Recombinant expression and purification of sugar beet Phytoglobins in *Escherichia Coli* BL21 (DE3) including metabolic engineering of the heme biosynthesis

Pennie Lindblom & Olivia Geel

M.Sc Thesis in Applied Biochemistry
Division of Pure and Applied Biochemistry
Lund University, Faculty of Engineering, LTH
2023-06-12

Examiner
Prof. Leif Bülow
leif.bulow@kilu.lu.se

Supervisor
Leonard Groth
leonard.groth@tbiokem.lth.se

Assistant Supervisor
Anna Barjuan Grau
anna.barjuan_grau@tbiokem.lth.se

---

pennie.lindblom@gmail.com
0768015066

olivia.geel@hotmail.com
0727229808

Populärvetenskaplig sammanfattning

Kan hemoglobin från växter lösa humana problem?

Dagligen tar vi för givet att vårt blod syresätts tillräckligt för att orka med alla sysslor. Detta möjliggörs av det livsviktiga proteinet hemoglobin som finns i vårt blod. I dagens samhälle råder det dock blodbrist samtidigt som många drabbas av järnbrist. Det är där forskningen på växthemoglobin kommer in i bilden. Experimenten som genomförts i denna rapport har gått ut på att producera hemoglobin från sockerbeta i bakterier och sedan rena upp det. Resultaten är goda och indikerar att det är möjligt att producera denna typ av hemoglobin i bakterier.

Hemoglobin finns i blodet hos nästan alla ryggradsdjur och har i uppgift att syresätta cellerna så att kroppen kan få energi till alla kroppsliga processer, såsom ämnesomsättningen och muskelfunktioner. Proteinet har också i uppgift att föra ut koldioxid ur kroppen. Det är de röda blodkropparna som bär på hemoglobinet och proteinet ger blodkropparna deras röda färg. Det finns däremot flera problem i samhället som lägger grunden för forskning kring hemoglobin.

Inom sjukvården är det livsviktigt att det finns tillräckligt med blod för att kunna genomföra transfusioner och därmed rädda människors liv. Men eftersom bloddonatorerna är för få i dagens läge så är det brist på blod. Detta är ett aktuellt problem som skriker efter en lösning. Ett annat vanligt blodrelaterat problem i samhället är järnbrist. Det kan exempelvis drabba personer med vegetarisk kost. Det brukar inte innebära alltför allvarliga konsekvenser, men ibland kan järnbrist leda till anemi. Detta innebär att blodet har för få röda blodkroppar vilket gör att cellerna inte får tillräckligt med syre. På grund av dessa utmaningar är studier på blodsubstitut och järntillskott ett aktuellt ämne inom forskarvärlden.

Något man har kommit fram till på senare tid är att växter, precis som människor, också har hemoglobin. En lösning på blodbristen i samhället kan därför vara artificiellt blod tillverkat av växthemoglobin. Dessa hemoglobiner från växter skulle även kunna ligga till grund för järntillskott som kan användas i vegetarisk mat.

Projektet i denna rapport baseras på att med hjälp av genteknik producera hemoglobin från sockerbetor i en annan värd, så kallad rekombinant produktion. I detta projektet används *E. coli* som värd för att det är en billig bakterie vars produktion av protein kan skalas upp till större volymer. Vi odlade upp *E. coli,* som innehåller en gen för hemoglobinet, i små flaskor. Därefter skördade vi cellerna och sedan renade vi upp proteinet. Det vi kom fram till är att denna rekombinanta produktion av hemoglobin i *E. coli* lyckades och likaså uppreningen av proteinet. Vi såg också att uttrycket av genen som kodar för hemoglobinet ökade mycket jämfört med naiva *E. coli*. Gener är en mall för proteiner, uttrycks genen mycket så produceras mycket protein. Detta är en bra grund att jobba vidare på i kommande studier för att förhoppningsvis komma ett steg närmare artificiellt blod och järntillskott från växter.

Abstract

The protein hemoglobin (Hb) is vital for us humans to ensure enough oxygen replacement in the body. Hb-resembling proteins also exist in plants, these are called phytoglobins (Pgbs). Pgbs are very interesting since they can play a considerable role in the development of blood substitutes and iron supplements for human use. A few years ago, genes that code for Pgb in sugar beet were found and identified. In this thesis, one of these genes, *BvPgb1.2*, was analyzed further. This was done by bacterial cultivations of *E. coli* BL21 (DE3) that recombinantly produced the sugar beet Pgb BvPgb1.2. Several cultivations were performed, one with native *E. coli* cells, one with *E. coli* cells with wildtype *BvPgb1.2* gene (rWt) and three different mutations of this gene (K34C, N7C and E120C). It could be concluded that the rWt and the mutants showed a 1000-folds upregulation of the *BvPgb1.2* gene in comparison to the native cells seen in the qPCR measurements. However, when investigating the *hemH* gene, the expression did not vary significantly between the cultivations. *HemH* is a key gene in heme biosynthesis. During a δ-ALA gradient experiment, an optimal addition of 1M δ-ALA was found to be 138 µl to a 150 ml cultivation flask since the *hemH* expression was the highest for this addition. Based on the results found in the SDS-PAGE, the QFF purification can be deemed successful. This indicates a thorough removal of other proteins and a fairly pure Pgb sample was received. Gene knockout of the *yfex* gene, coding for an enzyme that degrades heme, with CRISPR-Cas9 was planned but not finalized. This was due to obstacles during the experiments. A troubleshooting procedure showed that the likely error source was the ligation between the crRNA and the pCRISPR plasmid. All together, it has been shown that rWt and the mutant Pgbs can successfully be produced and purified. However, more trials are needed to ensure reliable results.

Abbreviations

Hb	Hemoglobin
Pgb	Phytoglobin
sHbs	symbiotic Hemoglobins
nsHbs	non-symbiotic Hemoglobins
trHbs	truncated Hemoglobins
δ-ALA	δ-Aminolevulinic acid
IPTG	Isopropyl-β-D-thiogalactopyranoside
OD_{600}	Optical Density at 600 nm
rWt	recombinant Wildtype
qPCR	quantitative real-time Polymerase Chain Reaction
SDS-PAGE	Sodium Dodecyl-Sulfate Polyacrylamide Gel Electrophoresis
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
PIPE	Polymerase Incomplete Primer Extension

Table of contents

1. Introduction to project **8**
1.1 Aim of the Master Thesis 8
2. Background **9**
2.1 Hemoglobin 9
2.1.1 The Heme group 10
2.2 Hemoglobin in plants: Phytoglobins 10
2.2.1 Sugar beet phytoglobins 11
2.2.2 Previous studies on sugar beet phytoglobins 12
2.2.2.1 Folding: Conformational dynamics 12
2.2.2.2 Cystein mutants 12
2.3 Recombinant protein production 13
2.3.1 Production host: Escherichia coli BL21 (DE3) 13
2.3.2 Vector used for the mutants 14
2.3.3 Transformation 14
2.3.4 Protein purification 14
2.3.5 Previous studies on hemoglobin production in E. coli 14
2.4 Heme biosynthesis in E. coli 15
2.4.1 Strategy of metabolic engineering 16
2.5 The CRISPR-Cas9 system 17
2.5.1 CRISPR-Cas9 for gene knockout 17
2.5.1 Design of CRISPR-Cas9 platform 17
3. Materials and methods **19**
3.1 Bacterial cultivation 19
3.1.1 Inoculum preparation 19
3.1.2 Induction and Cultivation 19
3.1.3 Harvesting 19
3.2 Downstream processing for qPCR 20
3.2.1 RNA extraction 20
3.2.2 NanoDrop 20
3.2.3 Gel electrophoresis 20
3.2.4 cDNA synthesis 20
3.2.5 qPCR & relative gene expression 21
3.3 Downstream processing for protein investigation 22
3.3.1 Cell resuspension and homogenization 22
3.3.2 Dialysis and filtration 22
3.3.3 ÄKTA chromatography 23
3.3.4 SDS-PAGE 23
3.4 Gene knockout using CRISPR-Cas9 24
3.4.1 Design of crRNA and donor DNA 24

3.4.2 Plasmid recovery from bacteria and plasmid purification 24
3.4.3 Creation of competent cells 24
3.4.4 Cloning crRNA into pCRISPR 25
3.4.5 Combination of plasmids using PIPE 25
3.4.6 Vector transformation 25
4. Results **26**
4.1 Pellet color analysis 26
4.2 qPCR: Relative gene expression 26
4.2.1 The phytoglobin mutants 26
4.2.2 The δ-ALA gradient 27
4.3 QFF purification chromatograms 28
4.4 SDS-PAGE 33
4.5 Genome editing with CRISPR-Cas9 35
4.5.1 Cloning crRNA into pCRISPR 35
4.5.2 Troubleshooting 35
4.5.3 Combination of plasmids using PIPE 40
5. Discussion **41**
5.1 Pellet color analysis 41
5.2 qPCR: Relative gene expression 41
5.2.1 The phytoglobin mutants 41
5.2.2 The δ-ALA gradient 42
5.3 QFF purification chromatograms 43
5.4 SDS-PAGE 43
5.4.1 Mutant N7C 43
5.4.2 Mutant E120C 45
5.5 Genome editing with CRISPR-Cas9 46
5.5.1 Troubleshooting 46
5.5.2 Combination of plasmids using PIPE 48
5.6 Future considerations 48
6. Conclusion **50**
Acknowledgements **51**
Contributions **51**
7. References **52**
8. Appendix **57**
8.1 NanoDrop Results 57
8.2 Gel electrophoresis 58
8.3 GeneRuler and PageRuler guide 59
8.4 Primer validation and efficiency calculations 60
8.4.1 Melt curve analysis 60
8.4.2 Primer efficiency 60
8.5 Relative gene expression calculations - Raw data 61

8.5.1 Phytoglobin mutants 61
8.5.2 The δ-ALA gradient 63
8.6 CRISPR-Cas9: Plasmid recovery from bacteria and plasmid purification 64
8.7 Primers for PIPE 66
8.8 Protocols 67
8.8.1 Cultivation protocol 67
8.8.2 Protocol for plasmid purification from bacterial stabs 71
8.8.3 Protocol for cloning crRNA into pCRISPR 72
8.8.4 Lab protocol for CRISPR-Cas9 75
8.8.5 Control transformation Protocol from the division 79
8.8.6 Ligation protocol from New England BioLabs 80

1. Introduction to project

There is a constant need for accessible blood in healthcare. In Sweden, one bag of blood per minute is needed but the number of blood donors decreases. This is a problem that once in a while gives acute shortages of blood in the hospitals [1]. Especially during the covid-19 pandemic, the shortage in Sweden had an adverse effect since it forced surgeries to be canceled [2]. The shortage of blood is one reason to facilitate research on artificial blood from other species that can be used in humans. Another societal problem is that vegetarians have a higher risk of suffering from iron deficiency due to their diet in contrast to omnivores. Iron deficiency is a condition where the iron supply to the tissues in the body is suppressed. Vegetarians run a higher risk of iron deficiency due to the low bioavailability of iron in plants. This due to that iron found in plant-based foods does not contain heme which makes it much less absorbable. If the condition regarding iron deficiency becomes really serious this can lead to anemia [3], which is a condition where the patient has a low amount of healthy red blood cells. This leads to blood not containing enough oxygen [4].

For about a century there has been on-going research on blood substitutes to treat anemic conditions all over the world. To this day there are no oxygen therapeutics that are globally accepted or clinically used. The reason that the research has not been as successful as wanted is due to that human blood transfusions bring many difficulties. These hindrances are such as contamination risks, specific storage demands and limitation in shelf-life [5].

Thereby new ways to handle problems like shortage of blood and anemia are necessary. The research about plant hemoglobins, called phytoglobin (Pgb), and their similarity to human hemoglobin is therefore interesting in this field. Recombinant protein production in different expressions hosts have been seen to work effectively to yield this kind of Pgbs. This type of production is promising due to its cost-effectiveness and scalability. These prerequisites also provide opportunities for use in future applications such as supplements in vegan food or as the core of artificial blood.

1.1 Aim of the Master Thesis

The aim of this thesis is to recombinantly produce the Pgb BvPgb1.2 from sugar beet (Beta vulgaris spp. vulgaris) in *Escherichia coli* BL21 (DE3) as well as three cysteine mutants of this protein. The gene expression for the key gene *hemH* in the heme biosynthesis and the Pgb gene *BvPgb1.2* will be investigated using qPCR. Additionally the gene expression of *hemH* will be studied with varying additions of δ-Aminolevulinic acid in the growth media. This is done to evaluate its role in heme production. The produced Pgbs will also be purified using ÄKTA chromatography. Another aim is to design a CRISPR-Cas9 system that intends to result in a gene knockout of *yfeX*, a gene coding for a deferrochelatase in the heme biosynthesis. This to try to increase the heme production which could possibly increase the Pgb yield. All with the end goal of increasing our understanding of Pgbs as well as the tools used in recombinant protein production. Moreover, this could be a step in the development towards new applications such as artificial blood and iron supplements made from plants.

2. Background

This report regards hemoglobins from plants and therefore some fundamental understanding of hemoglobins are needed and will thus be presented below. Additionally, plant hemoglobins will be presented further along with the chosen production host *Escherichia coli* BL21 (DE3). The metabolic pathway of heme biosynthesis will be explained for understanding the choices of genes investigated in the methods. Lastly, CRISPR-Cas9 will be elucidated since an implementation of this platform will be attempted.

2.1 Hemoglobin

Hemoglobin (Hb) is a vital protein that can be found in red blood cells. Its main purpose is to carry oxygen from the lungs to the tissues and to transfer carbon dioxide and hydrogen ions back to the lungs. Human Hb, also known as HbA, consists of four polypeptide chains with two identical α chains and two identical β chains. It is a pair of identical $\alpha\beta$ dimers that form the tetramer by associating. The chains bind one heme group each which can be seen in figure 1. These four chains bind oxygen in a cooperative manner, which means that the chance of one site in a chain binding oxygen increases if another chain already has bound oxygen. The oxygen-carrying capacity is also increased by the binding of carbon dioxide and hydrogen ions. Oxygen binds to Hb via the iron ion situated in the heme that binds tightly to Hb. The quaternary structure changes considerably when oxygen binds [6].

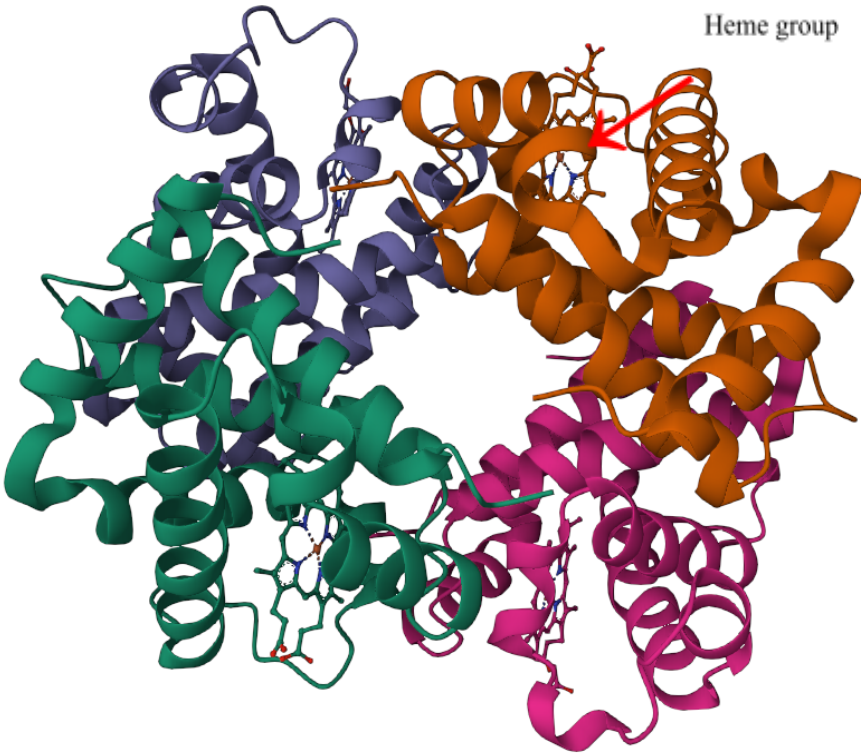

Figure 1. Human hemoglobin where each color represents one chain of the protein (two α chains and two β chains). The chains bind one heme group each, one of the heme groups is marked with a red arrow [7].

2.1.1 The Heme group

Heme is a prosthetic group that is essential for the function of Hb since it binds the oxygen. The heme consists of an iron ion situated in the center of the organic compound protoporphyrin. The iron can be in two different oxidation states, in ferrous state (Fe^{2+}) or in ferric state (Fe^{3+}). To enable the oxygen to bind, the iron needs to be in the ferrous state and the oxygen will bind to the sixth coordination site of the iron. Iron can form two additional bonds, where the sixth coordination site is one of them and the other one is the fifth coordination site [6].

Heme is not just essential for Hbs, there are other proteins that require heme for function; hemoproteins [8]. Examples of hemoproteins are Hb, myoglobin, catalase and cytochrome c. They all play an important role in oxygen transport as well as transferring electrons for energy generation [9]. Free heme in bacteria can lead to heme-mediated toxicity and therefore it is essential for the cell to incorporate it into hemoproteins. The intracellular heme concentration is hence strongly regulated [10].

2.2 Hemoglobin in plants: Phytoglobins

There are three types of Hbs existing in plants: symbiotic hemoglobins (sHbs), non-symbiotic hemoglobins (nsHbs) and truncated hemoglobins (trHbs). sHbs has for a long time been known as the proteins existing in roots of legumes that are in symbiosis with nitrogen-fixing bacteria and are called leghemoglobins [11]. Since 2014 the nomenclature has been altered to clarify the different sorts of plant Hbs. sHbs continue to be named leghemoglobins while nsHbs and trHbs are called phytoglobins (Pgbs) [12].

TrHbs are Pgbs where the alpha-helical fold is 2-on-2 fold instead of 3-on-3, like most globins [11]. This implies that the trHbs often have polypeptides of shorter length due to less residues [13]. Since the focus will be on Pgbs from sugar beets, the structure and function of nsHbs will be more thoroughly described. The nsHbs are hexacoordinate which means that a histidine reversibly binds the sixth coordination site of the iron in the heme. This happens in both the ferrous and the ferric state. This differs from sHbs and mammalian Hbs since they are pentacoordinate [11]. NsHbs can be divided into two different classes, class 1 (nsHbs1) and class 2 (nsHbs2). This depends on oxygen binding properties, phylogenetic analysis and expression patterns. Plants that are monocots have one or more nsHb1s but do not have any nsHb2s. Dicots on the other hand, often have at least one nsHb1 and one nsHb2 [14]. The difference between monocots and dicots is how many seed leaves are located in the embryo in the seed [15].

Pgbs do not play the exact same role in plants as Hbs does in vertebrates. Instead of only being a transporter for oxygen, Pgbs have been associated with the nitric oxide metabolism and energy maintenance. It has especially been connected to NO scavenging in hypoxic conditions and stress adaptation depending on the class [16].

2.2.1 Sugar beet phytoglobins

It has been shown that sugar beet (Beta vulgaris spp. vulgaris) have four separate genes coding for Pgbs. These are three nsHbs (BvPgb1.1, BvPgb1.2 and BvPgb2) and one truncated gene (BvPgb3). BvPgb1.1 and BvPgb1.2 codes for Pgbs that belong to class 1 and BvPgb2 to class 2 [14]. This sugar beet species is a dicot and belongs to the order Caryophylalles. Sugar beet can be found in temperate climates and is interesting for research due to its importance for the global sugar and energy production [17]. It has been shown that the BvPgb1.2 almost always appears in a dimeric form (see figure 2) with an theoretical weight at 38.4 kDa [18]. The three found nsHbs in sugar beet have been compared with Pgbs from monocots and eudicots. It was found that the Pgbs have different functions in different plants which indicate the diversity of these Pgbs. Because of the diversity, these proteins can be seen as future protein additives to improve food and agricultural conditions [19]. However in general, it seems that Pgbs can functionally be important for plant development and early stages of the plant such as germination, since BvPgb1.2 mostly is expressed in the seeds of the sugar beet [18].

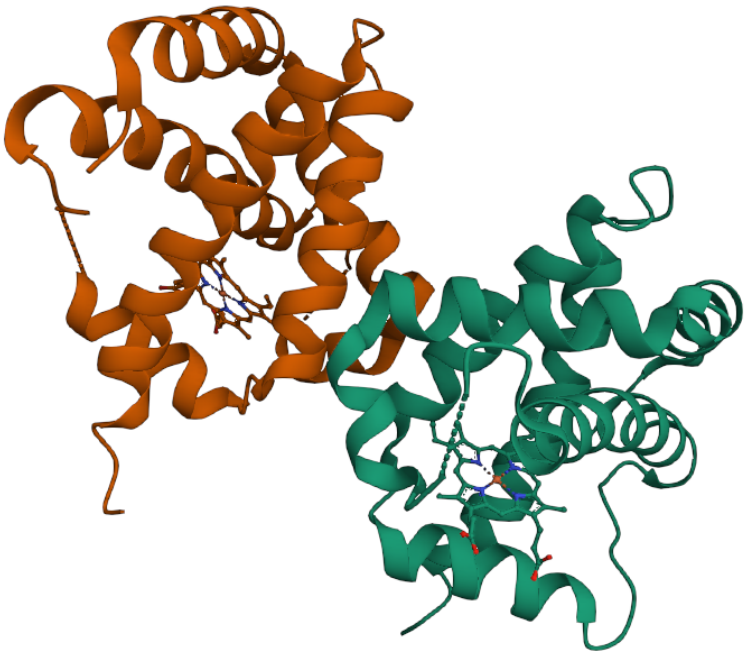

Figure 2. The dimeric phytoglobin BvPgb1.2 where each subunit is represented by a different color. In the figure, the two heme groups, one for each subunit, can be seen [20].

2.2.2 Previous studies on sugar beet phytoglobins

There are several studies made on sugar beet Pgbs at the Division of Pure and Applied Biochemistry at LTH. In this report one study about the folding of the Pgbs and one about cysteine mutants will be brought up.

2.2.2.1 Folding: Conformational dynamics

In the study "Conformational Dynamics of Phytoglobin BvPgb1.2 from *Beta vulgaris* ssp. *vulgaris"* it is investigated which residues are important for the dimeric conformational dynamics in BvPgb1.2 using NMR technique. They concluded that there are several non-assigned residues in BvPgb1.2 between the α-helixes G and H that are involved in the dimerization. This trend has also been seen in other Hbs. The results in this study confirms the dimeric structure of the BvPgb1.2. Moreover it was shown that it has higher flexibility around the heme pocket which might affect residue interactions and solvent accessibility [19].

2.2.2.2 Cystein mutants

In another study it is investigated how cysteine plays a functional and structural role in BvPgb1.2. This was based on the fact that Pgbs are highly diverse but remarkably have conserved cysteine residues at functionally important places. To investigate the role of cysteine they altered the BvPgb1.2 with a site directed mutagenesis substituting the conserved cysteine at position 86 for alanine (Cys86Ala). They could conclude that the autoxidation rate increased for the mutant compared to the wildtype suggesting that cysteine is important for function and oxidative stability. This without altering the tertiary and quaternary structure [18].

Since the above mentioned study concluded the importance of cysteine, mutants substituting other amino acids to cysteine is an interesting approach to try. In this thesis three mutants of sugar beet Pgbs were recombinantly produced and analyzed. These were K34C, N7C and E120C. In K34C the lysine at the position 34 of the protein was substituted with cysteine. In N7C the asparagine at position 7 was substituted with cysteine. Lastly, in the mutant E120C glutamic acid was replaced with cysteine. As an example of a visualization of the mutations, see figure 3 where the locations of K34 can be seen on the surface of the protein in green. These are the locations where cysteine has been inserted instead of lysine.

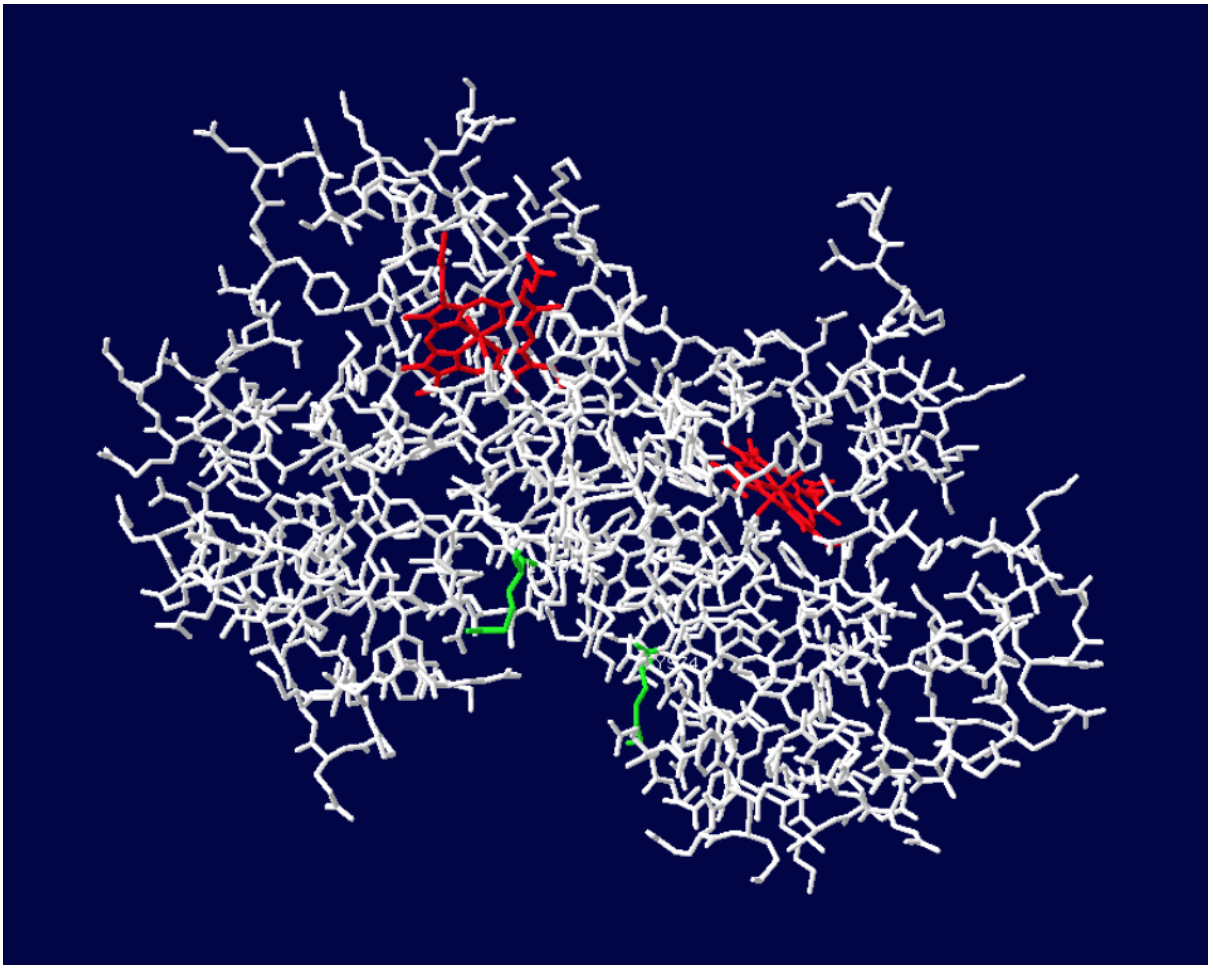

Figure 3. BvPgb1.2 with the K34 position seen in green on the surface, one on each subunit. The heme groups for each monomer can be seen in red. This figure is established by us using Swiss PDB-Viewer but based on the protein structure from figure 2.

2.3 Recombinant protein production

2.3.1 Production host: *Escherichia coli* BL21 (DE3)

Escherichia coli is an easy to work with enterobacterium that is the obvious choice for bacterial cultivations due to its diverse and hardy nature [21]. As an facultative anaerobe it can grow both with and without oxygen; it however grows optimally in an aerobe condition at 37℃ and a pH of 7.0. *E. coli* is subdivided into many commercially available and frequently used strains based on genotype and application [22]. The strain *E. coli* BL21 (DE3) is a well established microbial cell factory for biotechnical applications such as recombinant protein production. BL21 (DE3) is thoroughly annotated and its genotype is based on an addition of a prophage carrying the gene for a T7 RNA polymerase [23]. T7 synthesizes RNA much more efficiently compared to RNA polymerase in native *E. coli* which makes it a suitable host for expression of proteins [24].

2.3.2 Vector used for the mutants

The vector used for the sugar beet Pgb mutants is the pET-DEST42 vector. It contains a controllable T7 promoter under repression from the *lac* operon [25]. A gene coding for a *lac* repressor is present in the vector and when it is synthesized it binds to the T7 promoter which makes the expression of the inserted gene unavailable. During induction, isopropyl-*β*-D-thiogalactopyranoside (IPTG) is added and can bind to the repressor which makes it release from the promoter and induces the expression of the target protein. This type of controllable promoter is favorable for an enhanced recombinant protein production [26]. In addition to a controllable promotor, the pET-DEST42 vector also contains the *ampR* gene for ampicillin and carbenicillin resistance [25]. This for selection of successful transformants and to make sure that only the wanted cells can grow [26].

2.3.3 Transformation

To be able to transform a vector into cells they need to be competent. A competent cell has an altered cell wall and can thus more easily take up foreign DNA. There are mainly two types of competent cells: electro-competent and chemically competent cells. Electro-competent cells are created by a process called electroporation or electropermeabilization and is based on that an high voltage electrical current is applied on the cells which created pores in the cell wall. Through these pores, DNA can enter the cell and they will be transformed [26]. Chemically competent cells are created by a $CaCl_2$ treatment that makes the plasmid bind to the surface of the cell. This is followed by a heat shock that enables the plasmid to enter the cell [27]. Electroporation is more efficient than chemically competent cells but harder to perform in practice since it requires an electroporator. Chemically competent cells are more laborious to make but can be performed in almost any lab [28].

2.3.4 Protein purification

Purifying proteins are often performed using a chromatography system, in this thesis an *ÄKTA avant chromatography system* from Cytiva. The column used is a QFF which is an anion exchange chromatography column where negative charged proteins can attach. Q stands for Quaternary ammonium which is the positively charged ions that are packed in the column and will catch the anions. FF stands for Fast Flow. Due to the isoelectric point of the target protein (BvPgb1.2) it is negatively charged at pH 8.5 and the column is established based on this [29]. Fractionation during the ÄKTA chromatography starts at 412 nm since this is when heme absorbs light [30].

2.3.5 Previous studies on hemoglobin production in *E. coli*

Previous studies have shown a correlation between the choice of *E. coli* strain and yield of the produced Hb. One such study looking at a Hb from the deer mouse, *Peromyscus maniculatus* concluded that the strain BL21Star™ (DE3) gave the highest yield. This by using a low induction temperature over a longer time (12°C for 24 h) [31]. The main features of *E. coli* BL21Star™ (DE3) is the same as BL21 (DE3), however one additional mutation in

the *rne131* gene is added. This mutation knocks out an enzyme responsible for mRNA degradation which results in mRNA stability in the strain [32].

Another study has highlighted one important factor to take into consideration when producing Hb in *E. coli*. That a sufficient amount of heme must be present for Hb to be able to fold properly, otherwise it will be degraded. This has previously been addressed by external addition of heme to the medium. This study, additionally to heme in the media, tried to coexpress heme transport genes from *Plesiomonas shigelloides* with human Hb. This resulted in an 10-fold increase in produced soluble Hb [33].

2.4 Heme biosynthesis in *E. coli*

Heme synthesis in *E. coli* starts with the synthesis of δ-Aminolevulinic acid (δ-ALA), in figure 4 called 5-Aminolevulinate. This synthesis is performed through either the C4 or C5 pathway depending on species. Bacteria, like *E. coli*, uses the C5 pathway (blue in figure 4) that stems from α-ketoglutarate from the TCA cycle. It then undergoes several conversions that lead to δ-ALA. Several studies have indicated that the production of δ-ALA is a rate limiting step for heme production and upregulation of pathway enzymes or media supplementation of δ-ALA is beneficial for the biosynthesis [34, 9, 35]. When δ-ALA is externally added in the media it can permeate the outer membrane of *E. coli* through non-specific porins to the periplasm. It is then taken up by the bacteria to the cytoplasm through a specific uptake system, the dipeptide permease transport system [36, 37].

From δ-ALA, the pathway continues into the downstream heme biosynthesis pathway (pink in figure 4) starting with the conversion to Porphobilinogen by an enzyme coded by *hemB*. This is followed by several conversions performed by enzymes from the genes *hemC*, *hemD*, *hemE*, *hemF*, *hemG* and *hemH*. The last step performed by ferrochelatase (FECH) encoded by *hemH* is to incorporate the ferrous iron needed to create heme. Newly synthesized heme can then bind to proteins and create hemoproteins such as Hb [8, 9].

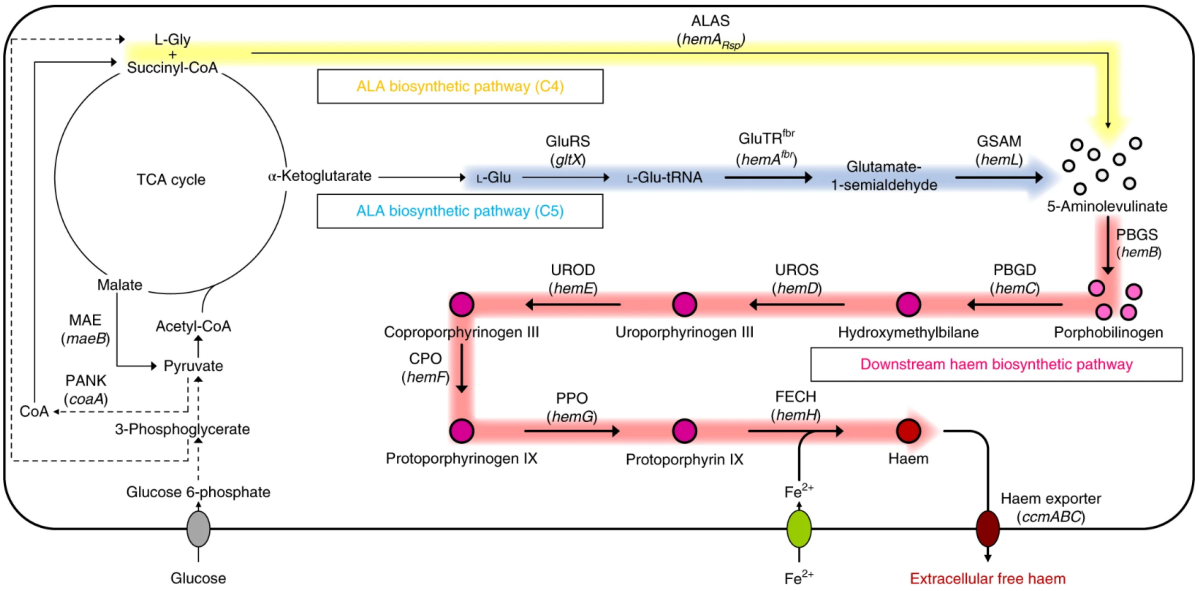

Figure 4. Metabolic pathways for the biosynthesis of heme in the mitochondria of E. coli. The yellow and blue pathways show the C4 and C5 routes for δ-ALA synthesis and the pink is the downstream heme biosynthesis pathway [9].

When investigating the expression of mRNA using qPCR, *hemH* is an interesting target to look at as the last step in the synthesis. An increased expression of this gene could indicate an increased heme synthesis.

2.4.1 Strategy of metabolic engineering

Increasing the heme synthesis gives rise to an increased hemoprotein production. Previous studies have attempted to engineer the metabolic pathways included in the heme biosynthesis to increase the production with various results. A common way is to overexpress the genes *hemC-H* in the downstream pathway to enhance the reaction and production rate and bypass bottlenecks in the synthesis [38, 9]. A continuation of this has been attempted by Zhao et. al who aimed to knock out the gene *yfeX* that codes for deferrochelatase. This enzyme catalyzes the reverse reaction of ferrochelatase and removes the iron from heme which initiates heme degradation (see figure 5). This knockout resulted in a higher heme titer compared to just overexpressing the pathway genes and a strain constituting the combined strategies was deemed the most successful in heme production [9].

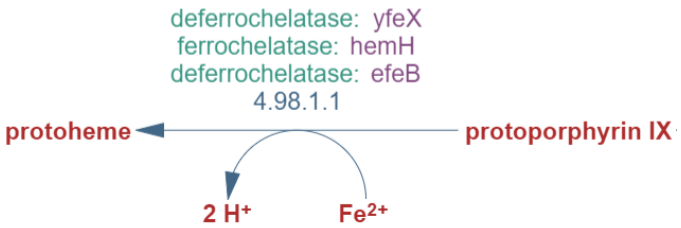

Figure 5. The last step in the heme biosynthesis. Ferrochelatase coded by hemH performs the reaction in the direction of the arrow while deferrochelatase coded by either yfeX (cytoplasmic enzyme) or efeB (periplasmic enzyme) does the reverse reaction [39].

As previously mentioned, heme can bind to proteins and create hemoproteins. Studies have shown that the enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is a heme chaperone meaning it aids in the transport and binding of heme to downstream proteins [40]. Because of this, other studies using mouse cell lines have investigated its role in heme maturation of globins. An overexpression of wildtype GAPDH has shown an increase in heme levels in globins such as myoglobin and Hb strengthening the theory on its role as a heme chaperone [41].

2.5 The CRISPR-Cas9 system

CRISPR-Cas9 is a revolutionary tool for genome editing that derives from an existing defense system found in bacteria and archaea. It is based on integration of DNA that comes from previous foreign infections from phages into a CRISPR (clustered regularly interspaced short palindromic repeats) loci. From this loci, a small RNA fragment is coded for. This fragment is complementary to the infectious DNA and can guide an endonuclease called Cas to cleave the target DNA. This is then used as a recognition system that will be able to fight off the infection the next time it enters the cell, like adaptive immunity [42].

2.5.1 CRISPR-Cas9 for gene knockout

This CRISPR-Cas9 system has in recent years been utilized for genome editing in a wide variety of cell lines by designing the system for a specific application. One such application can be a gene knockout and to achieve that the first step is to design the crRNA. The crRNA is a 20 bases long sequence that is complementary to a target sequence in the gene that shall be disabled. It must also be upstream of a PAM sequence, in Cas9: 5'-NGG-3'. The crRNA is then usually implemented in a vector together with tracrRNA and combined they create the guide RNA (gRNA). The gRNA vector is then transformed into a competent cell together with a vector containing the gene for the endonuclease Cas9 from *Streptococcus pyogenes*. When the gRNA and Cas9 are expressed together in the cell, the tracrRNA binds to Cas9 and the crRNA part of the gRNA guides it to the site that is complementary to its base pairs. After binding, Cas9 makes a double stranded break in the target sequence and the cells that cannot repair this break will die [42]. For the simplest way of achieving a gene knockout, the cell uses non homologous end joining to repair the DNA and this often results in indel errors in the gene making it not be able to express correctly. Another way is to co-transform the cells with a donor DNA that has homologous sequences up- and downstream of the target site. This enables the cell to perform homologous recombination to repair itself and an advantage is that it is possible to insert sequences between the homologous parts, like a stop codon or an entirely different gene [43].

2.5.1 Design of CRISPR-Cas9 platform

The implementation of the CRISPR-Cas9 system in this thesis is based on an article written by Zerbini et al from 2017. Four plasmids were ordered from Addgene [44]: pCRISPR (Addgene plasmid #42875), pCNA (Addgene plasmid #135191), pKM154 (Addgene plasmid

#13036) and pCas9 (Addgene plasmid #42876). pCRISPR contains an insertion site for implementing the crRNA and pKM154 contains the *sacB* gene coding for the *Bacillus subtilis* levansucrase. This enzyme creates toxic metabolites for *E. coli* when having access to sucrose. These two plasmids will be combined using the polymerase incomplete primer extension (PIPE) method creating the plasmid pCRISPR-sacB. The chosen crRNAs will then be cloned into pCRISPR-sacB to create pCRISPR-sacB-crRNA. The sacB gene is used for removal of the pCRISPR-sacB-crRNA plasmid after successful gene modification. This by growing the cells in a media containing 5% sucrose since only cells who lose the plasmid will be able to grow [45]. pCas9 contains the gene coding for the endonuclease Cas9 as well as the tracrRNA. pCNA has the genes coding for the λ-Red machinery which aids in homologous recombineering by protecting the donor DNA from degradation and promotes annealing to the target [46]. These two plasmids will be combined using PIPE to create pCasRed [45]. pCRISPR-sacB-crRNA and pCasRed will be co-transformed into native *E. coli* BL21 (DE3) cells with the goal of knocking out the gene *yfeX*.

3. Materials and methods

3.1 Bacterial cultivation

The cultivation was performed based on a previously seen successful protocol obtained from supervisors at the Division of Pure and Applied Biochemistry at LTH. For the detailed step by step protocol, see Appendix 8.8.1. Five different bacterial cultivations were made:

- Native *E. coli* BL21 (DE3)
- *E. coli* BL21 (DE3) transformed with wildtype BvPgb1.2 (hereafter referred to as recombinant wildtype - rWt)
- *E. coli* BL21 (DE3) transformed with BvPgb1.2 mutant K34C
- *E. coli* BL21 (DE3) transformed with BvPgb1.2 mutant N7C
- *E. coli* BL21 (DE3) transformed with BvPgb1.2 mutant E120C

3.1.1 Inoculum preparation

The first step in the bacterial cultivation is the preparation of the inoculums. 5 ml LB media was added into the desired number of inoculation tubes. Depending if the cells were transformed or not, ampicillin was added to the stock. With the use of a sterile pipette, one colony from the desired plate was transferred to each inoculation tube. The plates had been prepared with plate pouring of agar and plating with desired bacterial stock. The inoculation tubes were then incubated.

3.1.2 Induction and Cultivation

In cultivation for large shake flasks, the protocol was used. If the cultivation was performed with the small shake flasks, 111 ml of phosphate buffer was added into 1000 ml of C/N source and 1111 µl of carbenicillin was added if desired. 150 ml of the TB media was then transferred into each shake flask and 1 ml of the inoculums was added. The flasks were incubated while shaking until the $OD_{600} \geq 2.5$ a.u. The induction by IPTG and addition of δ-ALA, was performed according to protocol for large flasks. If small flasks were used, 76 µl IPTG and 46 µl δ-ALA were added. The flasks were then bubbled with carbon monoxide for 10 seconds for small flasks and 25 seconds for large flasks and sealed with parafilm over the corks and incubated.

One cultivation was made to investigate how different additions of δ-ALA affected the gene expression of *hemH* in native cells. This was made in small flasks as described above, but with different volumes of 1M δ-ALA: 23 µl, 46 µl, 92 µl, 138 µl and 184 µl. The flasks were not bubbled with carbon monoxide.

3.1.3 Harvesting

The OD_{600} of the cell suspensions was measured and 400 µl was transferred from each flask into two eppendorf tubes with 200 µl in each. These were centrifuged for 5 min at 13 000 rpm. The supernatants were removed and the tubes were flash-freezed and put into the -80℃

freezer, to enable RNA extraction later on. The remaining cell suspensions in the flasks were poured into centrifugation flasks, 1 L flasks for the large flask cultivation and 200 ml flasks for the small flask cultivation. Then the flasks were centrifuged, the pellet resuspended and then transferred to falcon tubes according to protocol. The empty falcon tubes had been weighed, to be able to acknowledge the weight of the pellets later on. The falcon tubes with the resuspended cells were centrifuged, supernatant was discarded and the pellets were weighed, photographed and flash-freezed.

3.2 Downstream processing for qPCR

3.2.1 RNA extraction

RNA extraction was performed on the samples taken before harvesting. The *"Quick-RNA™ MiniPrep"* from Zymo Research was used and the attached instructions were followed for all samples. Since 200 µl sample was used, it was estimated based on that one unit of OD corresponds to $8*10^8$ cells/ml [47], that over 10^7 cells were present in the eppendorf tube. Therefore 600 µl lysis buffer was used in the first step. For the ethanol in step three, 600 µl was added and then 1 ml of the pooled ethanol/lysis buffer was transferred to the next column. RNA was eluted using 70 µl DNase/RNase-free water instead of 100 µl. The extracted RNA was kept at 4℃ if it was to be used directly after, otherwise it was kept at -80℃.

3.2.2 NanoDrop

Nanodrop was used for determination of concentration and purity of the extracted RNA. A nanophotometer (Implen) was used with a lid factor of 10 and the predetermined settings. The samples were taken in duplicates with a volume of 1.5-2.0 µl. The purity was measured using the ratios A260/A280 and A260/A230 where A260/A280 shows the purity of RNA in relation to proteins and A260/A230 in relation to carbohydrates and phenols. A value of A260/A280>1.8 and A260/A230>2 is considered to be pure for RNA. These ratios can also be applied to DNA.

3.2.3 Gel electrophoresis

To determine if the RNA is intact, gel electrophoresis on 1-2% agarose gel was performed. 5 µl sample was stained with 1 µl TriTrack DNA loading dye from Thermo Scientific and the ladder was mixed with 1 µl GeneRuler 1 kb DNA Ladder, 4 µl Milli-Q water and 1 µl TriTrack DNA loading dye stain. The gel was run for 30 minutes at 120 V and was then visualized using the *Bio-Rad Gel Doc XR System w/ Universal Hood II*. If two bands were detected, the RNA was further used for cDNA synthesis.

3.2.4 cDNA synthesis

From the values (concentration in ng/µl) obtained in the NanoDrop analysis, the volume to obtain 300 ng cDNA was calculated. Then the synthesis was performed based on the

manufacturer's protocol for the *RevertAid RT Kit* from Thermo Scientific. Using the *SenoQuest labcycler* the samples were thermocycled in the following conditions: 5 min at 25℃, 60 min at 42°C, 5 min at 70℃ and then hold at 4°C if qPCR was run directly after otherwise they were stored at -80℃ for further use.

3.2.5 qPCR & relative gene expression

For qPCR, the cDNA stock was first diluted 20 times to a working stock of 0.375 ng/ul (300 ng/40 µl/20x dil.) which resulted in a total mass of 1.5 ng cDNA in each well. To each well reagent was added according to table 1 based on Applied Biosystems™ *SYBR™ Green PCR Master Mix* instructions.

Table 1. Components added to each well for the qPCR reaction.

Component	Volume
Milli-Q water	4 µl
SYBR™ Green PCR Master Mix	10 µl
Forward primer (1000x)	1 µl
Reverse primer (1000x)	1 µl
Sample 20x dilution (or Milli-Q water for blank)	4 µl
	Total volume: 20 µl

Two reference primers were used: rrsA and cygC that are housekeeping genes not affected by the cultivation conditions. Primers for the target genes *hemH* and *BvPgb1.2* were also used for investigation of the mRNA expression. All primers were validated using melt curves and their efficiencies were calculated according to Appendix 8.4 prior to use. All samples and blanks were added in triplicates for each primer pair.

The qPCR reactions were run with a *CFX96TM Real-Time System* in a *C1000 TouchTM Thermal Cycler* from Bio-Rad. The program used consisted of three main steps. Firstly the temperature was increased to 95°C for 10 min to activate the *AmpliTaq Gold® DNA Polymerase*. This was followed by 15 seconds at 95°C for denaturing of the cDNA as the second step. Thirdly, a decrease in temperature to 65°C for 1 min for primer annealing/extension. Step two and three were then repeated for 40 cycles (step 4 in figure 6). Lastly, melt curves for the primers to ensure specific binding were produced as the fifth step. This by increasing the temperature from 65°C to 95°C with a rate of 0.5°C/cycle and with each cycle being 5 seconds long.

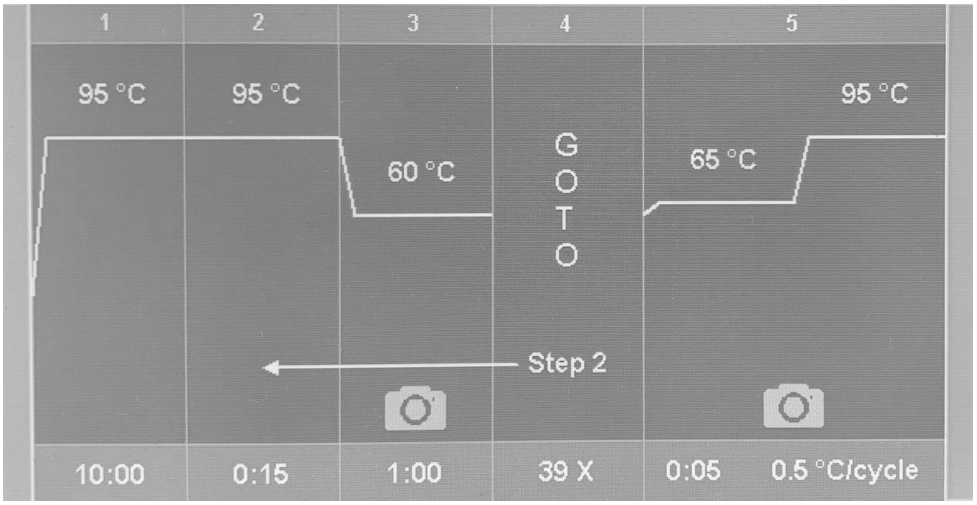

Figure 6. The qPCR program used for all samples in the C1000 TouchTM Thermal Cycler.

The obtained results were analyzed by calculating the relative gene expression using multiple reference genes (Equation 1).

$$Relative\ gene\ expression\ = \frac{(E_{GOI})^{\Delta Ct\ GOI}}{GeoMean[(E_{REF})^{\Delta Ct\ REF}]} \quad (1)$$

Where *E* is the base of exponential amplification, *GOI* stands for Gene Of Interest, *REF* is the reference genes and *Ct* the cycle threshold. *ΔCt* is the calibrator *Ct* subtracted with the sample *Ct*. The calibrator *Ct* is the *Ct* for the sample that the expression will be relative to. *GeoMean* is the geometric mean value of the E_{REF} between the two reference genes.

3.3 Downstream processing for protein investigation

3.3.1 Cell resuspension and homogenization

The cells from one cultivation were stored in falcon tubes and were brought out to thaw at room temperature. 2 ml of 50 mM TRIS-HCl (pH 8.5) was added per g cells in each tube. The resuspended cells were pooled into one schott flask and put on ice. Thereafter the homogenization was performed with the machine *Maximator Homogenizer* according to instructions provided by Christina Wennerberg at the Division of Biotechnology. The homogenization cycle was repeated four times. One sample from the homogenized cells was taken for SDS-PAGE analysis and stored in the -20°C freezer.

3.3.2 Dialysis and filtration

The homogenized cells were then prepared for dialysis. The homogenized solution was divided in half where one of the parts was centrifuged for 20 minutes at 4°C and 11 000 rpm. After centrifugation, one sample was taken from the supernatant and one from the pellet for the SDS-PAGE. Then the non-centrifuged solution was put in one dialysis bag and the supernatant of the centrifuged solution was put in another. The dialysis bags were called *Spectra/Por®1 Dialysis Membrane* with a cutoff of 6-8 kDa and 5.1 ml/cm in volume/length from Spectrum Laboratories. The bags were sealed and put in a four liter container with 50

mM TRIS-HCl (pH 8.5) and changed to a new TRIS-HCl container after six hours. They were left in the second container during the night. After the dialysis the solutions were filtered with *Minisart® Syringe Filter* for particle removal of 0.45-0.65 µm sizes and held on ice. There were also samples for the SDS-PAGE analysis taken after the dialysis and after the filtration.

3.3.3 ÄKTA chromatography

Before the chromatography with the QFF was started in the *ÄKTA avant chromatography system* from Cytiva buffers were filtered and degassed to be used in the system. The preparation of the ÄKTA chromatography started with running all the tubings with Milli-Q water and purging the bubbles in the system. The column was saturated with Milli-Q water using 1.2 column volumes. Then all the tubings were filled with the correct buffer. The 50 mM TRIS-HCl (pH 8.5, standard buffer) and the 50 mM TRIS-HCl with 200 mM NaCl (pH 8.5) had been supplemented with 25 µM Dithiothreitol (DTT) as a reducing agent to prevent disulfide bridges. A pre-made program customized for the protein in question and the column, was used. The fractionation started when the absorbance for the wavelength 412 nm exceeded 100 mAu. During the program the fractionated sample was divided into 15 ml tubes filled with 8 ml in each. When the program was done the column had been saturated with Milli-Q water and then all the tubing was flushed with Milli-Q water. The tubes for the standard buffer and the sample were also flushed with 20% ethanol and then left in it. Then a CIP program was performed and the whole system was left in 20% ethanol until the next run. The tubes with the fractionated sample were pooled from each row and one sample from each row were also taken for the SDS-PAGE analysis. The remaining pooled samples were flash-freezed and put in the -80°C freezer. This whole ÄKTA run was repeated with the centrifuged sample. However the CIP was not required between runs in close proximity.

3.3.4 SDS-PAGE

To enable loading of the samples in the SDS-PAGE gel, a SDS-loading buffer was prepared by mixing Milli-Q water, 0.25 M Tris-HCl (pH 6.8), Sodium dodecyl sulfate, glycerol, β-mercaptoethanol and 0.5% Bromphenolblue. The 15 well-gel used was the *Mini-PROTEAN TGX Gel* from Bio-Rad and was constructed in an electrophoresis chamber filled with SDS-PAGE running buffer. 50 µl of each sample was mixed with 50 µl SDS-loading buffer. The samples were then boiled at 95°C for 10 minutes. 5 µl of each sample and the ladder *Page ruler* was added into each decided lane. The gel was run for 40 minutes at 150 V and 400 mA. After the electrophoresis was done the gel was left in a staining solution for 30 minutes on an orbital shaker. Thereafter the gel was washed four times for 30 minutes with a destaining solution and left on the shaker. The gel was then left in Milli-Q water to avoid shrinkage and thereafter visualized with the *Bio-Rad Gel Doc XR System w/ Universal Hood II*.

3.4 Gene knockout using CRISPR-Cas9

3.4.1 Design of crRNA and donor DNA

The gene chosen to be knocked out is the *yfeX* gene coding for deferrochelatase. Firstly, a suitable target sequence in the gene was acquired using the available webtool CHOPCHOP [48]. By choosing the highest ranking target, the sequence was determined to: TGACGGTACAGAAAACCCGG (bottom strand 5' → 3'). The PAM site in this sequence is the three last bases CGG. Using Primer BLAST from NCBI, it could be decided that no off-targets were found. To be able to clone the crRNA into the pCRISPR plasmid it needs to be in the following form:

Forward:	AAACNNNNNNNNNNNNNNNNNNNG
Reverse:	AAAACNNNNNNNNNNNNNNNNNNN

So the crRNA was ordered as the following two oligos:

Forward:	AAACCCGGGTTTTCTGTACCGTCAG
Reverse:	AAAACTGACGGTACAGAAAACCCGG

The donor DNA (dDNA) was designed as a single stranded DNA with 40 nt homology arms on each side of an added stop codon (TGA):

5'-GAGCGTGACCTGAGCGGCTTTGTTGACGGTACAGAAAACC**TGA**CGGCGGGTG
AAGAGACGCGTCGCGAAGTGGCGGTTATCAA- 3'

3.4.2 Plasmid recovery from bacteria and plasmid purification

The plasmids pCRISPR (Addgene plasmid # 42875), pCNA (Addgene plasmid # 135191), pKM154 (Addgene plasmid # 13036) and pCas9 (Addgene plasmid # 42876) were delivered as agar stabs from Addgene. From the agar stabs, glycerol stocks were made according to the protocol in Appendix 8.8.2. The plasmids were then purified using the protocol and isolated according to the instructions from the kit *E.Z.N.A.® Plasmid DNA Mini Kit* I. The concentration and purity of the purified plasmids were measured using nanodrop and gel electrophoresis was used for verification. They were stored at -20℃ until further use.

3.4.3 Creation of competent cells

The competent cells were made using the protocol found in Appendix 8.8.4. Briefly, native cells were plated on agar plates grown overnight. LB media was inoculated with a single colony and allowed to grow overnight. The inoculum was transferred to new LB media and grown to a OD_{600} of 0.4-0.6 (about 3 h). The cultivation broth was then divided into two centrifugation flasks and centrifuged. The supernatant was discarded carefully and the pellets resuspended in $MgCl_2$ and incubated on ice. Another centrifugation at the same conditions were performed and the supernatant was discarded. The pellets were resuspended with $CaCl_2$

+ Glycerol solution, aliquoted and flash freezed. The competent cells were stored at -80℃ until further use.

3.4.4 Cloning crRNA into pCRISPR

The insertion of crRNA into pCRISPR was based on a protocol found at Addgene. The reworked detailed protocol can be found in Appendix 8.8.3. The protocol consists of six main steps with the first one being designing the crRNA which is described above. Step two was vector digestion where the restriction enzyme BsaI was mixed with purified pCRISPR, NEB buffer, BSA and Milli-Q water and then incubated. The digested pCRISPR was loaded on a 0.8% agarose gel and ran at 60V for about 2 h. The bands were visualized using the *Bio-Rad Gel Doc XR System w/ Universal Hood II* and cut out with a scalpel. The gel cutout was weighed and based on the weight, the amount of buffer needed for gel purification using the *NucleoSpin® Gel and PCR Clean-up* kit was calculated. The fourth step was phosphorylation of the crRNA oligos. The two oligos were mixed with T4 Ligase buffer, T4 Polynucleotide Kinase (PNK) and Milli-Q water and were incubated followed by a deactivation of the enzyme. The annealing at step five was initiated by addition of NaCl to the phosphorylated oligo pairs. The tube was then incubated for 5 min at 95℃ in a heat block and following the five minutes incubation the block was turned off and the tube was left to cool naturally for about 2.5 hours to room temperature. The last step was the ligation of crRNA into the digested pCRISPR. The annealed crRNA was mixed with the digested and gel purified pCRISPR together with T4 Ligase buffer, T4 Ligase and Milli-Q water. The mixture was incubated and then heat inactivated. The tube was held on ice and then the cells were transformed into competent *E. coli* BL21 (DE3).

3.4.5 Combination of plasmids using PIPE

To combine pCas9 and pCNA to create pCasRed and combine pKM154 with pCRISPR-crRNA to create pCRISPR-crRNA-sacB, Polymerase Incomplete Primer Extension (PIPE) was used. The detailed protocol can be found in Appendix 8.8.4. Briefly, the primers were diluted and the *PIPE Pfu Master Mix* was mixed. To the master mix, one sort of plasmid was added together with corresponding primers and this was done for all the plasmids. The tubes were then thermocycled at different conditions depending on if the plasmids were to undergo I-PIPE amplification or V-PIPE amplification. V-PIPE is for the recipient vector and I-PIPE for the insert. To combine the plasmids, the amplification products were mixed together. To validate, gel electrophoresis was run at 90V for 1 hour.

3.4.6 Vector transformation

The vector transformation was performed according to Appendix 8.8.4. Briefly, competent cells were mixed with the plasmid to be transformed. This was then incubated on ice and then heat shocked. The cells were recovered by adding LB media to each tube and incubating them. The transformed cells were then spreaded out on an agar plate with the correct antibiotics and incubated overnight.

4. Results

4.1 Pellet color analysis

A first verification that the production of Pgb was successful can be made by looking at the pellet colors after harvesting. Figure 7 below compares pellets from native *E. coli* BL21 (DE3) cells with transformed *E. coli* BL21 (DE3) containing the gene for Pgb, both the mutated genes and the wildtype gene. A big difference can be seen between the redness of the pellets when comparing the native to the mutants and the recombinant wildtype (rWt). There is no evident visual difference in redness between the mutants and rWt.

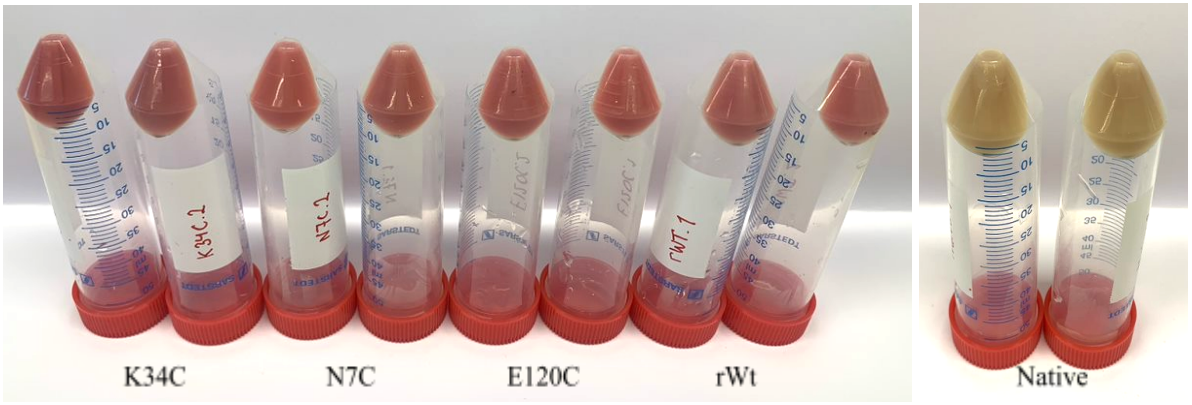

Figure 7. Shows the color of the pellets after harvesting. To the left, the mutants K34C, N7C and E120C as well as the recombinant wildtype cells (rWt) are displayed. To the right the native E. coli BL21 (DE3) is shown. All samples are shown as duplicates.

4.2 qPCR: Relative gene expression

4.2.1 The phytoglobin mutants

Table 2 shows the relative gene expression of *hemH* and *BvPgb1.2* for the three *E. coli* cultivations containing Pgb mutants K34C, N7C and E120C as well as the rWt. This is relative to native *E. coli* BL21 (DE3) not containing any Pgb gene. A standard deviation between the duplicates is also presented. N7C has a relative expression of 1.0 ± 0.3 for *hemH* meaning the gene is expressed around the same magnitude as the native. K34C and rWt have a relative expression of 0.9, slightly lower than the native. E120C has a higher expression of *hemH* compared to native, namely 1.8 with a standard deviation of 0.6. The relative expression of *BvPgb1.2* is 1000-folds higher for all samples compared to the native. E120C has the highest relative gene expression, however also the highest standard deviation.

Table 2. *hemH* and *BvPgb1.2* expression relative to native *E. coli* BL21 (DE3) cells.

Sample	Relative gene expression of *hemH*	Relative gene expression of *BvPgb1.2*
K34C	0.9 ± 0.3	6000 ± 1300
N7C	1.0 ± 0.3	2900 ± 1300
E120C	1.8 ± 0.6	9900 ± 7700
rWt	0.9 ± 0.04	7800 ± 900
Native	-	-

Table 3 shows the relative gene expression of *hemH* and *BvPgb1.2* for the three *E. coli* cultivations containing Pgb mutants as well as the native *E. coli* BL21 (DE3). This is relative to the samples with rWt Pgb. A standard deviation between the duplicates is also presented. The relative expression of *hemH* is around the same for K34C as for rWt. N7C as well as the native *E. coli* is slightly higher. The *hemH* expression is more than twice as high for E120C compared to the recombinant wildtype. For *BvPgb1.2*, the relative gene expression is much lower for native compared to rWt and also lower for K34C and N7C. E120C has a slightly higher expression relative to rWt.

Table 3. *hemH* and *BvPgb1.2* expression relative to recombinant wildtype *E. coli* BL21 (DE3) cells containing the non mutated *BvPgb1.2* gene.

Sample	Relative gene expression of *hemH*	Relative gene expression of *BvPgb1.2*
K34C	1.0 ± 0.4	0.8 ± 0.2
N7C	1.1 ± 0.4	0.4 ± 0.2
E120C	2.1 ± 0.7	1.3 ± 1.0
rWt	-	-
Native	1.2 ± 0.5	0.0002 ± 0.00009

4.2.2 The δ-ALA gradient

In table 4, the relative gene expression of *hemH* is shown for samples that were cultivated with different amounts of added 1M δ-ALA. The expression is relative to native *E. coli* BL21 (DE3) that was not supplemented with δ-ALA. Standard deviations between the samples are also displayed. The two lowest additions, 23 and 46 µl, have an expression lower and around the same as for the reference while the higher additions have a greater expression level. The sample with 92 µl has 40% higher expression and the 138 µl addition has more than double the expression level (however with a high standard deviation). The highest addition with 184 µl has a lower expression level than the 138 µl addition but still higher than the reference without δ-ALA.

Table 4. *hemH* expression of native *E. coli* BL21 (DE3) cells grown with different amounts of 1M δ-ALA relative to cells without addition of δ-ALA.

Sample	Relative gene expression of *hemH*
23ALA	0.7 ± 0.1
46ALA	1.0 ± 0.5
92ALA	1.4 ± 1.1
138ALA	2.2 ± 1.4
184ALA	1.7 ± 0.2

The relative gene expressions displayed in table 4 can also be seen in figure 8, which clearly shows that the 138 µl addition has the highest expression level but also the largest error bar.

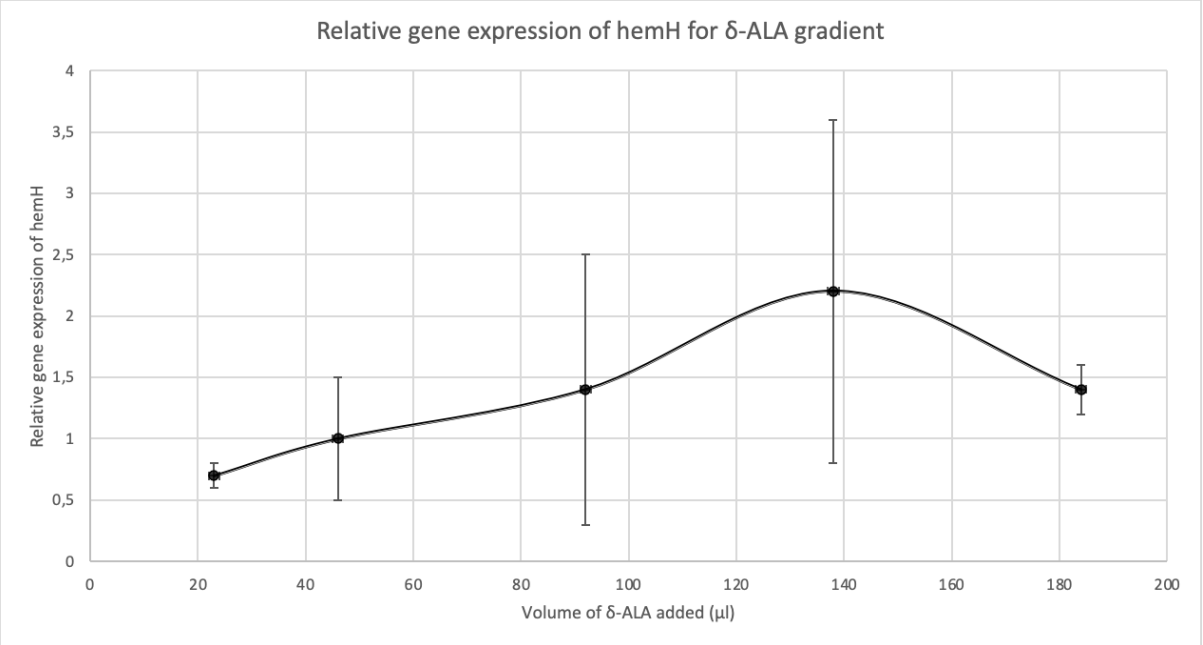

Figure 8. Gene expression of hemH for different additions of 1M δ-ALA relative to no addition of δ-ALA. The standard deviation can be seen as error bars.

4.3 QFF purification chromatograms

The four figures 9, 10, 11 and 12 below display the chromatograms for the QFF purifications using the ÄKTA system. This for the mutants N7C, E120C and the centrifuged corresponding mutants. On the y-axis the absorbance units (mAU) is shown and on the x-axis the run-volume (ml). The blue line shows the UV measurements at 412 nm which is the wavelength heme absorbs at. The purple line shows the UV measurements at 280 nm where proteins absorb and the orange line shows the conductivity. From zero to 200 ml in all chromatograms the column was washed and saturated with standard buffer resulting in washout of residual proteins that give rise to the high peak. Then the conductivity is increased

and the first elution step starts, at the second elution step the conductivity is increased more and here the elution peaks can be seen (between approximately 350-500 ml). After that the NaCl wash will elute the remaining proteins which can be seen as a high purple peak at about 550 ml.

Figure 9 below shows the chromatogram for the N7C purification. The elution peak at approximately 400 ml marked in blue shows the elution of the Pgb and the purple peak right next to it shows the elution of other proteins. The peaks are approximately the same height at about 700 mAU and partly overlapping.

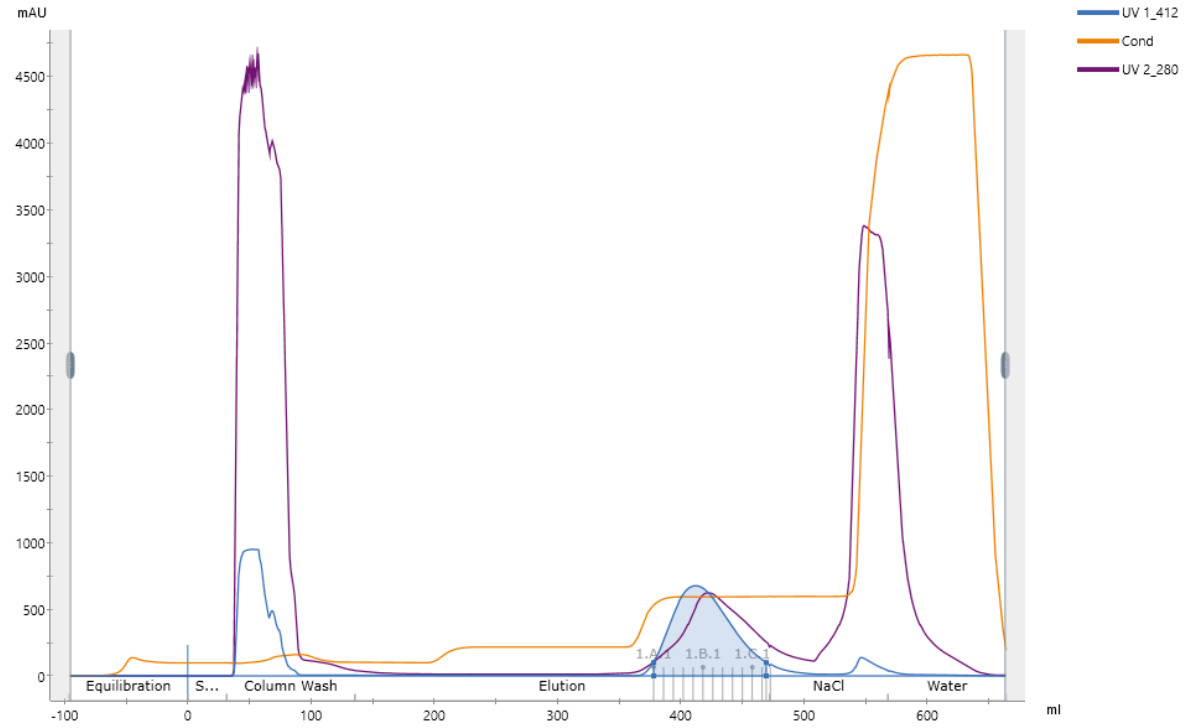

Figure 9. The ÄKTA chromatogram for purification of the N7C mutant. The y-axis shows the milli-absorbance units and the x-axis shows the run-volume in ml. On the x-axis can also the stages of the purification program be seen. The blue line shows UV_{412}, the purple line shows UV_{280} and the orange line shows the conductivity. The peak marked in blue shows the elution of the phytoglobin and the vertical lines the fractionation.

Figure 10 shows the chromatogram for the centrifuged N7C. The elution peaks look the same as for the non centrifuged sample in figure 9 above. The height is about 700 mAU and the heme-peak overlaps with the purple protein peak here as well.

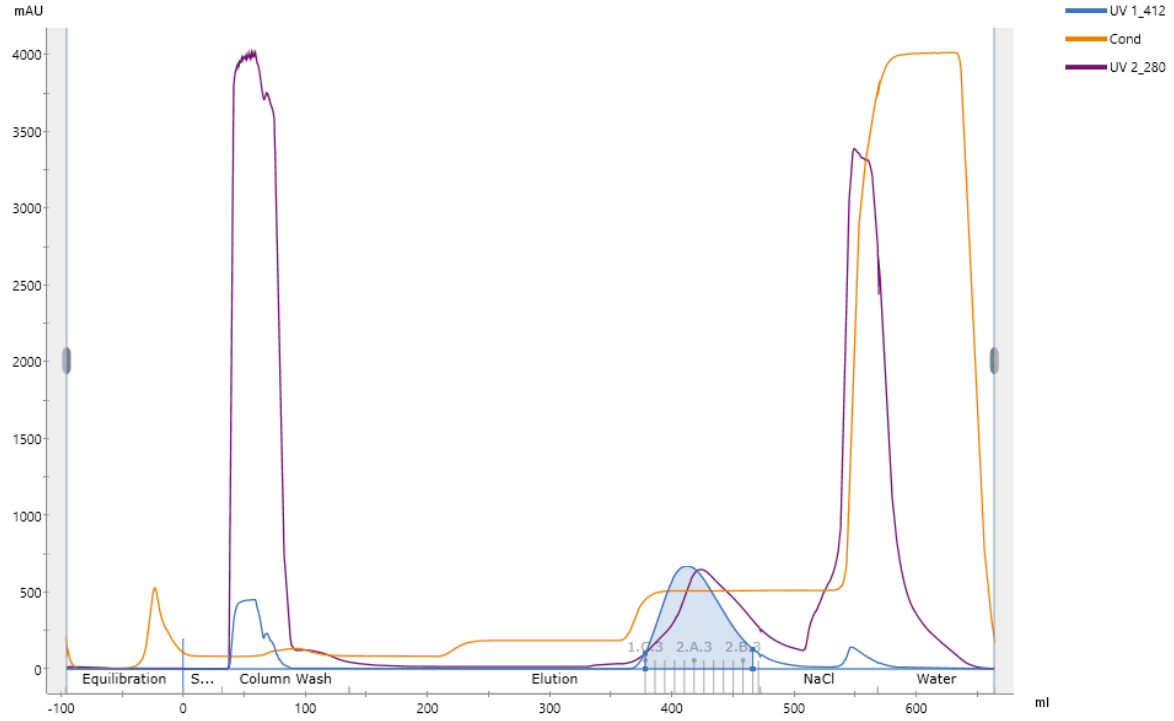

Figure 10. The ÄKTA chromatogram for purification of the centrifuged N7C mutant. The y-axis shows the milli-absorbance units and the x-axis shows the run-volume in ml. On the x-axis can also the stages of the purification program be seen. The blue line shows UV_{412}, the purple line shows UV_{280} and the orange line shows the conductivity. The peak marked in blue shows the elution of the phytoglobin and the vertical lines the fractionation.

The purification chromatogram for E120C can be seen in figure 11 below. The blue heme-peak is higher, approximately 1400 mAU, and more narrow compared to the corresponding peaks for N7C. The heme-peak is roughly double the height compared to the purple protein peak. The protein peak also consists of a second elution peak at about 440 ml.

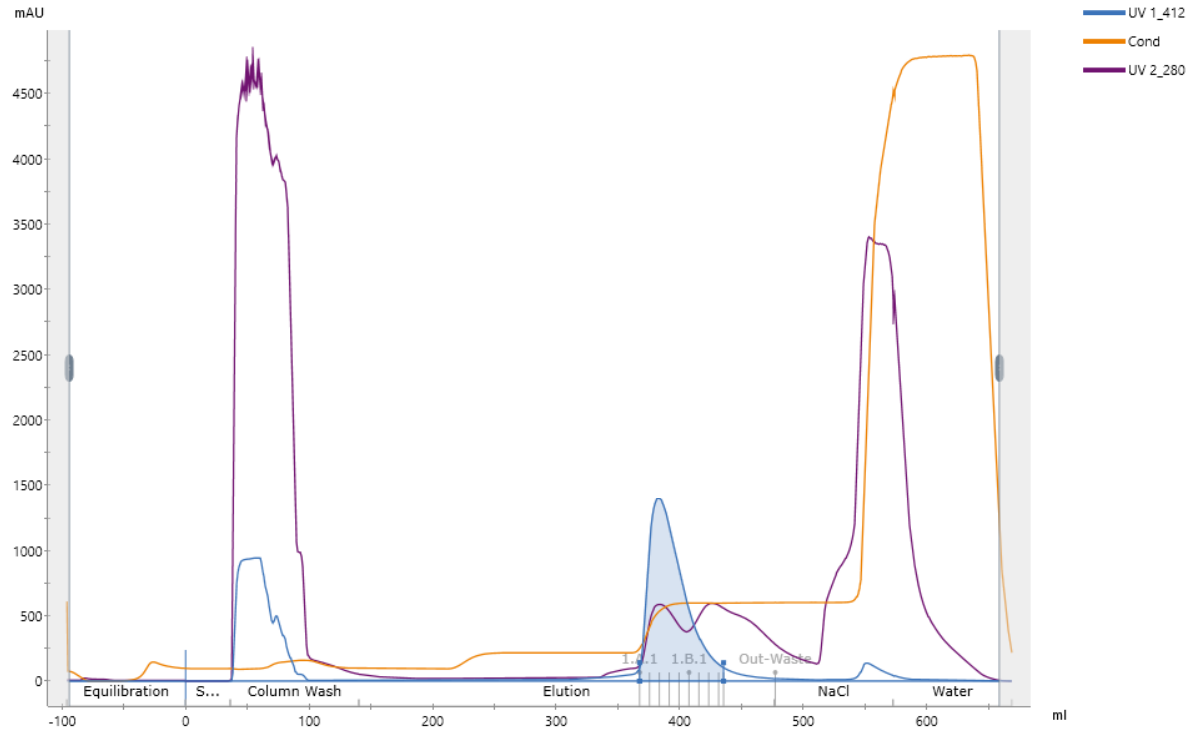

Figure 11. The ÄKTA chromatogram for purification of the E120C mutant. The y-axis shows the milli-absorbance units and the x-axis shows the run-volume in ml. On the x-axis can also the stages of the purification program be seen. The blue line shows UV_{412}, the purple line shows UV_{280} and the orange line shows the conductivity. The peak marked in blue shows the elution of the phytoglobin and the vertical lines the fractionation.

Figure 12 shows the chromatogram for the centrifuged E120C mutant and it shows the same trends as the non centrifuged sample.

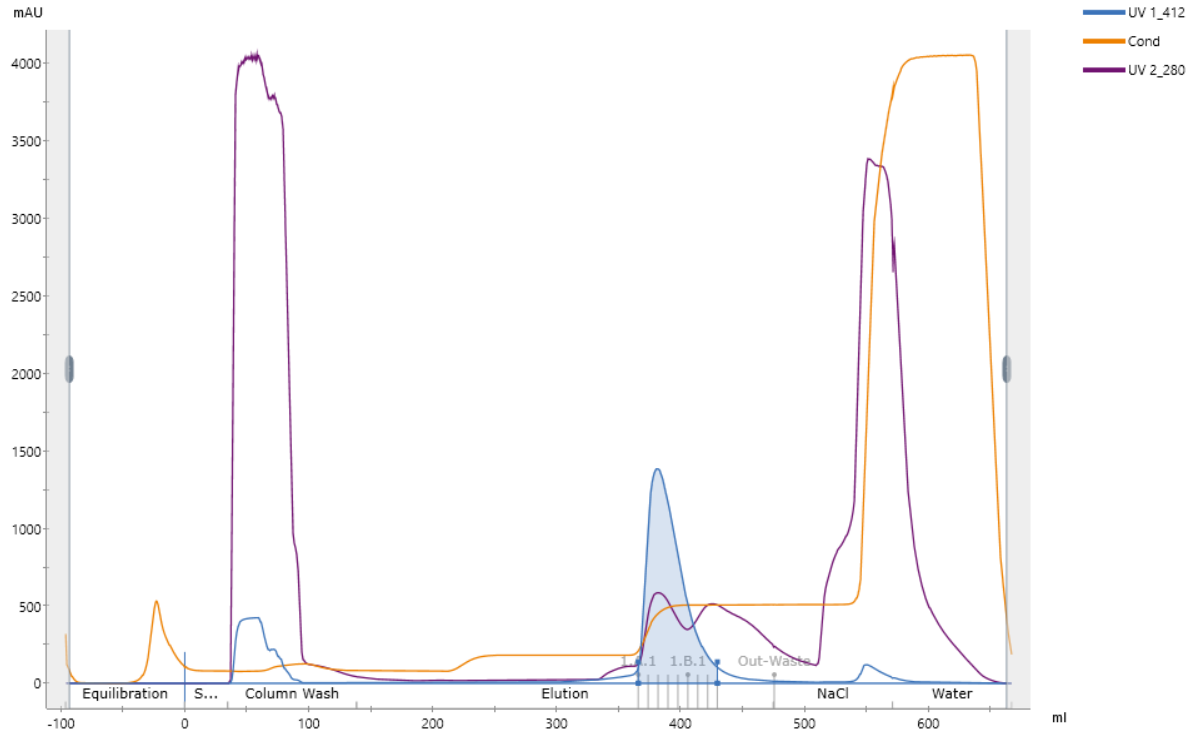

Figure 12. The ÄKTA chromatogram for purification of the centrifuged E120C mutant. The y-axis shows the milli-absorbance units and the x-axis shows the run-volume in ml. On the x-axis can also the stages of the purification program be seen. The blue line shows UV_{412}, the purple line shows UV_{280} and the orange line shows the conductivity. The peak marked in blue shows the elution of the phytoglobin and the vertical lines the fractionation.

4.4 SDS-PAGE

In figure 13 a visualization of the SDS-PAGE gel can be seen for the mutant N7C. Table 5 below can be used as a lane reference for the different lanes in figure 13.

Table 5. Lane references for SDS-PAGE in figure 13.

Lane	Sample
1	Ladder
2	After homogenization
3	After homogenized supernatant (centrifuged)
4	After homogenized pellet (centrifuged)
5	After dialysis
6	After dialysis (centrifuged)
7	After filtration
8	After filtration (centrifuged)
9	After QFF (row 1)
10	After QFF (row 2)
11	After QFF (row 3)
12	After QFF (row 1, centrifuged)
13	After QFF (row 2, centrifuged)
14	After QFF (row 3, centrifuged)

In figure 13, it can be seen that the lanes 2-8 have more bands compared to the lanes 9-14. In lane 4 the bands are less distinct and more smudged in comparison with the lanes 2, 3 and 5-8. The lanes 9-14 show a defined band at approximately 15-20 kDa, where lane 9-10 and 12-13 are more distinct in this band than the ones for lane 11 and 14. In the lanes 9-14 a few other faint bands can be seen.

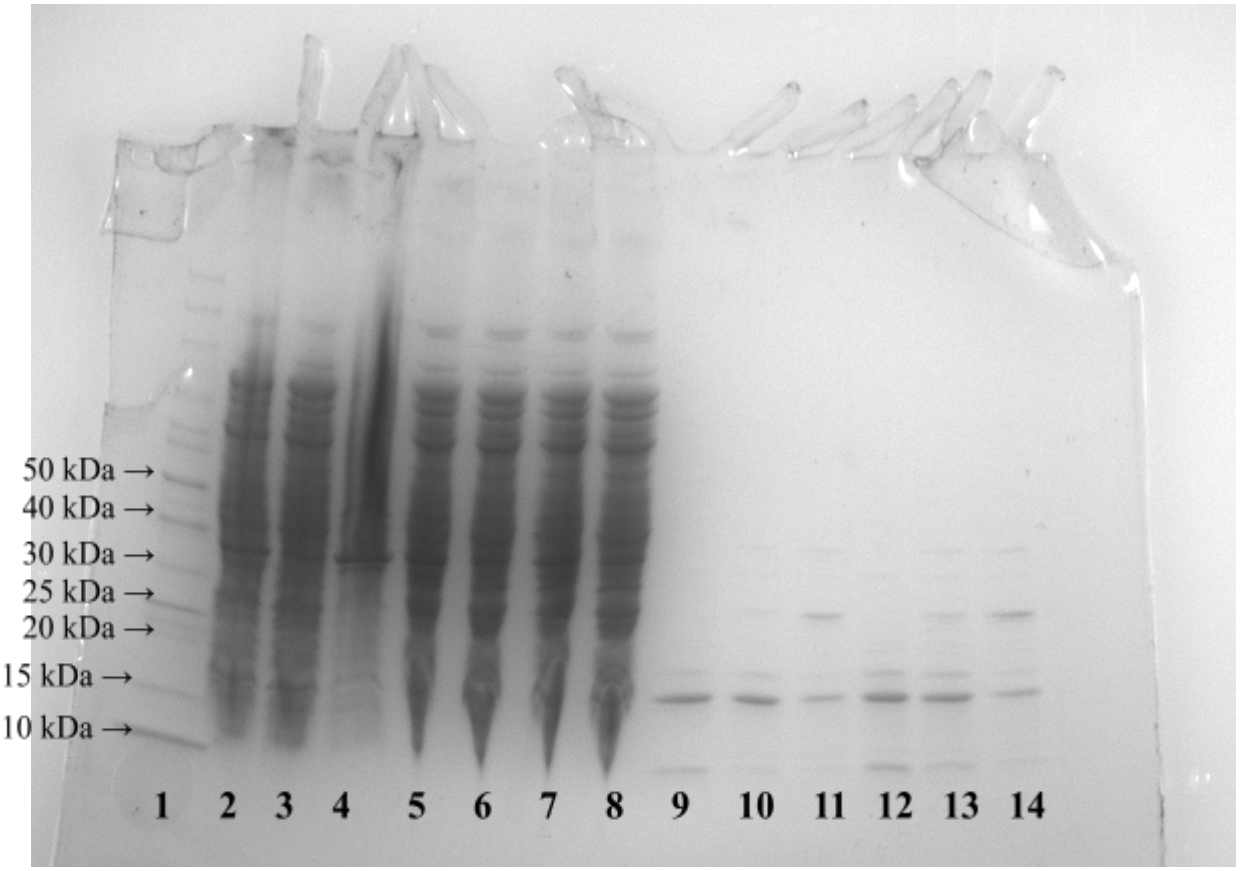

Figure 13. SDS-PAGE for different samples from mutant N7C, see table 5 for lane references.

In figure 14 a visualization of the SDS-PAGE gel can be seen for the mutant E120C. Table 6 below can be used as a lane reference for the different lanes in figure 14.

Table 6. Lane references for SDS-PAGE in figure 14.

Lane	Sample
1	Ladder
2	After homogenization
3	After homogenized supernatant (centrifuged)
4	After homogenized pellet (centrifuged)
5	After dialysis
6	After dialysis (centrifuged)
7	After filtration
8	After filtration (centrifuged)
9	After QFF (row 1)
10	After QFF (row 2)

11	After QFF (row 1, centrifuged)
12	After QFF (row 2, centrifuged)

In figure 14, it can be seen that the lanes 2-8 have more bands compared to the lanes 9-12. In lane 4 the bands are less distinct and more smudged in comparison with the lanes 2, 3 and 5-8. The lanes 9-12 show a defined band at approximately 15-20 kDa, where lane 9 and 10 are more distinct in this band than the ones for lane 11 and 12. In the lanes 9-12 a few other faint bands can be seen.

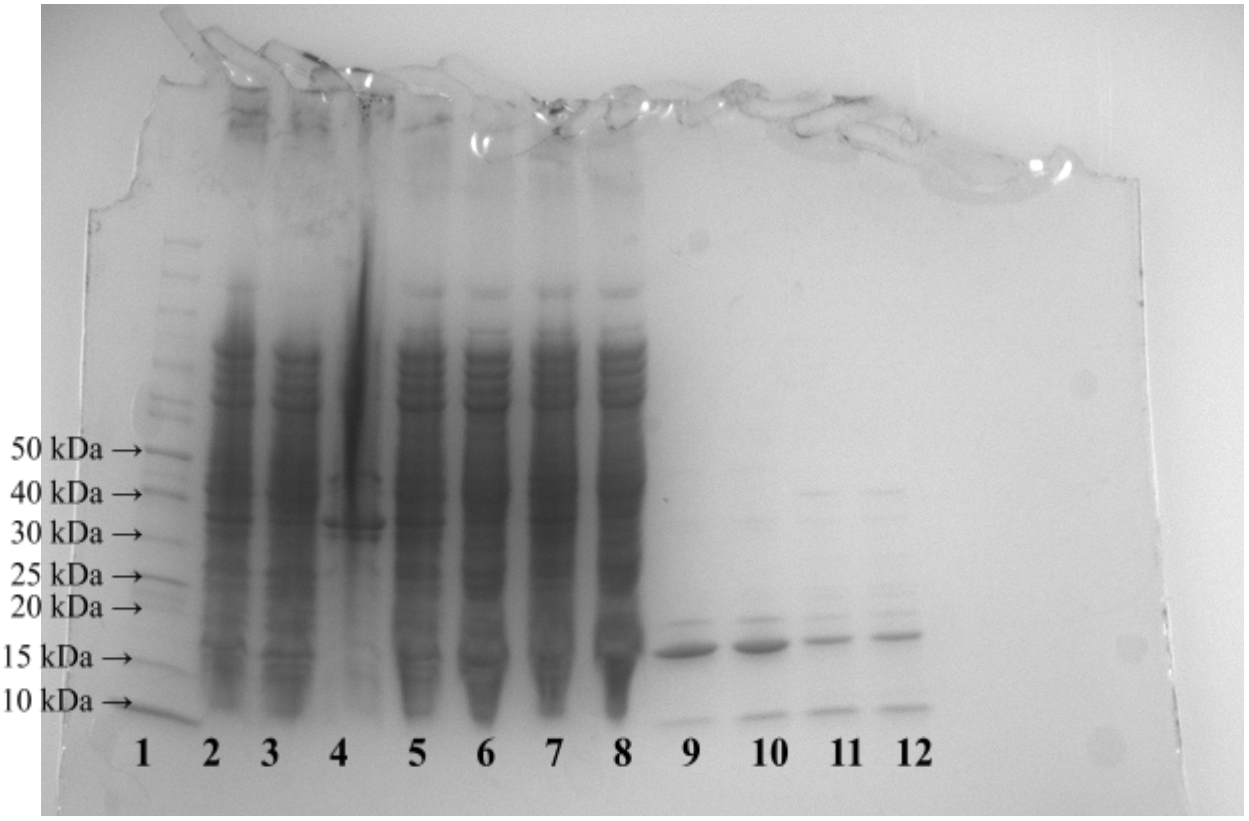

Figure 14. SDS-PAGE for different samples from mutant E120C, see table 6 for lane references.

4.5 Genome editing with CRISPR-Cas9

4.5.1 Cloning crRNA into pCRISPR

After the first experiment to try to insert crRNA into pCRISPR and transforming into native *E. coli* BL21 (DE3), no colonies were seen on the agar plates. Therefore the troubleshooting found in the section below was performed.

4.5.2 Troubleshooting

To be able to determine why no colonies were found on the plate after the first try to clone crRNA into pCRISPR, a troubleshooting table was created. It can be seen as table 7 below and is composed of a column with troubleshooting ID, a potential error source, a column with

a possible method that can be used to dismiss the error source, a timeframe of how long the method will take and lastly the likelihood of it being the error source. The likelihood of the error sources was considered due to time constraints. The most likely reasons were primarily tested.

Table 7. Shows the likely reasons that the cloning was unsuccessful and the reasonable methods to investigate the error sources. Troubleshooting ID is included to refer to the error sources.

Trouble-shooting ID	**Potential error source**	**Method to dismiss the error source**	**Time to perform the experiment**	**Likelihood of being the error source/Comments**
TS1	Failed to produce competent cells	Transform a positive control to verify competence	Two days	Likely since the lack of colonies is a direct consequence of the probable error source
TS2	Failed to transform the pCRISPR-crRNA plasmid into the competent cells due to a poorly constructed transformation protocol	Transform a positive control to verify a working transformation protocol	Two days	Likely since the transformation protocol used was made by rewriting other existing protocols. Something could have gone wrong in the rewriting
TS3	The incubation times and temperatures in our protocol are wrong	Try temperatures and times from other protocols that are known to work	Two days	Not very likely. The temperatures and times are not that different from other protocols that have been successful
TS4	The BsaI was not heat inactivated after vector digestion	Perform two parallel experiments: one with heat inactivation and one without	Two days	Not very likely, the gel purification removes the restriction enzyme instead. Also the manufacturer states that the BsaI does not have any star activity
TS5	Purification with the kit after gel cut-out failed	Redo the cloning protocol and test for DNA with NanoDrop	Two days	Not very likely, since kits in general are known to be successful

TS6	Failed to ligate the crRNA into the pCRISPR plasmid	Gel electrophoresis. One lane with only crRNA, one with digested pCRISPR and one with hopefully ligated pCRISPR-crRNA	One day (Two if we continue to transform)	Could be likely. If the ligation has not worked, the plasmid could still be linear and can be degraded by exonucleolytic enzymes [49]. This will result in no kanamycin resistance and the cells will not be able to grow on the agar plates
TS7	The agar plates are poor: dried out, too much antibiotics that killed the cells, wrong composition	Pour new agar plates and redo the cloning protocol	Two-three days	Some of the reasons could be likely. It could have been a bit dry since it was made approximately 4 weeks prior to plating
TS8	The pCRISPR plasmid was subjected to UV light for too long during gel purification	Redo the protocol and try to slice the gel faster	Two days	Not very likely, the UV light was not on for long periods of time

To troubleshoot both the competent cells (TS1) and the transformation protocol (TS2), a plasmid that was known to be circular, not cut and has a known selection marker was used. The plasmid was transformed using both our protocol and a protocol from the division that is known to work (see Appendix 8.8.5) into the produced competent cells. Figure 15 below shows the agar plates after the transformation. Plates A and B are the transformed cells using the control protocol from the division and the plates C and D are cells transformed with the protocol in Appendix 8.8.4. From the figure it can be seen that cells have grown on all four plates. Plates C and D were plated with 60 µl and have more colonies than the plates A and B that were plated with 50 and 20 µl respectively.

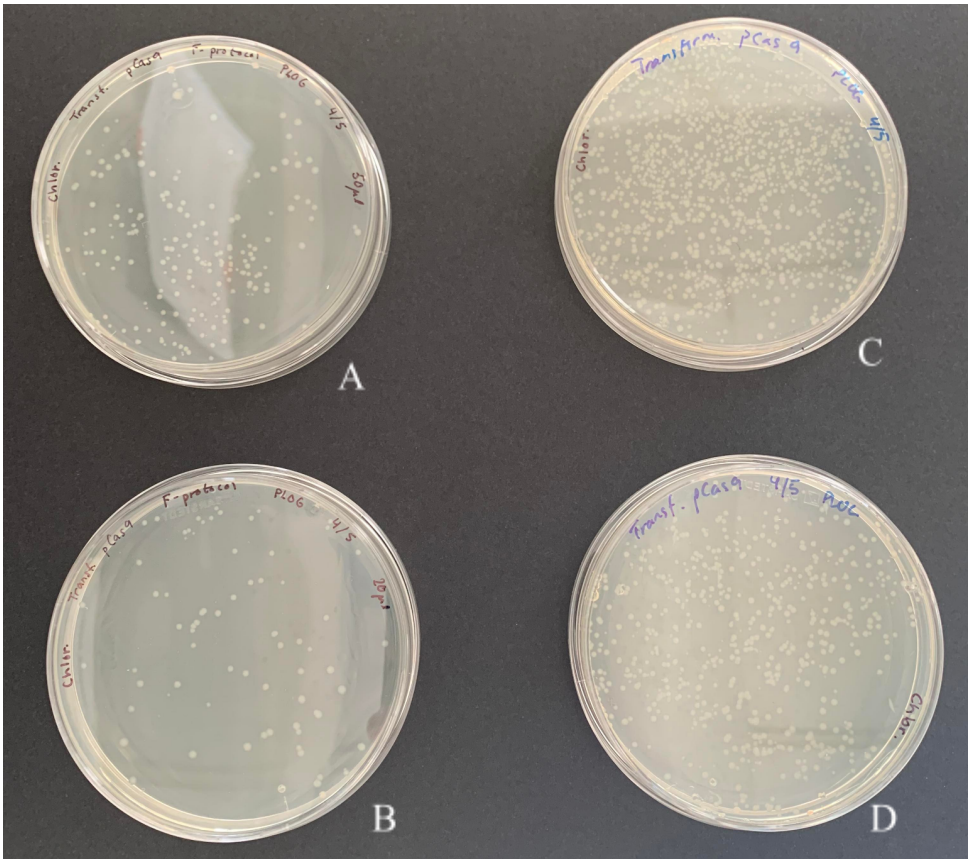

Figure 15. Shows the plates after investigation of TS1 and TS2. Plates A and B are made with the control protocol from the division. A was plated with 50 µl and B with 20 µl. Plates C and D have used the transformation protocol in this thesis and both were plated using 60 µl.

To investigate TS5, the protocol for cloning crRNA into pCRISPR was retested. The protocol was followed as described in Appendix 8.8.3 but Nanodrop was additionally used after the gel purification to ensure that the purification of the pCRIPSR plasmid was successful. The results of the Nanodrop can be seen in table 8, where the concentration of pCRISPR was about 6.5 ng/µl and the A260/A280 is slightly over 1.8 while the A260/A230 value is very low. Nanodrop was also implemented after annealing of the phosphorylated oligo pairs to ensure that the oligos were annealed. That can also be seen in table 8. These results show that the concentration is approximately 840 ng/µl for crRNA and also display high values of A260/A280 and A260/A230, with numbers around 3.8 and 2.6 respectively.

Table 8. The NanoDrop results for the purified pCRISPR and the crRNA. It displays both the concentration in ng/µl and the purity in the form of the ratios A260/A280 and A260/A230.

Sample	**Concentration (ng/µl)**	**A260/A280**	**A260/A230**
pCRISPR	6.46	1.88	0.0380
crRNA	838	3.81	2.59

To test TS6, a gel electrophoresis was implemented after the ligation step to verify if the ligation was successful or not. This was done with two different ligation compositions. The components in the first composition (Ligation 1) were mixed as the protocol in Appendix 8.8.3 describes. In the second composition (Ligation 2) the amount of the components were modified based on another protocol for the T4 DNA Ligase from New England BioLabs (see Appendix 8.8.6). According to this protocol the molar ratio should be 1:3 vector to DNA insert. To design this ratio to the specific insert and vector used, the NEBiocalculator was utilized [50]. In the ligation calculator the insert DNA length was stated as 35 bp, the vector DNA length was 2707 bp and vector DNA mass was 30 ng. From this the calculator gave the required insert DNA mass as 1.164 ng. Based on the concentration in table 8, 5 ul of pCRIPSR and 1.4 ul (1000x) of the annealed oligos (crRNA) were added. Thereby the Milli-Q water addition changed to 10.6 ul. This gave roughly a 1:3 molar ratio. The gel ran at 90 V for 1.5 hours. The results from the gel can be seen in figure 16.

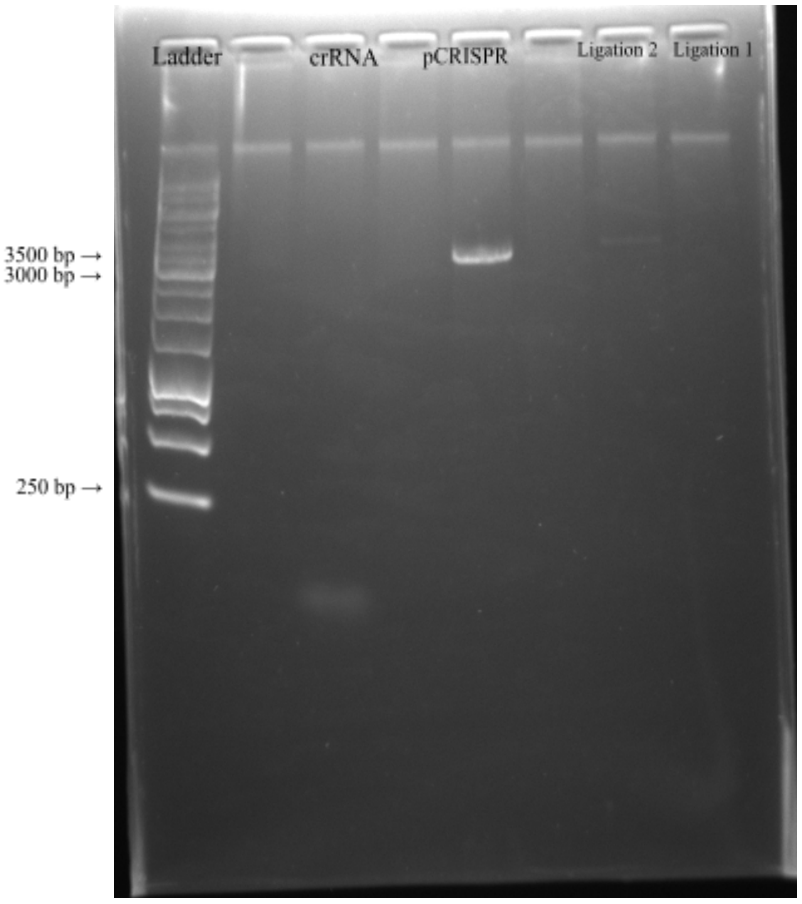

Figure 16. Gel showing ladder in lane 1, crRNA (10x) in lane 3, pCRISPR plasmid in lane 5, Ligation 2 in lane 7 and Ligation 1 in lane 8. The reference lengths for the ladder is based on figure A3, in Appendix 8.3.

A faint band can be seen in the lane for the crRNA in figure 16. This band has migrated past the ladder. The pCRISPR lane shows a more distinct band. At approximately the same height a very faint band for Ligation 2 can be seen but none for Ligation 1.

TS3-4 and TS7-8 were deemed not very likely and therefore not experimentally tested due to time limitations.

4.5.3 Combination of plasmids using PIPE

PIPE was used to try to combine the plasmids pCNA and pCas9. Figure 17 below shows the results after a gel electrophoresis of the plasmids. PIPE amplified pCNA (in duplicates) and PIPE amplified pCas9 (in duplicates) were loaded as well as the combined PCR products that should result in pCasRed. For both pCNA and pCas9 and for their duplicates, one line is visible in each lane. The gel was however run for too long which makes them blurry and somewhat distorted. The lanes where pCasRed was loaded have two bands per lane.

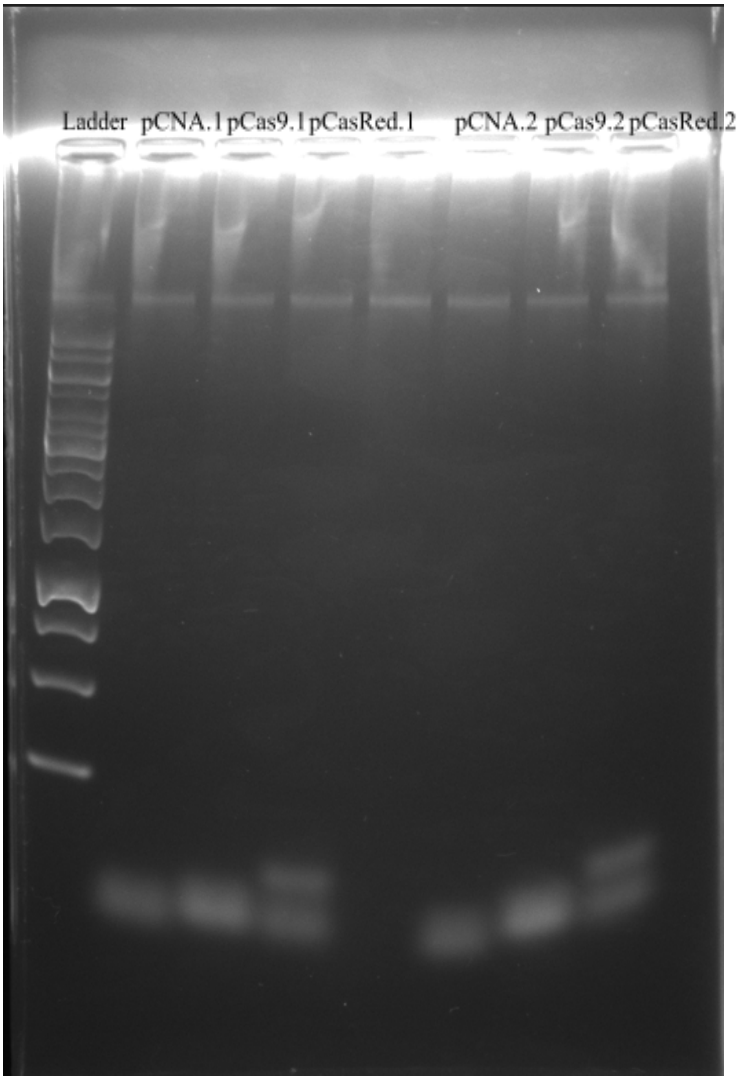

Figure 17. Shows the gel electrophoresis results after Polymerase Incomplete Primer Extension (PIPE) with plasmids pCNA (1 and 2) and pCas9 (1 and 2). The combined PCR products, pCasRed are also shown.

Since this thesis was time limited, there was no time to combine pCRISPR and pKM154. It was also decided that the incorporation of crRNA should be performed before the combination of pCRISPR with pKM154. Since the incorporation failed, we could not move on to create pCRISPR-crRNA-sacB using PIPE.

5. Discussion

5.1 Pellet color analysis

The pellet color analysis was made to verify successful production of Pgb in the cells. As seen in figure 7, the native cells are lacking the red color. This is reasonable since the native cells do not have the gene coding for the Pgb which gives the cells the red color. However, all the other cells have this color regardless if it is a mutant or rWt. There is no distinctive difference in the redness between the mutants and rWt, which means that these cells have produced approximately the same amount of Pgb. This is a first indication that the mutations does not have any effect on the amount of Pgb produced in comparison with rWt.

5.2 qPCR: Relative gene expression

5.2.1 The phytoglobin mutants

Starting with comparing the gene expression of *hemH* relative to native *E. coli* BL21 (DE3). *HemH* is the gene that codes for the enzyme that catalyzes the last reaction that gives us heme (see figure 4). We can see in table 2 that the mutants K34C and N7C as well as rWt have around the same expression for *hemH* as for the native cells. This even though it can be seen in the same table that the expression of *BvPgb1.2* is 1000-folds higher. This indicates that the synthesis of heme does not change even though the cells produce Pgbs. We hypothesized that the heme production would increase when the cells produce Pgb, and therefore also the expression of *hemH*. However, in the native cells there is still a heme synthesis where the heme goes to other hemoproteins such as catalase and cytochrome c, which we have stated in 2.1.1 *The Heme group*. It could explain why the gene expression is the same for native *E. coli* as for the cells that produce Pgbs. Maybe the heme gets redirected, from other hemoproteins to Pgb, but is still produced at the same rate. It would have been interesting in future work to look at the other hemoproteins to see if this theory is accurate. Another perspective on why the expression is the same could be that the enzyme coded by *hemH* is not that active during the harvesting where we measure the expression. The production of Pgbs may have decreased during harvesting resulting in the expression we see. To confirm this, a sample for qPCR analysis could have been taken during the production phase. Looking at the expression of E120C, it is higher than for native cells and the other mutants. This is most likely due to coincidence and that there are a lot of factors that are uncontrollable in regards to cultivating cells.

Looking at the expression of *BvPgb1.2* in table 2, for all samples the relative expression is 1000-folds higher compared to native cells. This is reasonable since native cells should not produce Pgbs, hence not express the gene *BvPgb1.2.* The magnitude of the expression for the other samples can be seen as successful since it is a big difference compared to native. The different mutants and the rWt have a substantial difference between expression levels. This is most likely not due to the different mutants but rather due to cultivation parameters, and

thereby depending on if the cultivations were more or less successful. If more time was available, we would have wanted to reproduce the cultivations several times and therefore be able to draw a more accurate conclusion. One sample that distinguishes from the rest is E120C with an expression level of 9900 ± 7700. The duplicates are very different (17600 and 2200) and it is hard to say why. Like we have theorized before, it probably has to do with how well we succeeded with the cultivation. Here however they were cultivated at the same time. There could have been different amounts of cells in the inoculum and this would have resulted in that more Pgb was produced in the sample with more cells. A theory on why the expression is higher than the other mutants and rWt could be that the mutation E120C is placed somewhere in the protein that has affected the stability of the Pgb. Maybe it enables disulfide bridges between the monomers which affects the stability leading to an abundant protein production. To investigate the 3D structure of this mutation would therefore be interesting.

When comparing the expression of *hemH* to the rWt cells in table 3, it can be seen that K34C, N7C and native have around the same expression level. This could possibly be explained by the same fact as before, that the enzyme is not as active during harvesting. Then it would make sense that native cells have the same expression level as the others. Once again, E120C has a higher expression level and the reasons have been explained before.

The expression of *BvPgb1.2* in table 3 shows that K34C has about the same expression level as rWt, N7C has a lower expression and E120C a higher. This is probably for the same reasons as before, that the cultivations have been more or less successful due to coincidence. The native ones have a much lower expression and this is reasonable since they should not produce any Pgb and hence not express *BvPgb1.2*.

5.2.2 The δ-ALA gradient

In this part of the experiment, five different amounts of δ-ALA were compared to no addition at all. Looking at table 4 and figure 8, it can be seen that the addition of 138 µl δ-ALA gives the highest relative gene expression of *hemH* with an expression of 2.2 ± 1.4. The expression of *hemH* increased with higher additions of δ-ALA until 138 µl, then it decreased again for 184 µl. This indicates that 138 µl is the optimal amount that should be supplemented. It is however hard to conclude since the standard deviation between the duplicates is remarkably high. The reason that the results are so different between the duplicates is the same reason as the section above, that it is a coincidence. Since we are dealing with living bacteria, it is hard to control every aspect of the cultivation. If we however proceed with that the mean value is correct, the shape of the curve in figure 8 could indicate that the production of δ-ALA is a bottleneck in the heme biosynthesis. When we have low additions of δ-ALA the expression of *hemH* is reduced which strengthens the theory that δ-ALA is a bottleneck in the metabolic pathways that makes heme be produced at a lower rate. With the highest addition of δ-ALA the expression decreases which could possibly be explained by clogged porins. Since δ-ALA is transported into the cells first through porins and then by a transport system [36, 37], too much δ-ALA might disturb this. Another possible reason could also be that too much δ-ALA

leads to an amount of free heme that causes heme-mediated toxicity in the cells [10]. If that's the case the cell needs to regulate the heme biosynthesis and might downregulate the enzyme coded for by *hemH*.

An unexpected result from the δ-ALA gradient is that the addition of 23 µl gives a lower gene expression than no addition at all. This feels somewhat wrong since it contradicts why supplementation of δ-ALA is needed in the first place. The reasonable result would be that it gives at least the same expression. Something might have gone wrong with these samples and that's why we see this kind of result.

5.3 QFF purification chromatograms

In figure 9-12, the chromatograms from the QFF purification are presented. The protocol in the method was not optimized when purifying K34C which resulted in an unsuccessful purification. Due to time limitations, we did not redo the experiment. Therefore, from now on we will only focus on N7C and E120C.

The figures from this section strengthen the results from the qPCR analysis. It can be seen in figure 11 and 12, that displays E120C, that the peak reaches a much higher absorbance when the wavelength is 412 compared to N7C in figure 9 and 10. This is in line with the qPCR results with E120C having a higher mean relative gene expression compared to N7C. Since heme absorbs at 412, it should be proportional to the amount of Pgb produced. It can also be seen that the heme-peak for E120C is more narrow compared to N7C which should result in a more concentrated sample. This is strengthened later on with the SDS-PAGE analysis. The peak that represents the other proteins is more overlapped with the N7C sample compared to E120C. This could result in a more contaminated sample and another purification step would be recommended.

Comparing the centrifuged sample to the non centrifuged for N7C, there is no visible difference in the chromatograms. The heme-peak reached the same altitude and is the same width. The same can be said for E120C. This indicates that the centrifugation does not affect the amount of Pgb.

5.4 SDS-PAGE

5.4.1 Mutant N7C

It can be seen a big difference in the amount of bands in the SDS-PAGE gel for N7C in figure 13. Most prominent is the difference between the samples before the QFF purification (lane 2-8) and the samples after the QFF (lane 9-14). This implies that the purification has been successful since many irrelevant proteins have been removed. In lane 9-14 it can be seen a band that is more distinct than the others, at approximately 15-20 kDa. Our Pgb is dimeric and has a weight of 38.4 kDa, one monomer should therefore be about 19 kDa. The gel is a bit skewed but because of its position and its intensity this should be the band for the target

protein. The reason that the protein go from dimer to monomer when running the SDS-PAGE is because the SDS makes the protein dissociate into its subunits. In lane 9-14 there are also other bands, however faint ones. These are probably other proteins that also are negatively charged at pH 8.5 and thereby attached to the column simultaneously as the Pgb. It is very difficult to identify these proteins since the number of proteins expressed by *E. coli* BL21 (DE3) is about 4156 [51].

It can also be seen in figure 13, that the band for the Pgb in lane 11 is weaker in intensity than for the lanes 9 and 10. The weaker band is reasonable since row 3 (lane 11) was at the end of the fractionation of the protein in the ÄKTA run. This means that most of the protein already came out in row 1 and 2 (lane 9 and 10) and not as much is left for row 3 (lane 11). Another band at approximately 20-25 kDa can be seen quite prominently in lane 11. However, this band is not as strong in lane 9 and 10. This is because the protein peak overlaps more with the heme peak in the later fractionations (see figure 9). Since we wanted to make sure that most of the Pgb was eluted, a few other proteins were eluted as well. The same trend as described for lane 9-11 can be seen for the centrifuged sample in lane 12-14.

In the gel in figure 13, lane 2 and 3 represent after homogenization, lane 5 and 6 after dialysis and lane 7 and 8 after filtration. More or less there are around the same number of bands for these lanes. This means that these purification steps did not remove many proteins. The dialysis bags that were used have a cutoff at 6-8 kDa which means that proteins larger than that stay in the bag, with the Pgb included. The probable removal of the smaller proteins is not visualized within the ladders range in the gel. Dialysis bags with a cutoff near 15-20 kDa could have been used to get a more effective purification. The filtration did not remove any proteins, but instead larger particles such as cell debris since the filtration size was 0.45-0.65 μm. This is not something that can be seen in the SDS-PAGE gel. In general, there is no significant difference between the lanes representing these purification steps, which is good since it means no protein in the range 10-200 kDa have gone missing.

Another aspect that was examined during the purification of the protein was if there was any difference in the results with the centrifuged samples compared to the non-centrifuged. The centrifuged samples can be seen in lanes 3, 4, 6, 8 and 12-14 in figure 13. In the same figure, the non-centrifuged samples can be seen in lane 2, 5, 7 and 9-11. The comparison between the centrifuged and non-centrifuged samples before the QFF purification show no difference in numbers of bands or the intensity. This means that the centrifugation did not affect the purification of the proteins. However during the filtration the centrifuged samples were easier to push through the filter than the non-centrifuged samples were. This indicates that the centrifugation removed bigger particles such as cell debris from the samples. Furthermore, the centrifugation is shown to make no difference for the final purified sample, that can be seen if lane 12-14 compared to lane 9-11 since these bands are quite similar. Anyhow, the centrifugation can be considered favorable for the column since it is not clogged or damaged by cell debris.

The centrifuged sample before the QFF purification resulted in a supernatant and pellet, which both were loaded in the SDS-PAGE gel. This can be seen in lane 3 and 4 in figure 13. The pellet sample in lane 4 is very smudged and difficult to interpret compared to lane 3. This is probably because there is cell debris and bigger particles that made it harder for the proteins to migrate. It is thereby difficult to interpret if the pellet sample contains numerous or few proteins. Lane 3 on the other hand, indicates more proteins. In both lanes the probable band for the Pgb can be seen, though a bit weaker for lane 4. This is good since it shows that the target protein was not completely lost during centrifugation. It is a cytosolic protein so it is reasonable that it should be in the supernatant for most part.

5.4.2 Mutant E120C

The SDS-PAGE gel for mutant E120C in comparison with the one for N7C, is very similar as can be seen in figure 14 compared to figure 13. It can be seen that the QFF purification has been successful when comparing the lanes 2-8 to 9-12 by the bands decreasing in number after the run. Just like the N7C mutant, it is believed that the band for the Pgb is the strong one at approximately 15-20 kDa, clearly shown in lane 9 and 10. In lanes 9-12 there are other faint bands that probably represent other proteins negatively charged at pH 8.5. Similarly to the N7C, it is very difficult to identify these proteins.

It can also be seen in figure 14 that the Pgb band has the same intensity for lane 9 and 10 as well as for 11 and 12, respectively. This indicates that there was no difference in how much Pgb there was in row 1 compared to row 2 during the fractionation. It can also be seen in figure 11 and 12 that the blue marked area is narrower than in figure 9 and 10. This means that less volume was needed to fractionate all Pgb and the E120C samples got more concentrated with Pgb than N7C.

The difference between the purification steps are just like N7C, not significant as can be seen in figure 14. Lane 2 and 3 represent after homogenization, lane 5 and 6 after dialysis and lane 7 and 8 after filtration. There is no direct difference between the number of bands which is due to the reasons described for N7C.

For the mutant E120C, the aspect centrifugation was also tested. When comparing the centrifuged samples with the non-centrifuged samples before the ÄKTA run, there is no significant difference in the number of bands. This can be seen in figure 14 when comparing lane 2, 5 and 7 with lane 3, 6 and 8. However, after the QFF purification the conclusion is different. The centrifuged samples in lane 11 and 12 have a lower intensity in the Pgb band than the lane 9 and 10. This indicates that Pgb has been lost during the centrifugation. Once again, it is probably good for the column to centrifuge the sample beforehand but in this case not favorable for the Pgb yield. It can also be seen that the centrifuged samples have several faint bands that the non-centrifuged samples do not have. This feels a bit odd since there should be fewer proteins in the centrifuged samples than in the non-centrifuged since some protein probably were left in the discarded pellet. One theory why this is the case could be that all these proteins could penetrate the filter during filtration because there was no cell

debris that clogged the filter. However, clogging in the filter could be the case for the non-centrifuged samples, which makes proteins get stuck and thereby there are fewer bands shown on the gel for those samples.

The centrifuged sample before the QFF purification resulted in a supernatant and pellet, which both were loaded in the SDS-PAGE gel. This can be seen in lane 3 and 4 in figure 14. Just as for N7C, lane 4 was smudged and difficult to interpret due to the same reasons. Lane 3 shows more distinct bands which indicate more proteins, among them the Pgb. It is very difficult to determine if the Pgb band is there or not in lane 4, compared to figure 13 where it was more clear. The reason for this can be that lane 4 is more smudged in figure 14 than in figure 13 and thereby the bands can not be as reliable.

5.5 Genome editing with CRISPR-Cas9

5.5.1 Troubleshooting

When troubleshooting TS1 and TS2, one of three different outcomes were expected. The first outcome could have been that cells did not grow on either of the plates. That would have meant that the competent cells were dead or that they were not competent at all. The second outcome could have been that it only grew on the plates where the cells plated were transformed using the control protocol from the division. In that case, it would have been the protocol in this thesis that was at fault for the unsuccessful experiment. The third outcome was that it grew on all the plates. The outcome for the troubleshooting experiment was the third one: it grew on all the plates which can be seen in figure 15. This means that it can be ruled out that there was something wrong with the competent cells. It can also be dismissed that the transformation protocol was at fault. Therefore the error source probably lies within the cloning protocol, most probably the ligation. Another possibility that caused there to not be any colonies on the first try could be human error, that we did something wrong without noticing it.

It can be concluded that the gel purification (TS5) of the pCRISPR plasmid was successful and not the error source of the failed experiment. This since the concentration was 6.46 ng/µl, which indicates that there is pCRISPR in the sample. This can be seen in table 8. Before the plasmid purification, the pCRISPR plasmid band was cut out from the gel, which should mean that sample more or less should only consist of the plasmid. The sample can be considered pure in relation to proteins since A260/A280 > 1.8 but not in relation to carbohydrates and phenols since A260/A230 is only 0.038 and a value over 2 is considered pure. The low purity could mean that there are compounds in the sample that will interfere later on. The results however clarified that there was pCRIPSR in our sample, but it is hard to know if it was enough. The Nanodrop measurement was also done to investigate if the crRNA oligos had successfully annealed. The Nanodrop was set to measure double stranded DNA to be able to determine this. There should be a high concentration of double stranded DNA if the oligos had annealed, which the Nanodrop showed since the concentration of the crRNA was 838 ng/µl, as can be seen in table 8. The purity in relation to proteins,

carbohydrates and phenols is very good and these components are considered to not be a problem in this case. With this it can be concluded that the purification of the plasmid nor the anneling between the crRNA oligos was the error source of the failed experiment.

The results from the investigation on TS6 can be seen in figure 16. The reason that crRNA has migrated past the ladder is probably because the ladder was not customized for small fragments like the crRNA. The actual size of the crRNA is 25 bp. The band of the crRNA is also smudged and difficult to interpret which can be due to that the conditions of the gel run were not customized for the crRNA. It was customized to the pCRIPSR, such as the run time, the voltage and the agarose content. It should be reasonable if the intensity of the crRNA band would be higher than the pCRIPSR band since the concentration of the crRNA is higher than for pCRISPR. When loading the gel the concentration of crRNA was about 84 ng/µl due to a 10x dilution but for pCRISPR the concentration was about 6 ng/µl, as can be seen in table 8. However, it is difficult to determine the intensity of the band since the crRNA band is diffuse. The position of pCRISPR is reasonable since theoretically the plasmid is 2707 bp and in the gel it seems that it is about 3000 bp. It is quite difficult to read the exact size since the gel seems to be a bit skewed. The band for ligation 2 is very faint and is about 3500 bp according to the ladder. The reason that the size is a bit bigger than pCRIPSR can be due to that the ligation has succeeded and the ligated plasmid has become a bit heavier. The reason why the band can be seen for ligation 2 but not for ligation 1 can be because the concentration of the crRNA and the plasmid was optimized for ligation 2. However, it is most likely that ligation 2 only shows pCRISPR itself since the *E. coli* plates did not have bacterial growth when transformed after being ligated. We have ruled out that the transformation protocol, the plasmid purification and the annealing of the crRNA oligos is the error source. Therefore, this result points to that the previous theory on ligation being the error source is true. This results in a still linear plasmid which is harder to transform and if it transforms, it can be degraded by exonucleolytic enzymes [49].

After the troubleshooting experiment regarding the failed ligation was made it was noticed that the insert DNA length stated in the NEBiocalculator was wrong. The length 35 bp was based on the insert template but the actual length was 25 bp. This made the inserted DNA mass higher than it should have been and gave a molar ratio about 1:4 instead of 1:3 (vector to DNA insert). This is something that probably had an effect on the results but not to such degree that the ligation failed because of this. Actually, the concentration of crRNA was higher than expected which should have given pCRISPR higher probability of encountering crRNA fragments to ligate.

The two error sources TS3 and TS4 in table 7 were deemed unlikely and were therefore not experimentally tested. As said in the table, the times and temperatures are not very different from other successful protocols and should therefore not be the source of error. That the enzyme BsaI was not heat activated should not be at fault since it is stated to not have any star activity meaning it should only cleave at the intended place and nowhere else. Furthermore, since the gel purification was made directly after, it would have removed the BsaI.

Another possible error source is that there was something wrong with the agar plates (TS7). This should however not be the case since the transformation experiment described under the section 4.5.2 *Troubleshooting*, showed growth. The transformation experiment did use plates with chloramphenicol instead of kanamycin but the plates were made at the same time and by the same person so they should be equally good.

The last possible error source in table 7 (TS8) is that the DNA was exposed to UV light for too long which might have resulted in degradation. The first time we tried to clone the crRNA into pCRISPR was the first time we tried gel purification. Therefore we might have left the gel on for too long. However the next time that the protocol was tried, we tried to limit the time the DNA was exposed to UV. That did not change the end result (that we got no growth) so DNA degradation was probably not at fault. It can also be seen from the NanoDrop experiments in table 8 that DNA was found after the purification. Therefore it is more likely, as we have stated above, that the ligation is the error source.

5.5.2 Combination of plasmids using PIPE

When trying PIPE to combine pCNA with pCas9 to create pCasRed, gel electrophoresis was used to validate a successful combination. In figure 17 the result of this gel electrophoresis can be found. It can be seen that the bands are very blurry and skewed, this is most probably due to a too long run time and a too high voltage. Both pCNA.1, pCNA.2, pCas9.1 and pCas9.2 have one band each meaning that only the plasmid is present. For pCasRed.1 and pCasRed.2 there are two distinct bands seen. The samples that were loaded in those lanes are the combined PCR products after PIPE meaning the plasmid should in theory be combined to one. That we got two bands could possibly mean a number of things. One of the reasons could be that the combination failed and the plasmids did not combine so the two bands seen are the two separate plasmids. It could also be that the combination worked but that pCasRed are in two different forms, supercoiled and circular. That would give rise to two bands [52]. Another possibility is that the combination worked but the two plasmids were mixed in different concentrations. That could give rise to one band being the combined plasmid and the other being residual plasmid. This would however probably give a larger distance between the bands. Since the early stages in CRISPR-Cas9 did not work out, we did not move forward with this investigation. If we would have continued, sending the plasmid for sequencing would have been the next step since gel electrophoresis did not give any conclusive results.

5.6 Future considerations

From the results in this master thesis, there are several things that would be interesting to follow up on. Firstly, the results obtained for the mutant E120C show really promising results, both with the qPCR, QFF purification and SDS-PAGE. It however has the highest standard deviation so to redo the cultivation with this mutant more times to more accurately determine the values should be done. It would also be interesting to investigate this mutation

in more detail. For example, as mentioned before, to determine its 3D structure and study the effect of the mutation. This to see if it results in some kind of stability, such as disulfide bridges, to the protein.

Since the results from the δ-ALA gradient shows that the *hemH* expression decreases when the addition of δ-ALA exceeds 138 µl in our gradient, this can be a sign of heme toxicity. This would be interesting to examine further to ensure total understanding of the impact of δ-ALA. This could be done with a long term cultivation where different amounts of δ-ALA are added to deeply understand the growth trend maintained from the cells and when or if cell death occurs.

A problem with the δ-ALA gradient is that we are looking at the gene *hemH* in the qPCR. By only looking at that gene we might not get a very accurate result on how much of the supplemented δ-ALA goes to the heme biosynthesis. For a more accurate result, it would have been interesting to investigate the gene *hemB* (see figure 4) that leads δ-ALA into the metabolic pathway. Additionally it would have been an interesting approach to research if δ-ALA can go into another metabolic pathway that is not the heme biosynthesis. By doing this, comparing the expression of a gene in that pathway to *hemB* might give insight into the distribution of δ-ALA in *E. coli*.

One problem that was noticed in the SDS-PAGE for both N7C and E120C was that there were several other proteins in the samples after the ÄKTA chromatography, which means that the run was not as accurate as wanted. Thereby it would be intriguing to target these proteins somehow in an additional purification step, such as using an extra column in an additional ÄKTA run. This would give a more pure final sample and a high yield.

The genome editing with CRISPR-Cas9 was in this thesis designed and initiated, but since problems arose during the experiments this part could not be finalized. This work would be intriguing to continue with to understand the effect of the *yfex* gene knockout. Additionally, a continuation to the metabolic engineering we tried to perform in this thesis could be to overexpress the enzyme GAPDH. This enzyme has been, as previously mentioned in section 2.4.1 *Strategy of metabolic engineering*, seen to work as a heme chaperone [40, 41]. We have not found that a combination of the knockout of *yfeX* and an overexpression of GAPDH has been done before. Therefore this could be something for the future to consider.

6. Conclusion

To conclude this thesis, the qPCR experiments show a successful upregulation of the gene *BvPgb1.2* in rWt and the mutants K34C, N7C and E120C compared to native *E. coli* BL21 (DE3). With relative gene expressions 1000 times higher than native cells, they ought to be investigated further. Especially E120C which gave the most promising results and even expressed the gene *BvPgb1.2* higher than the rWt. The expression of *hemH* did not vary noticeably between the samples and might not be the ideal gene to study according to our results. The results from the δ-ALA gradient indicate an optimal supplementation at 138 µl with lower additions leading the production to be a bottleneck. Higher additions might result in difficulties with transportation into the cell. However, other routes for δ-ALA and genes closer to its entry to heme biosynthesis is something to study in the future. Additionally, purifying the Pgbs with ÄKTA chromatography seems to work effectively as can be seen from the SDS-PAGE results. There is a considerable difference between the amounts of bands before and after QFF purification. Still, there are some faint bands left so an additional purification step should be considered. Furthermore, more targeted purification methods before the ÄKTA run might be an advantage in getting a higher purity. In regards to adding a centrifugation step before the QFF purification, it showed no visual difference in the chromatogram or the SDS-PAGE. However, it possibly helped the column not be subjected to unnecessary debris that might damage it.

In regards to the metabolic engineering using CRISPR-Cas9, a full plan was designed. It could however not be finalized due to obstacles that were encountered during the experiments. Troubleshooting the experiment resulted in the theory that the ligation of crRNA into pCRISPR is the error source. Overcoming this challenge should hopefully result in a strain that provides a higher yield.

Overall, the results point to the fact that the production and purification of Pgbs has been successful and is an indication that these kinds of mutants can be effectively produced. More work and more reproducibility is however needed to be able to draw more accurate conclusions. Numerous interesting new ideas have arisen during this thesis that could be explored further to getting a step closer to artificial blood and iron supplements made from plants.

Acknowledgements

First and foremost, a big thank you to our supervisor Leonard Groth for all the help and invaluable insight during this thesis. Thank you for believing in our ideas and letting us be independent and try our own things. We would also like to sincerely thank our assistant supervisor Anna Barjuan Grau for guidance in the lab and for always answering our questions regarding the practical elements. Additionally, we would like to thank Dr. Simon Christensen for providing us with the cysteine mutants and answering our questions regarding them and the ÄKTA. We also want to thank Christina Wennerberg for helping us with the homogenization of our cells. Last but not least, thank you to Professor Leif Bülow for allowing us to do our thesis in your group and for interesting discussions along the way.

Contributions

The initial plan for this thesis was established by Leonard Groth and Prof. Leif Bülow at the Division of Pure and Applied Biochemistry at LTH. The protocol for the bacterial cultivation was provided by Leonard Groth and the cysteine mutant stocks as well as the ÄKTA program was obtained from Dr. Simon Christensen. Leonard Groth and Anna Barjuan Grau also helped with instructions during the downstream purifications. The CRISPR-Cas9 execution was planned and the protocol was constructed by us, Pennie Lindblom and Olivia Geel. The experiments were prepared and performed solely by us, including the writing of the thesis. We both contributed equally in the lab and during the writing process.

7. References

1. M. Johanson, “Trenden med färre blodgivare fortsätter,” *Läkartidningen*, Jun. 14, 2022. https://lakartidningen.se/aktuellt/nyheter/2022/06/trenden-med-farre-blodgivare-fortsatter/ (accessed Mar. 15, 2023).

2. A.-C. Torkelsson, “Vården larmar om akut blodbrist:»Har aldrig sett något liknande«,” *Läkartidningen*, Jan. 17, 2022. https://lakartidningen.se/aktuellt/nyheter/2022/01/varden-larmar-om-akut-blodbrist-har-aldrig-sett-nagot-liknande/ (accessed Mar. 14, 2023).

3. R. Pawlak, J. Berger, and I. Hines, “Iron Status of Vegetarian Adults: A Review of Literature,” *American Journal of Lifestyle Medicine*, vol. 12, no. 6, pp. 486–498, Dec. 2016, doi: https://doi.org/10.1177/1559827616682933.

4. National Heart, Lung, and Blood Institute, “What Is Anemia? ,” *www.nhlbi.nih.gov*, Mar. 24, 2022. https://www.nhlbi.nih.gov/health/anemia (accessed Mar. 24, 2023).

5. L. Chen, Z. Yang, and H. Liu, “Hemoglobin-Based Oxygen Carriers: Where Are We Now in 2023?,” *Medicina*, vol. 59, no. 2, p. 396, Feb. 2023, doi: https://doi.org/10.3390/medicina59020396.

6. L. Stryer, J. Berg, J. Tymoczko, and G. Gatto, *Biochemistry*, 8th ed. WH Freeman and Company, 2015.

7. Worldwide PDB Protein Data Bank, “DEOXY HUMAN HEMOGLOBIN,” *www.wwpdb.org*, Jul. 13, 2011. https://www.wwpdb.org/pdb?id=pdb_00001a3n (accessed May 18, 2023).

8. A. Kumari, “Heme Synthesis,” *Sweet Biochemistry*, pp. 33–36, 2018, doi: https://doi.org/10.1016/b978-0-12-814453-4.00008-x.

9. X. R. Zhao, K. R. Choi, and S. Y. Lee, “Metabolic engineering of Escherichia coli for secretory production of free haem,” *Nature Catalysis*, vol. 1, no. 9, pp. 720–728, Sep. 2018, doi: https://doi.org/10.1038/s41929-018-0126-1.

10. L. L. Anzaldi and E. P. Skaar, “Overcoming the Heme Paradox: Heme Toxicity and Tolerance in Bacterial Pathogens,” *Infection and Immunity*, vol. 78, no. 12, pp. 4977–4989, Aug. 2010, doi: https://doi.org/10.1128/iai.00613-10.

11. J. A. Hoy and M. S. Hargrove, “The structure and function of plant hemoglobins,” *Plant Physiology and Biochemistry*, vol. 46, no. 3, pp. 371–379, Mar. 2008, doi: https://doi.org/10.1016/j.plaphy.2007.12.016.

12. R. A. Hill, M. S. Hargrove, and Raúl 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,” *F1000 Research*, vol. 5, p. 212, Feb. 2016, doi: https://doi.org/10.12688/f1000research.8133.1.

13. A. Kumar, M. Nag, and S. Basak, “TRUNCATED OR 2/2 HEMOGLOBINS : A NEW CLASS OF GLOBINS WITH NOVEL STRUCTURE AND FUNCTION,” *Journal of Proteins and Proteomics*, vol. 4, no. 1, Oct. 2013, Accessed: Apr. 13, 2023. [Online]. Available: https://www.researchgate.net/publication/261063125_Truncated_Hemoglobins_A_New_Class_of_Globins_with_Novel_Structure_and_Function

14. Nélida Leiva-Eriksson *et al.*, “Differential Expression Patterns of Non-Symbiotic Hemoglobins in Sugar Beet (Beta vulgaris ssp. vulgaris),” *Plant Cell Physiology*, vol. 55, no. 4, pp. 834–844, Apr. 2014, doi: https://doi.org/10.1093/pcp/pcu027.

15. The Editors of Encyclopaedia Britannica, “Monocotyledon ,” *Encyclopædia Britannica*. 2019. Accessed: Apr. 19, 2023. [Online]. Available: https://www.britannica.com/plant/monocotyledon

16. R. D. Hill, “Non-symbiotic haemoglobins—What’s happening beyond nitric oxide scavenging?,” *AoB PLANTS*, vol. 2012, Jan. 2012, doi: https://doi.org/10.1093/aobpla/pls004.

17. J. C. Dohm *et al.*, “The genome of the recently domesticated crop plant sugar beet (Beta vulgaris),” *Nature*, vol. 505, no. 7484, pp. 546–549, Dec. 2013, doi: https://doi.org/10.1038/nature12817.

18. S. Christensen, L. Groth, N. Leiva-Eriksson, M. Nyblom, and L. Bülow, “Oxidative Implications of Substituting a Conserved Cysteine Residue in Sugar Beet Phytoglobin BvPgb 1.2,” *Antioxidants*, vol. 11, no. 8, p. 1615, Aug. 2022, doi: https://doi.org/10.3390/antiox11081615.

19. S. Christensen, O. Stenström, M. Akke, and L. Bülow, “Conformational Dynamics of Phytoglobin BvPgb1.2 from Beta vulgaris ssp. vulgaris,” *International Journal of Molecular Sciences*, vol. 24, no. 4, p. 3973, Jan. 2023, doi: https://doi.org/10.3390/ijms24043973.

20. RCSB Protein Data Bank, “RCSB PDB - 7Z1U: Biochemical implications of the substitution of a unique cysteine residue in sugar beet phytoglobin BvPgb 1.2,” *www.rcsb.org*, Sep. 07, 2022. https://www.rcsb.org/structure/7Z1U (accessed May 18, 2023).

21. Z. D. Blount, “The unexhausted potential of E. coli,” *eLife*, vol. 4, no. e05826, Mar. 2015, doi: https://doi.org/10.7554/elife.05826.

22. A. R. Tuttle, N. D. Trahan, and M. S. Son, “Growth and Maintenance of Escherichia coli Laboratory Strains,” *Current Protocols*, vol. 1, no. 1, Jan. 2021, doi: https://doi.org/10.1002/cpz1.20.

23. S. Kim, H. Jeong, E.-Y. Kim, J. F. Kim, S. Y. Lee, and S. H. Yoon, “Genomic and transcriptomic landscape of Escherichia coli BL21(DE3),” *Nucleic Acids Research*, vol. 45, no. 9, pp. 5285–5293, Mar. 2017, doi: https://doi.org/10.1093/nar/gkx228.

24. F. W. Studier and B. A. Moffatt, “Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes,” *Journal of Molecular Biology*, vol. 189, no. 1, pp. 113–130, May 1986, doi: https://doi.org/10.1016/0022-2836(86)90385-2.

25. Thermo Fisher Scientific, “Gateway™ pET-DEST42 Vector.” https://www.thermofisher.com/order/catalog/product/12276010 (accessed Apr. 10, 2023).

26. T. A. Brown, *Gene cloning and DNA analysis : an introduction*, 8th ed. Chichester, West Sussex ; Hoboken, Nj: Wiley-Blackwell, 2020.

27. A. Chang, V. Chau, and J. Landas, “Preparation of calcium competent Escherichia coli and heat-shock transformation,” *The Undergraduate Journal of Experimental Microbiology and Immunology*, vol. 1, Jul. 2017, Accessed: Apr. 14, 2023. [Online]. Available: https://static.igem.org/mediawiki/2018/d/d2/T--NYMU-Taipei--protocol-competent-cell.pdf

28. N. Renzette, “Generation of Transformation Competent E. coli,” *Current Protocols in Microbiology*, vol. 22, no. 1, pp. A.3L.1–A.3L.5, Aug. 2011, doi: https://doi.org/10.1002/9780471729259.mca03ls22.

29. Bio-Rad, “Anion Exchange Chromatography ,” *www.bio-rad.com*. https://www.bio-rad.com/en-se/applications-technologies/anion-exchange-chromatography?ID=MWHAZ4C4S (accessed May 25, 2023).

30. M-T. Hopp, B. F. Schmalohr, T. Kühl, M. S. Detzel, A. Wißbrock, and D. Imhof, “Heme Determination and Quantification Methods and Their Suitability for Practical Applications and Everyday Use,” *Analytical Chemistry*, vol. 92, no. 14, pp. 9429–9440, Jun. 2020, doi: https://doi.org/10.1021/acs.analchem.0c00415.

31. C. Natarajan, X. Jiang, A. Fago, R. E. Weber, H. Moriyama, and J. F. Storz, “Expression and Purification of Recombinant Hemoglobin in Escherichia coli,” *PLoS ONE*, vol. 6, no. 5, p. e20176, May 2011, doi: https://doi.org/10.1371/journal.pone.0020176.

32. Thermo Fisher Scientific , “One Shot™ BL21 Star™ (DE3) Chemically Competent *E. coli.*” https://www.thermofisher.com/order/catalog/product/C601003 (accessed Apr. 10, 2023).

33. D. M. Villarreal *et al.*, “Enhancement of Recombinant Hemoglobin Production in *Escherichia coli* BL21(DE3) Containing the *Plesiomonas shigelloides* Heme Transport System,” *Applied and Environmental Microbiology*, vol. 74, no. 18, pp. 5854–5856, Sep. 2008, doi: https://doi.org/10.1128/aem.01291-08.

34. J. Zhang, Z. Kang, J. Chen, and G. Du, “Optimization of the heme biosynthesis pathway for the production of 5-aminolevulinic acid in Escherichia coli,” *Scientific Reports*, vol. 5, no. 1, Feb. 2015, doi: https://doi.org/10.1038/srep08584.

35. T. Harnois, M. Rousselot, H. Rogniaux, and F. Zal, “High-level Production of RecombinantArenicola MarinaGlobin Chains inEscherichia Coli:A New Generation of Blood Substitute,” *Artificial cells, Nanomedicine and Biotechnology*, vol. 37, no. 3, pp. 106–116, Aug. 2009, doi: https://doi.org/10.1080/10731190902908445.

36. E. Verkamp, V. M. Backman, J. M. Björnsson, D. Söll, and G. Eggertsson, “The periplasmic dipeptide permease system transports 5-aminolevulinic acid in Escherichia coli,” *Journal of Bacteriology*, vol. 175, no. 5, pp. 1452–1456, Mar. 1993, doi: https://doi.org/10.1128/jb.175.5.1452-1456.1993.

37. W. N. Abouhamad and M. D. Manson, “The dipeptide permease of Escherichia coli closely resembles other bacterial transport systems and shows growth-phase-dependent expression,” *Molecular Microbiology*, vol. 14, no. 5, pp. 1077–1092, Dec. 1994, doi: https://doi.org/10.1111/j.1365-2958.1994.tb01340.x.

38. M. J. Lee, “Effect of Gene Amplifications in Porphyrin Pathway on Heme Biosynthesis in a Recombinant Escherichia coli,” *Journal of Microbiology and Biotechnology*, vol. 23, no. 5, pp. 668–673, May 2013, doi: https://doi.org/10.4014/jmb.1302.02022.

39. BioCyc, “Escherichia coli BL21(DE3) (GCF_000022665.1) Pathway: heme b biosynthesis V (aerobic),” *biocyc.org*, 2023. https://biocyc.org/GCF_000022665/NEW-IMAGE?type=PATHWAY&object=HEME-BIOSYNTHESIS-II-1 (accessed May 11, 2023).

40. E. A. Sweeny *et al.*, “Glyceraldehyde-3-phosphate dehydrogenase is a chaperone that allocates labile heme in cells,” *The Journal of Biological Chemistry*, vol. 293, no. 37, pp. 14557–14568, Sep. 2018, doi: https://doi.org/10.1074/jbc.RA118.004169.

41. B. Tupta *et al.*, “GAPDH is involved in the heme‐maturation of myoglobin and hemoglobin,” *The FASEB Journal*, vol. 36, no. 2, Dec. 2021, doi: https://doi.org/10.1096/fj.202101237rr.

42. J. A. Doudna and E. Charpentier, “The new frontier of genome engineering with CRISPR-Cas9,” *Science*, vol. 346, no. 6213, pp. 1258096–1258096, Nov. 2014, doi: https://doi.org/10.1126/science.1258096.

43. J. Salsman and G. Dellaire, “Precision genome editing in the CRISPR era,” *Biochemistry and Cell Biology*, vol. 95, no. 2, pp. 187–201, Apr. 2017, doi: https://doi.org/10.1139/bcb-2016-0137.

44. “Addgene: Homepage,” *Addgene.org*, 2019. https://www.addgene.org/

45. F. Zerbini *et al.*, “Large scale validation of an efficient CRISPR/Cas-based multi gene editing protocol in Escherichia coli,” *Microbial Cell Factories*, vol. 16, no. 1, Apr. 2017, doi: https://doi.org/10.1186/s12934-017-0681-1.

46. B. Kenkel, “Lambda Red: A Homologous Recombination-based Technique for Genetic Engineering,” *blog.addgene.org*, Dec. 15, 2016. https://blog.addgene.org/lambda-red-a-homologous-recombination-based-technique-for-genetic-engineering (accessed May 12, 2023).

47. “Agilent Genomics : Tools - Bio Calculators,” *www.agilent.com*. https://www.agilent.com/store/biocalculators/calcODBacterial.jsp (accessed May 5, 2023).

48. “CHOPCHOP.” http://chopchop.cbu.uib.no/ (accessed May 10, 2023).

49. M. El Karoui, “Gene replacement with linear DNA in electroporated wild-type Escherichia coli,” *Nucleic Acids Research*, vol. 27, no. 5, pp. 1296–1299, Mar. 1999, doi: https://doi.org/10.1093/nar/27.5.1296.

50. “NEBioCalculator.” https://nebiocalculator.neb.com/# (accessed May 15, 2023).

51. “Escherichia coli (strain B / BL21-DE3),” 2019. https://www.uniprot.org/proteomes/UP000002032 (accessed May 23, 2023).

52. T. Kroemer, “How to Interpret DNA Gel Electrophoresis Results.” https://goldbio.com/articles/article/Interpreting-Gel-Electrophoresis-Results#_Toc37918944 (accessed Apr. 18, 2023).

53. A. Gallego, “How To Interpret RT-qPCR Results | GoldBio.” https://goldbio.com/articles/article/How-To-Interpret-RT-qPCR-Results (accessed May 3, 2023).

54. Addgene, “Addgene: Protocol - Bacterial Transformation,” Nov. 13, 2017. https://www.addgene.org/protocols/bacterial-transformation/ (accessed May 5, 2023).

8. Appendix

8.1 NanoDrop Results

Table A1, A2 and A3 below shows the results from the NanoDrop measurements. The values are mean values between duplicates. It shows the concentration in ng/µl as well as the purity in the ratios A260/A280 and A260/A230.

Table A1. The results from NanoDrop for the Pgb mutants which show the concentration of the RNA and the purity.

Samples	Concentration (ng/µl)	A260/A280	A260/A230
K34C.1	33.4	2.05	1.71
K34C.2	30.8	2.01	1.23
N7C.1	50.1	1.93	1.07
N7C.2	36.0	2.99	1.53
E120C.1	38.6	1.98	1.43
E120C.2	40.2	1.97	1.34
rWt.1	39.2	2.05	1.51
rWt.2	40.0	2.01	1.42
Native.1	28.3	1.94	0.833
Native.2	39.2	1.99	1.43

Table A2. The results from NanoDrop for the first δ-ALA gradient which shows the concentration of the RNA and the purity.

Samples	Concentration (ng/µl)	A260/A280	A260/A230
-ALA.1	84.9	1.97	1.80
-ALA.2	80.9	2.02	0.960
23.1	91.1	1.96	1.56
23.2	122	2.01	2.13
46.1	124	1.96	1.96
46.2	97.8	2.01	1.95

92.1	98.7	1.97	1.71
92.2	104	1.96	1.70

Table A3. The results from NanoDrop for the second δ-ALA gradient which shows the concentration of the RNA and the purity.

Samples	Concentration (ng/µl)	A260/A280	A260/A230
-ALA.1	93.6	1.95	1.95
-ALA.2	50.1	1.97	1.61
46.1	130	2.04	0.919
46.2	46.1	1.97	1.71
92.1	113	1.94	1.38
92.2	72.6	1.96	1.91
138.1	94.2	1.95	1.72
138.2	96.2	1.94	2.09
184.1	115	1.99	0.965
184.2	77.3	1.93	1.73

8.2 Gel electrophoresis

For intact RNA, two bands representing the 16S and 23S units of the ribosomal RNA should be seen. The gel below (Figure A1) is from the Pgb mutant cultivation and two ladders and duplicates for the samples K34C, N7C, E120C, rWt and Native cells are displayed. There are two distinct bands visible (by the red arrows in figure A1) that are expected to be seen when looking at RNA and this tells us that the RNA extraction was successful. The bright band (lowest for each sample lane) is RNA debris which is non intact RNA. This debris is however always expected and is okay to have in the sample when moving on with downstream processing. Figure A1 below is an example on how the gel should look when the RNA extraction is deemed successful and is an verification that cDNA synthesis can be performed.

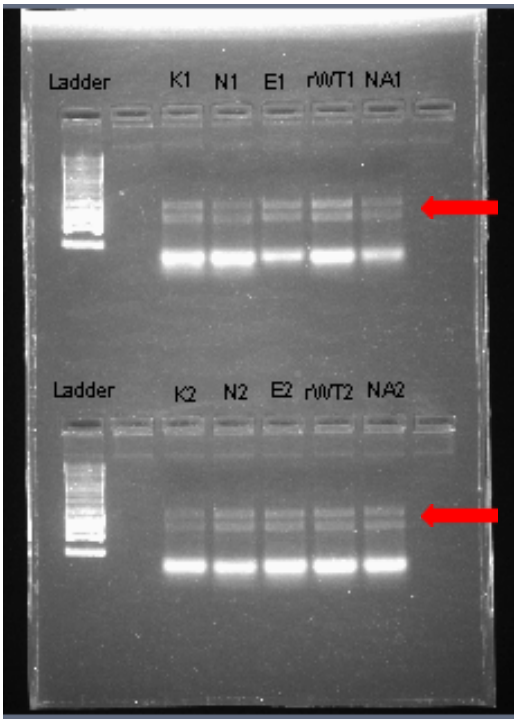

Figure A1. Example of a gel when a successful RNA extraction has been performed. The red arrows highlight the two bands that are searched for when determining intact RNA. K1 and K2 are duplicates of K34C, N1 and N2 are duplicates of N7C, E1 and E2 are duplicates of E120C, rWt1 and rWt2 are duplicates of rWt, NA1 and NA2 are duplicates of the Native cells.

8.3 GeneRuler and PageRuler guide

Figure A2 and A3 below shows the reference guides used to interpret the gel electrophoresis and the SDS-PAGE ladders.

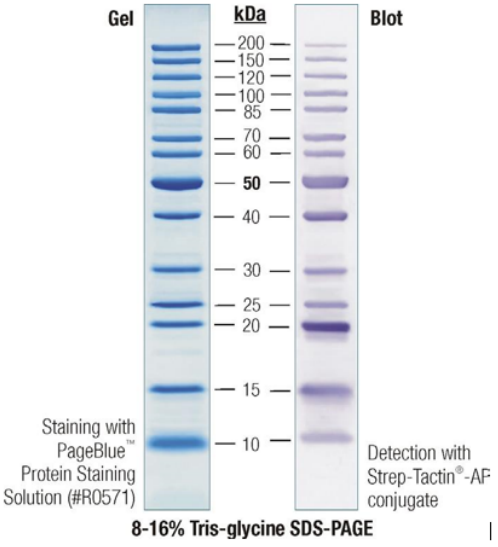

Figure A2. Shows the weight represented by the reference bands in the PageRule used in SDS-PAGE

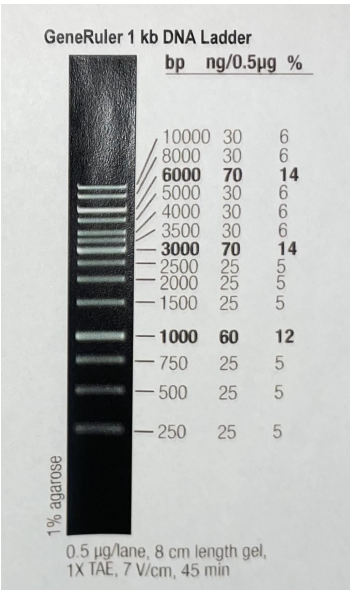

Figure A3. Shows the number of base pairs represented by the reference bands in the GeneRuler used in Gel electrophoresis

8.4 Primer validation and efficiency calculations

8.4.1 Melt curve analysis

To evaluate and investigate if the primers have off-targets, a melt curve analysis was performed. If they only display one peak in the curve it means that they only bind to one thing, i.e. the target, and not to other mRNAs. To have only one peak is what is aimed for to get very specific primer pairs. Figure A4 below shows the melt curve for the primer targeting *BvPgb1.2* and is an example of a good melt curve that indicates specific primers.

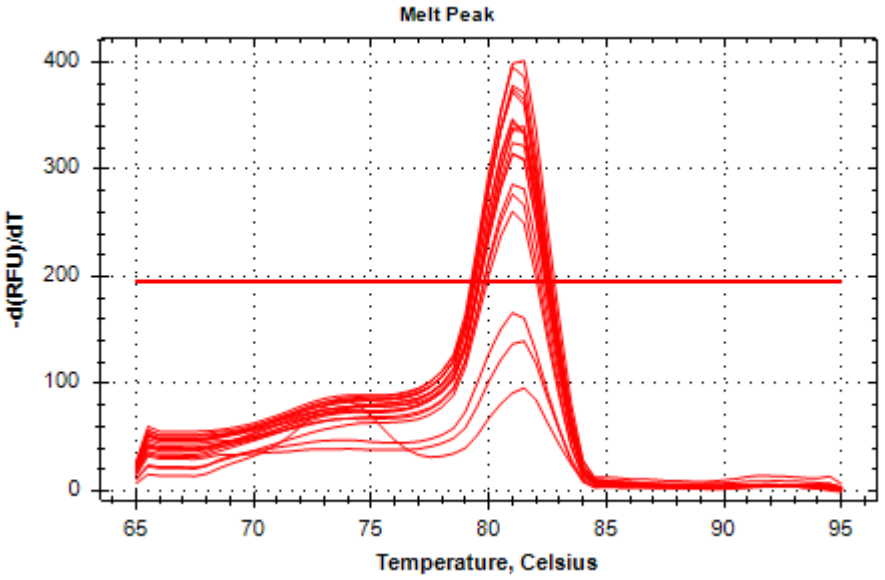

Figure A4. Melt curve for Bvpgb1.2. Shows the temperature in celsius on the x axis and the y axis the relative fluorescence unit (RFU) over time.

All of the primers (rrSA, cysG, hemH and BvPgb1.2) only displayed one peak which means that they are specific and they are considered validated.

8.4.2 Primer efficiency

Primer efficiency is calculated by analyzing a dilution series and plotting its Ct values against Log(10) of the dilution. Then the efficiency is calculated using equation A1.

$$Primer\ efficiency\ = 10^{\left(\frac{-1}{Slope}\right)} - 1 * 100\% \qquad \text{(A1)}$$

The slope is obtained from the plot described above. An acceptable primer efficiency is in the range 85-110% [53]. Table A4 below shows the primer efficiencies for the primers used in this thesis. A converted primer efficiency that is used for the relative gene expression calculation is also shown. The converted efficiencies are calculated using equation A2 below.

$$Converted\ primer\ efficiency\ = \frac{Primer\ efficiency\ (\%)}{100} + 1 \qquad \text{(A2)}$$

Table A4. Displays the primer efficiencies as well as the converted efficiencies for the primers rrsA, cysG, hemH and Bv2

Primer	Efficiency (%)	Converted efficiency
rrsA	97.6	1.98
cysG	101	2.01
hemH	109	2.09
BvPgb1.2	88.4	1.88

8.5 Relative gene expression calculations - Raw data

8.5.1 Phytoglobin mutants

Table A5-A8 displays the raw data from the relative gene expressions for the Pgb mutants. Two reference genes (REF) have been used, *rrsA* and *cysG*. GOI stands for gene of interest and the two genes investigated are *hemH* and *BvPgb1.2*. The calculations are relative to either native *E. coli* BL21 (DE3) or rWt.

Table A5. Shows the relative gene expression calculation for the gene *hemH* relative to native *E. coli*.

	REF rrsA			REF cysG			GOI hemH				
Sample	*Ct*	*ΔCt*	*RQ*	*Ct*	*ΔCt*	*RQ*	*Ct*	*ΔCt*	*RQ*	**GeoMean RQ REFs**	**Relative Gene expression**
K34C.1	8,79	1,19	2,249584494	26,49	0,085	1,060934601	27,04	-0,15	0,895570208	1,544882529	0,579701169
K34C.2	10,94	-0,96	0,519937424	27,35	-0,775	0,583150009	27,44	-0,55	0,667367939	0,55063737	1,21199173
N7C.1	9,48	0,5	1,405860235	26,83	-0,255	0,837402318	26,47	0,42	1,361824095	1,085021023	1,255113096
N7C.2	10,36	-0,38	0,771905203	28,1	-1,525	0,346031829	28,4	-1,51	0,329460248	0,516820829	0,637474788
E120C.1	11,99	-2,01	0,254256796	29,12	-2,545	0,170158122	27,8	-0,91	0,5121586	0,207999661	2,46230498
E120C.2	11,61	-1,63	0,329388628	28,33	-1,755	0,294853089	28,21	-1,32	0,378858306	0,311642831	1,21568112
rWt.1	9,25	0,73	1,644354816	27,16	-0,585	0,665582499	27,08	-0,19	0,869613369	1,046161454	0,831242029
rWt.2	9,62	0,36	1,277963193	27,29	-0,715	0,608013785	27,18	-0,29	0,807965158	0,881486947	0,916593446
Native.1	10,94	-0,96	0,519937424	26,82	-0,245	0,843249998	27,99	-1,1	0,445379966	0,662145929	0,672631132
Native.2	9,02	0,96	1,923308373	26,33	0,245	1,185887937	25,79	1,1	2,245273871	1,510241106	1,486698953
Control average Ct	9,98			26,575			26,89				

Table A6. Shows the relative gene expression calculation for the gene *hemH* relative to rWt.

Sample	REF rrsA			REF cysG			GOI hemH			GeoMean RQ REFs	Relative Gene expression
	Ct	*ΔCt*	*RQ*	*Ct*	*ΔCt*	*RQ*	*Ct*	*ΔCt*	*RQ*		
K34C.1	8,79	0,645	1,551834381	26,49	0,735	1,667750018	27,04	0,09	1,068415592	1,608748525	0,664128405
K34C.2	10,94	-1,505	0,358669244	27,35	-0,125	0,916690281	27,44	-0,31	0,7961702	0,573400916	1,388505281
N7C.1	9,48	-0,045	0,969806759	26,83	0,395	1,316365523	26,47	0,66	1,624656653	1,12987618	1,43790681
N7C.2	10,36	-0,925	0,532484571	28,1	-0,875	0,543949257	28,4	-1,27	0,393046199	0,538186387	0,730316129
E120C.1	11,99	-2,555	0,175394362	29,12	-1,895	0,267482284	27,8	-0,67	0,611005401	0,216598441	2,820913199
E120C.2	11,61	-2,175	0,227222671	28,33	-1,105	0,463498168	28,21	-1,08	0,4519781	0,324526257	1,392731991
rWt.1	9,25	0,185	1,134327848	27,16	0,065	1,046271111	27,08	0,05	1,037449074	1,089410142	0,952303484
rWt.2	9,62	-0,185	0,881579344	27,29	-0,065	0,955775219	27,18	-0,05	0,963902735	0,917927933	1,050085416
Native.1	10,94	-1,505	0,358669244	26,82	0,405	1,325557859	27,99	-0,86	0,531338465	0,689519278	0,770592617
Native.2	9,02	0,415	1,326758816	26,33	0,895	1,864172047	25,79	1,34	2,678612566	1,572675013	1,703220655
Control average Ct	9,435			27,225			27,13				

Table A7. Shows the relative gene expression calculation for the gene *BvPgb1.2* relative to native *E. coli*.

Sample	REF rrsA			REF cysG			GOI BvPgb1.2			GeoMean RQ REFs	Relative Gene expression
	Ct	*ΔCt*	*RQ*	*Ct*	*ΔCt*	*RQ*	*Ct*	*ΔCt*	*RQ*		
K34C.1	8,79	1,19	2,249584494	26,49	0,085	1,060934601	13,93	14,735	11303,14173	1,544882529	7316,505635
K34C.2	10,94	-0,96	0,519937424	27,35	-0,775	0,583150009	16,24	12,425	2616,861727	0,55063737	4752,423046
N7C.1	9,48	0,5	1,405860235	26,83	-0,255	0,837402318	15,34	13,325	4627,499548	1,085021023	4264,893906
N7C.2	10,36	-0,38	0,771905203	28,1	-1,525	0,346031829	18,07	10,595	821,1024907	0,516820829	1588,756577
E120C.1	11,99	-2,01	0,254256796	29,12	-2,545	0,170158122	15,71	12,955	3660,744446	0,207999661	17599,76163
E120C.2	11,61	-1,63	0,329388628	28,33	-1,755	0,294853089	18,39	10,275	670,4618047	0,311642831	2151,378878
rWt.1	9,25	0,73	1,644354816	27,16	-0,585	0,665582499	14,28	14,385	9055,734024	1,046161454	8656,1534
rWt.2	9,62	0,36	1,277963193	27,29	-0,715	0,608013785	14,91	13,755	6076,139439	0,881486947	6893,056622
Native.1	10,94	-0,96	0,519937424	26,82	-0,245	0,843249998	30,29	-1,625	0,357278217	0,662145929	0,53957625
Native.2	9,02	0,96	1,923308373	26,33	0,245	1,185887937	27,04	1,625	2,798939181	1,510241106	1,853306184
Control average Ct	9,98			26,575			28,665				

Table A8. Shows the relative gene expression calculation for the gene *BvPgb1.2* relative to rWt.

Sample	REF rrsA			REF cysG			GOI BvPgb1.2			GeoMean RQ REFs	Relative Gene expression
	Ct	*ΔCt*	*RQ*	*Ct*	*ΔCt*	*RQ*	*Ct*	*ΔCt*	*RQ*		
K34C.1	8,79	0,645	1,551834381	26,49	0,735	1,667750018	13,93	0,665	1,523784242	1,608748525	0,947186101
K34C.2	10,94	-1,505	0,358669244	27,35	-0,125	0,916690281	16,24	-1,645	0,352780913	0,573400916	0,615243024
N7C.1	9,48	-0,045	0,969806759	26,83	0,395	1,316365523	15,34	-0,745	0,623836369	1,12987618	0,552128083
N7C.2	10,36	-0,925	0,532484571	28,1	-0,875	0,543949257	18,07	-3,475	0,110693387	0,538186387	0,205678533
E120C.1	11,99	-2,555	0,175394362	29,12	-1,895	0,267482284	15,71	-1,115	0,493507454	0,216598441	2,278444168
E120C.2	11,61	-2,175	0,227222671	28,33	-1,105	0,463498168	18,39	-3,795	0,090385413	0,324526257	0,278514946
rWt.1	9,25	0,185	1,134327848	27,16	0,065	1,046271111	14,28	0,315	1,220809677	1,089410142	1,120615304
rWt.2	9,62	-0,185	0,881579344	27,29	-0,065	0,955775219	14,91	-0,315	0,8191285	0,917927933	0,892366896
Native.1	10,94	-1,505	0,358669244	26,82	0,405	1,325557859	30,29	-15,695	4,81649E-05	0,689519278	6,98529E-05
Native.2	9,02	0,415	1,326758816	26,33	0,895	1,864172047	27,04	-12,445	0,000377327	1,572675013	0,000239927
Control average Ct	9,435			27,225			14,595				

8.5.2 The δ-ALA gradient

Table A9 and A10 shows the raw data from the relative gene expressions for the δ-ALA gradient. Just as with the Pgb mutants two reference genes (REF) have been used, *rrsA* and *cysG*. The GOI is *hemH* and the calculations are relative to native *E. coli* BL21 (DE3) that had no addition of δ-ALA in the growth media. Table A9 was the first try at the δ-ALA gradient with addition of 23, 46 and 92 µl. Table A10 is an extension of the experiment where 46, 92, 138 and 184 µl were investigated. The mean values and standard deviation in table 4 are for 23ALA, 138ALA and 184ALA calculated from the duplicates in respective tables. For 46ALA and 92ALA that were used in both experiments, all 4 values (two from table A9 and two from A10) were used in the calculation of the mean value and standard deviation.

Table A9. Shows the relative gene expression calculation for the gene hemH relative to no addition of δ-ALA (-ALA1,2) for the first try.

Sample	REF rrsA			REF cysG			GOI hemH			GeoMean RQ REFs	Relative Gene expression
	Ct	*ΔCt*	*RQ*	*Ct*	*ΔCt*	*RQ*	*Ct*	*ΔCt*	*RQ*		
-ALA.1	9,9	-0,085	0,943734507	24,27	-0,62	0,649567413	24,64	-0,665	0,61325589	0,782955415	0,783257741
-ALA.2	9,73	0,085	1,059620044	23,03	0,62	1,539486095	23,31	0,665	1,630640678	1,277211934	1,276718949
23ALA.1	8,7	1,115	2,137523941	22,88	0,77	1,708868398	23,29	0,685	1,654797999	1,911216134	0,865835093
23ALA.2	9,14	0,675	1,583878627	23,99	-0,34	0,789306256	24,5	-0,525	0,679749243	1,118107915	0,607946008
46ALA.1	9,24	0,575	1,479563023	23,14	0,51	1,426039917	23,49	0,485	1,428492082	1,452554967	0,983434097

46ALA.2	8,91	0,905	1,852572811	24,13	-0,48	0,716036049	24,64	-0,665	0,61325589	1,151741688	0,532459575
92ALA.1	8,45	1,365	2,534440274	23,95	-0,3	0,811585547	24,43	-0,455	0,715652569	1,43419493	0,498992539
92ALA.2	10,16	-0,345	0,790532845	21,21	2,44	5,462787704	21,38	2,595	6,740254036	2,078103247	3,243464465
Control average Ct	9,815			23,65			23,975				

Table A10. Shows the relative gene expression calculation for the gene hemH relative to no addition of δ-ALA (-ALA1,2) for the second try.

	REF rrsA			**REF cysG**			**GOI hemH**			**GeoMean**	**Relative**
Sample	*Ct*	*ΔCt*	*RQ*	*Ct*	*ΔCt*	*RQ*	*Ct*	*ΔCt*	*RQ*	**RQ REFs**	**Gene expresion**
46.1	10,3	-0,14	0,909025778	24,19	0,275	1,210905414	24,02	0,82	1,827491279	1,049163588	1,741855416
46.2	9,7	0,46	1,368065135	25	-0,535	0,689148544	25,35	-0,51	0,687288012	0,970978937	0,70782999
92.1	8,24	1,92	3,699115096	23,35	1,115	2,172571951	23,84	1	2,086103	2,834888657	0,735867701
92.2	8,45	1,71	3,205989847	23,07	1,395	2,639948678	23,42	1,42	2,84090533	2,909235064	0,976512818
138.1	10,14	0,02	1,013719232	24,28	0,185	1,137392898	25,06	-0,22	0,850640655	1,073777004	0,79219489
138.2	11,85	-1,69	0,316195397	22,76	1,705	3,275543485	23,06	1,78	3,701839889	1,017699254	3,637459567
184.1	9,34	0,82	1,748336889	22,79	1,675	3,207870293	23,16	1,68	3,439410843	2,368214088	1,4523226
184.2	10,15	0,01	1,006836249	22,85	1,615	3,076689438	23,21	1,63	3,315257518	1,760034787	1,883631814
-ALA.1	9,07	1,09	2,101424931	23,87	0,595	1,512935091	24,3	0,54	1,487446327	1,783064643	0,834207741
-ALA.2	11,25	-1,09	0,475867582	25,06	-0,595	0,660966889	25,38	-0,54	0,672293166	0,560832163	1,198742173
Control average Ct	10,16			24,465			24,84				

8.6 CRISPR-Cas9: Plasmid recovery from bacteria and plasmid purification

After the plasmid recovery and purification, the plasmids were run on gel electrophoresis and tested using NanoDrop. The plasmids were recovered in duplicates. Table A11 below shows the NanoDrop results, both the concentration and the purity as a mean value of the sample duplicates. The plasmids have a concentration in the range between 34-127 ng/µl and a purity in relation to proteins (A260/A280) above 1.8. The purity related to carbohydrates and phenols (A260/A230) vary a bit more but with most of the samples above 1.85, the exception is the pCNA plasmids which have a bit lower purity.

Table A11. The NanoDrop results after plasmid recovery and purification. It displays both the concentration in ng/µl and the purity in the form of the ratios A260/A280 and A260/A230.

Plasmid	Concentration (ng/µl)	A260/A280	A260/A230
pCNA.1	42.5	1.92	1.69
pCNA.2	34.6	1.85	1.76
pCRISPR.1	45.5	1.85	1.86
pCRISPR.2	78.0	1.85	1.99
pCas9.1	54.7	1.82	1.88
pCas9.2	78.3	1.81	2.04
pKM154.1	127	1.84	2.18
pKM154.2	123	1.82	2.13

The results from the gel electrophoresis to verify that only the plasmid was purified are seen in figure A5 below. In the figure it can be seen that there is only one band per lane, however with a bit of smudging. This shows that the purification has worked and that only the plasmid is present in the sample.

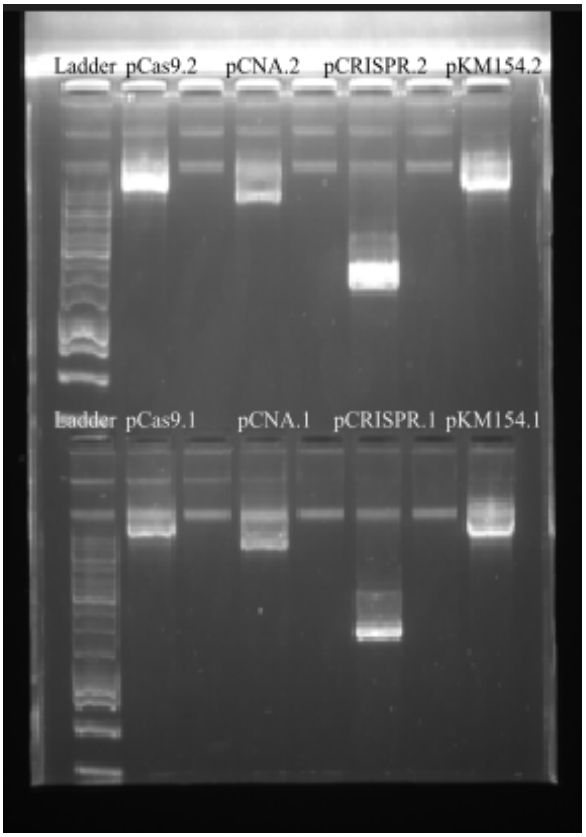

Figure A5. Gel electrophoresis results after plasmid purification.

8.7 Primers for PIPE

Table A12 states the primes that were used and was planned to be used in the PIPE experiment. How the primers will be used has previously been described in an article written by Zerbini et al from 2017 [45].

Table A12. Shows the primer names and sequence in 5'→3'. F in the names stands for forward while R stands for reverse.

Primer	Sequence
Pipe1 pCAS9-F	ctaacggattcaccactccaagaattggagccaatcaattc
Pipe1 pCAS9-R	cgaggtgccgccggcttccattcaggtcgaggtggcccgg
RedF	gccggcggcacctcgcatcgatttattatgacaacttg
RedR	tggtgaatccgttagtcatcgccattgctccccaa
pipeCRISPR-F	tgagcgggactctggggttcgagagctcgcttggactcctgttg
pipeCRISPR-R	tggcaattccgacgtctaagaaaccattattatcatgacattaacc
cat\sac-pipeF	acgtcggaattgccataagggcaccaataactgcc
cat\sac-pipeR	ccagagtcccgctcaatcggcattttcttttgcg
V-crSac F	tttagcttccttagctcctgaaaatctcg
V-crSac R	ttttttttaaggcagttattggtgccc
I-kanaF	actgccttaaaaaaaattgaacaagatggattgcacgc
I-kanaR	gctaaggaagctaaacgccttcttgacgagttcttc

8.8 Protocols

8.8.1 Cultivation protocol

Lysogeny Broth Medium Preparation

1. Add 250 g Milli-Q water to 2.5 g tryptone + 1.25 g yeast extract + 2.5 g NaCl
2. Dissolve on stir plate
3. Autoclave LB with 48X inoculum tubes + cotton plugs
4. Set aside in cold room until inoculation

Terrific Broth Medium Preparation

1. Phosphate buffer: add 500 g Milli-Q water + 9.24 g KH_2PO_4 + 50.16 g K_2HPO_4 in 1000 mL Schott flask
2. Dissolve phosphate buffer on stir plate
3. C/N source: add 3000 g Milli-Q water + 48 g tryptone + 96 g yeast extract in 4000 mL beaker
4. Dissolve C/N source on stir plate
5. Add 600 g Milli-Q water and 18.9 g glycerol to C/N source
6. Transfer to 5 000 mL Schott flask
7. Autoclave Phosphate Buffer (1000 mL Schott flask) and C/N Source (5000 mL Schott flask)
8. Autoclave plastic funnel + 12X 2 L baffled shake flasks + 12 X cotton plugs

Plate pouring

1. Autoclave agar plate-media

   Note: this is a good time to label each plate with name, date, strain and antibiotics. Perform step 3 first.
2. Allow autoclaved agar plate-media to settle to handling temperatures (<60 °C)
3. Sterilize workbench, set Bunsen burner flame set on high and work within 30 cm of flame
4. Supplement agar plate-media with ampicillin for a final concentration of 100 ug/mL ampicillin

   Note: ampicillin is stored as 1000X stocks of 100 mg/mL
5. Pour into sterile plates

   Note: use serological pipette to get a "feel" of how much media is needed per plate.
6. Swirl to remove bubbles and cover bottom of plate evenly, put on lid and place next place on top.
7. Stack plates as you pour.
8. Allow to media in plates to solidify, minimum 30 minutes.
   Note: If possible, allow plates to dry at room temperature overnight before storing them at 4 °C.
9. If media begins to partially solidify in beaker, stop pouring-process. If more plates are needed, start over.

Bacterial plating

1. Transfer the glycerol stock onto dry ice next to a flame.
2. Sterilise the loop in the flame.
3. Rub the loop in the glycerol stock.
4. Streak the bacteri in zigzag over half the agar plate.
5. Sterilise the loop.
6. Zigzag over half the plate which overlaps with the first half by a quarter of a plate.
7. Sterilise the loop.
8. Zigzag over half the plate which overlaps with the second half by a quarter of a plate.
9. Incubate overnight at 37 °C
10. The next day, you should have single colonies visible on a quarter of the plate or half the plate

Day 1: Inoculation

10. Sterilize workbench, set Bunsen burner flame set on high
11. Add 250 uL ampicillin (1 mg/mL) to LB medium stock
12. Transfer 5 mL LB + Amp per flask, cork and cover with foil
13. Inoculate with desired expression colony
14. Incubate at 37 °C for 16-18 h at 150 rpm
15. Autoclave shake-flasks and TB medium components for tomorrow

Day 2: Cultivation and induction

1. Sterilize workbench, set Bunsen burner flame set on high
2. Add 400 mL phosphate buffer to 3600 mL C/N source
3. Add 4000 uL carbenicillin (1 mg/mL) to 4000 mL room-tempered TB medium
4. Distribute 650 g TB medium + Carb per shake flask
5. Pour over incubated inoculums
6. Incubate at 37 °C for 3 – 4 h at 150 rpm
7. Cut 12X large strips of parafilm and sterilize with 70 % EtOH
8. Begin induction once $OD_{600} \geq 2.5$ a.u.
9. Sterilize workbench, set Bunsen burner flame set on high
10. Work in batches of 6 flasks around flame
11. Replace plugs with sterilized parafilm
12. Add 330 uL 1M IPTG per shake flask (yields 0.5 mM solution)
13. Add 200 uL 1M δ-ALA per shake flask (yields 0.3 mM solution)
14. Bubble with CO for 25 s and seal parafilm tightly
15. Incubate at 22 °C for 16-18 h at 150 rpm in dark

Day 3: Harvesting

1. Work in 3X batches of 4 flasks at a time
2. Transfer fermentation media to 4X 1 L centrifugation flasks, equilibrate mass within ± 0.05 g
3. Centrifuge at 3700 rpm at 4 °C for 40 min
4. Repeat fermentation media transfer, equilibrate, and keep ready by centrifuge
5. Autoclave flasks and plugs then place in dishwasher and run Program E
6. After centrifugation, discard supernatant and transfer to 5000 mL Schott flask

7. *Per batch:* Wash and dissolve cell-pellet with 30 mL 50 mM Tris-HCl, pH 8.5
8. *Per batch:* Transfer dissolved cells to 2X 50 mL Falcon tube with beaker, equilibrate within ± 0.05g
9. Centrifuge at 11000 rpm at 4 °C for 20 min
10. Discard supernatant in Bacterial Waste
11. Flash-freeze and place in -80 °C freeze

8.8.2 Protocol for plasmid purification from bacterial stabs

Create glycerol stocks for long term storage:

1. Create agar plates with the correct antibiotics and mix LB media
2. Using a sterile loop touch the bacteria growing within the punctured area of the stab culture.
3. Gently spread the bacteria over the plate - as before (sick sack)
4. Incubate overnight (12-18h) at 37℃
5. Add the correct antibiotics to the LB media
6. Transfer a single colony to 2-5 ml LB media
7. Grow the cultures for 12-18h at 37℃ and 150 rpm
8. Add 500 µL of the overnight culture to 500 µL of 50% glycerol in a 2 mL screw top tube and gently mix

   Note: Make the 50% glycerol solution by diluting 100% glycerol in dH20 and autoclave.
9. Freeze the glycerol stock tube at -80℃

Plasmid purification out of the bacteria

If you start from a glycerol stock:

1. Mix LB media and create agar plates if necessary
2. Scrape off some cells from the frozen glycerol stock
3. Spread on a plate as above
4. Let grow overnight (12-18h) at 37℃
5. Transfer a single colony to 2-5 ml LB media
6. Grow the cultures for 12-18h at 37℃ and 150 rpm
7. Then continue at step 7 below

In you start from a bacterial stab:

1. Perform step 1-7 above
2. Continue at step 7 below

7. Isolate plasmid DNA according to the instructions from the kit: E.Z.N.A.® Plasmid DNA Mini Kit I

8.8.3 Protocol for cloning crRNA into pCRISPR

1. Oligo (crRNA) design

Design two oligos with the following form:

Forward: 5'- AAACNNNNNNNNNNNNNNNNNNG - 3'
Reverse: 5' - AAAACNNNNNNNNNNNNNNNNNN - 3'

2. Vector digest

1. Mix the following in an eppendorf tube

Component	Amount
pCRISPR plasmid (1-2 µg)	x µl
BsaI (NEB)	1 µl
10X NEB Buffer	5 µl
100X BSA	0.5 µl
Milli-Q water	y µl
	Total: 50 µl

2. Incubate at 37℃ for 40 min.

3. Gel purification of digested pCRISPR

1. Prepare the gel with 0.7-1 % agarose content
2. Add 12.5 µl of the digested vector to an eppendorf and add 2 µl loading dye (do this x4)
3. Load the gel with a ladder and the samples (14.5 µl per well)
4. Run the gel at 60V for about 2 h
5. Weigh 4 empty eppendorf tubes (label them)
6. Locate the bands using the *Bio-Rad Gel Doc XR System w/ Universal Hood II*
7. Cut out the bands using a scalpel (use face protection)
8. Add the gel slices to different eppendorf tubes (the ones you previously weighed) and weigh again to calculate the amount of buffer needed
9. Follow the protocol in the kit *NucleoSpin® Gel and PCR Clean-up*

4. Phosphorylation

1. Mix the following in an eppendorf tube

Component	Amount
Oligo I (100 µM)	1 µl
Oligo II (100 µM)	1 µl
10X T4 Ligase buffer (NEB)	5 µl
T4 Polynucleotide Kinase (PNK), (NEB)	1 µl
Milli-Q water	42 µl

2. Incubate at 37℃ for 30 minutes
3. Incubate at 65℃ for 20 minutes to enable heat inactivation

5. Annealing

1. Add 2.5 µl of 1M NaCl to the phosphorylated oligo pairs
2. Incubate for 5 min at 95℃ in a heat block
3. Turn off the heat block and let cool naturally for 2-3 h to room temperature
4. Dilute annealed oligos 10 times

6. Ligation

1. Mix the following in an eppendorf tube

Component	Amount
Gel purified, BsaI digested pCRISPR	1 µl
Annealed oligos (crRNA)	2 µl
10X T4 Ligase buffer (NEB)	2 µl
T4 Ligase	1 µl
Milli-Q water	14 µl

2. Incubate at room temperature for 2 h (or 16℃ overnight)
3. Heat inactivate enzymes at 65℃ for 10 min
4. Chill on ice
5. Transform into competent cells

8.8.4 Lab protocol for CRISPR-Cas9

The following protocol is based on an article written by Zerbini et al from 2017.

Create chemically competent cells

Note: Prepare LB Media, $MgCl_2$ solution and $CaCl_2$ + Glycerol solution before

1. Grow native cells on agar plates overnight at 37℃
2. Inoculate 10 ml of LB Media with a single colony and incubate in 37℃ overnight
3. Pour over inoculum to 1 L LB Media
4. Incubate the culture at 37℃, 150 rpm until OD_{600} is 0.4-0.6 (about 3h)
5. In the following steps the samples should be held on ice when possible
6. Divide the 1 L culture into two centrifugation flasks and centrifuge for 20 min at 2500g and 4℃ (do not forget to equilibrate the flasks to ± 0.05g)
7. Discard the supernatant
8. Add 50 ml of $MgCl_2$ solution to each pellet and resuspend them gently.
9. Incubate the suspended cells on ice for 30 min.
10. Centrifuge for 20 min at 2500g and 4℃
11. Discard the supernatant
12. Add 10 ml of $CaCl_2$ + Glycerol solution to each pellet and resuspend them gently.
13. Divide the cells into 2 ml aliqutos, flash-freeze them and put in -80℃ freezer

Preparation of $MgCl_2$ solution

1. Add 20,33 g $MgCl_2$ * $6H_20$ in 80 ml MilliQ water and then adjust to 100 ml to make a 1M stock (Only do this the first time then use the stock and start from 2)
2. Add 20 ml 1M stock and 180 ml MilliQ water to a final concentration of 100 mM.
3. Sterile filtrate (0.22 μm) and store at 4℃

Preparation of $CaCl_2$ + Glycerol solution

1. Mix a glycerol + MilliQ water stock with 15% glycerol (7,5 ml glycerol + 42,5 ml MilliQ water)

2. Add $CaCl_2$ in 50 ml glycerol/MilliQ water to a final concentration of 100 mM (0.555 g)
3. Sterile filtrate (0.22 µm) and store at 4℃

Combine the plasmids using PIPE

Note: Create competent cells before this

1. Dilute the primers to 10 µM with Milli-Q water
2. Mix the *PIPE Pfu Master Mix* according to the table below

PIPE Pfu Master Mix

Component	Amount
10× Cloned Pfu DNA Polymerase Reaction Buffer	5 µl
PfuTurbo® DNA polymerase (2.5 U/µl)	1 µl
10 mM dNTPs	1 µl
MilliQ Water	28 µl
	Total: 35 µl

3. Add the following to a PCR tube:

Component	Amount
PIPE Pfu Master Mix	35 µl
Forward primer (V or I-PIPE)	5 µl
Reverse primer (V or I-PIPE)	5 µl
Template DNA (vector) (1–5 ng)	5 µl
	Total: 50 µl

4. Thermocycle according to conditions below:

(V-PIPE):
95°C for 2 min
25 cycles of 95°C for 30 s
55°C for 45 s and
68°C for 14 min
4°C hold.

(I-PIPE):
95°C for 2 min
25 cycles of 95°C for 30 s
55°C for 45 s
68°C for 3 min
4°C hold

5. Confirm successful amplification with gel electrophoresis.
6. Mix 2 µl from the V-PIPE and 2 µl I-PIPE reactions together
   Note: Combination of plasmids occur by mixing the PCR products
7. Transform into chemically competent cells using heat shock

Transformation of plasmids into competent cells:

Note: Make agar plates with antibiotics before this
Note 2: Work in a sterile environment

1. Thaw chemically competent cells on ice for 10-15 min
2. Mix together 20 µl competent cells with 4 µl Plasmid (pipet up and down once)
3. Incubate on ice for 15 min
4. Transfer the tubes to a 42 °C heat block and incubate for 45 s
5. Put the tubes back on ice gently
6. Add 100 µl LB media (no antibiotics) to each tube

7. Recover the cells by incubating them at 37°C for 1h and 250 rpm
8. Plate the cells on agar plates with the correct antibiotic (40 µl)
9. Incubate the plates for 12-16 h at 37°C
10. (Create glycerol stocks according to protocol.)

Greate glycerol stocks

1. Mix, autoclave and add the correct antibiotics to the LB media
2. Transfer a single colony to 2-5 ml LB media
3. Grow the cultures for 12-18h at 37℃ and 150 rpm
4. Add 500 µL of the overnight culture to 500 µL of 50% glycerol in a 2 mL screw top tube and gently mix

   Note: Make the 50% glycerol solution by diluting 100% glycerol in dH20 and autoclave.
5. Freeze the glycerol stock tube at -80℃

8.8.5 Control transformation Protocol from the division

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

Work in sterile environment!

2. Gently pipette 0.5 µl pure plasmid into 20 µl competent cells [54]. Handle with care.
3. Incubate 25 minutes on ice.
4. Transfer the tube to a 42 °C heat block and incubate for 45 s. Put the tube back on ice gently.
5. Add 900 µl sterile LB (no antibiotic) and incubate while shaking for 60 minutes at 37 °C.
6. 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).
7. Use two plates, one with 20 µl and 50 µl.
8. Let the plates dry before turning them upside down (~10-20 min).
9. Incubate upside down overnight at 37 °C.
10. Do not wait too long to take out the plates the day after. Do not let the colonies become too large.

8.8.6 Ligation protocol from New England BioLabs

Protocol

1. Set up the following reaction in a microcentrifuge tube on ice.
   (T4 DNA Ligase should be added last. Note that the table shows a ligation using a molar ratio of 1:3 vector to insert for the indicated DNA sizes.) Use NEBioCalculator to calculate molar ratios.

COMPONENT	20 µl REACTION
T4 DNA Ligase Buffer (10X)*	2 µl
Vector DNA (4 kb)	50 ng (0.020 pmol)
Insert DNA (1 kb)	37.5 ng (0.060 pmol)
Nuclease-free water	to 20 µl
T4 DNA Ligase	1 µl

   ** The T4 DNA Ligase Buffer should be thawed and resuspended at room temperature.*
2. Gently mix the reaction by pipetting up and down and microfuge briefly.
3. For cohesive (sticky) ends, incubate at 16°C overnight or room temperature for 10 minutes.
4. For blunt ends or single base overhangs, incubate at 16°C overnight or room temperature for 2 hours *(alternatively, high concentration T4 DNA Ligase can be used in a 10 minute ligation)*.
5. Heat inactivate at 65°C for 10 minutes.
6. Chill on ice and transform 1-5 µl of the reaction into 50 µl competent cells.