-20 (2010).

70. Weis, R., Gaisberger, R., Gruber, K. & Glieder, A. Serine scanning: a tool to prove the consequences of N-glycosylation of proteins. *Journal of biotechnology* **129**, 50-61 (2007).

Appendix 1: CalB activity assay – screening results

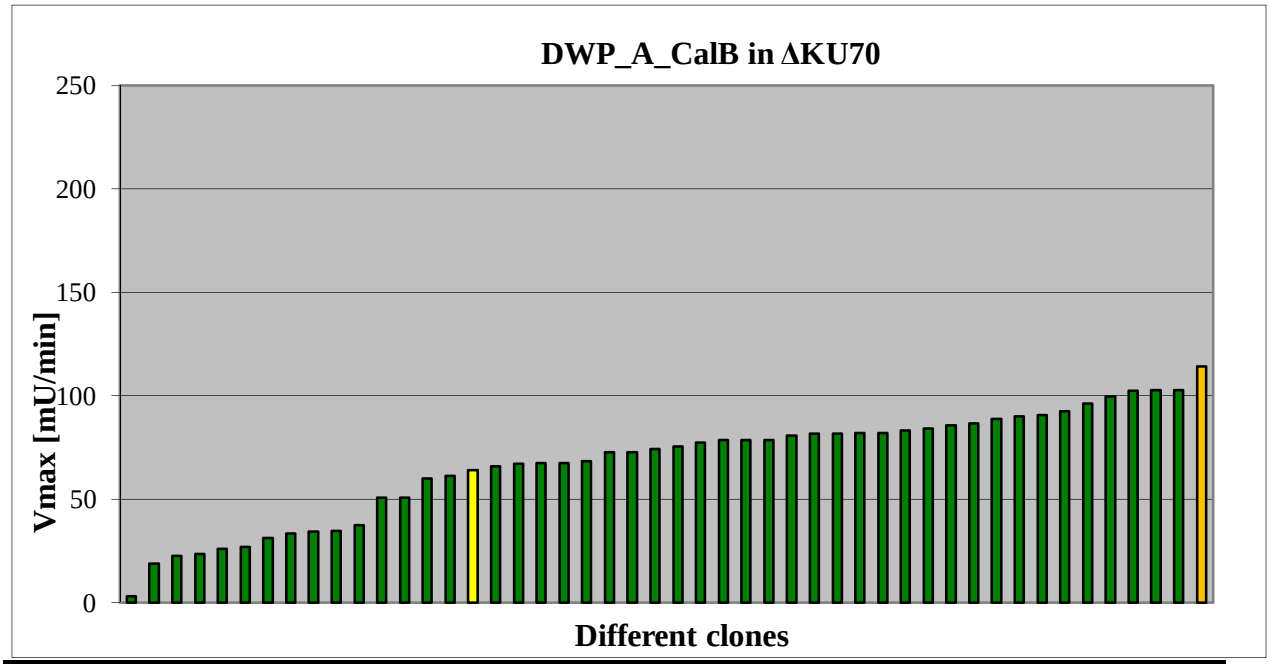

Figure 56: Landscape of photometric CalB activity screening of DWP_A. The orange colored bar presents the clone used as positive control with high activity (*Pp*Mut^{s}_CalB_E9), and the yellow colored bar presents the clone used as positive control with low activity (*Pp*Mut^{s}_CalB_A9).

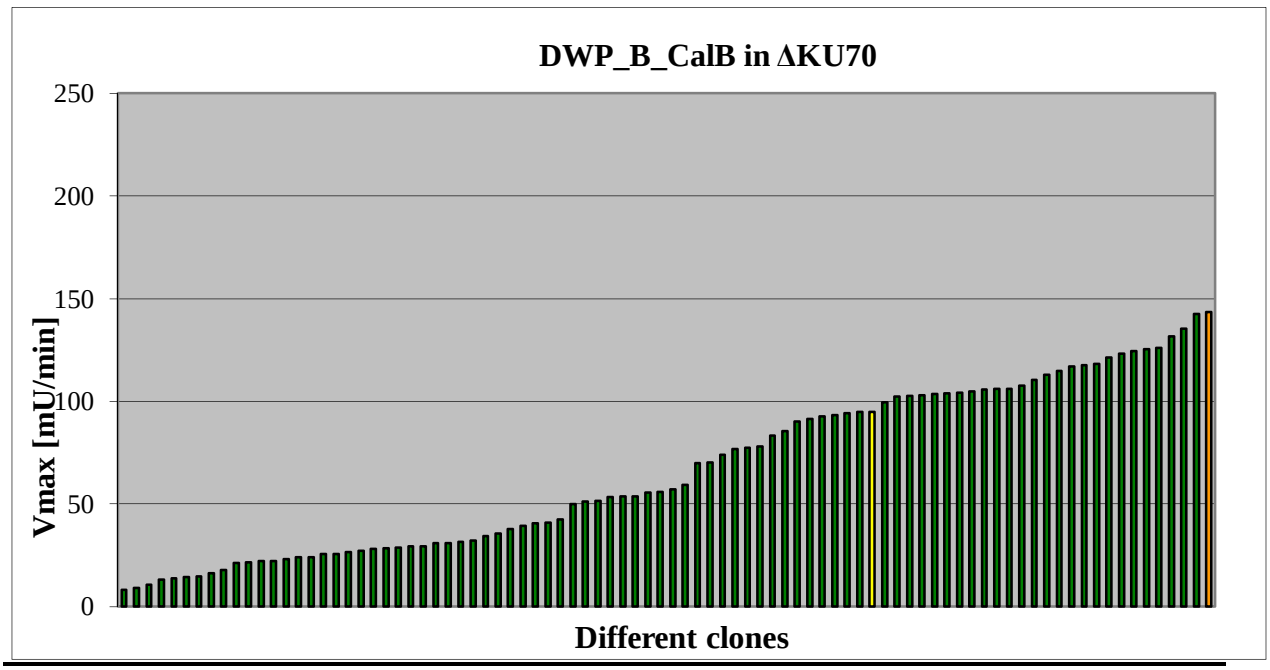

Figure 57: Landscape of photometric CalB activity screening of DWP_B. The orange colored bar presents the clone used as positive control with high activity (*Pp*Mut^{s}_CalB_E9), and the yellow colored bar presents the clone used as positive control with low activity (*Pp*Mut^{s}_CalB_A9).

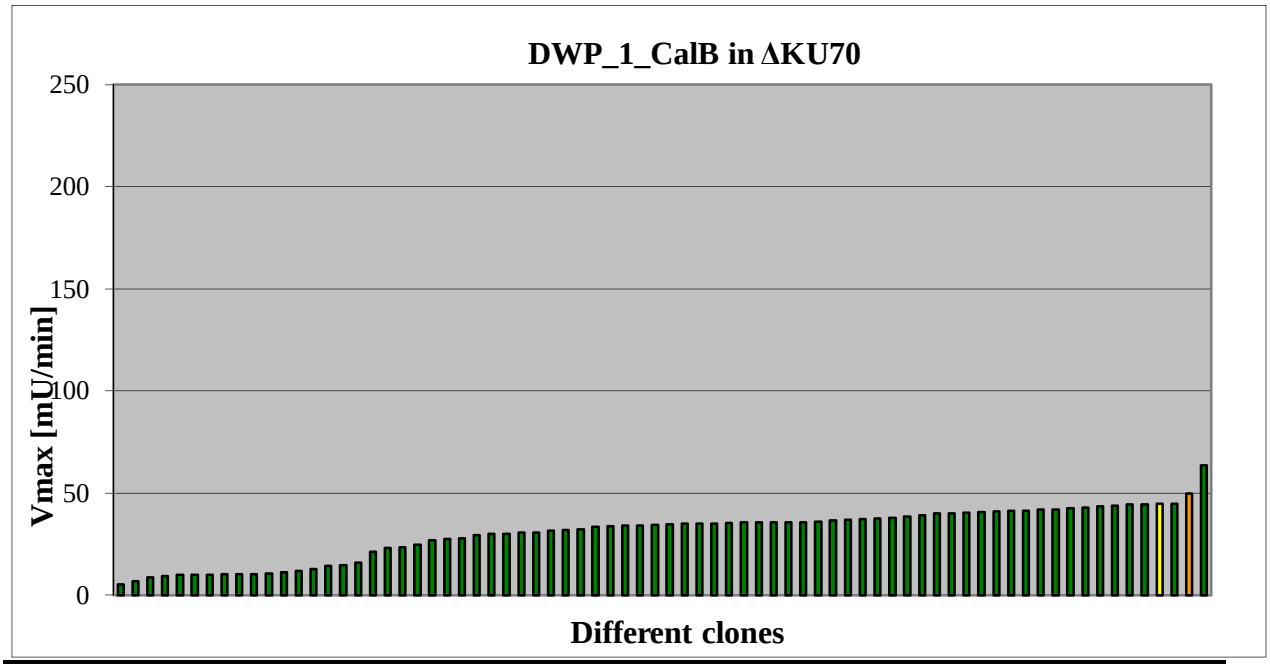

Figure 58: Landscape of photometric CalB activity screening of DWP_1. The orange colored bar presents the clone used as positive control with high activity (*Pp*Mut^{s}_CalB_E9), and the yellow colored bar presents the clone used as positive control with low activity (*Pp*Mut^{s}_CalB_A9).

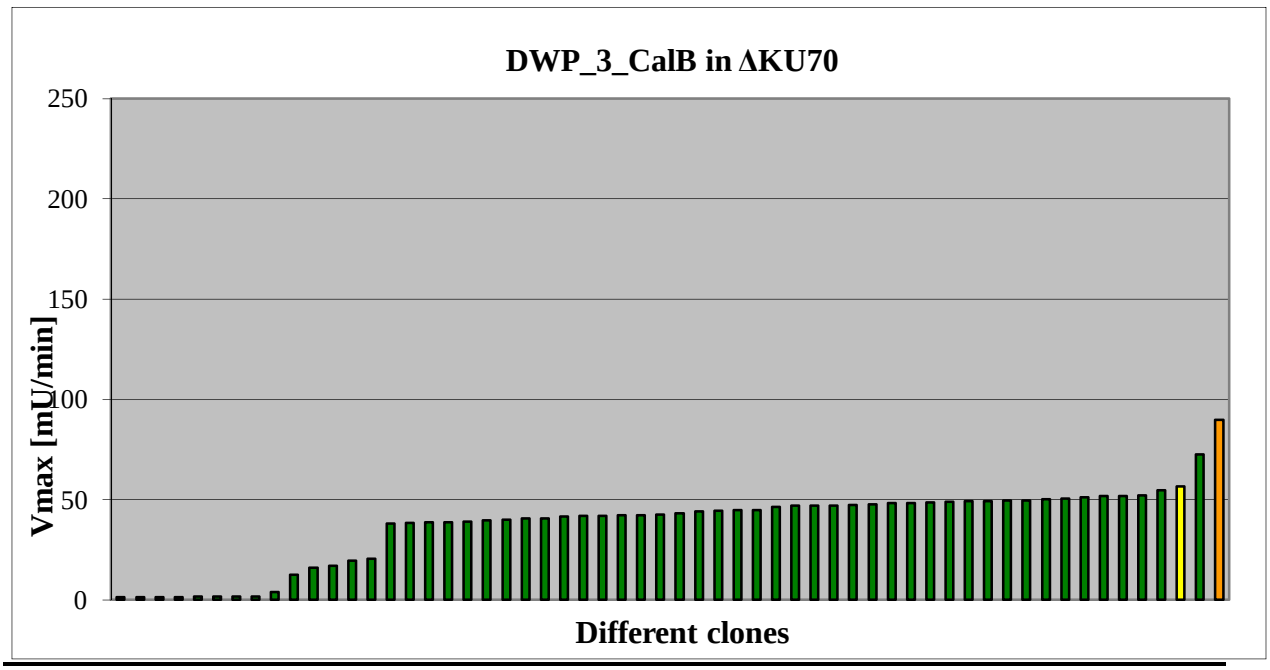

Figure 59: Landscape of photometric CalB activity screening of DWP_3. The orange colored bar presents the clone used as positive control with high activity (*Pp*Mut^{s}_CalB_E9), and the yellow colored bar presents the clone used as positive control with low activity (*Pp*Mut^{s}_CalB_A9).

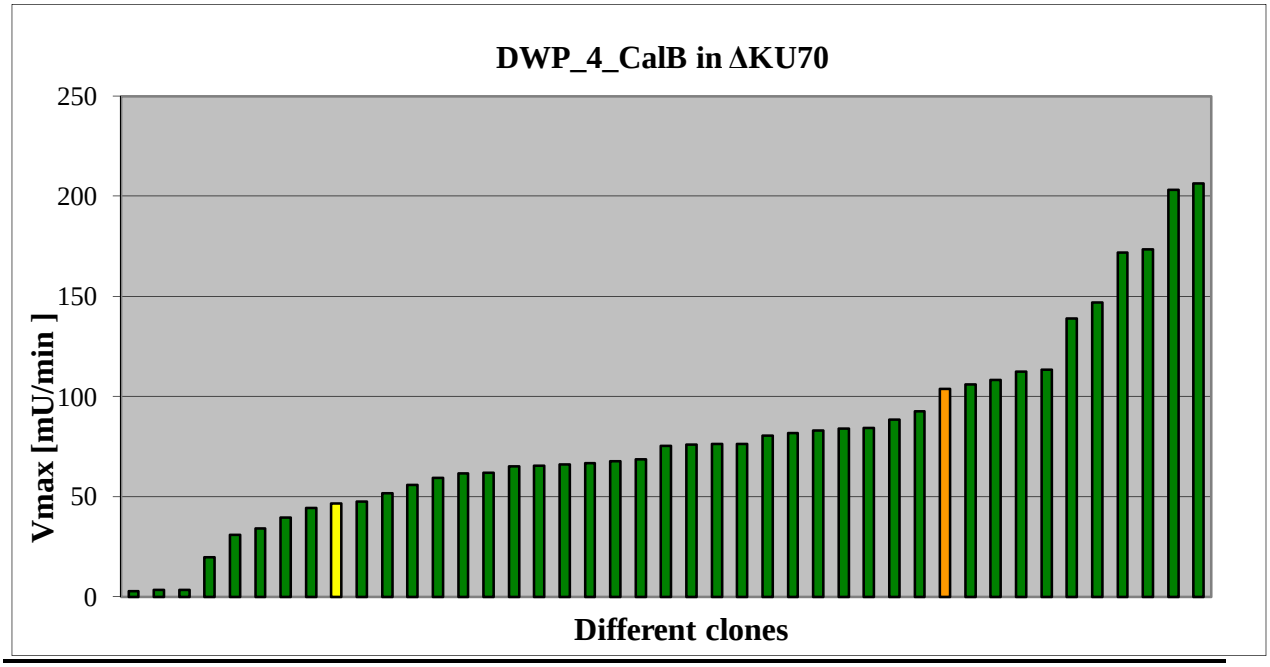

Figure 60: Landscape of photometric CalB activity screening of DWP_4. The orange colored bar presents the clone used as positive control with high activity (*Pp*Mut^{s}_CalB_E9), and the yellow colored bar presents the clone used as positive control with low activity (*Pp*Mut^{s}_CalB_A9).

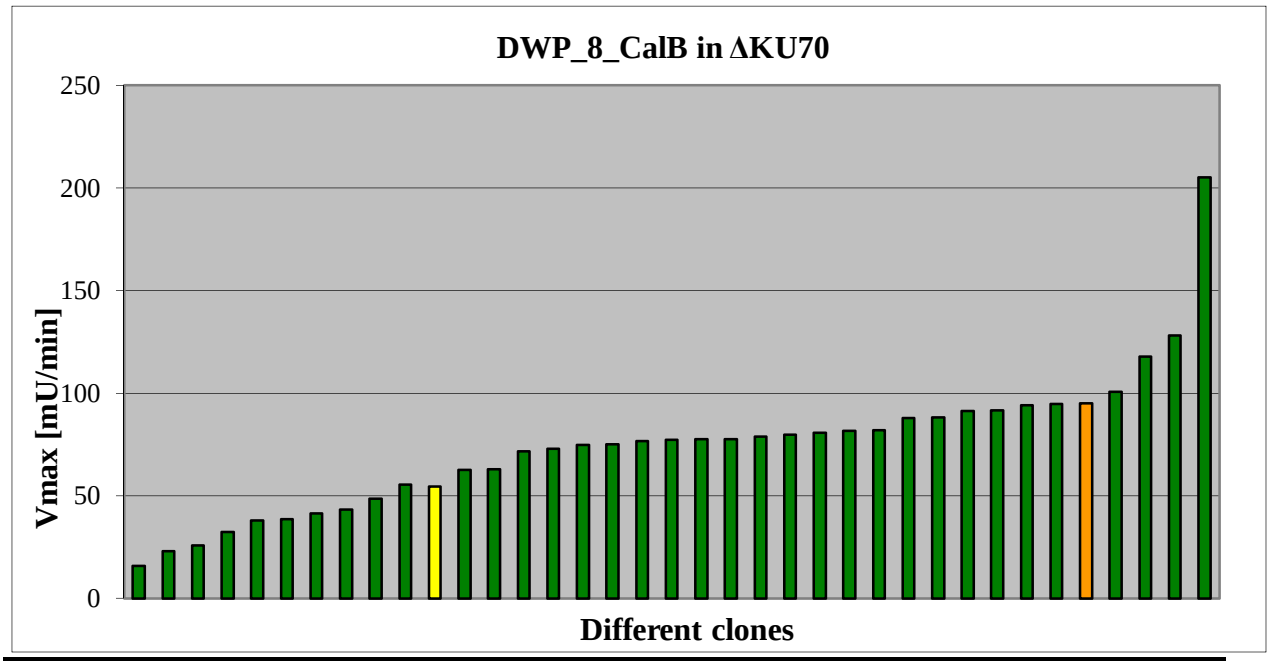

Figure 61: Landscape of photometric CalB activity screening of DWP_8. The orange colored bar presents the clone used as positive control with high activity (*Pp*Mut^{s}_CalB_E9), and the yellow colored bar presents the clone used as positive control with low activity (*Pp*Mut^{s}_CalB_A9).

Appendix 2: HRP activity assay – screening results

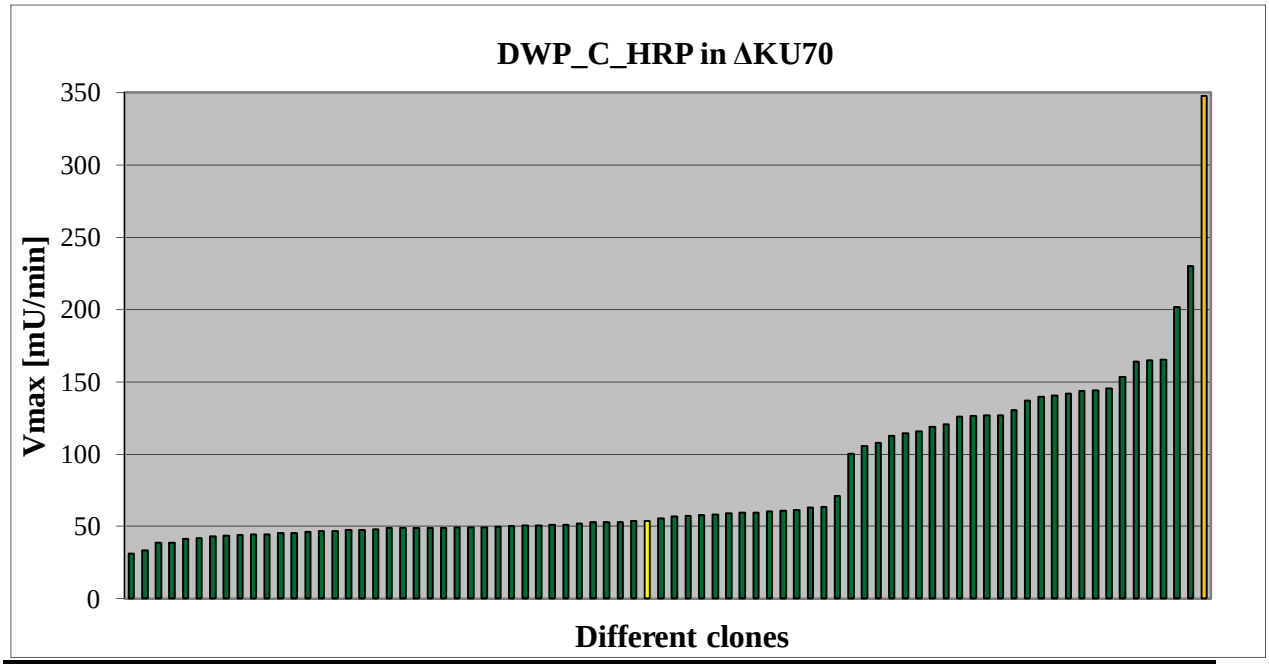

Figure 62: Landscape of photometric HRP activity screening of DWP_C. The orange colored bar presents the clone used as positive control with high activity (*Pp*Mut^{s}_HRP_H5), and the yellow colored bar presents the clone used as positive control with low activity (*Pp*Mut^{s}_HRP_H12).

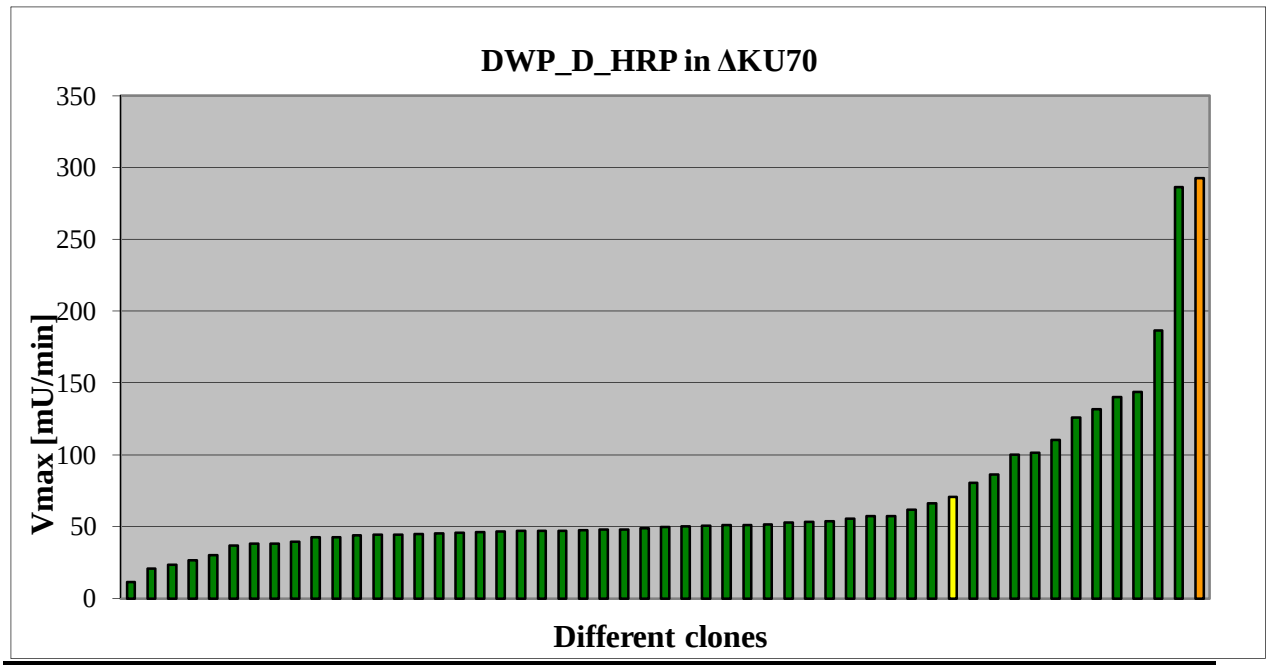

Figure 63: Landscape of photometric HRP activity screening of DWP_D. The orange colored bar presents the clone used as positive control with high activity (*Pp*Mut^{s}_HRP_H5), and the yellow colored bar presents the clone used as positive control with low activity (*Pp*Mut^{s}_HRP_H12).

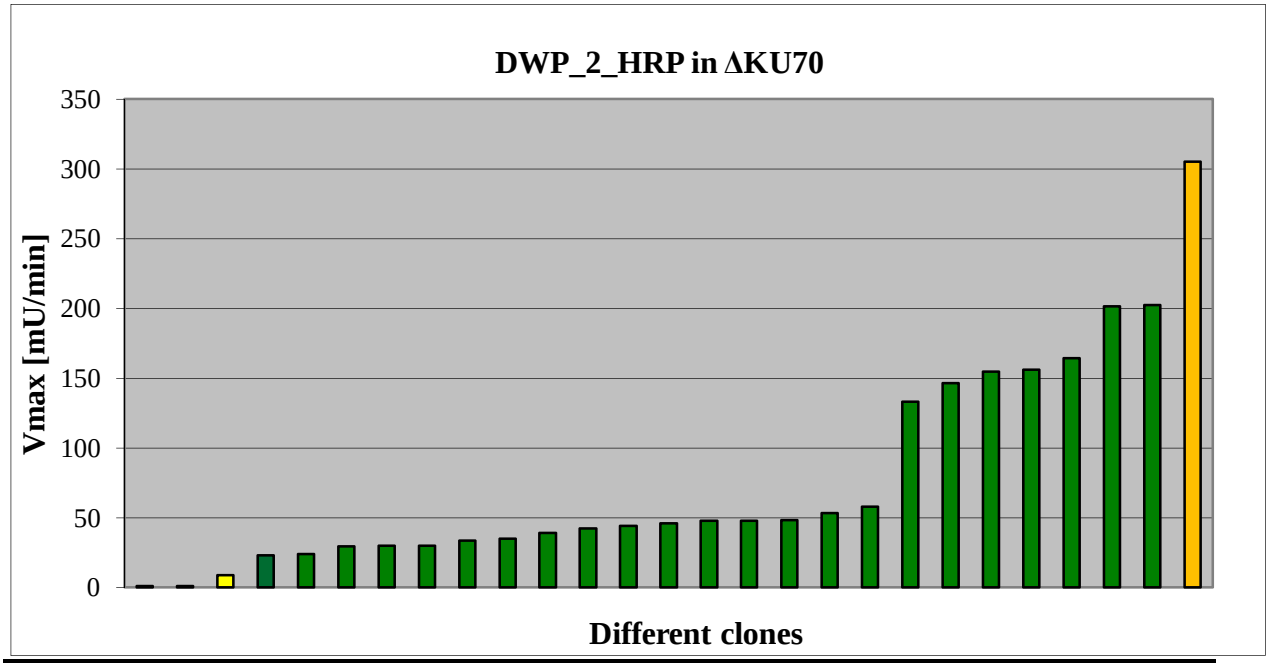

Figure 64: Landscape of photometric HRP activity screening of DWP_2. The orange colored bar presents the clone used as positive control with high activity (*Pp*Mut^{s}_HRP_H5), and the yellow colored bar presents the clone used as positive control with low activity (*Pp*Mut^{s}_HRP_H12).

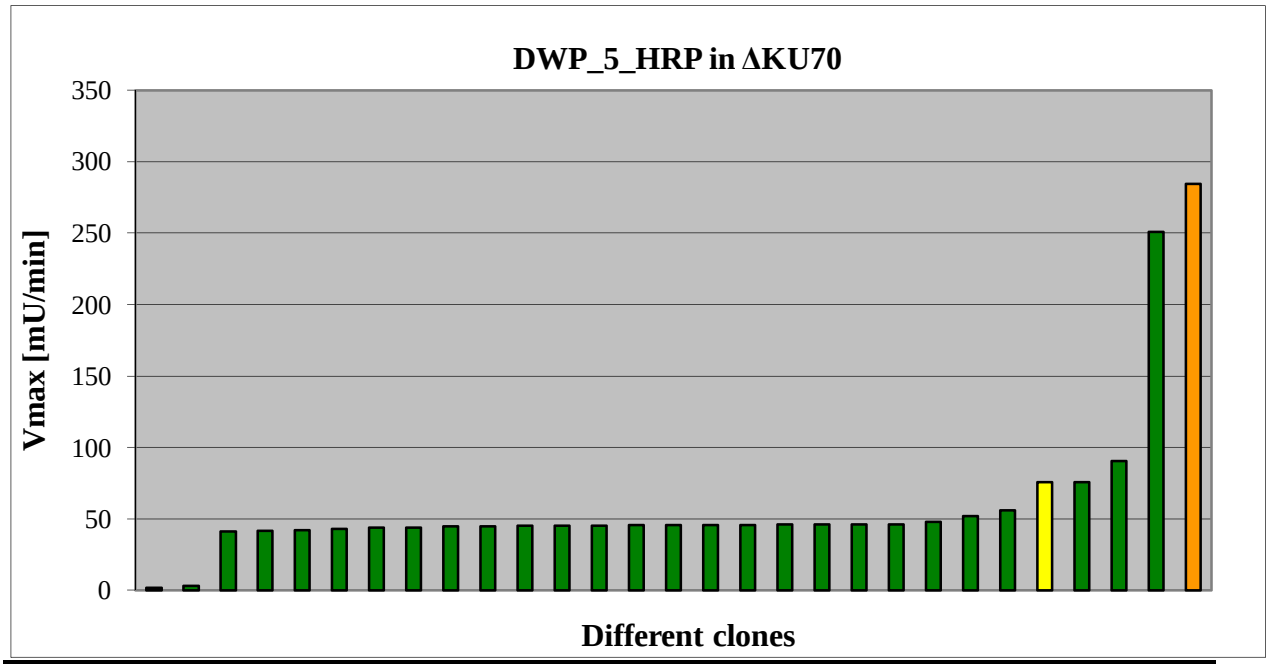

Figure 65: Landscape of photometric HRP activity screening of DWP_5. The orange colored bar presents the clone used as positive control with high activity (*Pp*Mut^{s}_HRP_H5), and the yellow colored bar presents the clone used as positive control with low activity (*Pp*Mut^{s}_HRP_H12).

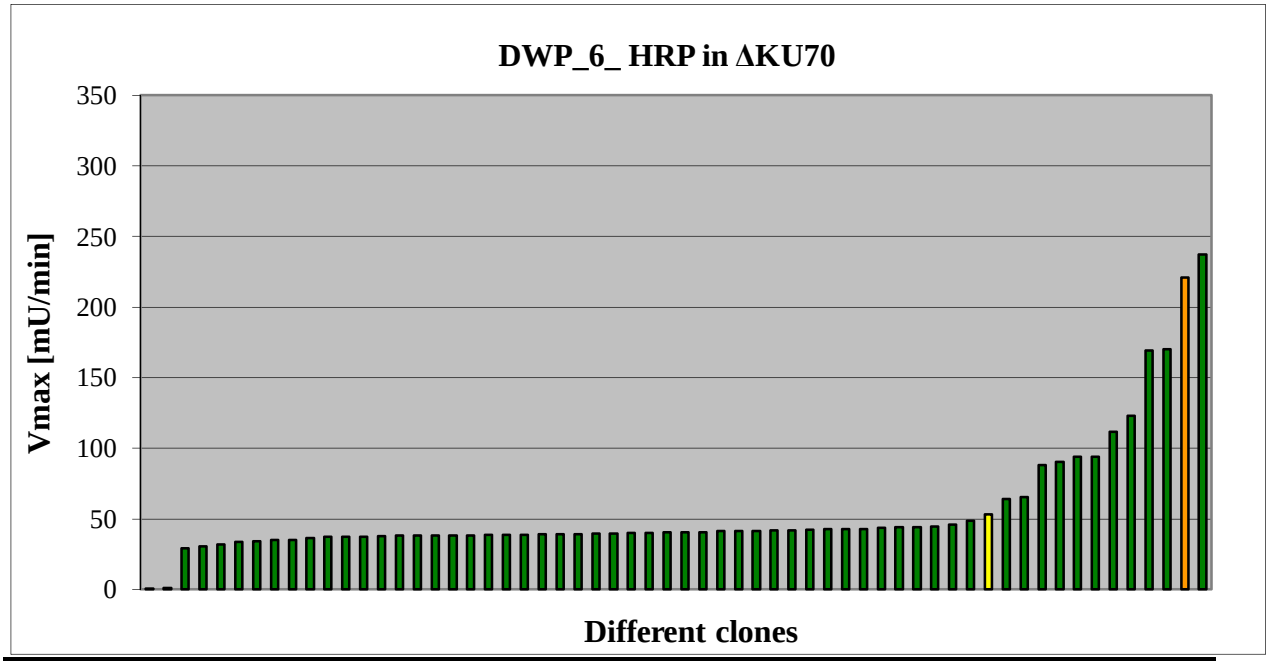

Figure 66: Landscape of photometric HRP activity screening of DWP_6. The orange colored bar presents the clone used as positive control with high activity (*Pp*Mut^{s}_HRP_H5), and the yellow colored bar presents the clone used as positive control with low activity (*Pp*Mut^{s}_HRP_H12).

Appendix 3: DNA sequences of relevance for this thesis

NCBI Accession number: CCA40283.1
Gene designation (GenDB): PP7435 Chr4-0107
EC number: 5.3.4.1
Gene product: K09580 prlyl 4-hydroxylase, beta polypeptide

3000bp upstream reagion / CDS (colored gray) / 1000bp downstream region

gttcaaacct agcgctaatg ttttcaggca tttgttggct cttttaactg aagttcctgg tgtttcctat
gatggagggg atcaagggct gattaactat gtattccaaa acaaatggtt gcgcactggt gatgacacaa
aacggtgtgg tgtatggtac aatcttagtt ttgcgttcaa tatgacactt tccaacaact atgagagtct
tccttcagta ctgcgaaatc tcacagacat caaattggtt catttcatag gaattgtgaa accatggatg
cttaaaccat cttttgtgaa cgactttcca gatggcagtc ttgacagttt tgttgcccaa tggtgggagc
agtttagctc ttttgagaac ggtgagtttc tacctttggt ttttaagaat gttgaaagcg aacgtattga
agaagactcc catgagacgg aagaaaaggt ggacgaagaa gtctccattt cagagcctca ggatgaaacc
actgatttca agtaccagtt tggccatcat tcctttgaag aacctgcacc ggtattagat tattcaactg
agggagaagc atggaagctt aatgaagaac aattgactaa tcagtgggat gttgatgcac cagctgagcc
gctgcctgtc cctgttgaag aagacgaacg cgaagagacg aaggccgaag ctgaattgga ggagctgcta
ccggatattg ttcagcctga acctccagca cctcatgttt ttccctggga agcttacaat gaaaagccga
cacgagtttt ccatgattac aggtaaaact ttaactaaat ataaggtata agatagcatc attttcgtat
ttagctggta cttagtcgaa ttttctactg catggtagcg tttgatacac gttgcccatt cttctcttta
agtgcgatta actgcttgga ccggtaagtc tcgtaatgaa cgaccgaagt agtctctatt aaatcatgga
ggtgggtacg tgttagaaag tcccgcaggt acacaaattc gcactgcgtt acatcttcca cgttgattag
accccatcgg gttctacgac ctctgatgac ctgtccatca attgtggaga tctcattttc agagccaatt
actgcaaacg gtaaaagact tcggatatct tgattcaaag tcagttcttc ttcactcaaa tcttcattgt
cgtacggaaa aaagtttagg tcgtgtttct caaattgagt gtgcaaaata cttttaaact tggtacgctc
ttcggtagtc aaggtgtcag ccttagcaat aacaggaata acgttcacaa cagtagtgag tttggaaaga
acttgtatgt caaggtttct cagtccctta ccgttaggtt gaatgaagta aagcacacaa tgaacccttg
tgtcaggtat ataccttttct ctctgtgctg tcaactcttt tcgcaagtag gtagagtttt gctctttgat
atactgcaat attggttccc aacatttgtc attgttgatc tgatctccaa aaccaggtgt gtcaatcacg
ttgagtttca gattaacgtt gttttcattg agaatttggg aggtgctttt aatctctaca gttctttcca
caggctcaaa agtacttttt cttccgttag agtccacaag atgtgtggca aaaagagtat tgactagagt
tgacttccca agcccagatt gaccaacaag cataacattg agctcaaatc cacgtttaag gagtctgttt
tcaatttggg tagtgatagt atcaaacccc acataattct taggttgaac cacttgactt gccatagttt
ttgtattatc tgacgaacaa aatgaaagag aataactgta accaacgcgt ggacaggtag gacacagaaa
tcgaggagaa aggcgattgc attcgcgact gtatctttcc gtgcagctct attacgagag tggtgttctt
agtttttccc tcttcaattt ctactaaagc tgctttggta aagggagaca tattcggcta ttgtttactt
tgcgcccaca gtagcgttaa gaaacattgt ttgttcgatt tattgggctg ttgataaatt caattgatta
cgttcgcata ctagctatca taaactaagc accaccttac accactttct cactgaagat tttcgacatc
aaatttctct tggatcacca tcaaccttgt gtctacatgt ccttgtcttt gaacctaaat cagatagccg
tgcgggttgt gggcatattg cctcgtattc cggagattca cattgccatt cctaatattt ttcagcgacg
caccgaagct tctacagaga ctcacgatcc tcgcatacta gagctgatag aaaatctaca ggatgccgag
gttcctccat tcttcattga taacggtata cttaaagcag caccaaaaaa gaaggtttct catatgaaaa
gacgccagaa attatatggt ccaggaaaaa aacaactctc tttactacaa aatttgaaca ggtgtcctgc
ctgcggaaac tacaaacgat cacacaccct ctgcatgcat tgcgtaggac aaatcaggag acattggaac
gactctgttc ctcaacagga ggcatttcgt gaagagtttg ttaatccttt ggatgagaag attctttatc
caggaaagaa agaactgccc gatgaacgaa ctttacgtaa gaaggagtgg ctgaagagaa gaccccgaac
actccctgtt gaatagaaca cgaacactgt aaatagaata aaagaaaact tggatagtag aacttcaatg
tagtgtttct attgtcttac gcggctcttt agattgcaat ccccagaatg gaatcgtcca tctttctcaa
cccactcaaa gataatctac cagacatacc tacgccctcc atcccagcac cacgtcgcga tcacccctaa
aacttcaata attgaacacg tactgatttc caaaccttct tcttcttcct atctataaga atgcaattca
actggaatat taaaactgtg gcaagtattt tgtccgctct cacactagca caagcaagtg atcaggaggc
tattgctcca gaggactctc atgtcgtcaa attgactgaa gccacttttg agtctttcat caccagtaat
cctcacgttt tggcagagtt ttttgcccct tggtgtggtc actgtaagaa gttgggccct gaacttgttt

ctgctgccga gatcttaaag gacaatgagc aggttaagat tgctcaaatt gattgtacgg aggagaagga
attatgtcaa ggctacgaaa ttaaagggta tcctactttg aaggtgttcc atggtgaggt tgaggtccca
agtgactatc aaggtcaaag acagagccaa agcattgtca gctatatgct aaagcagagt ttacccccctg
tcagtgaaat caatgcaacc aaagatttag acgacacaat cgccgaggca aaagagcccg tgattgtgca
agtactaccg gaagatgcat ccaacttgga atctaacacc acattttacg gagttgccgg tactctcaga
gagaaattca cttttgtctc cactaagtct actgattatg ccaaaaaata cactagcgac tcgactcctg
cctatttgct tgtcagacct ggcgaggaac ctagtgttta ctctggtgag gagttagatg agactcattt
ggtgcactgg attgatattg agtccaaacc tctatttgga gacattgacg gatccacctt caaatcatat
gctgaagcta acatcccttt agcctactat ttctatgaga acgaagaaca acgtgctgct gctgccgata
ttattaaacc ttttgctaaa gagcaacgtg gcaaaattaa ctttgttggc ttagatgccg ttaaattcgg
taagcatgcc aagaacttaa acatggatga agagaaactc cctctatttg tcattcatga tttggtgagc
aacaagaagt ttggagttcc tcaagaccaa gaattgacga acaaagatgt gaccgagctg attgagaaat
tcatcgcagg agaggcagaa ccaattgtga aatcagagcc aattccagaa attcaagaag agaaagtctt
caagctagtc ggaaaggccc acgatgaagt tgtcttcgat gaatctaaag atgttctagt caagtactac
gccccttggt gtggtcactg taagagaatg gctcctgctt atgaggaatt ggctactctt tacgccaatg
atgaggatgc ctcttcaaag gttgtgattg caaaacttga tcacactttg aacgatgtcg acaacgttga
tattcaaggt tatcctactt tgatccttta tccagctggt gataaatcca atcctcaact gtatgatgga
tctcgtgacc tagaatcatt ggctgagttt gtaaaggaga gaggaaccca caaagtggat gccctagcac
tcagaccagt cgaggaagaa aaggaagctg aagaagaagc tgaaagtgag gcagacgctc acgacgagct
ttaagaatgt ccggggtttc tatatttact gtaactaggt tttattcttt tgtatagtcc aacaaaccat
atggattagc atcacttcgc gcaatacgtt ctagttgccc acttgctaaa accaatttag caatgattca
acagcaatta ttatcaccat gactaactgg aaagcgatat tgactcccgc tcaataccaa gtcctccgtt
tgggcggaac agaaagaccg tataccggac agtatgtgaa cttcaagaaa aatggaacct acttgtgtag
tgggtgtcaa actccgcttt acaaaagtgg cacaaaattt gattcatctt gtggttggcc tgcattctat
gaagcattac ctggagcagt taaacgaata gaagacaatt cgcttggaat gcgaagaata gaaatcagat
gctccaaatg tgatggacat cttggccatg tttttgaggg tgagggattt gacactccaa cagattccag
acattgtgtc aacagcatca gcctaaaatt tcaaggtgaa gaagagaact aagcgggggg aatgactaaa
taaataaagt aaaaattttg tttattgacg ccttttataa tttgtttcct tggtctttaa gaaaactatc
attctggaga atttctagaa gatacccatc agggtccttg aagaaggcaa gcttgtcgct aatctctccc
aatttaaaga cccattctac atcattctca agctctttgg caagtttggc gacatccttt gttttaatgc
ctgtatggcc atatccaaaa ggctcagtgg tgttggtatg gtaacctttg aagctggagt cactttctgt
accccagtta tgggtcaact caataataga ttccctctca gcttgtggac gaggtttttc ctggttttcc
acgtattgat cgtcatgctc gtagcccaga aaatagaggg tgaacttagc ttccggaaat tctctagtag
aataaagttt catacctaaa actt

NCBI Accession number: CCA36295.1
Gene designation (GenDB): PP7435 Chr1-0128
EC number: /
Gene product: putative secreted protein

3000bp upstream reagion / CDS (colored gray) / 1000bp downstream region

ctggtcaggc tcggtttcat cctgaggttc attcgttggc ttctgcgacg atattgagtt cttattccaa
atgtatatca tcacctcatt ctttcttaca ctcagggtaa tgccacaaat gtcgtcttga aactcttcaa
agtcagcatc tgaatagttt gagaaatagg tacctccaca aagtctcaat atcaatcgtt cccagatcaa
agaaaatctc tccatggttg aggcatcgca ttgagctctc ctataacggg acttgattaa ccatcttcca
ccagccttat tagcaggatc ctcccaaact ggcttgattc cacttttaaa caaaaagagc tctgtgccaa
ctggaaggtc agatactttc ctcaaattgc tcaatgattg ccacaactgt tcgcatgagt caatattggt
agtagtagta gcaccgccat tcgccactga tggaaactct acttgtttaa cggtctgtaa atattggtca
agtgtgtttt gccagtcatt tctggaagcc tccgaacttg tctctttatt ttcctgttca gaaggctcag
ctacactacc agatgtagtt acgttttcgg tctccttgcc cctcacatga tgccaaaaag tccaagagta
ttgcattgga tgaggagcgg caacttgttt cgtcgtccga gatccattga tatcctcaaa gaacttggac
attgcagttt tgggaaacct ggagaaatat gttcgaagct tggtgtcgga tggacgattt tcaagttgga
gcgcgcatgc agagaaaggt caatgctagt agcgctaggt atgccatttg ttcgccttgt aggatcaaag
agttgagagt tggagccggg taatactgac cattaattcc attaattcat acctatatcg tttatatacc
aaacttaagc cagtctagat cttccgcaat gtttctattg ttcggggttg gacctgcagt gccagggtct
ttggaatttc gtctgggacc gcatgttttg tcctctggga agccctcctg tttgacatca aaaccagtaa
caacgtgctg agtttgaaga ctaagactgg aaggacgtct ggcagtctct ggaaatctct gctgaaacga
gtttttcaga ataggggaag atggtatacc gcgcttcatt tggaattctt ggtgcaaaga gctggagatg

gagttggaga acccatgccc atagtcgttg ctgggagata taatacctag gtctgagtta attggcgagg
cagcgttcgt aaaatacgaa gcaacagcca ctggttgctg tgaagtaaaa ctggaaacag aggaaggtga
tgccggtata gatactccag gattattacc tgaaagactg atttgtgact ttctgcgctt gaatcctgct
gaatgcgggg agaacgaagt atagggagag tcttgttgtt gtggcatcat gggaacggga atgtttttat
tgctgggacc cctggaaggc ttgggcaaaa tgggaacatt cttatgcttt tgagcgttgc tgttgagact
gttgcctctc attagataat aggaatcttg tgaacctccg tagggagaaa atgatacggc tgttgttgga
agagaattat tataggggta aggataaggc tgttgctgtt gttgttgtat ttgcttttga gtttgagtgc
tatgagaagg ttgaatatgt tcatgttgtt gtggctggaa ttgaacttgt tgttgttgct gttgcatctt
gcagagcccg tgtgttaaag aagattgatg ttgttgtctc ttggaatttc gtttcctaat gacgtccgtt
ttcaaagaca atggcctcac agtaccatga agtttttgga aaagaccaca agcattgcac aacgtattgc
cacttccatc tcttctccaa agaggagtct tttcagtctt acaatttgag cattgagttc ttgttatact
ttcgtcattg tcagctttgg gggatgatga gtcactcata tctttctgaa cagtcactgt tgttgtgata
ctagtagtct cggaagtagg tttcgctaac tgtagtttat gagaagctcc gtcactcaag aattcttgca
aagacttgtc ttggataata tcactaaaat ctcctcccat accatcttcc tcaccatcat cattgctatc
gaacttccac atcgagttta gatccaaatt gttggagtca tgcaagggac tatcattcat agatgagggc
gtaaccatca ttgaaggttt tggttttcga agttcggatt taaccttgga atctgagtaa ctagacggag
ccccagaatt attttccaac gaatctggag aataatacag gccaggagca cttgaaataa tctccgtcaa
atcagcagag tcatctagtt gcaattcgga aaaatgtagt ttgccagaaa ggctgtcttg ttttgtcatc
tgggttttct tggattcatt tcttaatgtt ctgttttcaa ctcgggggta taaatctagt agctttttag
ccgaagaata tacccccccat atggacggcg aatcttcaag atccttcaaa gatcctgtat tatttcttct
gtttccttcc atctaaatga ggccttaggg catataatta gagaaaggac aggaccagtt gagcttcaat
gctagggtgg tgtaaagtaa aaattttgac cctcagggat ctccaagtca tgttaggaga agtgtttttt
tggaacagac acgaccgggg tatttgtggg ccaagactaa gcaaacggca aacgaaatgt gggtttgaaa
tggagaagcg ctgcttgaaa tgagcagccg tagaaggttg gatggcacta catgtatgaa ctagaaggcg
ttgaattttc tagatgtaca gagggagatc ttgattgcat gcgaagcatc gtcgtcaggc tccccttta
tttattttcc caccgttttt tgtccccttt ctcgctaaat ctcttccctc tcgacagcac atgaagttac
tatccttggc acttctggtg tctttggtgt ctgcggatac tttctacact ccaaaggatg atgtaatcca
gttgaatgct tataatttca aggatgtcgt tttcaactca aactactctt cggttgtgga attttatgct
ccttggtgtg gccattgtca gaacttgaaa aatcccttca agaaggcagc agccgtttcc aaagactacc
ttcaggtggc cgcaattgac tgcgatgctg ccgaaaacaa gaagctttgc tctgattacc gtattcaagg
atttcctacg atcatggttt tcagaccccc caagtttgat cccacatcaa gcaccaatag aagatctggt
gctcatgcta atgaggtata cagtggtgct agagatacta agtcaattgt tgagtttgga gtgtctcgta
taaagaacta cgtcaagcga gtgtcaccca ataatattaa ccaaaccctg ggaaattctg agaagaccca
gcttctactt gtgacagata aggccaaacc tagtgctctg ataaagtcaa ttgccttaga ctttttgaat
gacatagagt ctttttacta tcctttcaac gacaaaacga agaaagcatt gactacacga ttagaggagt
atcaacaatc gttttctgga gagagcataa cttcgccttc gattttggtg ctacatgaaa atgaaatcca
cattttgat gggaaactgg ataagctgag cattagcaag tttttagccg agttttcaac tcccctggaa
ggacctctca gtaagagagg taagtttcta gagcacattc gcaggggaat taaacccgga agaaaagcga
agaagggcaa gaagggcaag caaaccaaaa atcacgacga attatagata catacgaagt tttttatttt
ttttttgct tctttcattc atatatttac aagttgaata tttttacaca attgtagaag tcaccctaac
cctggatgga ttcttccaac tgaagtttc tggaatggcg atcaaacagg ttgctaagct cgttgaagtt
caacgatgtt gggttgacat ctaacgtctt tttgtctttg taggtcacct caatgatagg ttttttgttt
gactggctgt tcaagacttc gtagtttatt gtgcatgatt tcttcaagga actgggcaaa gctgcaagaa
aaagtcttgc gcctttagct tctggaccaa aagggttgaa tttgaccgat actttggtaa agtacttggt
aatcatggag atgcttctgc aagaattggg tcgtgatgag agagggtgga ttttttcgaa gagtatccga
tattgagttt cgcccactac catgatatat tttttttct gacttacctc cctctaaata agaatcactt
cacccataga tatctacatg aaggctcaga aattttccaa ggcttcgtta ctggatgatg aggatgacta
ctcaactcag cttgacagaa ttcttgctcg gaggaagaag agtcagaagg aagactctgg agatctagat
acgatttctt ttggctcctt atcaaaagct caagcaagat tacaaactga ggagagagaa tccaagaaga
aggtaaaaaa agtcaagaaa ccgaaagtat tgaagaaaca ggaagaagat tttcaacctc ccagtgacat
ttctgatgac ggtgagtttt ttgaagagcc atctcaagac catcaaagtc gaacttcaga taggcattcc
aaggaggaac ctagatctaa aagtaagcac gcaccctcag agtctagtgc taagaagaga gtgtcaaagg
tgagggagat tcctggactc ttgaacggtt ctggctcaag ttcattgtac caggatgtca gattcga

NCBI Accession number: CCA36622.1
Gene designation (GenDB): PP7435 Chr1-0470
EC number: 5.3.4.1
Gene product: K09584 protein disulfide isomerase family A, member 6

3000bp upstream reagion / CDS (colored gray) / 1000bp downstream region

cccatgatct ggatccatta ttccagacta ctgactgtcc ttcttcattt gtgaatagaa aaaatgtgga
attattcatc ttgaatgaac tagtgtcaga aacggtgaaa ttagtatctg ccaacgataa gacattggaa
tctctcaaaa ggaagatgga tgaattgctg atgacggtaa tgtcttcgcc aactgttgat gttagactat
ccacggttcc attgaattgc aaaagcctca ttccctcagt gttaaaagaa ttatcggtgt tgttgatatt
ttgaatagca acttccaaga ggttaccttc gtcattagca gacagtttac caccaacata aagtttacca
tcctgaaaaa ctaacccatt cactgtagca ttctcaagta aggaggatcc aagtagattt atggaggagt
ttccgacaaa agtattgttt gataacctta aagtgtcaaa atctgtaaag ttgttgaggg aatactcaaa
aataccattt aaaggcactg tcaaattcaa ggatctggta ctattagttg agttgatagc cgcatctgca
tcaatgtaca gtatgtctac gtacaactga tatgcggcaa attccaactc ggtttctgaa tttcccggta
tagttgattc gacctctaag gtgacatacg gatgatcttc caaggttgca ttgatacgac cgttgtaaag
aatgtcatac ttgtcaaaat cattgttttg atatattaag atggatgcga gttgctcatt gtcgattcca
ttgtgaagtg tagcgttgac aattcctctg gtgctacatg agtcggttgc taaacagcca ggtgtgtaaa
cataaaattc gtaatcacca ggaaaactaa tatttggata gaatgtcaca cccactgtgg agtctggaat
ttcaccattg atcaatgata tggtcgtggt aatggaccca tcttcgtggg gggtccagga tccaggtgcc
tctacggata cactggaaga atgaaactca aaatcatcac aggtaggttc gttgaaactt tcgttaccat
aaacgtagat tcctgattgg tagatttgaa gtcctttaag agcaccgatt tcattgtaag cggacaatac
ttggatggtt acttgagtaa cgtcgagttc attgtggaac tcgaattctt ggtaatctcc agaccacgtt
aggtagcctg ctgttctggt tttgtcatta aagacggtgt cttcgtatac atctgaagag tttgcgttct
gaagagcgtc ttccaactgt gattttgaga ggagtggaca ccacgcgtca caatattcat attcgttggt
ttgtggattc acatatgtca agttcatgat actgtttgaa gggccagtta tgagcctaaa gtactcgaca
ccataaccct ctgcagaagt tgattggaat agacggagct tggatggaat tattgtgtgg tggagcatag
ctagccaagt cccttgggag cctgaaagaa tccattcatt gtcttgatcg ggacaagtga ttgacttggg
gtctggcgaa agggcttcgt tggtacctgc tgttgacacg tcggcgatat taaaacttgt tagttgatct
ggactgacat catatatggt caactcagtt gagtttgatg aattggtgga cacattatga tgaagtaagg
attgattccc caaggagtaa aagtttccac cgaaaatcaa gttctcatca tcaatgctaa caattgcgtt
cactgatgcg ttttcattaa atccttggaa gggaatggac tcgaaggagg aagattcata gttgaagcag
accaaaccaa aaatagaact gttcaacgtc aaatttcccc caaaatatat caacttggac gtttggttga
caaacacggt ttgaatagat ccttgaactt gattactcca attttcggtc aggttcgata ttgaaacctg
agaagcttcc tcgttcactt ctaggatcag aggagcttca ttttcttcga tggggattac caacacgaag
ttattcccca atgagtagga ctgttttatt gtgccgttga atgtggaaat gaacgataac gactcatctt
gacctagttt gaaaatgcct aaactctttg aatcttcgtt cctaacagat gtcctgtact gattggtatt
attgggtgta gaaagtccgt tgaaccgtcc cagcaaagag atatcatttg aaaagttgat cgcctggcaa
gagacgatta ttgagaccag taggaacaac aacgccgcca tagtttcaat ctgaagttga cttgtgtgat
ttggtgttgg ggggattgag ggtcaaaaac gggggtcaga aacagaataa aaaactagct tggaactcga
agtaggagaa gagaagtggt gaaaaaagaa gaggggacgt gacgttaaag acctgaactc tgcgtgggtg
caatgtggtt ggaaggtgaa gtgcaaggta tggtttttt ttcctgccag gaatcaaaaa tatcattctc
atttctgagg aaaggcatgg cagaatattt acctgttgcc agatctggat atgaagatta cataataaac
acgtagacag tgtagattct gcgaggcgtg ccagacgttg aggaggacaa gttggggctc cgtaaggtgt
gcgtccatat gcggacagag tgtacctatt ggctgattgg caacacgtta cattgttgca ttcctgattg
gcgttggtgt gctcttgttg aaacagggat gcatacatat tattatcaaa catggactga ggatcgagga
tctgcaggtc ggacagccac ccgacatcgt ttaagcttac gtaagatttg aaggtcccaa catcttgaca
gatctctcaa ctttgcttaa tcttacataa ccccaagagc caaccctcac caggttattt ttttgccttc
cgattgtctt ctcctccccc cgtttgctgt ccaaccactt gaaagggtct gtctaggata atgaaaatat
taagtgcatt gcttcttctt tttacgttgg cctttgctga ggtcattgag ctgaccaaca agaactttga
tgacgtggtt ctaaagtccg gaaagtacac cttagtgaag ttttatgccg attggtgttc gcattgcaag
cgaatgaatc cagagtatga aaagctggcc gaagaactga agccaaagag tgatctgatc cagattgccg
ccattgatgc taacaaatac tcaaagtaca tgaaggtgta cgatattgat ggatttccga cgatgaaatt
gttcacaccc aaggacatat ctcatccgat cgaattttct ggatcaagag acagtgaaag ctttttgaac
tttttggagt caactactgg tttgaagttg aagaagaagg cggaagtaaa tgagccttcg ttagttcaat
caattgatga ttcaacaata gatgaccttg ttgggaagga caggtttatt gcagttactg cttcgtggtg
tggatattgc aaaagattgc atcctgaatg ggagaagtta gccaaagctt ttggcaatga cgatattgtc
atcggaaacg ttgttaccga tgttgtggaa ggtgagaata ttaaggcgaa gtataaagtt caatctttcc
cgactatcct gtacttcaca gcaggctcag atgaaccaat aagatatgaa tctccagata gaactgttga
aggtttggtt aaatttgtca atgaacaagc tggcttattt cgtgatccag atggaacttt gaatttcaac
gccggtctaa ttccaggagt gagtgataaa cttaccaatt acattaagga aaaagaccaa agtttattgg
agtcaacgtt agacttgcta agcaaccatg aacatatcaa ggacaaattc agtgtcaaat accacaagaa
ggtcatagaa aagttgttga agggagagaa tgaattcctc aacaatgaag ttgagaggct atcaaaaatg
ctgaatacaa agctatcggc aaacaattca gactctgtga ttaaaagact gaatatcttg cgcaatttta

ttgaggccaa aactgagtca aaaccccagt tattacacca agagctataa gtaaatgaaa aagatacctt
tctagataag taaataaata acgctataac tctaccactg tagaggcttt gtagatttct ggaccaatct
cacatgctcc ggccacactt tgccgccccc taggatggtt tcttctatga taggtttgac attaaagggg
ttacatcttg ccaaccagat attagcccca gacttcaagt aaatgagaac atttgctgca tacttgtgac
ctccgacatg attgataaaa tgaacttgca ctccgcctgg tctatcatct ccgagatctc tgtagagatc
gtgatcacga aggtgaatgc acatctcctt cttaatgata ggtgcggtga ttccgcaacg tttatccctt
gttttatgcg agcacaagaa cacataggac ttgtttgtgt ctttaacgat ctcataacct ttgatttcgg
cgggaatatc tattggttcc acttgggtag tttcgataga gaaatttgaa aaaagaatgg gcactagctc
gtccaatacg gcatcaacat cctccataca caactgcttg acccatacaa agtacggaag aattaaaata
tccccagctg tctgttcggc aattcttgga tccttggcat tagactgggg tatggatgat gtggagactt
taaccgcttt tggaaccagc ttggtagatg aagcccaatg agagattgat tttaaaatcg tgccatcaac
atcagtggcg tcacgcttcc aatcggactt acccgttgga atgataatgt gtaaagagta cggtttagtt
gatttccaaa ggagatcgtt tttggagcct gggtcatctt tgtaactgtt tggatacctt gtgaagcact
cagcggccga acaagtagcg gggcactgtg aaagaggaac tatttctgag atctcctcag ccgcctgggc
ttcgccggat ttgcccccca agaacttcga caaaaaagcc atactgattg gacttaagga tgaaaagcgg

NCBI Accession number: CCA40358.1
Gene designation (GenDB): PP7435 Chr4-0183
EC number: 5.3.4.1
Gene product: K09584 protein disulfide isomerase family A, member 6

3000bp upstream reagion / CDS (colored gray) / 1000bp downstream region

ttttcaaatc tccaacttta atcgaatcga agattgtttg gtatcgattt ttcacttttg aaaactggct
gtacaatttc tggctgggta gttttcttgt gactagaagc tgcaacggag ccaataatag tattatacat
tgcatatttt ttgtgctggt ttttaaacat gtgctgagag cacttgtcag atttttatga gcttcctcaa
acttattgtg aaaacagttg aggacagcat tgaggtagag gtatgcagta ctgtgagatt tggggacttt
tcgaaggctc ggtaaagtag atgatgctga aagtaagact ttctgtaatg agctggccag gtcgaacttg
ttcatggtga agtaaagttt gaacagtatt ccagcaaaat agtagactgc tattttcttt gaagttgcaa
gatcagggtt cttgtcactt aaacaaattt tgaacgacct attgaataca tcagcaattt tctccaaagg
tgcccttgaa tgctcgggag tggattgatc acctttgtta ctattataca aatttccgac atatttcagc
tcagtagaaa ggaggatcag aagaccgata atccagtttg tttcattatc agcgatacgg ttgagctggg
aaactagatc aagacttgtt tcaaacgcct ttttgaggtc tccatgtttg gtaattgatt tgagaagcga
tagatatgtt gagaccaaat cggtccattt ttggtcaaca atccggtttt gttgcggtaa tgcgaagttt
gatctttctt gaagaaggtt atttggaatg cttatggaaa ggagattact cagtgatgag ccatcttgaa
accttacagc ttgagacact tgttggaagt actgcagctc tgacattgaa ttggaggtcc aactggagag
agatagattt gatggtagga aatgacggtg ttgggatgtt gttcccttgg gacagaatga tgtttaacga
tgcaattttg atacaatata aggatttgat agacttaaat ccagtcaatt acttggagat ctatccgttt
tacggtgtaa ggttaaaggt gcaaaatatc atttattata tagcagttgc attttaaaga tttacaaatg
gtctcgcttg aaacaagtaa agttgacgat gcaggcaact taatatccag tggccaaaat aaaactgaaa
tgaaatccat ttgagttgat taactattca gttttgatgg acaaatcatg ttttgtcaag cacagccctt
ttctttaacg tctgtacttc aacctcatct ccagatatat tataaaagtg aaatataaat attttaattc
gctatttaca tttaacccaa ttcggaaccc agaatcctta agatcctatc aaaaagatcg cacaccattc
tgtaccgtaa acaacacaac attggtatat ggccgtccca agtacaccac caagaggcat acaatacaag
aaatcgatac ccattctggc taacaatgat gcactcctca acacccaacc agtaacccct gtcacacaac
gacaagttag cgagtcaaca ggactgttga agttttccat aggcaaaaat gcggcatctc cggtgagaga
tggtctcccc atgtaccttc cattgaaatc tgctgtacat gaggacaagg aaaacgatat cctgtttcaa
ctggctaaga agcaacgaca ggtggtggac ctggctcaaa cattaaagat agccaaacag gagcttgcgg
aactggaacg aaagtaccaa tccgagttcg attctagaaa cccgtcccct gttcaaaaga aggatcagaa
tccaccaacg gcggctaaga atccattgac tgaactaaaa cggatcatgg agaaaaacaa tgaaaaaaaa
ccatctgcac tcaagagcag gaaaagtttg ttcaatttgc aacaagactt tgaggataac agagcaatcc
aaaaaatccg cacggatttt actactctct cttcgaactt tttaaatgat tctaagaaat ttcagcaggg
ttttgtgaca aaagccaata gtttctttaa caactttaac acctcaggca gcaactccgt agctagcgac
caaaattctt tggagcgcaa agcatctgga agttgcctgt atgatgttcg aaagcatgac gcaaaccaga
cattgatgag ctctgaagat gaagatgaag aagattatgg cggaaaagca taccgactag acttgtagtc
aaatagcata caaacagcat caagaccatt aaatactcgc ctcaactaaa aactcgtctt ccttgattga
ctttactttt tcgattagat caccaagcgg tgtatcaaga gacgggtact gtctacgcta ttcacggtta
gcgtcaacaa aatgtaaaaa aaaagaaaaa aagactgtaa agcccaactt acttttgcca aaaagccaaa

caactcagtt ctgctcatct cttccactgg cttagcctca atgtcaatgg gaatgtcatc cttgaccacc
tctgtgtcct tcttcttgga ttgctttttc gatatcgttc gtcggctaga gggcttgaag gtggtaaaaa
acattgaata caatgtagcc cccacaacga ggcccacaaa cccataaaag aatttcttct ttacttggtg
taccattgca atggaagagg ttgatatatt taaagaggtc caaggacccg tggaagcaaa ataccttcaa
ttccctaccg cgattgactt tctgatgttt gattagatgg acaatcagca aatgtacagt aaagcgaaat
gaaacgaaat tgtcacagcg tttggatgga cttaatcaag gacagctagc aatgccaaga cagtaactaa
tacactgcat ttctcttaac agtactcaac ttccacactt cgtcaggttc tggttcctgg cccatgcgcg
actcatcccc cgcctgatga aatagataaa tttttgcatc tatgaacttt tcactcaaaa atgaagtcgt
tactgctact tctattagct ctctgtcaaa ctgtctttgg tacgatatgg gaactcacgg acaaaaactt
tgaaaaaaaa gcttttgggc agcaaggtat gtactccttc gtgtacatct actcccctta ttgcaattat
tgtaacgaga tgacacctca attcgctgca ctagctgatt tatacgacga caccaagttg caactattcc
agattaatgg gtatgccaac aagagggtct caaagaaata cgaggtagtc gggtttcctg ttttgaagat
tttttcatct gatggaaccg acatggggtc ttataccggt gtcagaggca ctcagaactt cattgattat
attcatgaag ttacaggggt tactccaagc tttccagact cttatgtcaa acagcctact tcagaggagc
tagaagatat cataaaagac actaagaggg atattctcgt tgctttttca cagccttggc ttagaggctg
ggagttccct tacacaaact tttacgaatc tctggcaaga tattatgctg aagagcttga tgactctttg
actaccttcg tcagaataga cgtctccgac gccaagaatg ctgagattgt ttcgaagttt caagtttcaa
aaactccttc agttttccat tttgcctctt atcgtgaata ttatgatcaa acacataagc ttttcaaaga
agatttgggc cccgtggaaa ttgttcaatt gctagagggt gctgaggact tagttactga gagcattagg
ctgtcaacat tatcagactc tcgccacaaa gatgttgagg agtctgggga agcagataat tcctttgaag
agtatgccca actaagggaa ttgtagatag aagtcttccc atttgtgaag gcaagataga aaccaatcac
atacatttgg tgtagactag aatggatcta tctggttatg taatcacgat acctttcgga attgtctcgt
gtagcaggcg cgccccttcg catctcaaat caacagttgg caagtgtcac tttctaagtt agcactctac
tttaagaaca aaatgaacat ctttagaatc ctaggtaagt ttccgctctg ctcaattgaa gaatcaaacc
caattttcta accacgcagg tgatatctct catttgataa gcatcttgat cctaatacat gctgtatcaa
caacaaagtc tgctcaaggt atctcattaa aaactcaagc agcttatgtc gtagtctttg tgagcaggta
tcttgacctc tttttcagat atgtgtcact ttacaacacg ctgatgaaac tctttttcat tgtctcttcc
atttacgttg tgtttttgtt aaatcggtac aaaaagaaca gtgaaaagac caatgatatt ttcccagtat
ggtatctctt tggaggagcg ttcgttttcg ctttgatctt caactacgaa tttagtattg tggagatttt
gtggagcttc agtctttggt tagaaagtgt tgctatcatt ccgcaattag tggttcttca gagaacggga
gaagctcaat tactaaccac acattacata tttgcacttg gattatacag agcgttatac atccccaatt
ggatctaccg ttactacaca gagggtcgca tggacaagat tgcagtcatt tctggtatct tacagactat
tgtctattct gacttttct acatttacta caagaaagta gttaagaact tgaaattcaa gctgccagaa
tgaacatgga aatatagaaa aacaaaagat agtgtatcaa taagtatata tgtgttgcag gaataaaaat
cagtccttct tcaaaggata atattcagtc ctgacgccat ctttag

NCBI Accession number: CCA36419.1
Gene designation (GenDB): PP7435_Chr1-0258
Gene name: *OCH1*
EC number: 2.4.1.232
Gene product: K05528 alpha 1,6-mannosyltransferase

1000bp upstream reagion / CDS (colored gray) / 1000bp downstream region

ccgctaaaag acccggaaaa ccgagagaac tctggattag cagtctgaaa aagaatcttc actctgtcta
gtggagcaat taatgtctta gcggcacttc ctgctactcc gccagctact cctgaataga tcacatactg
caaagactgc ttgtcgatga ccttggggtt atttagcttc aagggcaatt tttgggacat tttggacaca
ggagactcag aaacagacac agagcgttct gagtcctggt gctcctgacg taggcctaga acaggaatta
ttggctttat ttgtttgtcc atttcatagg cttggggtaa tagatagatg acagagaaat agagaagacc
taatattttt tgttcatggc aaatcgcggg ttcgcggtcg ggtcacacac ggagaagtaa tgagaagagc
tggtaatctg gggtaaaagg gttcaaaaga aggtcgcctg gtagggatgc aatacaaggt tgtcttggag
tttacattga ccagatgatt tggctttttc tctgttcaat tcacattttt cagcgagaat cggattgacg
gagaaatggc ggggtgtggg gtggatagat ggcagaaatg ctcgcaatca ccgcgaaaga aagactttat
ggaatagaac tactgggtgg tgtaaggatt acatagctag tccaatggag tccgttggaa aggtaagaag
aagctaaaac cggctaagta actagggaag aatgatcaga ctttgatttg atgaggtctg aaaatactct
gctgcttttt cagttgcttt ttccctgcaa cctatcattt tccttttcat aagcctgcct tttctgtttt
cacttatatg agttccgccg agacttcccc aaattctctc ctggaacatt ctctatcgct ctccttccaa
gttgcgcccc ctggcactgc ctagtaatat taccacgcga cttatattca gttccacaat ttccagtgtt
cgtagcaaat atcatcagcc atggcgaagg cagatggcag tttgctctac tataatcctc acaatccacc

cagaaggtat tacttctaca tggctatatt cgccgtttct gtcatttgcg ttttgtacgg accctcacaa
caattatcat ctccaaaaat agactatgat ccattgacgc tccgatcact tgatttgaag actttggaag
ctccttcaca gttgagtcca ggcaccgtag aagataatct tcgaagacaa ttggagtttc attttcctta
ccgcagttac gaacctttc cccaacatat ttggcaaacg tggaaagttt ctccctctga tagttccttt
ccgaaaaact tcaaagactt aggtgaaagt tggctgcaaa ggtccccaaa ttatgatcat tttgtgatac
ccgatgatgc agcatgggaa cttattcacc atgaatacga acgtgtacca gaagtcttgg aagctttcca
cctgctacca gagcccattc taaaggccga ttttttcagg tatttgattc tttttgcccg tggaggactg
tatgctgaca tggacactat gttattaaaa ccaatagaat cgtggctgac tttcaatgaa actattggtg
gagtaaaaaa caatgctggg ttggtcattg gtattgaggc tgatcctgat agacctgatt ggcacgactg
gtatgctaga aggatacaat tttgccaatg ggcaattcag tccaaacgag gacacccagc actgcgtgaa
ctgattgtaa gagttgtcag cacgacttta cggaaagaga aaagcggtta cttgaacatg gtggaaggaa
aggatcgtgg aagtgatgtg atggactgga cgggtccagg aatatttaca gacactctat ttgattatat
gactaatgtc aatacaacag gccactcagg ccaaggaatt ggagctggct cagcgtatta caatgcctta
tcgttggaag aacgtgatgc cctctctgcc cgcccgaacg gagagatgtt aaaagagaaa gtcccaggta
aatatgcaca gcaggttgtt ttatgggaac aatttaccaa cctgcgctcc cccaaattaa tcgacgatat
tcttattctt ccgatcacca gcttcagtcc agggattggc cacagtggag ctggagattt gaaccatcac
cttgcatata ttaggcatac atttgaagga agttggaagg actaaagaaa gctagagtaa aatagatata
gcgagattag agaatgaata ccttcttcta agcgatcgtc cgtcatcata gaatatcatg gactgtatag
ttttttttt gtacatataa tgattaaacg gtcatccaac atctcgttga cagatctctc agtacgcgaa
atccctgact atcaaagcaa gaaccgatga agaaaaaaac aacagtaacc caaacaccac aacaaacact
ttatcttctc ccccccaac accaatcatc aaagagatgt cggaaccaaa caccaagaag caaaaactaa
ccccatataa aaacatcctg gtagataatg ctggtaaccc gctctccttc catattctgg gctacttcac
gaagtctgac cggtctcagt tgatcaacat gatcctcgaa atgggtggca agatcgttcc agacctgcct
cctctggtag atggagtgtt gtttttgaca ggggattaca agtctattga tgaagatacc ctaaagcaac
tgggggacgt tccaatatac agagactcct tcatctacca gtgttttgtg cacaagacat ctcttcccat
tgacactttc cgaattgaca agaacgtcga cttggctcaa gatttgatca atagggccct tcaagagtct
gtggatcatg tcacttctgc cagcacagct gcagctgctg ctgttgttgt cgctaccaac ggcctgtctt
ctaaaccaga cgctcgtact agcaaaatac agttcactcc cgaagaagat cgttttattc ttgactttgt
taggagaaat cctaaacgaa gaaacacaca tcaactgtac actgagctcg ctcagcacat gaaaaaccat
acgaatcatt ctatccgcca cagatttcgt cgtaatcttt ccgctcaact tgattgggtt tatgatatcg
atccattgac caaccaacct cgaaaagatg aaaacgggaa ctacatcaag gtacaagatc ttcca

---

NCBI Accession number: CCA39540.1
Gene designation (GenDB): PP7435_Chr3-0582
Gene name: *HOC1*
EC number: 2.4.1.-
Gene product: K05534 mannan polymerase II complex HOC1 subunit

1000bp upstream reagion / CDS (colored gray) / 1000bp downstream region

acccttccca cttagtaagt tggacgttat gcataatttt gggatgtacg accaaatatt gcctaaagta
gaataagtaa tatttcttta taaatagtaa tgataatttt ccatttaatg gggtactcta gactgtcatc
aaaaataaat atatggtgtt gctatgtagg ccgtaacacg gctcttcacc tgtatggagc atcggcatct
aaagcgtact ggatttcctg ttcccatttg tcaccattag caagcaattt ttccttgtta tacaacagct
cctctctagt ttcggtggca tactcaacgt taggggtttc aacaatggca tcaactggac aactttcctg
acaataaccg cagtagatac acttggtcat gtcaatgtcg tacttgtaag ttcttcggga accgtcaata
cgttcttcgg cttcgatggt gatagcctga gcaggacaaa ctgcttcgca taacttacag gcaatgcaac
gttcttctcc ggatggatat cttctcaaag cgtgttctcc tctgaaacgg ggagaaactg gacctttttc
aaaagggtag tagattgtat atggggcacg aaagtacatt tccaaagtga tataaaggcc tctaaagatc
tcagacagta ggtaccattt ggtggccttg gaaagagcgc tttcactgct ttcttcccac gtctttggtc
ttgggagtct gaatccttct ggatgtgtgc cttggatcga accagaagtt aaagccaatg tgggtttctg
tgcaatgagt cccgtcgaaa acctgcaaac tggtattctg tttatggcac ggcaggaagg cttcaatgct
agtggtctaa acattgtgtt ataaatgtgt aggtagaaac ctataagttt gtgttggggt tcgataatca
aacttgcctt gaaggtaatt ggctaggtgc ttgttgagtg tcgcacgtct cggtattttt ttacttgtcg
attctctctt actgtccaag atgtacgaaa gatgtgaagt ttctttgttg gcgattttct tgcagcgcga
tcgcggtcgg tatgatttcc cttcagacca aaaacagata tagatattca attgaaactc gcccttccaa
caccctgaac accacccccc aatgcaacaa cagctgttta gaaaagtaat ttggctcacg gttggactca
taacggtgat attggtaata ataaagattt catccagcaa gtccaccgcc acagaccttc agaaagtatt
gaagaatgcc aatattttac cgcaggacgt gatcaattac aatagcagaa aagtaacgga tgaactcgct

tcaaagctgg acgagattca aaagaagtat ctctcgaagc aagatgatag aattagcaag ctcgaagctg
aacgggcaga tctactggaa caggttagat ttctaaggaa ccccccctgca ggatcaagtt taagagaaaa
attggcttat ctgtttcctt ataatgaaaa cggcaaattc cctgcttata tatggcaaac atggaagtat
ggcttgaatg acgatcggtt tggagaaaag ttcaaagaag gcgaaactca gtgggcttcg aagaatcctg
gtttcgttca tgagttgttt aacgatgata cttccggtgt gtttattcac catctgtata tcaatgttcc
agaagtgatc aaagcatacg agctgcttcc caacataatc ttgaaaatgg acttcttcag atatttggtt
ttatacgcca agggaggtgt ctatgcagac gttgatacta tgcctcttca gcctgtacca aactggattc
ctgaaaatgt ctccccaaaa agcatcggga tgatcattgg aatacaaaac gatgctaaca atccagattg
gaaaaagatt acgtacatcg tctacaattt tccaattggt gtattcaagc gaagcctggt cacccaatct
tgagagaatt aatcgccaaa atcacagagg atacactgca acgagccgag tcaaactcac tggaactagc
tgacattagc gaagaaggcg gcctgtctga taagaatttg tccattatgc aatggacggg tactggtatt
tttacagatg ccatatttac ctattttaat gactacattc aaagtagtat ctataccaaa gttacttgga
aagaattctc caaattgaga aagcccaagc ttgtcagtga tgtattggta ctgccgatta tcagcttctc
ggccggtgca ggtagtggaa aatcgactga actgaacgat cccttagcat tcgtacaaca ttattttgaa
agattacata acgacaacca ctaagggtca gaaccattta gattgtctgg atctatcatt atggccttgt
ttatagacaa agaattgtat cctggactga agggaagttt atagagtaat acccctctga caaccaactc
gaatgggtat tttgagataa tttcccatat attattgtct tcactgggag acttacttct tctttgtagt
tccggtatcg gcagcaaact tatggtggtc acatcttcca aacttctcgg tatagtagaa ctatttgtgg
ttgatacggt ctgggtatcg gtgataatgc ttgtaataat ggattttgtg ggtgtaactg aagctggtgt
attagtggta gtggtagtgg tggttgttac tgtgctatct gtaggttctg ggagtggttt ctgagccggg
tttaaaattt tgatctctgt ggtatcttcg tctggatttc gactaaaaaa tccgccttga aatacgagtg
gaacaggtaa acgattgatg gagcccagaa gcactgtgtc caaggaagat tcatcaaaag tgaaaggagt
gccatcttcg tttccatcgt taagatgaga acttttggca aaaaaatcaa tcagaacgtc atctatttct
atggcgacaa atccgggatt gaa

Appendix 4: Amino acid sequences of relevance for this thesis

NCBI Accession number: CCA40283.1

MQFNWNIKTV ASILSALTLA QASDQEAIAP EDSHVVKLTE ATFESFITSN PHVLAEFFAP WCGHCKKLGP
ELVSAAEILK DNEQVKIAQI DCTEEKELCQ GYEIKGYPTL KVFHGEVEVP SDYQGQRQSQ SIVSYMLKQS
LPPVSEINAT KDLDDTIAEA KEPVIVQVLP EDASNLESNT TFYGVAGTLR EKFTFVSTKS TDYAKKYTSD
STPAYLLVRP GEEPSVYSGE ELDETHLVHW IDIESKPLFG DIDGSTFKSY AEANIPLAYY FYENEEQRAA
AADIIKPFAK EQRGKINFVG LDAVKFGKHA KNLNMDEEKL PLFVIHDLVS NKKFGVPQDQ ELTNKDVTEL
IEKFIAGEAE PIVKSEPIPE IQEEKVFKLV GKAHDEVVFD ESKDVLVKYY APWCGHCKRM APAYEELATL
YANDEDASSK VVIAKLDHTL NDVDNVDIQG YPTLILYPAG DKSNPQLYDG SRDLESLAEF VKERGTHKVD
ALALRPVEEE KEAEEEAESE ADAHDEL

NCBI Accession number: CCA36295.1

MKLLSLALLV SLVSADTFYT PKDDVIQLNA YNFKDVVFNS NYSSVVEFYA PWCGHCQNLK NPFKKAAAVS
KDYLQVAAID CDAAENKKLC SDYRIQGFPT IMVFRPPKFD PTSSTNRRSG AHANEVYSGA RDTKSIVEFG
VSRIKNYVKR VSPNNINQTL GNSEKTQLLL VTDKAKPSAL IKSIALDFLN DIESFYYPFN DKTKKALTTR
LEEYQQSFSG ESITSPSILV LHENEIHIFD GKLDKLSISK FLAEFSTPLE GPLSKRGKFL EHIRRGIKPG
RKAKKGKKGK QTKNHDEL

NCBI Accession number: CCA36622.1

MKILSALLLL FTLAFAEVIE LTNKNFDDVV LKSGKYTLVK FYADWCSHCK RMNPEYEKLA EELKPKSDLI
QIAAIDANKY SKYMKVYDID GFPTMKLFTP KDISHPIEFS GSRDSESFLN FLESTTGLKL KKKAEVNEPS
LVQSIDDSTI DDLVGKDRFI AVTASWCGYC KRLHPEWEKL AKAFGNDDIV IGNVVTDVVE GENIKAKYKV
QSFPTILYFT AGSDEPIRYE SPDRTVEGLV KFVNEQAGLF RDPDGTLNFN AGLIPGVSDK LTNYIKEKDQ
SLLESTLDLL SNHEHIKDKF SVKYHKKVIE KLLKGENEFL NNEVERLSKM LNTKLSANNS DSVIKRLNIL
RNFIEAKTES KPQLLHQEL

NCBI Accession number: CCA40358.1

MKSLLLLLLA LCQTVFGTIW ELTDKNFEKK AFGQQGMYSF VYIYSPYCNY CNEMTPQFAA LADLYDDTKL
QLFQINGYAN KRVSKKYEVV GFPVLKIFSS DGTDMGSYTG VRGTQNFIDY IHEVTGVTPS FPDSYVKQPT
SEELEDIIKD TKRDILVAFS QPWLRGWEFP YTNFYESLAR YYAEELDDSL TTFVRIDVSD AKNAEIVSKF
QVSKTPSVFH FASYREYYDQ THKLFKEDLG PVEIVQLLEG AEDLVTESIR LSTLSDSRHK DVEESGEADN
SFEEYAQLRE L

NCBI Accession number: CCA36419.1

MAKADGSLLY YNPHNPPRRY YFYMAIFAVS VICVLYGPSQ QLSSPKIDYD PLTLRSLDLK TLEAPSQLSP
GTVEDNLRRQ LEFHFPYRSY EPFPQHIWQT WKVSPSDSSF PKNFKDLGES WLQRSPNYDH FVIPDDAAWE
LIHHEYERVP EVLEAFHLLP EPILKADFFR YLILFARGGL YADMDTMLLK PIESWLTFNE TIGGVKNNAG
LVIGIEADPD RPDWHDWYAR RIQFCQWAIQ SKRGHPALRE LIVRVVSTTL RKEKSGYLNM VEGKDRGSDV
MDWTGPGIFT DTLFDYMTNV NTTGHSGQGI GAGSAYYNAL SLEERDALSA RPNGEMLKEK VPGKYAQQVV
LWEQFTNLRS PKLIDDILIL PITSFSPGIG HSGAGDLNHH LAYIRHTFEG SWKD

NCBI Accession number: CCA39540.1

MYERCEVSLL AIFLQRDRGR KVIWLTVGLI TVILVIIKIS SSKSTATDLQ KVLKNANILP QDVINYNSRK
VTDELASKLD EIQKKYLSKQ DDRISKLEAE RADLLEQVRF LRNPPAGSSL REKLAYLFPY NENGKFPAYI
WQTWKYGLND DRFGEKFKEG ETQWASKNPG FVHELFNDDT SGVFIHHLYI NVPEVIKAYE LLPNIILKMD
FFRYLVLYAK GGVYADVDTM PLQPVPNWIP ENVSPKSIGM IIGIQNDANN PDWKKITYIV YNFPIGVFKR
SLVTQS

Appendix 5: DNA multiple sequence alignmets

Alignment 3: Multiple sequence alignment of the coding domain sequences of all four *PpPDIs.*

```
CLUSTAL 2.1 multiple sequence alignment


CCA40283.1      ATGCAATTCAACTGGAATATTAAAACTGTGGCAAGTATTTTGTCCGCTCTCACACTAGCA 60
CCA40358.1      ATGAAGTC----------GTT---ACTGCTAC-----TTCTATTAGCTCT----CTGTCA 38
CCA36295.1      ATGAAGTT----------ACTA--TCCTTGGCA---CTTCTGGTGTCTTTGGTGTCTGCG 45
CCA36622.1      ATGAAAAT----------ATTA--AGT---GCA---TTGCTTCTTCTTTTTACGTTGGC- 41
                *** *              *          *     *  *      * *        *


CCA40283.1      CAAGCAAGTGATC----------AGGAGGCTATTGCTCCAGAGGACTCTCATGTCGTCAA 110
CCA40358.1      --AACT-GTCTTT----------GGTACGATAT-------GGGAACTCACG----GACAA 74
CCA36295.1      GATACT----TTCTACACTCCAAAGGATGATGTAATC-----CAGTTGAAT----GCTTA 92
CCA36622.1      ----CT----TT-------------GCTGAGGTCATT-----GAGCTGACC----AACAA 71
                    *      *               *   *            *            *


CCA40283.1      ATTGACTGAAGCCACTTTTGAGTCTTTCATCACCAGTAATCCTCACGTTTTGG---CAGA 167
CCA40358.1      AAACTTTGAAAAAA-----AAGCTTTTGGGCAGCAA-GGTATGTACTCCTTCG---TGTA 125
CCA36295.1      TAATTTCAAGGATGT-----CGTTTTCAACTC---AAACT----ACTCTTCGGTTGTGGA 140
CCA36622.1      GAACTTTGATGACGT-----GGTTCTAAAGTCCGGAAAGT----ACACCTTAG---TGAA 119
                         *            *   *              *    **   *  *      *


CCA40283.1      GTTTTTTGCCCCTTGGTGTGGTCACTGTAAGA----AGTTGGGCCCTGAACTTGTTTCTG 223
CCA40358.1      CATCTACTCCCCTTATTGCAATTATTGTAACG----AGATGACACCTCAATTCGCTGCAC 181
CCA36295.1      ATTTTATGCTCCTTGGTGTGGCCATTGTCAGAACTTGAA-AAATCCC---TTCAAGAAGG 196
CCA36622.1      GTTTTATGCCGATTGGTGTTCGCATTGCAAG----CGAATGAATCCAGAGTATGAAAAGC 175
                  * *   *   **  **     * **  *            **


CCA40283.1      CTGCCGAGATCTTAAAGGACAATGAGCAGGTTAAGATTGCTCAAATTGATTGTACGGAGG 283
CCA40358.1      TAGCTGA---TTTATACGACGACA----------------CCAAGTTGC----------- 211
CCA36295.1      CAGC-------------AGCCGTTT--------------CCAAAGACT----------- 217
CCA36622.1      TGGCCGA--------AGAACTGAAG--------------CCAAAGAGTG---------- 202
                  **             *                    ***


CCA40283.1      AGAAGGAATTATGTCAAGGCTACGAAATTAAAGGGTATCCTACTTTGAAGGTGTTCCATG 343
CCA40358.1      ------AACTATTCCAG---------ATTAATGGGTATGCCAAC---AAGA--------- 244
CCA36295.1      ------ACCT---TCAGGTGGCCGCAATTGACTGCGATGCTGCC--GAAAA--------- 257
CCA36622.1      ------ATCTGATCCAGATTGCCGCCATTGA------TGCTAAC--AAATA--------- 239
                      *  *    **          *** *      * *       **


CCA40283.1      GTGAGGTTGAGGTCCCAAGTGACTATCAAGGTCAAAGACAGAGCCAAAGCATTGTCAGCT 403
CCA40358.1      ---------GGGTCTCAAAGAAATA-CGAGGT---------------------------- 266
CCA36295.1      -------------CAAGAAGCTTTGCTCTGAT---------------------------- 276
CCA36622.1      -------------CTCAAAG-TACATGAAGGTG---------------------------- 258
                             *   *          * *


CCA40283.1      ATATGCTAAAGCAGAGTTTACCCCCTGTCAGTGAAATCAATGCAACCAAAGATTTAGACG 463
CCA40358.1      ---------AGTCGGGTTT----CCTGTT---------------TTGAAGATTTTTTC- 296
CCA36295.1      -TACCGTATTCAAGGATTT----CCTACGA---------------TCATGGTTTTCAG- 314
CCA36622.1      -TACGATATTGATGGATTT----CCGACGA---------------TGAAATTGTTCAC- 296
                             *  ***    **                   *   * *


CCA40283.1      ACACAATCGCCGAGGCAAAAGAGCCCGTGATTGTGCAAGTACTACCGGAAGATGCATCCA 523
CCA40358.1      -----ATC-----------------TGAT--------------GGAA-------CCG 310
CCA36295.1      -----ACCCCC----------CAAGTTTGAT-------------------------CCC 333
CCA36622.1      -----ACCCAAGGA-------CATATCTCAT-------------------------CCG 318
                     * *                * **                          **


CCA40283.1      ACTTGGAATCTAACACCACATTTTACGGAGTTGCCGGTA-CTCTCAGAG-----AGAAAT 577
CCA40358.1      ACATGGGGTCTTATACC---------GGTGTCAGAGGCA-CTC----------AGAACT 349
```

```
CCA36295.1      AC------------ATCA--------AGCACCAATAGAAGATCT-----------GGTGC 362
CCA36622.1      ATCGAATTTTCTGGATCA--------AGAGACAGTGAAAGCTTTTTGAACTTTTTGGAGT 370
                *            * *          *          *  *             *

CCA40283.1      TCACTT-TTGTCTCCA--CTAAGTC-TACTGATTATGCCAAAAAATACACTAGCGACTCG 633
CCA40358.1      TCATTGATTATATTCA--TGAAGTTACAGGGGTTACTCCAAGCTTTCCA----------G 397
CCA36295.1      TCA-TGCTAAT--------GAGGTA--TACAGTGGTGCTAGA----------------- 393
CCA36622.1      CAACTACTGGTTTGAAGTTGAAGAA--GAAGGCGGAAGTAAAT---------------G 412
                  * *  *  *         * *               *

CCA40283.1      ACTCCTGCCTATTTGCTTGTCAGACCTGGCGAGGAACCTAGTGTTTACTCTGGTGAGGAG 693
CCA40358.1      ACTCTTA----------TGTCAAAC---------AGCCTACT--TCA-------GAGGAG 429
CCA36295.1      ------------------GATA-CTA--------AGTCAATT-----------GTTGAG 414
CCA36622.1      AGCCTTC-----------GTTAGTTC--------AATCAATT-----------GATGAT 441
                              *  *            *  * * *            *  **

CCA40283.1      T------TAGATGAGACTCATTTGGTGCACTGGATTGATATTGAGTCCAAACCTCTATTT 747
CCA40358.1      C------TAGAAGATA-TCAT-----------AAAAGACACTAAG--------------- 456
CCA36295.1      TT-----TGGAGTGTC-TCGT----------ATAAAGA-ACTACGTCAA----------- 446
CCA36622.1      TCAACAATAGATGACC-TTGTT---------GGGAAGG-AC-AGGTTTAT---------T 480
                       * **      *  *              *  *    *

CCA40283.1      GGAGACATTGACGGATCCACCTTCAAATCATATGCTGAAGCTAACATCCCTTTAGCCTAC 807
CCA40358.1      AGGGATATTCTCG---TTGCTTTT---TCACAG----------------CCTTGGCTTA- 493
CCA36295.1      GCGAGTGTCACCCAA-TAATATTAAC----CAAA--------------CCCTGGGA--- 483
CCA36622.1      GCAGTTACTGCTTCG-TGGTGTGGATATTGCAAA--------------AGATTGCATCC 524
                              *        *                    * *

CCA40283.1      TATTTCTATGAGAACGAAGAACAACGTGCTGCTGCTGCCGATATTATTAAACCTTTTGCT 867
CCA40358.1      ---------GAGGCTGGGA-------------GTTCCC----TTACACAAACTTTTAC- 525
CCA36295.1      ---------AATTCTGAGAA-------------G--ACC--------CA--GCTTCTACT 509
CCA36622.1      T--------GAATGGGAGAA-------------GTTAGC--------CAAAGCTTTTGGC 555
                          *    *                *    *           *** *

CCA40283.1      AAAGAGCAACGTGGCAA-----AATTAACTTTGTTGGCTTAGATGCCGTTAAATTCGGTA 922
CCA40358.1      ---GAATCTC-TGGCAA-----GAT--ATTATGCTG----------------------- 550
CCA36295.1      TGTGACAGATAAGGCCA-----AACCTAGTGCTCTG----------------ATAAAGTC 548
CCA36622.1      AATGAC-GATATTGTCATCGGAAACGTTGT-TACCG----------------ATGTTGTG 597
                   **        *  *      *     *     *

CCA40283.1      AGCATGCCAAGAACTTAAACATGGATGAA-GAGAAACTC---CCTCT---ATTTGTCAT- 974
CCA40358.1      --------AAGAGCTT-------GATGA----------C---TCTTT---GACTACC-T- 577
CCA36295.1      AATTGCCTTAGACTTTTTG----AATGA--CATAGAGTC---TTTTT----ACTATCCT- 594
CCA36622.1      GAAGG--TGAGAATATTAA----GGCGAAGTATAAAGTTCAATCTTTCCCGACTATCCTG 651
                         ***   *         **                * *      *  * *

CCA40283.1      ------TCATGATTTGGTGAGCAACAA--GAAGTTTGGAGTTC-----------CTCAA 1014
CCA40358.1      ------TCGTCA----------------GAA---TAGACGTC-----------TCCGA 599
CCA36295.1      --TTCAACGACAA---------AACGAAGAAAGCATTGACTAC--------ACGATTAGA 635
CCA36622.1      TACTTCACAGCAGGCTCAGATGAACCAATAAGATATGAATCTCCAGATAGAACTGTTGAA 711
                       *   *                  *    *  *   *                *

CCA40283.1      GACCAAG-----------AATTGACGAACAAAGATGTGACCGAGCT----GATTGAGAAA 1059
CCA40358.1      CGCCAAG-----------AAT----------------GCTGAGAT----TGTTTCGAAG 627
CCA36295.1      GG---AGT-----------AT---------------CAACAA----------TCGTTT 654
CCA36622.1      GGTTTGGTTAAATTTGTCAAT---------------GAACAAGCTGGCTTATTTCGTGA 755
                      *           **                  *           *

CCA40283.1      TTCATCGCAGGAGAGGCAGAACCAATTGTGAAATCAGAGCCAATTCCAGAAATTCAAGAA 1119
CCA40358.1      TT-------------------------------TCA-----AGTTTCAAAAACTC----- 646
CCA36295.1      TC-------------------------------TGGAGAGAGCAT---AACTTCGC--- 676
CCA36622.1      TC-------------------------------CAGATGGAACTTT-GAATTTCAA--- 779
                *                                     *      **  **

CCA40283.1      GAGAAAGTCTTCAAGCTAGTCGGAAAGGCCCACGATGAAGTTGTCTTCGATGAATCTAAA 1179
CCA40358.1      --------CTTCAGTTT----------TCCATT------TTGCCTCTTAT-----CGTG 676
CCA36295.1      --------CTTCGATTT--------------TGGTG----------CTACAT----GAA 699
CCA36622.1      --------CGCCGGTCTAATTCCAGGAGTGAGTGATAAACTTACCAATTACATTAAGGAA 831
```

```
                        *  *    *                                  *

CCA40283.1      GA-----------TGTTCTAGTCAAGTAC-TACGCCCCTTGGTGTGGTCACTGT------ 1221
CCA40358.1      AA-----------TATTATGATCAAACACATAAGCTTTTCA------------------- 706
CCA36295.1      AA--------------TGAAATCCA-CAT-T---TTTGATGGGAAACT------------ 728
CCA36622.1      AAAGACCAAAGTTTATTGGAGTCAA-CGT-TAGACTTGCTAAGCAACCATGAACATATCA 889
                 *              *    ** *     *

CCA40283.1      AAGA-GAAT--GGCTCCTGCT--TATGAGGAATT----GGCTACTCTTTACGCCAATGAT 1272
CCA40358.1      AAGA-AGATTTGGGCCCCGT-------GGAAATT----GTTCAATTGCTA---------- 744
CCA36295.1      -GGATAAGCTGAGCATTAGC---------AAGTTTTTAGCCGAGTTTTCA----ACTCCC 774
CCA36622.1      AGGACAAATTCAGTGTCAAATACCACAAGAAGGTCATAGAAAAGTTGTTG----AAGGGA 945
                  **        *                  *  *    *   * *

CCA40283.1      GAGGATGCCTCTTCAAAGGTTGTGATTGCAAAACTTGATCACACTTTGAACGATGTCGA- 1331
CCA40358.1      GAGGGTGCT------GAGGACTTAGTTAC------TGAGAGCATT----AGGCTGTCAA- 787
CCA36295.1      CTGGAAG--------GACCTCTCAGTAA--------GAGAG-GT-----AAGTTTCTAGA 812
CCA36622.1      GAGAATG--------AATTCCTCAACAATGAAG-TTGAGAG-GCTATCAAAAATGCTGAA 995
                  *   *         *                    **          *   *

CCA40283.1      --CAACGTTGATATTCAAGGTTATCCTACTTTGATCCTTTATCCAGCTGGTGATAAATCC 1389
CCA40358.1      --CA----------------TTATCAGACTCTCGCC----------------ACAAA--- 810
CCA36295.1      -GCA----------------CATTCGCAGGGGAAT----------------TAAACC- 836
CCA36622.1      TACAA-------------AGCTATCGGCAAACAATT----------------CAGACT- 1024
                  **                **                                * *

CCA40283.1      AATCCTCAACTGTA-TGATGGATCTCGTGACCTAGAATCATTGGCTGAGTTTGTAAAGGA 1448
CCA40358.1      --------GATGT--TGAGGAGTCTGGGGAAGCAGA-TAATT----------------- 841
CCA36295.1      ---------CGGA----AGAAAAGCGAAGAA---GGGCAAG------------------ 861
CCA36622.1      ---------CTGTGATTAAAAGACTGAATATCTTGCGCAATT----------------- 1057
                           *     *           *    *    *

CCA40283.1      GAGAGGAACCCACAAAGTGGATGCCCTAGCACTCAGACCAGTCGAGGAAGAAAAGGAAGC 1508
CCA40358.1      --------CCTTTGAAGAGTATGCCC---AACT-------------------AAGGGAAT 871
CCA36295.1      --------------AAGGGCAAG--C---AAACCAAAA--------ATCACGAC--GAAT 892
CCA36622.1      ---------TTATTGAGGCCAAAACT---GAGTCAAAACCCCAGTTATTACACCAAGAGC 1105
                           **   *         *                              *

CCA40283.1      TGAAGAAGAAGCTGAAAGTGAGGCAGACGCTCACGACGAGCTTTAA 1554
CCA40358.1      TGTAG------------------------------------------ 876
CCA36295.1      TATAG------------------------------------------ 897
CCA36622.1      TATAA------------------------------------------ 1110
                *  *
```

Alignment 4: Multiple sequence alignment of the 3000bp long upstream regions of all four *PpPDIs.*

```
CLUSTAL 2.1 multiple sequence alignment


CCA40283.1      ------GTTCAAACCT--AGCGCTAATGTTTTCAGG--CATTTGTTGGCTCTTTTA---- 46
CCA40358.1      -----TTTTCAAATCTCCAACTTTAATCGAATCGAA--GATTGTTTGGT----------- 42
CCA36295.1      ---CTGGTCAGGCTCGGTT---TCA---TCCTGAGGTTCATTCGTTGGCTTCTGCG---- 47
CCA36622.1      CCCATGATCTGGATCCATTATTCCAGACTACTGACTGTCCTTC-TTCATTTGTGAATAGA 59
                       *      *         *      *        **  **

CCA40283.1      ------ACTGAAGTTCCTGGTGTTTCCTATG-ATGGAGGGGAT--CAAGGGCTGATTAAC 97
CCA40358.1      ------ATCGATTTTTC----ACTTTTGAAA-ACTGGCTGTA---CAATTTCTGGCTGGG 88
CCA36295.1      -ACGATATTGAGTTCTT----ATTCCAAATGTA-TATCATCACCTCATTCTTTCTTACAC 101
CCA36622.1      AAAAATGTGGAATTATTC---ATCTTGAATGAACTAGTGTCAG---AAACGGT---GAAA 110
                         **  *             *   *        *    *     *

CCA40283.1      TATGTATTCCAAAACAAATGGTTGCGCACTGGTGATGACACAAAACGGTGTGGTGTATGG 157
CCA40358.1      TAGTTTTCTTGTGACTAGAAGCTGCA-----ACGGAGCCA-ATAATAGTATTATACATTG 142
CCA36295.1      TCAGGGTAATGCCACAAAT-GTCGTCT-----TGAAACTCTTCAAAGTCAGCATC--T-G 152
```

```
CCA36622.1      TTAGTATC-TGCCAACGAT-AAGACAT-----TGGAATCTCTCAAAAGGAAGATGGAT-G 162
                *     *      *                    *         **        *   * *

CCA40283.1      TACAATCTTAG----TTTTGCGTTC--AATATG--------ACACTTTCCA-ACAACTAT 202
CCA40358.1      CATATTTTTTG----TGCTGGTTTTTAAACATGTGCTGAGAGCACTTGTCAGATTTTTAT 198
CCA36295.1      AATAGTTTGAGAAATAGGTACCTCCACAAAGTCT--------CAAT-ATCA----ATC-- 197
CCA36622.1      AATTGCTGATGACGGTAATGTCTTCGCCAACTGT--------TGAT-GTTAGACTATCCA 213
                 *        *       *   *     *  *             *    *

CCA40283.1      GAGAGTCTTCCTTCAG--TACTGCGAAATC---------------TCACAGACATCAAAT 245
CCA40358.1      GAG---CTTCCTCAAACTTATTGTGAAAACA-------------GTTGAGGACAGCA--T 240
CCA36295.1      --G----TTCC--CAG---ATCAAAGAAA----ATCTCTCCATGGTTGAGG----CA--T 236
CCA36622.1      CGG----TTCCATTGA---ATTGCAAAAGCCTCATTCCCTCAGTGTTAAAAGAATTA--T 264
                  *    ****        *      **                 *          *  *

CCA40283.1      TGGTTCATTTCATAGGAAT-TGTGAAACCATGGATGCTTAAACCATCTTTTGTGAACGAC 304
CCA40358.1      TGAG-------GTAGAGGTATGCAGTACTGTG-------AGA------TTTGGG---GAC 277
CCA36295.1      CG--------CATTGAGCTCTCCTATAACGGGA--------------CTTGATTAACCAT 274
CCA36622.1      CGGTG----TTGTTGATATTTTGAATAGCAACT--------------TCCAAGAGGTTAC 306
                 *          * *   * *     *                               *

CCA40283.1      TTTC--CAGATGGCAGTCTTGACAGTTTTGTTGCCCAATGGTGGGAGCAGTTTAGCTCTT 362
CCA40358.1      TTTT--CGA-----AGGCTCGGTAAAGTAGATG----ATGCTGAAAGTAA---GACTTTC 323
CCA36295.1      CTTC--CA----------CCAGC------CTTAT----TAGCAGGAT--------CCTCC 304
CCA36622.1      CTTCGTCA----------TTAGCAGACAGTTTAC----CACCAACATAAAGTTTACCATC 352
                 **   *                          *           *         *

CCA40283.1      T-TGA-GAAC-GGTGAGTTTCTACCTTTGGTTTTTAAGAATGTTGAAAGCGAACGTATTG 419
CCA40358.1      TGTAATGAGCTGGCCAGGTCGAACTT---GTTC------ATGGTGAA-GTAAA-GT-TTG 371
CCA36295.1      C-----AAACTGGCTTGATTC-------CACTT---------TTAAA-----------CA 332
CCA36622.1      C-TGAAAAACTAACCC-ATTCACTGTAGCATTC---------TCAAG-----------TA 390
                       * *        *            *            *

CCA40283.1      AAGAAGACTCCC-ATGAGACGGAAGA---------------AAA------GGTGGACGA- 456
CCA40358.1      AACAGTATTCCA-GCAAAATAGTAGAC-TGCTATTTTCTTTGAA------GTTGCAAGAT 423
CCA36295.1      AAAAGAGCTCTGTGCCAACTGGAAGGTCAGATACTTTCCTCAAA----TTGCTCAATGAT 388
CCA36622.1      AGGAGGA-TCCAAGTAGATTTATGG---AGGAGTTTCCGACAAAAGTATTGTTTGATAAC 446
                *  *    **               *                  **      * *  *  *

CCA40283.1      ----AGAAGTCTC--CATT--TCAG----AGCCTCAGGATGAAACCACTGATTTCAAGTA 504
CCA40358.1      ---CAGGGTTCTTGTCACT--TAAAC---AAATTTTGAACGAC-CTATTGAATACATCAG 474
CCA36295.1      TGCCACAACTGTTCGCATGAGTCAATATTGGTAGTAGTAGTAG--CACCGC---CATTCG 443
CCA36622.1      --CTTAAAGTGTCAAAATCTGTAAA-----GTTGTTGAGGGAA--TACTCA---AAAATA 494
                        * *    *    * *          *    *    *        *

CCA40283.1      CCAGTTTGGCCATCATT-CCT-TTGAAGAACCTG----CACCGGT--ATTA-----GATT 551
CCA40358.1      CAATTTTCTCCAAAGGTGCCC-TTGAA-----TG----C-TCGGG--AGTG-----GATT 516
CCA36295.1      CCACTGATGGAAACTCTACTTGTTTAACGGTCTGTAAATATTGGTCAAGTGTGTTTTGCC 503
CCA36622.1      CCATTTAAAGGCACTGT-CAAATTCAAGGATCTGGTACTATTAGTTGAGT------TGAT 547
                * * *          * *   ** **     **         *   * *

CCA40283.1      ATTCAACTGAGGGAGAAGCATGGAAGCTT---AATGAAGAACAATTGACTAATCAG---- 604
CCA40358.1      GATCACCTTTGTTACTATTATACAAATTTC-CGACATATTTCAGCTCAGTAGAAAG---- 571
CCA36295.1      AGTCATTTCTGG-A--AGCCTCCGAACTTGTCTCTTTATT-TTCCTGTTCAGAAGG---- 555
CCA36622.1      AGCCGCATCTGC-ATCAATGTAC-AGTATGTCTACGTACAACTGATATGCGGCAAATTCC 605
                   *   *  *  *  *   *   *  *        *       *

CCA40283.1      ---TGGGATGTTGATGCACCAGCTG-AGCCGCTGCCTGTCCCTGTTGAAGAAGACGAACG 660
CCA40358.1      ---GAGGAT-CAGAAG-ACC-GATA-ATCCA--GTTTGTTTC-ATTATCAGCGAT--ACG 619
CCA36295.1      --CTCAGCTAC---ACTACCAGATGTAGTTACGTTTTCGGTC--TCCT---TGCCCCTCA 605
CCA36622.1      AACTCGGTTTCTGAATTTCCCGGTATAGTTGA---TTCGACC--TCTAAGGTGACATACG 660
                      * *        ** * *  *         *    *  *       *    *

CCA40283.1      CGAAGAGACGAAGGCCGAAGCTGAATTGGAGGAGCTGCTACCGGATATTGTTCAGCCTGA 720
CCA40358.1      -GTTGAGCTGG-----GAAACTAGATC--AAGACTTGTT-TCAAACGCCTTTTTG---AG 667
CCA36295.1      -CATGATGCCAA----AAAGTC-CAAG---AGTATTGC-ATTGGATGA----------GG 645
CCA36622.1      -GATGATCTTCC----AAGGTTGCATT---GATACGACCGTTG--TAA----------AG 700
                    **          *      *
```

```
CCA40283.1      ACCTCCAGCACCTCATGT--TTTTC-CCTGGGAAGCTTACAATGAAAAGCCGACACGAGT 777
CCA40358.1      GTCTCCA-TGTTTGGTAA--TTGAT-T-TGAGAAGCG-ATA--GATATGTTGAGACCAAA 719
CCA36295.1      AGCGGCA--ACTTGTT----TCGTCGTCCGAGA----TCCATTGATATCCTCAAAGAACT 695
CCA36622.1      AATGTCA-TACTTGTCAAAATCATTGTTTT-GA----TATATTAAGATGG--ATGCGAGT 752
                     **     *        *          **       *   * *     *    *

CCA40283.1      TT--TCCATGATTACAGGTAAAACTTTAACTAAATATAAGGTA-TAAGATAGCATCATTT 834
CCA40358.1      TCGGTCCAT--TTTTGGTCAACAATCCGGTTTTGT-TGCGGTAATGCGA-AGTTTGATCT 775
CCA36295.1      TG--GACAT---------TG-CAGTT---TTGGGA--AACCTG--GAGA-AAT---ATGT 732
CCA36622.1      TG--CTCAT---------TGTCGATTCCATTGTGA--AGTGTA--GCGTTGAC---AATT 794
                *     ***              *     *          *     *        *  *

CCA40283.1      TCGTATTTAGCTGGTACTTAGTCGAATTTTCTACTGCATGGTAGCGTTTGATACACGTTG 894
CCA40358.1      T----TCTTGAAGAAGGTTATTTGGAATGCTTATGGAAAGGAGAT------TACTCAGTG 825
CCA36295.1      TCGAAGCTTGGTG----TCGGATGGA-CGATTTTCAAGTTGGAGCGC---GCATGCAGAG 784
CCA36622.1      CC----TCTGGTG----CTACATGAGTCGGTTGCTAAACAGCCAGGT---GTGTAAACAT 843
                          *  *          *       *        *

CCA40283.1      CCCATTCTT--CT-----CTTTA-AGTGCGATTAACTGCTTGGAC---CGGTAAG-TCTC 942
CCA40358.1      ATGAGCCAT--CTTGAAACCTTACAGCTTGAGACACTTGTTGGAAGTACTGCAGC-TCTG 882
CCA36295.1      AAAGGTCAATGCTAGTAGCGCTA-GGTATG--CCATTTGTTCGCC---TTGTAGGATCAA 838
CCA36622.1      AAAATTCG----TAAT--CACCA-GGAAAA--CTAATATTTGG-------ATAGAATGTC 887
                       *     *     *   *  *         * *  ** *         *   *

CCA40283.1      GTAATGAA---------CGACCGAAGTAGTCTCTAT--TAAATCAT-GGAG---GTGGGT 987
CCA40358.1      ACATTGAATTGGAGGTCCAACTGGAGAGAGATAGAT--TTGATGGTAGGAA---ATG--- 934
CCA36295.1      AGAGTTGAGAG----TTGGAGCCGGGTAATACTGACCATTAATTCCATTAATTCATACCT 894
CCA36622.1      ACACCCACTG------TGGAGTCTGG-AAT-TTCACCATTGAT--CAATGAT--ATGGTC 935
                  *                 *     *        *   *  **            *

CCA40283.1      ACG-TGTTAGAAAGTC---CCGCA-GGTACACAAATTCGCACTGCGTTACATCTTCCACG 1042
CCA40358.1      ACGGTGTTGGGATGTTGTTCCCTT-GGGACAGAATGATGTTTAACGATGCAATT------ 987
CCA36295.1      ATATCGTTTATATACCAAACTTAA--GCCAGTCTAGATCTTCCGCAATGTTTCT-----A 947
CCA36622.1      GT---GGTAATGGACCCATCTTCGTGGGGGGTCCAGG--ATCCA-GGTGCCTCT-----A 984
                     * *          *      *                     *     *

CCA40283.1      TTGATTAGACCC--CATCGGGTTCTACGACCTCTGATGACCTGTCCATCAATTGT-GGAG 1099
CCA40358.1      TTGATACAATAT--AA--GGATTTGATAGACTTAAAT--CCAGTCAATTACTTG---GAG 1038
CCA36295.1      TTGTT----CGG--GGTTGGACCTG-CAGTGCCAGGGTCTTTGG-AATT--TCGTCTGGG 997
CCA36622.1      CGGATACACTGGAAGAATGAAACT--CAAAATCA---TCACAGGTAGGT--TCGT-TGAA 1036
                  * *           *                        *        * *   *

CCA40283.1      A-------TCTCATTTTC---AGAGCCAATTACTGC---AAACGGTAAAAGACTTCGGAT 1146
CCA40358.1      A-------TCT-ATCC--------GTT--TTACGGT---GTAAGGTTAAAGGTGCAAAAT 1077
CCA36295.1      ACCGCATGTTTTGTCCTCTGGGAAGCC--CTCCTGTTTGACA---TCAAAACC----AGT 1048
CCA36622.1      AC------TTTCGTTACCATAAACGTAGATTCCTGATTGGTAGATTTGAAGTCCTTTAAG 1090
                *       * *  *          *     * * *      *   *  **

CCA40283.1      ATC------TTGATTCAAAGTCAGTTCTTCTTCACTCAAATCTTCA--TTGTCGTACG-G 1197
CCA40358.1      ATCA-----TTTATTATATAGCAGTTGCATTTTA--AAGATTTACAAATGGTCTCGCTTG 1130
CCA36295.1      AACAACGTGCTGAGTTTGAA--GACTAAGACTGGA-AGGACGTCTG-GCAGTCT--CTGG 1102
CCA36622.1      AGCACCGATTTCATTGTAAGCGGACAATACTTGGA-TGGTTACTTGAGTAACGT--C--G 1145
                * *       * * *                 *                      *  *

CCA40283.1      AAA-----AAAGTTTAGGTCGT--GTTTCTCAA-ATTGAGTGTGCAAAATA--------- 1240
CCA40358.1      AAACAAGTAAAGTTGACGATGCAGGCAACTTAATATCCAGTGGCCAAAATAAAACTGAAA 1190
CCA36295.1      AAAT---CTCTGCTGAAA-CGA--GTTTTTCAGAAT----AGGGGAAGATGGT------- 1145
CCA36622.1      AGTT---CATTGTGGAACTCGA--ATTCTTGGTAATCTCCAGACCACGTTAG-------- 1192
                *          *   *    *        *    **     *   *   *

CCA40283.1      ------CTTTTAAACTTGG----TACGCTCTTCGG---TAGTCAAGGT---GTCA-GCCT 1283
CCA40358.1      TGAAATCCATTTGAGTTGA----TTAACTATTCAG---TTTTGATGGACAAATCATGTTT 1243
CCA36295.1      -ATACCGCGCTTCATTTGGAATTCTTGGTGCAAAGAGCTGGAGATGGAGTTGGAGAACCC 1204
CCA36622.1      -GTAGCCTGCTGT-TCTGG---TTTTGTCAT-------TAAAGACGGTGTCTTCGTATAC 1240
                          *     **                    *    * **

CCA40283.1      -TAGCAATAACAGGAAT-------AACGTTCACAA--CAGTAGTGAGTT----TGGAAAG 1329
```

```
CCA40358.1      -TGTCAAGCACAGCCCTTTTCTTTAACGTCTGTACTTCAACCTCATCTC----CAGATAT 1298
CCA36295.1      ATGCCCATAGT---CGTTG-CTGGGAGATATAATACCTAGGTCTGAGTT-AATTGGCGAG 1259
CCA36622.1      ATCTGAAGAGTTTGCGTT--CTGAAGAGCGT--CTTCCAACTGTGATTTTGAGAGGAGTG 1296
                 *    *         *                     *       *       *

CCA40283.1      AACT-TGTATGTCAAG-----GTTTC---TCA-GTCCCTTACCGTTA-GGTTGAATGAAG 1378
CCA40358.1      ATTA-TAAAAGTGAAATATAAATATT---TTA-ATTCGCTAT--TTA-CATTTAACCCAA 1350
CCA36295.1      G-CAGCGTTCGTAA-A-----ATACGAAGCAACAGCCACTGG--TTGCTGTGAAGTAAAA 1310
CCA36622.1      GACACCACGCGTCACA-----ATAT----TCATATTCGTTGG--TT--TGTGGATTCACA 1343
                          ** *        *       *    *  *    **    *  *

CCA40283.1      TAAAGCACACAATGAACCCTTG------TGTCAGGT-ATAT-ACCTTTCTCTCTGTGCTG 1430
CCA40358.1      TTCGGAACCCA--GAATCCTTAAGATCCTATCAAAA-AGATCGCACACCATTCTGTACCG 1407
CCA36295.1      C-TGGAA-ACAGAGGAAGGTGA------TGCCGGTATAGATACTCCAGGATTATTA-CC- 1360
CCA36622.1      TATGTCA-A-----GTTCATGA------TACTGTTT--GAAGGGCCAG--TTATGAGCC- 1386
                       *           *        *          *           * *   *

CCA40283.1      T---CAACTC--TTTTCGCAAGTAGGTA-----GAGTTTTGCTCTTTGA----TATACTG 1476
CCA40358.1      TAAACAACACAACATTGGTATATGGCCGTCCC-AAGTACACCACCAAGAGGCATACAATA 1466
CCA36295.1      TGAAAGACT-GAT-TT-GTGACTTTCTGCGCTTGAATCCTGCTGAATGCGG---GGAGAA 1414
CCA36622.1      TAAAGTACTCGACACC-ATAACCCTCTGCA---GAA----GTTGATT--GG---AATAGA 1433
                *     **                          *

CCA40283.1      CAATA--TTGGTTCCCAACATTTGTCATTGTTGATCTGA------------TCTCCAA- 1520
CCA40358.1      CAAGAAATCGATACCCA-TTCTGGCTAACAATGATGCAC------------TCCTCAAC 1512
CCA36295.1      CGAAGTATAGGGA--GAGTC-TTGT---TGTTGTGGCAT------------CATGGG- 1453
CCA36622.1      CGGAGCTTGGATG--GAATTATTGTG--TGGTGGAGCATAGCTAGCCAAGTCCCTTGGG- 1488
                *      * *      *    * *       **                    *

CCA40283.1      ----AACC---AGGTG---------TGTCA--ATCACGTTGAGTTT-CAGATTAAC---- 1557
CCA40358.1      ACCCAACC---AGTAACCC-----CTGTCAC-ACAACGACAAGTTAGCGAGTCAACAGGA 1563
CCA36295.1      ----AAC----GGGAATGTTTTTATTGCT------GGGACCCCTGGAAGGCTTGGGCAAA 1499
CCA36622.1      ----AGCCTGAAAGAATCCATTCATTGTCTTGATCGGGACAAGTGATTGACTTGGG--GT 1542
                    * *                  **          *     *       *

CCA40283.1      --GTTG----TTTTCATTG---AGAATTTGGGAGGTGCTTT------TAATCTCTACA-G 1601
CCA40358.1      CTGTTGAAGTTTTCCATAGGCAAAAATGCGGCATCTCCGGTGAGAGATGGTCTCCCCATG 1623
CCA36295.1      ATGG-GAACATTCTTAT------GCTTTTGAGCGTTGCTGT------TGA----GACTGT 1542
CCA36622.1      CTGGCGAAAGGGCTT---------CGTTGGTAC-CTGCTGT------TGA----CAC-GT 1581
                  *  *                    *  *     * *  *      *       *

CCA40283.1      TTCTTTCCACAGGCTCAAAAGTACTTTTTCTTCCG-TTAGAGTCCACAAGATG-TGTGGC 1659
CCA40358.1      TACCTTCCAT----TGAAATCTGCTGTACATGAGG-ACAAGGAAAACGATATCCTGTTTC 1678
CCA36295.1      TGCCTCTCATTAGATAATAGGAATCTTGTGAACC--TCCGTAGGGAGAAAATGATACGGC 1600
CCA36622.1      CGGC--------GATATTA--AAACTTGTTAGTTGATCTGGACTGACATCAT-ATATGGT 1630
                        *   *       *                   *    **  *

CCA40283.1      AA------AAAGAGTATTGAC-TAG-AGTTGACTT--CCCAAGCCC---AGATTGACCAA 1706
CCA40358.1      AACTGGCTAAGAAGCAACGAC-AGGTGGTGGACCTGGCTCAAACATTAAAGATAGCCAAA 1737
CCA36295.1      TGTTGTTGGAAGAGAATTATTATAGGGGTAAGGAT-----AAGGCT----GTTG---CTG 1648
CCA36622.1      CAAC-TCAGTTGAGT-TTGATGAATTGGTGGACAC-----ATTA-T----GATG---AAG 1675
                     **          **          *       * *

CCA40283.1      CAAG----CATAACATTG-------AGCTCAAATCCACGTTTAA-----GGAGTCTGTTT 1750
CCA40358.1      CAGGAGCTTGCGGAACTGGAACGAAAGTACCAATCCGAGTTCGATTCTAGAAACCCGTCC 1797
CCA36295.1      TTGT----TGTTGTATTTGCTTTTGAGTTTGAGTGCTATGAGAAGGTT--GAATATGTTC 1702
CCA36622.1      TAAG----GATTG-ATTCCCCAAGGAGTAAAAGTTTCCACCGAAAATC--AA----GTTC 1724
                     * *        **    * *         *       *    **

CCA40283.1      TCAATTTGGGTAGTGATAGTATCAAACCCCAC--ATA----ATTCTTAGGTT-----GAA 1799
CCA40358.1      CCTGTTCA---AAAGA-AGGATCAGAATCCACCAACG----GCGGCTAAG--------AA 1841
CCA36295.1      ATG--TTGT-TGTGGCTGG-AATTGAACTTGTTGTTG--TTGCTGTTGCATCTTGCAGAG 1756
CCA36622.1      TCA--TCAT-CAATGCTAACAATTGCGTTCACTGATGCGTTTTCATTAAATCCTTGGAAG 1781
                     *       *     *                        *         *

CCA40283.1      CCACTTGACTTG-CCATAGTTTTTGT-ATTATCT-GACGAACAAAATGAAAGAGAATAA- 1855
CCA40358.1      TCCATTGACTGAACTAAAAC----GG-ATCATG--GAGAAAAACAATGAAAAAAAACCAT 1894
CCA36295.1      CCCGTGTGTT-AAAGAAGATTGATGTTGTTGTCTCTTGGAATTTCGTTTCCTAATGACGT 1815
```

```
CCA36622.1      GGAATGGACTCGAAGGAGGAAGAT-TCATAGT----TGAAGCAGACCAAACCAAAAATAG 1836
                    *    *                  *  *       *           *

CCA40283.1      CTGTAACCAAC-GCGTGGACAGGTAGGACA-----CAGAAA----TCGAGGAGAA--AGG 1903
CCA40358.1      CTGCACTCAAGAGCAGGAAAAGTTTGTTCAATTTGCAACAAGACTTTGAGGATAACAGAG 1954
CCA36295.1      CCGTTTTCAAAGACA--A--TGGCCTCACAGTAC-CATGAAGTTTTT--GGAAA---AGA 1865
CCA36622.1      AACTGTTCAACGTCA--AATTTCCCCCAAAATAT--ATCAA---CTT--GGACGTTTGGT 1887
                       ***   *              *      *  **    *   ***

CCA40283.1      CGATT----GCATT-CGCGACTGTATCTTTCCGTGCAGCT--CTATTACGAGAGTG-GTG 1955
CCA40358.1      CAATCCAAAAAATC-CGCA--CGGATTTTACTAC---TCT--CTCTT-CGAA------CT 1999
CCA36295.1      CCACA---AGCATTGCACAA-CGTATTGCCACTTCCATCT--CTTCTCCAAAGAGGAGTC 1919
CCA36622.1      TGACA---AACACGGTTTGA----ATAGATCCTTGAACTTGATTACTCCAA--------- 1931
                  *        *             **             *   *  * * *

CCA40283.1      TTCTTAGTT------TTTCCCTCTTCAATTTCTACTAAAGCTGC-----------TTTG 1997
CCA40358.1      TTTTAAATG------ATTC--TAAGAAATTTCAGC--AGGGTTT-----------TGTG 2037
CCA36295.1      TTTTCAGTCTTACAATTTGAGCATTGAGTTCTTGTTATACTTTCGTCATTGTCAGCTTTG 1979
CCA36622.1      TTTTCGGTC----AGGTTCGATATTGAAACCT-GAGAAGCTTCC-TCGT--TCACTTCTA 1983
                ** *   *        **        *             *             * *

CCA40283.1      GTAAAGGGAGA------CATATTCGGCT-ATTGTTTACT-TTG-CGCCC---ACAGTAGC 2045
CCA40358.1      ACAAAAGCCAA------TAGTTTCT-----TTAACAACT-TTAACACCTCAGGCAGCAAC 2085
CCA36295.1      GGGGATGATGAGTCACTCATATCTTTCTGAACAGTCACTGTTGTTGTGA-TACTAGTAGT 2038
CCA36622.1      GGATCAGAGGAGC--TTCATTTTCTTC--GATGGGGATTACCAACACGA-AGTTATT--C 2036
                      *   *       *  *             * *             *

CCA40283.1      ------GTTAAGAAACATTGTTTGTTCGATTTATTGGGCTGTTGATAAATTCAATTGATT 2099
CCA40358.1      TCCGTAGCTAGCGACCAAAATTC-TTTGGAGCGCAAAGCATCTGG-AAGTTGCCTGTATG 2143
CCA36295.1      C------TCGG-AAGTAGGTTTCGCTAACTGTA---GTTTA-TGAGAAGCTCCGTCACTC 2087
CCA36622.1      C------CCAATGAGTAGGACTGTTTTATTGTGCC-GTTGAATGTGGAAATGAACGAT-- 2087
                           *  *    *   *                 **   *  *

CCA40283.1      ACG--TTCGCATACTAGC-TATCATAAACTAAGC----ACCACCTTACACCACTTTCTCA 2152
CCA40358.1      ATG--TTCGAA----AGCATGACGCAAACCAGAC----A-----TTGATGAGCTCTGAAG 2188
CCA36295.1      AAGAATTCTTGCAAAGACTTGTCTTGGATAATATC---ACTA---AAATCTCCTCCCATA 2141
CCA36622.1      AACGACTCATCTT--GACCTAGTTTGAA-AATGCCTAAACTCTTTGAATCTTCGTTCCTA 2144
                *     **         * *       *  *       *           *

CCA40283.1      CTGAAGATTT-----TCGACATCAAATTTC------------TCTTGGAT-----CACC 2189
CCA40358.1      ATGAAGAT---------GAAG--AAGATTA------------TGGCGGAA-----AAGC 2219
CCA36295.1      CCA--TCTTC----CTCACCATCATCATTG------------CTATCGAAC--TTCCA-C 2180
CCA36622.1      ACAGATGTCCTGTACTGATTGGTATTATTGGGTGTAGAAAGTCCGTTGAACCGTCCCAGC 2204
                      *               *   **                  * *       * *

CCA40283.1      AT--CAAC----CTTG--TGTCTACATGTCCT---------TGTCTTTGAACCTAAATCA 2232
CCA40358.1      ATACCGACTAGACTTG--TAGTCAAATAGCATACAAA----CAGCATCAAGACCA--TTA 2271
CCA36295.1      ATCGAGTTTAGATCCAAATTGTTGGA-GTCATGCAAGGGACTATCATTCATAGATGAGGG 2239
CCA36622.1      AAAGAGATATCATTTGAAAAGTTGATCGCCTGGCAAGAGACGATTATTGAGACCAGTAGG 2264
                *                       *                   *  *

CCA40283.1      GATAGCCGTGCGGGTTGTGGGCATATTGCCTCGTATTCCGGAGATTCAC--------ATT 2284
CCA40358.1      AATACTCGCCTCAACTAAAAA--------CTCGTCTTCCTT-GATTGAC--------TTT 2314
CCA36295.1      CGTAACCATCATTG--AAGGT--------TTTGGTTTTCGA-AGTTCG-------GATTT 2281
CCA36622.1      AACAACAA-CGCCGCCATAGT--------TTCAATCT--GA-AGTTGACTTGTGTGATTT 2312
                   *                         *     *       **           **

CCA40283.1      GCCATTCCTAA---TATTTTTC---AGCGACGCAC-----CGAAG--------CT-TCTA 2324
CCA40358.1      ACTTTTTCGAT---TAGATCACCA-AGCGGTGTAT-----CAAGAGACGGGTACTGTCTA 2365
CCA36295.1      AACCTTGGAAT--CTGAGTAACTAGA-CGGAGCCC-----CAGAAT--------TATTTT 2325
CCA36622.1      GGTGTTGGGGGGATTGAGGGTCAAAAACGGGGGTCAGAAACAGAATA-AAAAACTAGCTT 2371
                    **        *      *   * **  *        *             *   *

CCA40283.1      CAGAGACTCACGATCCTCGCATACTAGAGCTGATAGAAAATCTACAGGATGC--CG--AG 2380
CCA40358.1      C-GCTATTCACGGTTAGCGTCAACAAAATGTAAAAAAAAAAGA-AAAAAAGAC--TGT-AA 2420
CCA36295.1      C--CAAC--GAATCTGGAGAATAATACAGGCCAGGAGCACTTGAAATAATCTC-CGTCAA 2380
CCA36622.1      G--GAACTCGAAGTAGGAGAAGAGAAGTGGTGAAAAA-AGAAGAGGGGACGTGACGTTAA 2428
                     *           *   *  *      *     *    *    *      *  *
```

```
CCA40283.1      GTTCCTCCATTCTTCATTG--ATAACGGTAT----ACTTAAAGCAGCACCAAAAAAGAAG 2434
CCA40358.1      AGCCCAACTTACTT--TTG--CCAA-------------AAAGC-----CAAACAACTCA 2457
CCA36295.1      ATCAGCAGAGTCATC-TAGTTGCAATTCGG---AAAAATGTAGTT-TGCCAGAAA-GGCT 2434
CCA36622.1      AGACCTGAACTCTGCGTGGGTGCAATGTGGTTGGAAGGTGAAGT---GCAAGGTATGGTT 2485
                          *    * *    **                **      *   *

CCA40283.1      GTTT--CT----CATAT-----------GAAAAGACGCCAGAAATTATATGGTCCA---- 2473
CCA40358.1      GTTCTGCT----CATCT-----------CTTCCACTGGCTTAGCCTCAATG-TCAA---- 2497
CCA36295.1      GTCTTGTTTTGTCATCTGGGTTTTCTTGGATTCATTTCTTAATGTTCTGTTTTCAACTCG 2494
CCA36622.1      TTTTTTTCCTGCCA------------GGAATCA------AAAATATCATTCTCATTTCT 2526
                 *         **                          *       *  **

CCA40283.1      GGAA---AAA---AACAACTCTCTTTAC-------TACAAAATTTGAACAGGTGTCCTGC 2520
CCA40358.1      TGGG---AAT---GTCATC---CTTGAC-------CACC---TCTG------TGTCCTTC 2532
CCA36295.1      GGGGTATAAATCTAGTAGCT-TTTTAGCCGAAGAATATACCCCCCATATGGACGGCGAAT 2553
CCA36622.1      GAGG---AAA---GGCA-------TGGC---AGAATATTTACCT------GTTGCCAGAT 2564
                       **       *       *  *        *                * *

CCA40283.1      CTGCGGAAAC----TACAAACG----ATCACACACCCT--------CTGCATGCATTGCG 2564
CCA40358.1      TTCTTGGATTGCTTTTTCGAT-----ATCGTTCGTCGG--------CTAGAGGGCTTGAA 2579
CCA36295.1      CTTCAAGATCC-TTCAAAGATCCTGTATTATTTCTTCTGTTTCCTTCCATCTAAATGAGG 2612
CCA36622.1      CT---GGATA----TGAAGATTACATAATAAACACGTAGA------CAGTGTAGATTCTG 2611
                 *     *          *      *                    *       *

CCA40283.1      T---AGGACAAATCAGGAGACATTGGA-ACGAC----TCTGTTCCT-CAACAGGAGGCAT 2615
CCA40358.1      G---GTGGTAAA-----AAACATTGAATACAAT----GTAGCCCCCACAACGAGGCCCAC 2627
CCA36295.1      CCTTAGGGCATATAATTAGAGAAAGGACAGGACCAGTTGAGCTTCAATGCTAGGGTG-GT 2671
CCA36622.1      C---GAGGCGTGC---CAGACGTTGAGGAGGACAAGTTGGGGCTCCG---TAAGGTGTGC 2662
                       *          * *    *   *  *        *   *        *

CCA40283.1      ---TTCGT-----GAA-GAGTTTGTTAATCCTTTGGATGAGAAGATTCTTTATCCAGGAA 2666
CCA40358.1      AAACCCAT-----AAAAGAATTTCTTCTTTACTTGG---------TGTACCATT--GCAA 2671
CCA36295.1      G--TAAA-----GTAAAAATTTTGACCCTCAGG-GA---------TCTCCAAGT----CA 2710
CCA36622.1      G--TCCATATGCGGACAGAGTGTACCTATTGGCTGA---------TTGGCAA------CA 2705
                            *   * * *     *     *          *     *       *

CCA40283.1      AGAAAGAACTGCCCGATGAACGAACTTTACGTAAG---AAGGAGTGGCTGAAG-AGAAGA 2722
CCA40358.1      TGGAAGAGGT---TGATATA-----TTTAAAGAGGTCCAAGGACCCGTGGAAGCAAAATA 2723
CCA36295.1      TGTTAGGAGA----AGTGTT-----TTTTTGGAAC----AGACACGACCGGGG------- 2750
CCA36622.1      CGTTAC-ATT----GTTGCA-----TTCCTG--AT----TGGCGTTGGTGTGC------- 2742
                 *  *          *        **            *       *

CCA40283.1      CCCCGAAC--------------ACTCCCTG----TTGAATAGA-ACACG--AACACTGT 2760
CCA40358.1      CCTTCAATTCCCTACCGCGATTGACTTTCTGATGTTTGATTAGATGGACA--ATCAGCAA 2781
CCA36295.1      ----------------------TATTTGTGGGCCAAGACTAAGCAAACGGCAA-ACGAA 2786
CCA36622.1      ----------------------TCTTGTTGAAACAGGGATGCATACATATTATTATCAA 2779
                                     *   **      *  *      *    *  *

CCA40283.1      AAATAGAATAAA----AGAAAACTTGGAT--AGTAGAACT-TCAATGTAGTGTTTCTATT 2813
CCA40358.1      ATGTACAGTAAA----GCGAAAT---GA---AACGAAATTGTCA---CAGCGTTTGGATG 2828
CCA36295.1      ATGTGGGTTTGAAATGGAGAAGCGCTGCTTGAAATGAGCAGCCG--TAGAAGGTTGGATG 2844
CCA36622.1      ACATGGACTGAGGATCGAGGATC--TGC--AGGTCGGACAGCCA--CCCGACATCGTTTA 2833
                *  *    *          *     *                *         *    *

CCA40283.1      GTCTTA--CGCGGCTCTTTAGATTGCAA-TCCCCAGAATGGAATCG------TCCATCTT 2864
CCA40358.1      GACTTAATCAAGGA----CAGCTAGCAA-TGCCAAGACAGTAACTAATACACTGCATTTC 2883
CCA36295.1      GCACTA--CATGTAT---GAACTAGAAGGCGTTGA-ATTTTC-TAGATGTA-CAGA---- 2892
CCA36622.1      AGCTTA--CGTA------AGATTTGAAGGTCCCAACATCTTGACAGATCTC-TCAACTTT 2884
                    **  *            * * *       * *                    *

CCA40283.1      ------TCTCA-ACC-CACTCAAAGATAATCTACCAGACATACCTACGCCCTCCATCCCA 2916
CCA40358.1      ------TCTTA-ACAGTACTCAA-------CTTCCACACTTCGTCAGGTTCTGGTTCCTG 2929
CCA36295.1      GGGAGATCTTG-ATTGCATGCGAA--------GC--ATCGTCGTCAGGCTCCCCTTTTAT 2941
CCA36622.1      GCTTAATCTTACATAACCCCAAGA--------GCCAACCCTCACCAGGTT----ATTTTT 2932
                      ***   *                    *    * *    * *       *

CCA40283.1      GCACCACGTCGCGA-TCACCCCTAAAACTTCAATAATTGAACACGTACTGATTTCCAAAC 2975
```

```
CCA40358.1      GC-CCATG-CGCGACTCATCCCC----CGCC--TGATGAAATAGATA--AATTTTTGCAT 2979
CCA36295.1      ---TTATTTTCCCA-CCGTTTTTTG-TCCCC--TTTCTCGCTA---A---ATCTCT---T 2985
CCA36622.1      ---TTGCCTTCCGA-TTGTCTTCTCCTCCCC--CCGTTTGCTGTCCA---ACCACT---T 2980
                           * *             *  *               *   *

CCA40283.1      CTTCTTCTTCTTCCTATCTATAAGA- 3000
CCA40358.1      CTATGAACTTTTC--ACTCAAAA--- 3000
CCA36295.1      C------CCTCTC-----GACAGCAC 3000
CCA36622.1      GAAAGGGTCTGTC-----TAGGATA- 3000
                     **      *
```

Alignment 5: Multiple sequence alignment of the 1000bp long downstream regions of all four *PpPDIs.*

```
CLUSTAL 2.1 multiple sequence alignment


CCA40283.1      --GAATGTCCGGGGT--TTCTATATTTACTGTAACTAGGTTTTATTCTTTTGTATAGTCC 56
CCA40358.1      ----ATA---GAAGT--CTTCCCATTTG-TGAAGGCAAG--------------ATAGA-- 34
CCA36622.1      GTAAATGAAAAAGATACCTTTCTAGATA-AGTAAATAAATAACGCT-------ATAACTC 52
CCA36295.1      --ATACATACGAAGT---TTTTTATTTTTTTT--------------------TTTGCTTC 35
                    *         *   *    *  *                          *

CCA40283.1      AACAAACCATATGGATT-AGCATCACTTCGCGCAATACGT------TCTAGTTGCCCA-C 108
CCA40358.1      AACCAATCACATACATTTGGTGTAGACTAGAATGGATCTA------TCTGGTTATGTA-A 87
CCA36622.1      TACCA-CTGTAGAGGCTTTGTAGATTTCTGGACCAATCTCACATGCTCCGGCCACAC--T 109
CCA36295.1      TTTCATTCATAT--ATT---TACAAGTT-GAATATTTTTA------CACAATTGTAGAAG 83
                    *     *     *            *

CCA40283.1      TTGCTAAAACCAATTTAGCAATGATTCA----ACAGCAA---------TTATTATCACCA 155
CCA40358.1      TCACGATACCT--TTCGGAATTGTCTCGT---GTAGCAGGCGCGCCCCTTCGCATCTCAA 142
CCA36622.1      TTGCCGCCCCC----TAGGATGGTTTCTTCT-ATGATAGGT-------TTGACATTAAAG 157
CCA36295.1      TCACCCTAACC----CTGGATGGATTCTTCCAACTGAAGTT-----TTCTGGAATGGCGA 134
                *  *     *       * *  *  **          *          *   **

CCA40283.1      ---TGA--CTAACTGGAAAGC-GATATTGA-CTCCCGCTCAATACCAAGTCCTCCGTTTG 208
CCA40358.1      A--TCA--ACAGTTGGCAAGT-GTCACT---TTCTAAGTTAGCAC-------TCTACTTT 187
CCA36622.1      GGGTTA--CATCTTGCCAACCAGATATTAG-CCCCAG-----------------ACTTC 196
CCA36295.1      ---TCAAACAGGTTGCTAAGC--TCGTTGAAGTTCAA---------------CGATGTT 173
                   * *       **  **        *                            *

CCA40283.1      GGCGGAACAGA-AAG-ACCGTATACCGGACAGTATGTGAACTTCAAGAAAAATG------ 260
CCA40358.1      A--AGAACAAA-ATG-AACATCTTTAGAATCCTAGGTAAGTTTCCGC------------- 230
CCA36622.1      A----AGTAAATGAG-AACATTTGCTGCATACT-TGTGACCTCCGACATGATTGATAAAA 250
CCA36295.1      G----GGTTGACATCTAACGTCTTTTTG--TCTTTGTAGGTCACCTCAATGATA------ 221
                          *     * * * *         *  **      *

CCA40283.1      -GAACCTAC---TTGTGTAGTGGGTGTCAAACTCCG---CTTTACA----AAAGTGG-CA 308
CCA40358.1      ----TCTGC---TCAATTGAAGAAT--CAAACCCAA---TTTTCTA----ACCACG--CA 272
CCA36622.1      TGAACTTGCACTCCGCCTGGTCTAT--CATCTCCGAGATCTCTGTAGAG-ATCGTGATCA 307
CCA36295.1      -GGTTTT-----TTGTTTGACTGGC----TGTTCAAGA-CTTCGTAGTTTATTGTG--CA 268
                      *          *               *      *    *    *    *  **

CCA40283.1      CAAA-----------ATTTGATTCATC-TTGTGGTTGG-CCTGCATTCTATGAAGCATTA 355
CCA40358.1      GGTG-----------ATATCTCTCATT-TGATAAGCAT-CTTG-ATCCTAATACATGCTG 318
CCA36622.1      CGAAGGTGAATGCACATCTCCTTCTTAATGATAGGTGC-GGTG-ATTCCGCAACGT-TTA 364
CCA36295.1      TG-------------ATTTC-TTCAAGGAACTGGGCAAAGCTGCAAG--AAAAAGTCTTG 312
                           ** *   **       *         ** *       *     *

CCA40283.1      C--CTGGAGCAGTTAAACGAATAGAAGA---CAA----TTCGCTTGGAATGCGAAGAATA 406
CCA40358.1      T--ATC-AACAACAAAGTCTGCTCAAGG---TAT----CTCATTAAAAACTCAAGCAGCT 368
CCA36622.1      T--CCCTTGTT-TTATGCGAGCACAAGAACACATAGGACTTGTTTGTGTCTTTAACGATC 421
CCA36295.1      CGCCTTTAGCTTCTGGACCAA---AAGGGTTGAA----TTTGACCGATACTTTGGTAAAG 365
                                  ***     *      *
```

```
CCA40283.1      GAAATCAGATGCTCCAA-ATGTGA-TGGA-CATCT--TGGCCATGTTT-------TTGAG 454
CCA40358.1      TATGTCGTAGTCTT-----TGTGAGCAGG-TATCT--TGACCTCTTTT-------TCAGA 413
CCA36622.1      ----TCATAACCTTTGA-TTTCGGCGGGAATATCTATTGGTTCCACTTGGGTAGTTTCGA 476
CCA36295.1      TACTTGGTAATCATGGAGATGCTTCTGCAAGAATTG--GGTCGTGATG---------AGA 414
                    *   *  *       *           *  *   *       *

CCA40283.1      GGTGAGGGATTT---GACACTCCAACAGA--T---TCCAGACATTGTGTCAACAGCATCA 506
CCA40358.1      TATGTGTCACTTTACAACACGCTGATGAAACT---CTTTTTCATTGTCTCTTC--CATTT 468
CCA36622.1      TA-GAGAAATTT---GAAAAAAGAATGGGCACTAGCTCGTCCAATACGGCATCAACATCC 532
CCA36295.1      GAGGGTGGATTT-------TTTCGAAGAGTAT-----CCGATATTGAGTTTCGCCCACTA 462
                   *    * **             *                 * *         **

CCA40283.1      GCCTAAAATTTC----AAGGTG--AAGAAGAGAACTAAGCGGGGGGAATGACTAAATAAA 560
CCA40358.1      ACGTTGTGTTTTTGTTAAATCGGTACAAAAAGAAC--AGTGAAAAGACCAATGATATTTT 526
CCA36622.1      TCCATACACAACTGCTTGACCC--ATACAAAGTAC-----GGAAGAATTAA---AATATC 582
CCA36295.1      CCATGATATATTTTTTTTCTG--ACTTACCTCCCTCTAAATAAGAATCACTT---CACC 517
                 *                      *   *     *           *

CCA40283.1      TAAAGTAAAAATTTTGTTTATTGACGCCTTTTATAATTTGTT---TCCTTGGTCTTTAAG 617
CCA40358.1      CCCAGTATGGTATCTCTTTGGAGGAGCGTTC----GTTT-TC---GCTTTGATCTTCAAC 578
CCA36622.1      CCCAGC----TGTCTGTTCGGCAA----TTC------TTGGA---TCCTTGGCATTAGAC 625
CCA36295.1      CATAG----ATATCTACATGAAGG----CTCAGAAATTTTCCAAGGCTTCGTTACTGGAT 569
                   **       * *              *       **       * * *    *  *

CCA40283.1      ---AAA-ACTATCATTCTGGAGAATTTC-TAGAAGATACCCATCAGGGTCCTTGA--AGA 670
CCA40358.1      TACGAA-TTTAGTATTGTGGAGATTTTG-TGGAG------CTTCA--GTCTTTGGTTAGA 628
CCA36622.1      TGGGGT-ATGGATGATGTGGAGACTTT-----AA--------CC---GCTTTTGG--AAC 666
CCA36295.1      GATGAGGATGACTACTCAACTCAGCTTGACAGAA-------TTCT---TGCTCGG--AGG 617
                              *      *  **     *          *       * *   *

CCA40283.1      AGGCAAGCTTGTC----GCTAATCTCTCCCAATTTAAA---GACCCATTCTACATCATTC 723
CCA40358.1      AAG------TGTT----GCTA-TCATTCCGCAATTAGT---GGTTCTTCAGAGAACGGGA 674
CCA36622.1      CAGC-------TT----GGTAGATGAAGCCCAATGAGA---GATTGATTTTAAAATCGTG 712
CCA36295.1      AAGAAGA---GTCAGAAGGAAGACTCTGGAGATCTAGATACGATTTCTTTTGGCTCCTTA 674
                  *        *     *  *          *   *     *     *

CCA40283.1      TCAA--GCTCTTTGGCAAGTTTGGCG--ACATCCTT-TGTTTTAATGCCTGTATGGC--- 775
CCA40358.1      G-AA--GCTCAATTACTAACCACACATTACATATTTGCACTTGGAT-TATACAGAGCGTT 730
CCA36622.1      CCA-----TCAAC-ATCAGT--GGCGT--CACGCTTCCAATCGGACTTACCCGTTGGAAT 762
CCA36295.1      TCAAAAGCTCAA--GCAAG------ATTACAAACTGAGGAGAGAGAATCCAAGAAGAAGG 726
                  *     **       *           **   *                    *

CCA40283.1      ---CATATCCAAAAGGCTCAGTGGTGTTGGTAT------GGTAACCTT--TGAAGCTGGA 824
CCA40358.1      ATACATCCCCAATTGGATC--TACCGTTACTACACAGAGGGTCGCATGGACAAGATTGCA 788
CCA36622.1      GATAATGTGTAAAGAGTACGGTTTAGTTGATTTCCAAAGGAGATCGTTTTTGGAGCCTGG 822
CCA36295.1      TAAAAAAAGTCAAGAAACCGAAAGTATTGAAGAAACAGGAAGAAGATTTTCAACCTCCCA 786
                     *      *      *       **                   *

CCA40283.1      GT--CACTTTCTGTACCCCA---GTTATGG----GTCAACTCAA----TAATAGATTCCC 871
CCA40358.1      GT--CATTTCTGGTATCTTACAGACTATTG----TCTATTCTGACTTTTTCTACATTTAC 842
CCA36622.1      GT--CATCTTT-GTAACT-----GT--TTG----GATACCTTG-----TGAAGCACTCAG 863
CCA36295.1      GTGACATTTCTGATGACGGTGAGTTTTTTGAAGAGCCATCTCAAGAC-CATCAAAGTCGA 845
                **  **  *    *  *          * *       *             * *

CCA40283.1      T--CTCAGCTTGTGGAC-----GAGGTTTT----TCCTGGTTTTCCACGTATTGAT---- 916
CCA40358.1      T--ACAAGAAAGTAGTT-----AAGAACTTGAAATTCAAGCTGCCAGAATGAACAT---- 891
CCA36622.1      CGGCCGAACAAGTAGCG-----GGGCACTG-----------TGAAAGAGGAACTAT---- 903
CCA36295.1      ACTTCAGATAGGCATTCCAAGGAGGAACCT--------AGATCTAAAAGTAAGCACGCAC 897
                           *           *           *           *

CCA40283.1      CGTCATGCTCGTA-GCCCAGAAAATAGAG--------GGTGAACTTAGCTTCCGGAAAT- 966
CCA40358.1      GGAAATA-TAGAA-AAACAAAAGATAGTGTATCAATAAGTATATATGTGTTGCAGGAATA 949
CCA36622.1      TTCTGAGATCTC---CTCAGCCGCCTGGGCTTC----GCCGGATTTGCCCCCCAAGAAC- 955
CCA36295.1      CCTCAGAGTCTAGTGCTAAGAAGAGAGTGTCAAA---GGTGAGGGAGATTCCTGGAC--- 951
                         *       *       * *

CCA40283.1      ------------TCTCTAGTAG-----AATAAAGTTT--CATACCTAA-----AACTT-- 1000
CCA40358.1      AAAATCAGTCCTTCTTCAA-AG-----GATAATATT---CAGTCCTGACGCC-ATCTTTA 999
```

```
CCA36622.1      -----------TTCGACAAAAA-----AGCCATACTGATTGGACTTAAGGATGAAAAGCG 999
CCA36295.1      ------------TCTTGAACGGTTCTGGCTCAAGTTCATTGTACCAGGATGTCAGATTCG 999
                           **   *             *   *       *         *

CCA40283.1      -
CCA40358.1      G 1000
CCA36622.1      G 1000
CCA36295.1      A 1000
```

Alignment 6: Multiple sequence alignment of the 1000bp long upstream regions of *PpOCH1* and *PpHOC1*.

```
CLUSTAL 2.1 multiple sequence alignment


CCA36419.1      -------CCGCTAAAAGACCCGGAAAACCGAGAGAACTCTGGATTAGCAGTCTGAAAAAG 53
CCA39540.1      ACCCTTCCCACTTAGTAAGTTGGACGTTATGCATAATTTTGGGAT----GTACGACCAAA 56
                       ** ** *   *   ***        * ** * ***  *    **  **  **

CCA36419.1      AATCTTCACTCTGTCTA--GTGGAGCAATTAATGTCT------------TAGCGGCACTT 99
CCA39540.1      TAT--------TGCCTAAAGTAGAATAAGTAATATTTCTTTATAAATAGTAATGATAATT 108
                 **        ** ***  ** **  ** **** * *           **  *  * **

CCA36419.1      CCTGCTACTCCGCCAGCTACTCCTGAATAGATCACATACTGCAAAGACTGCTTGTCGATG 159
CCA39540.1      TT--CCATTTAATGGGGTACTCTAGACT--GTCAT------CAAAAA-TAAATAT--ATG 155
                    * * *      * *****  ** *   ***       **** * *   * *  ***

CCA36419.1      ACCTTGGGGTTATTTAGCTTCAAGGGCAATTTTTGGGACATTTTGGACACAGGAGACTCA 219
CCA39540.1      GTGTT---GCTATGTAGGCCGTAACACGGCTCTT----CACCT---GTAT-GGAGCATCG 204
                   **   * *** ***     *   *   * **    **  *     *  ****  **

CCA36419.1      GAAACAGACACAGAGCGTTCTGAGTCCTGGTGCTCCTGAC-GTAGGC-CTAGAACAGGAA 277
CCA39540.1      GCATC-----TAAAGCGTACTGGATTTCC-TGTTCCCATTTGTCACCATTAGCA-AGCAA 257
                * * *      * ***** ***  *     ** ***     **   *  *** * ** **

CCA36419.1      TTATTGGCT-TTATTTGTTTGTCCATTTCATAGGCTTGGGGTAATA------GATAGATG 330
CCA39540.1      TTTTTCCTTGTTATACAACAGCTCCTCTC-TAGTTTCGGTGGCATACTCAACGTTAGGGG 316
                ** **   * ****      *  * * ** ***  * ** *  ***      * ***  *

CCA36419.1      ACAGAGAAATAGAGAAGACCTAATATTTTTTGTTCATGGCAA--ATCGCGG---GTTCGC 385
CCA39540.1      TTTCAACAATGGCATCAACTGGACAACT----TTCCTGACAATAACCGCAGTAGATACAC 372
                    *  *** *     **   * *  *    *** ** ***  * *** *    * * *

CCA36419.1      --GGTCGGGTCA----CACACGGAGAAGTAATGAGAAGAGCTGGTAATCTGGGGTAAAAG 439
CCA39540.1      TTGGTCATGTCAATGTCGTACTTGTAAGTTCTTCGG-GAACCGTCAATACGTTCTTC--- 428
                  ****  ****    *  **    ****  *  *  ** * *  ***  *   *

CCA36419.1      GGTTCAAAAGAAGGTCGCCTGGTAGGGATGCAATACAAGGTTGTCTTGGAGTTTACATTG 499
CCA39540.1      GGCTTCGATGGTGATAGCCTGA-------GCAGGACAAA-CTGCTTCGCA---TAACTTA 477
                ** *   * *  * * *****        ***  ****   **  * * *   **  **

CCA36419.1      ACCAGATGATTTGGC-TTTTTCTCTGTTCAATTCACATTTT-TCAGCGAGAATC------ 551
CCA39540.1      --CAGGCAATGCAACGTTCTTCTCCG---GATGGATATCTTCTCAAAGCGTGTTCTCCTC 532
                  ***   **    * ** ***** *    **  * ** ** ***  * *  *

CCA36419.1      -GGATTGACGGAGAAA-TGGC----------GGGGT-GTGGGGTGGATAGATGGCAGAA 597
CCA39540.1      TGAAACGG-GGAGAAACTGGACCTTTTTCAAAAGGGTAGTAGATTGTATATGGGGCACGA 591
                 * *  *  ******* ***             **** ** *  ** ***   ****  *

CCA36419.1      ATGCTCGCAATCACCGCGA----AAG------AAAGACTTTATGGAATAGA-ACTACTGG 646
CCA39540.1      AAGTACATTTCCAAAGTGATATAAAGGCCTCTAAAGATCTCAGACAGTAGGTACCATTTG 651
                * *  *     **  * **    ***      *****  * *   * ***  ** * * *

CCA36419.1      GTGG---TGTAAGGA----TTACATAGCTAG--TCCAATGGAGTCCGTTGGAAAGGTAAG 697
CCA39540.1      GTGGCCTTGGAAAGAGCGCTTTCACTGCTTTCTTCCCACGTCTTTGGTCTTGGGAGTCTG 711
                ****   ** ** **    ** **  ***    *** * *   *  **       **  *
```

```
CCA36419.1      AA--------GAAG-----CTAAAACCGGCTAAGTAACTAGGGAAGAATGATCAGACTTT 744
CCA39540.1      AATCCTTCTGGATGTGTGCCTTGGATCGAACCAGAAGTTAAAGCC-AATG--TGGGTTTC 768
                **        ** *     **   * **    ** *  **  *   ****    *  **

CCA36419.1      GATTTGATGAGGTCTGAAAATACTCTGCTG-CT--TTTTCAGTTGCTTTTTCCCTGCAAC 801
CCA39540.1      TGTGCAATGAGTCCCGTCGAAAACCTGCAAACTGGTATTCTGTT--TATGGCACGGCAG- 825
                  *   *****  * *   * *  ****   **  * *** ***  * *  * * ***

CCA36419.1      CTATCATTTTCCTTTTCATAAGCCTGCCTTTTCTGTTTTCACTTATATGAGTTCCGCCGA 861
CCA39540.1      --GAAGGCTTCAATGCTAGTGGTCTAAACATTGTGTTATAAAT-GTGTAGGT-------A 875
                        ***  *   *   * **     ** **** * * *  * *  **       *

CCA36419.1      GACTTCCCCAAATTCTCTCCTGGAACATTCTCTA-TCG--CTCTCCTTCCAAGTTGCGCC 918
CCA39540.1      GAAACCTATAAGTT-TGTGTTGGGG--TTCGATAATCAAACTTGCCTTGAAGGTAAT--- 929
                **   *   ** ** * *  ***    ***  ** **   **  ****  * **

CCA36419.1      CCCTGGC--ACTGCCTAGTAATATTACCACGCGACTTA-TATTCAGTTCCACAATTTCCA 975
CCA39540.1      ---TGGCTAGGTGCTTGTTGAG--TGTCGCACGTCTCGGTATTT--TTTTAC--TTGTCG 980
                   ****    *** *  * *   *  * * ** **   ****   **  **  **  *

CCA36419.1      GTGTTCGTAGCAAATATCATCAGCC 1000
CCA39540.1      ATTCTCTC--TTACTGTC-CAAG-- 1000
                 *  **      * * **   **
```

Alignment 7: Multiple sequence alignment of the 1000bp long downstream regions of *PpOCH1* and *PpHOC1*.

```
CLUSTAL 2.1 multiple sequence alignment


CCA36419.1      -AGAAA-----GCTAGAGTAAAATAG-ATATAGCGAGATTAGA-GAATGAATACCTTCTT 52
CCA39540.1      GAGAATTAATCGCCAAAATCACAGAGGATACACTGCAACGAGCCGAGTCAA-ACTCACT- 58
                 ****      ** * * * * * ** *** *  *  *  **  ** * ** **   **

CCA36419.1      CTAAGCGATCGTCCGTCATCATAGAATATCATGGACTGT---ATAGTTTTTTTTTTGTAC 109
CCA39540.1      ----GGAACTAGCTGACATTAGCGAAGAAGGCGGCCTGTCTGATAAGAATTTGTCCAT-T 113
                    *  *    * * *** *  *** *    ** ****   ***    *** *   *

CCA36419.1      ATATAATGATTAAACGGTCATCCAACATCTCGTTGACAGATCTCTCAGTACGCGAAATCC 169
CCA39540.1      ATGCAATG----GACGGGTA-CTGGTATTT---TTACAGATGCCATATTTACCTATTTTA 165
                **  ****     ****  * *    ** *   * ******  *  * *   * *  *

CCA36419.1      CTGACTA---TCAAAGCAAGAACCGATGAAGAAAAAAACAACAGTAACCCAAACACCACA 226
CCA39540.1      ATGACTACATTCAAAGTA-GTATCTAT----------ACCAAAGTTACTTGGA------- 207
                 ******   ****** * * * * **          ** * *** **    *

CCA36419.1      ACAAACACTTTATCTTCTCCCCCCCCAACACCAATCATCAAAGAGATGTCGGAACCAAAC 286
CCA39540.1      AAGAATTCTCCAAATTGAGAAAGCCCAA----GCTTGTCAGTGATGTATTGGTACTG--- 260
                *  **  **  *  **       *****      *  ***  **  * * ** **

CCA36419.1      ACCAAGAAGCAAAAACTAACCCCATATAAAAACATCCTGGTAGATAAT--GCTGGTAACC 344
CCA39540.1      -CCGATTATCAGCTTCTCGGCCGGTGCAGG-------TAGTGGAAAATCGACTG--AACT 310
                 ** *  * **    **   **  *  *         * ** ** ***   ***  ***

CCA36419.1      ---CGCTCTCCTTCCAT--ATTCTGGGCTACTTCACGAAGTC---TGAC--CGGTCTCAG 394
CCA39540.1      GAACGATCCCTTAGCATTCGTACAACATTATTTTGAAAGATTACATAACGACAACCACTA 370
                   ** ** * *  ***   * *     ** **    *  *    * **  *   * *

CCA36419.1      TTGATCAACATGATCCTCGAAATGGGTGGCAAGATCGTTCCAGACCTGCCTCCTCTGGTA 454
CCA39540.1      AGGGTCAGAACCAT--TTAGATTGTCTGGATCTATCATTATGGCCTTGTTT------ATA 422
                  * ***  *  **  *   * **  ***    *** **   * * **  *       **

CCA36419.1      GATGGAGTGTTGTTTTT--GACAGGGGATTACAAGTCTATTGATGAAGATACC-CTAA-- 509
CCA39540.1      GACAAAGAATTGTATCCTGGACTGAAGGG---AAGTTTATAGAGTAATACCCCTCTGACA 479
                **   **  **** *    *** *  *     **** *** **  ** *  ** ** *
```

```
CCA36419.1      AGCAACTGGG--GGACGTTCCAATATACAGAGACTCCTTCATCTACCAGTGTTTTGTGCA 567
CCA39540.1      ACCAACTCGAATGGGTATTTTGAGATA-----ATTTCC-CATATATTATTGTCTTCACTG 533
                * ***** *   **   **   * ***     * * *  *** **  * *** **

CCA36419.1      CAAGACATCTCTTCCCATTGACACTTTCCGAATTGACAAGAACGTCGACTTGGC------ 621
CCA39540.1      GGAGAC-TTACTTCTTCTTTGTAGTTCCGGTATCGGCAGCAAACTTATGGTGGTCACATC 592
                  **** *  ****   **   * ** * * ** * **  **  *     ***

CCA36419.1      -TCAAGATTT---GAT-CAATAGGGCCCTTCAAGA--------GTCTGTGGATCA----- 663
CCA39540.1      TTCCAAACTTCTCGGTATAGTAGAACTATTTGTGGTTGATACGGTCTGGGTATCGGTGAT 652
                 ** * * **   * *  * ***  *  **   *         ***** * ***

CCA36419.1      ------TGTCAC------TTCTGCCAGCACAGCTGCAGCTGCTGCTGTTGTTGTCGCTAC 711
CCA39540.1      AATGCTTGTAATAATGGATTTTGTGGGTGTAACTGAAGCTGGTG-TATTAGTGGTAGTGG 711
                      *** *       ** **   *   * *** ***** ** * **  **    *

CCA36419.1      CAACGGC--CTGTCTTCTAAACCAGACGCTCGTACTAGCAAAATACAGTTCACTCCCGAA 769
CCA39540.1      TAGTGGTGGTTGT-TACTGTGCTATCTGTAGGTTCTGGGAG--------TGGTTTCTGAG 762
                 *  **    *** * **   * *   *   ** ** * *         *   * * **

CCA36419.1      GAAGATC----GTTTTATTCT------TGACTTTGTTAGGAGAAATCCTAAACGAAGAAA 819
CCA39540.1      CCGGGTTTAAAATTTTGATCTCTGTGGTATCTTCGTCTGGA-----TTTCGACTAAAAAA 817
                   * *      ****  ***      *  *** **  ***       *  ** ** ***

CCA36419.1      CACACATCAACTGTAC-ACTGAGCTCGCTCAGCACATGAAAAACCATACGAATCATTCTA 878
CCA39540.1      TCCGCCTTGAA-ATACGAGTGGAACAGGTAAACGATTGATGGAGCCCA-GAAGCACTGTG 875
                  * * *  *   *** * **     * * * *   ***   * *  * *** ** * *

CCA36419.1      TCCGCCACAGATTTCGTCGTAATCT-------TTCCGC--TCAACTTGATTGGGTTTATG 929
CCA39540.1      TCCAAGGAAGATT-CATCAAAAGTGAAAGGAGTGCCATCTTCGTTTCCATCG---TTAAG 931
                ***     ***** * **  **          * **    **   *  ** *   *** *

CCA36419.1      ATATCGATCCATTGACCAACCAACCTCGAAAAGATGAAAACGGGAACTACATCAA-GGTA 988
CCA39540.1      ATGA-GAACTTTTGGCAAAAAAATC-------AATCAGAACGTCATCTATTTCTATGGCG 983
                **   ** *  *** * **  ** *        ** * ****  * ***  ** * **

CCA36419.1      CAAGATCT----TCCA- 1000
CCA39540.1      ACAAATCCGGGATTGAA 1000
                  * ***     *  *
```

Appendix 6: Amino acid multiple sequence alignments

Alignment 8: Multiple amino acid sequence alignment of all four *Pp*PDIs.

```
CLUSTAL 2.1 multiple sequence alignment


CCA40283.1      MQFNWNIKTVASILSALTLAQASDQEAIAPEDSHVVKLTEATFESFIT--SNPHVLAEFF 58
CCA36295.1      ----------MKLLSLALLVSLVSADTFYTPKDDVIQLNAYNFKDVVFN-SNYSSVVEFY 49
CCA36622.1      ----------MKILSALLLL-------FTLAFAEVIELTNKNFDDVVLK-SGKYTLVKFY 42
CCA40358.1      ----------MKSLLLLLLA------LCQTVFGTIWELTDKNFEKKAFGQQGMYSFVYIY 44
                           . *    *              : :*.  .*..     ..   .. ::


CCA40283.1      APWCGHCKKLGPELVSAAEILK-DNEQVKIAQIDCTEEKELCQGYEIKGYPTLKVFHG-E 116
CCA36295.1      APWCGHCQNLKNPFKKAAAVSK-DYLQVAAIDCDAAENKKLCSDYRIQGFPTIMVFRP-P 107
CCA36622.1      ADWCSHCKRMNPEYEKLAEELKPKSDLIQIAAIDANKYSKYMKVYDIDGFPTMKLFTPKD 102
CCA40358.1      SPYCNYCNEMTPQFAALADLYD--DTKLQLFQINGYANKRVSKKYEVVGFPVLKIFSS-D 101
                : :*.:*:.:       *   .     :     :    ..  . * : *:*.: :*


CCA40283.1      VEVPSDYQGQRQSQSIVSYMLKQSLPPVSEINATKDLDDTIAEAKEPVIVQVLPEDASNL 176
CCA36295.1      KFDPTSSTNRRSG-------------------------------------------- 120
CCA36622.1      ISHPIEFSGSRDSESFLNFLESTTGLKLKKKAEVNE-------------------PSLV 142
CCA40358.1      GTDMGSYTGVRGTQN-----------------------------------------FI 118
                     .  . *


CCA40283.1      ESNTTFYGVAGTLREKFTFVSTKSTDYAKKYTSDSTPAYLLVRPGEEPS-VYSGEELDET 235
CCA36295.1      AHANEVYSGARDTKSIVEFGVSRIKNYVKRVSPNNINQTLGN-SEKTQL-LLVTDKAKPS 178
CCA36622.1      QSIDDSTIDDLVGKDRFIAVTASWCGYCKRLHPEWEKLAKAFGNDDIVIGNVVTDVVEGE 202
CCA40358.1      DYIHEVTGVTPSFPD----------SYVKQPTSEELEDIIKDTKRDILV-------AFSQ 161
                              .          .* *:  .:           .


CCA40283.1      HLVHWIDIESKPLFGDIDGSTFKSYAEANIPLAYYFYENEEQRAAAADIIKPFAKEQRGK 295
CCA36295.1      ALIKSIALD---FLNDIES-------------FYYPFNDKTKKALTTRLEEYQQSFSGE 221
CCA36622.1      NIKAKYKVQSFPTILYFTAGSDEP---------IRYESPDRTVEGLVKFVNEQAGLFRDP 253
CCA40358.1      PWLRGWEFPYTNFYESLAR-----------------YYAEELDDSLTTFVRIDVSDAKN- 203
                       .       :                     :.   . .  :.       .


CCA40283.1      INFVGLDAVKFGKHAKNLNMDEEKLPLFVIHDLVSNKKFGVPQDQELTNKDVTELIEKFI 355
CCA36295.1      S------------------------------------------------------------ 222
CCA36622.1      DGTLNFNAG---------------------------------------------------- 262
CCA40358.1      ------------------------------------------------------------


CCA40283.1      AGEAEPIVKSEPIPEIQEEKVFKLVGKAHDEVVFDESK-DVLVKYYAPWCGHCKRMAPAY 414
CCA36295.1      -------ITSPSILVLHENEIHIFDGKLDKLSIS-----KFLAEFSTPLEGPLSKRG--- 267
CCA36622.1      ---LIPGVSDKLTNYIKEKDQSLLESTLDLLSNHEHIKDKFSVKYHKKVIEKLLKGENEF 319
CCA40358.1      ---------AEIVSKFQVSKTPSVFHFASYREYYDQTHKLFKEDLGPVEIVQLLEGAEDL 254
                              :: ..   .               .  .          .


CCA40283.1      EELATLYANDEDASSKVVIAKLDHTLNDVDNVDIQGYPTLILYPAGDKSNPQLYDGSRDL 474
CCA36295.1      ------------------------------------------------------------
CCA36622.1      LN------------------------------------------------------NEV 324
CCA40358.1      VT------------------------------------------------------ES 258


CCA40283.1      ESLAEFVKERGTHKVDALALRPVEEEKEAEEEAESEAD--AHDEL 517
CCA36295.1      -KFLEHIRR---------GIKPGRKAKKGKKGKQTKN----HDEL 298
CCA36622.1      ERLSKMLNTKLSANNSDSVIKRLNILRNFIEAKTESKPQLLHQEL 369
CCA40358.1      IRLSTLSDS-----------RHKDVEESGEADNSFEEYAQLREL 291
                  :                   .  .:  :           **
```

Appendix 7: Most frequently used abbreviations within this thesis

Abbreviation	Explanation
DWP	96-well footprint deep well plate
*AOX*1	alcohol oxidase gene 1
*AOX*2	alcohol oxidase gene 2
CalB	*Candida antartica* lipase B
CBS	Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands
DNA	deoxyribonucleic acid
ER	endoplasmatic reticulum
FRT	recombinase target sequence
gDNA	genomic DNA
HNL	hydrolynitrile lyase
HOC	*P. pastoris* gene coding for α-1,6 mannosyltransferase
HRP	horseradish peroxidase
MeOH	methanol
MUT	methanol utilization
Mut^-	methanol utilization minus phenotype
Mut^+	methanol utilization plus phenotype
Mut^S	methanol utilization slow phenotype
OCH	*P. pastoris* gene coding for α-1,6 mannosyltransferase
OD_{600}	optical densitiy measured at 600nm
OE-PCR	overlap extension PCR
PCR	polymerase chain reaction
PDI	Protein disulfide isomerase
ΔKU70	*P. pastoris* KU70 deletion strain
	All other terms are introduced and defined in the text.

Appendix 8: List of new designed strains

1. **PDI4_up_gen_down_pJet/KlonC (No. 3972)**
2. **PDI3_up_gen_down_pJet/KlonA (No. 3973)**
3. **PDI2_up_gen_down_pJet/Klonb (No. 3974)**
4. **PDI1_up_gen_down_pJet/KlonA (No. 3975)**
5. **HNL_wt_pPT4/KlonB (No. 3976)**
6. **HRP_C1_#0_pPKan/Klon2 (No. 3977)**
7. **CalB#1_pPKan/KlonB (No. 3978)**
8. **HNL_wt_CBS7435_F7_3 (No. 3979)**
9. **HNL_wt_ΔKU70_C10_3 (No. 3980)**
10. **CalB_ ΔKU70_E12_4 (No. 3981)**
11. **CalB_ ΔKU70_E9_1 (No. 3982)**
12. **HRP_C1_ ΔKU70_H11_C (No. 3983)**
13. **HRP_C1_ ΔKU70_D7_C (No. 3984)**
14. **PDI_4.1_FRT_cass_pJet_TOP10 F‘ (No. 6533)**
15. **PDI_4.2_FRT_cass_pJet_TOP10 F‘ (No. 6534)**
16. **PDI_4.3_FRT_cass_pJet_TOP10 F‘ (No. 6535)**
17. **PDI_4.4_FRT_cass_pJet_TOP10 F‘ (No. 6536)**
18. **OCH249_pJet_TOP10 F‘ (No. 6537)**
19. **OCH636_pJet_TOP10 F‘ (No. 6538)**