EXPRESSION OF SUGAR BEET HEMOGLOBIN IN PICHIA PASTORIS

Master Thesis, Department of Pure and Applied Biochemistry

Supervisors:
Simon Christensen
simon.christensen@tbiokem.lth.se, Tel: +46 76-3949950

Mats Larsson
mats.larsson@lantmannen.com, Tel: +46 70-5578339

Examiner:
Leif Bülow
leif.bülow@tbiokem.lth.se, Tel: +46 46-2229594

Johan Pettersson

muv14jpe@student.lu.se

Populärvetenskaplig sammanfattning

När man tänker på hemoglobin tänker man vanligtvis på det protein som finns i blodet och som hjälper till att transportera syre från lungorna ut till kroppens alla delar. Men det är inte alla som vet att varianter av hemoglobin också finns i växter, där det hjälper till med att bland.annat reglera halten kväveoxid eller transportera syre hos växten.

Hemoglobin är ett proteinkomplex med en järnatom i mitten som kan binda molekyler som exempelvis syre, men också koldioxid eller kolmonoxid. Detta examesarbete handlar framför allt om att försöka uttrycka (producera) ett växthemoglobin från sockerbeta i en jästsort som heter *Pichia pastoris.* Det attraktiva med att kunna uttrycka en växtbaserad variant av hemoglobin är möjligheten då få tillgång till en vegansk och vegetarisk järnkälla helt utan animalisk inblandning. Detta skulle göra det enklare att producera vegetabiliska livsmedel med en hög halt av järn som är ett grundämne som många veganer och vegetarianer har brist på.

I detta arbete introducerades den genen som kodar för sockerbetshemoglobin till det befintliga DNAt hos *Pichia pastoris* och genom att tillsätta metanol så startades produktionen av sockerbetshemoglobinet. Eftersom *Pichia pastoris* i normalfallet producerar egna protein också, så renades alla protein upp och analyserades i en metod som ger en översikt på storleken av de protein som finns i ett specifikt prov.

Det kunde bevisas att genen för hemoglobinet hade integrerats korrekt i genomet hos *Pichia pastoris* men man kunde inte se några tydliga tecken på att hemoglobin hade producerats i någon större utsträckning. Detta behöver dock inte betyda att det inte har producerats alls. För att driva forskningen på området vidare skulle det behövas utföras analyser som är känsligare än den metod som användes i detta arbete, för att kvantifiera den grad hemoglobin som har producerats.

Acknowledgements

Firstly, I would like to give appreciation for the Division of Pure and Applied Biochemistry where I performed my master thesis project. I felt very welcomed by the scientists, students and all the other people who work there.

I would like to thank Professor Leif Bülow for helping me finding a project at his department, and always making me feel inspired and making the world of science feel less intimidating and full of opportunity. I would also like to thank Mats Larsson for the idea for the project and helping me getting it off the ground.

I would like to my supervisor Simon Christensen for enduring a constant barrage of questions and teaching me almost all the things I know about working in a biochemistry lab, from the very basics to the more difficult stuff. I would also like to give a thank you to Johan Svensson Bonde for lending his expertise about restriction enzymes and all that crucial early information I needed to get my expression vectors completed.

Furthermore, I would like to give a big heartfelt thank you to my friends and family for supporting me and an especially large thank you to my wonderful girlfriend Linnéa who kept encouraging me and always cheered me up when things felt overwhelming. Without all of you guys, this master thesis would have felt like an insurmountable task.

And lastly, I want to give a big thank you to YOU, the reader, for downloading this document and taking some time to read it, or some parts of it. I hope you will find it interesting and can gain some bits of insight from it.

Johan Pettersson

12 june, 2022

Karlshamn

Abstract

People who don't eat meat can suffer from iron deficiency due to the lower efficiency of absorption of non-heme iron compared to heme iron. The aim of this master thesis was to explore the possibilities of producing a plant-based alternative of heme-iron by evaluating the efficiency of expression of sugar beet hemoglobin in the methylotrophic yeast *Pichia pastoris.*

The gene from sugar beet hemoglobin were inserted into three different *Pichia pastoris* expression vectors, one for intracellular expression and two for secreted expression using both the native secretion signal from *Pichia pastoris* and the α-factor signal sequence from *Saccharomyces cerevisiae*. The three expression vectors were integrated into the *Pichia pastoris* genome by homologous recombination and the successful transformants were induced with methanol. The expressed proteins were then analyzed using SDS-PAGE.

The SDS-PAGE analysis showed no clear evidence of hemoglobin production in *Pichia pastoris*; however, a more sensitive analysis would need to be performed to determine if hemoglobin was produced or not. Further studies (e.g., ELISA or northern blot) would need to be performed to establish to what degree sugar beet hemoglobin is expressed in the KM71 strain of *Pichia pastoris*.

Table of Contents

Abbreviations

AOX1 – Alcohol Oxidase Coding Gene 1
AOX2 – Alcohol Oxidase Coding Gene 2
BMGY – Buffered Glycerol-Complex Medium
BMMY – Buffered Methanol-Complex Medium
CO – Carbon Monoxide
CO_2 – Carbon Dioxide
ExPHO1 – Extracellular Expressiong, Using the PHO1 Signal Sequence
Exα – Extracellular Expression, Using the α Signal Sequence
Fe^{2+} - Ferrous Oxidation State of Iron
Fe^{3+} - Ferric Oxidation State of Iron
Hb – Hemoglobin
H_D – Distal Histidine
HIS4 – Histidinol Dehydrogenase Coding Gene
H_P – Proximal Histidine
ISK – Intracelllular Expression, Standard Kozak
K_m – Binding Constant to the Distal Hemolobin
MGY – Minimal Glycerol Medium
Mut+ - Methanol Utilization Regular
MutS – Methanol Utilization Slow
NO – Nitric Oxide
nsHb – Non-Symbiotic Hemoglobin
RD – Regeneration Dextrose Medium
RDH – Regeneration Dextrose Histidine Medium
sHB – Symbiotic Hemoglobin
trHb – Truncated Hemoglobin
YPD – Yeast Peptone Dextrose Medium

Scientific Background

In Sweden, the number of people choosing to eat a vegetarian meal instead of a meal containing meat has been increasing for the last six years. More than half of the population eats a meat-free dish at least once a week, and only 12% eat meat every day. Considering this, the demand for vegan and vegetarian alternatives for meat has increased, especially affordable basics and foods at the forefront of product development. (Axfood, 2022)

One issue that affects vegans and vegetarians, to a greater extent than meat-eaters, is iron deficiency anemia (Pawlak, 2016). Dietary iron mainly comes from two forms: heme iron and non-heme iron, where heme iron only can be obtained from foods from the animal kingdom, such as meat and fish (Moustarah, 2021). 25-30% of dietary heme iron is bioavailable for the body to absorb, compared to non-heme iron that only has 1-10 % bioavailability, this means that vegetarian would approximately need more than two times higher dietary iron intake, compared to those who eat a mixed diet (Skolmowska and Głabska, 2019). On the other hand, an abundance of free iron in the body could cause reactions that would form toxic free radicals (Imam *et al*, 2017).

If heme iron could be produced using vegan and vegetarian-friendly methods, this could potentially open a possibility to create vegetarian and vegan products that are both in the forefront of development and supply non-meat eaters with heme iron. Meat alternatives, that are more similar in both taste and look to real meat, could possibly be developed using established methods and incorporating plant-based heme iron.

One possible way to create plant-based heme iron could be to express the gene from hemoglobin from plants using a suitable expression system, such as *Pichia pastoris. Pichia pastoris* is a methylotrophic yeast that is similar to *Saccharomyces cerevisiae* but bring advantages such as easy protein purification and is easy to scale up to industrial scale. *Pichia pastoris* also has the ability to perform post-translational modifications such as O and N-linked glycosylation and the creation of disulphide bridges, which ensures correct folding of the expressed protein. (Karbalaei, 2020)

Anemia

Anemia is a health condition where the body lacks a sufficient amount of red blood cells, or that the hemoglobin that is present in the red blood cells are too low in concentration. According to WHO, 42% of children under the age of five and 40% of pregnant women worldwide suffer from anemia. WHO states that anemia is commonly caused by dietary reasons, such as low amount of iron (or other vitamins e.g. vitamins B12 and A) in the diet but other potential causes could be infectious diseases, such as HIV. (WHO, 2022)

The low concentration of hemoglobin affects the bloods' ability to bind and transport oxygen to the tissues within the body. This can lead to symptoms such as fatigue, dizziness, headaches, and tinnitus. It can also lead to psychological symptoms such as depression and difficultly with cognition. (Soundarya & Suganthi, 2016)

The treatment of anemia is largely dependent on diet, and it is important for people suffering from anemia to increase their intake of dietary iron. A combination of dietary iron and B-vitamins have been shown to be especially effective. (Soundarya & Suganthi, 2016)

Therefore, making iron more accessible for people, in the form of food or similar, could be an important step in preventing and/or treating the high number of people suffering from anemia today.

Brief history of hemoglobin

In 1840, a member of the German Biochemistry Association named Friedrich Ludwig Hünefeld, was in his laboratory pressing two glass slides with earthworm blood placed between them (Van Buren, J. 2017). When the blood had dried, Hünefeld noticed small bright red rectangular crystalline structures who was later named hemoglobin by Hoppe-Seyler in 1864. By 1870, Claude Bernard discovered its function as an oxygen carrier and in 1962 the three-dimensional structure was determined by Preutz with the use of X-ray crystallography. Preutz was later awarded the Nobel prize together with Sir John Kendew (Saha, D. 2014).

From 1962 and until today, hemoglobin has had a big impact on the development and maturation of modern medicine, and our knowledge of the molecular basis for the structure and function of hemoglobin is probably equal to none compared to molecules in the same area (Schechter, 2018).

Hemoglobin in humans

Hemoglobin is present in the red blood cells in our bloodstream (as well as in almost all other vertebrates) and it has the job of supplying the other cells in our bodies with oxygen, and to transport waste products, such as carbon dioxide (CO_2) and nitric oxide (NO), away from the cells. It is a protein consisting of four polypeptide chains, each with the potential to reversibly bind one O_2 molecule and carry it from the lungs to tissues throughout the body. On the way back to the lungs it can bring waste products such as carbon dioxide and hydrogen atoms with it. Preforming the same transport in the muscles is a similar protein called myoglobin. (Stryer *et al*, 2012)

The ability for hemoglobin and myoglobin to be able to bind ligands in this manner is due to a prosthetic group called *heme*. Heme consists of a protoporphyrin ring with an iron atom in the center, bound by non-covalent forces. The iron atom can be either in the ferrous (2+) or the ferric (3+) oxidation state, with only the ferrous being able to reversibly bind O_2, CO_2 and NO. In *Figure 1*, the structure of heme is presented. (Schechter, 2018)

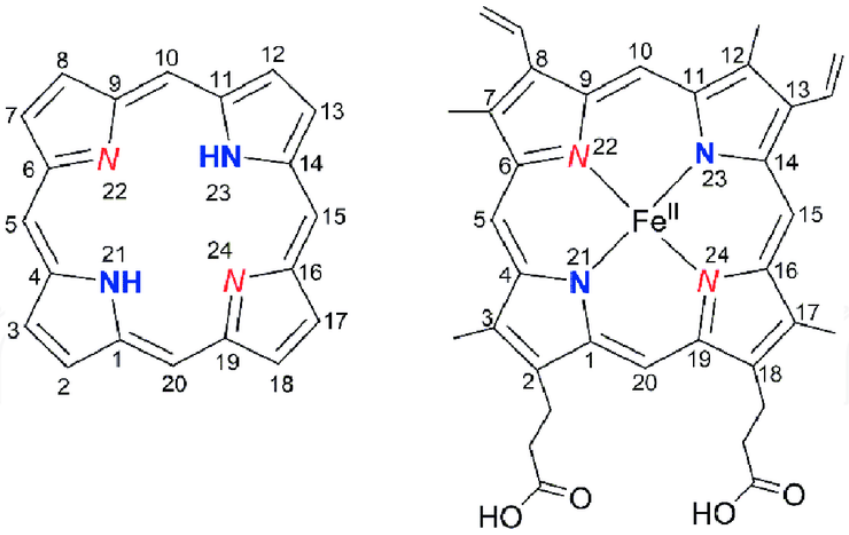


Figure 1: The most basic structure of porphyrin (left) and heme (right), The right structure shows the iron atom at the center of the porphyrin ring (Walke, 2020).

Regarding the difference between hemoglobin and myoglobin, the myoglobin protein only contains one subunit while the bigger hemoglobin contains four. Two α-globins and two β-globins, with each subunit having one big central space with a heme group in it. The alpha and beta globins subunits have a similar number of amino acids in their sequence, but the constellations of amino acids are different. When an oxygen is bound to a heme group in one of the subunits, bonding with further oxygen molecules is made easier. This is called cooperative binding and this effect causes a variance in the structure of hemoglobin, based on how many oxygens it has bound at that present moment. (Marengo-Rowe, 2006)

The iron atom in the heme prosthetic group potentially has a total of six coordination sites. The four fist coordination sites are bound with pyrrole rings from the heme group, but iron can also form two additional bonds that are called the fifth and sixth coordination sites. *Figure 2* shows the structure of the penta- and hexacoordinate states. (Gupta *et al*, 2011)

Without any ligands bound to it, myoglobin is pentacoordinate (Kakar *et al*, 2010), which means that the fifth coordination site is occupied by the imidazole ring of the *proximal histidine* and the sixth site is free, and when all six sites are occupied, it is called that the globin is *hexacoordinate*. (Gupta *et al*, 2011)

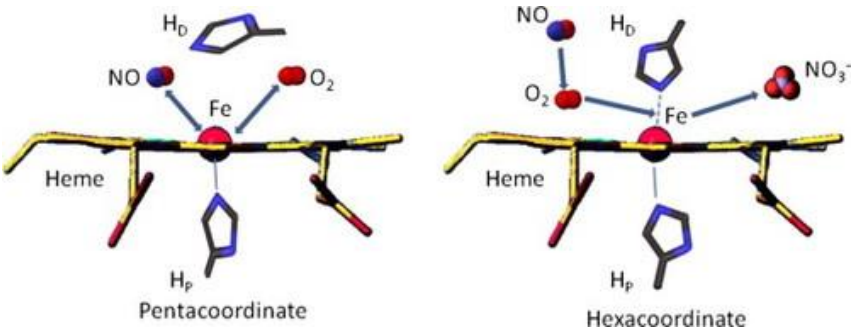


Figure 2: The figure above shows how the heme group can be both penta-and hexacoordinate. H_P = proximal histidine and H_D = distal histidine. The pentacoordinate version of hemoglobin is open for binding of O_2 and NO, while the hexacoordinate structure enables a tight binding of O_2, with the potential to accept electrons and oxygenate NO (Gupta et al, 2011).

Hemoglobin in plants

Leghemoglobin and Symbiotic hemoglobin

Hemoglobin was first discovered in plants in 1939 by Kubo (Hoy & Hargrove, 2018), almost 100 years after Hünefeld discovered the crystalline hemoglobin structures in human blood (Van Buren, J. 2017). It was found in the root nodules of legumes in symbiosis with the bacteria *Rhizobium*. This group of plant hemoglobin was therefore aptly named Leghemoglobin. *Rhizobium* aids with nitrogen fixation within the root nodule, while the hemoglobin regulates the O_2 levels for the bacterium in return (Gupta *et al*, 2011). Symbiotic hemoglobin (sHB) has since been discovered in non-legume plants, such as *Ulmaceae Parasponia andersonii* (Hoy & Hargrove, 2018). Plant hemoglobins can be placed in groups based on the coordination sites of the iron ion in the heme. Pentacoordinate hemoglobins are typically named symbiotic hemoglobin and hexacoordinate are typically named non-symbiotic hemoglobin (Gupta *et al*, 2011).

Non-symbiotic hemoglobin

When the studies of plant hemoglobin emerged, symbiotic hemoglobin was the only known type of plant hemoglobin. Back then, it was speculated that symbiotic hemoglobin originated in animals and was spread to plants by the phenomena of horizontal gene transfer (Hoy and Hargrove, 2018). This speculation was disproven in the late 1980s when a hemoglobin was found in a plant called *Trema torentosa*, a plant without any root-nodules, and thus very different from the leghemoglobin that were known before (Hill, 2012). Non-symbiotic plant hemoglobins were later found in an abundance of plants such as barley, maize, rye and many more (Hoy and Hargrove, 2018).

Classes of plant hemoglobin

The grouping of hemoglobin in accordance with their heme coordination sites is useful in some regards, but it fails to fully represent the evolutionary origins and structure of plant hemoglobins. In response to this, a new concept of "classes" was introduced, originating phylogenetic analysis. The three new classes were called class 1, class 2 and truncated (class 3) (Gupta *et al*, 2011).

By investigating the physical structure of the molecules and preforming DNA sequencing, it has been concluded that class 1 and class 2 plant hemoglobins are distantly related to animal myo- and hemoglobin. Regarding truncated hemoglobins, gene sequencing shows that they are decedents from bacterial hemoglobin, most likely spread to plants via horizontal gene transfer (Gupta *et al*, 2011).

One other key difference between class 1 and 2 versus truncated hemoglobin can be observed in the physical structure of the proteins. Hemoglobin class 1 and 2 has a structure that can be described as *"3-on-3"*, which is based on the alpha-helical "sandwich" fold, where three alpha-helixes are folded over on three other alpha-helixes, whereas truncated hemoglobin have an alpha-helical sandwich with a *"2-on-2"* fold. Hence the name truncated hemoglobin. (Gupta *et al*, 2011).

Class 1 non-symbiotic plant hemoglobin

A distinct feature of class 1 non-symbiotic plant hemoglobin is the low value of the binding constant (K_m) to the distal histidine. This leads to an equilibrium between the penta-and hexacoordinated state and furthermore leads to a tight reversible binding of O_2. This property of class 1 nsHb make it very efficient at scavenging NO at low oxygen concentrations (Gupta *et al*, 2011). It also makes class 1 nsHbs have a lower degree of hexacoordination than class 2 nsHBs (Kakar *et al*, 2010)

This process begins with an interaction with oxygenated class 1 nsHbs and NO, and in this reaction NO is converted to nitrate (NO^{3-}) and then further converted to nitrite (NO^{2-}) by nitrite reductase (Gupta *et al*, 2011). Another potential function of class 1 nsHbs could be to oxidize NADH during ATP synthesis in low oxygen environments (Hoy and Hargrove, 2008).

Structurally, class 1 nsHs is a dimer consisting of two identical subunits. Compare this to human hemoglobin, which is a tetramer and myoglobin, which is a monomer (Gupta *et al*, 2011).

Class 2 non-symbiotic plant hemoglobin

Class 2 nsHb has a higher binding constant to the distal histidine, which increases the amount of hexacoorination than that of class 1 and this, in turn, leads to a lower affinity for oxygen. Class 2's function is obscure and not a lot is known about it (Hoy and Hargrove, 2018), but it might have a function regarding oxygen storage and facilitating oxygen diffusion. (Gupta *et al*, 2011)

Pichia pastoris

Development

Pichia pastoris is a methylotropic yeast that was first isolated by A. Guilliermond in 1920 in France. He purified it from the exudates of a chestnut tree and named *Zygosacchromyces pastori*. Three decades later, in 1956, California, USA, H. Phaff took the species further and developed several new strains, isolated from the black oak tree and he gave the species the name *Pichia pastoris*, which is the name we know it by today. *P. pastoris* was then developed as a production system for single cell proteins by the Phillips Petroleum company. (PichiaGenome.org, 2021)

In the 1980's the Phillips Petroleum Company made a deal with the Salk Institute of Biotechnology/Industrial Associates to further develop the *P. pastoris* system and create expression vectors and protocols for genetic engineering. In 1993, Phillips Petroleum sold the *P. pastoris* patent to Research Corporation Technologies and licensed Invitrogen Corporation to sell and distribute the expression system (Cereghino, 2000).

Similarity to Saccharomyces cerevisiae

Both *Pichia pastoris* and *Saccharomyces cerevisiae* are yeast organisms and share similar characteristics, and several of the laboratory techniques that has been developed for *S. cerevisiae* can be used when working with *P. pastoris*. Some examples of these include

transformation by complementation, gene disruption and gene displacement. It is also believed that there is no difference in codon usage between the two different yeast types.

P. pastoris and *S. cerevisiae* share several notable genes, such as the HIS4 gene that codes for the enzyme histidinol dehydrogenase, which is an enzyme that catalyzes the production of L-histidine. The secretion factor in *S. cerevisiae* can also be used to secrete a produced protein in *P. pastoris*. (Invitrogen, 2020)

Advantages of *Pichia pastoris*

One factor that makes *P. pastoris* attractive as an expression system for recombinant proteins is that it offers many advantages compared to other eukaryotic expression systems while still showing high similarity to them. (Karbalaei, 2020).

It is a cheap expression system but has relatively quick expression times with fast co- and post-translational processing. It simplifies the purification process a fair bit by being able to secrete the produced protein directly to the media (using its native signal sequence or a secretion signal from *Saccharomyces cerevisiae*) and producing a low amount of endogenous proteins (Karbalaei, 2020). This also make it possible to harvest protein without destroying the yeast cells themselves (Li, P *et al* 2007).

Pichia pastoris is also a good option for expressing therapeutic proteins because of the ability to perform post-translational modifications such as O and N-linked glycosylation and the creation of disulphide bridges. This is essential for the protein to obtain correct folding, stability, and activity. With the help of industrial bioreactors, the production process is easily scaled up to industrial scale (Karbalaei, 2020).

The Aox1 and Aox2 promotors

As previously mentioned, *P. pastoris* is a methylotropic yeast. This means that it can use methanol as its sole carbon source. The metabolism of methanol is catalyzed by an enzyme called alcohol oxidase and produced by the coding genes AOX1 and AOX2 and is pictured in *Figure 3*. (Karbalaei, 2020)

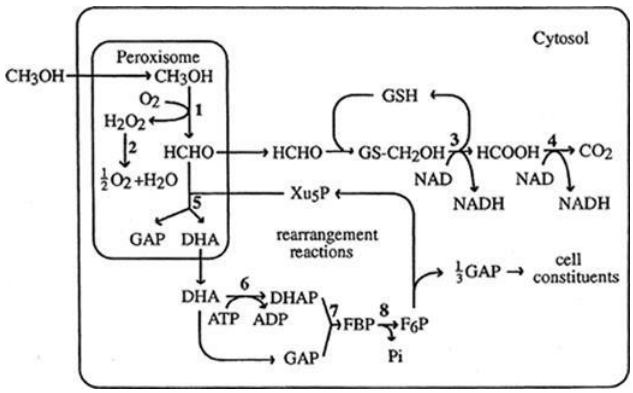


Figure 3: Methanol metabolism in P. pastoris. This flowchart shows the breakdown of methanol in the peroxisome and the metabolism of the byproducts in the cytosol. (Cereghino, 2000).

When subjected to methanol, the transcription of these genes is induced, and alcohol oxidase is produced. The main difference of the two genes is that AOX1 produces significantly more

enzyme than AOX2, so one way to regulate the production of alcohol oxidase is to knock out AOX1 or AOX2. When AOX1 is knocked out, the growth of *P. pastoris* on methanol is slowed significantly. This variation is commonly referred to as a "methanol utilization slow" (MutS). If only AOX2 is disrupted, the growth is not significantly reduced (This is referred to as Mut+), and a disruption of both AOX genes will prevent *P. pastoris* from growing with a methanol carbon source. A complete knockout is referred to as Mut-. (Karbalaei, 2020)

KM71

The *P. pastoris* strain used in this project is called KM71. It, along with the strain GS115, is included in the Invitrogen Original Expression kit (2020). Both GS115 and KM71 has a mutation in the histidinol dehydrogenase(*his4*) gene, which does not allow them to grow in the absence of histidine. This selective pressure is used to confirm integration of expression vectors in the DNA of *P. pastoris*, since the vectors contain the full *HIS4* gene.

The difference between GS115 and KM71 is that KM71 has a disrupted AOX1 gene. This makes it KM71 MutS His- strain, whereas GS115 is a Mut+ His- strain. It is difficult to predict beforehand what phenotype of strain will better express a certain protein, but in the context of this master thesis work, KM71 was chosen based on the benefit of reduced amount of screening (Invitrogen, 2020). This made for a more effective and concentrated workflow. Also, for industrial applications, there can be a benefit from using a MutS strain in the regard that it uses less amounts of methanol, which can be considered a fire hazard when stored in large amounts.

Intracellular and secreted expression

Proteins produced by *Pichia pastoris* can either be produced intracellularly or secreted. *P. pastoris* produces a low amount of native proteins, so secreting an expressed protein could be a valuable step in reducing the amount of work in the purification process. (Cereghino, 2000)

The native secretion signal of *P. pastoris* is called the *Pichia pastoris acid phosphatase (PHO1)* and is included in the pHIL-S1 expression vector from the P. Pastoris expression kit from Invitrogen. Another vector included with the kit is the pPIC9 vector, which includes the α-factor: The secretion signal of *S. cerevisiae*. (Invitrogen, 2020)

General experimental outline

In general, the expression of a gene in *Pichia pastoris* has three distinct steps: a) clone the gene into a *Pichia pastoris* expression vector, b) transform the gene into the genomal DNA of a selected host strain, and c) trial of potential expression strain. (Karbalaei, 2020)

Parent vectors and creation of expression vectors

The parent expression vectors used in this thesis are called pPIC3.5, pPIC9 and pHIL-S1. All of them share several common traits as shown in *Figure 4*.

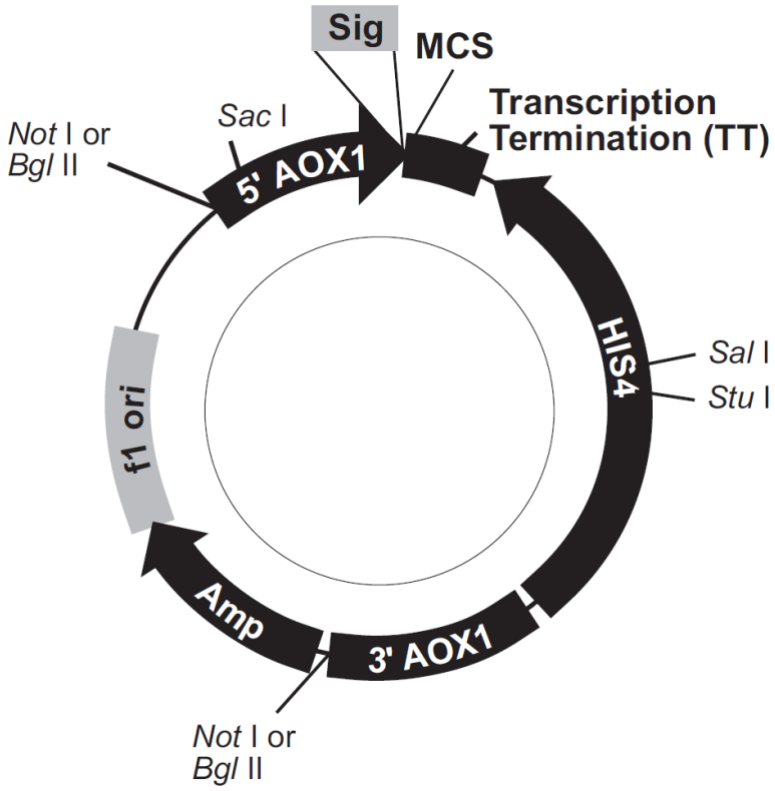


Figure 4: Generic expression vector structure (Invitrogen,2020). As shown, the vector contains a HIS4 gene and ampicillin resistance, along with fragments from the AOX1 promotor.

All vectors include fragments from the AOX1 promoter, which allows high-level induction in *P. pastoris*, it is also used as a target for plasmid integration in the AOX1 locus. They all also have the HIS4 gene which, as mentioned previously, introduces a selection marker that allows colonies with integrated expression vectors to grow in a histidine free medium. An ampicillin resistance gene is included which allows for easy multiplication of the vector by cloning, using *Escherichia coli* without a native ampicillin resistance. (Invitrogen, 2020)

pPIC3.5, pPIC9 and pHIL-S1 differ in a few regards: pPIC3.5 is used for intracellular expression and does not contain any secretion signal sequence, whereas pPIC9 contains the alpha-factor secretion signal from *Saccharomyces cerevisiae*, and pHIL-S1 contains the native *P. pastoris* PHO1 secretion signal. They also vary somewhat in the presence of different restriction sites in the multiple cloning site that is used for the insertion of the hemoglobin gene. (Appendix 1.2). (Invitrogen, 2020)

Expression vectors for hemoglobin expression were created by the use restriction digestion. A hemoglobin gene from *Beta vulgaris* (sugar beet) was inserted into each of the three parent vectors, creating two expression vectors to evaluate ability for *Pichia pastoris* to secrete sugar beet hemoglobin, and one vector for evaluation of the production of intracellular hemoglobin.

Transformation by spheroplasting

The method of transformation used in this master thesis project is transformation by spheroplasting. Yeast cells are treated with the lytic enzyme Zymolyase, the cell walls are digested by hydrolyzation of the glucose polymers, and spheroplasts are created. The

breakdown is monitored by observing the cells sensitivity to SDS (Invitrogen, 2020). The spheroplasts are then incubated with linearized expression vector. The gene of interest is then integrated into the DNA of the host strain by homologous recombination, and the very fragile transformed spheroplasts are plated in top agar to protect them from lysis. The transformed *Pichia pastoris* cells can then proliferate and the colonies can be picked for expression trials (White, 1994).

Aim of thesis

The aim of the thesis is to insert the hemoglobin gene from sugar beet (*Beta vulgaris)* into the genomic DNA of the methylotrophic yeast *Pichia pastoris*, evaluate the level of expression and further the knowledge about the initial level of expression of a foreign protein in *Pichia pastoris.*

The results of this master thesis could also be used to develop processes to produce plant hemoglobin that could be used as food additives and play a role in the development of vegan and vegetarian meat substitutes.

Methods

In silico design and amplification of hemoglobin gene inserts

The first step in designing the insert sequence was to acquire a suitable hemoglobin gene. For this purpose, plant hemoglobin from sugar beet (*Beta vulgaris*) was selected based on previous experience by the Department of Pure and Applied Biochemistry. The sequences were received, and codon optimized for *Pichia pastoris* (Kazusa, 2021). Suitable restriction sites were also added to the end of the sequences to make them compatible for ligation with the vectors of choice (Invitrogen, 2020).

For pPIC3.5 and pPIC9 the flanking restriction sites *Eco*RI and *Not*I were added, and for pHIL-S1, *Eco*RI and *Bam*HI were added. To then be able to amplify the gene sequences with PCR, primer sequences were generated using the Oligo Analysis tool from Eurofins (2021) and added to the ends of the flanking restriction sites. This gives the additional benefit of giving some extra base pairs at the end of the restriction sites so that the restriction enzymes can cleave efficiently. A Kozak sequence of AAAA was added before the gene for the intracellular expression based on the lab manual from Invitrogen (Invitrogen, 2020). The three insert vectors where named ISK (Intracellular expression), Exα (extracellular expression, α-factor signal sequence) and ExPHO1 (extracellular expression, PHO1 signal sequence).The sequences for the extracellular inserts are identical apart from the flanking restriction sites. The signal sequences are integrated into the expression vectors, therefore the suffixes α and PHO1 are derived from the destination of the insert rather than the content of the insert.

The full insert sequences were synthesized, along with matching primers, by Integrated Technologies and can be found in appendix A along with maps of the expression vectors (appendix B)

All three inserts were amplified using Phusion™ High-Fidelity DNA Polymerase from Thermo Scientific using a master-mix kit and standard protocol (Thermo Scientific, 2021), with annealing temperatures of 52 °C for both the forward primer and reverse primer (appendix A). The full primer specification sheets can be found in appendix C. The inserts were collected and purified from the PCR reaction mixture with the NucleoSpin Gel and PCR Clean-up kit from Macheray-Nagel using the standard protocol (Macherey-Nagel, 2021). The concentration of the inserts was measured spectrophotometrically with NanoDrop™.

Transformation, cloning and harvesting parent vectors in *Escherichia coli*

The three parent vectors (pPIC3.5, pPIC9 and pHIL-S1) were acquired from Invitrogen with the Original Pichia expression kit and transformed into competent *E. coli* using heat-shock. The heat-shock protocol can be found in appendix D.

The parent vectors where then cloned by picking one colony (for every vector), suspending it in 5 ml LB-medium, containing 100 µg/mL ampicillin, and grown overnight (16-18h) in a shaking incubator at 150 rpm and a temperature of 38 °C. The plasmids where then purified using a E.Z.N.A® plasmid purification kit from Omega BIO-TEK, using 50 µl elution buffer and standard protocol (Omega BIO-TEK, 2019). The concentrations of the vectors were measured spectrophotometrically using NanoDrop™.

Creation of expression vectors

The optimal digestion time for each restriction enzyme was determined by using FastDigest restriction enzymes (Thermo Scientific, 2019), by preforming repeated digestions and studying the resulting fragments on an agarose gel. It was important to ensure near-complete to complete digestion of the restriction sites for both the vectors and the inserts, to make the ligation process as effective as possible. The incubation duration was assumed to be identical for both vectors and inserts.

The determination of optimal digestion times for the *Eco*RI and *Not*I restriction enzymes (Thermo Scientific, 2019) was initiated by preforming two digestions on the parent vector pPIC9 with incubation times of 10 and 40 minutes respectively. The resulting fragments were run on a 1% agarose gel. Based on the results from the first digestion, the digestion was repeated with digestion times of 9 minutes for *Eco*RI and 20 minutes for *Not*I. The undigested plasmid generally presents as three bands on the agarose gel: Nicked circular, linear and supercoiled plasmid, with the supercoiled plasmid migrating faster in the agarose gel due to its compactness. The different plasmid forms are shown in *Figure 5* (ThermoFisher SCIENTIFIC, n.d).

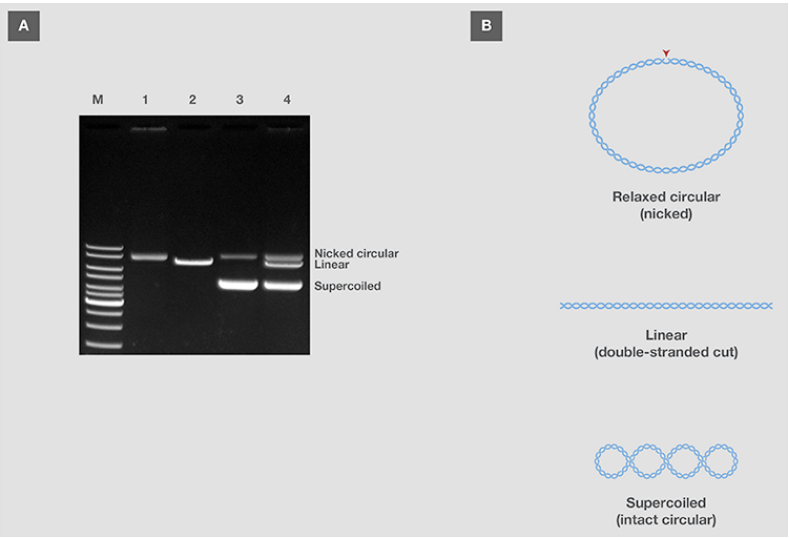


Figure 5: Different types of plasmid conformations move at different gels in an agarose gel. Supercoiled plasmid moves the fastest followed by linear plasmid and nicked circular plasmid (ThermoFisher SCIENTIFIC, n.d).

The optimal digestion time was determined for *Bam*HI and pHIL-S1 aswell. The incubation times trialed was times in both ends of the spectrum specified in the fast digest protocol: 5 and 15 minutes (Thermo Scientific, 2019).

The vectors and inserts were then digested with the determined optimal incubation time, and the restriction enzyme was then heat inactivated by placing the tubes in a heat block at 80 °C for five minutes. The vectors (not inserts) were dephosphorylated using FastAp™ (Thermo Scientific, 2020) to prevent self-ligation.

The digested DNA was then run on an 0.7% agarose gel and purified from the gel with the the NucleoSpin Gel and PCR Clean-up kit from Macheray-Nagel (2021). This was to make sure that minimal amount of undigested vector and insert were carried on to the ligation step.

The concentrations of purified digested vector- and insert DNA was measured, yet again, with the NanoDrop™.

The vectors and inserts were ligated together using T4 DNA ligase from Thermo Scientific (2012) with a ratio of 1:5 (30 ng and 150 ng) vector to insert; pPIC3.5 was ligated together with the ISK insert, pPIC9 with the Exα insert, and pHIL-S1 with the ExPHO1 insert. The DNA was incubated along with ligase, buffer, and water for two hours instead of the 10 minutes stated in the protocol for T4 DNA ligase (Thermo Scientific, 2012).

10 µl of the ligation mixture was transformed into *E. coli* with the same protocol that was used to transform the parent vectors (appendix D). Successful transformants were then grown overnight in LB-medium. The plasmids were then purified using the E.Z.N.A® kit (Omega BIO-TEK, 2019) and 400-500 ng DNA was sent for sequencing to Euofins to ensure correct ligation.

Linearizing and purifying expression vectors

To stimulate integration of the expression vector into the genomic DNA of *Pichia pastoris*, the circular vectors (both expression vectors and parent vectors) were cleaved at the HIS4 site by the FastDigest restriction enzyme SalI (Thermo Scientific, 2019). SalI will cleave the expression vector at the HIS4 gene and generate His+, MutS transformants in KM71 by integrating in the HIS4 site in the genomal DNA of *Pichia pastoris* (Invitrogen 2020).

Transformation with spheroplasting and plasmid linearization is a procedure with a high success rate regarding how many HIS+ transformant carry the insert. According to Cereghino and Cregg (2000), a HIS+ transformant has a 50-80% chance to have the DNA insert integrated into its DNA. The remainder of the HIS+ transformants that doesn't carry the insert will have undergone a gene transformation event where only the selective marker has been integrated. (Cereghino, 2000)

The vectors and parent vectors were linearized by the restriction enzyme SalI using a 15-minute digestion time. After the vectors were linearized by SalI, they were purified from the FastDigest buffer (Thermo Scientific, 2019) using the the NucleoSpin Gel and PCR Clean-up kit from Macheray-Nagel (2021). The DNA was resuspended in 70 µl of TE buffer according to the instructions from the instructions from the Original *Pichia* Expression Kit (Invitrogen, 2020).

Generating spheroplasts for transformation

The following methods are largely based on the instructions given in the *Pichia Expression Kit USER GUIDE* (Invitrogen, 2020).

Pichia pastoris of the strain KM71 was received in the Original *Pichia* Expression Kit (2020) from Invitrogen and plated on YPD plates and incubated for two days at 30 °C. One colony was then picked and inoculated in a 100 mL baffled shake flask with 10 ml liquid YPD medium overnight at 30°C at 300 rpm. The contents of YPD can be found in appendix E, along with the contents of the reagents used in appendix F.

Three 500 mL baffled flasks with 200 mL YPD were each inoculated with 5, 10 and 20 µl of the inoculum solution from the previous step. These flasks were grown overnight and one

flask with an OD600 of 0.2-0.3 was selected and the cells from that flask were harvested. The different inoculation volumes were to ensure that at least one bottle would have an OD600 of 0.2-0.3 in the morning after incubation.

The cells were washed first in water, then in SED (19 ml SE + 1 ml DTT), and lastly in 1M sorbitol. They were then resuspended in SCE buffer and divided into two 50 mL tubes of 10 mL cell solution each.

7.5 µl Zymolyase were added to one of the tubes and it was then incubated at 30°C for 50 minutes. Samples were taken at 0, 4, 8, 12, 15, 20, 25, 30, 35, 40, 45, 50 minutes and the OD800 was read. The point was to determine a suitable incubation time to achieve 70% spheroplasting of the yeast cells using the equation presented below:

$$(1)\ \%\ Spheroplasting = 100 - \left(\frac{OD800\ at\ time\ t}{OD800\ at\ time\ 0}\right) x\ 100$$

This trial was repeated many times due to varying success, but a general incubation duration where 70% spheroplasting had been achieved could be determined for the specific Zymolyase batch that was used.

7.5 µl Zymolyase were added to the second tube and incubated for the time calculated with equation 1. The cells were then washed in 1M sorbitol and CaS, and then resuspended in 0.6 mL CaS.

Transformation of Pichia pastoris

To have a negative control for expression, the three vectors carrying the hemoglobin genes and their respective parent vectors were linearized and prepared for transformation.

100 µl of the spheroplasted cells were added to six 15 mL falcon tubes. One for each transformant; pPIC3.5, pPIC9, pHIL-S1 (parent vectors), ISK, Exα and ExPHO1 (parent vectors with integrated inserts). 5-10 µl purified linearized DNA was added to every tube and incubated for 10 minutes.

The transformation solutions were washed with 1:1 PEG/CaT solution and resuspended in SOS medium and 1M sorbitol. 100-300 µl spheroplast solution were then mixed with 10 mL of molten RD agarose and poured on RDB plates. The spheroplast solution, without transformants, and a sample with only purified expression vector was also plated on RDHB plates to check for viability control and potential contaminants. These were then incubated for 4-6 days, until HIS+ transformants could be observed. The contents of RD and RDH top agarose, and RDB and RDHB plates can be found in appendix G.

Expression

One colony from a plate of successful transformants was picked from the top agarose and added to 100 mL BMGY in a 1L baffled flask, it was then grown at 30 °C and 300 rpm until an OD600 of 2-6 was reached. The cells were harvested and washed, then added to 20 mL of BMMY medium in a 100 mL baffled flask and covered with heat sterilized aluminum foil. This was done for all six transformants. The contents of BMGY and BMMY can be found in appendix H.

Methanol, to a concentration of 0.5%, was added every 24 hours during expression to maintain induction, and samples from the expression cultures was taken at times: 0, 24h, 48h, 72h, 96h and 144h.

The samples taken were centrifuged at room temperature at 13 000 rpm for three minutes. When analyzing secreted expression in the case of Exα and ExPHO1 the cell pellet and supernatant was saved in separate tubes for later analysis, when analyzing intracellular expression, only the cell pellet was saved. To destroy the yeast cells in the cell pellet, to be able to analyze the intracellular contents, the cells were lysed using acid-washed glass beads and repeatedly vortexed to cause a physical break down of the cell walls. This was done in accordance with the protocol from Invitrogen (2020) that can be found in appendix I.

The cell pellets and supernatants were run on an 4-20% SDS-PAGE gel from Bio-Rad to determine if any notable amount of plant hemoglobin had been produced (Bio-Rad, 2021). The gels were then inspected for any bands in the 19.2 kDa range, where produced hemoglobin is expected to be found (Leiva-Eriksson *et al*, 2014). The protein weight ladder used was the PageRuler Unstained Protein Ladder (Thermo Scientific, 2017)

To ensure that the plasmid had integrated in *Pichia pastoris,* a culture PCR was performed. The direct PCR screening was performed according to the protocol from Invitrogen (2020), except that a lyticase concentration of 10 U/µl was used instead of 5 U/µl and the cells were lysed for 30 minutes instead of 10 minutes. The incubation duration was decided based of a trial with 10-, 30- and 60-min incubation time and with the result loaded on a 1% agarose gel. The primer that was used was the primer synthesized for the multiplication of the insert sequences, thus binding to a sequence that was cleaved of during the creation of the expression vectors. This implicates that the primer would not anneal to the expression vectors that were harvested from the expression cultures. The protocol for the culture PCR can be found in appendix J.

Results

Creation of expression vectors

Restriction digestion of parent vectors

The resulting fragments of the first digestion of pPIC9 with *Eco*RI and *Not*I, with incubation times of 10 and 40 minutes respectively, were separated on a 1% agarose gel and can be seen in *Figure N1* in the appendix.

The undigested plasmid, in the lane next to the ladder, shows two strong bands (linear and supercoiled plasmid), while the digested plasmids show a strong band with a various number of smaller bands. The following lanes show only one strong band, which suggests that the circular plasmid has been linearized by the restriction enzymes. The weaker bands in lanes 3-8 are most certainly different forms of undigested vector.

The digestion was repeated with an incubation time of 9 minutes for *Eco*RI, and 20 minutes for *Not*I. The fragments were run on a 0.7% agarose gel and can be seen in *Figure N2*.

Figure N2 shows that the bands that represent the linearized plasmid are clearer and that the amount of uncleaved plasmid is minimal. From this, it can be suggested that an incubation time of 20 minutes for *Not*I and 9 minutes for *Eco*RI gives an optimal digestion.

pHIL-S1 was digested with *Bam*HI to determine the optimal incubation time for *Bam*HI. The incubation durations were 5 minutes and 15 minutes; the minimal and maximal recommended duration stated in the FastDigest manual (Thermo Scientific, 2019).

The 5 and 15-minute digestion times show a similar amount of digestion, and in *Figure N3* only a small amount of undigested plasmid can be detected in lane three and four. Therefore, the lowest duration of 5 minutes was chosen to minimize the star activity of the restriction enzyme.

From these experiments the optimal digestion times were determined and the results are shown in **Table 1**:

Table 1*: Optimal digestion times for the* FastDigest (Thermo Scientific, 2019) *enzymes.*

Enzyme	**Optimal digestion time (minutes)**
*Bam*HI	5
*Eco*RI	9
*Not*I	20

All three vectors were digested with their appropriate restriction enzymes: pHIL-S1 with *Eco*RI and *Bam*HI, and pPIC3.5 and pPIC 9 with *Eco*RI and *Not*I. They were then run on a 0.7% agarose gel and purified using the NucleoSpin Gel and PCR Clean-up kit from Macheray-Nagel using the standard protocol and an elution volume of 30 µl (Macherey-Nagel, 2021). The concentrations of the purified digested parent vectors can be found in **Table 2.**

***Table 2**: Concentration and purities of digested parent vectors after purification measured with NanoDrop™.*

Vector	Concentration (ng/µl)	A260/A280	A260/A230
pHIL-S1	5.964	1.714	0.128
pPIC3.5	8.946	1.636	0.818
pPIC9	8.449	1.700	0.895

PCR of inserts

The insert sequences for ISK, Exα, and ExPHO1 was amplified using Phusion™ High-Fidelity DNA Polymerase from Thermo Scientific with their master-mix kit and standard protocol (Thermo Scientific, 2021). The amplified inserts produced during the PCR reaction can be seen in *Figure N4* in the appendix.

In *Figure N4*, the three amplified inserts can be found in lanes 3-5 along with a ladder in lane one and a blank in lane two. Clear bands can be observed in the lanes with the amplified insert. The concentrations of insert sequences, after they were purified, can be found in **Table 3**.

***Table 3**: Concentrations after PCR-multiplication of the inserts measured with NanoDrop™.*

Insert	Concentration (ng/µl)	A260/280	A260/230
ISK	138	1.834	0.785
Exα	126	1.789	0.638
ExPHO1	119	1.853	0.552

The inserts were then digested with their corresponding restriction enzymes (with the same digestion times as for the parent vectors) and run on a gel. They were then purified from the gel using the NucleoSpin Gel and PCR Clean-up kit from Macheray-Nagel with the standard protocol (Macherey-Nagel, 2021). The concentrations of the purified inserts can be found in **Table 4**.

***Table 4**: Concentration and purities of digested inserts after purification measured with NanoDrop™.*

Insert	Concentration (ng/µl)	A260/A280	A260/A230
ISK	13.9	1.647	0.459
Exα	12.9	1.857	0.420
ExPHO1	16.9	1.700	0.436

Ligation and sequencing

The digested inserts and parent vectors were ligated together, and the ligation mixture was transformed into *E. coli*. Images of the transformants can be seen in *Figure 6*. The figure shows a reasonable number of colonies on every plate. All three combinations of parent vector and insert gave successful clones.

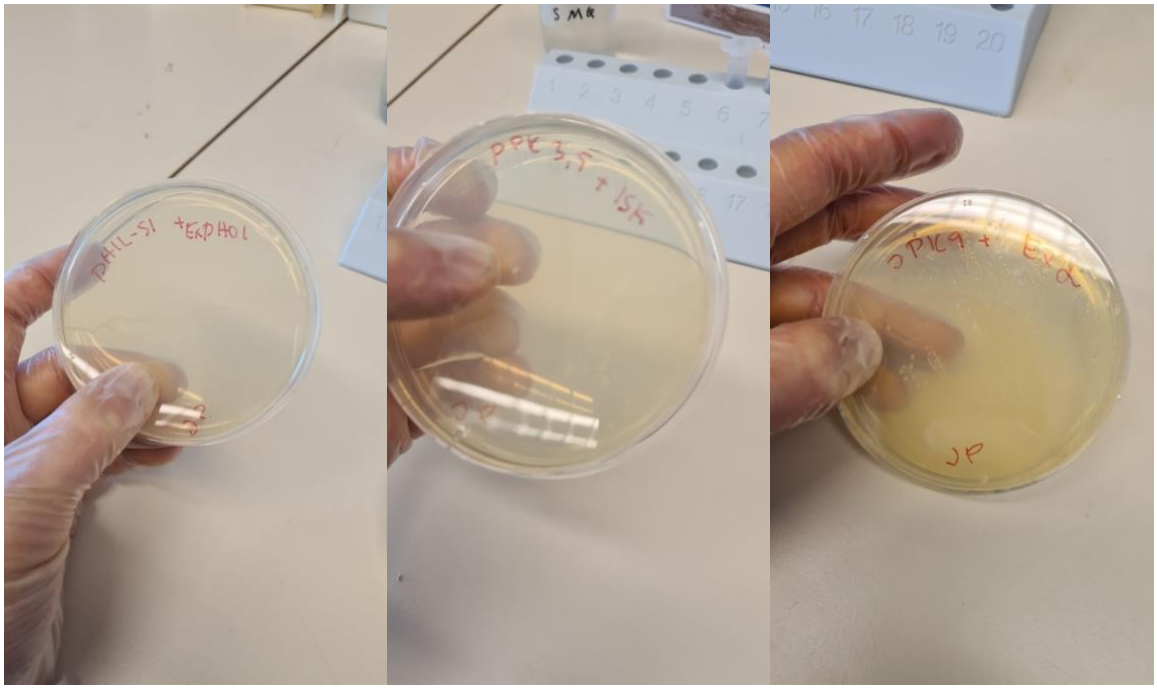


Figure 6. Images of plates with transformants of the ligation reaction between insert and corresponding parent vector. Small colonies can be seen on every plate.

Colonies from each plate were grown in LB-medium overnight and sent to Eurofins for sequencing. The alignments between the in silico designed insert sequences and the sequenced expression vectors, that was created in the previous step, were preformed using Benchling (2022) and shows an 100% alignment for all three expression vectors and their respective inserts. The full alignments can be found in *Figure K1, K2 and K3*.

Transformation of expression vectors into the genomal DNA of *Pichia pastoris*

Linearization of expression vectors

Expression vectors and parent vectors were linearized with SalI to ensure a stable integration into *Pichia pastoris.* The digestion times used were 15 minutes for all vectors and the digested vectors were run on a gel to visualize the degree of digestion and can be found in *Figure 7*. In the gel there are two lanes per vector, these two digestions are pooled together and used together during the transformation. This was done because of the maximum amount of DNA per digestion was 5 µg, whereas the transformation was recommended to performed with DNA up to 10 µg.

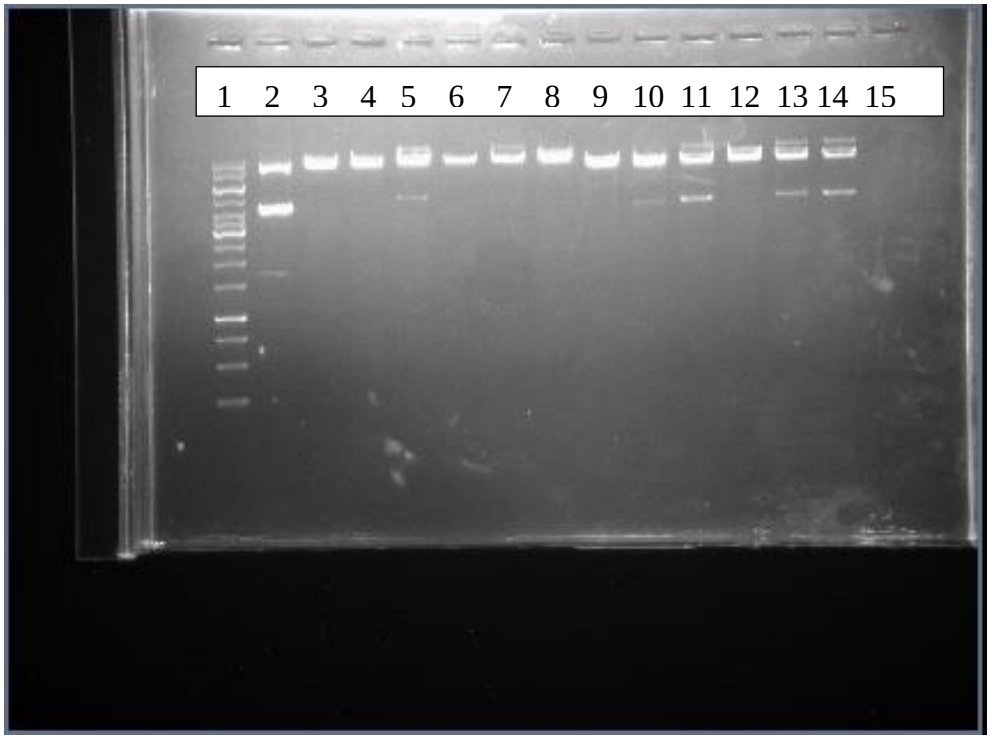


Figure 7: Linearization of expression vectors and parent vectors. The lanes from left to right represent: lane 1: ladder, lane 2: undigested pPIC3.5, lane 3-4: pPIC3.5 + ISK, lane 5-6: pPIC9 + Exα, lane 7-8: pHIL-S1 + ExPHO1, lane 9-10: pPIC3.5, lane 11-12: pPIC9, and lane 13-14: pHIL-S1.

Most lanes show a complete- to almost complete digestion. One digestion of pPIC9 + Exα, one digestion of pPIC3.5, one digestion of pPIC9 and both digestions of pHIL-S1 shows that some of the plasmid remains circular. This degree of digestion was decided to be sufficient for transformation into *Pichia pastoris.*

Spheroplasting trials

Several different spheroplasting trials were done to determine the best incubation duration to achieve 70% spheroplasting. These results, represented in graph form, are presented below. On the y-axis, the OD800 is presented, and the x axis presents the duration of incubation where the cell sample was taken.

Figure 8 shows the first spheroplasting trial, the following four trials can be found in appendix L.

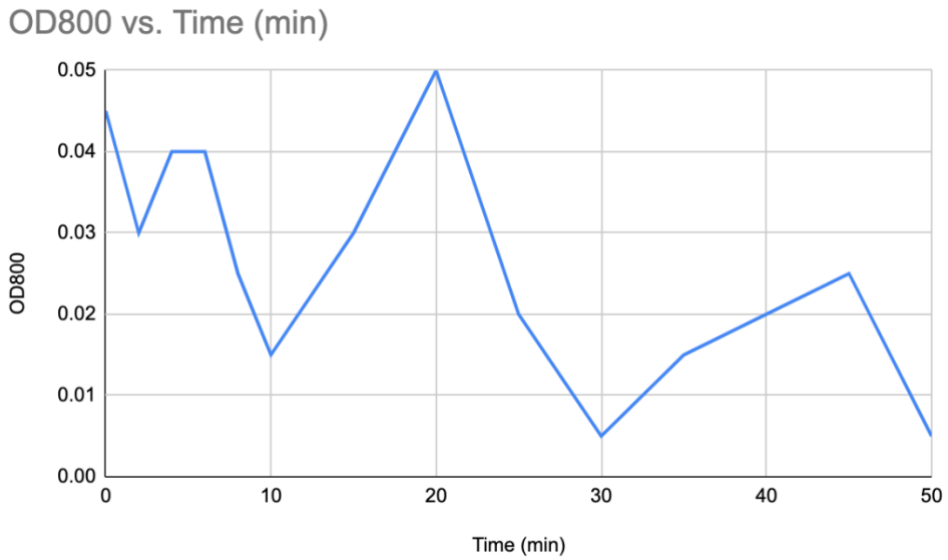


Figure 8: First spheroplasting trial. The OD600 decreases with time with some unexpected peaks and valleys.

With the help of the graph presented above and in Appendix L and equation (1), the time where 70% spheroplasting had occurred was determined to be around 30 minutes.

Results of transformations of *Pichia pastoris*

Two control plates were prepared; one for viability check of the spheroplasts, and one to check for contaminations in the purified expression vectors used for transformation.

Colonies can be observed on the spheroplasted KM71 plate indicating that the spheroplasted cells were able to proliferate and were not completely destroyed during the spheroplasting process. The negative control without cells had no colonies, which indicates that no potential colonies are from the expression vector samples. The plates of KM71 transformed with the expression vectors can be seen in Appendix M, along with pictures of the transformants containing the parent vectors. All plates show HIS+ transformants with a higher amount of successful transformants with the ExPHO1 vector, and a lower amount of transformants with the ISK and Exα vectors.

Trial of expression

Expression of ISK

When expressing ISK, the parent plasmid pPIC3.5 and three colonies (A, B and C) from the transformed KM71 were expressed. The starting OD600 of the cultures can be found in **Table 5**.

***Table 5**: Starting OD600 of expression cultures of pPIC3.5 and three ISK cultures.*

Culture	**OD600**
pPIC3.5	3.45
ISK A	2.70*
ISK B	5.69
ISK C	2.42*

**This is an approximated calculated value based on a measured OD600 and the growth rate of KM71.*

In the first gel, in *Figure 9* below, samples from pPIC3.5 and ISK A can be observed, along with a PageRuler™ with the expected size of the hemoglobin monomer marked with a red line at 19.2 kDA. This line is present at the same place in all the expression figures below. Samples from pPIC3.5 and ISK were taken at 0h, 25h, 48h, 72h, and 96h. All samples show expressed proteins at various strength and sizes that increases with the expression duration.

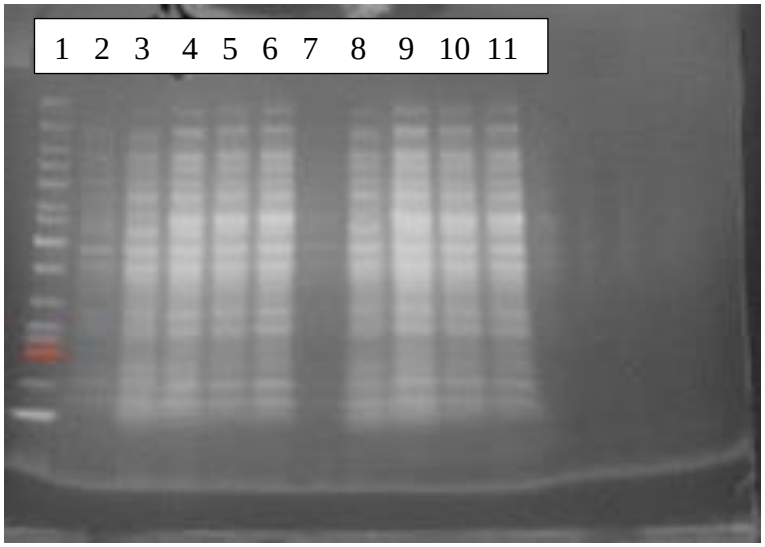


Figure 9: PageRuler™, pPIC3.5 and ISK expression culture A. Lane 1: protein ladder, lane 2: pPIC3.5 at 0h, lane 3: pPIC3.5 at 24h, lane 4: pPIC3.5 at 48h, lane 5: pPIC3.5 at 72h, lane 6: pPIC3.5 at 96h, lane 7: ISK A at 0h, lane 8: ISK A at 24h, lane 9: ISK A at 48h, lane 10: ISK A at 72h, and lane 11: ISK A at 96h.

In the second gel, presented in *Figure 10*, samples from ISK B and C can be observed, along with a PageRuler™. Samples from both cultures of ISK was taken at 0h, 25h, 48h, 72h and 96h.

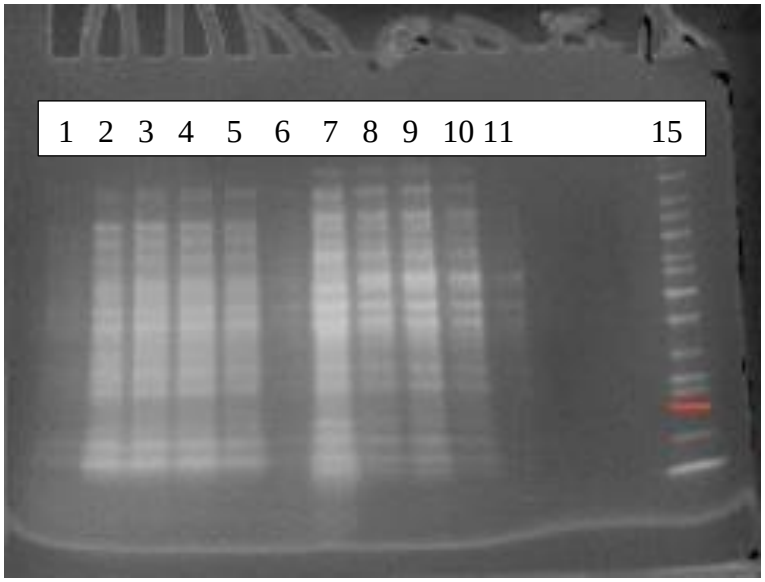


Figure 10: In this gel, samples from ISK B and ISK C are loaded in a similar fashion as for ISK gel number one. Lane 1: ISK B at 0h, lane 2: ISK B at 24h, lane 3: ISK B at 48h, lane 4: ISK B at 72h, lane 5: ISK B at 96h, lane 6: ISK C at 0h, lane 7: ISK C at 24h, lane 8: ISK C at 48h, lane 9: ISK C at 72h, lane 10: ISK C at 96h, and the last lane to the right: PageRuler™.

Expression of Exα

The results of the expression of Exα can be seen in *Figure 11* and *Figure 12* below. For this round of expression, the parent plasmid pPIC9 along with four cultures from four different colonies were expressed. The starting OD600 for all five samples can be found in **Table 6**.

***Table 6**: Starting OD600 of expression cultures of pPIC9 and* four Exα cultures.

Culture	**OD600**
pPIC9	1.96
Exα A	2.41
Exα B	3.56*
Exα C	2.53
Exα D	1.59

**This value was diluted from OD600 = 17.8 to achieve a starting OD600 of ≈ 2-6.*

Figure 11 shows, from left to right, the PageRuler™ ladder followed by sample pPIC9 at times: 0h, 24h, 48h, 72h, and 96h. Sample Exα A follows with samples taken at the same time points. In the last lane on the right is the first sample at t = 0 for sample B. Sample A shows a clear band, but not at the expected size of hemoglobin.

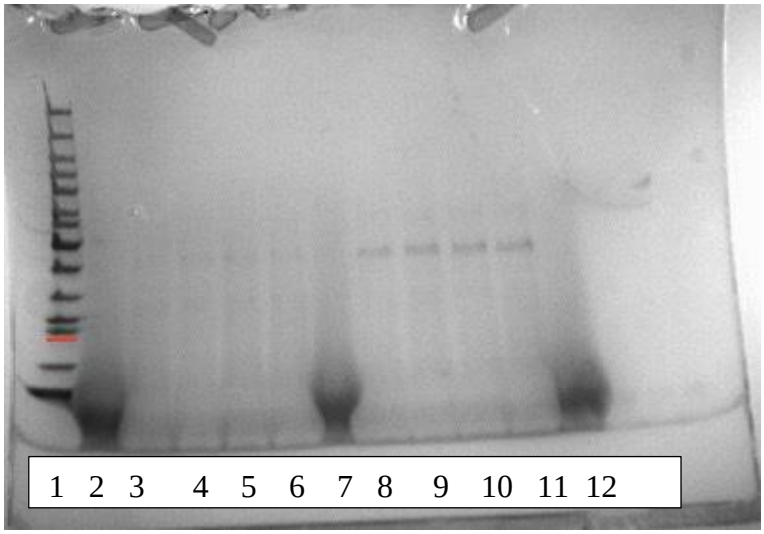


Figure 11: This gel contains samples from pPIC9 and Exα A. In lane 1: PageRuler™, lane 2: pPIC9 at 0h, lane 3: pPIC9 at 24h, lane 4: pPIC9 48h, lane 5: pPIC9 at 72h, lane 6: pPIC9 at 96h, lane 7: Exα A at 0h, lane 8: Exα A at 24h, lane 9: Exα A at 48h, lane 10: Exα A at 72h, lane 11: Exα A at 96h, and lane 12: Exα B at 0h.

In *Figure 12*, Exα B samples from t = 24h, 48h, 72h, and 96h can be observed in the first four lanes. Following in the next five lanes are extractions from sample C at the time points 0h, 24h, 48h, 72h, and 96h. The next five lanes contain the same sample points for sample D. In the last lane on the right is a PageRuler™ ladder. Sample B displays many bands, some stronger than others, with one band around the size of 19-21 kDa. Neither sample C nor sample D have any discernable bands.

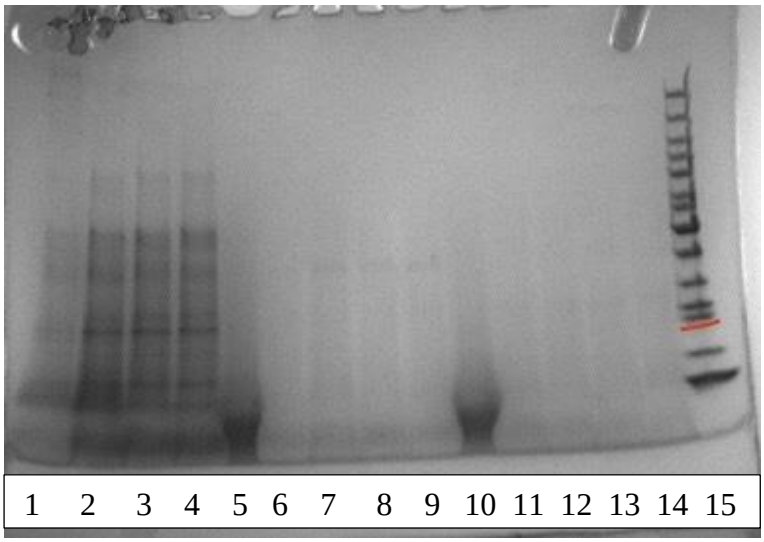


Figure 12: This gel contains samples from Exα B, C, D. In lane 1: Exα B at 24h, lane 2: Exα B at 48h, lane 3: Exα B at 72h, lane 4: Exα B at 96, lane 5: Exα C at 0h, lane 6: Exα C at 24h, lane 7: Exα C at 48h, lane 8: Exα C at 72h, lane 9: Exα C at 96h, lane 10: Exα D at 0h, lane 11: Exα D at 24h, lane 12: Exα D at 48h, lane 13: Exα D at 72h, lane 14: Exα D at 96h, and lane 15: PageRuler™.

Expression of ExPHO1

Results of the first round of expression of ExPHO1 can be seen in *Figure 13* In this expression, the parent plasmid and two cultures of ExPHO1 transformants were expressed. On the left is PageRuler™ ladder. To the right of the ladder, is the parent plasmid pHIL-S1 at 0 hours, followed by sample A and sample B at 0 hours. In the following three lanes are

pHIL-S1, ExPHO1 A and ExPHO1 B at 24 hours. Next three are samples from the 48h mark. This is then repeated for sample points at 96h. The last two lanes are samples from ExPHO1 A and ExPHO1 B at 140h. Starting OD600 for each culture can be found in **Table 7**.

***Table 7**: Starting OD600 for the first round of expression of ExPHO1.*

Culture	**Starting OD600**
pHIL-S1	5.60
ExPHO1 A	4.25
ExPHO1 B	3.75

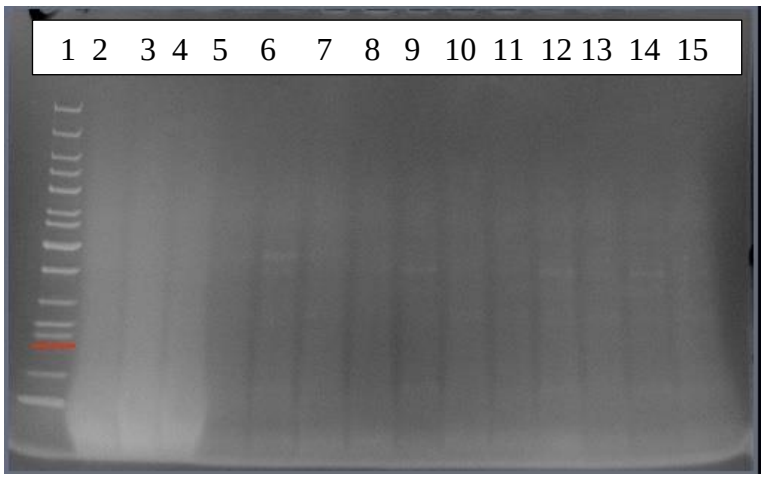


Figure 13: Expression of ExPHO1. The parent plasmid and two cultures from two colonies are expressed. Lane 1: PageRuler™, lane 2: pHIL-S1 at 0h, lane 3: ExPHO1 A at 0h, lane 4: ExPHO1 B at 0h, lane 5: pHIL-S1 at 24h, lane 6: ExPHO1 A at 24h, lane 7: ExPHO1 B at 24h, lane 8: pHIL-S1 at 48h, lane 9: ExPHO1 A at 48h, lane 10: ExPHO1 B at 48h, lane 11: pHIL-S1 at 96h, lane 12: ExPHO1 A at 96h, lane 13: ExPHO1 B at 96h, lane 14: ExPHO1 A at 140h, and lane 15: ExPHO1 B at 140h.

No apparent bands can be seen at the expected size for the hemoglobin monomer. Other clear bands can be seen in sample ExPHO1 A at all time points except for t = 0h. At t = 0h, the supernatant was not separated from the rest of the cells by means of centrifugation until after they were frozen. This would implicate the results seen in lanes two to four.

ExPHO1 was expressed for a second time, this time with the parent vector pHIL-S1, and four different EXPHO1 expression cultures were expressed. The starting OD600 of these cultures can be observed in **Table 8**:

***Table 8**: Starting OD600 for the second round of expression of ExPHO1.*

Culture	**OD600**
pHIL-S1	1.25
ExPHO1 A	2.49
ExPHO1 B	1.80
ExPHO1 C	5,45*
ExPHO1 E	2.3

** This value was diluted from an OD600 of 10.9.*

As can be seen above, the OD600 for pHIL-S1 and ExPHO1 B was not in the range of OD600 2-6.

In *Figure 14*, samples of pHIL-S1 and ExPHO1 A can be observed, along with a PageRuler™. In this expression, sampes taken at 0h, 24h, 72h and 120h were loaded on the gel. In well two to five, samples from pHIL-S1 were loaded, and in wells six to nine, samples from ExPHO1 A were loaded. In wells three to five and seven to nine, faint bands can be observed in the 10 kDa-area. The wells with the samples from t = 0h (well one and five) a smear of protein can be seen.

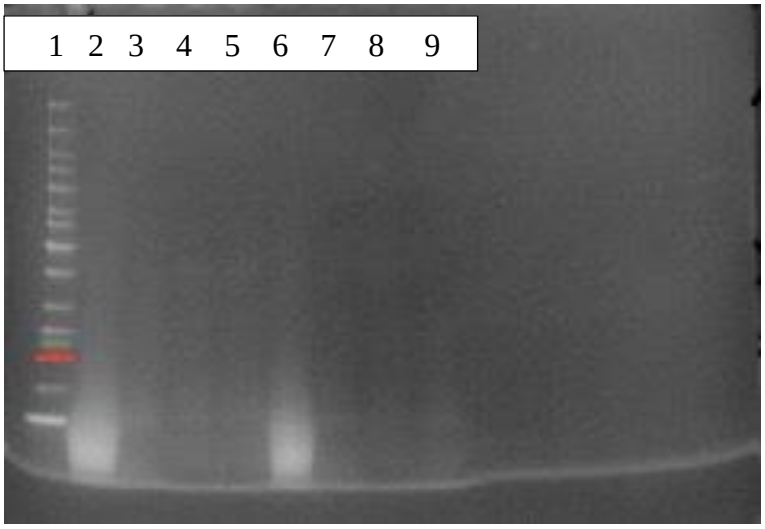


Figure 14: ExPHO1 expression round two. Lane 1: PageRuler™, lane 2: pHIL-S1 at 0h, lane 3: pHIL-S1 at 24h, lane 4: pHIL-S1 at 72h, lane 5: pHIL-S1 at 120h, lane 6: ExPHO1 A at 0h, lane 7: ExPHO1 A at 24h, lane 8: ExPHO1 A at 72h, and lane 9: ExPHO1 A at 120h.

The three remaining expression cultures were run on an additional gel that can be observed in *Figure 15*. Wells one to four contains ExPHO1 B with samples taken 0h, 24h, 72h and 120h, with the following four lanes containing ExPHO1 C at the same time points. After ExPHO1 C, the next four lanes contain ExPHO1 E at the same sample points. All wells at t = 0h show the same protein smear that can be observed in pHIL and ExPHO1 A from the previous gel. One band can be observed at about 17-19 kDa in well number six, which represents ExPHO1 C at 24h. The other wells show no clear bands

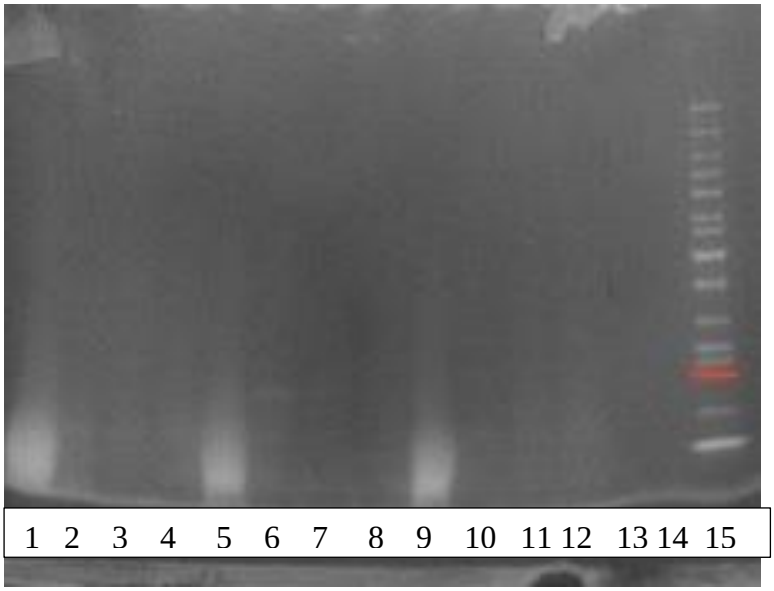


Figure 15: ExPHO1 expression round two. Lane 1: ExPHO1 B at 0h, lane 2: ExPHO1 B at 24h, lane 3: ExPHO1 B at 72h, lane 4: ExPHO1 B at 120h, lane 5: ExPHO1 C at 0h, lane 6: ExPHO1 C at 24h, lane 7: ExPHO1 C at 72h, lane 8: ExPHO1 C at 120h, lane 9: ExPHO1 E at 0h, lane 10: ExPHO1 E at 24h, lane 11: ExPHO1 E 72h, lane 12: ExPHO1 E at 120h, and lane 15: PageRuler ™.

Colony PCR of expression cultures

To test the viability of the lyticase, a trial with three different incubation times was performed. These times were: 10 minutes, 30 minutes, and 60 minutes. The result can be observed in *Figure O1* (in the appendix). On this 1% agarose gel, a GenRuler™ ladder was loaded in the first well and with an unlysed sample of pPIC3.5. The following three lanes are 10 minutes incubation, then 30 minutes and 60 minutes.

As *Figure O1* show, no streaks of DNA can be observed in the unlysed control, whereas all three incubation times show lysed cell-material. From this trial, an incubation time of 30 minutes was chosen to ensure complete cell destruction.

In *Figure 16* below, the result of the direct PCR can be observed. The samples are flanked with GenRuler™ ladders on either side. In the following four lanes, samples from ISK A, B, and C were loaded, along with their parent plasmid pPIC3.5. In the next five lanes the four expressed Exα cultures along with their parent plasmid pPIC9 were loaded. In the following five lanes, the second-round cultures of ExPHO1 were loaded along with their parent plasmid pHIL-S1. In the last three lanes, before the second ladder, controls with the insert DNA were loaded in the order of ISK, Exα and ExPHO1.

Figure 16: The agarose gel with the direct PCR. Lane 1: GenRuler™, lane 2: ISK A, lane 3: ISK B, lane 4: ISK C, lane 4: pPIC3,5, lane 5: Exα A, lane 6: Exα B, lane 7: Exα C, lane 8: Exα D, lane 9: pPIC9, lane 9: ExPHO1 A, lane 10: ExPHO1 B, lane 11: ExPHO1 C, lane 12: ExPHO1 E, lane 13: pHIL-S1, lane 14: ISK insert, lane 15: Exα, lane 16: ExPHO1 insert, and lane 17: GenRuler™.

No bands have been amplified from the cultures from the expression, but two out of three of the control samples have been amplified.

Discussion

Benefits of using hexacoordinated plant hemoglobin as an iron source

There could be potential benefits of using the more stable hexacoordinated hemoglobin from sugar beet over the pentacoordinate myoglobin variant from meat or the iron salts from iron supplements. The hemoglobin could degrade in the body and create free iron, which is reactive and could cause toxic reactions which the more stable heme iron would not. As stated, heme iron has a much higher degree of absorption than for example iron salts, and this would lead to a lower amount of free iron circulating in the body.

Therefore, an argument for introducing a stable hexacoordinated plant hemoglobin over traditional supplements could be made. This also further the importance of creating a dietary iron that is as stable as possible with a high level of absorption, reducing the risk of iron overload.

Creation of expression vectors

During the first digestion, the parent vector pPIC9 was digested with the two restriction enzymes *Eco*RI and *Not*I. This was done to create a linearized vector that would match in the ends with the insert sequence Exα, who has the restriction sites *Eco*RI and *Not*I flanking the hemoglobin gene. When preforming the digestion with an incubation time of 10 minutes per enzyme (***Figure N1***) the agarose gel shows that the lane with pPIC9 and *Eco*RI is fully digested. This conclusion can be drawn, since for the undigested vector, the DNA shows two bands. These bands represent linearized and supercoiled circular vector. For the digested vector only the band representing the linearized vector remains. From this, it can be concluded that 10 minutes is enough for complete digestion of the *Eco*RI restriction site. In the 10-minute lane with pPIC9 and *Not*I, a small amount undigested supercoiled vector remained. The trial was repeated a second time with enzyme-specific digestion times: 9 minutes for *Eco*RI and 20 minutes for *Not*I. As can be observed in ***Figure N2,*** this resulted in a complete digestion of both restriction sites. As the restriction sites are the same for pPIC9 and pPIC3.5, the same incubation duration was deemed appropriate for both parent vectors.

The third parent vector, pHIL-S1, and its corresponding insert, ExPHO1, has the flanking restriction sites *Eco*RI and *Bam*HI. To achieve a good ligation reaction, the optimal digestion time for *Bam*HI also was investigated. The tried incubation times for *Bam*HI was 5 minutes and 15 minutes, as per the instructions on the lab manual from ThermoFisher. The 5-minute incubation showed the same amount of digestion as the 15-minute incubation; therefore the 5-minute incubation was chosen because a longer incubation time could possibly create cuts in the DNA in unwanted locations. This gel image can be observed in ***Figure N3.***

The inserts were successfully amplified using Phusion™ High-Fidelity DNA Polymerase (***Figure N4)*** from Thermo Scientific and were digested using the same restriction enzymes with the same incubation duration to create matching sites at their respective ends. These digested parent plasmids and vectors was then incubated along with T4 DNA ligase to create circular vectors with the hemoglobin gene integrated into the multiple cloning site (MCS) of the parent vectors.

These circular DNA expression vectors were transformed into *E. coli* and the resulting colonies were grown in LB-medium and purified. These DNA sequences were sent for sequencing, along with primers of the respective MCS for all three expression vectors, to confirm integration of the hemoglobin genes in the sequenced DNA. The insert was found, with a 100 percent identity, in all three expression vectors multiple cloning site and can be observed in the appendix in ***Figures K1, K2*** and ***K3***. This could be determined by aligning the sequence received from Eurofins of the MCS with the hemoglobin insert sequence. The alignment shows the restoration of the restriction sites and the insertion of the hemoglobin gene. From this it can be concluded that the three expression vectors contain the hemoglobin gene, and the ligation was successful

Transformation of Pichia pastoris

Linearization of expression vectors

Cleaving the expression vectors at a site that also exist in the genomic DNA of Pichia pastoris will stimulate the integration of the vector into the DNA by homologous recombination. Figure 12 shows all expression vectors and parent vectors being digested with SalI, which causes the vector to be digested at the HIS4 gene. The time used, 15 minutes, is the maximum recommended digestion time when using FastDigest enzymes and was therefore chosen to improve the amount of DNA that would be cleaved. When 15 minutes of digestion was used, almost all vectors were completely digested, except for a few exceptions (***Figure 7)***.

It can be argued that a complete digestion of all vectors would, further down the line, increase the amount of HIS+ transformants whereas a lower amount of cleaved vectors would cause a lower amount of postitive tranformants. When observing the number of transformants received from the transformation reactions of *Pichia pastoris* KM71, the transformation with the parent vector pHIL-S1 showed one of the most successful transformations in regard to the amount of HIS+ colonies, while being one of the vectors with the highest ratio of uncleaved vector left after linearization. The transformations with the other parent vectors, pPIC3.5 and pPIC9 both show a lower number of colonies, while having a higher degree of digestion. From this it can be concluded that a higher amount of digested vector is not the only factor influencing the amount of successful transformants, and that there might be other more important factors determining the success of the transformation reactions.

Spheroplasting

With the use of Zymolyase, the cells were broken down to facilitate the integration of the expression vectors into the DNA of *Pichia pastoris.* The degree of cell wall destruction increased with the incubation duration of Zymolyase and an optimal digestion time, where 70% spheroplasting had occurred, was determined by performing repeated digestion trials with KM71 *Pichia pastoris* cells and Zymolyase.

The first digestion, observed in ***Figure 8***, had a starting OD800 value of 0.045. This would imply that 70% spheroplasting would have occurred at a OD800 of 0.0135 and at t ~ 26 minutes. The credibility of this trial seem fairly high, since the starting value is reasonable and the graph shows a general trendline downwards, which should be expected.

The fourth spheroplasting (***Figure L3)*** trial had a starting OD800 of 0.195 and a quite stable linear trajectory downwards with decreasing OD800 with increasing time. A starting OD800

of 0.195 would imply 70% speroplasting at an OD800 of 0.0585 which was reached at an incubation time of ~ 25 – 30 minutes. It can be speculated that the highest peak at t = 3 minutes could occur based on a measuring or sampling error.

In the fifth trial (***Figure L4)***, the curve is once again quite linear with a downwards trajectory. Just as the theory behind the Zymolyase and cell wall breakdown and their relation to the OD800 value suggests. This trial had a starting value of 0.105 and reached 70% spheroplasting at 25 minutes. The measurement after this point seemed not to decrease anymore but stay linear at an OD800 about 0.025.

The second and third spheroplasting trial (***Figure L1*** and ***L2)*** were deemed to be defective based on the shape of the curves generated.

In hindsight, a spheroplasting duration of 25 minutes would seem to be sufficient in comparison to the 30 minutes used during the laboratory practical. This could possibly lead to more potential transformants and lower amount of "over digested" cells.

Controls and HIS+ transformants

The viability control show that the digested cells were able to proliferate since colonies grew on the plate with only spheroplasts (***Figure M7***). This plate had a medium containing histidine, so the cells themselves do not need the HIS+ gene, that the transformed cells growing on the other plates needed. Another control with just the expression plasmid was preformed to make sure that the vector DNA did not introduce any contaminant DNA, and as expected, no colonies grew on this plate.

All plates with HIS+ colonies show a similar number of successful colonies, except for the plates with the ExPHO1 expression vector (***Figure M1, M2*** and ***M3)***. These plates show a lot more HIS+ colonies. This result suggests that the transformation with the ExPHO1 vector was for some reason much more effective than the other two expression vectors. This could be do simple technical reasons in the lab, such as uncareful handling of spheroplasted *Pichia pastoris* cells for the plates with fewer colonies. It could also be a coincidence that just that transformation worked so well. More research on this specific occurrence needs to be performed to draw a reasonable conclusion.

Expression of transformed KM71

The starting OD600 of the different expression colonies were quite different in the beginning of each expression (**Table 5,6,7** and **8**). The suggested OD600 of 2-6 in the lab manual from ThermoFisher was followed. This span is quite large, so it could be speculated that starting OD600 does not carry great importance in the efficiency of the expression. Since the expression goes on for 120 hours, the OD600 of all the samples would probably reach a similar amount of cell density after a while. This conclusion can be drawn since all cultures contain the same amount of medium with the same amount of nutrients for the cells. It can be assumed that the cells will grow until the nutrients are used up and will then maintain the same OD600. This was also the case based on ocular observation, but no OD600 measurements was taken during expression, so this statement cannot be confirmed in any way in hindsight.

When inspecting the SDS-pages for all expressions (***Figure 9-15)***, no clear bands can be found in the 19.2 kDa range for any of the expressions. One could argue that a faint band can be seen in the first ISK expression, when comparing the expression of the parent plasmid with the expression of ISK A. Due to the low resolution of the image and the high amount of bands in all of the lanes, it's hard to determine if there are any differences what so ever. These factors make this statement carry a low degree of trust.

When inspecting the last expression, ExPHO1, one can see a band that could be in the 18-19 kDa range in the 24h lane of sample C (***Figure 15)***. This band is not present in any of the following lanes. This could be because the protein in question is broken down in the cell between 24h and 48h, or that the band in the lane is there because of a sheer coincidence.

It should be stated that using an SDS-page assay as the only method of detection carries a high margin of error. If the protein is expressed, but in very low concentrations, it would be incredibly difficult to see it on the SDS-page and other methods of detection, such as ELISA, could be used instead to detect the expression of hemoglobin. Hemoglobin could very well be expressed in these experiments, but in very low concentrations.

It could be that the KM71 strain is not optimal for this certain protein and that the GS115 strain would express hemoglobin better. Possibly at the level that it could be detected using an SDS-page assay. It's also possible that with some condition optimization, KM71 also could express the hemoglobin to the level of it being detected using only SDS-page.

It could also be possible that the hemoglobin gene itself was defective and could not produce any hemoglobin, or that the match between the hemoglobin gene from sugar beet and *Pichia pastoris* is a bad match in practice.

At this stage of research, it's difficult to say anything for certain if any amount of hemoglobin is expressed or not, since a limited amount of testing was preformed to determine the level of expression. It can be said for certain, that no high levels of hemoglobin are produced using this strain with this specific strain and expression conditions.

Culture PCR

When inspecting the gel from the culture PCR (***Figure 16)***, no bands are produced for the lanes with the lysed cell culture samples and only two out of three PCR-controls containing the insert has been amplified. Only two out of three controls being amplified suggest that the PCR procedure was not 100% successful in all cases and would probably benefit from optimization. Since this procedure was only performed on one occasion, errors such as pipetting mistakes, and other random errors could play a large role in the outcome of the PCR.

It was discovered in hindsight that the primers that was used in the culture PCR were the primers synthesized for the multiplication by PCR of the insert sequences. They were designed to anneal to the sequences that was added to the insert sequences to make them multiply during the original PCR reaction (to amplify the insert sequences for the creation of the expression vectors). These sequences were subsequently cleaved off the insert sequences when creating the expression vectors and thus, are not present in expression vectors that were harvest from the expression cultures. This explains why only the insert sequences was

amplified and not even one of the expression cultures and makes the result of the colony PCR irrelevant.

It is important to know if the HIS+ colonies contained the insert or not when discussing the efficacy of the transformation reactions and the expressions, but with the inadmissible results of the PCR it is difficult to draw a well-educated conclusion about the contents of the expression cultures. If more time was allowed in this thesis project, it would benefit the conclusions from this project if more careful analysis of the expression cultures could be performed.

If the research of Cegg is used as a basis for discussion, it can be approximated that 50-80% of the HIS+ transformants carry the insert. If using this fact as a basis for discussion, conclutions could be drawn that some of the cultures probably did contain the hemoglobin gene (since they all are HIS+) but this cannot be confirmed without further analysis.

Future considerations

To derive a more clear and decisive result from this project, it could be beneficial to perform a number of further trials. One thing would be to use a more sensitive assay to determine the concentration of produced hemoglobin. SDS-PAGE is a fast and easy way to see the proteins expressed from a cell, but if a protein is expressed at a low level, it could possibly be difficult to see the bands on the gel. An assay, such as an ELISA or a western blot, could be performed to exclude the possibility of the protein being expressed at a lower rate of detection than the SDS-PAGE offers.

Another easy way of optimizing the expression could be to perform the expression trials with the other strain of *Pichia pastoris* included in the *Pichia Expression Kit* (Invitrogen, 2020). In these experiments only the KM71 strain was tested for expression, but in the instructions for the kit it's stated that one strain could potentially express "the protein of choice" better than the other. Therefore, an expression using the GS115 strain could show better results, even when using SDS-PAGE as an evaluation method for expression.

If, potentially, the cells were prevented from growing when inoculating a HIS+ colony embedded in the top agarose to the expression medium, a different method (such as electroporation) could possibly change the result of the expression. This conclusion is based on the result of the colony PCR, and with that result being of rather low value of trust, this step could be redundant. One benefit of using electrophoresis over the spheroplasting method used in this project would, generally, be a higher rate of transformation and spare the analyst the trouble of working with top agarose.

Conclusions

In conclusion, this work has established several insights. It is possible to integrate the gene of sugar beet hemoglobin into the expression vectors from the *Pichia pastoris* ORIGINAL Expression Kit. Furthermore, the expression vector has partially or completely integrated into the Pichia genome. If only partly, the gene for histidinol dehydrogenase was integrated but the transformed cells went through a gene conversion event, thus removing the hemoglobin gene itself. The latter seems highly unlikely according to the statistics from Cereghino and Cegg. Lastly, hemoglobin was not expressed to the extent that it was present as a band on the

SDS-PAGE, further experiments are needed to establish the degree of expression in the KM71 strain of *Pichia pastoris*.

Sources

Axfood. (2022). *Vegobarometern: Klimat och miljö allt viktigare när vegetariskt fortsätter öka.* Avaliable at : https://www.axfood.se/nyhetsrum/pressmeddelanden/2022/01/vegobarometern-klimat-och-miljo-allt-viktigare-nar-vegetariskt-fortsatter-oka/ (Accessed 2022-03-24).

Benchling. (2022) *Cloud-Based Informatics for Life Science R&D. Available at: https://www.benchling.com/* (Accessed 2022-01-18).

Bio-Rad. (2021) Mini-PROTEAN® TGX Precast Gels. Avaliable at: https://www.bio-rad.com/en-se/product/mini-protean-tgx-153-precast-gels?ID=N3GRW04VY (Accessed: 2021-12-22)

Cereghino, JL. Cregg, JM. (2000) 'Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*', *FEMS Microbiology Reviews*, 24(1), pp. 45–66. doi: 10.1111/j.1574-6976.2000.tb00532.x

Charoenrat, T. (2005) 'Oxygen-limited fed-batch process: An alternative control for Pichia pastoris recombinant protein processes´, *Bioprocess and Biosystems Engineering*, 27(6), pp. 399-406. doi: 10.1007/s00449-005-0005-4

Eurofins, (2021) OLIGO ANALYSIS TOOL. Avaliable at https://eurofinsgenomics.eu/en/ecom/tools/oligo-analysis/. (Accessed: 2021-05-24)

Gupta, K. Hebelstrup, KH. Mur, LAJ. Igamberdiev, AU. (2011) 'Plant hemoglobins: Important players at the crossroads between oxygen and nitric oxide', *FEBS Letters*. 585(24), pp. 3843-3849. doi: 10.1016/j.febslet.2011.10.036.

Hill, R. (2012) 'Non-symbiotic hemoglobins - Whats happening beyond nitric oxide scavenging?', *AoB Plants*, published online 2012-03-02. doi: 10.1093/aobpla/pls004

Hoy, J & Hargrove, M. (2008) 'The structure and function of plant hemoglobins'. *Plant Physiology and Biochemistry*. 46(3), pp. 371-379. doi: 10.1016/j.plaphy.2007.12.016.

Imam MU, Zhang S, Ma J, Wang H, Wang F. (2017) ´Antioxidants Mediate Both Iron Homeostasis and Oxidative Stress´, *Nutrients*, Jun 28;9(7), pp. 671. doi: 10.3390/nu9070671

Integrated Technologies. (2021). *Integrated DNA Technologies. Avaliable at: https://eu.idtdna.com/pages (Accessed 2022-04-01)*

Invitrogen. 2020. *Pichia Expression Kit USER GUIDE*. Life Technologies Corporation. Carlsbad, CA 92008.

Kakar, S. Hoffman, F. Storz, J.F, Fabian, M. Hargrove, M.S. (2010) 'Structure and reactivity of hexacoordinate hemoglobins', *Biophysical Chemistry*, 152(1-3), pp 1-14. doi: 10.1016/j.bpc.2010.08.008

Karbalaei, M. Rezaee, S. Farsiani, H. (2020) 'Pichia pastoris: A highly successful expression system for optimal synthesis of heterologous proteins', *Journal of Cellular Physiology*, 235(9), pp. 5867-5881. doi: 10.1002/jcp.29583

Karbalaei, M. Rezaee, S. Farsiani, H. (2020) 'Pichia Pastoris: A highly successful expression system for optimal synthesis of heterologous proteins', *Journal of Cellular Physiology*, 234(9), pp. 5867-5881. doi: 10.1002/jcp.29583.

Kazusa. (2021) *Pichia pastoris codon usage table.* Avaliable at: https://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=4922&aa=15&style=N (Acessed: 2022-02-16).

Leiva-Eriksson, N. Pin, P. Kraft, T. Dohm, J. Minochi, A. Himmelbauer, H. Bülow, L. (2014) 'Differential Expression Patterns of Non-Symbiotic Hemoglobins in Sugar Beet', *Plant and Cell Physiology*, 55(4), pp. 834-844. doi: 10.1093/pcp/pcu027

Macauley Patrick, S. Fazenda, M. McNeil, B. Harvey, L. (2021) 'Heterologous protein production using the Pichia pastoris expression system', *Yeast*, 22(4):249-70. doi: 10.1002/yea.1208.

MACHEREY-NAGEL. 2021. *PCR clean-up and Gel extraction instructions. Rev 0.6.*

Marengo-Rowe, A. (2006) 'Structure-funcion relations of human hemoglobins', *Proc (Bayl Univ Med Cent)*, 19(3), pp. 239-245. doi: 10.1080/08998280.2006.11928171

Moustarah, F. Mohiuddin, S. 2021. 'Dietary Iron', Tresure Island (FL). StatPearls Publishing

Omega BIO-TEK, 2019. *E.Z.N.A® plasmid DNA mini kit instructions*. *Rev 5.2*

Pawlak, R. Berger, J. Hines, I. 2018, 'Iron Status of Vegetarian Adults: A Review of Literature', *American Journal of Lifestyle Medicine*, 12(6), pp 486-498. doi: 10.1177/1559827616682933

PichiaGenome. (2021) *Taxonomy and natural isolates of Pichia Pastoris*. Avaliable at: http://pichiagenome-ext.boku.ac.at:8080/apex/f?p=100:1 (Accessed 2021-03-27).

Pingzuo, L. Anukanth, A. Xiu-Gong, G. Kuppusamy, I. Vincent V, S. Nejat, D. Renugopalakrishnan, V. (2007) 'Expression of recombinant proteins in Pichia pastoris', *Applied Biochemistry and Biotechnology*, 142(2), pp. 105-124. doi:10.1007/s12010-007-0003-x

Saha, D. Patgaonkar M. Shroff, A. Ayyar, K. Bashir, T. Reddy, K.V.R. (2014) 'Hemoglobin Expression in Nonerythoid Cells: Novel or Ubiquitous?',*_International Journal of Inflammation*, published online 2014-11-05. doi: 10.1155/2014/803237

Schechter, A. (2008) 'Hemoglobin research and the origins of molecular medicine', *Blood*, 112(10), pp. 3927-3938. doi: 10.1182/blood-2008-04-078188

Skolmowska, D. Głabska, D. (2019) ‘Analysis of Heme and Non-Heme Iron Intake and Iron Dietary Sources in Adolescent Menstruating Females in a National Polish Sample’, *Nutrients*, doi: 10.3390/nu11051049

Soundaraya, N. Suganthi, P. (2016) ´A Review on anaemia – types, causes, symptoms and their treatments´, *Journal of science and technology investigation*, 1(1), pp. 10-17.

Stryer, L. Berg, JM. Tymoczko, JL. (2012) *Biochemistry*. 7th edition. New York. Freeman and Company.

Thermo Scientific, 2012. *DNA insert ligation (sticky-end and blunt-end) into vector DNA.*

Thermo Scientific, 2018. *FastDigest restriction enzymes protocol*.

Thermo Scientific, 2020. *FastAp™ Thermosensitive Alkaline Phosphatase protocol. Rev C.00*

Thermo Scientific, 2021. *Phusion™ High–Fidelity DNA Polymerase protocol. Rev E.0*

Thermo Scientific. 2017. *Thermo Scientific PageRuler Unstained Protein Ladder. Rev. B.00*

ThermoFisher SCIENTIFIC. (n.d) *Nucleic Acid Electrophoresis Additional Considerations – 7 Aspects.* Avaliable at: https://www.thermofisher.com/se/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/na-electrophoresis-education/na-electrophoresis-considerations.html. (Accessed: 2022-01-14)

Van Buren, J. (2017) *Who discovered hemoglobin*. Avaliable at: https://sciencing.com/organs-make-up-circulatory-system-8566255.html. (Accessed: 2021-03-02).

Walke A *et al*, (2020) ‘Protoporphyrin IX Analysis from Blood and Serum in the Context of Neurosurgery of Glioblastoma’, *Mass Spectrometry in Life Sciences and Clinical Laboratory*, published online 2020-12-12. doi: 10.5772/intechopen.95042

White, C E. Kempi, NM. Komives, EA. (1994) ‘Expression of highly disulfide-bonded proteins in Pichia pastoris’, *Ways and Means*, 2, pp. 1003-1005. doi: 10.1016/s0969-2126(94)00103-0

WHO. (2022) *Anaemia. Avaliable at:* https://www.who.int/health-topics/anaemia#tab=tab_1. (Accessed: 2022-09-09).

Appendix

A. Full insert sequences and primers

Below are the full insert sequences for the hemoglobin genes, also included below the name of the insert, the component parts for the sequence are listed. They all contain a primer sequence that allows for amplification and restriction sites. The intracellular insert also contains a Kozak sequence per the instructions of Thermo Fisher.

Insert for pPIC3.5 (Intracellular)

Primer sequence - *Eco*RI – Kozak sequence – Gene (BvHb.1.2) – Stop - *Not*I – Reverse complement of primer sequence

BvHb.ISK *(Beta vulgaris, Intracellular, Standard Kozak)*

GAACTAGAACTTGGTCTACCTCGAATTCAAAAATGTCTTTTACTAACGTTAACTACCCAGCT
TCTGATGGTACTGTTATTTTTACTGAAGAACAAGAAGCTTTGGTTGTTCAATCTTGGAACGT
TATGAAGAAGAACTCTGCTGAATTGGGTTTGAAGTTGTTTTTGAAGATTTTTGAAATTGCTC
CAACTGCTAAGAAGATGTTTTCTTTTGTTAGAGATTCTGATGTTCCATTGGAACAAAACCAA
AAGTTGAAGGGTCATGCTATGTCTGTTTTTGTTATGACTTGTAAATCTGCTGCTCAATTGAG
AAAGGCTGGTAAGGTTACTTTTGGTGAATCTTCTTTGAAGCATATGGGTTCTGTTCATTTGA
AGTACGGTGTTGTTGATGAACATTTTGAAGTTACTAGATTTGCTTTGTTGGAAACTATTAAA
GAAGCTGTTCCAGAAATGTGGTCTCCAGAAATGAAAAACGCTTGGGCTGAAGCTTTTAACCA
TTTGGTTGCTGCTATTAAGGCTGAAATGCAAAGATTGTCTACTCAACCATAAGCGGCCGCGA
GTCAGCATATAGTGATTCCG

Insert for pPIC9 (Extracellular, α-factor)

Primer sequence - *Eco*RI – Gene (BvHb.1.2) – Stop - *Not*I – Reverse compliment of primer sequence

BvHb.Exα *(Beta vulgaris, Extracellular, α-factor)*

GAACTAGAACTTGGTCTACCTCGAATTCATGTCTTTTACTAACGTTAACTACCCAGCTTCTG
ATGGTACTGTTATTTTTACTGAAGAACAAGAAGCTTTGGTTGTTCAATCTTGGAACGTTATG
AAGAAGAACTCTGCTGAATTGGGTTTGAAGTTGTTTTTGAAGATTTTTGAAATTGCTCCAAC
TGCTAAGAAGATGTTTTCTTTTGTTAGAGATTCTGATGTTCCATTGGAACAAAACCAAAAGT
TGAAGGGTCATGCTATGTCTGTTTTTGTTATGACTTGTAAATCTGCTGCTCAATTGAGAAAG
GCTGGTAAGGTTACTTTTGGTGAATCTTCTTTGAAGCATATGGGTTCTGTTCATTTGAAGTA
CGGTGTTGTTGATGAACATTTTGAAGTTACTAGATTTGCTTTGTTGGAAACTATTAAAGAAG
CTGTTCCAGAAATGTGGTCTCCAGAAATGAAAAACGCTTGGGCTGAAGCTTTTAACCATTTG
GTTGCTGCTATTAAGGCTGAAATGCAAAGATTGTCTACTCAACCATAAGCGGCCGCGAGTCA
GCATATAGTGATTCCG

Insert for pHIL-S1 (Extracellular, PHO1)

Primer sequence - *Eco*RI – Gene (BvHb.1.2) – Stop – *Bam*HI – reverse compliment to primer sequence

BvHb.ExPHO1 *(Beta vulgaris, Extracellular, PHO1 signal sequence)*

```
GAACTAGAACTTGGTCTACCTCGAATTCATGTCTTTTACTAACGTTAACTACCCAGCTTCTG
ATGGTACTGTTATTTTTACTGAAGAACAAGAAGCTTTGGTTGTTCAATCTTGGAACGTTATG
AAGAAGAACTCTGCTGAATTGGGTTTGAAGTTGTTTTTGAAGATTTTTGAAATTGCTCCAAC
TGCTAAGAAGATGTTTTCTTTTGTTAGAGATTCTGATGTTCCATTGGAACAAAACCAAAAGT
TGAAGGGTCATGCTATGTCTGTTTTTGTTATGACTTGTAAATCTGCTGCTCAATTGAGAAAG
GCTGGTAAGGTTACTTTTGGTGAATCTTCTTTGAAGCATATGGGTTCTGTTCATTTGAAGTA
CGGTGTTGTTGATGAACATTTTGAAGTTACTAGATTTGCTTTGTTGGAAACTATTAAAGAAG
CTGTTCCAGAAATGTGGTCTCCAGAAATGAAAAACGCTTGGGCTGAAGCTTTTAACCATTTG
GTTGCTGCTATTAAGGCTGAAATGCAAAGATTGTCTACTCAACCATAAGGATCCGAGTCAGC
ATATAGTGATTCCG
```

Below are the primers used to amplify the insert sequences. These were attached to the the insert sequences along with the sequences the restriction sites, as shown above.

Forward primer

```
GAACTAGAACTTGGTCTACCTC
```

Reverse primer

```
CGGAATCACTATATGCTGACTC
```

B. Maps of expression vectors

Below, in *Figure B1, B2 and B3*, genetic maps of all three expression vectors are presented.

Comments for pPIC3.5:
7751 nucleotides

5′ *AOX1* promoter fragment: bases 1-937
5′ *AOX1* primer site: bases 855-875
Multiple Cloning Site: bases 938-968
3′ *AOX1* primer site: bases 1055-1075
3′ *AOX1* transcription termination
(TT) fragment: bases 981-1314
HIS4 ORF: bases 4242-1708
3′ *AOX1* fragment: bases 4598-5354
pBR322 origin: bases 6436-5764
Ampicillin resistance gene: bases 7442-6582

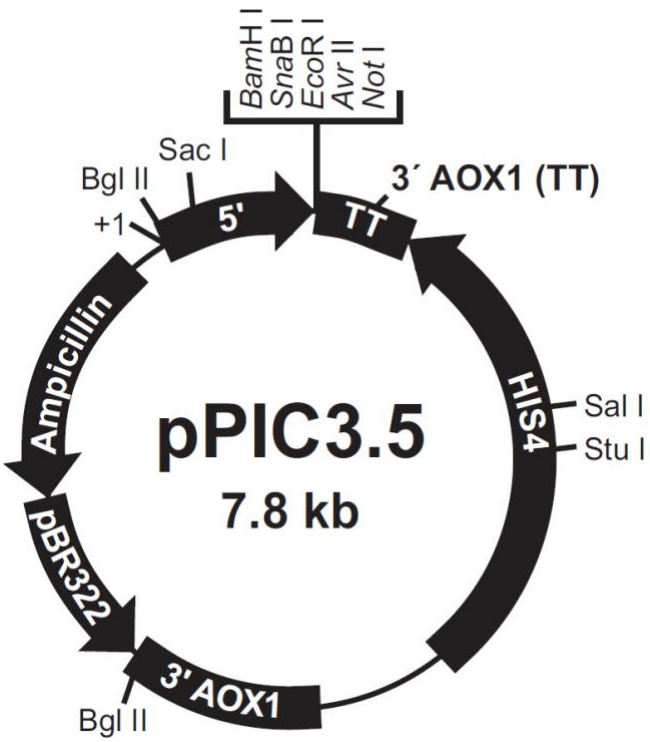


Figure B1: Map of pPIC3.5. The vector contains a HIS4 gene, the AOX1 promotor, ampicillin resistance and various restriction sites. (Invitrogen, 2020)

Comments for pPIC9:
8023 nucleotides

5′ *AOX1* promoter fragment: bases 1-948
5′ *AOX1* primer site: bases 855-875
α-Factor secretion signal(s): bases 949-1215
α-Factor primer site: bases 1152-1172
Multiple Cloning Site: bases 1192-1241
3′ *AOX1* primer site: bases 1327-1347
3′ *AOX1* transcription
termination (TT): bases 1253-1586
HIS4 ORF: bases 4514-1980
3′ *AOX1* fragment: bases 4870-5626
pBR322 origin: bases 6708-6034
Ampicillin resistance gene: bases 7713-6853

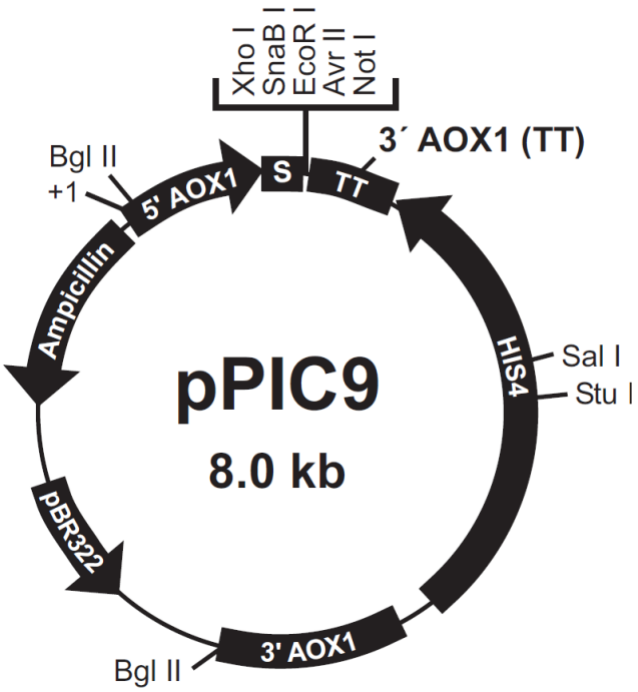


Figure B2: Map of pPIC9. *The vector contains a HIS4 gene, the AOX1 promotor, ampicillin resistance and various restriction sites* (Invitrogen, 2020).

Comments for pHIL-S1:
8212 nucleotides

5' *AOX1* promoter: bases 1-940
5' *AOX1* primer site: bases 855-875
PHO1 secretion signal (S): bases 941-1006
Multiple Cloning Site: bases 1005-1025
3' *AOX1* primer site: bases 1098-1118
3' *AOX1* transcription termination (TT): bases 1024-1189
HIS4 ORF: bases 4273-1740
3' *AOX1* fragment: bases 4628-5376
pBR322 origin: bases 6536-5917
f1 origin of replication: bases 6545-7000
Ampicillin resistance gene: bases 7902-7042

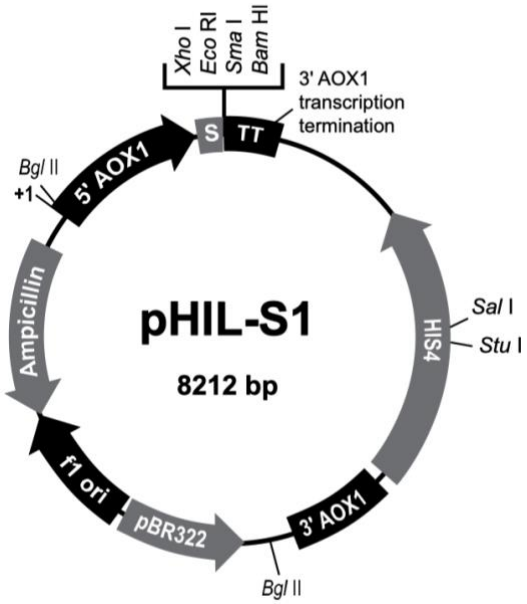
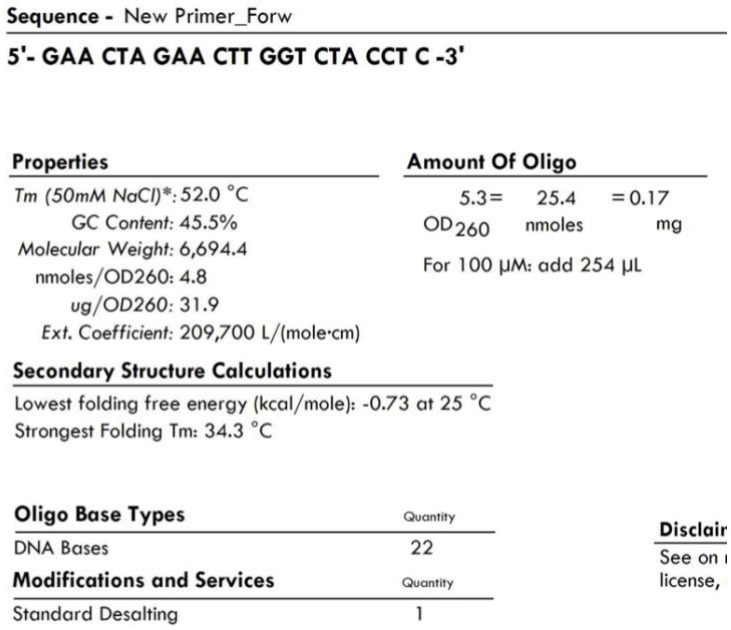


Figure B3: Map of pHIL-S1. The vector contains a HIS4 gene, the AOX1 promotor, ampicillin resistance and various restriction sites (*Invitrogen, 2020).*

C. Spec. sheets of primers for amplification of insert sequences with Phusion PCR.

Below, in *Figure C1* and *C2*, are the spec. sheets for the primers used for the amplification of the insert sequences containing the hemoglobin genes synthesized by Integrated Technologies. (2021).



Sequence - New Primer_Forw

5'- GAA CTA GAA CTT GGT CTA CCT C -3'

Properties	**Amount Of Oligo**
Tm (50mM NaCl):* 52.0 °C	5.3= OD_{260} 25.4 nmoles = 0.17 mg
GC Content: 45.5%	For 100 µM: add 254 µL
Molecular Weight: 6,694.4	
nmoles/OD260: 4.8	
ug/OD260: 31.9	
Ext. Coefficient: 209,700 L/(mole·cm)	

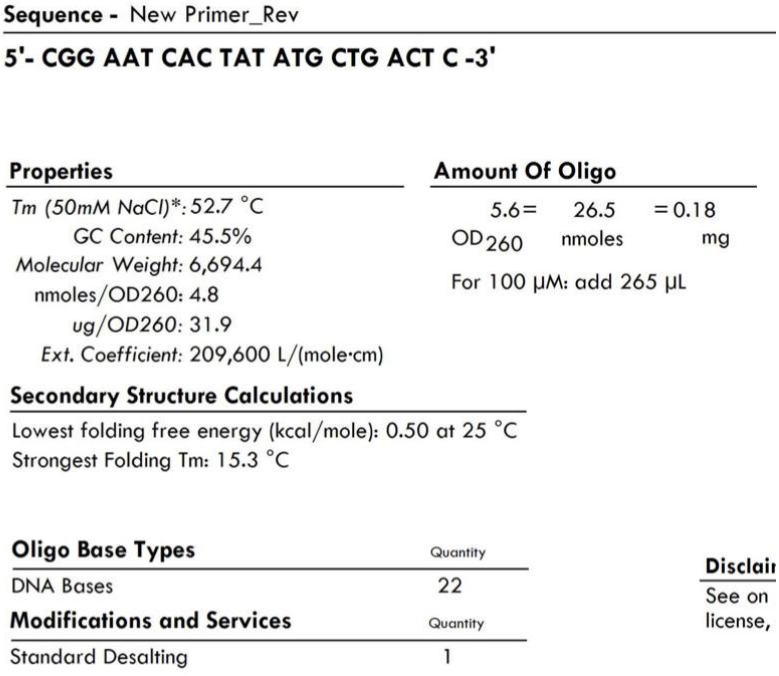
Secondary Structure Calculations

Lowest folding free energy (kcal/mole): -0.73 at 25 °C
Strongest Folding Tm: 34.3 °C

Oligo Base Types	Quantity
DNA Bases	22
Modifications and Services	Quantity
Standard Desalting	1

Disclair
See on
license,

Figure C1: Specs for forward primer. These spec. sheets contain the sequence along with some properties of the primer. (Integrated Technologies, 2021)



Sequence - New Primer_Rev

5'- CGG AAT CAC TAT ATG CTG ACT C -3'

Properties	**Amount Of Oligo**
Tm (50mM NaCl):* 52.7 °C	5.6= OD_{260} 26.5 nmoles = 0.18 mg
GC Content: 45.5%	For 100 µM: add 265 µL
Molecular Weight: 6,694.4	
nmoles/OD260: 4.8	
ug/OD260: 31.9	
Ext. Coefficient: 209,600 L/(mole·cm)	

Secondary Structure Calculations

Lowest folding free energy (kcal/mole): 0.50 at 25 °C
Strongest Folding Tm: 15.3 °C

Oligo Base Types	Quantity
DNA Bases	22
Modifications and Services	Quantity
Standard Desalting	1

Disclair
See on
license,

Figure C2: Specs for reverse primer. These spec. sheets contain the sequence along with some properties of the primer. (Integrated Technologies, 2021).

D. Heat shock transformation in *E.coli*

Below is the protocol used when transforming *E. coli.*

T**ransformation of E. coli**

1. Thaw the competent cells gently on ice (~10 min).Work in sterile environment!

2. Gently pipette 15 µl PCR (0.5 µl of pure plasmid) reaction into the competent cells.

3. Incubate 25 minutes on ice.

4. Transfer the tube to a 42°C heat block and incubate for exactly 45 s. Put the tube back on ice gently.

5. Add 900 microliter sterile LB (no antibiotic) and incubate while shaking for 60 minutes at 37°C.

6. Pellet the cells by centrifugation and remove all but ~50 ml LB. Resuspend the cells and spread the cell suspension onto agar plates containing appropriate antibiotic.

7. Let the plates dry before turning them upside down (~10-20 min).

8. Incubate upside down overnight at 37°C.

E. YPD medium

Presented in *Figure E1* are the instructions for making the YPD medium used for growing *Pichia pastoris.*

YPD or YEPD

Yeast Extract Peptone Dextrose Medium (1 liter)

1% yeast extract

2% peptone

2% dextrose (glucose)

Note: If you are using the YP Base Medium or the YP Base Agar medium pouches included with the Multi-Copy *Pichia* Expression Kit, follow the directions on the pouch.

1. Dissolve 10 g yeast extract and 20 g of peptone in 900 mL of water.

   Note: If making YPD slants or plates, add 20 g of agar.

2. Autoclave for 20 minutes on liquid cycle.
3. Add 100 mL of 10X D.
4. Store the liquid medium at room temperature. Store the YPD slants or plates at 4°C. The shelf life is several months.

Figure E1: Components of the YPD medium. (Invitrogen, 2020)

F. Contents of reagents in the *Pichia ORIGINAL Expression Kit*

In *Figure F1* is the content of reagents of the *Pichia ORIGINAL Expression Kit* presented. (Invitrogen, 2020)

Reagent	Amount	Components
SOS medium	20 mL	1 M Sorbitol 0.3X YPD 10 mM $CaCl_2$
Sterile Water	2 × 125 mL	Autoclaved, deionized water
SE	2 × 125 mL	1 M Sorbitol 25 mM EDTA, pH 8.0
SCE	2 × 125 mL	1 M Sorbitol 10 mM Sodium citrate buffer, pH 5.8 1 mM EDTA
1 M Sorbitol	2 × 125 mL	–
CaS	2 × 60 mL	1 M Sorbitol 10 mM Tris-HCl, pH 7.5 10 mM $CaCl_2$
40% PEG	25 mL	40% (w/v) PEG 3350 (Reagent grade) in water
CaT	25 mL	20 mM Tris-HCl, pH 7.5 20 mM $CaCl_2$

Figure F1. Reagents included in the Pichia ORIGINAL Expression kit. (Invitrogen, 2020).

G. Top agarose and agar plates for *Pichia* transformants

In *Figure G1* and *G2*, the instructions for making RDB, RD, RDHB and RDH plates are presented.

RDB and RDHB agar plates

1. Dissolve 186 g of sorbitol in 700 mL of water and add 20 g of agar.
2. Autoclave for 20 minutes on liquid cycle.
3. Place the autoclaved solution in a 60°C water bath before the addition of pre-warmed mixture of stock solutions. This will keep the medium from becoming too thick to mix reagents.
4. Prepare the pre-warmed (45°C) mixture from "RD and RDH liquid media" on page 75, Step 3, then add to the sorbitol/agar solution. **If you are selecting for His^+ transformants, do not add histidine.**
5. Pour the plates immediately after mixing the solutions in Step 4. Store the plates at 4°C. Shelf life is several months.

Figure G1: Instructions for RDB and RDHB plates (Invitrogen, 2020).

RD and RDH top agar

1. Dissolve 186 g of sorbitol in 700 mL of water and add 10 g of agar or agarose.
2. Autoclave 20 minutes on liquid cycle.
3. Place the autoclaved solution in a 60°C water bath before the addition of pre-warmed mixture of stock solutions. This will keep the medium from becoming too thick to mix reagents.
4. Prepare the pre-warmed (45°C) mixture from "RD and RDH liquid media" on page 75, Step 3, then add to sorbitol/agar solution. **If you are selecting for His^+ transformants, do not add histidine.**
5. Place the solution to 45°C after adding the solutions in Step 4. During transformation, use as a molten solution at 45°C.
6. Store the top agar at 4°C. Shelf life is several months.

Figure G2: Instructions for RD and RDH plates (Invitrogen, 2020).

H. BMGY and BMMY medium

In *Figure H1* below, the instructions for preparing BMGY and BMMY medium are presented.

BMGY and BMMY

Buffered Glycerol-complex Medium

Buffered Methanol-complex Medium (1 liter)

1% yeast extract

2% peptone

100 mM potassium phosphate, pH 6.0

1.34% YNB

4×10^{-5} % biotin

1% glycerol or 0.5% methanol

1. Dissolve 10 g of yeast extract, 20 g peptone in 700 mL of water.
2. Autoclave the solution for 20 minutes on liquid cycle.
3. Cool to room temperature, then add the following and mix well:
   100 mL 1 M potassium phosphate buffer, pH 6.0
   100 mL 10X YNB
   2 mL 500X B
   100 mL 10X GY
4. For BMMY, add 100 mL 10X M (methanol) instead of GY (glycerol).
5. Store the media at 4°C. The shelf life of this solution is approximately two months.

Figure H1: Instructions for BMGY and BMMY expression medium (Invitrogen, 2020).

I. Lysis of cell pellets using acid-washed glass beads

Below, *in Figure I1*, is the protocol for lysis using acid-washed glass beads is presented.

Prepare cell pellets (intracellular and secreted expression)

1. Thaw cell pellets quickly and place on ice.
2. For each 1 mL sample, add 100 µL Breaking Buffer to the cell pellet and resuspend.
3. Add an equal volume of acid-washed glass beads (size 0.5-mm). Estimate equal volume by displacement.
4. Vortex 30 seconds, then incubate on ice for 30 seconds. Repeat for a total of 8 cycles.
5. Centrifuge at maximum speed for 10 minutes at 4°C. Transfer the clear supernatant to a fresh microcentrifuge tube.
6. Take 50 µL of supernatant (cell lysate) and mix with an appropriate volume of denaturing PAGE Gel Loading buffer (Sample Buffer).
7. Heat the sample as recommended and load 10–20 µL per well. Thickness of the gel and number of wells determines the volume loaded. You can store the remaining sample at –20°C for western blots, if necessary. You can store the cell lysates at –80°C for further analysis.

Figure I1. Protocol for lysis of yeast cells using glass beads (Invitrogen, 2020)

J. Direct PCR screening of *Pichia* colonies.

Below, in *Figure J1*, the protocol for the direct PCR screening of the *Pichia pastoris* expression cultures is presented.

Direct PCR screening of *Pichia* clones

Introduction

The following protocol has been reported in the literature to directly test *Pichia* clones for insertion of your gene by PCR (Linder *et al.*, 1996). Briefly, the cells are lysed by a combined enzyme, freezing, and heating treatment. You can use the genomic DNA directly as a PCR template.

Materials needed

- A culture or single colony of a *Pichia* transformant
- 1.5-mL microcentrifuge tube
- 5 U/µL solution of Lyticase (Sigma)
- 30°C water bath or heat block
- Liquid nitrogen
- Reagents for PCR

Perform PCR

1. Place 10 µL of a *Pichia pastoris* culture into a 1.5-mL microcentrifuge tube. For relatively dense cultures, dilute 1 µL of the culture into 9 µL water. Alternatively, pick a single colony and resuspend it in 10 µL of water.
2. Add 5 µL of a 5 U/µL solution of lyticase to the cells, then incubate at 30°C for 10 minutes.
3. Freeze the sample at –80°C for 10 minutes or immerse in liquid nitrogen for 1 minute.
4. Set up a 50-µL PCR for a hot start:

10X Reaction Buffer	5 µL
25 mM $MgCl_2$	5 µL
25 mM dNTPs	1 µL
5′ *AOX1* primer (10 pmol/µL)	1 µL
3′ *AOX1* primer (10 pmol/µL)	1 µL
Sterile water	27 µL
Cell lysate	5 µL
Total volume	45 µL

5. Place the reaction mixture in the thermocycler and incubate at 95°C for 5 minutes.
6. Add 5 µL of a 0.16 U/µL solution of *Taq* polymerase (0.8 units).
7. Cycle 30 times using the following parameters:

Step	Temperature	Time
Denaturation	95°C	1 minute
Annealing	54°C	1 minute
Extension	72°C	1 minute

Include a final extension of 7 minutes at 72°C.

8. Analyze a 10-µL aliquot by agarose gel electrophoresis.

Figure J1: Protocol for direct PCR of Pichia pastoris expression cultures (Invitrogen, 2020).

K. Sequencing of expression vectors

Alignment of the ISK sequence with the sequenced pPIC3.5 + ISK ligation plasmid. The alignment was made using Benchling.com and is shown in *Figure K1, K2 and K3* (2021).

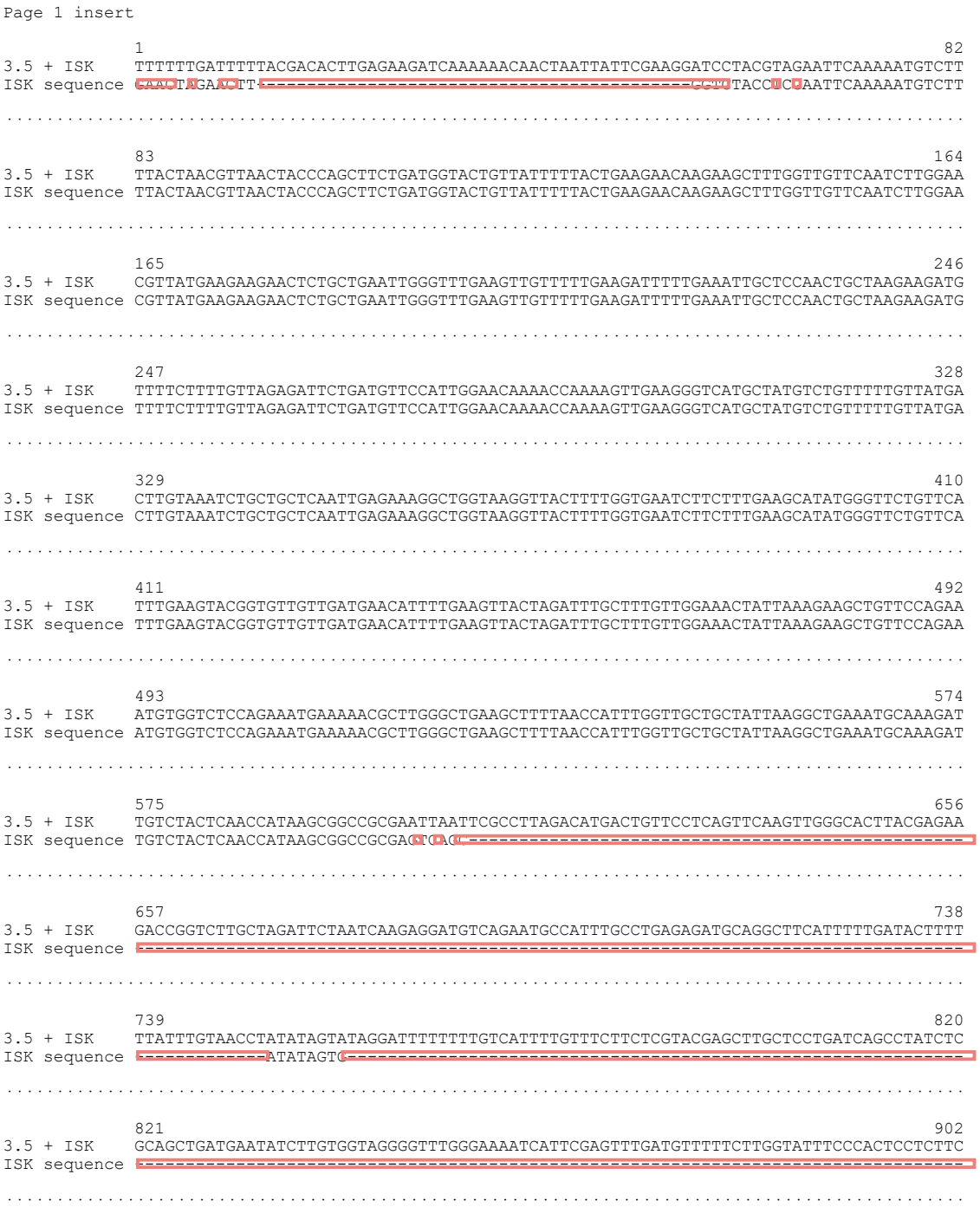


Figure K1: Alignment of ISK sequence and expression vector pPIC3.5 + ISK. A 100% correct alignment starting with restriction site EcoRI and ending with restriction site NotI. The mismatches before and after the restriction site is the primer sequence included in the in silico sequence, which is cleaved off during restriction digestion.

Alignment of the Exα sequence with the sequenced pPIC9 + Exα ligation plasmid. The alignment was made using Benchling.com (2021).

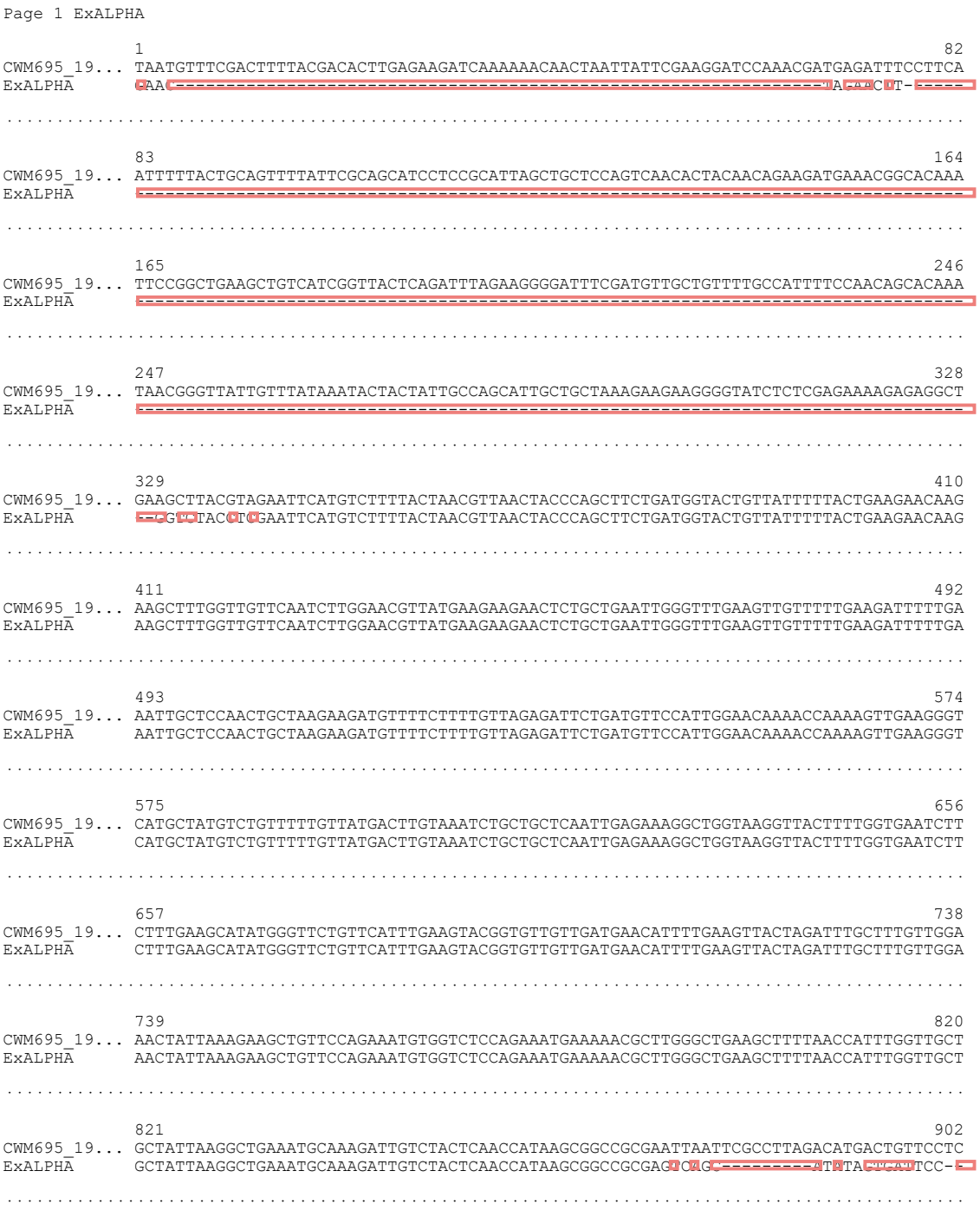


Figure K2: Alignment of Exα sequence and expression vector pPIC9 + Exα. A 100% correct alignment starting with restriction site EcoRI and ending with restriction site NotI. The mismatches before and after the restriction site is the primer sequence included in the in-silico sequence, which is cleaved off during restriction digestion.

Alignment of the PHO1 sequence with the sequenced pHIL-S1 + PHO1 ligation plasmid. The alignment was made using Benchling.com (2021).

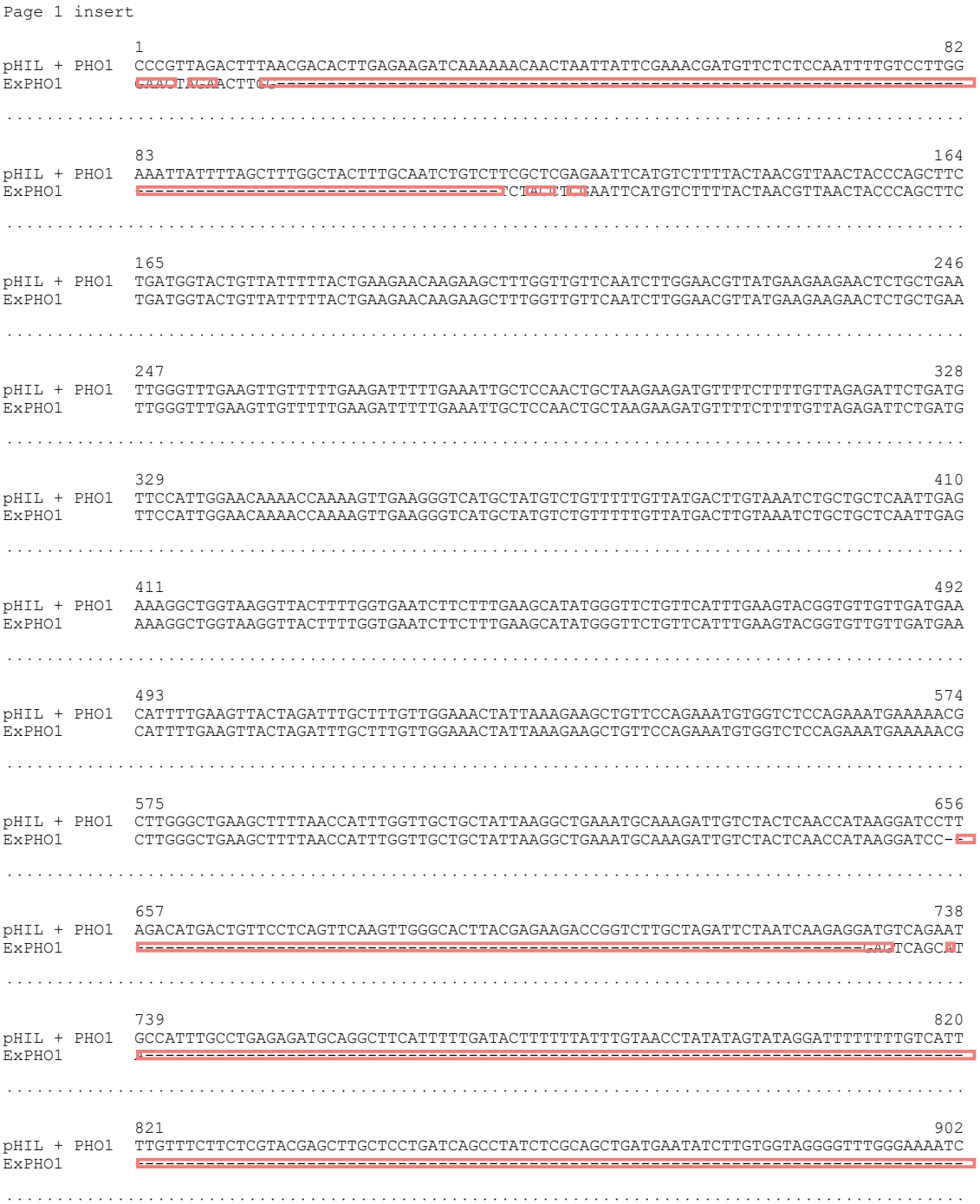


Figure K3: Alignment of ExPHO1 sequence and expression vector pPHIL-S1 + ExPHO1. A 100% correct alignment starting with restriction site EcoRI and ending with restriction site BamHI. The mismatches before and after the restriction site is the primer sequence included in the in silico sequence, which is cleaved off during restriction digestion.

L. Graphs from spheroplasting trials

In the four graphs presented below the results of four out of the five spheroplasting trials are presented. The graphs in *Figure L3* and *Figure L4* was used to determine the incubation duration, while the graphs presented in *Figure L1 and L2* was not used due to their appearance which suggest that they are defect. A correct graph should start att the highest value and then decrease rather linearly with time.

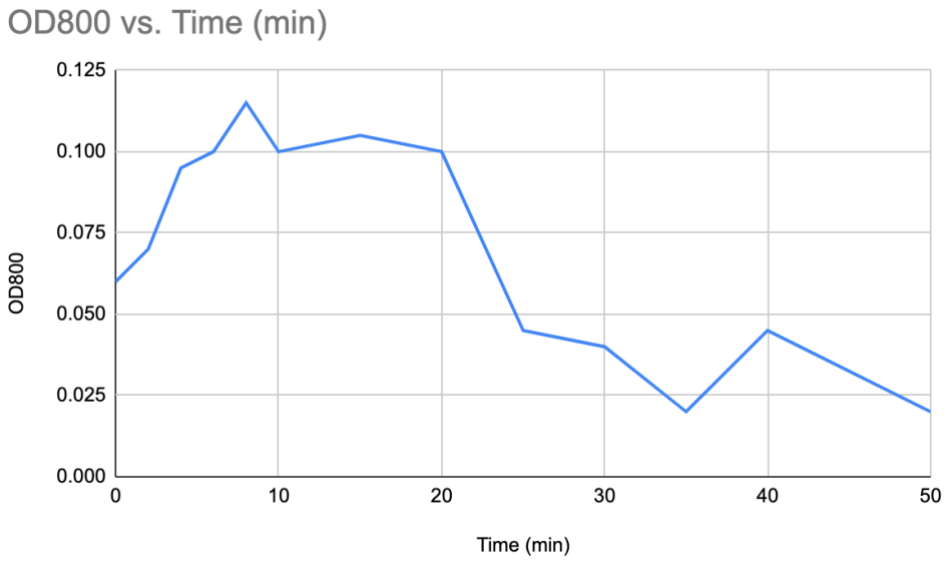


Figure L1: Second spheroplasting trial. This graph has a very low starting value but continues, after about 10 minutes, to decrease in an almost linear fashion.

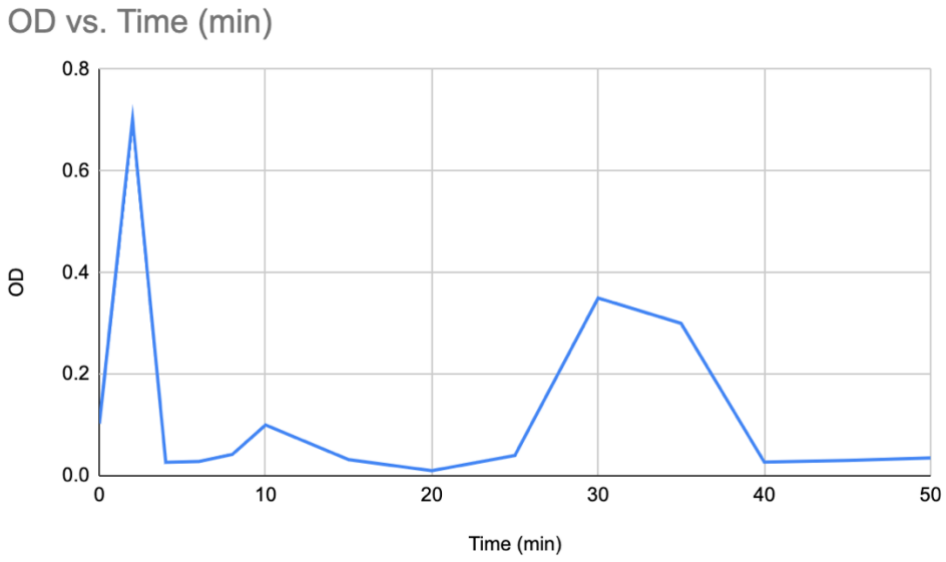


Figure L2: Third spheroplasting trial. This graph shows irregular values with a low starting value and seemingly random OD800 measurements. This trial was determined to be defective and did not influence the incubation duration calculations.

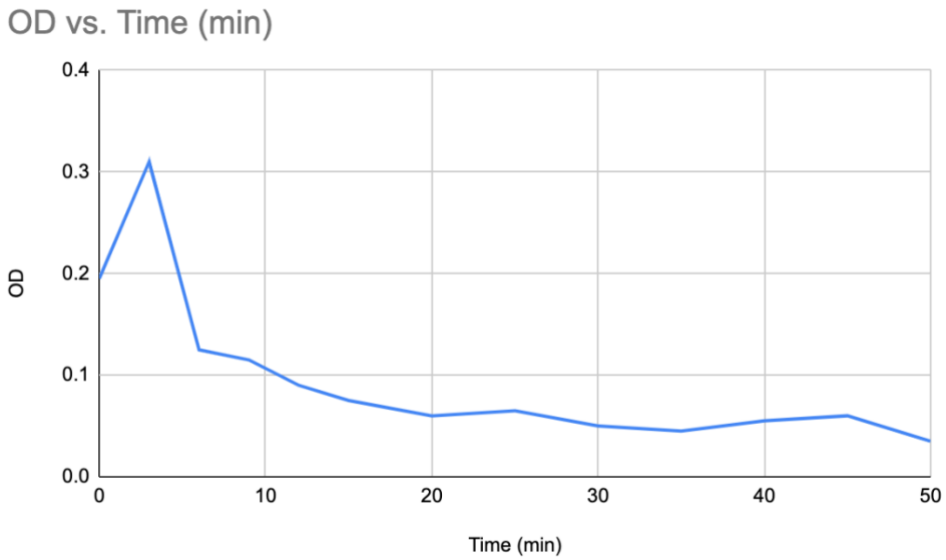


Figure L3: Fourth spheroplasting trial. A quite linear decrease in OD800 except for the second measurement, which peaks at an OD800 of 0.31. If that value would be disregarded as an anomaly, the fourth spheroplasting trial would look like a quite stable curve with a downwards trajectory.

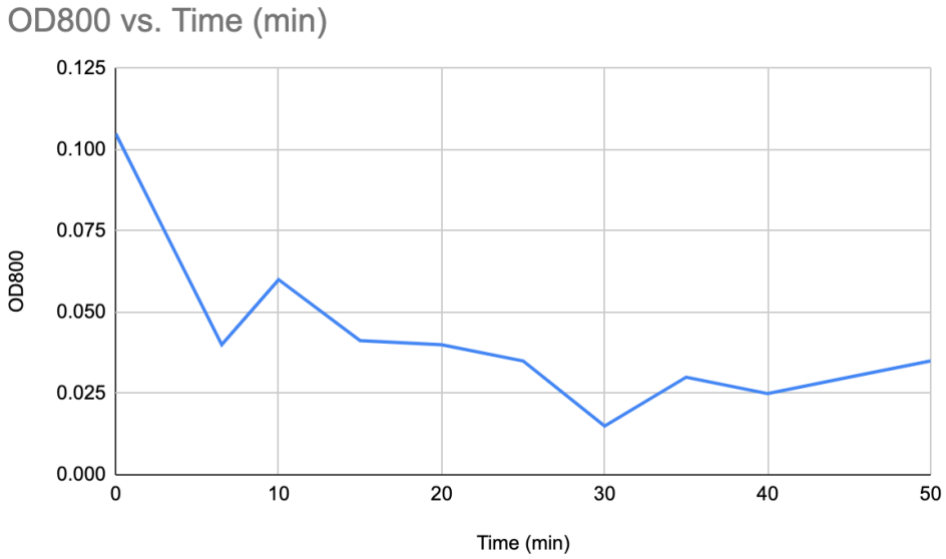


Figure L4. Fifth spheroplasting trial. Starting with a reasonable value, continuing down to a lowered OD800. The curve is increasing in the end, which, regarding the degrading nature of the Zymolyase, should not happen, and should thus be interpreted accordingly.

M. Pichia pastoris transformant plates

In *Figure M1-M3*, shown below, pictures of KM71 transformed with the expression vectors are presented.

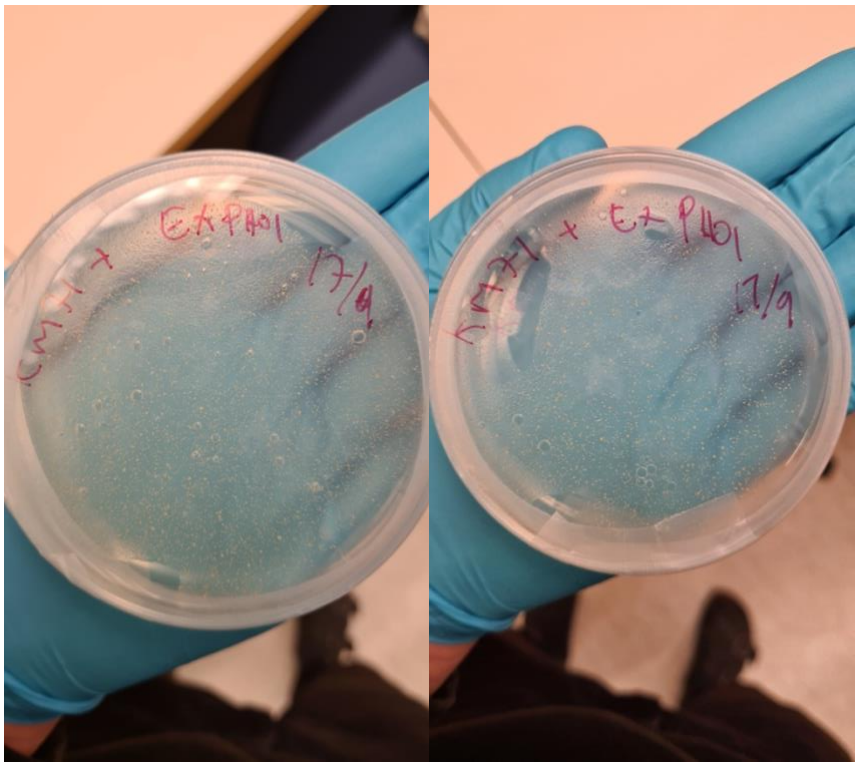


Figure M1: Both plates with pHIL-S1 + ExPHO1 shows HIS+ colonies.

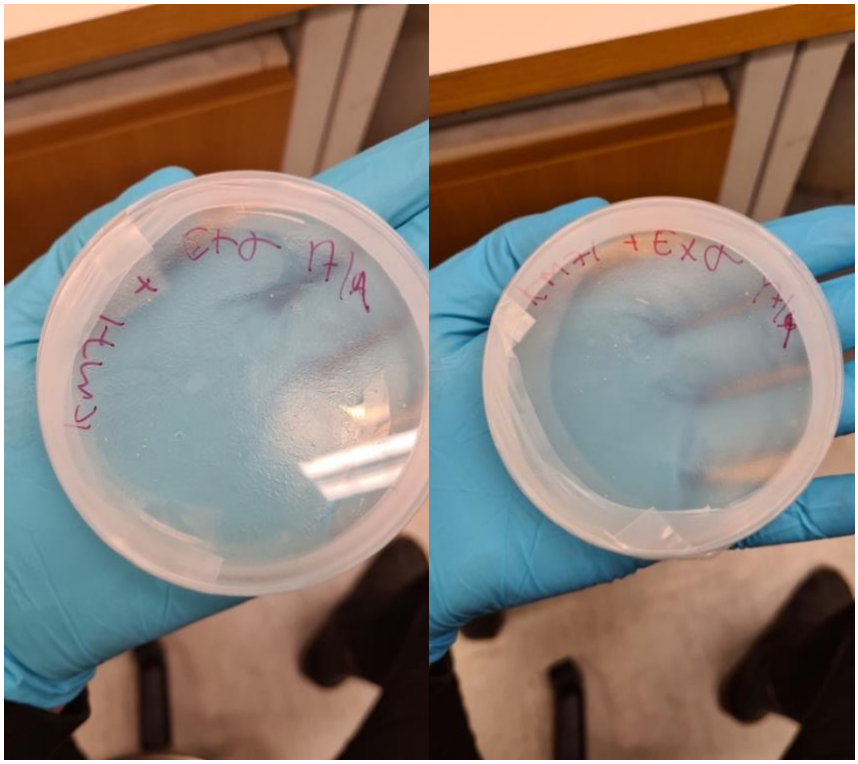


Figure M2: Both plates show a small amount of HIS+ colonies.

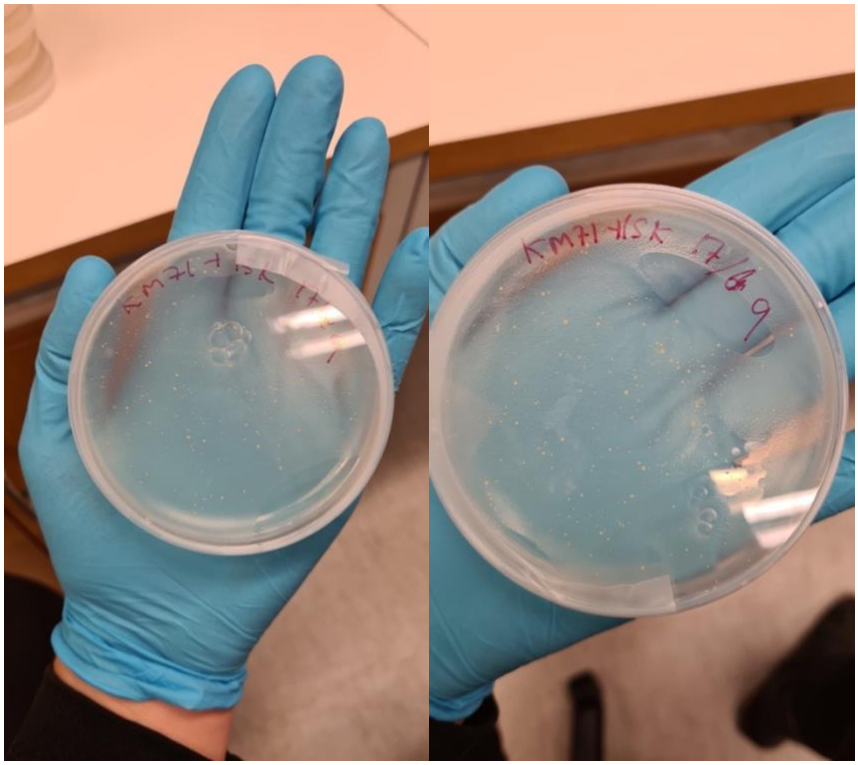


Figure M3: Both plates show several HIS+ transformants.

Presented below, in *Figure M4-M6*, are the transformants of KM71 with the parent vectors containing no insert. These were used as a control for no expression.

pPIC3.5

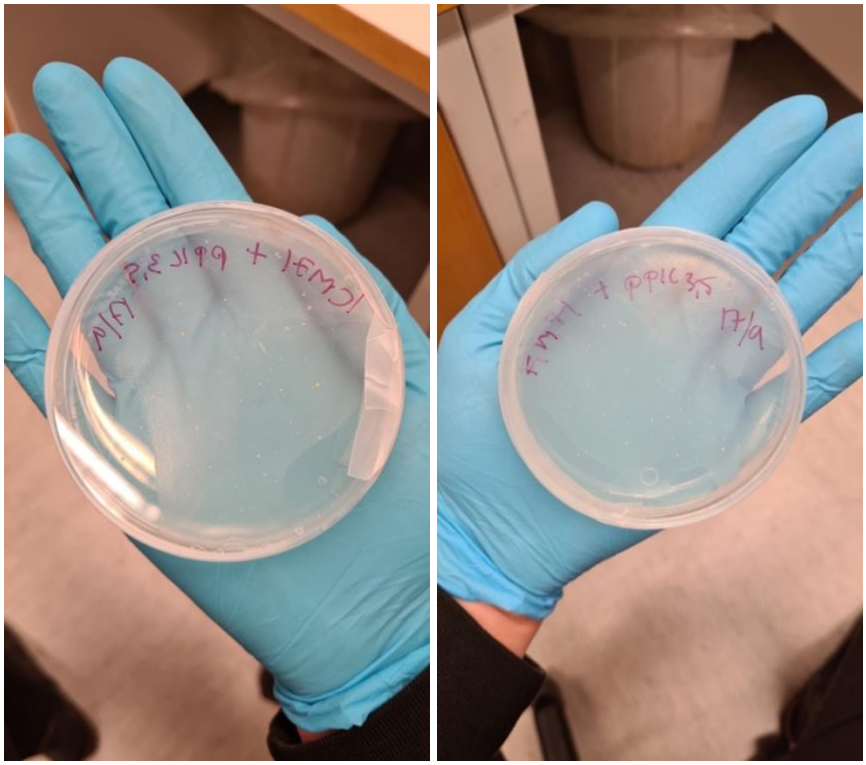


Figure M4: A low number of HIS+ transformants can be seen on both plates.

pPIC9

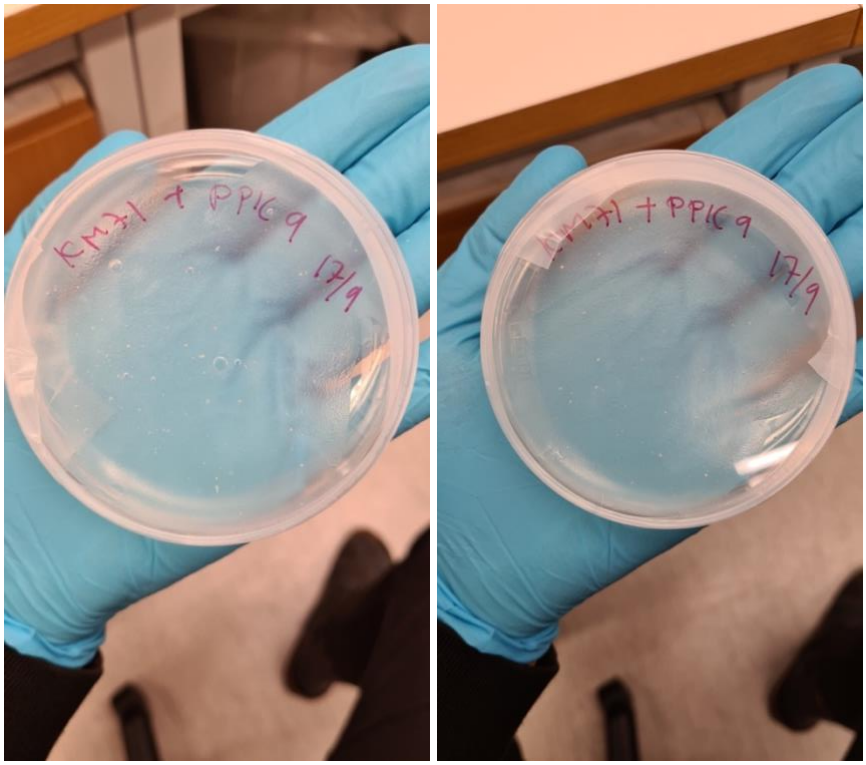


Figure M5: A low number of HIS+ transformants can be seen on both plates.

pHIL-S1

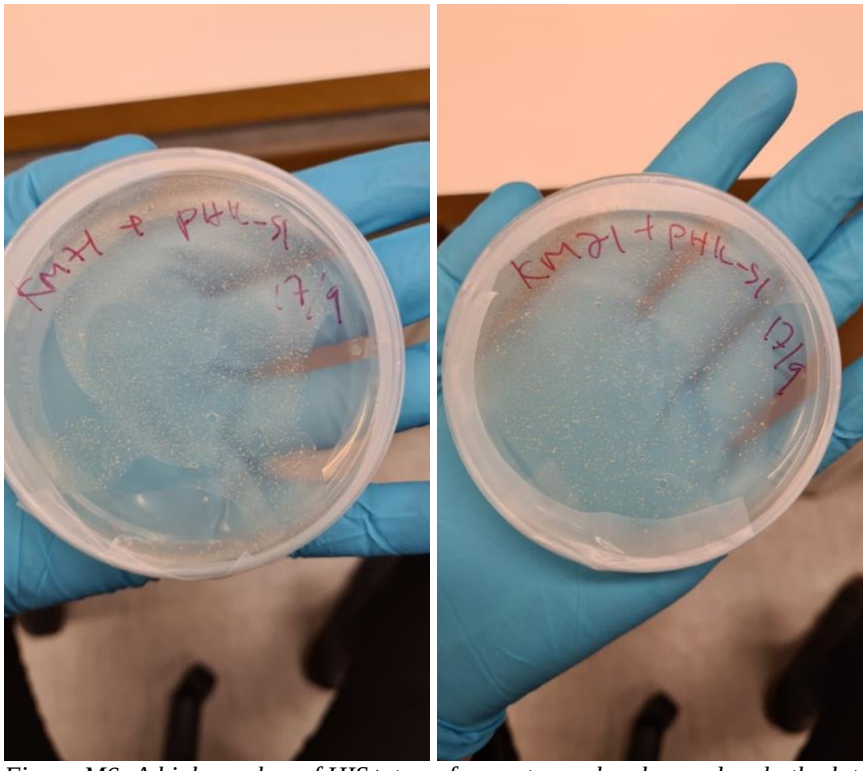


Figure M6: A high number of HIS+ transformants can be observed on both plates.

In *Figure M7* the viability control and the contaminant control plates are presented. The high amount of colonies on the viability plate shows that the cells are able to proliferate after spheroplasting (these cells are *his*- on a medium containing histidine). The second plate confirms that no contaminants were able to grow on the medium without histidine.

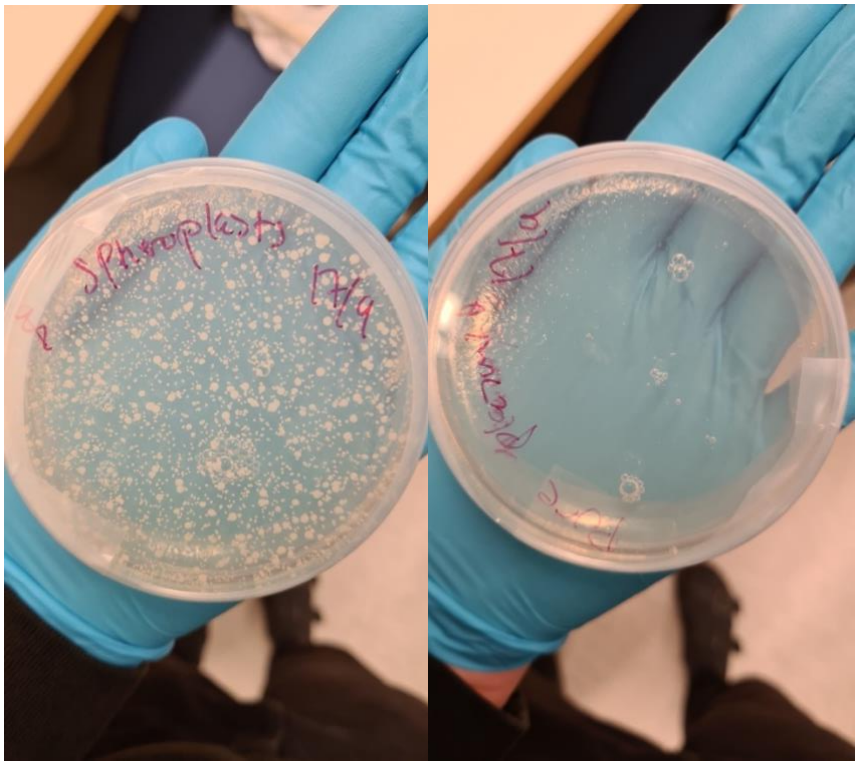


Figure M7: Control samples. The left plate is a viability check of the untransformed Pichia pastoris spheroplasts, and to the right is a negative control with only one of the purified expression vectors to check for contamination.

N. Gel images from the creation of the expression vectors

In *Figure N1-N4* below, the gel images from the various steps in creating the expression vectors are presented. These results are presented and discussed in the main text.

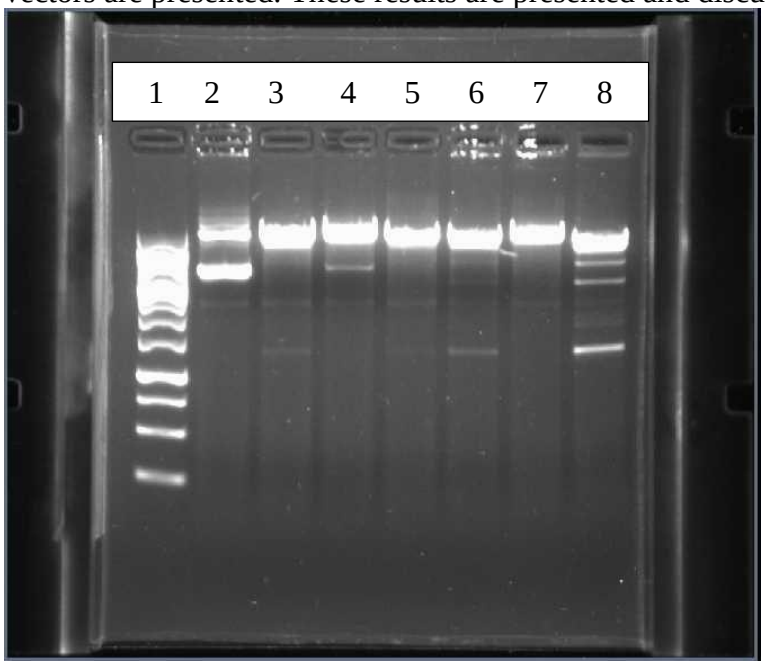


Figure N1: The lanes on the gel represent, starting at the left-hand side: lane 1: GenRuler™, lane 2: uncleaved pPIC9, lane 3: pPIC9 + EcoRI (10 minutes), lane 4: pPIC9 + NotI (10 minutes), lane 5: pPIC9 + EcoRI + NotI (10 minutes), lane 6: pPIC9 + EcoRI (40 minutes), lane 7: pPIC9 + NotI (40 minutes), and lane 8: pPIC9 + EcoRI + NotI (40 minutes)

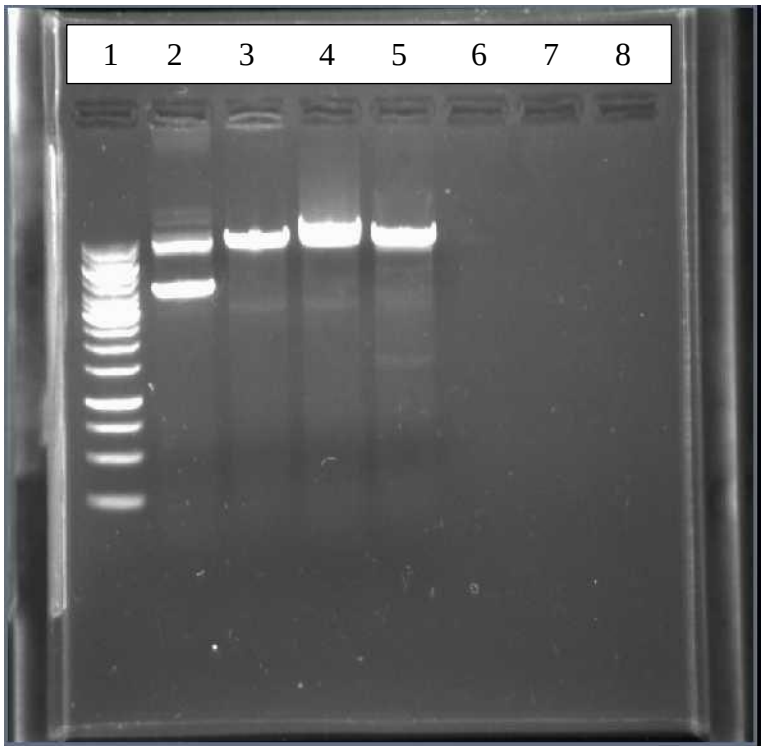


Figure N2: Lanes from left to right are as follows: Lane 1: GenRuler™, lane 2: undigested pPIC9, lane 3: plasmid + EcoRI, lane 4: pPIC9 + NotI, and lane 5: pPIC9 + EcoRI + NotI.

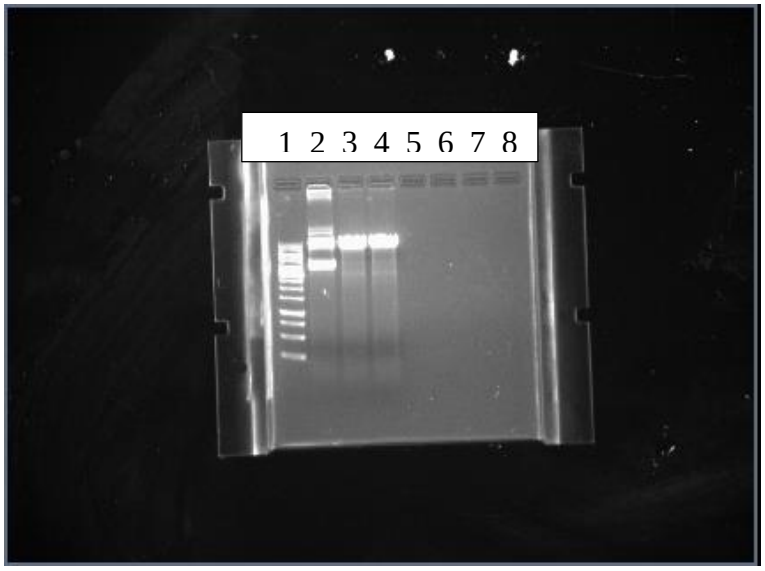


Figure N3: pHIL-S1 digested with BamHI. The lanes from left to right are: lane 1: GenRuler ™, lane 2: undigested pHIL-S1, lane 3: pHIL-S1 + BamHI (5 minutes), and lane 4: pHIL-S1 + BamHI (15 minutes).

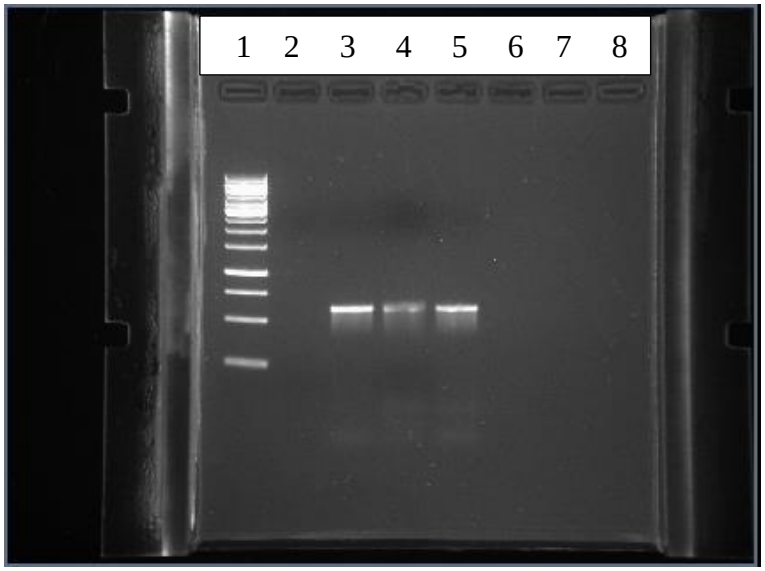


Figure N4: PCR reactions run on a gel. The lanes from left to right represent: lane 1: GenRuler ™, lane 2: blank, lane 3: ISK, lane 4: Exα, and lane 5: ExPHO1.

O. Test of lyticase efficiency

In *Figure O1*, the gel image of the lyticase test is presented. This was to ensure that the lyticase actually destroyed the yeast cells prior to the colony PCR.

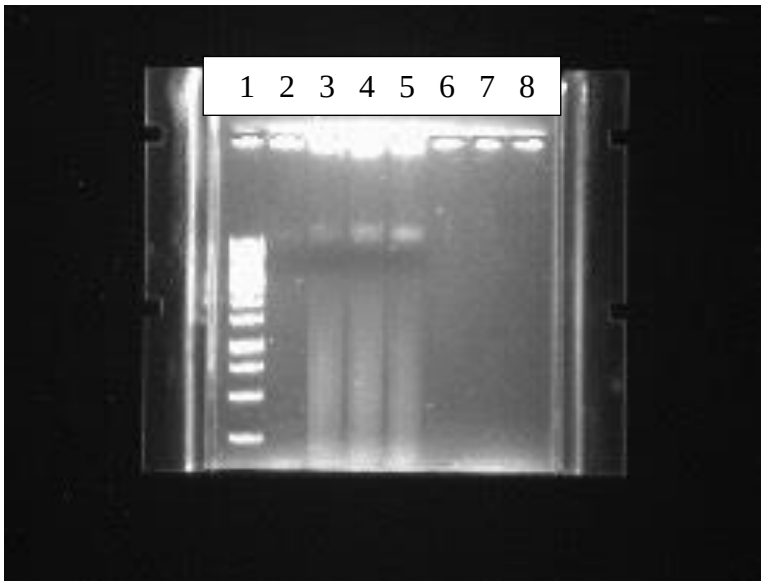


Figure O1: Test of lysis. Lane 1: GenRuler™, lane 2: unlysed sample of pPIC3.5, lane 3: pPIC3.5 (10 minutes), lane 4: pPIC3.5 (30 minutes), and lane 5: pPIC3.5 (60 minutes).