Recombinant Expression of Novel Wild Type Oat Hemoglobins in *Escherichia Coli*

Master's Thesis in Applied Biochemistry, KBKM05,

At the Division of Pure and Applied Biochemistry,

Faculty of Engineering, LTH,

Lund University

By Frank Rummukainen

Supervisors: Simon Christensen & Leonard Groth

Contacts: Simon.Christensen@tbiokem.lth.se & Leonard.Groth@tbiokem.lth.se

Examiner: Leif Bülow

Contact: Leif.Bulow@kilu.lu.se

LUND UNIVERSITY

LTH

FACULTY OF ENGINEERING

Frank.Rummukainen@gmail.com

Tel: +46738985037

Duration: 14 Feb 2022 – 19 Aug 2022

Populärvetenskaplig Sammanfattning

Hemoglobin från växtriket: en lösning på blod- och järnbrist i samhället?

Många känner till hemoglobin som det protein som ger vårt blod dess röda färg och som håller oss vid liv genom att transportera syre till våra celler. Mindre känt är att nästintill alla organismer har gener i sitt DNA som kodar för liknande proteiner, även om de inte tar formen av blod så som vi är vana vid. I det här projektet har tre nya hemoglobiner identifierats i havres DNA. De har sedan uttryckts och producerats i bakterier med hjälp av genteknik.

I vardagen påminns vi ofta om att det råder blodbrist i samhället, och privatpersoner uppmanas donera sitt blod. I trauma-situationer, vid operationer och vid uppvisad blodbrist, ges idag donerat blod till patienter. Det är ett system som fungerar, men som också har många svagheter. Ett av de största problemen idag är bristen på donatorer. En annan att blod är en färskvara och måste förvaras på specifika sätt. Att individer har olika blodgrupper nödvändiggör också en större beredskap. I de områden med allra störst behov, som i krigs- och katastrofområden, är det som mest komplicerat att lösa de här problemen. Ett syntetiskt alternativ till donerat blod skulle kunna vara en lösning på många av dessa utmaningar. Idag produceras redan komplexa proteinläkemedel mot ett flertal svårbehandlade sjukdomar. Varför skulle vi inte också kunna producera konstgjort blod så länge vi hittar rätt protein? Skulle det rätta proteinet, eller hemoglobinet i det här fallet, kunna finnas gömt i just havre-växtens DNA?

Ett annat utbrett samhällsproblem idag är järnbristanemi, blodbrist som orsakas av för lågt intag av järn. Järnbristanemi behandlas ofta med järntabletter, som kan ge biverkningar så som kräkningar, diarré och förstoppning. Riskgrupper för järnbristanemi är främst kvinnor som har mens eller ammar, gravida, och individer som äter vegetarisk eller vegansk kost. Det järn som finns i vegetarisk kost är mindre tillgängligt för upptag i kroppen i jämförelse med det järn som vi hittar i animaliska produkter och som existerar bundet till hemoglobin. En möjlig lösning på järnbristproblemet är ”Impossible Foods”, helt säkra, vegetariska produkter innehållande hemoglobin som framställts med hjälp av genteknik i mikroorganismer. Det är produkter som efterliknar köttprodukter i smak, men som också innehåller det mer tillgängliga järnet som många vegetarianer saknar i sin vanliga kost. Havre är en växt som nyligen fått mycket god publicitet för sina hälsofrämjande egenskaper och är den störst växande grödan i marknaden för växtbaserade drycker. Havre är därmed en mycket intressant kandidat för nya produkter rent marknadsekonomiskt.

Havres genom, dess DNA, är ytterst komplext och har först nyligen sekvenserats. Det här har möjliggjort vidare studier av havres proteiner. Det här projektet syftade till att optimera produktionen av de tre havre-hemoglobiner som identifierats i havres genom. Under arbetets gång visade det sig att det här var svårt. Frågan kvarstår fortfarande om det, med absolut säkert, är fungerade hemoglobiner som producerats i det här projektet. De gentekniska metoder som använts i det här fallet, kloning och transformering, har validerats i flera steg. Havre-generna finns på plats i bakterien, och experiment har påvisat att bakterien läser av den anvisade genetiska koden.
De resulterande cellerna uppvisade dock bara en rosa färg och inte den blodröda färg som en effektiv hemoglobin-produktion hade gett upphov till. Mer arbete är nödvändigt innan dessa hemoglobiner är redo för att användas i vidare applikationer.

Acknowledgements

This thesis was carried out as part, and marks the end of, my master studies at Lund University (LTH). More specifically, this thesis was performed at the Division of Pure and Applied Biochemistry. From my first year at LTH, when I first heard of the hemoglobin-project at this division in a lecture from Professor Leif Bülow, my interest was sparked. I would like to personally thank Professor Leif Bülow for recognizing my interest in this project and for giving me the opportunity to be a small part of it myself. Everyone involved in this project alongside me have always been very keen on designing and managing the project to maximize my learning outcomes, for which I am very grateful. This learning experience has been unmatched!

This project has revolved around oat hemoglobins, an area that is mostly unexplored, and for that very reason also challenging. I am very thankful towards my very knowledgeable supervisors Simon Christensen and Leonard Groth who provided the support needed to complete this project and taught me many things during its course. I highly valued the discussions we had on how to proceed with the project at different stages, and how to interpret the results gained along the way. I would also like to thank both of you for listening to my ideas on how to further develop this master thesis project.

Since this project marks the end of my studies at LTH, I would like to direct a big thank you to my family for supporting me during the past 5 years. Also, without the friends I made along the way it would not have been nearly the same and amazing experience that it was. Thank you all!

Abstract

Hemoglobin is a protein transporting oxygen through our bloodstreams and is as such essential for aerobic life. Hemoglobins also exist in plants, with a wide variety of functions other than oxygen transport. The study of plant hemoglobins could potentially be interesting for the development of synthetic blood substitutes and have also shown promise as an alternative dietary source of the more bioavailable heme-iron, other than conventional meat. Oat has recently been sequenced, which has allowed insight into its inner proteomic workings. In this master thesis, three novel hemoglobins were identified in the oat genome, transformed into *Escherichia coli* and recombinantly expressed. The project aimed to maximize the recombinant production of these hemoglobins. A high cell density and the addition of a heme-precursor (δ-ALA) were identified as key parameters for effective expression. Analysis by SDS-PAGE of the proteins produced by *E. coli* failed to unambiguously identify the sought after hemoglobins following expression, however, transcriptomic analysis by RT-qPCR confirmed oat-hemoglobin transcription. These results suggest that the folding of the target proteins was sub-optimal, resulting in unsatisfactory production of the proteins of interest. More work is needed to elucidate why these oat hemoglobins failed to fold properly and if they hold any promise for relevant applications using other methodology or a different recombinant host. This project also confirmed an upregulation (16-fold) of the key heme biosynthesis gene (*hemH*) upon successful expression of a sugar beet hemoglobin in *E. coli.*

Sammanfattning

Hemoglobin är ett essentiellt protein välkänt för att transportera syre genom våra kroppar. Hemoglobiner finns även i växter, där de kan utföra ett antal andra funktioner. Studier av dessa växt-hemoglobiner är potentiellt intressant för utvecklingen av syntetiska blodsubstitut. Rekombinant växt-hemoglobin har också visat sig vara en lovande alternativ kostkälla av heme-bundet järn, annan än konventionellt kött. Havreväxtens genom har nyligen sekvenserats, vilket har gett insyn i dess tidigare okända gener och proteiner. Sekvenseringen lade som så också grunden för det här projektet. I det här examensarbetet har tre nya havre-hemoglobiner identifierats, transformerats till *Escherichia coli* och uttryckts rekombinant. Projektet syftade till att optimera det här proteinuttrycket. Faktorer som hög celldensitet och tillsatsen av kritiska metaboliter för heme-syntesen (δ-ALA) identifierades som nyckelfaktorer för ett mer effektivt proteinuttryck. Analys med SDS-PAGE av proteinerna i de transformerade *E. coli*-cellerna lyckades inte påvisa att hemoglobinerna av intresse hade producerats korrekt, men analys av transkriptionen med hjälp av RT-qPCR påvisade att generna uttrycktes efter induktion. De här resultaten tyder på att havre-hemoglobinerna inte veckades korrekt, och i förlängning, att den rekombinanta produktionen inte var så effektiv som man kunde ha hoppats. Mer arbete krävs för att förstå mekanismerna bakom den misslyckade veckningen, och varför de här proteinerna inte veckades på samma sätt som andra liknande växt-hemoglobiner har gjort tidigare. Det här projektet fastställde också att transkriptionen av ett nyckel-enzym (*hemH*) i biosyntesen av heme uppregleras så mycket som 16 gånger när ett hemoglobin från sockerbeta uttrycks effektivt i *E. coli*.

Abbreviations

Mb	Myoglobin
Hb	Hemoglobin
NO	Nitrogen Monoxide
CO	Carbon Monoxide
sHB	symbiotic Hemoglobin
nsHB	non-symbiotic Hemoglobin
trHB	truncated Hemoglobin
RBC	Red Blood Cell
HBOC	Hemoglobin-Based Oxygen Carrier
δ-ALA	δ-Aminolevulinic acid
IPTG	IsoPropyl β- d-1-ThioGalactopyranoside
OD_{600}	Optical Density at 600 nm
CO	Carbon Monoxide
RT-qPCR	Reverse Transcription – quantitative real-time Polymerase Chain Reaction

Table of Contents

1. Introduction

1.1 Hemoglobin

The evolution towards aerobic life granted access to a previously untapped source of energy. Vertebrates evolved two mechanism of oxygen transport, a circulatory system bringing oxygen directly to the cells, and oxygen-transport by proteins such as hemoglobin and myoglobin. Myoglobin (Mb) is present in the muscle and provides oxygen to the muscle in times of need. Hemoglobin (Hb), present in red blood cells, is a protein known for transporting oxygen from the lungs to various tissues, and carbon dioxide and hydrogen atoms from the tissue back to the lungs [1]. Human hemoglobin is structurally comprised of four polypeptide chains, with two identical α-chains and two identical β-chains, effectively making it a tetramer. The subunits exist as two dimers: $\alpha_1\beta_1$ and $\alpha_2\beta_2$. Each polypeptide chain subunit folds into a so-called globin fold, which is characterized by several α-helices. This fold is identical with how the monomeric myoglobin folds. Furthermore, the structure of Hb includes four prosthetic heme-groups, each one tightly bound to one polypeptide chain [1, 2]. The structure can be seen in **Figure 1** below.

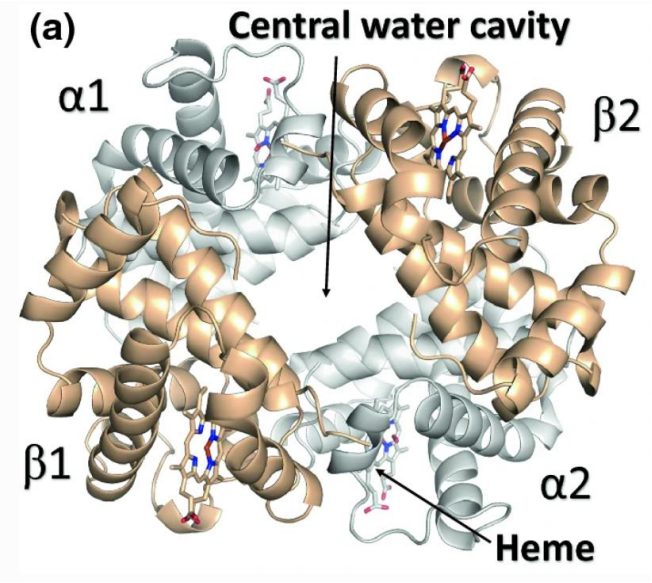

Figure 1. The structure of Hb with α-chains colored in grey and β colored in tan. Figure taken from Ahmed et al. (2020): *Hemoglobin: structure, function and allostery* [3].

Heme is an organic molecule (protoporphyrin), comprised of four pyrrole rings bound together by methine bridges, with a central iron atom. The iron atom is bound in four of its six coordination sites by four separate nitrogen atoms situated in respective pyrrole rings. The iron atom commonly exists in two states within the heme-group; ferrous (Fe^{2+}) or ferric (Fe^{3+}), but only the ferrous state can bind oxygen. On each side of the heme-plane additional bonds can be formed, on the fifth and sixth coordination sites respectively. The fifth coordination site is occupied by a histidine side chain, known as the proximal histidine. If hemoglobin binds oxygen, this occurs at the sixth coordination site [1]. Other compounds such as nitric oxide (NO) and carbon monoxide (CO) can also bind to hemoglobin. CO binds more tightly than oxygen to the ferrous iron atom, which means that oxygen

cannot displace it [2]. A bond can also be formed on the sixth coordination site by an additional histidine residue, the distal histidine. When oxygen binds to the iron atom it shifts its position slightly, placing it more in line with the heme-plane [1]. If the proximal histidine binds to the fifth coordination site of heme, this is known as a pentacoordinate Hb. If the distal histidine reversibly occupies the sixth coordination site, it is known as a hexacoordinate Hb [4]. The structure of the heme-complex in a Hb can be seen in **Figure 2** below.

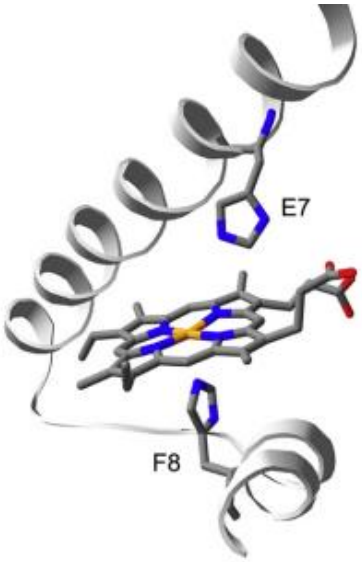

Figure 2. The structure of an open pentacoordinate distal pocket of a soybean leghemoglobin, with a prosthetic heme-group, distal histidine (E7) and proximal histidine (F8). Figure taken from Hoy & Hargrove (2008): *The structure and function of plant hemoglobins* [4].

1.2 Plant Hemoglobins

In addition to existing in vertebrates, Hb has also been found to exist in nearly all organisms, including plants. Three types of plant hemoglobins exist: symbiotic, non-symbiotic and truncated hemoglobins. Studied plants have been shown to contain at least one but also several of the different types of Hb. The Hb-types vary both in structure and function and have been suggested to have various and/or multiple different biological functions in plants [4].

Since 2014, consensus exists among researchers studying heme-containing proteins to describe and call Hb-like plant proteins by the name “Phytoglobins” [5]. In this report, different terminology might be used to describe the same plant Hbs, including plant-Hb, oat-Hb and phytoglobin. The DNA-sequences found in oat which are the key aspect of this report, the *AsPgbs*, have been named according to their origin: *Avena sativa* (As-) and their traits as phytoglobins (-Pgb). More about these Hb-encoding oat genes later.

1.2.1 Symbiotic Hemoglobins

Symbiotic hemoglobins (sHbs), were first found in legumes (more specifically in their root-nodules) that were in symbiosis with *Rhizobium*, a bacterium. SHBs have for this reason previously been

known as leghemoglobins but have since then also been found in non-legumes. The discovery of hemoglobin genes in unrelated plant families led researchers to suspect that hemoglobin genes could be present the genomes of all plants, and that it could be related also to vertebrate-Hb [6].

The symbiotic relationship mentioned above between plants and bacteria is closely related to the main function of the sHbs. The bacteria carry out nitrogen-fixation, which benefits the plant, and in return the plant regulates the oxygen levels for optimal bacterial function. The diffusion of the oxygen needed for nitrogen-fixation is facilitated by sHbs in the root nodules [4, 6]. SHbs have a higher affinity for oxygen compared to Mb, which is important for this particular application as the bacterial nitrogenase enzymes require a low free oxygen concentration [4]. Symbiotic hemoglobins may also be active in the root nodules with binding and controlling the levels of nitrogen oxide (NO) which has been reported to accumulate at the start of a symbiotic relationship as described above [7].

1.2.2 Non-Symbiotic Hemoglobins

Non-symbiotic hemoglobins (nsHBs) were discovered in plants not involved in any symbiotic relationships with bacteria. NsHbs are different from other types of Hb since they are often hexacoordinate in both oxidation states, ferric and ferrous. This is because of the distal histidine reversibly binding and occupying the sixth coordination site. The nsHbs are classified as class 1 and class 2. Class 1 nsHbs show high affinity for oxygen and low dissociation constants due to stabilization by the distal histidine. The class 2 nsHbs show more similarities to sHbs but have lower oxygen affinity. The class 1 nsHbs have previously been studied in detail in monocots such as barley and rice but have also been found in dicots such as soybean or cotton. Their presence has been linked to maintaining stable ATP levels under different stresses, e. g. flooding stress or hypoxic conditions. However, many different functions can be possible, another example given for possible mechanism of activation is nitric oxide scavenging [4]. Class 2 nsHbs have been studied to a lesser extent, but its expression has been linked to somatic embryogenesis [8].

1.2.3 Truncated Hemoglobins

The alpha helices in the globin fold of truncated hemoglobins (trHbs) are truncated, which is the cause for their nomenclature. The functions of trHbs are quite poorly understood but have been suggested to be connected to nitrogen scavenging. They have been divided phylogenetically into three groups and have been found in green alga, thale cress and barrelclover [4].

1.3 Applications of Hemoglobin

In situations of severe trauma, whole blood and blood components are usually given to patients suffering from uncontrolled hemorrhage to increase the chance of survival. The loss of red blood cells (RBCs) and their oxygen-carrying component hemoglobin is what leads to suboptimal oxygenation and possibly death. Conventionally, treatment with donated whole blood is performed, which efficiently mitigates the consequences of blood loss. However, treatment with whole blood is limited in several ways. The pharmaceutical access to whole blood is dependent on donations made by the public. In addition, blood has a limited shelf life, which makes it a logistical challenge to provide it to remote areas. There are also issues such as the need for type matching and pathogenic

contamination risks. Development of safe and efficient hemoglobin-based oxygen carriers (HBOCs) could potentially solve many problems related to treatment of hemorrhage with whole blood. These HBOCs can be semisynthetic and synthetic alternatives to RBCs, with encapsulated Hb as a key part. Some HBOCs have been in clinical trials, but none have yet been accepted for human use by the FDA. This is due to physiological risks and limitations, including NO scavenging and heme toxicity [9]. Nitric oxide is a molecule with many biological functions and interacts with all kinds of functions in the human body. NO regulates blood flow and the vascular smooth muscles. NO also controls brain blood flow, controls the cells' redox states, and promotes neuronal survival. A notable consequence to a decreased NO availability includes hypertension, high blood pressure [10].

Another potential application of recombinant Hb could be as a supplement or additive in a food product. Impossible™ Foods produces their hemoglobin recombinantly in engineered yeast. The hemoglobin, and its key component heme, is included in their vegetarian products to give them a meaty resemblance in flavor and aroma. The source of the heme in this case is a myoglobin (leghemoglobin) from soybean. The process is said to save 87% in greenhouse gas emissions, need 95% less land and 75% less water than producing the heme using meat from cows [11].

1.3.1 Metabolism of Heme-Iron

Heme-bound iron is absorbed to a degree of 15-35% and is the most easily absorbable form of iron. Heme-bound iron is conventionally derived from animal food sources such as meat or fish (Hb or Mb). Non heme-bound iron is less absorbable and usually derived from plants and iron-fortified foods [12]. Non-heme iron serves as the source of iron for vegetarians and vegans who do not consume flesh foods. Benefits of a vegetarian/vegan diet include lesser likelihood of ischemic heart disease, lower risk of cancer, reduced risk of type 2 diabetes and healthier weights in general. However, vegetarians need to consume up to 1.8 times more iron compared non-vegetarians due to the lesser bioavailability of non-heme iron. Non-heme iron is less available for absorption due to the natural inhibitors of iron absorption which exist in the sources of non-heme iron. These include phytate, oxalate and different types of polyphenols. Phytate is one of the most effective iron absorption inhibitors. Consumption of grains constitute 50% of the entire phytate intake. However, grains are also one of the most significant sources of dietary iron for vegetarians. Simultaneous intake of inhibitors such as phytate and non-heme iron is the cause for the discrepancy in iron requirements between vegetarians and non-vegetarians [13].

The absorption of iron in the human body is tightly regulated due to serious side effects with excessive uptake, while the excretion of iron on the other hand is unregulated. The limitations in uptake but not excretion can lead to iron deficiency, which in turn can lead to anemia. There are several potential causes for anemia; including hemorrhage, decreased dietary iron and decreased iron absorption. There are also physical conditions during which an individual has an increased iron requirement, such as pregnancy and breastfeeding. Deficiency of iron is the most common deficiency in the world [12].

Iron deficiency anemia can cause different symptoms such as for example tiredness, dizziness, and pale skin. Left untreated, it can lead to more serious complications such as heart problems and

pregnancy complications. The most common treatment is iron supplements, which are often given during pregnancy. This can however give raise to side effects such as vomiting, diarrhea and constipation. Other manners of treatment could be intravenous iron or blood transfusions, but these types of treatments are reserved for more serious cases [14].

1.4 Oat

Avena sativa, Oat, is a very well-known ingredient in plant-based beverages such as oat milk, which was first developed by a food scientist at Lund University. Oat based beverages are since 2020 the most popular of the plant-based beverages in Europe, also compared to soy alternatives which have the second highest sales. The market for oat beverages has a combined annual growth rate of 13.4% (2021). The consumer interest in oats is said to be due to increased sustainability compared to other options, its richness in fibers, fortification with nutrients, and its taste [15].

Oat is a global crop placed seventh in the production ranking of cereals. Oat requires less treatments with pesticides and fertilizers than other cereals and is known as a healthy source of many important nutrients. Oat contains antioxidants, polyunsaturated fatty acids, dietary fiber, and proteins. Oat stores globular proteins in their grain, compared to for example wheat, which instead stores gluten proteins. Oat is an allohexaploid, and only recently a high-quality reference genome of *A. sativa* was presented. The unveiling and publication of a high-quality reference genome allows for further studies into the traits of oat and the proteins present in the organism [16].

1.5 Recombinant Protein Production

Previous studies have produced a range of plant-Hbs recombinantly in *E. coli* and in other host organisms.

1.5.1 Previous Methodology for Recombinant Hemoglobin Expression

A Barley (*Hordeum* sp.) non-symbiotic hemoglobin fusion protein was successfully expressed in DH5-α *E. coli* with resulting extracts containing up to 5% total soluble protein. The transformed cells showed a distinct red color following fermentation. Addition of IPTG had no effect on total protein expression. The red color, together with the results from analyzing the absorbance spectra, was said to be a strong indication that a globin heme-protein could be produced recombinantly in *E. coli* [17].

A study attempting to optimize the recombinant production of deer mouse hemoglobin showed that the choice of *E. coli* strain can enhance the quality and quantity of the produced Hbs. BL21Star™ (DE3) produced the highest yield of soluble Hbs as compared to five other strains. The highest yield of soluble Hbs was produced at low temperatures with an extended induction time, with the optimal being 12 °C for 24h. *E. coli* is said to be able to maintain production in a temperature range of 10°C - 49°C, but low temperatures seem to be beneficial for protein production as cold shock proteins with important roles may be produced. Notably, the JM109 (DE3) strain was more efficient at heme incorporation during production of recombinant Hbs [18].

To enhance the recombinant production of Hb in *E. coli,* one study attempted to co-express a heme transport system from *Plesiomonas shigelloides.* The transport system is employed by *P. shigelloides* to transport heme-iron into the cytoplasm from the surroundings. Without enough intracellular heme present in *E. coli*, expressed Hb folds improperly and is degraded. To produce human Hbs, heme-permeable *E. coli* and addition of heme to the medium has previously been the solution. BL21 DE3 *E. coli* were used in this study. Spectrophotometric analysis of the cultures revealed that co-expression of Hb and heme transport system genes produced 5.8 times more Hb compared to the culture only expressing Hb. Analysis of the soluble fraction supported these results, showing a 5.3-fold increase in expression of Hb upon co-expression. No increase in expression was noted when only the transport genes were expressed. When Hb and the transport genes were co-expressed as part of the same plasmid, the study saw a 10-fold increase in soluble Hb production as compared to Hb-genes alone [19].

1.5.2 Previous Experiments on Oat Hemoglobins

In a previous master thesis [20] performed by Simon Christensen at the Division of Pure and Applied Biochemistry, oat-Hbs were identified by *in silico* analysis. Five potentially interesting Hbs showing close similarity to other monocot-Hbs were identified and three of them were expressed in *E. coli* using Gateway Recombination Cloning ([21]). An attempt to optimize the recombinant protein expression revealed OD_{600} as the most important parameter for expression, with IPTG- and δ-ALA-concentrations potentially less important but still relevant. The transformed cells yielded after expression were studied by visual inspection (red/pink color), with SDS-PAGE and with a spectrophotometric assay. The plasmid constructed was suggested to be leaking, since the Hb-genes were expressed even without induction. A possible explanation was said to be that the lac repressor protein might not be produced to the extent that transcription of Hb genes was blocked completely. Ultimately, expression was successful but in limited amounts. The thesis concluded that the purification needs to be optimized in order to study the discovered oat-Hbs further [20].

1.5.3 Safety of Recombinant Plant Hemoglobins

Leghemoglobin from *Glycine max* (soy) expressed in *Pichia pastoris* showed no mutagenic or clastogenic effects in *in vivo* and *in vitro* tests. Reproductive systems in rats were unaffected for both males and females. No toxicological effects or adverse effects were notable under treatment in rats with 750 mg/kg/d leghemoglobin (100 times more than the 90^{th} percentile daily intake) during the 28-day dietary study [22].

1.6 Biosynthesis of Heme

Heme is a central part of hemoglobin [2]. The biosynthesis of heme takes place in the mitochondria after which its transported to the cytosol to be incorporated in proteins [23]. δ-Aminolevulinic acid, also known as 5-aAminolevulinic acid (ALA), is a critical intermediate in the biosynthesis of heme. It is generally accepted that for the biosynthesis of heme in organisms such as *E. coli*, the biosynthesis of δ-Aminolevulinic acid (δ-ALA) is the main rate-limiting step. There exist two separate pathways in which δ-ALA is produced, the C4- and C5-pathways. The C4-pathway is active in mammals, fungi and purple nonsulfur bacteria, while the C5-pathway is active in most bacteria, archaea and plants. The heme biosynthesis-pathway and the C5-pathway generating δ-ALA can be seen in **Figure 3** below. δ-ALA which is produced can go on to form heme, or other metabolites such as vitamin B12 or

chlorophyll [24]. Another pathway which branches out from the heme biosynthesis pathway is the conversion to siroheme via uroporphyrinogen III, a prosthetic group present in sulphite reductases. The conversion to siroheme is catalysed by a *CysG* protein [23].

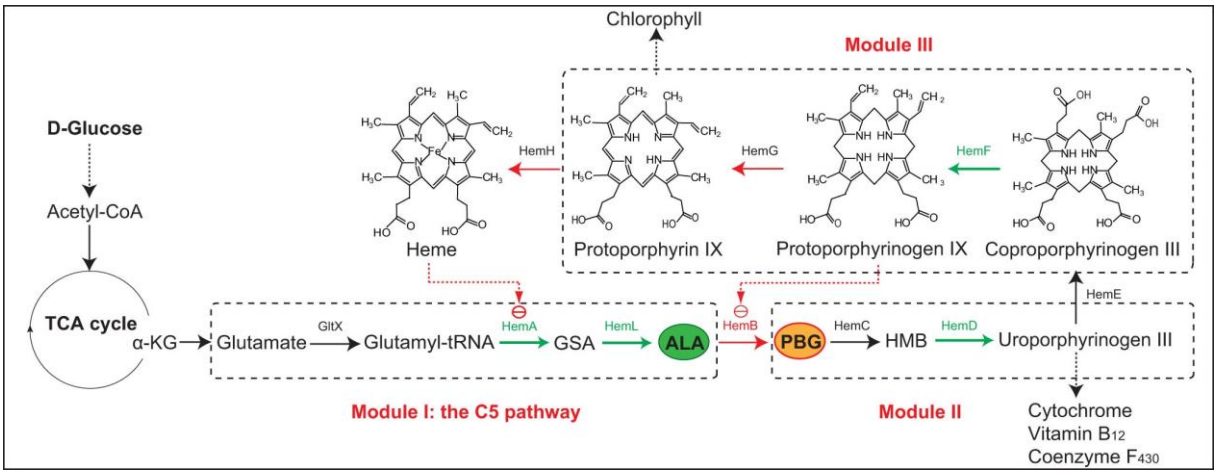

Figure 3. The heme biosynthesis and its regulation in *E. coli*. Green and red arrows indicate if the enzymes are positive or negative to heme accumulation and dotted red arrows show feedback-inhibition. Figure taken from Zhang et al. (2015): *Optimization of the heme biosynthesis pathway for the production of 5-aminolevulinic acid in Escherichia coli* [24].

Studies have shown that an upregulation of *hemD* and *hemF* lead to an increased accumulation of δ-ALA, whilst increasing expression of *hemB, hemG* and *hemH* decreased it. In addition, ALA dehydratase, encoded by *hemB,* has been shown to be feedback inhibited by protoporhyrinogen IX. As such, *hemB, hemD, hemF, hemG* and *hemH* are the main regulatory elements in the heme biosynthesis [24]. Another study showed that the final conversion of protoporphyrin IX to heme (*hemH*) could be rate-limiting since accumulation of protoporphyrin IX was observed. The *hemH*-gene encodes for the enzyme ferrochelatase which incorporates ferrous iron into protoporphyrin IX. The addition of more ferrous iron increased the synthesis of heme, but the total production of heme still did not amount to the expected carbon flow from δ-ALA. It is unclear if this was still due to limitations in the ferrochelatase-step or due to other reasons [23]. In response to heme limitations under aerobic conditions, *hemeH* and *hemeCD* showed a 4- and 2.2-fold increase in expression respectively [25].

1.7 Aim of the Master Thesis

The aim of this master thesis is to evaluate the possibilities to recombinantly produce novel hemoglobins discovered in the genome of *A. sativa*. In addition, it is the aim of this thesis to optimize the recombinant expression and, if possible, produce a large enough amount of recombinant oat hemoglobins for purification and further study. Furthermore, this project aims to increase the understanding of recombinant expression of plant hemoglobins and possibly examine if they could be candidates for relevant applications within the pharmaceutical or food industry.

2. Materials and Methods

2.1 DNA Sequences

Three DNA-sequences corresponding to hemoglobins were identified in oat's genome and codon optimized for expression in *E. coli*. In addition, *att*B sites were added to the DNA-sequences for the relevant enzymes to recognize recombination sites, more about this below. The sequences were ordered from and produced by Integrated DNA Technologies. The sequences were codon optimized for recombinant expression in *E. coli* and named *AsPgb1.1, AsPgb1.2* and *AsPgb1.3.* The sequences can be found in **Appendix 7.1 DNA Sequences**.

2.2 Cloning & Transformation

To generate functional *E. coli* expression clones for all the three different oat hemoglobins, "Gateway® Technology with Clonase® II" [21] (Invitrogen) was used. This technology is based upon a bacteriophage lambda site-specific recombination system. The technology consists of two major recombination reactions after which a functional expression vector is yielded and can be transformed into a competent cell-line of choice for final expression of the wanted gene. The two reactions are called BP and LR recombination reactions respectively. The enzymes used in the recombination reactions recognize specific *att* sites where the DNA is cleaved and where recombination is allowed to happen. After both BP and LR reactions, the plasmids produced were purified and subsequently sequenced to ensure that the correct recombination has occurred.

2.2.1 BP-reaction

In the BP-reaction, catalyzed by Gateway® BP Clonase® II enzyme mix, the *att*B-flanked gene of interest undergoes recombination with an *att*P-containing donor vector, forming an entry vector where the gene is flanked by *att*L-regions. The gene which is removed by recombination from the donor vector is a lethal gene meaning that only successfully transformed cells survive. In this project the pDONR™221 entry vector was used. This donor vector also carries a kanamycin resistance gene, which allows for selection of the entry clones after transformation. One Shot™ OmniMAX™ 2 T1R *E. coli* (Invitrogen) were transformed with the plasmid from the BP-reaction. Transformed cells were grown on selective kanamycin plates (100 µg/ml) overnight (37°C), after which a colony was chosen from the plate to be inoculated in LB media supplemented with kanamycin (100 µg/ml), again to be grown overnight (37°C). From this resulting culture of entry clones, the plasmid DNA was then purified, sequenced, and further used in the LR reaction after confirmation that the correct genes had entered the plasmids by sequencing. Negative controls without addition of the BP enzyme mix were also plated on selective media. The BP reaction was set up according to **Table 1** below.

Table 1. Setup of the BP-reaction.

Components	**Volume added to sample (µl)**	**Volume added to negative Control (µl)**
*att*B-PCR product	4	2
pDONR™ vector	1	0.5
TE Buffer	3	2.5
BP Clonase® II enzyme mix	2	-

2.2.2 LR-reaction

The LR-reaction, catalyzed by Gateway® LR Clonase® II enzyme mix, facilitates the recombination of the *att*L donor vector with an *att*R destination vector, creating a resulting *att*B expression vector. Again, a lethal gene is substituted to the gene of interest in the final expression clone. The destination vector used was the pET-DEST42 which carries an ampicillin resistance gene. One Shot™ OmniMAX™ 2 T1R *E. coli* (Invitrogen) were transformed with the plasmid from the LR-reaction. The transformed cells are grown overnight on selective plates with ampicillin (100 µg/ml, 37°C), after which a colony was inoculated into LB medium with ampicillin (100 µg/ml) to be grown overnight (37°C) yet again. Negative controls without addition of the LR enzyme mix were also plated on selective media. The LR reaction was set up according to **Table 2** below.

Table 2. Setup of the LR-reaction.

Components	**Volume added to sample (µl)**	**Volume added to negative Control (µl)**
Entry clone	1	1
Destination vector	1	1
TE Buffer	6	8
LR Clonase® II enzyme mix	2	-

The above procedure was performed as is described in the protocol given in the user guide "Gateway® Technology with Clonase® II" (Invitrogen) [21], and was followed accurately with only a few exceptions.

- No positive controls were made since sequencing was performed to validate the recombination-reactions.
- The DNA added as *attB*-PCR product (*AsPgb1.1, AsPgb1.2* and *AsPgb1.3* respectively) was added in excess compared to the recommendations in the protocol.

2.3 Plasmid Isolation

Following both BP- and LR-reactions, colonies were taken sterilely from the plates with the transformed colonies and grown in 5 ml LB media supplemented with the appropriate antibiotic (100 µg/ml) overnight. These cultures were used for the plasmid isolation.

Plasmid isolation was done following the BP- and LR-reactions respectively. This was done using an E.Z.N.A Plasmid DNA mini kit 1 (Omega Bio-tek). The quantity and quality of the plasmid DNA was analyzed using a nanodrop with a nanophotometer (Implen). The samples with the highest concentrations were sent for sequencing, as long as they were of adequate purity (absorbance rations >1.8). The selected samples were subsequently sent to Eurofins Genomics where the sequencing of the plasmids was performed by GATC© (LightRun Tube). The samples were prepared according to the specifications for LightRun Tube sample requirements [26].

The raw data from the Sanger sequencing was studied to ensure that the peaks indicating different bases were unambiguous. The given sequencing-results were aligned with the codon optimized sequences in **Appendix 7.1 DNA sequences** using the Clustal Omega Multiple Sequence Alignment online-tool [27].

2.4 Transformation of Competent Cells

Competent *E. coli* BL21 DE3 cells were, upon confirmation that the right gene was in place in each of the three separate expression vectors, separately transformed with respective expression vector resulting from the LR-reactions according to the protocol in **Appendix 7.2 Transformation Protocol**. The selective plates (100 µg/ml ampicillin) with each expression clone were kept in a cold room in preparation for fermentation and hemoglobin-expression. One colony was picked from each plate with the different expression clones and grown in 5 ml sterile LB media with ampicillin (100 ug/ml) at 37°C and 150 rpm overnight (16-18h). Glycerol stocks were made for each of the expression clones by mixing 0.15 ml of sterile glycerol with 0.85 ml of the inoculum, vortexing briefly and frozen into -80°C.

2.5 Expression and Recombinant Protein Production

In general, the expression-phase of this study can be divided into three parts: inoculum preparation, fermentation, and cell harvesting. Note that key parameters were varied in an attempt to optimize the expression of the recombinant oat hemoglobins. These parameters were varied in different rounds of cultivations. The parameters which were varied included: concentrations of isopropyl β-d-1-thiogalactopyranoside (IPTG) and δ-aminolevulinic acid (δ-ALA), OD_{600} following growth as well as temperature and time of fermentation. The reference protocol was provided by Simon Christensen, the supervisor of this project, and was based upon previous successful expression of other hemoglobins at the same department as this master thesis was performed at.

2.5.1 Inoculum Preparation

In preparation for starting a new round of fermentations for Hb-expression, a colony was picked from the appropriate plate containing the transformed BL21 DE3 *E. coli* cells and inoculated in 5 ml LB medium with 100 µg/ml ampicillin. These cells were grown overnight at 37°C and 150 rpm in an incubator.

2.5.2 Fermentation

The fermentations were run in different rounds (1-9) with different settings each time. Each round attempted to optimize a certain part of the fermentation-protocol. A table complete with the settings of each fermentation-round can be found in **Appendix 7.4 Expression Protocol.** Parameters that were altered between different round of fermentations in the optimization are marked with an asterisk (*) below.

Fermentations were performed in 500 ml baffled flasks with 100 ml of TB media. For each fermentation, carbenicillin was added to a final concentration of 100 µg/mL and 1 ml of inoculum was added to the baffled flask. The cells were allowed to grow (at 37°C and 150 rpm) for the time

needed to reach the wanted optical density OD_{600}*, the reference was slightly above 2.5, which would take around 3-4 hours. OD_{600} was measured with a nanophotometer (Implen). The expression of Hb was induced by addition of IPTG, and δ-ALA was added as a pre-cursor to heme to a final wanted concentration*, with reference concentrations of 0.5 mM IPTG and 0.3 mM δ-ALA. To any negative controls, IPTG and δ-ALA were not added. The flasks were bubbled with carbon monoxide (CO) for 10 seconds to saturate the solution and prevent oxidation of expressed Hb. The flasks were sealed with parafilm. Negative controls were not bubbled with CO but were sealed in the same way. The cells were subsequently fermented at a certain temperature* and time* in darkness, reference was 22°C and 18h.

Samples for RT- qPCR were taken before induction (addition of IPTG, δ-ALA and bubbling with CO) as well as after fermentation. The sample volumes taken were calculated based on the OD_{600} – value (as taken before induction and post-fermentation). This was based upon the estimation that one unit of OD corresponds to 8*10^8 cells/ml [28]. Furthermore, an *E. coli* cell contains 0.1 pg of RNA [29], which means that 10^9 cells would be needed to gather 100 µg of total RNA for the RNA-purification, which was done. The cells in these samples were pelleted by centrifugation at 13 000 rpm for 5 minutes in a Biofuge pico (Heraeus) centrifuge, flash frozen in liquid nitrogen and placed in -80°C until later use. The positive control *BvHb1.2* (for *Beta Vulgaris,* i. e. sugar beet hemoglobin, see **Appendix 7.1 DNA Sequences** for sequence), known to produce a lot of Hb and generate distinctively red cells, was fermented according to the reference protocol, whilst *AsPgb1.1-1.3* were fermented according to the optimized protocol.

2.5.3 Cell Harvesting

To harvest the cells, the fermentation media was centrifuged at 12 000 rpm and 4°C for 20 min in a Sorvall Lynx 4000 centrifuge (Thermo Scientific™). The supernatant was removed, and the pellet moved to a smaller tube (50 ml). The pellet was washed with approximately 20 ml of 50 mM TRIS-HCl buffer with pH 8.5. The re-suspended cells in TRIS-HCl were centrifuged again at 11 000 rpm and 4°C for 25 min in either the Sorvall centrifuge or a 5920R centrifuge from Eppendorf, depending on the number of tubes in the fermentation-round. The supernatant was thoroughly removed, and the resulting pellet studied, photographed, and weighed. The color of the pellet was visually inspected. The pellet was then flash-frozen in liquid nitrogen and placed in a -80 °C freezer.

2.6 SDS-PAGE

The frozen pellets from the cell harvesting-step were used for the SDS-PAGE. The cell pellets were re-suspended by shaking and vortexing in 50 mM TRIS-HCl pH 8.5 to a certain volume so that all the samples had the same concentration (as far as possible, with the lowest possible volume of 8 ml). The cells were re-suspended to a concentration of 0.0785 g cells/ml (if nothing else is stated). The resulting cell suspension was kept on ice and sonicated for 40s effectively (5s pulse, 3s break) with an amplification of 50% with a Q500 ultrasonicator from QSonica. The sonication was repeated twice. If the volume needed for re-suspension was >18 ml the cell suspension was divided into two 50 ml tubes during sonication, and then pooled immediately after sonication. The tubes with cell lysates were centrifuged at 11 000 rpm and 4°C for 20 min to remove cell debris. A small volume of the supernatant (30-100 µl) was mixed 1:1 with 2x SDS loading buffer. The mixture was boiled for 5 minutes at 100°C, vortexed and centrifuged briefly. Mini-PROTEAN® TGX™ (Bio-Rad) gels (4-20%)

were loaded with samples and PageRuler Unstained Protein Ladder (Thermo Scientific™). 15-well gels were loaded with 5 µl sample and 3 µl protein ladder. The 10-well gels were loaded with 10 µl sample and 7 µl protein ladder. The gels were run at 110V for 5 minutes, and until finish at 160 V (around 40 minutes). The gels were analyzed using a GelDoc XR System (Bio-Rad).

2.7 RNA - Extraction & cDNA Synthesis

RNA-extraction was performed on the non-induced and induced samples from the fermentation step. The RNA extraction was done using a Quick-RNA™ MiniPrep kit (Zymo Research). The RNA extraction was done according to the instructions provided with the kit [30]. Since the full load of RNA was used (100 µg), 600 µl of RNA lysis buffer was added and should in this case be considered "one volume" in accordance with the manufacturer's instruction. The quality and concentration of the RNA was analyzed using a nanodrop with a nanophotometer (Implen) prior to cDNA-synthesis. If the RNA-concentration was too low for it to generate 350 ng per 15 µl, the kit was re-used to concentrate RNA, starting at step 3: addition of ethanol. Concentration was attempted by elution with a smaller volume of DNase/RNase-free water. RNA was analyzed on 1% agarose gel. The gel was loaded with 4 µl of RNA sample solution and 2 µl of 6x DNA loading dye, as well as 5 µl ladder and 1 µl 6x DNA loading dye. The gel was run for 30 minutes at 120V. The extracted RNA was kept in -80°C until further use.

Using the RNA concentrations as given by the nanodrop, the appropriate volume to generate a total of 350 ng of RNA was calculated and used for the cDNA-synthesis. The RNA solution was diluted to a total volume of 15 µl using MilliQ water prior to cDNA-synthesis. The cDNA synthesis was performed using an iScript™ cDNA Synthesis Kit (Bio-Rad) in accordance with the manufacturer's instruction. The synthesis was performed with 4 µl of 5x iScript Reaction mix, 1 µl of iScript reverse transcriptase and 15 µl of diluted RNA-sample. The reaction protocol was run in a SensQuest thermal cycler (labcycler). The resulting stock solution was diluted 1:20 to a working solution used for the qPCR.

2.8 RT – qPCR

Primers for qPCR were designed using the coding sequences for the codon optimized target sequences *AsPgb1.1-1.3*, the positive control *BvHb1.2*, and the heme-synthesis protein ferrochelatase (*hemH*). The sequence used for *hemH* primer design was found on NCBI [31]. The primers were designed using primer3web (version 4.1.0) [32-34] with standard settings except for melting temperatures, T_m, which were set to: min 59°C, opt 60°C and max 61°C, and product size range which was set to 100-300 bp. Primers for the reference genes *rrsA*, *idnT* and *hcaT* were taken from a study on novel reference genes in *E. coli* for qPCR. The gene *rrsA* encodes for ribosomal 16S RNA, *idnT* for an idonate transporter and *hcaT* for a 3-phenypropionate transporter [35]. The T_m of the reference genes were calculated using a calculator by ThermoFisher [36] with a Taq polymerase. The primers were ordered from and produced by Integrated DNA Technologies. The primer sets were centrifuged briefly at 10 000 rpm for 30s and suspended in MilliQ-water to a stock solution with a concentration of 100 µM as per the manufacturer's instruction. The stock solution was briefly vortexed, spun down and diluted 1:10 in MilliQ water for a 10 µM working solution. The suspended and diluted primer sets were kept in 20°C at all times.

The qPCR was performed using Applied Biosystems™ SYBR™ Green PCR Master Mix and set up according to **Table 3** below. The order in which the components were added to the tube was consistent with that of the table. Technical triplicates were performed where 3 times the volumes of the table below (and an excess volume of 20%) were mixed first in one tube and then pipetted with an automatically dispensing pipette to three wells in the 96-well PCR-plate. This was done to minimize pipetting errors. Each well in the 96-well plate contained a total volume of 20 µl at the time of running the qPCR in accordance with the table below.

Non-induced and induced samples (*AsPgb1.1-1.3* and *BvHb1.2*) were prepared according to the recipe in the table below with one triplicate for their respective hemoglobin-genes, a heme synthesis gene and with the two reference genes, *rrsA* and *idnT*. In addition, triplicates of blanks ("No templ") were run where no cDNA was added (MilliQ water was added to the total volume of 20 µl instead). This was done for each primer set that was used on the same plate.

Table 3: The PCR-reaction was set up according to the following protocol.

Component	**Volume**
MilliQ vatten	4 µl
Master Mix (SYBR™ green)	10 µl
Forward primer (10 uM)	1 µl
Reverse Primer (10 uM	1 µl
cDNA (1:20 dilution)	4 µl → 4 * (1/20) * 350/20 ng = 3.5 ng cDNA
Total	20 µl

The qPCR was run in with a CFX96™ Real-Time System in a C1000 Touch™ Thermal Cycler (Bio-Rad). The real-time qPCR was run with the following conditions: an enzyme activation step at 95°C for 10 minutes followed by 40 cycles of denaturation at 95°C for 15s and annealing/extension at 60°C for 1 minute. A melting curve was produced following the qPCR-run by increasing the temperature from 65°C to 95°C with 0.5°C/cycle. The temperature was held for 5s per cycle and the ramp was 0.5°C/s.

Melt curve analysis was performed to ensure no non-specific amplification occurred. Quantification cycle values (C_q-values) for the reference genes were averaged and the "Livak method", as described by Gallego A. [37], was used to calculate the relative change in expression between non-induced and induced and fermented samples.

3. Results

3.1 Comparative Analysis of DNA Sequences

The DNA sequences of interest in this project can be found in **Appendix 7.1 DNA Sequences**. A multiple sequence alignment with the non-codon optimized sequences run with the Clustal Omega online-tool [27] reveals a >95 % identity between the sequences. This can be found in **Appendix 7.1.1 Multiple Sequence alignment**. Running nucleotide BLAST-searches (BLASTn) [38] on the native sequences versus common nucleotide-databases reveals homology with many different non-symbiotic hemoglobins. Some species with Hbs with a percentage identity >90% compared to the *AsPgbs* are *Lolium rigidum* (grass), *Triticum aestivum and dicoccoides* (wheat), *Zea mays* (mays), and *Hordeum vulgare* (barley). Translation of the codon optimized *AsPgbs* to their amino acid sequences and calculation of respective isoelectric points and molecular weights with the Expasy [39] "Compute pI/Mw"- tool gives the data for each sequence as given below in **Table 4.**

Table 4. Isoelectric points and molecular weights for each oat hemoglobin.

Gene	*Isoelectric point*	*Molecular weight (kDa)*
AsPgb1.1	8.65	18.3
AsPgb1.2	8.65	18.4
AsPgb1.3	8.67	18.6

3.2 Cloning & Transformation

Transformed cells following the BP- and LR-reactions yielded colonies when plated upon selective agar plates. The plates following the BP-reaction can be seen in **Figure 4** and the plates following the LR-reaction in **Figure 5**. Observe that all plates with transformed cells produced colonies while the negative control in both cases did not.

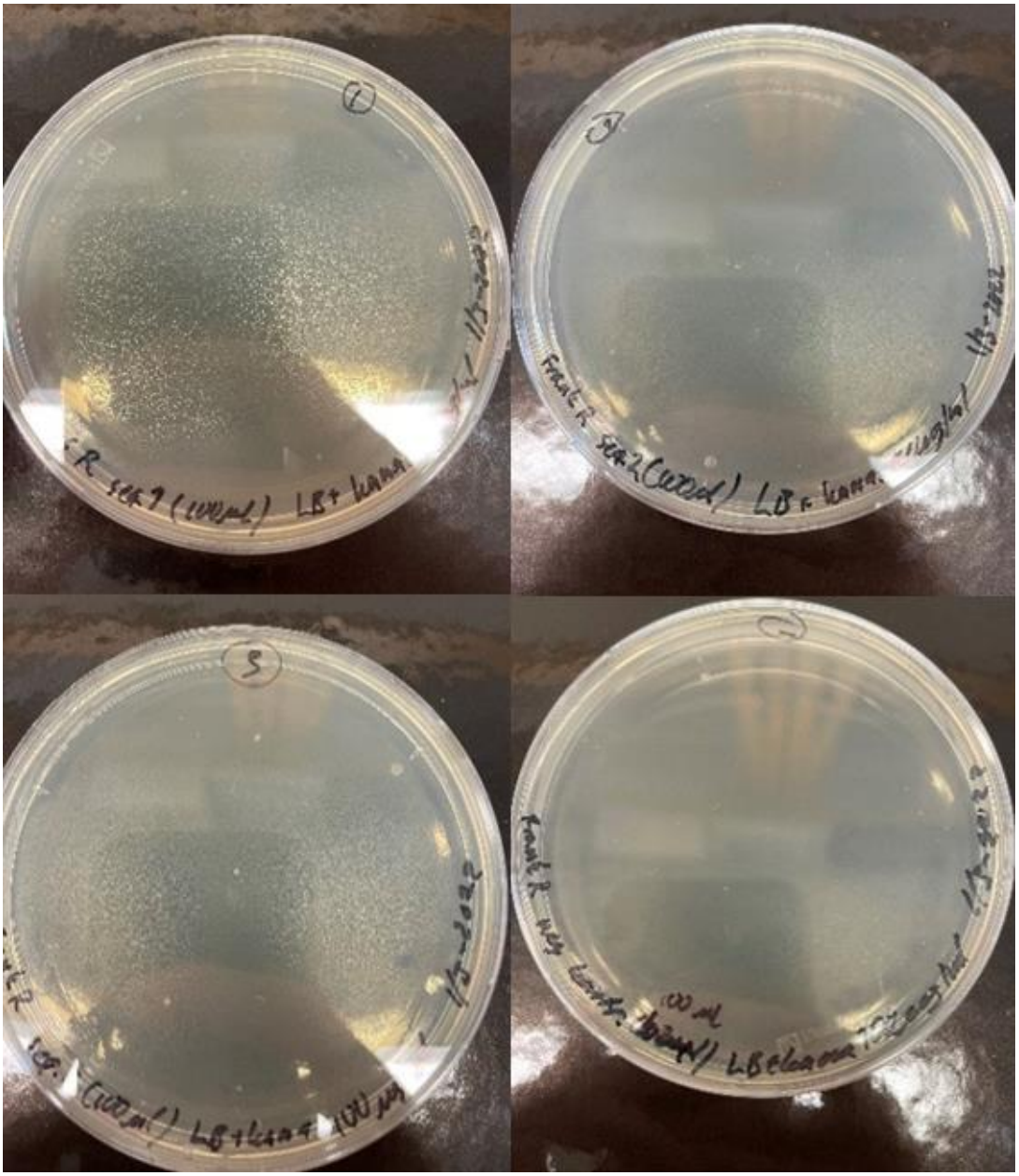

Figure 4. Transformed OmniMAX™ *E. coli* cells following the BP-reaction plated on LB-agar supplemented with 100 µg/ml kanamycin. Top left: *AsPgb1.1,* top right: *AsPgb1.2,* left bottom: *AsPgb1.3* and the negative control on the bottom right.

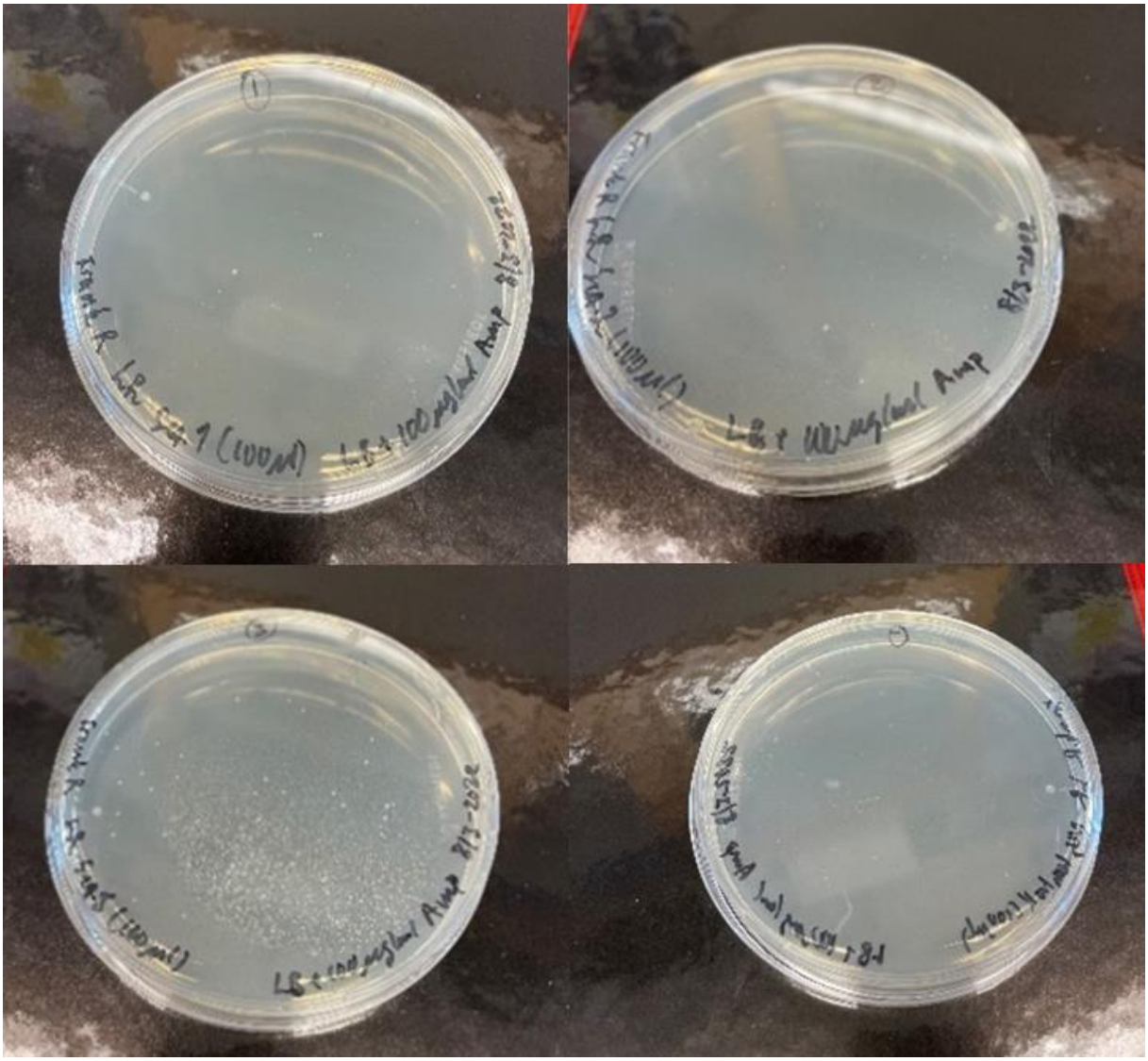

Figure 5. Transformed OmniMAX™ *E. coli* cells following the LR-reaction plated on LB-agar supplemented with 100 µg/ml ampicillin. Top left: *AsPgb1.1,* top right: *AsPgb1.2,* left bottom: *AsPgb1.3* and the negative control on the bottom right.

The quantification and quality of genomic material following plasmid purification after both BP- and LR-reactions as well as sequence alignments can be found in **Appendix 7.3 Sequencing Results**. The purification yielded plasmid DNA sufficient both in amount and quality, after both BP- and LR-reactions respectively, for it to be sent for sequencing. Good quality DNA in this case refers to the A_{260}/A_{280}- and A_{260}/A_{230}-ratios being >1.8 for all samples following both reactions. These values indicate a high DNA purity and a low presence of contaminants such as salts, sugars, or solvents.

Alignment of the sequencing results with the codon optimized *AsPgb*-sequences following both reactions confirmed that the right genes were inserted, first into the entry vectors and later into the expression vectors. Examples of the Clustal Omega sequence alignments using the sequenced plasmids as well as the template gene can be found in **Appendix 7.3 Sequencing Results** for both BP- and LR-reactions.

BL21 DE3 *E. coli* were successfully transformed with the expression vector and plated upon selective plates with 100 µg/ml ampicillin. All three strains with *AsPgb1.1, AsPgb1.2* and *AsPgb1.3* respectively yielded a large number of colonies. The plates can be seen in **Figure 6** below.

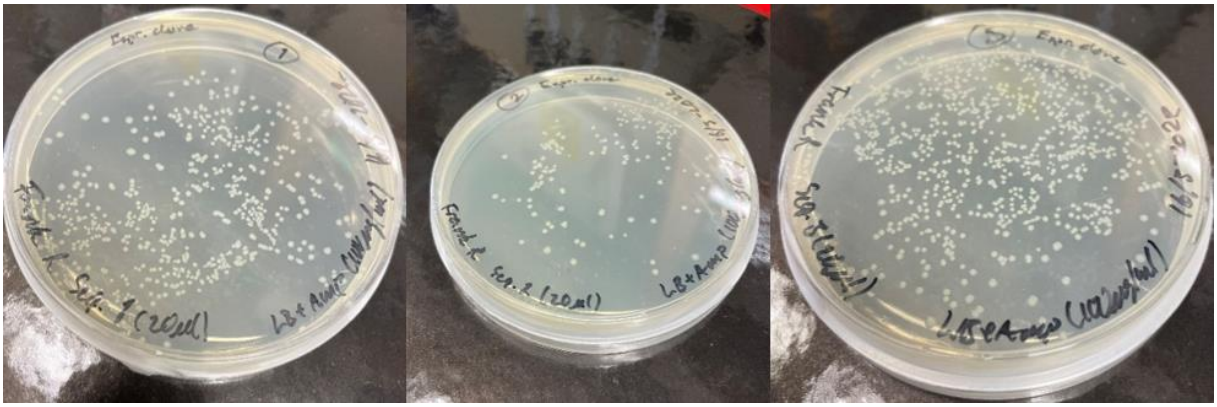

Figure 6. Selective plates with transformed BL21 DE3 *E. coli* cells. From left to right: *AsPgb1.1, AsPgb1.2* and *AsPgb1.3*.

3.3 Optimization

The optimization of the recombinant Hb-expression was done by performing several rounds of fermentations, nine in total. The eight round was performed to express all three genes, *AsPgb1.1-1.3*, with the optimized protocol. The ninth round was done to prepare samples for the RT-qPCR (**3.5 RT-qPCR**). For rounds one to seven, each round attempted to optimize a certain part of the fermentation protocol. The parameters that were optimized were IPTG-concentration, δ-ALA-concentration, OD_{600} before induction, as well as fermentation temperature and time. All the settings used, and the weights of the resulting cell pellets, for each round can be found in **Appendix 7.4 Expression Protocol**.

The first fermentation round tried all three expression clones, *AsPgb1.1, AsPgb1.2* and *AsPgb1.3* induced according to the reference protocol, and non-induced (-) without the addition of IPTG, δ-ALA or bubbling with CO. Visual inspection of the resulting pellets following cell harvesting showed that the all the induced cultures had a darker, more pink color tone, with *AsPgb1.*3 showing the clearest pink color. The induced culture with *AsPgb1.3* also weighed less compared to the *AsPgb1.1*- and *AsPgb1.2*-cultures whilst the non-induced cultures showed similar weights. The pellets resulting from this round of fermentation can be seen in **Figure 7** below. *AsPgb1.3* was chosen for the continuous optimization due to its pinker color tones, potentially indicating more Hb-expression. As such, the following fermentations for optimization (rounds 2-7) were done with cells transformed with *AsPgb1.3*.

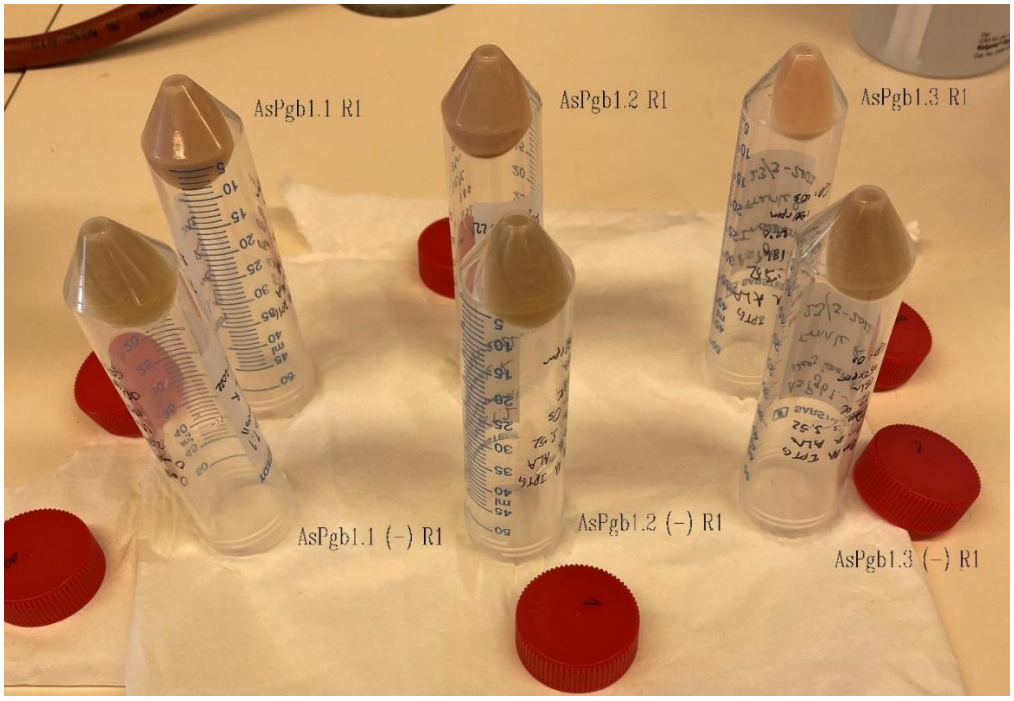

Figure 7. Cell pellets from the first fermentation round (R1). Position of the pellets indicated in the picture. (-) indicates non-induced sample. Fermentation conditions given in **Table A3** in **Appendix 7.4 Expression Protocol.**

The second round of the fermentations explored how addition of different amounts of IPTG, and δ-ALA affected the color and expression of Hb. The IPTG concentrations tried ranged from 0.2-0.8 mM as final concentrations in the media. δ-ALA concentrations ranged 0.1-1 mM. The pellet that was visually most promising, most pink/red, was the combination of 0.5 mM IPTG and 1 mM δ-ALA. IPTG concentration had no significant effect on color. The third round elaborated on the effect of additional δ-ALA with concentrations of 2-5 mM (at 0,5 mM IPTG). In **Figure 8** below, a picture of the cells from round 2 can be seen, and in **Figure A4** in **Appendix 7.5 Cell Pellets** the cells from round 3 can be seen. By visual inspection, and by weighing in the cost/benefit of using higher concentrations of δ-ALA, 0.5 mM IPTG and 1 mM δ-ALA were deemed most promising, and these concentrations were further used. The addition of δ-ALA above a concentration of 1 mM gave little noticeable increase in pinkness and δ-ALA is quite expensive.

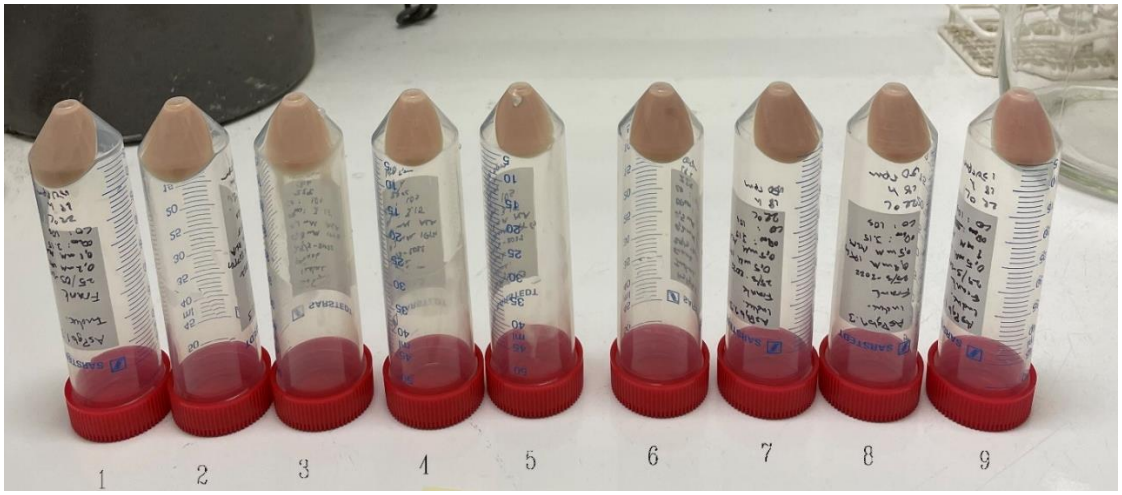

Figure 8. Cell pellets from the second round of fermentations (R2). The numbers in the picture correspond to the number of each pellet as given in **Table A3** in **Appendix 7.4 Expression Protocol**. From left to right: increasing δ-ALA concentration (1-3; 0.1 mM, 4-5; 0.3 mM, 6-8; 0.5 mM, 9; 1mM). IPTG concentration in series of three (1, 4, 6; 0.2 mM, 2, 7, 9; 0.5 mM, 3, 5, 8; 0.8 mM).

In the fourth round the OD_{600} (pre-induction) was varied. OD_{600}-values in a range of 0.1-9 were tried. Sample 1 was induced without prior growth and not allowed to grow before fermentation, sample 2 was induced with IPTG and δ-ALA allowed to grow for 3.5 h at 37°C and 150 rpm and then induced with CO as it usually would, and sample 3 was treated the same as sample 2 but was also bubbled with CO prior to growth. Samples 4-10 were prepared and treated as usual but stopped in their growth at a target OD_{600}, which would give different ODs prior to induction and fermentation. The result of the fourth round can be seen in **Figure 9**.

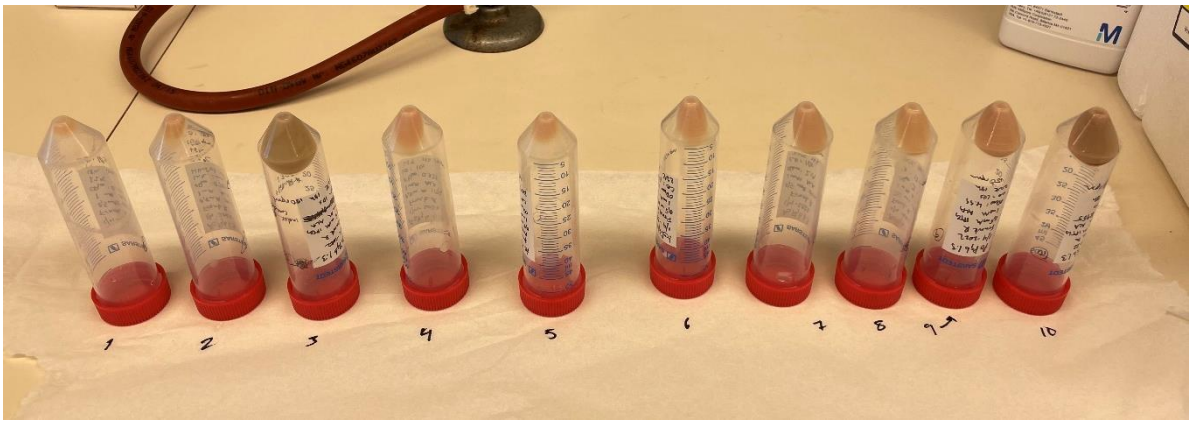

Figure 9. The cells harvested from the fourth fermentation round (R4). The numbers correspond to the number of each pellet as given in **Table A3** in **Appendix 7.4 Expression Protocol**. From left to right increasing OD_{600}. 1; 0.104, 2; no measurement, 3; no measurement, 4; 0.339, 5; 0.92, 6; 1.376, 7; 2.19, 8; 3.9, 9; 4.99, 10; 8.955.

From **Figure 9** above, and from the visual inspection done upon cell harvesting, sample 9 (OD_{600}=4.99) was deemed the most interesting in terms of color (pinkness). Samples 1-8 were bleaker and less colorful and sample 10 was brown rather than pink. As such, the target OD_{600} for the following rounds was increased to around 4-5 instead of slightly above 2.5.

In round 5 the temperature was varied, and fermentations were run at 12°C, 26°C and 30°C, as compared to the reference temperature of 22°C. No significant difference in color was noted when the temperature was varied, except for lower temperatures (12°C) where the color was less pink compared to the reference of 22°C. For higher temperatures (26°C & 30°C) the pellet seemed less pink the higher the temperature was. The optimal temperature following these rounds was still regarded as 22°C as before. The cell pellets from round 5 can be found in **Figure A5** in **Appendix 7.5 Cell Pellets.** In round 6 the time of fermentation was varied, with 12h and 24h of fermentation as opposed to the reference time of 18h. A picture of these cells can be found in **Figure A6** in **Appendix 7.5 Cell Pellets.** The 24h-culture appeared more pink than the 12h-culture, but compared with the reference 18h-culture no significant change in color could be observed. Round 7 featured an elaboration of rounds 5 and 6 with different combinations of times and temperatures, including a culture fermented at reference-conditions of 22°C and 18h for easier comparison. The 22°C- and 24°C-cultures had more color as compared to 18°C but the 18h-culture at 22°C had more color than the 16h-culture at 22°C (see **Figure A7** in **Appendix 7.5)**.

From the eight round of fermentations one can observe in **Figure A8** (**Appendix 7.5)** that the optimized protocol (from this optimization) produced cell pellets from all the transformed cells (*AsPgb1.1-1.3*) more similar to one another in pinkness as compared to the first round, where only *AsPgb1.3* had pink elements of color. In the fermentation performed for the RT-qPCR, round 9, *BvHb1.2* transformed cells fermented according to the reference protocol produced a lot more pink-/red-like cells than *AsPgb1.3*-cells with the optimized protocol (**Figure A9** in **Appendix 7.5**).

3.4 SDS-PAGE

The transformed *AsPgb*-genes have a molecular weight of around 18.3-18.6 kDa, and should as such, following the SDS-PAGE, present themselves close to the level of the 20 kDa reference band. SDS-PAGE run on the samples from the first round of fermentations (reference protocol) with induced and non-induced cultures as well as from the eight round where all the cultures were run with the optimized protocol can be seen below in **Figure 10**. Around 20 kDa there is quite a thin band (marked with an arrow in the figure) which could potentially correspond to *AsPgb*-proteins. This band looks like it is present for both induced and non-induced samples, with a bit higher intensity for all the induced cultures. The bands are potentially a bit more determined for the samples from round 8 run with the optimized protocol.

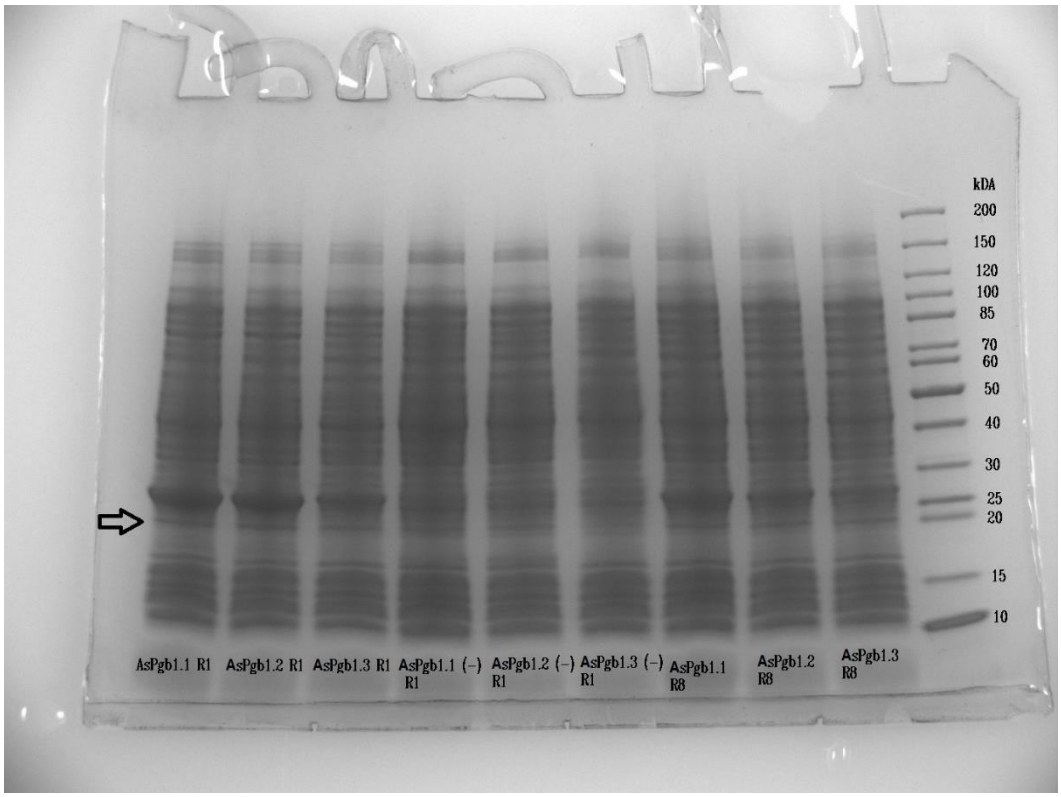

Figure 10. Gel from SDS-page analysis of samples from round 1 (R1), with induced and non-induced (-) samples, and round 8 (R8) of fermentations. Position as indicated in the figure. Protein ladder on the far-right lane with corresponding weight of each band. Arrow indicating possible bands of *AsPgb*-proteins.

SDS-PAGE results from the other rounds of fermentation did not produce any clear results, as most cultures produced a very similar band around the 20 kDa mark. The other gels can be found in **Appendix 7.6 SDS-PAGE Gels**. Notably at very low OD_{600} (sample 2, round 4) at very low temperatures (12°C, sample 1, round 5) and by bubbling with CO before growth (sample 3, round 4) the distinct band at around 20 kDa seems to disappear. Sample 3 from round 4 also produced a very different expression pattern, with darker and broader bands.

3.5 RNA – Extraction & Purification

The results of the RNA purification can be found in **Table A4** in **Appendix 7.7 RNA – Extraction**. The second round of RNA – extractions were done since the first round produced RNA-extracts with too low concentrations for some of the samples to proceed to cDNA synthesis, even after a second concentration. The second round of extractions produced RNA with much lower A_{260}/A_{230} – ratios compared to the first round and compared to the concentrated samples. The A_{260}/A_{280} were quite similar for the first round, the concentration, and the second round, and all of them were >1,8 indicating a high degree of purity.

The RNA resulting from the first round of extractions was analyzed on an agarose gel. The resulting gel showed bands rather than smears of RNA, which indicates that the RNA was not degraded during extraction. A picture of the agarose gel can be seen in **Figure A13** in **Appendix 7.7 RNA – Purification.** The synthesis of cDNA from the extracted RNA was successful.

3.6 RT – qPCR

The primer sets designed for this experiment are listed in **Table A5** in **Appendix 7.8 Primer Sequences**. Two separate 96-well plates were run. The positions of each sample are given in **Appendix 7.8 qPCR Plates.** By studying the melt curves on the CFX96^{TM} Real-Time System (Bio-Rad) from the first qPCR-plate and noticing no double-peaks, indicating that no unspecific priming or primer-primer annealing occurred, it was deemed reasonable to run the second plate with the same primer pairs. The peaks of the melt curves are given in **Appendix 7.10 Melt Curve Results.** From the first plate, no qPCR-results for the negative control of *AsPgb1.1* – transformed culture for reference gene *rrsA* was produced. On the second qPCR - plate results were successfully produced for all samples with *AsPgb1.2, AsPgb1.3* and *BvHb1.2* in technical triplicates which could proceed to relative quantification by the Livak method.

Raw data on C_q, the quantification cycle i. e. the number of cycles needed for the fluorescent signal to cross the threshold, for each sample are presented in **Appendix 7.11 Data – qPCR**. For most induced samples the C_q-values were below 29, indicating abundant target nucleic acid, and for non-induced samples the C_q-values were generally higher, indicating less target DNA present. For no template reactions to which no cDNA was added, the qPCR yielded C_q-values mostly above 35 and generally closer to 40, indicating minimal amounts of nucleic acid. The reference genes had very similar C_q-values before and after induction. The schematic overview of the calculations done according to the Livak method on the relative expressions of the different genes can be found in **Table A12**. The results of the calculations are summarized in **Table 5** below**.** The levels of expression for the hemoglobin genes in each transformed culture increased 12-25 times upon induction and following fermentation. The heme synthesis gene was upregulated mainly for the *BvHb1.2*-transformed culture, the positive control, with a 16-fold upregulation. The *AsPgb*-cultures showed inconclusive results on potential upregulation of the *hemH*-gene.

Table 5. Relative expression of Hb-genes and *hemH* as compared to reference genes *rrsA* and *idnT* after induction and fermentation.

Sample	**GOI1: Hemoglobin gene (*AsPgb1.1-1.3* & *BvHb1.2*)**	**GOI2: Heme synthesis gene (*hemH*)**
AsPgb1.2	25,11	1,29
AsPgb1.3	12,73	0,35
BvHb1.2	14,27	16,06

4. Discussion

4.1 Oat Hemoglobins

The three hemoglobins *AsPgb1.1-1.3* that were found in the newly sequenced oat genome showed more than a 95% identity between one another. They also showed a high percentage of identity (>90%) with other plant-Hbs (see **3.1 Comparative Analysis of DNA Sequences**). One of those is for example monocot barley-Hb, of which a variant has previously been successfully cloned, transformed, and expressed recombinantly in *E. coli*, generating distinctively red cells [17]. The sequences recently discovered in oat's genome are of interest since previous studies have attempted to explore possible oat-Hbs only by *in silico* analysis [20]. In the case of this study, the phytoglobins of interest have actually been identified directly from the genome of *A. sativa.*

4.2 Cloning & Transformation

Cloning and transformation of the three Hb-genes found in oat's genome using the Gateway® Technology (Invitrogen) yielded successful results. This is evident since transformed cells following the BP- and LR-reactions survived and produced colonies on selective medium whilst the negative controls following both reactions produced no colonies (**Figure 4** & **Figure 5**). The correct insertion into both the entry vector and expression vector of the genes of interest was also confirmed by plasmid purification, sequencing, and alignment thereof (**Appendix 7.3 Sequencing Results**). The sequencing results were deemed to be reliable by studying the raw data from the Sanger sequencing and noticing no ambiguity in the base identifications. This is also an indication that the plasmid DNA purification was efficient since sequencing would produce less clear results in the presence of contaminants. The genes in their entireties were confirmed to be present in both vectors by sequence alignments. The transformation of the expression vector into the expression strain of *E. coli* was deemed a success since BL21 DE3 *E. coli* could produce colonies on selective medium supplemented with ampicillin following transformation (**Figure 6**). These transformed BL21 DE3 *E. coli* could also grow in LB media with ampicillin and TB media with carbenicillin, a semi-synthetic ampicillin analog, which further increases the likelihood of successful transformation. As such, since the expression vector was seemingly correctly constructed and BL21 DE3 *E. coli* were successfully transformed, the project could progress into the optimization phase where the oat-Hbs were expressed.

4.3 Optimization of Recombinant Expression

The effect of different alterations to the expression protocol was primarily studied visually by studying the color of the harvested cells. Previous studies [17, 20] have identified red cell pellets resulting from effective Hb-expression, and as such the more red/pink tones were produced the more Hb was assumed to be expressed by the transformed cells. One might argue this subjective, which it very well could be, but pictures were also taken to follow the process and for the reader of this report to study. In addition, SDS-PAGE was performed in an attempt to validate any conclusions drawn from the visual inspection.

The first fermentation, performed with the reference protocol, in which induced and non-induced cultures were harvested (**Figure 7**), produced vaguely positive results in the sense that there was a notable difference between the induced and the non-induced cultures for all three different genes.

This could mean that the Hb-genes were expressed and yielded the wanted proteins. The color of the induced cells was however near the intensity of the *BvHb1.*2-transformed cells (**Figure A9**). *AsPgb1.3* produced cells showing the most resemblance to *BvHb1.2* (more red/pink cells), possibly due to more Hb being expressed as compared to *AsPgb1.1-1.2*. The lower pellet weight of induced *AsPgb1.3* compared to induced *AsPgb1.1* and *AsPgb1.2* could also be an indication of increased expression of Hb, in the sense that more Hb was produced instead of biomass (**Table A3**). *AsPgb1.3* was as such deemed most promising of the three oat Hb-sequences and worked with in the optimization.

The reference protocol was developed and optimized at the Division of Pure and Applied Biochemistry (Lund University) for the recombinant expression of other plant-Hbs and was as such deemed a good place to start studying the expression of the newly discovered oat-Hbs. The development of the reference protocol was said to be quite a cumbersome process, which motivated a continuation of the work with optimizing the expression of the new oat-genes, even though the initial fermentation produced less than optimal results. Previous studies identified strain-choice [18], fermentation temperature and time [18], the optical density (OD_{600}) and possibly concentrations of IPTG and δ-ALA [20] as possible factors affecting expression of recombinant Hb. The strain used in this study is very similar to the previously reported most effective *E. coli* strain for recombinant Hb-expression and as such the strain choice was not part of the optimization.

The optimization eventually yielded unsatisfactory results. Alterations of fermentation time and temperature and IPTG concentration showed little to no effect on the color of the cells. With an increased addition of δ-ALA (1 mM), and at 0.5 mM IPTG, a darker pink pellet was observed (**Figure 8**). The reason for this could potentially be an increased production of recombinant oat-Hb, or it could simply be due to an increased synthesis of heme and other heme-containing proteins or molecules (e. g. siroheme). OD_{600} has previously been noted as one of the most important parameters for Hb expression. The optimization done in this project seems to confirm that. An OD_{600} of around 5 seems to produce the most promising cells in terms of color. At OD_{600} closer to 9 we see a more brown-like colored pellet, which could indicate that the iron in the produced Hb is oxidized, and perhaps not as interesting for further applications. Expression of all the three Hbs using the optimized protocol with higher concentration of δ-ALA and higher OD_{600}, produced cells more similar to each other in color compared to when they were expressed the first time with the reference-protocol (**Figure A8**). If this was because of increased expression of Hb overall or simply due to chance is hard to evaluate. The optimization concluded that variations of the evaluated parameters did not increase the production of recombinant *AsPgb*-proteins to the level of *BvHb1.2*, or even close. The results from the SDS-PAGE, are discussed related to these findings in the following section. Possible reasons for the limited expression of *AsPgbs* are also discussed below.

4.4 SDS-PAGE

Running the SDS-PAGE, all samples were diluted to have the same cell concentration (unless stated otherwise) and as such comparable in amounts of total protein. Since cells are lysed, all intracellular and extracellular soluble protein should be present in the centrifuged lysate used for SDS-PAGE.

The *AsPgb*-Hbs have projected molecular weights of around 18.5 kDa. The only band that could potentially correspond to these proteins is the one marked in **Figure 10** around the 20 kDa mark. This band, as can be seen from **Figure 10**, differs between induced and non-induced cultures, with clearer bands following induction and potentially more defined bands after optimization. For the non-induced cultures, a band corresponding to the Hbs should normally not be expected. In this case, the presence of a fainter band at the target weight can be due to several reasons. A previous study [20] made similar observations using the same expression vector but with different Hb-sequences. It was discussed in said study that a reason could be a leaking expression vector. Meaning that protein was recombinantly expressed also without addition of IPTG for induction. The explanation given that a lack of lac repressor protein would result in insufficient blockage of transcription seems as unlikely in this case as in the previous study, but no alternative explanation can be given in this case. Since the samples used for SDS-PAGE in this study contained a lot of different proteins, it is also possible that the 20 kDa band corresponds to some other protein. This could also explain the unexpected bands of the non-induced cells around the 20 kDa mark. The induced and non-induced cultures show several differences in their protein contents, also other than the above mentioned. Another clear difference between the non-induced samples and induced samples is a band between 25-30 kDa, which is a lot more defined and bigger for induced samples compared to non-induced samples.

For most other samples from other rounds of fermentations, the band at around 20 kDa is present quite consistent in shade and size (**Appendix 7.6**). This could be an indication that the changes to the expression protocol performed in the optimization had very limited effect on protein expression, which is also consistent with the conclusion that can be drawn from the visual inspection of the cells as discussed in **4.3 Optimization**. However, assuming that the bands at 20 kDa are the *AsPgbs* of interest, some samples which present fainter bands can be used to determine which parameter-settings were less favorable for recombinant Hb-expression. Although a very low cell concentration was used for sample 2 in round 4 (**Figure A11**) due to the low OD_{600}, the band around 20 kDa is less distinguishable. At the lowest temperature (12°C) tried (sample 1, round 5, **Figure A12**) the band is indistinguishable, even though the same cell concentration was used as for most other samples and the fact that it seems to have a higher level of protein expression overall (darker, broader bands). A higher level of protein expression overall could be due to induction of cold shock proteins. The sample that was bubbled with CO before growth (sample 3, round 4, **Figure A11**) shows no particularly distinguishable band around the 20 kDa mark. The protein expression seems to be higher overall in this sample however, which is interesting, and suggests flushing with CO is effective in terms of either inducing cell growth, protein expression or serves to stabilize a range of different proteins. The pattern of this sample differentiates itself in overall protein amount from the culture which was treated the same except it was bubbled with CO first after growth (sample 2, round 4, **Figure A11**). The OD_{600} of this sample was not taken since it was bubbled with CO before growth and allowed to stay sealed for the fermentation as well, making it harder to discuss if the increased protein expression is due to increased cell mass or not. The weight of this pellet was quite a lot higher though, indicating a higher cell mass and growth.

4.5 RT – qPCR

The RT-qPCR was performed to elucidate whether the genes of interest were expressed or not in the transformed cells. This was thought to be able to bring clarity into why there wasn't more distinct expression of oat-Hbs. If gene expression could be shown, this could give clues to in which step the production of the recombinant oat-Hbs had failed. In addition, the expression of a key enzyme in the heme synthesis was studied in an attempt to link any findings of recombinant Hb-expression to the biosynthesis of heme.

The enzyme ferrochelatase (*hemH*) was chosen as an interesting target for this part of the experiment since it is the last step in the heme synthesis. It being active in the final step could decrease the risk of increased *hemH* expression upon δ-ALA supplementation since other by-products had branched off. *HemH* has also previously [24] been identified as a main regulatory element in heme synthesis. The enzyme has been shown to be active in potentially rate-limiting step [23], making the study thereof interesting from a purely metabolic point of view. It has also been shown to be prone to differ in levels of expression under specific conditions [25], indicating regulatory features.

The idea behind the experimental design of the RT-qPCR was to mimic key aspects of the optimization. This was achieved by performing a ninth round of fermentation to sample properly and consistently for this experiment. The *AsPgb*-transformed cells were expressed with the optimized protocol, since this showed the greatest promise of recombinant expression. *BvHb1.2*-transformed cells were expressed using the reference protocol, which was known previously to function well and produce cells expressing enough recombinant Hb to give a distinct, and red, cell color. Samples taken before induction, should be regarded as negative controls in this case, as no gene expression should be expected without induction. Samples taken after fermentation should theoretically show expression of target sequences. *BvHb1.2* was known to be expressed and could as such be used as a positive control in this fashion.

The reference genes used to evaluate the relative expression of target genes were taken from a study on the very subject of reference genes in *E. coli* for quantifying protein overexpression. The study identified three novel reference genes: *hcaT, idnT* and *cysG*. It also discussed *rrsA* as a commonly used reference gene [35]. *CysG* was deemed unfit for use in this report since it encodes a protein taking part in a conversion closely related to the metabolism of δ-ALA [23], and supplementation with δ-ALA could therefore affect transcription levels. Primers were taken from the article for *idnT, hcaT* and *rrsA*, but in the end only those for *idnT,* and *rrsA* were used since they were deemed to be unaffected by the experimental conditions. Results showed similar levels of expression of the reference genes before and after induction and fermentation, suggesting they were quite stable in expression during the changing conditions. In addition, the Livak method used for calculating the relative expression normalizes the expressions of the GOIs to that of the reference genes, which compensates for any general differences in protein expression when conditions are varied, and samples are taken at different times. The reference genes were as such deemed fit for purpose. Primers were also successfully and specifically designed for the other target genes, *AsPgb1.1-1.3, BvHb1.2* and *hemH*. For *hemH* the same primers were used for all the cultures, for the results to be comparable.

As samples were taken for RNA extraction to be further used in the qPCR, the proper amount of RNA was calculated from OD_{600}-values. This ensured the proper amount of RNA was used in the RNA-extraction. The RNA-extraction was done as fast and accurately as possible with the purified RNA flash frozen as quickly as possible following the extraction to minimize the risk of degradation. By running the RNA-extracts from the first extraction on agarose gel it was possible to confirm that no degradation occurred when following the protocol in question (**Figure A13**). The concentration of the first round and the preparation of the second round of samples were performed according to the same protocol as before. Resulting extracts were not run on agarose, but assumed to be as non-degraded as the first round of RNA-extracts. The main factor for not double-checking this and running all extracts on agarose gel was a lack of time. The purity of the extracted RNA varied quite a lot, especially between the two rounds (see **Table A4**). However, the A_{260}/A_{280} was high enough to proceed. The low A_{260}/A_{230} would be problematic for sequencing, but for the further application of qPCR, this should not affect the results significantly. The RNA-extracts used for the cDNA synthesis were chosen primarily based upon their higher concentrations, which had to be high enough to generate 350 ng of RNA per 15 µl. If several samples met that criterion the sample with the highest A_{260}/A_{280} and A_{260}/A_{230} was chosen, signifying a higher degree of purity.

Melt curves from the first qPCR-plate showed no double-peaks, which can be seen from **Table A8** and **Table A9**. This is a strong indication that the primers were specific enough and did not anneal to one another. As such, the same primers were used in the second qPCR-plate were the same as those used on the first plate. The samples for *AsPgb1.1* were also run on the first plate, which for some wells yielded no results (see **Table A10**). This eventually led to the exclusion of *AsPgb1.1* – transformed samples from the results, as there was deemed to be a lack of time for re-working them. An argument could be made that *AsPgb1.1* could be comparable to *AsPgb1.2* and *AsPgb1.3* in the RT-qPCR. This could be regarded as reasonable since *AsPgb1.1* cultures and samples had been produced in the same way as those for *AsPgb1.2* and *AsPgb1.3.* In addition, since all the *AsPgb*-sequences are very similar, it is likely that any issues related to their expression would be closely related, if they are not due to any experimental conditions. The samples from the different cultures had also behaved very similarly and produced comparable results in previous experiments (optimization and SDS-PAGE).

An important aspect of running the qPCR in this master thesis project is that the PCR efficiency was not determined since no DNA dilution series was prepared due to a lack of time. This also means that no absolute quantification could be done, but rather only a relative quantification. However, a relative quantification is also what is most interesting in this case. The relative quantification still allows for comparison between non-induced and induced samples and was deemed to be able to bring clarity into some of the problems with the other results. Since no PCR efficiency was calculated, it was assumed to be 100%. Note, to be able to use the Livak method reliably the efficiency should be near 100%, but since it was not calculated it was the only assumption that could be made.

During the qPCR some “No templ”, without cDNA, reactions yielded C_q-values, which was unexpected (compare **Tables A6** & **A7** with **Tables A10** & **A11)**. If this was due to contamination of some wells or how the machine calculated the values is hard to tell. These C_q-values were however high in general,

which means that many cycles needed to be run before the signal crossed the threshold above the baseline, and that any contamination amounts were quite low. C_q-values appearing for "No templ" reactions seemed to be quite random, indicating random errors rather than a systematic one. The baseline and threshold were set by the machine and could have perhaps been adjusted to further enhance the qPCR-results and remove some ambiguity associated with these "No templ"- reactions.

The qPCR yielded results which aligned quite well with the expected results. As of before the qPCR, the gene insertion was confirmed, and the transformation of expression cells was deemed a success. However, since the cells did not turn red this raised the question of whether or not the gene was expressed or if the limitations of expression related to some other factor. The results of the relative quantification (**Table 5**) revealed recombinant gene of interest expression and a transcriptional upregulation thereof upon induction, similar for *AsPgbs* as that of the positive control *BvHb1.2.* A conclusion from this could thus be that limitations in expression had nothing to do with the induction and transcription of the recombinant Hb-genes.

Another interesting factor is the heme synthesis gene. *HemH* showed increased transcription (see **Table 5**) for the positive control *BvHb1.2* which produced red cells, but not for the *AsPgbs* which did not produce red cells. Notably, the *BvHb1.2* cells were also given less δ-ALA than *AsPgbs* in accordance with the reference protocol. In general, more substrate-availability would upregulate forward reactions, in this case the biosynthesis of heme. There could be several reasons for this observed difference. It seems like the expressed *AsPgbs*-proteins do not fold properly, and thus the cells do not turn red upon expression. This improper folding could potentially be connected to the heme biosynthesis in several ways: either the heme is not synthesized in large enough amounts, or the expressed proteins do not incorporate the produced heme effectively, and therefore cannot fold properly. The first reason is countered by the fact that *BvHb1.2*-cells managed to produce enough heme to form red cells. However, perhaps the *AsPgb*-proteins are not as readily recognized by *E. coli* as a hemoglobin needing heme to fold properly, and thus heme biosynthesis is not upregulated and expressed protein is degraded instead of folding properly. If the case for the discrepancy in *hemH* transcription is due to the second reason of the above-mentioned reasons the sequences might have to be mutated for the proteins to fold properly in *E. coli,* or a different host organism would have to be used with has more advanced folding mechanisms. An example of such an organism could be for example *P. Pastoris* which has been used successfully in the past. However, the *AsPgbs* show significant sequence similarities with certain barley-Hbs, and barley-Hb has previously been expressed and most certainly successfully folded properly in *E. coli* ([17]).

4.6 Future Considerations

From the results of this master thesis, it is evident that the oat-Hbs of interest were not effectively produced in large enough amounts for further study, and expression was only partially confirmed. It would be interesting to further analyze the results of the recombinant expression system used here by either doing a spectrophotometric assay or a western blot to confirm that the proteins of interest were present in a properly folded state. This could also be used to confirm if the Hbs are in ferrous or ferric oxidation state, which would be interesting for future applications. Regarding the folding, it could also be interesting to do more advanced computer modelling on the *AsPgb*-sequences to gain a deeper understanding of which mechanisms could be of specific importance. Another alternative

could be to study expression-levels and protein expression of said phytoglobins *in vivo*, in oat directly, to elucidate whether the plant itself expresses the *AsPgbs* in a functional and properly folded state. Since the *AsPgbs* are very similar to other monocot nsHbs, which have been shown to be expressed under hypoxic conditions ([4]), a way to study this could be to apply these conditions to oats *in vivo* and studying their protein expression and contents in similar ways as was done in this study. This could be done for example using RT-qPCR or SDS-page.

It would also be interesting to study different aspects of the heme biosynthesis in relation to Hb-expression and -folding. Previous studies [19] on the subject have shown that increased Hb-expression can be achieved by upregulating key steps to ensure sufficient intracellular heme is present. If the *E. coli* cells do not readily recognize the transformed phytoglobins, and their need for heme to fold properly, an upregulation of the heme biosynthesis could kick-start the folding process. Further studies on this subject would also increase the general understanding of the link between heme biosynthesis and Hb-folding. A first step in this direction could be to either employ a system for transporting heme directly into the cell, similar to what has been done previously ([19]), or to upregulate rate-limiting steps in the biosynthesis by co-expression genes of interest. The latter possibly in the same plasmid as the Hb-genes since this has been shown to be more efficient in other cases.

5. Conclusion

To conclude this report, three novel Hbs were identified in the newly sequenced oat-genome, successfully transformed into *E. coli,* and subsequently expressed. Optimization of the Hb-expression was attempted, and resulting cultures were analyzed by SDS-PAGE. Optimization revealed OD_{600} and δ-ALA concentration as two important parameters for expression. The results from the SDS-PAGE possibly confirmed recombinant production of the oat-Hbs, but that would also imply that the expression vector was leaking. Since expression could not be confirmed with certainty, RT-qPCR was performed to ensure the genes of interest were transcribed. This was shown to be the case, which led to speculation that the phytoglobins were not properly folded in *E coli*, either due to a lack of heme or due to problems in the folding process relating to sequence, host organism or incorporation of heme. More work is required with identifying the potential problems with *AsPgb*-protein folding. Future work could focus on computer modelling of folding, sequence mutations or upregulation of heme biosynthesis. Further understanding could also be gained by studying the expression-levels of the phytoglobins *in vivo*.

Future research into the field of plant-Hb and specifically oat-Hb production could be interesting from both an economic and social point of view. There is still a need for effective and safe blood substitutes, HBOCs, but closer in time and easier to develop are plant-based food products based on recombinant plant-Hb. Plant food-based diets show numerous health benefits. Products based upon recombinant plant Hb-production have shown to be safe and attractive alternatives to conventional meat-products, in addition to being a potential source of highly bioavailable iron. Oat specifically is a crop attracting increasing global interest and succeeding in the endeavor of effectively expressing oat-Hbs recombinantly could, as such, potentially rival other similar products on the market today.

6. References

1. Berg, J.M., et al., *Biochemistry*. 8 ed. 2015, New York: W. H. Freeman and Company.
2. Marengo-Rowe, A.J. *Structure-function relations of human hemoglobins*. in *Baylor University Medical Center Proceedings*. 2006. Taylor & Francis.
3. Ahmed, M.H., M.S. Ghatge, and M.K. Safo, *Hemoglobin: structure, function and allostery.* Vertebrate and invertebrate respiratory proteins, lipoproteins and other body fluid proteins, 2020: p. 345-382.
4. Hoy, J.A. and M.S. Hargrove, *The structure and function of plant hemoglobins.* Plant Physiology and Biochemistry, 2008. **46**(3): p. 371-379.
5. Hill, R., M. Hargrove, and R. Arredondo-Peter, *Phytoglobin: a novel nomenclature for plant globins accepted by the globin community at the 2014 XVIII conference on Oxygen-Binding and Sensing Proteins.* F1000Res, 2016. **5**: p. 212.
6. Appleby, C., et al., *A role for haemoglobin in all plant roots?* Plant, Cell & Environment, 1988. **11**(5): p. 359-367.
7. Vázquez-Limón, C., et al., *The evolution of land plant hemoglobins.* Plant Science, 2012. **191**: p. 71-81.
8. Hendriks, T., et al., *A nonsymbiotic hemoglobin gene is expressed during somatic embryogenesis in Cichorium.* Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression, 1998. **1443**(1-2): p. 193-197.
9. Gupta, A.S., *Hemoglobin-based oxygen carriers: current state-of-the-art and novel molecules.* Shock (Augusta, Ga.), 2019. **52**(1): p. 70.
10. Luiking, Y.C., M.P. Engelen, and N.E. Deutz, *Regulation of nitric oxide production in health and disease.* Current opinion in clinical nutrition and metabolic care, 2010. **13**(1): p. 97.
11. Eisen, M. *HOW GMOS CAN SAVE CIVILIZATION (AND PROBABLY ALREADY HAVE)*. 2018 [cited 2022 June 22]; Available from: https://impossiblefoods.com/blog/how-gmos-can-save-civilization-and-probably-already-have.
12. Ems, T., K. St Lucia, and M.R. Huecker, *Biochemistry, iron absorption*, in *StatPearls [Internet]*. 2021, StatPearls Publishing.
13. Pawlak, R., J. Berger, and I. Hines, *Iron status of vegetarian adults: a review of literature.* American journal of lifestyle medicine, 2018. **12**(6): p. 486-498.
14. *Iron-Deficiency Anemia*. 2022 [cited 2022 June 22]; Available from: https://www.nhlbi.nih.gov/health/anemia/iron-deficiency-anemia.
15. *Oat beverages are booming. Here's why*. n.d. [cited 2022 June]; 22]. Available from: https://www.tetrapak.com/insights/cases-articles/oat-beverages-are-booming.
16. Kamal, N., et al., *The mosaic oat genome gives insights into a uniquely healthy cereal crop.* Nature, 2022. **606**(7912): p. 113-119.
17. Duff, S.M., J.B. Wittenberg, and R.D. Hill, *Expression, purification, and properties of recombinant barley (Hordeum sp.) hemoglobin: optical spectra and reactions with gaseous ligands.* Journal of Biological Chemistry, 1997. **272**(27): p. 16746-16752.
18. Natarajan, C., et al., *Expression and purification of recombinant hemoglobin in Escherichia coli.* PLoS One, 2011. **6**(5): p. e20176.
19. Villarreal, D., et al., *Enhancement of recombinant hemoglobin production in Escherichia coli BL21 (DE3) containing the Plesiomonas shigelloides heme transport system.* Applied and environmental microbiology, 2008. **74**(18): p. 5854-5856.
20. Christensen, S., *IDENTIFICATION AND EXPRESSION OF OAT HEMOGLOBIN.* LUP Student Papers, 2019.
21. Invitrogen. *Gateway® Technology with Clonase® II*. 2012 [cited 2022 January 3]; Available from: https://tools.thermofisher.com/content/sfs/manuals/gateway_clonaseii_man.pdf.

22. Fraser, R.Z., et al., *Safety evaluation of soy leghemoglobin protein preparation derived from Pichia pastoris, intended for use as a flavor catalyst in plant-based meat.* International journal of toxicology, 2018. **37**(3): p. 241-262.
23. Pranawidjaja, S., et al., *Analysis of heme biosynthetic pathways in a recombinant Escherichia coli.* Journal of Microbiology and Biotechnology, 2015. **25**(6): p. 880-886.
24. Zhang, J., et al., *Optimization of the heme biosynthesis pathway for the production of 5-aminolevulinic acid in Escherichia coli.* Scientific Reports, 2015. **5**(1): p. 1-7.
25. McNicholas, P.M., et al., *Expression of the heme biosynthetic pathway genes hemCD, hemH, hemM and hemA of Escherichia coli.* FEMS microbiology letters, 1997. **146**(1): p. 143-148.
26. Eurofins. *LightRun Tube*. n.d. [cited 2022 March]; Available from: https://eurofinsgenomics.eu/en/custom-dna-sequencing/gatc-services/lightrun-tube/.
27. Madeira, F., et al., *Search and sequence analysis tools services from EMBL-EBI in 2022.* Nucleic Acids Research, 2022.
28. Agilent. *E. coli Cell Culture Concentration from OD600 Calculator*. n.d. [cited 2022 May 23]; Available from: https://www.agilent.com/store/biocalculators/calcODBacterial.jsp.
29. ThermoFisher. *Macromolecular Components of E. coli and HeLa Cells*. n.d. [cited 2022 May 23]; Available from: https://www.thermofisher.com/se/en/home/references/ambion-tech-support/rna-tools-and-calculators/macromolecular-components-of-e.html.
30. ZymoResearch. *Quick-RNA™ Miniprep Kit*. n.d. [cited 2022 June 22]; Available from: https://files.zymoresearch.com/protocols/_r1054_r1055_quick-rna_miniprep_kit.pdf.
31. NCBI. *hemH ferrochelatase [Escherichia coli str. K-12 substr. MG1655]*. 2022 [cited 2022 June 22]; Available from: https://www.ncbi.nlm.nih.gov/gene/947532.
32. Untergasser, A., et al., *Primer3—new capabilities and interfaces.* Nucleic acids research, 2012. **40**(15): p. e115-e115.
33. Koressaar, T. and M. Remm, *Enhancements and modifications of primer design program Primer3.* Bioinformatics, 2007. **23**(10): p. 1289-1291.
34. Kõressaar, T., et al., *Primer3_masker: integrating masking of template sequence with primer design software.* Bioinformatics, 2018. **34**(11): p. 1937-1938.
35. Zhou, K., et al., *Novel reference genes for quantifying transcriptional responses of Escherichia coli to protein overexpression by quantitative PCR.* BMC molecular biology, 2011. **12**(1): p. 1-9.
36. ThermoFisher. *Tm Calculator*. n.d. [cited 2022 May 5]; Available from: https://www.thermofisher.com/se/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html.
37. Gallego, A. *How To Interpret RT-qPCR Results*. n.d. [cited 2022 June 1]; Available from: https://www.goldbio.com/articles/article/How-To-Interpret-RT-qPCR-Results#_lnxbz9.
38. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of molecular biology, 1990. **215**(3): p. 403-410.
39. Gasteiger, E., et al., *Protein identification and analysis tools on the ExPASy server.* The proteomics protocols handbook, 2005: p. 571-607.

7. Appendices

7.1 DNA Sequences

AsPgb1.1:

ATGTCTGCCGTGGAGGGAAGCAGCGCGTCCGGGGCAGCCGTCGTCTTCAGCGAGGAGCAGGAGGCGCTGGT
GCTCAAGTCATGGGCCATCATGAAGAAGGACTCCGCCAACCTTGGCCTCCGCTTCTTCCTGAAGATCTTCGAGA
TCGCGCCGTCGGCGAAGCAGATGTTCCCGTTCCTCCGCGACTCCGACGTGCCGCTCGAGACCAACCCCAAGCT
CAAGACCCACGCCGTCTCTGTCTTCGTCATGACGTGCGAGGCGGCTGCGCAGCTGCGGAAAGCCGGCAAGAT
CACCGTGAGGGAGACCACCCTGAAGAGGCTGGGTGGCACGCACGTGAAATACGGCGTCGCAGATGGCCACTT
CGAGGTGACGAGGTTCGCTCTCCTTGACACGATCAAGGGGGCGGTACCGGCTGACATGTGGGGCCCGGAGAT
GAAGACCGCGTGGGGCGAAGCCTACGACCAGCTCGTCGCGGCCATCAAGCAGGAGATGAAGCCTTCTGCTTA
G

Codon optimized *AsPgb1.1:*

ATGAGCGCGGTGGAAGGCAGCAGCGCGAGCGGCGCGGCGGTGGTGTTTAGCGAAGAACAGGAAGCGCTGG
TGCTGAAAAGCTGGGCGATTATGAAAAAAGATAGCGCGAACCTGGGCCTGCGTTTTTTTCTGAAAATTTTTGA
AATTGCGCCGAGCGCGAAACAGATGTTTCCGTTTCTGCGTGATAGCGATGTGCCGCTGGAAACCAACCCGAAA
CTGAAAACCCATGCGGTGAGCGTGTTTGTGATGACCTGCGAAGCGGCGGCGCAGCTGCGTAAAGCGGGCAAA
ATTACCGTGCGTGAAACCACCCTGAAACGTCTGGGCGGCACCCATGTGAAATATGGCGTGGCGGATGGCCATT
TTGAAGTGACCCGTTTTGCGCTGCTGGATACCATTAAAGGCGCGGTGCCGGCGGATATGTGGGGCCCGGAAA
TGAAAACCGCGTGGGGCGAAGCGTATGATCAGCTGGTGGCGGCGATTAAACAGGAAATGAAACCGAGCGCG
TAA

AsPgb1.2:

ATGTCTGCCGTGGAGGGGAACAGCGCGGCCGGGGGAGCCGTCGTCTTCAGCGAGGAGCAGGAGGCGCTGGT
GCTCAAGTCGTGGGCCATCATGAAGAAGGATTCCGCCAACCTTGGTCTCCGCTTCTTCCTGAAGATCTTCGAGA
TCGCGCCGTCGGCGAAGCAGATGTTCCCGTTCCTCCGCAACTCCGACGTGCCGCTCGAGACCAACCCCAAGCT
CAAGACTCACGCCGTTTCCGTCTTCGTCATGACGTGCGAGGCGGCTGCGCAGCTGCGGAAAGCCGGCAAGATC
ACCGTGAGGGAGACCAGCCTGAAGAGGCTGGGCGGCACGCACGTGAAATACGGCGTCGCAGACGGCCACTT
CGAGGTGACGAGGTTCGCTCTCCTTGACACGATCAAGGAGGCGGTGCCGGCTGACATGTGGGGCCCGGAGAT
GAAGAACGCGTGGGGCGAAGCCTACGACCAGCTCGTCGCGGCCATCAAGCAAGAGATGAAGCCTTCTGCTTA
G

Codon optimized *AsPgb1.2*:

ATGAGCGCGGTGGAAGGCAACAGCGCGGCGGGCGGCGCGGTGGTGTTTAGCGAAGAACAGGAAGCGCTGG
TGCTGAAAAGCTGGGCGATTATGAAAAAAGATAGCGCGAACCTGGGCCTGCGTTTTTTTCTGAAAATTTTTGA
AATTGCGCCGAGCGCGAAACAGATGTTTCCGTTTCTGCGTAACAGCGATGTGCCGCTGGAAACCAACCCGAAA
CTGAAAACCCATGCGGTGAGCGTGTTTGTGATGACCTGCGAAGCGGCGGCGCAGCTGCGTAAAGCGGGCAAA
ATTACCGTGCGTGAAACCAGCCTGAAACGTCTGGGCGGCACCCATGTGAAATATGGCGTGGCGGATGGCCAT
TTTGAAGTGACCCGTTTTGCGCTGCTGGATACCATTAAAGAAGCGGTGCCGGCGGATATGTGGGGCCCGGAA

ATGAAAAACGCGTGGGGCGAAGCGTATGATCAGCTGGTGGCGGCGATTAAACAGGAAATGAAACCGAGCGC
GTAA

AsPgb1.3:

ATGTCTGCCGTGGAGGGAAACATCGCGTCCGGCGGAGGGGGAGCCGTCGTCTTCAGCGAGGAGCAGGAGGC
GCTGGTGCTCAAGTCATGGGCCATCATGAAGAAGGATTCCGCCAACCTTGGCCTCCGTTTCTTCCTGAAGATCT
TTGAGATCGCGCCGTCGGCGAAGCAGATGTTCCCGTTCCTCCGCAACTCCGACGTACCGCTCGAAACCAACCCC
AAGCTCAAGACCCACGCCGTCTCCGTCTTCGTCATGACGTGCGAGGCGGCTGCGCAGCTGCGGAAAGCCGGC
AAGATCACCGTGAGGGAGACCACCCTGAAGAGGCTGGGTGGCACGCACCTGAAATACGGCGTCGCAGATGG
CCACTTTGAGGTGACGAGGTTCGCCCTGCTCGAGACGATCAAGGAGGCGGTACCGGCTGACATGTGGGGCCC
GGAGATGAGGAACGCGTGGGGCGAAGCCTACGACCAGCTGGTCGCGGCCATCAAGCAAGAGATGAAGCCTT
CTGCTTAG

Codon optimized *AsPgb1.3*:

ATGAGCGCGGTGGAAGGCAACATTGCGAGCGGCGGCGGCGGCGCGGTGGTGTTTAGCGAAGAACAGGAAG
CGCTGGTGCTGAAAAGCTGGGCGATTATGAAAAAAGATAGCGCGAACCTGGGCCTGCGTTTTTTTCTGAAAAT
TTTTGAAATTGCGCCGAGCGCGAAACAGATGTTTCCGTTTCTGCGTAACAGCGATGTGCCGCTGGAAACCAAC
CCGAAACTGAAAACCCATGCGGTGAGCGTGTTTGTGATGACCTGCGAAGCGGCGGCGCAGCTGCGTAAAGCG
GGCAAAATTACCGTGCGTGAAACCACCCTGAAACGTCTGGGCGGCACCCATCTGAAATATGGCGTGGCGGAT
GGCCATTTTGAAGTGACCCGTTTTGCGCTGCTGGAAACCATTAAAGAAGCGGTGCCGGCGGATATGTGGGGC
CCGGAAATGCGTAACGCGTGGGGCGAAGCGTATGATCAGCTGGTGGCGGCGATTAAACAGGAAATGAAACC
GAGCGCGTAA

BvHb1.2:

Atgagctttaccaatgtcaattatccggctagcgatggcaccgtaatttttaccgaagagcaagaagctctggtggttcaaagctggaatgtcatg
aaaaagaacagcgctgaactgggcctgaagttgttcctgaagatttttgagattgcaccgaccgcaaagaaaatgtttagctttgtacgcgatag
cgacgttccgctggagcagaatcaaaaactgaaaggccatgctatgagcgtttttgtcatgacctgtaaaagcgcagctcagctgcgcaaagctg
gcaaggttacctttggcgaaagcagcctgaaacatatgggcagcgtccacctgaaatatggcgttgtagatgaacattttgaggtgacccgctttg
cgctgctggaaaccattaaagaggcagtgccggagatgtggagcccggagatgaagaatgcttgggctgaagcttttaatcacctggtggctgct
attaaggctgaaatgcagcgcctgagcacccaaccgtaa

7.1.1 Multiple Sequence Alignment

```
CLUSTAL O(1.2.4) multiple sequence alignment


AsPgb1.3      ATGTCTGCCGTGGAGGGAAACATCGCGTCCGGCGGAGGGGGAGCCGTCGTCTTCAGCGAG  60
AsPgb1.1      ATGTCTGCCGTGGAGGGAAGCAGCGC------GTCCGGGGCAGCCGTCGTCTTCAGCGAG  54
AsPgb1.2      ATGTCTGCCGTGGAGGGGAACAGCGC------GGCCGGGGGAGCCGTCGTCTTCAGCGAG  54
              ***************** * ** ***          **** ********************

AsPgb1.3      GAGCAGGAGGCGCTGGTGCTCAAGTCATGGGCCATCATGAAGAAGGATTCCGCCAACCTT  120
AsPgb1.1      GAGCAGGAGGCGCTGGTGCTCAAGTCATGGGCCATCATGAAGAAGGACTCCGCCAACCTT  114
AsPgb1.2      GAGCAGGAGGCGCTGGTGCTCAAGTCGTGGGCCATCATGAAGAAGGATTCCGCCAACCTT  114
              ************************* ******************* ************

AsPgb1.3      GGCCTCCGTTTCTTCCTGAAGATCTTTGAGATCGCGCCGTCGGCGAAGCAGATGTTCCCG  180
AsPgb1.1      GGCCTCCGCTTCTTCCTGAAGATCTTCGAGATCGCGCCGTCGGCGAAGCAGATGTTCCCG  174
AsPgb1.2      GGTCTCCGCTTCTTCCTGAAGATCTTCGAGATCGCGCCGTCGGCGAAGCAGATGTTCCCG  174
              ** ***** ***************** ********************************

AsPgb1.3      TTCCTCCGCAACTCCGACGTACCGCTCGAAACCAACCCCAAGCTCAAGACCCACGCCGTC  240
AsPgb1.1      TTCCTCCGCGACTCCGACGTGCCGCTCGAGACCAACCCCAAGCTCAAGACCCACGCCGTC  234
AsPgb1.2      TTCCTCCGCAACTCCGACGTGCCGCTCGAGACCAACCCCAAGCTCAAGACTCACGCCGTT  234
              ********* ********** ******** ******************* ********

AsPgb1.3      TCCGTCTTCGTCATGACGTGCGAGGCGGCTGCGCAGCTGCGGAAAGCCGGCAAGATCACC  300
AsPgb1.1      TCTGTCTTCGTCATGACGTGCGAGGCGGCTGCGCAGCTGCGGAAAGCCGGCAAGATCACC  294
AsPgb1.2      TCCGTCTTCGTCATGACGTGCGAGGCGGCTGCGCAGCTGCGGAAAGCCGGCAAGATCACC  294
              ** ************************************************************

AsPgb1.3      GTGAGGGAGACCACCCTGAAGAGGCTGGGTGGCACGCACCTGAAATACGGCGTCGCAGAT  360
AsPgb1.1      GTGAGGGAGACCACCCTGAAGAGGCTGGGTGGCACGCACGTGAAATACGGCGTCGCAGAT  354
AsPgb1.2      GTGAGGGAGACCAGCCTGAAGAGGCTGGGCGGCACGCACGTGAAATACGGCGTCGCAGAC  354
              ************* *************** ********* ******************

AsPgb1.3      GGCCACTTTGAGGTGACGAGGTTCGCCCTGCTCGAGACGATCAAGGAGGCGGTACCGGCT  420
AsPgb1.1      GGCCACTTCGAGGTGACGAGGTTCGCTCTCCTTGACACGATCAAGGGGGCGGTACCGGCT  414
AsPgb1.2      GGCCACTTCGAGGTGACGAGGTTCGCTCTCCTTGACACGATCAAGGAGGCGGTGCCGGCT  414
              ******** ***************** ** ** ** ********** ****** ******

AsPgb1.3      GACATGTGGGGCCCGGAGATGAGGAACGCGTGGGGCGAAGCCTACGACCAGCTGGTCGCG  480
AsPgb1.1      GACATGTGGGGCCCGGAGATGAAGACCGCGTGGGGCGAAGCCTACGACCAGCTCGTCGCG  474
AsPgb1.2      GACATGTGGGGCCCGGAGATGAAGAACGCGTGGGGCGAAGCCTACGACCAGCTCGTCGCG  474
              ********************** ** ************************** ******

AsPgb1.3      GCCATCAAGCAAGAGATGAAGCCTTCTGCTTAG    513
AsPgb1.1      GCCATCAAGCAGGAGATGAAGCCTTCTGCTTAG    507
AsPgb1.2      GCCATCAAGCAAGAGATGAAGCCTTCTGCTTAG    507
              *********** ********************


Percent Identity  Matrix - created by Clustal2.1

     1: AsPgb1.3    100.00   95.07   95.07
     2: AsPgb1.1     95.07  100.00   96.45
     3: AsPgb1.2     95.07   96.45  100.00
```

7.2 Transformation Protocol

- Thaw the competent cells gently on ice (~10 min).

Work in sterile environment!

- Gently pipette 0.5 µl pure plasmid into the competent cells. *Handle with care.*
- Incubate 25 minutes on ice.
- Transfer the tube to a 42 °C heat block and incubate for 45 s. Put the tube back on ice gently.
- Add 900 µl sterile LB (no antibiotic) and incubate while shaking for 60 minutes at 37 °C.
- Pellet the cells by centrifugation and remove all but ~100 µl LB. Resuspend the cells and spread the cell suspension onto agar plates containing appropriate antibiotic (ampicillin).
- Use two plates, one with 20 µl and 50 µl.
- Let the plates dry before turning them upside down (~10-20 min).
- Incubate upside down overnight at 37 °C.
- Don´t wait too long to take out the plates the day after. Don´t let the colonies become too large.

Note: The transformation protocol was provided by supervisors at the Division of Pure & Applied Biochemistry.

7.3 Sequencing Results

The sequencing results were obtained as raw data from the Sanger sequencing, as seen below in **Figure A1**:

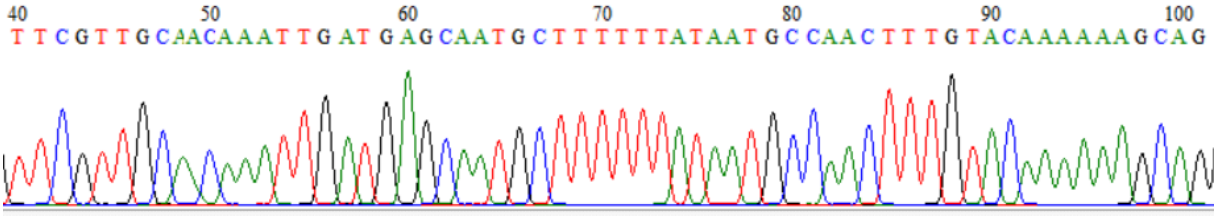

Figure A1. Example of raw data output from Sanger sequencing of plasmids.

and the resulting sequence could be used for sequence alignment with the target sequence.

7.3.1 Entry Vector following BP-reaction

The quantity and quality of plasmid DNA following the BP-reaction can be seen in **Table A1** below. The purifications were done in duplicates for each sequence and one measurement was performed per purification.

Table A1. Concentration and absorbance of purified plasmid material following the BP-reaction. The samples marked with an asterisk (*) were sent for sequencing.

Sample	Concentration (ng/µl)	A_{260}/A_{280}	A_{260}/A_{230}
AsPgb1.1 (1)	61.1	1.783	2.016
AsPgb1.1 (2) *	110	1.811	2.255
AsPgb1.2 (1)	97.4	1.798	2.178
AsPgb1.2 (2) *	98.4	1.817	2.200
AsPgb1.3 (1) *	134	1.818	2.261
AsPgb1.3 (2)	120	1.799	2.231

An example of the analysis done on the sequencing results following the BP-reaction can be seen below in **Figure A2**. The figure shows a Clustal Omega sequence alignment between the sequenced plasmid and the codon optimized *AsPgb*-template sequence.

```
BP_seq_AsPgb1.1_forward     ATGCTTTTTTATAATGCCAACTTTGTACAAAAAAGCAGGCTATGAGCGCGGTGGAAGGCA
120
Codon_opt_AsPgb1.1          ----------------------------------------ATGAGCGCGGTGGAAGGCA
                                                                     ******************

BP_seq_AsPgb1.1_forward     GCAGCGCGAGCGGCGCGGCGGTGGTGTTTAGCGAAGAACAGGAAGCGCTGGTGCTGAAAA
180
Codon_opt_AsPgb1.1          GCAGCGCGAGCGGCGCGGCGGTGGTGTTTAGCGAAGAACAGGAAGCGCTGGTGCTGAAAA
                            ************************************************************

BP_seq_AsPgb1.1_forward     GCTGGGCGATTATGAAAAAAGATAGCGCGAACCTGGGCCTGCGTTTTTTTCTGAAAATTT
240
Codon_opt_AsPgb1.1          GCTGGGCGATTATGAAAAAAGATAGCGCGAACCTGGGCCTGCGTTTTTTTCTGAAAATTT
139
                            ************************************************************

BP_seq_AsPgb1.1_forward     TTGAAATTGCGCCGAGCGCGAAACAGATGTTTCCGTTTCTGCGTGATAGCGATGTGCCGC
300
Codon_opt_AsPgb1.1          TTGAAATTGCGCCGAGCGCGAAACAGATGTTTCCGTTTCTGCGTGATAGCGATGTGCCGC
199
                            ************************************************************

BP_seq_AsPgb1.1_forward     TGGAAACCAACCCGAAACTGAAAACCCATGCGGTGAGCGTGTTTGTGATGACCTGCGAAG
360
Codon_opt_AsPgb1.1          TGGAAACCAACCCGAAACTGAAAACCCATGCGGTGAGCGTGTTTGTGATGACCTGCGAAG
259
                            ************************************************************

BP_seq_AsPgb1.1_forward     CGGCGGCGCAGCTGCGTAAAGCGGGCAAAATTACCGTGCGTGAAACCACCCTGAAACGTC
420
Codon_opt_AsPgb1.1          CGGCGGCGCAGCTGCGTAAAGCGGGCAAAATTACCGTGCGTGAAACCACCCTGAAACGTC
319
                            ************************************************************

BP_seq_AsPgb1.1_forward     TGGGCGGCACCCATGTGAAATATGGCGTGGCGGATGGCCATTTTGAAGTGACCCGTTTTG
480
Codon_opt_AsPgb1.1          TGGGCGGCACCCATGTGAAATATGGCGTGGCGGATGGCCATTTTGAAGTGACCCGTTTTG
379
                            ************************************************************

BP_seq_AsPgb1.1_forward     CGCTGCTGGATACCATTAAAGGCGCGGTGCCGGCGGATATGTGGGGCCCGGAAATGAAAA
540
Codon_opt_AsPgb1.1          CGCTGCTGGATACCATTAAAGGCGCGGTGCCGGCGGATATGTGGGGCCCGGAAATGAAAA
439
                            ************************************************************

BP_seq_AsPgb1.1_forward     CCGCGTGGGGCGAAGCGTATGATCAGCTGGTGGCGGCGATTAAACAGGAAATGAAACCGA
600
Codon_opt_AsPgb1.1          CCGCGTGGGGCGAAGCGTATGATCAGCTGGTGGCGGCGATTAAACAGGAAATGAAACCGA
499
                            ************************************************************

BP_seq_AsPgb1.1_forward     GCGCGTAAACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCATTGCTTATCAAT
660
Codon_opt_AsPgb1.1          GCGCGTAA--------------------------------------------------
507
                            ********
```

Figure A2. Clustal Omega sequence alignment between the entry vector (BP_seq_AsPgb1.1_forward) sequenced with the addition of the forward primer and the codon optimized *AsPgb1.1*.

7.3.2 Expression Vector following LR-reaction

The quantity and quality of plasmid DNA following the LR-reaction can be seen in **Table A2** below. The purifications were done in duplicates for each sequence and one measurement was performed per purification.

Table A2. Concentration and absorbance of purified plasmid material following the LR-reaction. The samples marked with an asterisk (*) were sent for sequencing.

Sample	Concentration (ng/µl)	A_{260}/A_{280}	A_{260}/A_{230}
AsPgb1.1 (1)	39.3	1.837	2.548
AsPgb1.1 (2)	40.8 *	1.822	2.278
AsPgb1.2 (1)	37.3	1.829	2.278
AsPgb1.2 (2)	41.7 *	1.867	2.154
AsPgb1.3 (1)	55.7 *	1.806	2.154
AsPgb1.3 (2)	38.3	1.878	2.333

An example of the analysis done on the sequencing results following the LR-reaction can be seen below in **Figure A3**. The figure shows a Clustal Omega sequence alignment between the sequenced plasmid and the codon optimized *AsPgb*-template sequence.

```
LR_seq_AsPgb1.1_forward      TTGTACAAAAAAGCAGGCTATGAGCGCGGTGGAAGGCAGCAGCGCGAGCGGCGCGGCGGT
120
Codon_opt_AsPgb1.1           ----------ATGAGCGCGGTGGAAGGCAGCAGCGCGAGCGGCGCGGCGGT
                                        ****************************************

LR_seq_AsPgb1.1_forward      GGTGTTTAGCGAAGAACAGGAAGCGCTGGTGCTGAAAAGCTGGGCGATTATGAAAAAAGA
180
Codon_opt_AsPgb1.1           GGTGTTTAGCGAAGAACAGGAAGCGCTGGTGCTGAAAAGCTGGGCGATTATGAAAAAAGA
101
                             ************************************************************

LR_seq_AsPgb1.1_forward      TAGCGCGAACCTGGGCCTGCGTTTTTTTCTGAAAATTTTTGAAATTGCGCCGAGCGCGAA
240
Codon_opt_AsPgb1.1           TAGCGCGAACCTGGGCCTGCGTTTTTTTCTGAAAATTTTTGAAATTGCGCCGAGCGCGAA
161
                             ************************************************************

LR_seq_AsPgb1.1_forward      ACAGATGTTTCCGTTTCTGCGTGATAGCGATGTGCCGCTGGAAACCAACCCGAAACTGAA
300
Codon_opt_AsPgb1.1           ACAGATGTTTCCGTTTCTGCGTGATAGCGATGTGCCGCTGGAAACCAACCCGAAACTGAA
221
                             ************************************************************

LR_seq_AsPgb1.1_forward      AACCCATGCGGTGAGCGTGTTTGTGATGACCTGCGAAGCGGCGGCGCAGCTGCGTAAAGC
360
Codon_opt_AsPgb1.1           AACCCATGCGGTGAGCGTGTTTGTGATGACCTGCGAAGCGGCGGCGCAGCTGCGTAAAGC
281
                             ************************************************************

LR_seq_AsPgb1.1_forward      GGGCAAAATTACCGTGCGTGAAACCACCCTGAAACGTCTGGGCGGCACCCATGTGAAATA
420
Codon_opt_AsPgb1.1           GGGCAAAATTACCGTGCGTGAAACCACCCTGAAACGTCTGGGCGGCACCCATGTGAAATA
341
                             ************************************************************

LR_seq_AsPgb1.1_forward      TGGCGTGGCGGATGGCCATTTTGAAGTGACCCGTTTTGCGCTGCTGGATACCATTAAAGG
480
Codon_opt_AsPgb1.1           TGGCGTGGCGGATGGCCATTTTGAAGTGACCCGTTTTGCGCTGCTGGATACCATTAAAGG
401
                             ************************************************************

LR_seq_AsPgb1.1_forward      CGCGGTGCCGGCGGATATGTGGGGCCCGGAAATGAAAACCGCGTGGGGCGAAGCGTATGA
540
Codon_opt_AsPgb1.1           CGCGGTGCCGGCGGATATGTGGGGCCCGGAAATGAAAACCGCGTGGGGCGAAGCGTATGA
461
                             ************************************************************

LR_seq_AsPgb1.1_forward      TCAGCTGGTGGCGGCGATTAAACAGGAAATGAAACCGAGCGCGTAAACCCAGCTTTCTTG
600
Codon_opt_AsPgb1.1           TCAGCTGGTGGCGGCGATTAAACAGGAAATGAAACCGAGCGCGTAA----------
507
                             **********************************************
```

Figure A3. Clustal Omega sequence alignment between the expression vector (LR_seq_AsPgb1.1_forward) sequenced with the addition of the forward primer and the codon optimized *AsPgb1.1*.

7.4 Expression Protocol

In **Table A3** below, the scheme for the different round of fermentations, their settings and the resulting pellet weights are given. Settings marked in blue are analogous to the "reference"-values. An "x" means that the value was not taken or could not be taken.

Table A3. The parameters of the different fermentation-rounds 1 to 9. Blue indicates the parameter is equivalent to the reference protocol.

Round	Gene (Sample nr)	IPTG (mM)	δ-ALA (mM)	OD_{600}	CO (s)	Temp. (°C)	Time (h)	Shaking (rpm)	Weight (g)
1									
	AsPgb1.1	0.5	0.3	3.515	10	22	18	150	2.055
	AsPgb1.2	0.5	0.3	3.515	10	22	18	150	2.233
	AsPgb1.3	0.5	0.3	3.515	10	22	18	150	0.718
	AsPgb1.1 (-)	0	0	3.515	0	22	18	150	0.9923
	AsPgb1.2 (-)	0	0	3.515	0	22	18	150	1.29
	AsPgb1.3 (-)	0	0	3.515	0	22	18	150	1.198
2									
	AsPgb1.3 (1)	0.2	0.1	3.15	10	22	18	150	1.501
	AsPgb1.3 (2)	0.5	0.1	3.15	10	22	18	150	1.196
	AsPgb1.3 (3)	0.8	0.1	3.15	10	22	18	150	1.074
	AsPgb1.3 (4)	0.2	0.3	3.15	10	22	18	150	0.852
	AsPgb1.3 (5)	0.8	0.3	3.15	10	22	18	150	0.746
	AsPgb1.3 (6)	0.2	0.5	3.15	10	22	18	150	0.931
	AsPgb1.3 (7)	0.5	0.5	3.15	10	22	18	150	1.123
	AsPgb1.3 (8)	0.8	0.5	3.15	10	22	18	150	1.101
	AsPgb1.3 (9)	0.5	1	2.93	10	22	18	150	0.98
3									
	AsPgb1.3 (1)	0.5	2	3.14	10	22	18	150	0.891
	AsPgb1.3 (2)	0.5	3	3.14	10	22	18	150	0.788
	AsPgb1.3 (3)	0.5	4	3.14	10	22	18	150	0.825
	AsPgb1.3 (4)	0.5	5	3.14	10	22	18	150	0.766
4									
	AsPgb1.3 (1)	0.5	1	0.104	10	22	18	150	0
	AsPgb1.3 (2)	0.5	1	0.385	10	22	18	150	0.136
	AsPgb1.3 (3)	0.5	1	Not taken	10	22	18	150	1.77
	AsPgb1.3 (4)	0.5	1	0.339	10	22	18	150	0.24
	AsPgb1.3 (5)	0.5	1	0.492	10	22	18	150	0.382
	AsPgb1.3 (6)	0.5	1	1.376	10	22	18	150	0.502
	AsPgb1.3 (7)	0.5	1	2.19	10	22	18	150	0.646
	AsPgb1.3 (8)	0.5	1	3.9	10	22	18	150	0.699
	AsPgb1.3 (9)	0.5	1	4.99	10	22	18	150	1.211
	AsPgb1.3 (10)	0.5	1	8.955	10	22	18	150	1.456

5								
	AsPgb1.3 (1)	0.5	1	4.835	10	12	18	150	0.915
	AsPgb1.3 (2)	0.5	1	4.835	10	26	18	150	1.291
	AsPgb1.3 (3)	0.5	1	4.835	10	30	18	150	1.051
6								
	AsPgb1.3 (1)	0.5	1	5.6	10	22	12	150	1.051
	AsPgb1.3 (2)	0.5	1	5.6	10	22	24	150	1.283
7								
	AsPgb1.3 (1)	0.5	1	4.945	10	18	18	150	0.791
	AsPgb1.3 (2)	0.5	1	4.945	10	22	16	150	1.105
	AsPgb1.3 (3)	0.5	1	4.945	10	22	18	150	1.0052
	AsPgb1.3 (4)	0.5	1	4.945	10	24	18	150	0.89
8								
	AsPgb1.1	0.5	1	3.985	10	22	18	150	0.628
	AsPgb1.2	0.5	1	4.745	10	22	18	150	0.941
	AsPgb1.3	0.5	1	5.155	10	22	18	150	0.8811
9								
	AsPgb1.1	0.5	1	5.065	10	22	18	150	x
	AsPgb1.2	0.5	1	5.0275	10	22	18	150	x
	AsPgb1.3	0.5	1	4.2225	10	22	18	150	x
	BvHb1.2	0.5	0.3	2.725	10	22	18	150	x

7.5 Cell Pellets

Pictures of harvested cells following fermentations that are not included in the main report.

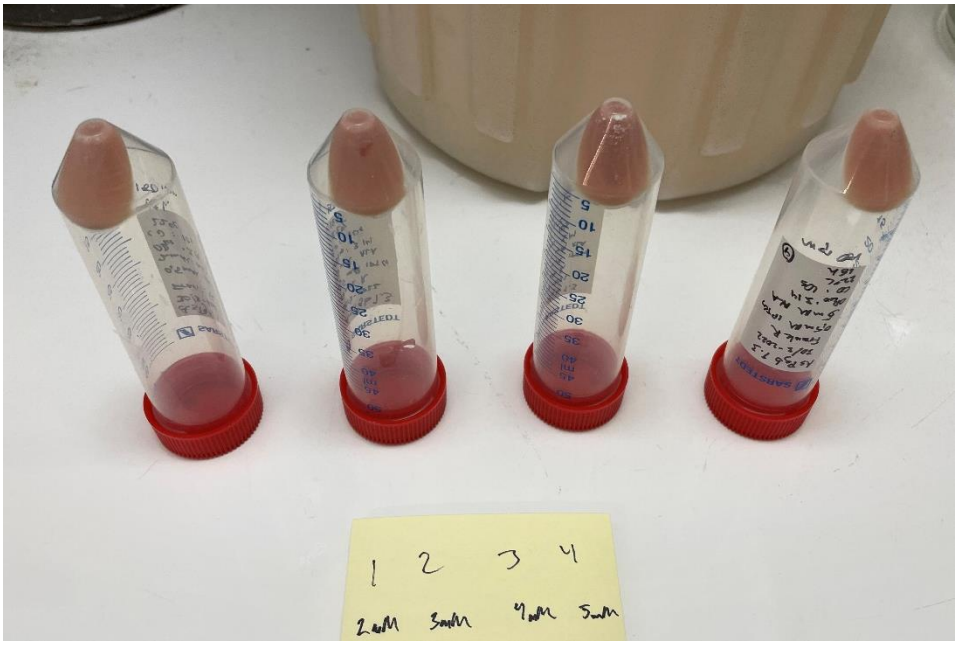

Figure A4. Cell pellets from the third round of fermentations. The numbers correspond to the number of each pellet as given in **Table A3** in **Appendix 7.4 Expression Protocol**. From left to right: increasing δ-ALA concentration (see picture). IPTG concentration in this case was 0.5 mM for all cultures.

Figure A5. Cell pellets from the fifth round of fermentations. The temperatures of fermentation for each pellet is given below it, these correspond to, from left to right, samples 1 , 2 and 3 in **Table A3** in **Appendix 7.4 Expression Protocol**.

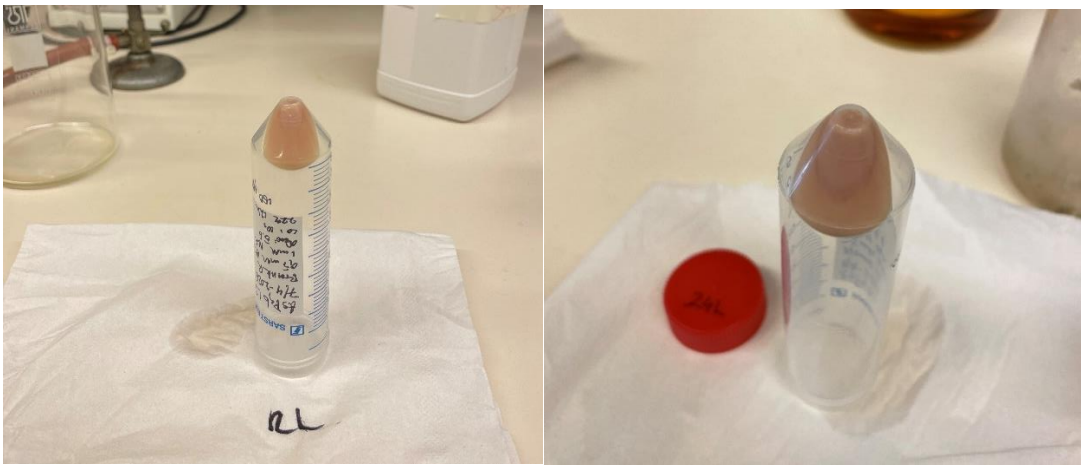

Figure A6. Cell pellets from the sixth round of fermentations. The times of fermentation for each pellet is given in the picture, these correspond to, from left to right, samples 1 and 2 in **Table A3** in **Appendix 7.4 Expression Protocol**.

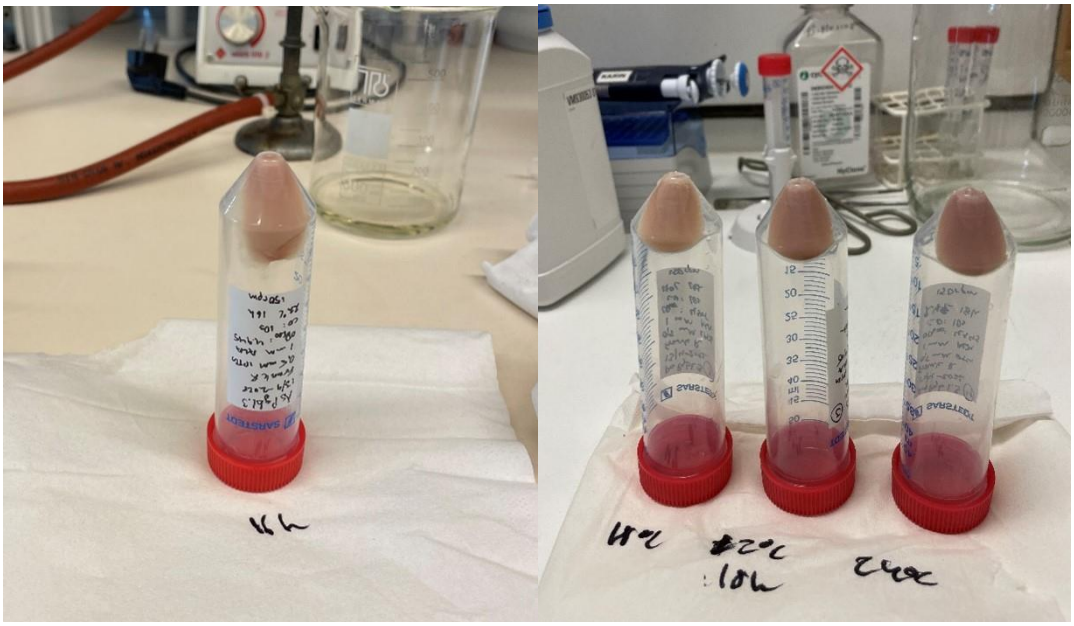

Figure A7. Cell pellets from the seventh round of fermentations. Left picture: sample 2 (16h, 22°C), right picture: samples 1 (18h, 18°C), 3 (18h, 22°C), and 4 (18h, 24°C) from left to right. The sample numbers correspond to thos in **Table A3** in **Appendix 7.4 Expression Protocol**.

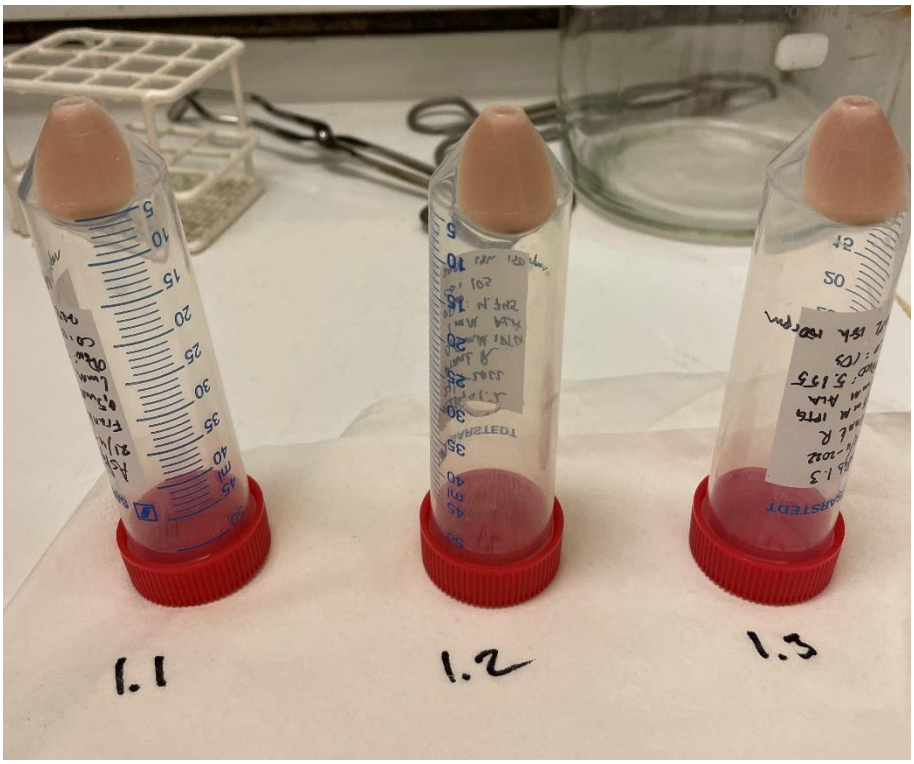

Figure A8. Cell pellets from the eight round of fermentations. All three transformed cells *AsPgb1.1-1.3* with the optimized protocol. See **Table A3** in **Appendix 7.4 Expression Protocol** for the specfications.

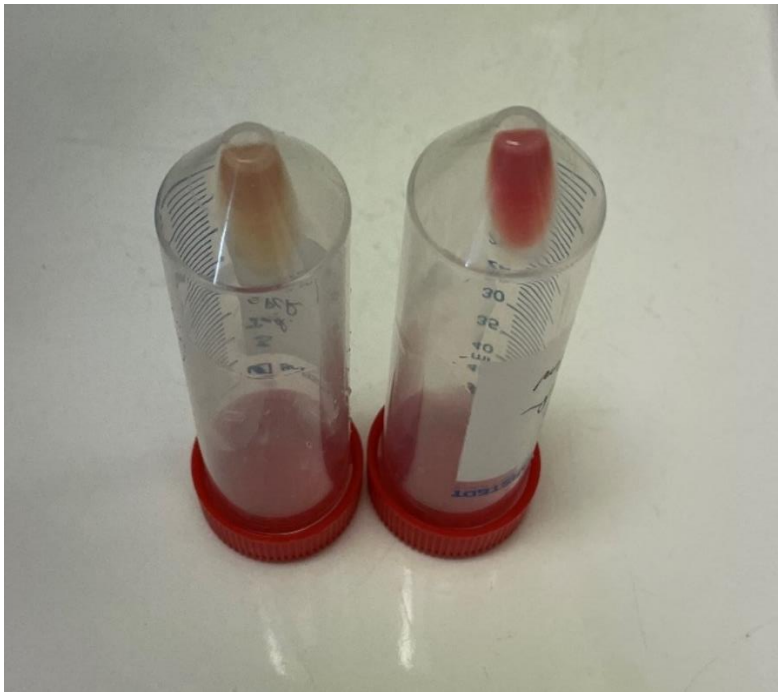

Figure A9. Cell pellets from the ninth round of fermentations. A small, unwashed, cell pellet from *AsPgb1.3* (left) and from *BvHb1.2* (right), fermented with the optimized protocol. See **Table A3** in **Appendix 7.4 Expression Protocol** for the specfications.

7.6 SDS-PAGE Gels

Pictures of produced SDS-PAGE gels that are not included in the main report.

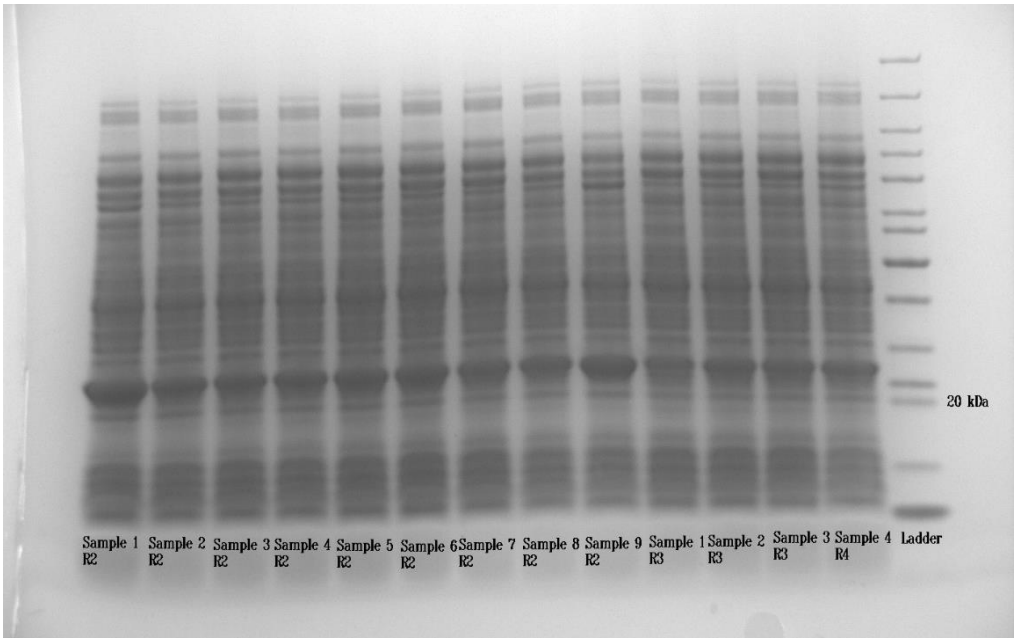

Figure A10. Gel from SDS-page analysis of samples from round 2 (R1), and round 3 (R3) of fermentations. Position as indicated in the figure. Sample numbers match those in **Table A3** in **Appendix 7.4 Expression Protocol.** The 20 kDa reference is marked.

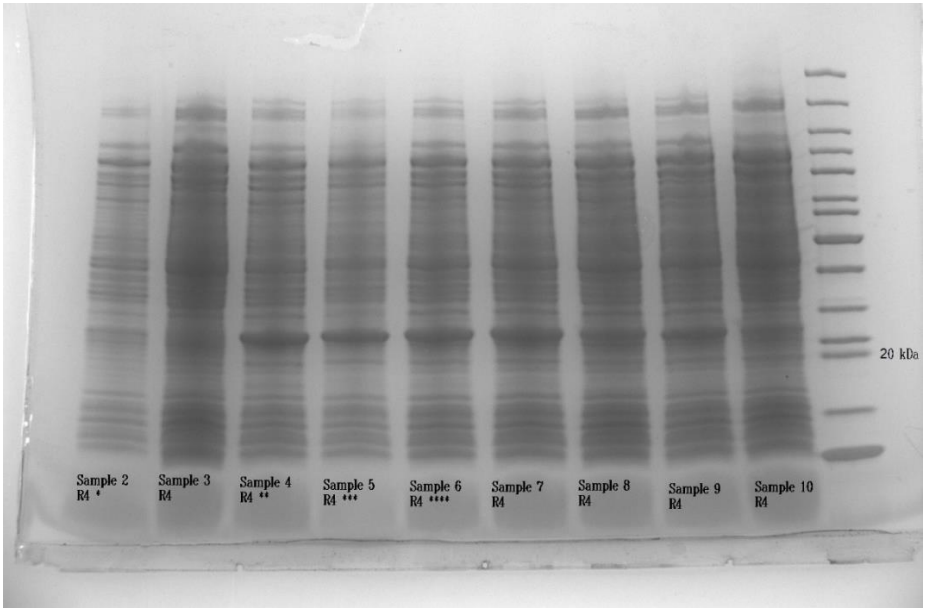

Figure A11. Gel from SDS-page analysis of samples from round 4 (R4) of fermentations. Position as indicated in the figure. Sample numbers match those in **Table A3** in **Appendix 7.4 Expression Protocol.** The 20 kDa reference is marked. Note: sample 1 was not run on SDS-PAGE due to its low cell count. Cell concentration was *0.017 g/ml, **0.03 g/ml, ***0.04775 g/ml, and ****0.06275 g/ml.

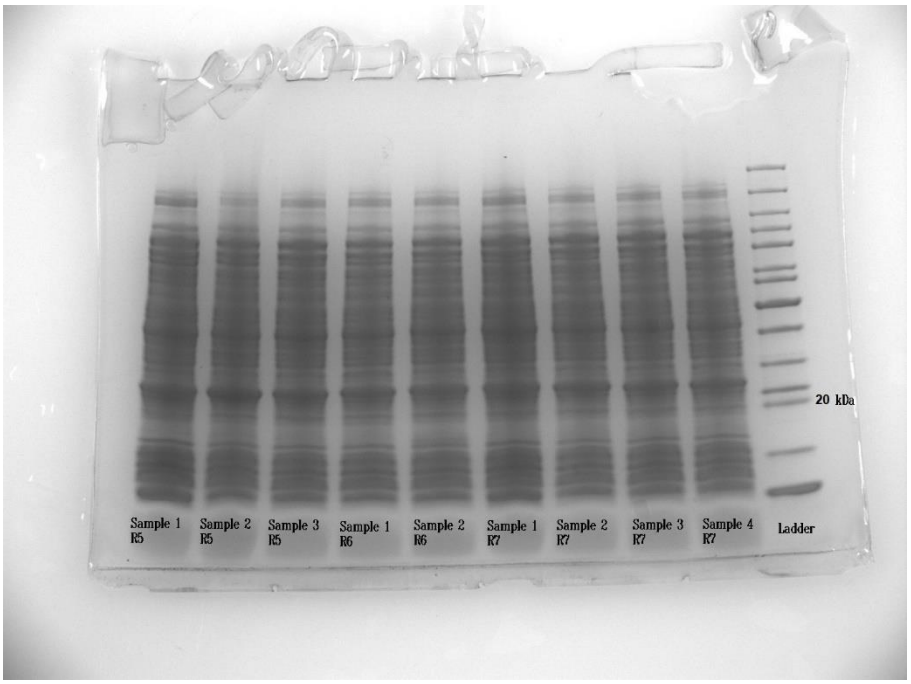

Figure A12. Gel from SDS-page analysis of samples from round 5 (R5), round 6 (R6), and round 7 (R7) of fermentations. Position as indicated in the figure. Sample numbers match those in **Table A3** in **Appendix 7.4 Expression Protocol.** The 20 kDa reference is marked.

7.7 RNA – Extraction

Table A4 below shows the results from the absorbance measurements of extracted RNA and **Figure A13** the RNA-extracts run on agarose gel.

Table A4. Results from the RNA – extraction. The measurements were done in duplicates, and the average of the measurements are given. Two rounds were done with duplicate samples from the same fermentation. If the sample was re-concentrated this is given. Only the final values following re-concentration are included. The sample proceeding to cDNA synthesis are given.

Round 1	**Sample**	**Av Conc. (ng/µl)**	**Av A_{260}/A_{280}**	**Av A_{260}/A_{230}**	**Concentrated?**	**Chosen for cDNA synthesis?**
	AsPgb1.1 (-)	35	1.978	1.978	No	**Yes**
	AsPgb1.2 (-)	27	2.0325	1.707	**Yes**	**Yes**
	AsPgb1.3 (-)	13.9	2.0015	1.893	**Yes**	No
	BvHb1.2 (-)	44.7	2.009	2.143	No	**Yes**
	AsPgb1.1	16.3	1.907	1.2085	**Yes**	No
	AsPgb1.2	24.85	1.9545	1.6065	No	**Yes**
	AsPgb1.3	16.9	1.9785	1.5745	**Yes**	No
	BvHb1.2	23.5	1.934	1.3755	**Yes**	**Yes**
Round 2						
	AsPgb1.1 (-)	28.85	2.0445	2.14	No	No
	AsPgb1.2 (-)	30.2	2.027	0.418	No	No
	AsPgb1.3 (-)	62	2.053	0.6815	No	**Yes**
	BvHb1.2 (-)	120	2.061	0.5715	No	No
	AsPgb1.1	24.45	2.0155	1.225	No	**Yes**
	AsPgb1.2	41.55	2.0295	0.46	No	No
	AsPgb1.3	33.8	2.0245	0.124	No	**Yes**
	BvHb1.2	42.5	2.0775	0.9425	No	No

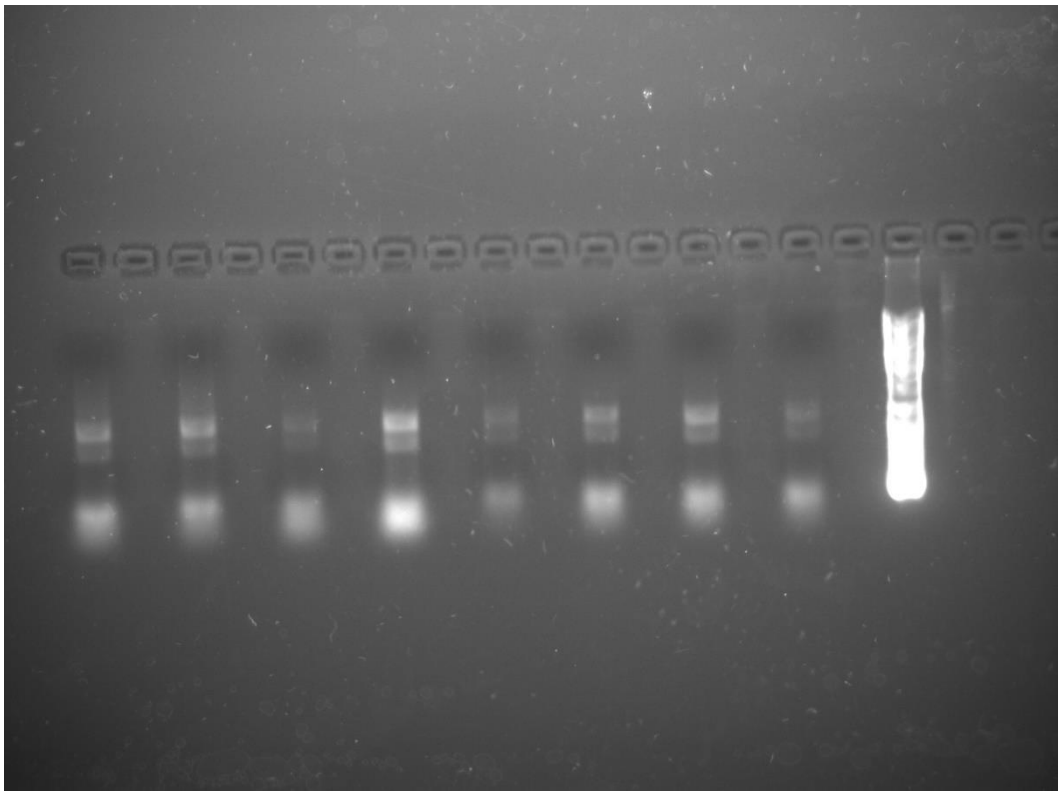

Figure A13. Agarose gel showing the RNA – extracts from the first extraction. Ladder to the right.

7.8 Primer Sequences

Below in **Table A5** the primer sets used in the qPCR are listed.

Table A5. Primer sets used in the RT – qPCR with melting temperatures (T_m). Orientation is either forward (F) or reverse (R). The primer sequences for *hcaT, rrsA* and *idnT* were taken from a study [35].

Gene & Orientation	Primer Sequence	T_m (°C)
hcaT_F	GCTGCTCGGCTTTCTCATCC	60.4
hcaT_R	CCAACCACGCTGACCAACC	61.8
idnT_F	CTGTTTAGCGAAGAGGAGATGC	60
idnT_R	ACAAACGGCGGCGATAGC	60.1
rrsA_F	CTCTTGCCATCGGATGTGCCCA	66.6
rrsA_R	CCAGTGTGGCTGGTCATCCTCTCA	67.4
hemH1_F	TCTCCTGTTCTACGGTCGGT	59.96
hemH1_R	TTGGCAAAAGAAGCGCGTAC	60.04
hemH2_F	GTCTGGATGGAAGGTGGCTC	60.11
hemH2_R	ACCGACCGTAGAACAGGAGA	59.96
1.1_F	GCGTGAAACCACCCTGAAAC	59.97
1.1_R	CCACCAGCTGATCATACGCT	59.89
1.2_F	CGTGTTTGTGATGACCTGCG	60.11
1.2_R	ATTTCCGGGCCCCACATATC	59.89
2.1_F	CGTGAAACCAGCCTGAAACG	60.04
2.1_R	CGGCACCGCTTCTTTAATGG	59.9
2.2_F	CGTGTTTGTGATGACCTGCG	60.11
2.2_R	CCACCAGCTGATCATACGCT	59.89
3.1_F	GCGTGAAACCACCCTGAAAC	59.97
3.1_R	CGGCACCGCTTCTTTAATGG	59.9
3.2_F	CGTGTTTGTGATGACCTGCG	60.11
3.2_R	CCACCAGCTGATCATACGCT	59.89
Bv1_F	gcaccgaccgcaaagaaaat	60.04
Bv1_R	tggacgctgcccatatgttt	60.03
Bv2_F	gggcctgaagttgttcctga	59.89
Bv2_R	atgtttcaggctgctttcgc	59.76

7.9 qPCR-plates

The positioning of samples on the qPCR-plates run are shown below, with **Table A6** for the first plates and **Table A7** for the second plate.

Table A6. First qPCR plate used for melt curve analysis of the primers and for the culture with *AsPgb1.1*. 1, 2 and 3 indicate sequences *AsPgb1.1-1.3*. "Bv" indicates *BvHb1.2*. "Pool" indicates pooled DNA from all the induced samples. (+) indicates an induced sample, and (-) a non-induced sample. "No templ" indicates that no cDNA was added. The primers added are given below the sample, and match the names in **Table A5** above, e. g. P1.1 is primer par 1.1_F and 1.1_R.

	1	2	3	4	5	6	7	8	9	10	11	12
A	1 (+) P1.1	1 (+) P1.1	1 (+) P1.1	No templ P1.1	No templ P1.1	No templ P1.1	Pool hemH1	Pool hemH1	Pool hemH1	1 (-) P1.1	1 (-) P1.1	1 (-) P1.1
B	1 (+) P1.2	1 (+) P1.2	1 (+) P1.2	No templ P1.2	No templ P1.2	No templ P1.2	No templ hemH1	No templ hemH1	No templ hemH1	1 (-) hemH1	1 (-) hemH1	1 (-) hemH1
C	2 (+) P2.1	2 (+) P2.1	2 (+) P2.1	No templ P2.1	No templ P2.1	No templ P2.1	No templ hemH2	No templ hemH2	No templ hemH2	1 (+) hemH1	1 (+) hemH1	1 (+) hemH1
D	2 (+) P2.2	2 (+) P2.2	2 (+) P2.2	No templ P2.2	No templ P2.2	No templ P2.2	Pool hemH2	Pool hemH2	Pool hemH2			
E	3 (+) P3.1	3 (+) P3.1	3 (+) P3.1	No templ P3.1	No templ P3.1	No templ P3.1	Pool rrsA	Pool rrsA	Pool rrsA	1 (-) rrsA	1 (-) rrsA	1 (-) rrsA
F	3 (+) P3.2	3 (+) P3.2	3 (+) P3.2	No templ P3.2	No templ P3.2	No templ P3.2	No templ rrsA	No templ rrsA	No templ rrsA	1 (+) rrsA	1 (+) rrsA	1 (+) rrsA
G	Bv (+) Bv1	Bv (+) Bv1	Bv (+) Bv1	No templ Bv1	No templ Bv1	No templ Bv1	Pool idnT	Pool idnT	Pool idnT	1 (-) idnT	1 (-) idnT	1 (-) idnT
H	Bv (+) Bv2	Bv (+) Bv2	Bv (+) Bv2	No templ Bv2	No templ Bv2	No templ Bv2	No templ idnT	No templ idnT	No templ idnT	1 (+) idnT	1 (+) idnT	1 (+) idnT

Table A7. Second qPCR plate used for cultures with *AsPgb1.2, AsPgb1.3,* and *BvHb1.2*. 1, 2 and 3 indicate sequences *AsPgb1.1-1.3*. "Bv" indicates *BvHb1.2*. "Pool" indicates pooled DNA from all the induced samples. (+) indicates an induced sample, and (-) a non-induced sample. "No templ" indicates that no cDNA was added. The primers added are given below the sample, and match the names in **Table A5** above, e. g. P2.1 is primer pair 2.1_F and 2.1_R.

	1	2	3	4	5	6	7	8	9	10	11	12
A	No templ P2.1	No templ P2.1	No templ P2.1	No templ Bv2	No templ Bv2	No templ Bv2						
B	No templ P3.1	No templ P3.1	No templ P3.1	No templ hemH1	No templ hemH1	No templ hemH1	No templ rrsA	No templ rrsA	No templ rrsA	No templ idnt	No templ idnt	No templ idnt
C	2 (-) P2.1	2 (-) P2.1	2 (-) P2.1	2 (-) hemH1	2 (-) hemH1	2 (-) hemH1	2 (-) rrsA	2 (-) rrsA	2 (-) rrsA	2 (-) idnt	2 (-) idnt	2 (-) idnt
D	2 (+) P2.1	2 (+) P2.1	2 (+) P2.1	2 (+) hemH1	2 (+) hemH1	2 (+) hemH1	2 (+) rrsA	2 (+) rrsA	2 (+) rrsA	2 (+) idnt	2 (+) idnt	2 (+) idnt
E	3 (-) P3.1	3 (-) P3.1	3 (-) P3.1	3 (-) hemH1	3 (-) hemH1	3 (-) hemH1	3 (-) rrsA	3 (-) rrsA	3 (-) rrsA	3 (-) idnt	3 (-) idnt	3 (-) idnt
F	3 (+) P3.1	3 (+) P3.1	3 (+) P3.1	3 (+) hemH1	3 (+) hemH1	3 (+) hemH1	3 (+) rrsA	3 (+) rrsA	3 (+) rrsA	3 (+) idnt	3 (+) idnt	3 (+) idnt
G	Bv (-) Bv2	Bv (-) Bv2	Bv (-) Bv2	Bv (-) hemH1	Bv (-) hemH1	Bv (-) hemH1	Bv (-) rrsA	Bv (-) rrsA	Bv (-) rrsA	Bv (-) idnt	Bv (-) idnt	Bv (-) idnt
H	Bv (+) Bv2	Bv (+) Bv2	Bv (+) Bv2	Bv (+) hemH1	Bv (+) hemH1	Bv (+) hemH1	Bv (+) rrsA	Bv (+) rrsA	Bv (+) rrsA	Bv (+) idnt	Bv (+) idnt	Bv (+) idnt

7.10 Melt Curve Results

The results from the melt curve analysis are given below.

Table A8. Melt curve plate view results from the first qPCR-plate. Samples in the wells match those in **Table A6.**

		1	2	3	4	5	6	7	8	9	10	11	12
A	Peak 1	84.00	84.00	None	None	None	None	None	None	None	84.50	84.50	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
B	Peak 1	85.00	85.00	None	None	None	None	None	None	None	None	None	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
C	Peak 1	82.00	82.00	82.00	None	None	None	None	None	None	None	None	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
D	Peak 1	84.50	84.50	84.50	None	None	None	None	None	None	None	None	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
E	Peak 1	82.00	82.00	81.50	81.50	81.50	None	82.00	82.00	None	None	None	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
F	Peak 1	None	None	None	None	None	None	None	None	None	82.00	82.00	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
G	Peak 1	82.00	82.00	None	82.00	None	None	80.50	80.50	80.50	80.00	None	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
H	Peak 1	81.50	81.50	81.50	None	None	None	None	None	None	80.50	80.50	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None

Table A9. Melt curve plate view results from the second qPCR-plate. Samples in the wells match those in **Table A7.**

		1	2	3	4	5	6	7	8	9	10	11	12
A	Peak 1	None	None	None	None	None	None	None	None	None	None	None	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
B	Peak 1	None	None	None	None	None	None	None	None	None	None	None	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
C	Peak 1	81.50	81.50	81.50	None	None	None	82.00	82.00	None	80.00	80.00	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
D	Peak 1	82.00	82.00	82.00	None	None	None	82.00	82.00	82.00	80.50	80.50	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
E	Peak 1	82.00	82.00	82.00	None	None	None	82.00	82.00	None	80.50	80.50	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
F	Peak 1	82.00	82.00	None	None	None	None	82.00	82.00	82.00	80.50	80.50	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
G	Peak 1	81.50	81.50	81.50	None	None	None	82.00	82.00	None	80.50	80.50	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None
H	Peak 1	81.50	81.50	None	None	None	None	82.00	82.00	82.00	80.50	80.50	None
	Peak 2	None	None	None	None	None	None	None	None	None	None	None	None

7.11 Data – qPCR

The data from the qPCR is presented, with raw data given and calculations done to determine the relative expression of hemoglobin and heme synthesis genes.

7.11.1 Raw Data

Table A10. Raw data from first qPCR - plate. If blank, no C_q-value was returned. Samples in the wells match those in **Table A6.**

		1	2	3	4	5	6	7	8	9	10	11	12
A	Cq	19.37	19.29	21.31	36.61	37.17		25.99	27.32		25.53	25.45	27.51
B	Cq	20.33	20.32	22.95				38.58	37.35	39.20	29.54	29.49	33.42
C	Cq	17.66	17.44	18.72				38.95	36.93		26.16	26.37	30.17
D	Cq	18.49	18.48	19.36				32.04	32.37	32.62			
E	Cq	17.68	17.68	18.85	32.43	32.71	34.30	13.30	13.37	15.26			
F	Cq										13.80	13.98	16.93
G	Cq	20.16	20.16	23.27	30.11			24.28	24.30	25.16	30.18	31.45	35.92
H	Cq	19.96	20.07	20.70				39.70	38.17		25.44	25.66	26.62

Table A11. Raw data from second qPCR-plate. If blank, no C_q-value was returned. Samples in the wells match those in **figure A7.**

		1	2	3	4	5	6	7	8	9	10	11	12
A	Cq	38.02					39.04						
B	Cq	36.42	36.72	36.63	37.13	36.68		36.64	35.33		37.12		
C	Cq	23.81	23.68	24.35	28.52	28.17	29.04	11.47	11.73	14.66	27.84	28.01	29.78
D	Cq	20.02	19.93	20.15	28.38	28.26	30.19	14.80	14.80	15.71	26.97	27.15	28.45
E	Cq	23.77	23.62	24.81	25.67	25.75	27.35	15.45	15.50	17.35	25.57	25.76	27.93
F	Cq	18.41	18.58	20.43	25.72	25.97	27.87	14.66	14.62	15.20	24.48	24.66	26.38
G	Cq	26.39	26.36	27.20	30.77	31.23	35.89	15.69	16.09	17.65	28.75	28.61	29.78
H	Cq	20.23	20.22	21.73	26.96	26.89	29.06	14.97	14.81	15.23	25.50	25.56	27.99

7.11.2 Calculation of Relative Expression

The calculations on relative gene expression can be followed using **table AX** below.

Table A12. Calculation of relative expression. The raw data from the qPCR was used to generate average C_q-values (**Table A6** & **Table A7**), and the Livak method to calculate relative expression. Fields colored in green show expression following induction and fermentation relative to before induction (yellow).

	C_q GOI1: *Hb*	C_q GOI2: *hemH*	C_q ref. gene 1: *rrsA*	C_q ref. gene 2: *idnT*	Av C_q of ref. genes	ΔC_q GOI1	ΔC_q GOI2	$\Delta\Delta C_q$ GOI1	$\Delta\Delta C_q$ GOI2	rel expression GOi1	rel expression GOI2
AsPgb1.1(-)	26.17	30.82	x	32.52	x	x	x	x	x	x	x
AsPgb1.2 (-)	23.95	28.58	12.62	28.54	20.58	3.37	8.00	0.00	0.00	1.00	1.00
AsPgb1.3 (-)	24.07	26.26	16.1	26.42	21.26	2.81	5.00	0.00	0.00	1.00	1.00
AsPgb1.1	19.99	27.57	14.91	25.91	20.41	x	x	x	x	x	x
AsPgb1.2	20.03	28.94	15.1	27.52	21.31	-1.28	7.63	-4.65	-0.37	25.11	1.29
AsPgb1.3	19.14	26.52	14.83	25.17	20	-0.86	6.52	-3.67	1.52	12.73	0.35
BvHb1.2 (-)	26.65	32.63	16.48	29.04	22.76	3.89	9.87	0.00	0.00	1.00	1.00
BvHb1.2	20.73	26.54	15	26.35	20.675	0.05	5.87	-3.84	-4.01	14.27	16.06