IDENTIFICATION AND EXPRESSION OF OAT HEMOGLOBIN

Master Thesis, Department of Pure and Applied Biochemistry

Supervisor: Nélida Leiva-Eriksson

Contact: nelida_leiva.eriksson@tbiokem.lth.se, Tel: +46(0)462228259

Examiner: Leif Bülow

Contact: leif.bulow@tbiokem.lth.se, Tel: +46-46-2229594

Duration: 20/8-2018 - 28/1-2019

Simon Christensen
bte13sch@student.lu.se

Populärvetenskaplig sammanfattning

Möjlig lösning av bristen på bloddonationer med hjälp av outforskat havrehemoglobin.

Hemoglobin är främst förknippat med blod i oss människor och andra djur, men tänk om någon berättade att det syretransporterade proteinet också finns i växter. Underligt, eller hur? Växter har ju inget blod men ändå finns det där. I denna rapport beskrivs identifieringen av tre nyfunna havrehemoglobin som har likande struktur som ris- och kornhemoglobin. Dessa har producerats i aktiv och fungerade form men det har visat sig svåra att separera från andra proteiner, åtminstone under detta arbetes korta tidsram.

I dagens samhälle är det få av oss som inte blivit uppmanade till att donera blod och detta med all rätta. I Skåne uppges det vara 3 av 100 personer som donerar blod och därför finns en mycket stor brist på vissa blodgrupper. Detta är även ett globalt problem, vilket Röda Korset återkommande informerar om. Denna brist innebär sämre överlevnadsmöjligheter för människor som är i akut behov av nytt blod, som vid en allvarlig olycka eller operation. Det kan handla om allt från en bilolycka till en oplanerad operation där det är bråttom med färskt blod för att hjälpa kroppen att återställa sig. Detta kan hända vem som helst när som helst, vilket innebär att detta problem är inget man skämtar bort eller kan blunda för. Tänk den ökade tryggheten med vetskapen att om något oförutsägbart händer mig eller närstående, så finns det en universell lösning på detta problem. Ett sätt att hjälpa hjärtat fortsätta pumpa ut syre i kroppen och låta det kämpa vidare. Detta behöver inte vara en utopisk tanke. Och tänk om denna lösning kommer från växtriket. Vad konstigt, men spännande!

För runt 70 år sedan upptäcktes det första hemoglobinet inom växtriket, alltså levande varelser utan blod. Sedan dess har mycket forskning lagts på att utforska detta relativt nya forskningsområde och flera olika typer av hemoglobin har hittats bland olika växter. Men varför skulle detta vara intressant? Tänk om denna nya typ av hemoglobin har egenskaper som skulle kunna utnyttjas som ett tillfälligt hjälpmedel vid blodbrist. En intressant kandidat till detta är havrehemoglobin, dels för att detta aldrig tidigare har utforskats och dels för de andra hälsofrämjande effekter havre har visats sig ha. Därför skulle de tre aktiva hemoglobin som beskrivs i denna rapport vara en grund för tillämpningen av växthemoglobin, där ett fungerande blodsubstitut för alla vore målet att uppnå. Ett alternativ kan även vara att använda proteinet som ett kosttillskott, vilket skulle ge en ny och utmärkt källa av protein och järn i ett.

Även om denna rapport har lagt grunden till framtida möjligheter att arbeta med havrehemoglobin, krävs det mycket tid och resurser innan denna utopiska lösning också blir verklighet. Detta med tanke på att i stort sätt bara identifiering, kloning, transformation och försök till upprening är det som har gjorts med dessa hemoglobin i nuläget. Men även om vägen till målet må vara lång och kantad med utmaningar är detta en möjlighet som måste utforskas. Varje framtida liv som kan räddas från en död på grund av bristande blodtillgångar vore en vinst och motivation nog. Det dags att låta havre vara basen för en ny livräddare och förse våra lungor med morgondagens syre.

Acknowledgments

This thesis was carried out at the Pure and Applied Biochemistry Department at the Kemicentrum of Lund University. I am grateful for this opportunity that was offered to me, to carry out this interesting and important project involving oat hemoglobin, which has never been done before. To be thrown into this unexplored area has both been scary and a great learning experience, where my knowledge in biochemistry has certainly expanded.

I would first like to thank Professor Leif Bülow for giving me this opportunity and setting up this project from the start. I have always felt welcome to talk to him about academic work and future career goals, which I have always considered valuable. I am also very thankful towards my supervisor Dr. Nélida Leiva-Eriksson, which has been providing constant support during the course of the project. With her great experience in this area, it has truly been a learning experience going into the world of hemoglobin research. We have had great discussion how to move forward throughout the project which I value highly. I would also give a special thanks to Karin Kettisen, who has helped me out in the lab when I felt lost. A big thank you also goes out to the rest of the department for creating a good working environment and made me feel welcome from the start.

Last but not least, I would like to thank my loving family and friends. I have always got great support and understanding from my loved ones from back home, especially from mom, dad and my siblings. Also, without my wonderful friends that I have got to known during over five years of studies, I would have never made it. I am extremely grateful to be a part of your lives and I hope you know how much I appreciate you all. Thank you!

Abstract

Hemoglobin is mainly viewed as an oxygen and carbon dioxide carrier in the blood of animals and thus essential for life as we know it. During the course of evolution, plants have also developed versions of this protein, for similar and different purposes, such as regulation of nitric oxide levels. A lot of research has been put into characterization of hemoglobin from monocots, including rice, barley and maize. However, another monocot, oat, has been a hot topic during recent years within nutritional sciences due to its discovered health benefits, leading to intense research on this crop and several oat-based products on the market. Due to its complex genomic structure, the genome was just recently elucidated. Therefore, this thesis has been focusing on identifying and expressing oat hemoglobin for the first time and investigating its potential future use as a Hemoglobin-based oxygen carrier and/or nutritional additive. In essence, three non-symbiotic hemoglobin were found with high similarity to hemoglobin from rice and barley (~83% and ~92%, respectively). During condition optimization for the expression, high cell density (~3) was the most important parameter to achieve reduced hemoglobin, while precise concentrations of induction compounds were less important. Spectrophotometric assays revealed expression of active hemoglobin. For purification, both cation and anion exchange chromatography was tested without successful results. Therefore, optimization of the purification protocol is must moving forward for future characterization of the hemoglobin. Despite this, these hemoglobin are a new and interesting research area that could be a benefit for modern medicine and nutritional sciences.

Sammanfattning

Hemoglobin är främst sedd som en bärare av syre och koldioxid i blod och är därför livsviktig för liv som vi känner till det. Under evolutionen har även flera växter utvecklat versioner av detta protein, både för liknande and andra anledningar som reglering av halten kväveoxid i växten. Mycket forskning har fokuserat på att karaktärisera hemoglobin från monocotyledon, där bland annat ris, korn och majs ingår. Dock har havre, en annan av denna typ, blivit ett hett samtalsämne under senare år tack vare dess hälsogynnande egenskaper, vilket har lett till intensiv forskning på denna planta och flera havre-baserade produkter på marknaden. På grund av dess komplicerade genomiska struktur har dess genom bara nyligen blivit sekvenserat. Detta examensarbete har därför fokuserat på att identifiera och uttrycka havre hemoglobin för första gången och undersöka dess potentiella användning som ett framtida hemoglobin-baserad syre bärare och/eller kosttillskott. Huvudsakligen identifierades tre icke-symbiotiska hemoglobin med högt överensstämmande med hemoglobin från ris (~83%) och korn (~92%). Under förhållandeoptimeringen för protein expressionen visade sig att hög celldensitet (~3) var den enskilt viktigaste parametern för att producera reducerat hemoglobin, medan exakta koncentration av induktionsämnen var av less betydelse. Spektrofotometriska analyser visade också att aktivitet hos det uttryckta hemoglobinet. Gällande uppreningen av hemoglobinet gav både kat- och anjon utbyteskoreografi inga lyckade resultat. Därför är optimering av uppreningsstegen ett måste i nästkommande fas för en framtida karaktärisering av hemoglobinet. Trots detta är dessa nyupptäckta hemoglobin ett nytt intressant forskningsområde som kan bli ett framtida framsteg inom den moderna medicin- och nutritionskunskapen.

Abbreviations

ALA	δ-Aminolevulinic Acid
CO	Carbon Monoxide
Fe^{2+}	Ferrous oxidation state of Iron
Fe^{3+}	Ferric oxidation state of Iron
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Hb	Hemoglobin
HBOC	Hemoglobin-based Oxygen Carriers
K_H	Hexacoordination Equilibrium Constant
k_{off}	Oxygen dissociation constant
NO	Nitric Oxide
nsHb	non-symbiotic Hemoglobin
OD_{600}	Optical Density at 600 nm
Sc233	Truncated Hemoglobin from scaffold 233
Sc445	Truncated Hemoglobin from scaffold 445
Sc485	Oat Hemoglobin from scaffold 485 (AsHb1z)
Sc1780	Oat Hemoglobin from scaffold 1780 (AsHb1x)
Sc313051	Oat Hemoglobin from scaffold 313051 (AsHb1y)
sHB	symbiotic Hemoglobin
trHb	truncated Hemoglobin

Table of Contents

1. Introduction

1.1 Hemoglobin: Structure and function

An essential basis for life among vertebrates is the transportation of oxygen from the surrounding environment to different cells in an organism and exportation of carbon dioxide and hydrogen ion in the opposite direction. In these organisms, this task is carried out by the protein hemoglobin (Hb), which increases the carrying capacity of mentioned ligands in a reversible manner (Jensen *et al.*, 1998). Due to the importance of hemoglobin, this is one of the most heavily studied proteins in vertebrates.

In order for hemoglobin to carry out these life essential functions, the protein has certain important characteristics. The general human variant consists of two identical α chains and two identical β chains forming two α-β dimers that are connected to each other, thus occurring as a tetrameric protein (Stryer *et al.* 2012). The structure of human hemoglobin can be seen in **Figure 1**, where the α- and β-chains are shown in gray and yellow, respectively. A unique characteristic for proteins from the globin superfamily, as hemoglobin originates from, is the α-helical secondary structure. This structure consists of 7 α-helices (helix A-H) and are arranged in a so called 3-on-3 sandwich fold between helices A, B, C and E over F, G and H (Vazquez-Limon *et al*, 2012). The mentioned type of structure is called the myoglobin fold and is conserved throughout the family of Hbs and similar proteins (Leiva-Eriksson *et al*, 2014).

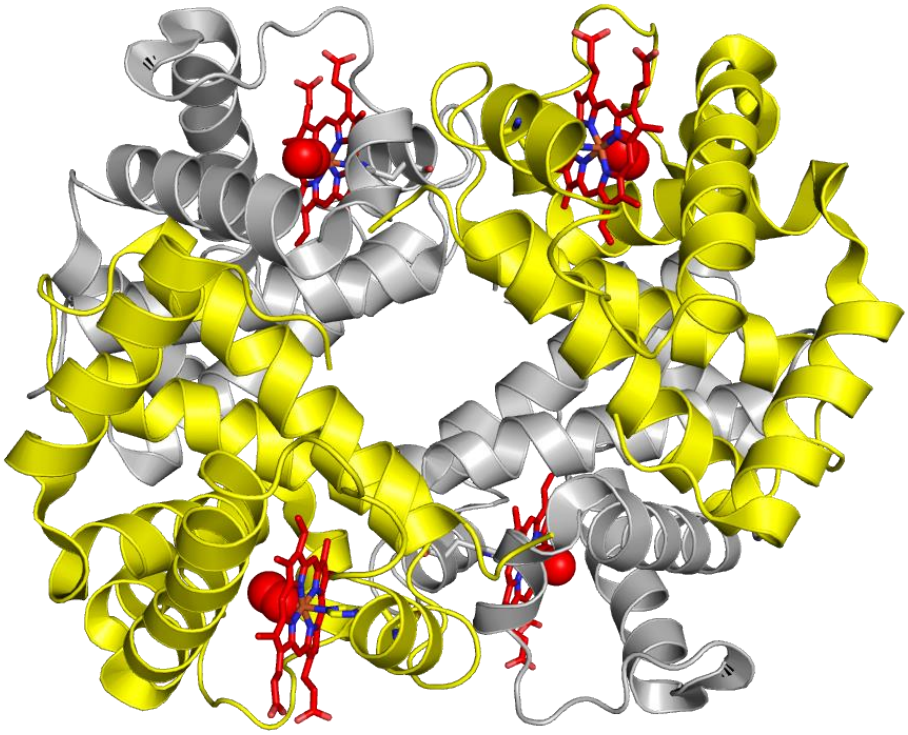


Figure 1. Structure of human hemoglobin A (HbA). The α- and β-chains are shown in gray and yellow, respectively. The four heme groups are represented by red and blue molecular structures with bound oxygen (red blobs). (Source: Alayash, A.I, 2012)

The reason why Hb can carry oxygen and other ligands is due to the presence of a prosthetic group called *heme.* These are indicated by the red and blue molecular structure in **Figure 1**, where oxygen (red blobs) are bound. The group consists of four pyrrole rings linked by methine bridges to form a tetrapyrrole ring with an iron atom centered in between these rings, which generates the red color of both muscles and blood among the animal kingdom. Different oxidation states is important in binding and transporting of the oxygen. For instance, ferrous oxidation state (Fe^{2+}) is capable of binding the oxygen while the ferric one (Fe^{3+}) is not (Stryer *et al*. 2012).

The iron atom can form two additional bonds on either side of the heme plane and these binding sites are referred to as the fifth and sixth coordination sites. In pentacoordinated Hb, five coordination sites of the iron is occupied without bound ligand. In contrast, all six coordination sites are occupied in hexacoordinated Hb. In both cases, the heme group is linked to the protein backbone by a histidine residue, termed the proximal histidine, and this residue has been shown to be important for the function of heme. In addition, the hexacoordinated Hb have an additional histidine residue in the binding pocket, referred to as the distal histidine (Stryer *et al.,* 2012). The distal histidine is crucial for stabilization of the bound oxygen by donation of hydrogen bond. Thus, hemoglobin can be classified by penta- or hexacoordination depending on the absence or presence of the distal histidine, respectively. Although, Hbs may have different degrees of distal histidine involvement in oxygen stabilization, giving them features from both classes. **Figure 2** displays the interactions of the proximal and distal histidine.

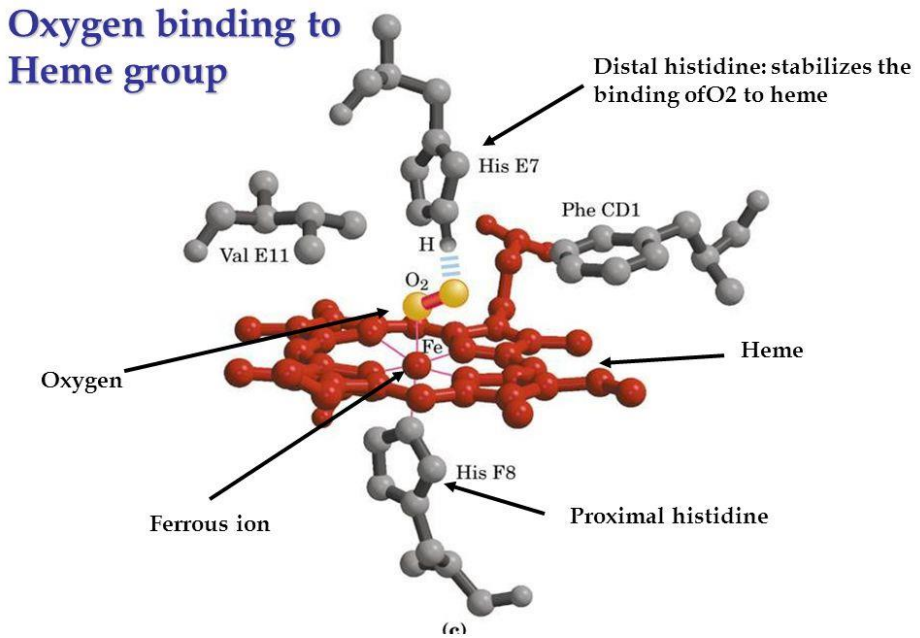


Figure 2. Interaction of the proximal histidine (penta- and hexacoordinated hemoglobin) and distal histidine (hexacoordinated) with the heme group. (Source: Chapman, 2016)

Furthermore, the biosynthesis of the heme group is dependent on eight different enzymes, where four of them are located in the mitochondria. One of the most important precursors of heme is δ-aminolevulinic acid (ALA), which is synthesized from glycine and succinyl-CoA by the enzyme Aminolevulinic acid synthase (Flora, 2014). Thus, the presence of δ-aminolevulinic acid is crucial for heme biosynthesis and production of functioning hemoglobin.

1.2 Plant Hemoglobin: Classes and features

Hemoglobin was firstly isolated and analyzed from mammals but it has now been shown that the protein can be found in nearly all organisms, ranging from archaea, eubacteria and eukaryotes (Hoy and Hargrove, 2008). During recent years, the interest in hemoglobin from plants has increased due to their different properties in comparison to human hemoglobin. Plant hemoglobin can be divided into three classes: symbiotic hemoglobin (sHbs), non-symbiotic hemoglobin (nsHbs, class I and II) and truncated hemoglobin (trHbs). This classification and known hemoglobin of different plants can be seen in **Figure 3**.

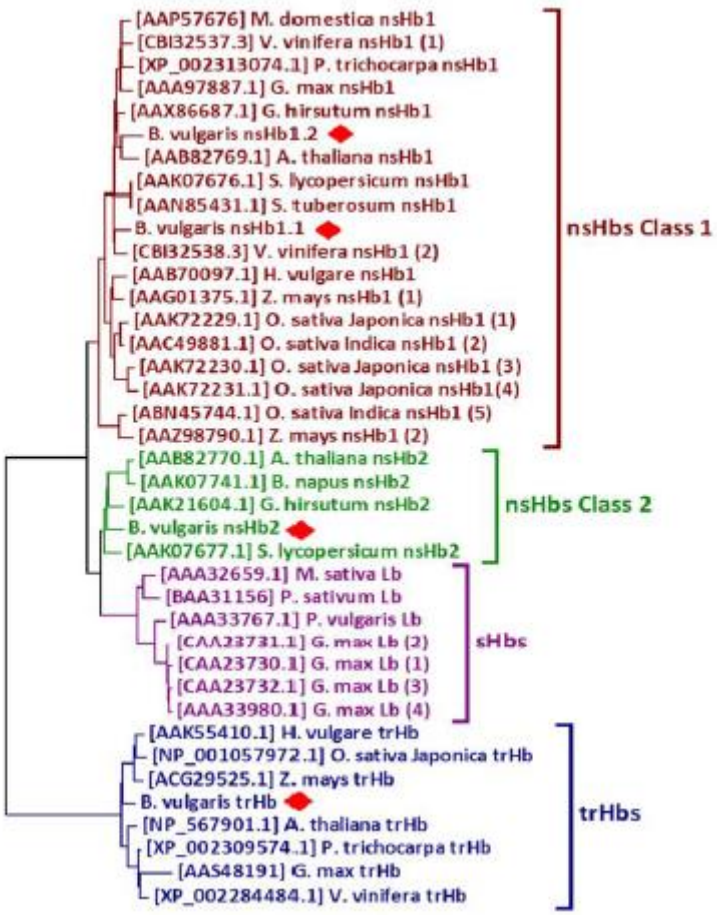


Figure 3. Classification of plant hemoglobin. The three major classes are non-symbiotic, symbiotic and truncated hemoglobin. (Source: Leiva-Eriksson, 2014).

1.2.1 Symbiotic Hemoglobin

The sHb exemplifies a class of hemoglobin that is mainly pentacoordinated. This kind is more adapted to effective oxygen buffering and delivery due to reversible binding in the pentacoordinated state which makes it suitable for storage and transportation of oxygen (Gupta *et al.*, 2011). Both symbiotic and non-symbiotic hemoglobin show high resemblance to human myo- and hemoglobin, which has been confirmed by structural analysis.

The sHbs can usually be found in legumes and were the first studied plant hemoglobin, now referred to as called Leghemoglobin. It can be found in root nodules of plants known to be nitrogen-fixing and shown to be important in the symbiotic nitrogen fixation (Gupta *et al.*, 2011). Thus, high levels of sHb in the nodules facilitates diffusion of the necessary oxygen needed nitrogen-fixation. Therefore, the function of sHbs has similar responsibilities as myoglobin in humans, which is in agreement with the close resemblance between these two proteins (Hoy and Hargrove, 2008).

1.2.2 Non-Symbiotic Hemoglobin

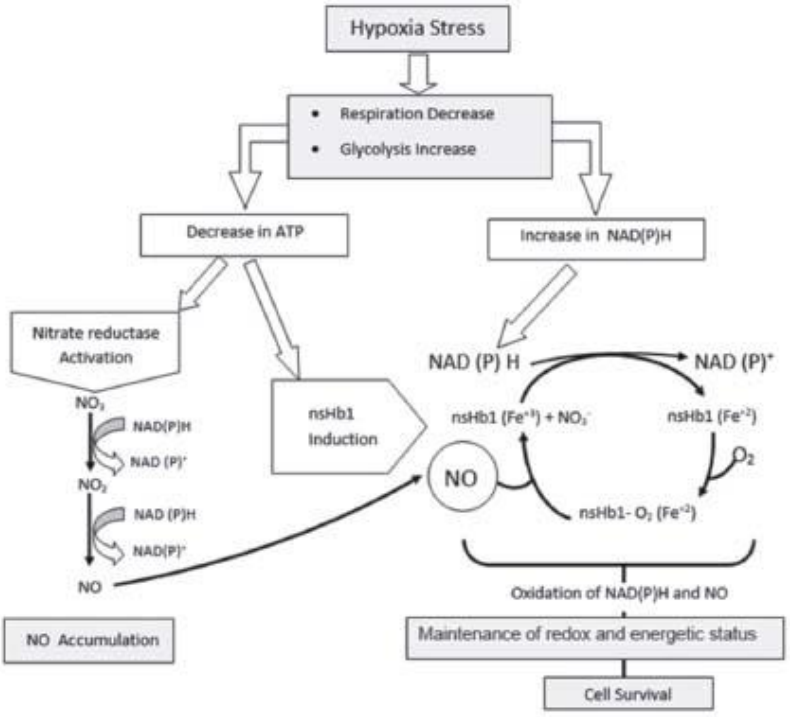
The more recently discovered nsHbs are divided into class I and class II where the most studied nsHbs can be found in monocots such as barley, rice and maize. It has been shown that nsHbs from plants are expressed differently under different stress conditions, such as hypoxia, nutrient deprivation and osmotic stress but also in response to nitric oxide (NO) and hormones. Thus, it has been suggested that nsHbs have a biological role in hormone signal transduction mainly by regulating the levels of NO produced under stress conditions. In addition, a feature of nsHbs is the presences of the classical myoglobin fold (Leiva-Eriksson *et al*, 2018).

1.2.2.1 Class I non-symbiotic hemoglobin

One of the most critical features of class I nsHbs that distinguishes them from other classes of Hb is a low value of the hexacoordination equilibrium constant (K_H). This constant describes the binding constant of the distal histidine, which allows equilibrium between the penta- and hexacoordinated states, facilitating ligand binding (Gupta *et al*., 2011). Due to this, these Hbs have a profoundly high avidity of oxygen that generates a tight and slow binding of oxygen, which is optimally designed for oxygen-dependent nitric oxide scavenging in low oxygen-containing environments.

One example is Hb1 in rice that has a 78-fold higher oxygen affinity when compared to Lb*a* in soybean. The reason for this is thought to be the low oxygen dissociation constant (k_{off}) suggesting oxygen stabilization of the gas when bound to this Hb (Garrocho-Villegas *et al*, 2007). Furthermore, the small oxygen dissociation rate points to the unlikeliness for class I nsHbs to function as sufficient oxygen transporters, but instead important components in electron transport due to their high oxygen affinity and redox potential, thus maintain the energy status during oxygen deprivation (Hoy and Hargrove, 2008) (Garrocho-Villegas *et al*, 2007). However, this reasoning is not universally accepted. Instead, recent research of nsHbs form sugar beet suggests that even if the Hbs have intrinsic properties to carry out certain enzymatic reactions, the location and concentration of ligands, such as oxygen, NO and nitrite will ultimately determine the activity of the Hbs (Leiva-Eriksson *et al*, 2018).

Even though the precise function of nsHbs is still under investigation, proposed models exist which tries to demonstrate the possible role of this class of Hbs. In **Figure 4**, one such model is presented. Here, hypoxia generates production of NO and induction of nsHb production. The nsHb ultimately oxidizes NAD(P)H and NO for maintenance of redox and energy status in the cells, promoting cell survival.



Modified from Dordas (2009).

Figure 4. Proposed model of function for nsHbs during hypoxia. During this condition, NO is accumulated and nsHb expression induced. The Hb ultimately binds oxygen and oxidizes NAD(P)H and NO to maintain the redox and energy status of the cell, promoting survival (Source: Riquelme and Hinrichsen, 2015).

1.2.2.2 Class II non-symbiotic hemoglobin

As of today, little is known regarding the biochemistry of class II nsHbs where few of these has been characterized. However, it has been shown that this class is not commonly expressed in hypoxic conditions, unlike class I nsHbs. In addition, class II nsHbs generally have a tighter hexacoordination (high K_H) in comparison to class I, leading to a lower oxygen affinity which makes them less suitable for NO scavenging. In contrast, these properties indicates possible functions in sensing low oxygen levels, as well as oxygen storage and diffusion (Gupta *et al*., 2011)(Hoy and Hargrove, 2008).

1.2.3 Truncated Hemoglobin

In the trHbs, the usual "3-on-3" α- helical sandwich fold, found in the other hemoglobin classes, is truncated into a "2-on-2" version of the protein and has been found in a range of different plants. Some trHbs have shown to be in a pentacoordinated state when oxygen is bound which forms an intermediate hexacoordinated structure after reduction and deoxygenation, followed by a regression to the pentacoordinated state again. This class of hemoglobin has been conserved throughout evolution and can be found in many different

types of plant. However, the precise function of these truncated versions is still unclear (Gupta *et al*., 2011).

1.3 Possible usage of Hemoglobin

In today's society, there is a great need for excess blood and blood substitutes that can be used in acute situations e.g emergencies, chronic or acute pathologies or during surgical interventions (Mozzarelli *et al.*, 2010). In these situations, blood transfusion might be a life-saving measure providing sufficient oxygen transport to the organs and thus reducing the risk of dying of blood loss. However, due to health risks coupled to transfusions and the huge need of new and suiting blood for each individual, there are a need of technologies to obtain these oxygen-carrying properties in another way. One alternative is to use Hemoglobin-based oxygen carriers (HBOC), which have been engineered to replace or enhance the oxygen-carrying properties of hemoglobin in erythrocytes (Silkstone *et al.*, 2016) and in this way reducing the need of new blood donors. In low doses, these oxygen carriers are thought to have the ability to reach and deliver oxygen to sites in the body where red blood cells have difficulties reaching due to different complications, such as regional tissue hypoxia, abnormalities of red cell deformability or disordered microvascular circulation (Cooper *et al.*, 2018). However, complications such as oxidative stress *in vivo*, has been linked to usage of modern HBOCs, thus increasing the need for engineered versions of human hemoglobin or finding other sources of hemoglobin with more desirable properties.

In recent research, different engineering strategies has been deployed where mostly human hemoglobin has been modified to obtain more suitable properties for use as a HBOC e.g increased stability and reduced oxidative toxicity (Silkstone *et al.*,2016). In addition, some plant hemoglobin has been shown to have desirable qualities in terms of stability and oxygen-carrying capacity (Hoy and Hargrove, 2008), which would make them interesting candidates to be used for this purpose.

1.4 Interest in Oat-based products

A possible way to obtain a new type of HBOC is to use hemoglobin from oat (*Avena sativa*). Oat as a grain has several health benefits including, hypocholesterolaemic and anti-cancerous properties as well as high content of dietary fiber, phytochemicals and nutritional value (Rasane *et al.*, 2015), leading to several oat-based products available on the current market. In addition, oat are in close relationship with both barley and rice (Institute for Agriculture and Trade Policy, 2007), which have more extensively studied hemoglobin products. However, due to the hexaploid genome structure in oat it has been a difficult task to elucidate this genome until recent efforts (ScanOats, 2018). With this in consideration, even though no hemoglobin from oat have ever been identified and characterized, it is reasonable that oat contains hemoglobin in the same class as both barley and rice, which corresponds to mostly class I nsHbs and possibly some truncated versions. Therefore, there is an interest in identifying the hemoglobin genes in oat and express these in bacteria to compare the properties with more researched hemoglobin, both human and plant versions.

1.5 Aim of Thesis

The aim of the thesis is to identify the different genes in oat that code for different Hbs and later express the top candidates in *Escherichia coli*. This is done to investigate the level of expression and in what state (reduced or oxidized) the Hbs are being expressed in. These results would provide important information regarding how to optimize the cultivation for maximal expression of oat hemoglobin. Furthermore, if time allows, the next step is to purify the oat hemoglobin in order to characterize it in the future. This would provide new biochemical knowledge describing how this Hb differs from others regarding e.g stability and oxygen-carrying capacity. Since oat Hbs have not been yet investigated, by examining these protein new insights in this area may be provided. For example, the number and structure of the genes that code for hemoglobin in the oat genome and what classes they belong to. The end goal of this project in the future is to evaluate if the oat Hbs would be a potential source of a new Hemoglobin-based oxygen carrier and/or as a source of protein/iron additive in different foods.

2. Materials and Methods

2.1 *In silico* determination of Oat Hemoglobin genes

Three scaffold containing genomic sequences from the oat genome along with five transcripts was received from the ScanOat consortium. These were generated from different time periods during seed development and showed the highest scores in the first homology search with barley and rice Hbs. Therefore, these sequences were the top candidates to contain one or several Hb genes and were used to determine the exact sequence of the oat Hb genes. In addition, two additional scaffolds were also analyzed but where it was more likely to find other variants of hemoglobin genes e.g truncated hemoglobin.

In the *in silico* analysis, several different tools was used to obtain the most likely sequence of the Hb genes. First of, ExPaSy translation tool (SIB, 2018) was deployed in order to analyze possible Hb in the different transcripts. These results were compared to a global alignment using GeneWise (EBI, 2018), where the generated peptides were aligned with the different genomic scaffolds. In addition, several tools specialized in finding exon-intron boundaries was needed to obtain these exact boundaries in a more statistically reliable way. The primary tools used were Augustus (University of Greifswald, NA), GenScan (MIT, 2005) and Fgenesh (Softberry, 2018). Finally, alignments of individual DNA and protein sequences was done in Clustal Omega (EBI, 2018) during the entire process. This tool was also used when comparing candidate Hb genes with known sequences of the corresponding protein in barley and rice.

In addition, the online tools Phyre^2 (Kelley *et al*, 2015) and I-Tasser (Roy *et al*, 2010) was used to predict the three-dimensional structure of the Hbs. These models base their estimation on the amino acid sequence as well as comparison with more studied proteins with similar characteristics e.g oxygen transport.

2.2 Transformation and cloning

The oat Hb genes were synthesized and obtained from Integrated DNA Technologies®. Prior to transformation, the BP reaction was performed. In this reaction, the Hb genes were inserted into the entry vector by homologous recombination in exchange for a lethal gene. In addition, kanamycin was used as a selective marker. The cells used for transformation was Invitrogen OmniMax® chemically competent *Escherichia coli* cells and isolated colonies were used to perform plasmid isolation to ensure the presence of the correct plasmid. The sequencing of the plasmids was performed by GATC©. In the following step, the Hb genes were inserted into a destination vector which will later be used as expression vector. This was done by the LR reaction, where the lethal gene was substituted for the Hb genes again. This time, the selective marker was ampicillin and the surviving colonies were used for plasmid isolation/sequencing as before. The complete protocol for the Gateway recombination cloning and transformation was provided by Invitrogen© (Invitrogen, 2012). The protocol was followed except following steps:

- Both BP and LR reactions were incubated at 25 Cº overnight.
- After LR reaction, the complete 10 µl was used in the transformation.
- 25 µl (half vial) of OmniMax® *E. coli* cells was used for BP reaction transformation.
- 100 µl (half vial) of chemically competent BE21-DE3 cells was used for LR reaction transformation.
- Three glycerol stocks of cells containing each hemoglobin producing plasmid were made by mixing 0.24 ml glycerol with 1.36 ml cell culture and were stored in -80 ºC.

2.3 Plasmid Isolation

For plasmid isolation, the Plasmid DNA Purification kit from Macherey-Nagel was used (Macherey-Nagel, 2012). The protocol was followed except the following steps:

- The samples were centrifuged at 11000x g for 15 minutes to separate plasmids from remaining cell components.
- The silica membrane was dried from ethanol by 11000x g centrifugation for 3 minutes.
- The elution of plasmid DNA was performed twice, with the same 50 µl elution buffer.

2.4 Expression of hemoglobin

The oat Hbs were expressed in *E. coli* (BL21-DE3) using the pET-DEST42 expression vector. Each gene (485, 1780 and 313051) was separately inserted into this vector and used for expression. The Hb expression protocol consisted of 4 steps: inoculum preparation, fermentation, cell harvesting and protein extraction. In addition, glycerol stocks of BL21-DE3 cells were made and can be used as starting cultures at later stages of the project.

2.4.1 Inoculum preparation

For the preparation of the inoculum, one colony is selected from a plate containing expression clones and used to inoculate 5 ml of LB broth containing ampicillin (100 µg/mL). Cells are grown overnight (16-18 h) at 25 °C and 100 rpm.

2.4.2 Fermentation conditions

When doing initial tests in the fermentation phase, 100ml TB media was prepared in 500ml baffled E-flasks (two for each hemoglobin gene). Carbenicillin was added to a final concentration of 100 mg/L and every vessel was inoculated with 1 mL of inoculum. Cells were allowed to grow at 37°C and 150 rpm for about 3-4 hours or until optical density (OD_{600}) reached a value equal or slightly higher than 2.5 units. Hb expression was induced by adding IPTG and delta-aminolevulinic acid (δ-ALA) to a final concentration of 0.5 mM and

0.3 mM, respectively. The flasks were bubbled with carbon monoxide for about 5-10 seconds and sealed with parafilm. Cells were subsequently cultivated at 22 °C and 150 rpm overnight (16-18 h) under dark conditions.

This was also done for the main production, but with 500 ml in 2L E-flasks where 5 ml inoculum was added. In addition, the flasks was bubbled with carbon monoxide for 25 seconds.

2.4.3 Cell harvesting

At the end of the cultivation, the fermentation media was centrifuged (10 000 rpm, 10 min) in a Sorvall RC-5B centrifuge (DuPont) to harvest the hemoglobin-containing bacterial cells. The supernatant was thoroughly removed and the pelleted cells are washed with 60 mL (depending on amount of cells obtained) of buffer Tris-HCl 50 mM pH 8.5 and collected into a smaller tube. The cells were again centrifuged (4000 rpm, 40 min). The supernatant was discarded and the weight of the pellet recorded. The harvested cells were flash frozen with liquid nitrogen and stored at -80 °C for later use. All centrifugation steps were performed at 4 °C

2.4.4 Protein extraction

The harvested cells were thawed and gently re-suspended in lysis buffer. For every gram of biomass, 2 mL of buffer was added. The suspended cells were kept on ice and disrupted with three consecutive 2 minute-long cycles of sonication using a Q500 ultrasonicator (QSonica) working at 70% (depends on sample size) amplitude. Cell debris was removed by mean of centrifugation (12 000 rpm, 40 min, 4 °C) using a Sorvall RC-6 Plus centrifuge (Thermo Electron Corporation). The supernatant was saved as this is where Hb is expected to be found.

2.4.5 Hemoglobin purification

The harvested cells were thawed and resuspended in lysis buffer. For every gram of biomass, 2 mL of buffer was added. The suspended cells were kept on ice and disrupted with three consecutive 2 minute-long cycles of sonication using a Q500 ultrasonicator (QSonica) working at 40%. The resulting mixture was centrifuged at 20,000 g twice at 20 min each, where the supernatants were collected after each centrifugation. An ÄKTA Explorer system was used for purification. In this case, cation exchange was deployed using a Capto S Hi Screen™ column. The supernatants were pre-dialyzed in starting buffer (10 mM NaP pH 6) and subjected to the system. The threshold for sample collection was 50 mAu at 412 nm. Collected samples were used in SDS-PAGE and activity analysis.

In the case of anion exchange, two different columns were used. These were BabyBio™ Q and BabyBio™ DEAE which work as a strong and weak anion exchange, respectively. The supernatants were pre-dialysed in starting buffer (50mM TRIS-HCl pH 8.5). Apart from the solution used, the procedures were the same as in the case of cation exchange.

2.5 Quantification of Hemoglobin

The concentrations of Hb were determined using a spectrophotometric assay (UV-Vis spectrophotometer Agilant Cray 60). The real absorbance used to calculate the Hb concentration was obtained by using the 3-point drop method. In this method, the absorbance of the three different points in **Figure 5** are used in Equation 1:

$$Real\ Absorbance = Abs\ point\ (c) - \frac{Abs\ point\ (a) + Abs\ point\ (b)}{2} \qquad Eq\ 1.$$

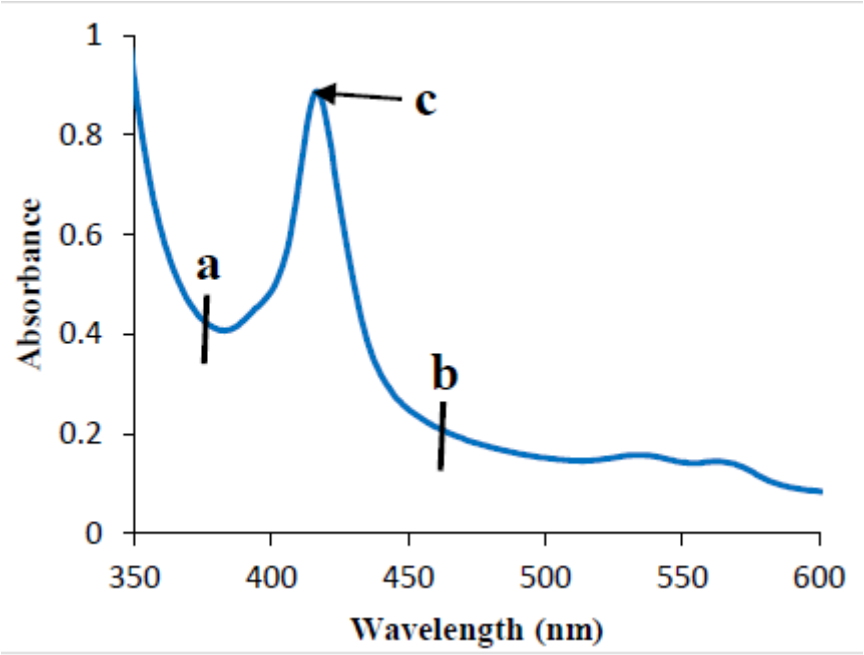


Figure 5. Three point drop method for determination of the real absorbance. The estimated absorbance for point a, b and c were put into Eq 1. to obtain the real absorbance (Chukwubueze-Okeke, 2017).

The calculated real absorbance at ~424 nm where used in Equation 2 to obtain the Hb concentration (mM), which can be used in Equation 3 to ultimately obtain the amount of Hb expressed:

$$Concentration\ (mM) = \frac{Real\ Abs * dilution\ factor}{\epsilon} \qquad Eq\ 2.$$

Where dilution factor is how many times the sample were diluted and ϵ is the extinction coefficient (mM^{-1} cm^{-1}), which was assumed to be 190 (measured coefficient for reduced barley Hb). The cuvette length is assumed to be 1 cm, resulting in the concentration in mM. The obtained concentration from Equation 2 can be converted into mg Hb by Equation 3:

$$Amount\ (mg) = \frac{Concentration * Mol\ Weight * Volume}{1000} \qquad Eq\ 3.$$

Where Mol Weight is the molecular weight (Da) of the protein, volume is the total volume (mL) of the supernatant and 1000 is a correction factor to get the correct unit. By using these equations, the amount of Hb produced was estimated.

2.6 Condition optimization of hemoglobin expression

In order to increase the possibility of maximal Hb expression, optimization of expression conditions were used. Here, optical density before induction, concentration of ALA and IPTG and temperature was altered to investigate which of these conditions would provide the most expression. The scaffold of 485 was used in the following conditions, where sample 1 was the negative control, thus with no induction of Hb expression and the flasks were incubated in 17ºC at 150 rpm overnight. The set up can be seen in **Table 1**.

Table 1. Condition optimization to increase expression levels of oat Hb. The "x" indicates what condition each culture was cultivated in. As before, every flask contained 100 mg/L carbenicillin and 1 ml inoculum.

	0.1 mM δ-ALA	**0.5 mM δ-ALA**	**0.1 mM IPTG**	**0.5 mM IPTG**	**1 mM IPTG**	**OD ~0**	**OD~1**
*1.						x	
2.	x		x			x	
3.	x			x		x	
4.	x				x	x	
5.		x	x			x	
6.		x		x		x	
7.		x			x	x	
*8.							x
9.	x		x				x
10.	x			x			x
11.	x				x		x
12.		x	x				x
13.		x		x			x
14.		x			x		x

In similar fashion, this was also done for 0.3 mM ALA and changing the amount of IPTG and optical density due to the obtained results. The corresponding set up can be seen in **Table 2**.

Table 2. Condition optimization to increase expression levels of oat Hb. The "x" indicates what condition each culture was cultivated in. As before, every flask contained 100 mg/L carbenicillin and 1 ml inoculum.

	0.3 mM δ-ALA	**0.2 mM IPTG**	**0.8 mM IPTG**	**OD ~0 min**	**OD~ 0.5h**	**OD~ 1h**	**OD~ 1.5h**	**OD~ 2h**	**OD~ 2.5h**	**OD~ 3h**
*1.				x						
2.	x	x		x						
3.	x		x	x						
4.	x	x			x					
5.	x		x		x					
6.	x	x				x				
7.	x		x			x				
8.	x	x					x			
9.	x		x				x			
10.	x	x						x		
11.	x		x					x		
12.	x	x							x	
13.	x		x						x	
14.	x	x								x
15.	x		x							x

2.7 SDS-PAGE form analysis of protein expression

The SDS-samples were prepared by thawing the cells and dissolve a small amount in LB medium. The OD_{600} was measured and 50 µl of cells were taken into a clean tube. These cells were centrifuged (6500 rpm, 5 min) and excess medium was discarded. Different amounts of 2x SDS loading buffer was added depending on the OD_{600}-values to a final OD of 10 in every sample. The samples were boiled (95ºC, 5min), vortexed and centrifuged. Mini-PROTEAN® TGX™ gels (15 well comb, 15 µl) were used were 5 µl of every sample was used to load the gels. The gels were analyzed using GelDoc XR system (BioRad). When similar samples were analyzed, one part of supernatant was added to one part 2x SDS loading buffer. These samples were loaded at 3-5 µl, depending of visualization of distinctive bands or not.

3. Results

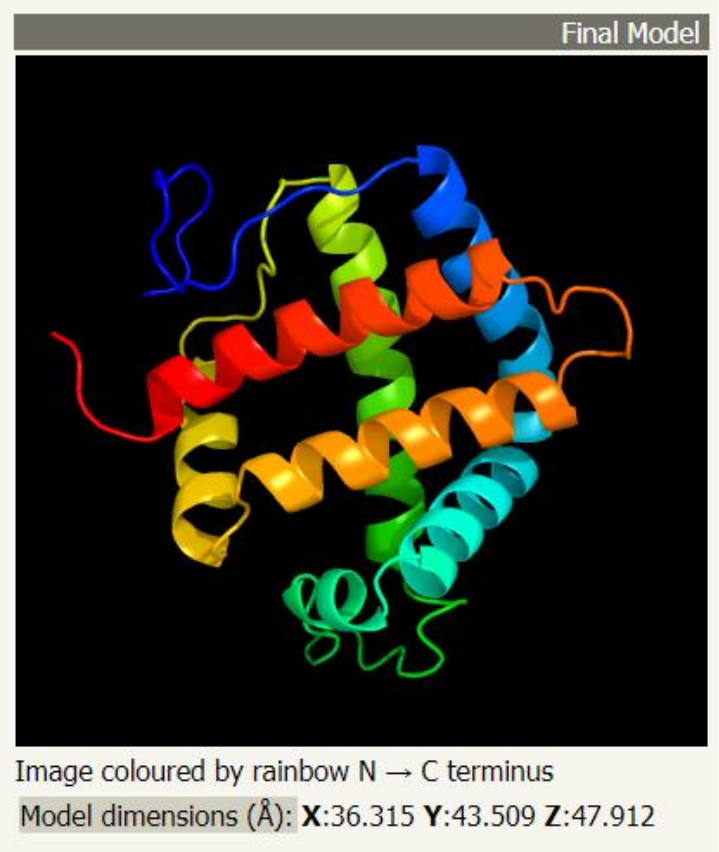
3.1 Hemoglobin genes from *in silico* analysis

In total, three Hb genes were found in the provided genomic sequences, one in each scaffold. In this section, the main findings are not presented here due to copyright complications. This includes the entire genes including exon-intron boundaries, the exons and the translated proteins. The gene and amino acid sequences are not shown here.

3.1.1 Scaffold 1780

Based on the online tools Phyre[2] (Kelley *et al,* 2015) and I-Tasser (Roy *et al*, 2010), the three-dimensional structure of the oat Hb could be predicted. These model use the amino acid sequence of the Hb and compare it with known proteins with similar characteristics. The predicted model of scaffold 1780 from Phyre[2] can be seen in **Figure 6**. Here, the characteristic α- helix motif consisting of 7 helices can be seen in close proximity. The confidence key (**Figure 6b**) indicates that the majority of the amino acids can be placed with high confidence, except for the ends of the N- and C-terminal. **Figure 6c** displays the predicted secondary structure, where every amino acid is placed in helices, strands or coils, and the predicted accessibility of the different amino acids. Other information regarding this model and the other scaffolds can be found in the **Appendix 7.6**. This includes top candidates which the model is based on as well as predicted ligand binding.

(a)



(b)

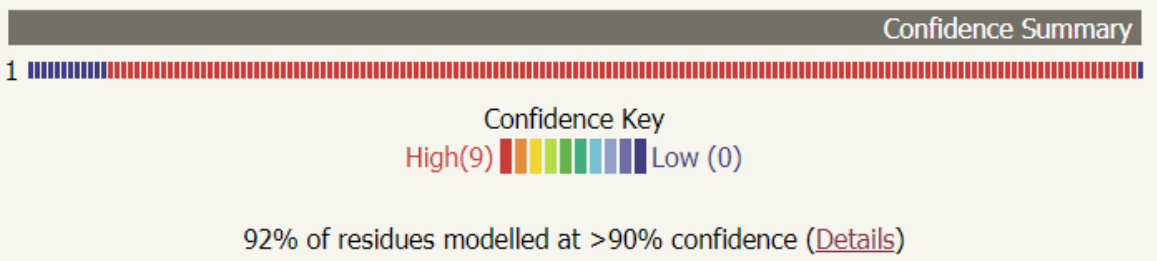


Figure 6. Predicted model of the oat Hb (a) and the confidence summary of this prediction (b) from Sc1780 (Kelley *et al*, 2015). In addition, the predicted secondary structure and accessibility of the amino acids (c) was generated by I-Tasser (Roy *et al*, 2010). These are not presented here.

3.1.2 Scaffold 485

Due to copyright complications, no DNA sequences from the ScanOat consortium or found amino acid sequence are not presented in the public version.

3.1.3 Scaffold 313051

Due to copyright complications, no DNA sequences from the ScanOat consortium or found amino acid sequence are not presented in the public version.

3.1.4 Truncated hemoglobin genes in scaffolds 233 and 445

In addition to the three Hb genes described above, two other potential candidates were also found in scaffolds 233 and 445. However, these genes encode shorter proteins in comparison to the three main candidates, 157 (17.31 kDa) and 143 (15.54 kDa) amino acids, respectively, and show substantially lower identity to corresponding barley and rice hemoglobin. Thus, these hemoglobin are probably examples of truncated hemoglobin present in the oat genome and are less interesting regarding known characteristic Hb properties.

3.2 Cloning and transformation

Three cloning steps and plasmid purifications were performed in order to create the expression clone used to reach a higher expression of Hb in *E. coli* BL21-DE3 cells. Results for creating the entry clone (BP reaction) and the first expression clone in OmniMax® cells (chemically competent *E. coli*), as well as the sequencing results of the plasmids, can be found in the **Appendix 7.2**. In this section, only the final expression clone is presented.

3.2.1 LR cloning in competent *E.*coli BL21-DE3 cells

The result from the final cloning step, where the expression vectors were transformed into competent *E.coli* BL21-DE3, can be seen in **Figure 8.** Here, defined colonies did appear on ampicillin-containing agar plate (right side) while the negative control (left side) did not show any sign of growth. These results in combination with the sequencing output of the plasmids (**Appendix 7.2)** indicate a successful cloning procedure. This was the case for every scaffold. Thus, these colonies were used as inoculum in the expression phase of the thesis.

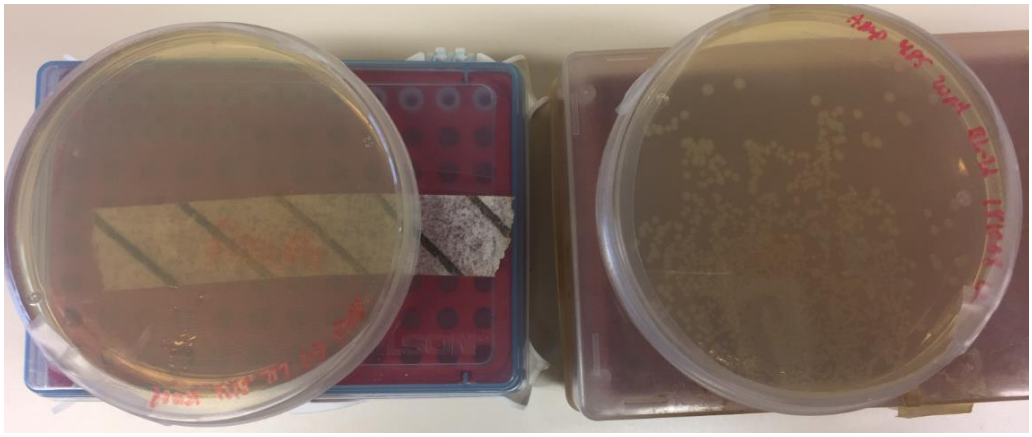


Figure 8. LR cloning in competent *E.coli* BL21-DE3 cells. The negative control (left side) showed no growth while the cells containing Hb vectors (right side) showed defined colonies.

3.3 Expression of oat hemoglobin

The expression of the oat Hbs were done by induction with IPTG and ALA. In the first attempt, the color of the cell pellets were white/brown, indicating that the expression of hemoglobin was low and the hemoglobin that was expressed was likely oxidized (**Appendix 7.4)**. However, SDS-PAGE of the supernatants from the initial cultures for all scaffolds revealed expression of oat Hb when compared to non-induced samples (-), which can be seen in **Figure 9.** Although, the amount of Hb seemed low according to this result and the non-induced sample also displayed a less defined band. Thus, condition optimization was deployed in order to increase the expression of Hb and obtain the characteristic pink/red color of non-oxidized protein.

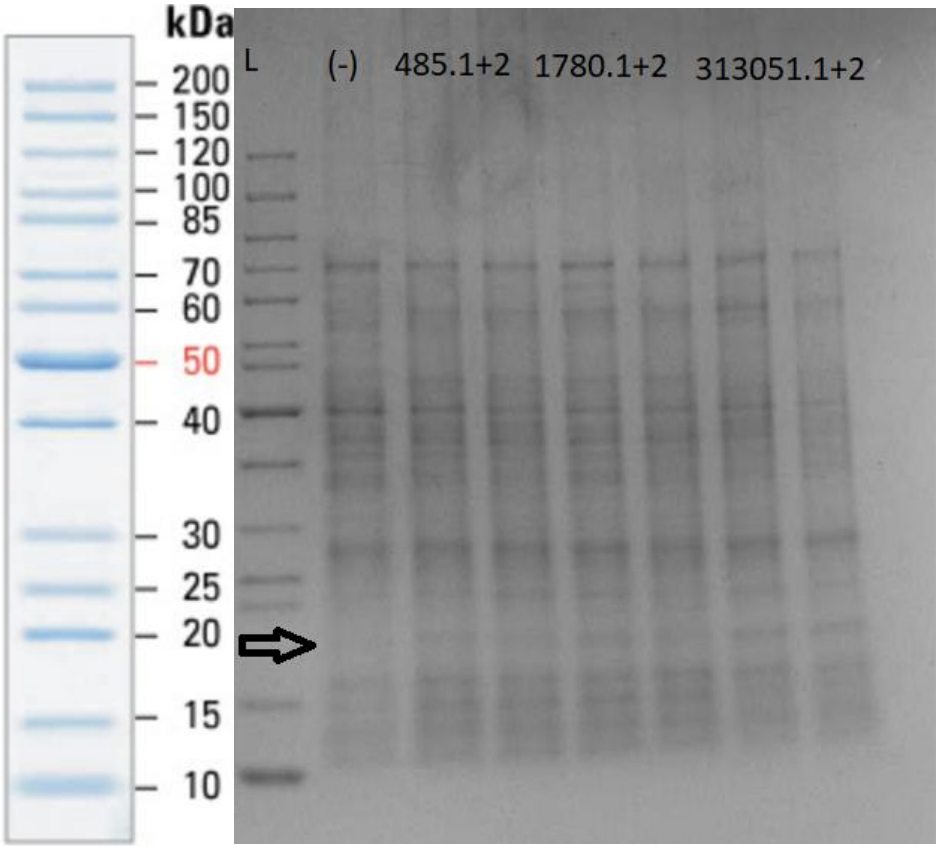


Figure 9. First around of condition optimization with all scaffolds. The cultures were induced at OD_{600}~2 with 0.3 mM of ALA and 0.5 mM IPTG. The arrow indicates the location of the projected band of hemoglobin, around 18 kDa.

3.3.1 SDS-PAGE of scaffold 485 initial cultures

In the first round of condition optimization of scaffold 485, the different cell cultures were induced at different OD_{600}, 0.1 or 0.5 mM ALA and 0.1, 0.5 or 1 mM IPTG (**Materials and Methods 2.5**). The resulting SDS-PAGE result is shown in **Figure 10**. The expected Hb band is located at approximately 18 kDa (indicated in the figure). As can be seen, higher OD_{600} (samples 8-13) tends to show a more distinctive band at this location in comparison with the samples induced at low OD_{600} (samples 2-7) and the negative control. In addition, this in combination with higher concentration of IPTG (samples 12 and 13 in particular) show the most distinctive band around 18 kDa. This result in combination with previous seen results (**Appendix 7.4.2**), suggests that the most important parameter to take into account are a high OD_{600} (>2) before induction but also IPTG concentration between 0.5-1 mM to increase the Hb expression. The precise concentration of ALA did not seem to have any significant impact on the expression.

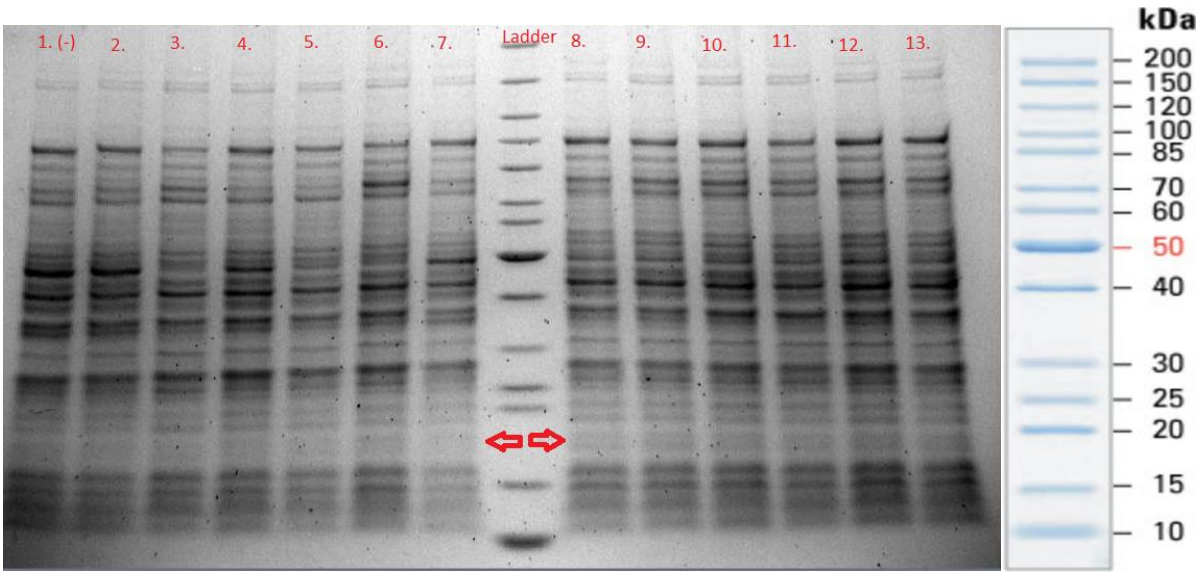


Figure 10. First around of condition optimization with scaffold 485. The cultures were induced at different OD_{600} with varying concentrations of ALA and IPTG. Apart from the negative control (sample 1), the remaining samples were induced with 0.1 or 0.5 mM ALA in combination with 0.1, 0.5 or 1 mM IPTG (see Materials and Methods section 2.5 for complete set-up).

3.4 Quantification of Hemoglobin

In order to determine Hb concentration and activity, a spectrophotometric assay analyzing the absorbance at different wavelengths was used. Here, the absorbance for unaltered Hb, reduced Hb (+NaD) and reduced Hb bubbled with carbon monoxide were measured for all scaffolds. This was done for unpurified Hb (supernantants) and purified Hb (**Materials and Methods 2.4**).

3.4.1 Quantification of unpurified hemoglobin

The absorbance at different wavelengths for scaffold 485 and determined Hb concentration can be seen in **Figure 11** and **Table 3**, respectively. Two peaks can be distinguished, located at 560 and 424 nm, where the most distinctive one was found at 424 nm. The amount of expressed Hb was estimated to 0.573 mg in this culture containing 5.735 g of cells (based on the +NaD curve). In addition, there was a change in maximum absorbance between the standard, reduced and CO-bound Hb. For instance, a 1 nm shift between the +NaD and +CO curves. Similar results were observed in the remaining scaffolds (**Appendix 7.5)**. Also, the negative control (no induction) showed Hb expression but to less extent in comparison with most of the induced cultures.

Table 3. Analyzed data from spectrophotometric assay. Among the different parameters, the real absorbance was determined using the 3-point drop method, leading to the estimated Hb concentration and amount produced.

Scaffold	Max peak +NaD (nm)	Max peak +CO (nm)	Real Abs (+NaD)	Real Abs +CO	Shift Nad-CO (nm)	Shift NaD-Standard (nm)	Conc. (mM, based on +NaD)	Conc. (mM based on +CO)	Amount (mg) (based on +CO)
485	424	423	0.114	0.105	1	1	0.00241	0.0020	0.573 (0.496)

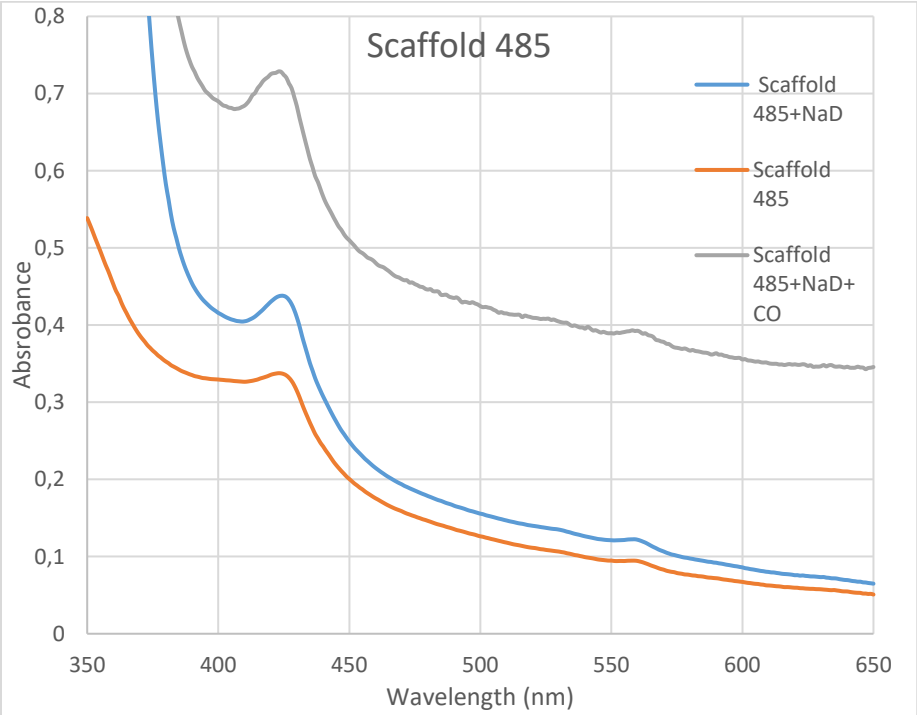


Figure 11. Absorbance as a function of wavelength for Sc485. Three curves were generated, standard Sc485 (nothing added, red curve), added NaD (blue) and added NaD bubbled with CO (gray). In all cases, two peaks could be seen, located at ~424 and ~560 nm.

3.4.2 Quantification of purified Hb using cation exchange

In the first attempt to purify the oat Hbs, cation exchange by the CaptoS HiScreen™ was used. The most important data for the attempted purification of Sc485 can be seen in **Table 4** and **Figure 12**, where the unpurified and fractionated Sc485 can be found. Both fractionated and unfractionated Hb shifted the location of the maximum peak, as was seen in previous results. In addition, the calculated Hb concentration and amount differs substantially between these two samples. The measured amount of the fractionated Hb was 46.3% in comparison with the untreated sample, meaning that 53.7% of Hb was bound to the column. However, this

fraction was not collected during the elution phase of the purification, pointing to that the remaining 53.7% were still bound to the column after elution.

Table 4. Analyzed data from spectrophotometric assay for Sc485 using cation exchange. Here, calculated real absorbance, concentrations and amounts of Hb for fractionated and unfractionated Sc485 can be found.

Scaffold	Max peak +NaD (nm)	Max peak +CO (nm)	Real Abs (+NaD)	Real Abs +CO	Conc. (mM, based on +NaD)	Conc. (mM based on +CO)	Amount (mg)	Difference Abs (for NaD)	Percentage in column (%,for NaD)
485	426	422	0.0793	0.0577	0.00209	0.00143	0.404	0.0537	53.7
Fraction 485	422	424	0.0267	0.0298	0.00070	0.00074	0.186		

This difference in amount of Hb can also be seen in **Figure 12**, where the three top curves belonged to the unpurified Sc485, while the three lower curved belonged to the fractionated Sc485. Altogether, the results indicated that the bound Hb was eluted during the cleaning of the column instead of in the elution phase, leading to this uncollected fraction of Hb.

In addition, the data for Sc1780 and Sc313051 indicated that the large majority of the Hbs went through the column (~99% and ~86%, respectively) and no Hb was detected in the elution phase of the purification (**Appendix 7.5**). Lastly, the shift in wavelength between untreated, reduced and CO-bound Hb was more clearly detected here, indicating presence of active Hb.

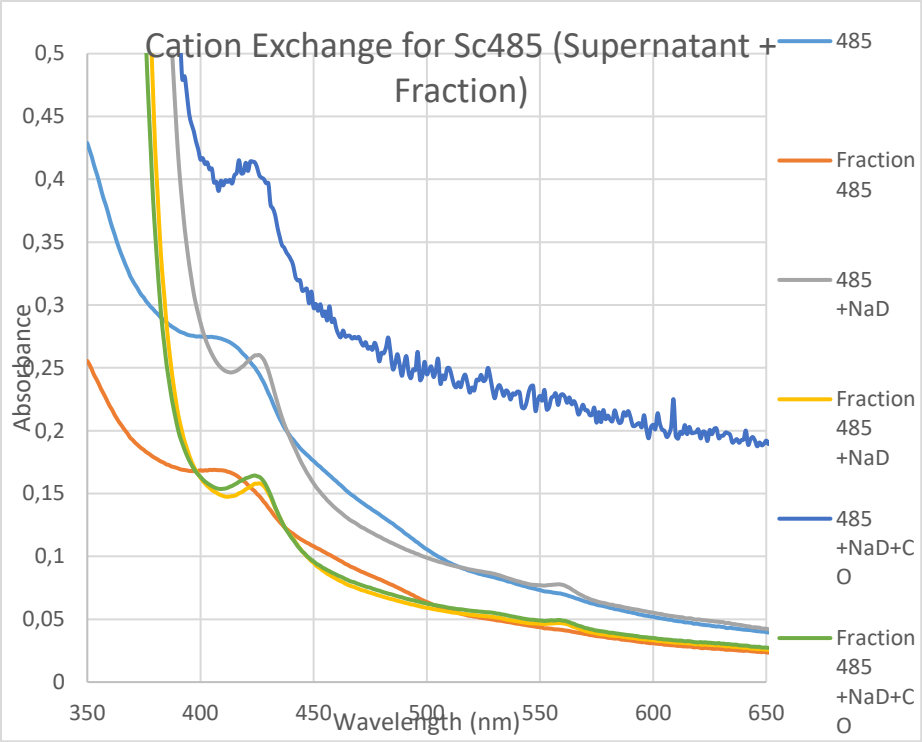


Figure 12. Absorbance as a function of wavelength for the cation exchange of Sc485. The three top curves (light blue, gray and dark blue) belonged to the unpurified Sc485 (no cation exchange) and the three lower curves (red, yellow and green) belonged to the fractionated Sc485.

3.4.3 Quantification of purified Hb using anion exchange

In addition to the attempted purification of oat Hb by cation exchange, anion exchange was also used. In this case, two different columns were used, a strong (BabyBio™ Q) and a weak (BioBaby™ DEAE) anion exchanger were deployed. The calculations were based on the +NaD extinction coefficient as before. All plots not presented here for every scaffold can be found in **Appendix 7.5.**

3.4.3.1 Quantification of purified Sc485 using anion exchange

The calculated data for the untreated and fractionated Sc485 as well as the resulting plot from the spectrophotometric assay can be seen in **Table 5** and **Figure 13**, respectively. The fractionated samples were divided into eluted Sc485 (eluted during the elution phase) and run-through sample (fractionated before the elution phase). The majority of Hb was collected as a run-through fraction (72.2%) for both columns in the case of Sc485, while they differ in the eluted fraction. The remainder of Hb was found in the eluted fraction for column Q (27.8%), but Hb was found in both the eluted fraction (21.1%) and after cleaning of the column (6.7%) for DEAE. Thus, more Hb were collected in the elution phase using anion exchange in comparison with previous results from the cation exchange, although the Hb was still eluted with other proteins present in the supernatant, not resulting in purified Hb.

Table 5. Analyzed data from spectrophotometric assay for Sc485 using anion exchange. Real absorbance, concentrations and amounts of Hb for the strong (Q) and weak (DEAE) anion exchanger can be found, as well as different percentages of the location of the detected Hb.

Sample	Max peak (nm)	Real Abs (+NaD)	Conc (mM)	Amount (mg)	% Eluted (DEAE)	% Eluted (Q1)	% Through (DEAE)	% Through (Q1)	% In column (DEAE)
485	409	0.0503	0.00132	0.727	21.1	27.8	72.2	72.2	6.7
485 Elute DEAE	412	0.0158	0.00042	0.153					
485 Elute Q	412	0.0208	0.00055	0.202					
485 Through DEAE	412	0.0363	0.00096	0.525					
485 Through Q	412	0.0363	0.00096	0.525					

These results are also evident in the spectrophotometric plot for the different fractions, shown in **Figure 13**. The sample not used in the anion exchange (485 Supernatant, blue) show substantially higher absorbance in comparison to the eluted fractions (red and grey curves), indicating that the majority of the Hb is not collected during the elution phase. Instead, this majority was detected in the run-through fractions (yellow and dark blue curves) for both columns.

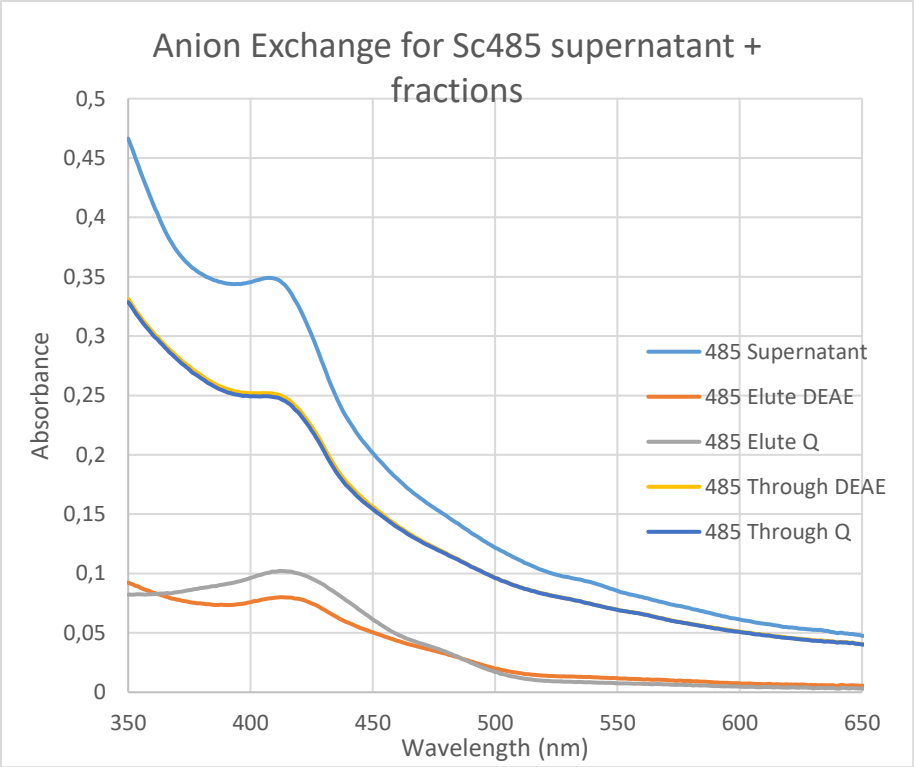


Figure 13. Absorbance as a function of wavelength for Sc485 using anion exchange. The curves for the eluted fractions (red and grey), run-through fractions (yellow and dark blue) and the untreated supernatant (light blue) can be seen in this figure. It should be noted that the yellow and dark blue curves are closely intertwined.

4. Discussion

4.1 *In silico* analysis

In order to determine the amino acid sequence for the present Hb in oat, *in silico* analysis was necessary. Here, transcripts of developing oat seeds were compared with the recently sequenced oat genome and compared with other Hbs from monocots. In this analysis, three non-truncated Hbs, called Sc485, Sc1780 and Sc313051, alongside two truncated versions, Sc233 and Sc445, were identified (**Results 3.1** and **Appendix 7.1**). The three non-truncated oat scaffolds showed the highest similarity with nsHb from barley (~92%), while the same parameter for rice nsHb was ~83%. Thus, it is reasonable to believe that these three Hbs also belong to the class 1 of nsHb and have similar characteristics with these other monocotic Hbs, even if this have to be proved experimentally by characterization.

The number of Hbs found in oat can also be compared with other monocots. For instance, in rice (*Oryza sativa*) five class I nsHbs and one trHbs have been found, while a single class I nsHb and trHb has been found in barley (Rodriguez-Alonso and Arredondo-Peter, 2013). Other monocots also show different numbers of Hb genes, indicating there is no general number of Hbs. However, since oat has a complex hexaploid gene structure divided into three diploid genomes, it is possible that one nsHb has evolved from each of the diploid genomes. Therefore, it is also reasonable that there should be three trHbs for the same reason but the last one may have been neglected in this analysis. Thus, three nsHbs and two or three trHbs are reasonable numbers of Hbs found in oat, resulting from a long and complex evolution of plants likely originating from a low oxygen-containing environment.

In the simulation of the 3D-structure for the different scaffolds it became more evident that the found oat Hbs were probably class 1 nsHbs. This due to the fact that proposed model of Sc485, Sc1780 and Sc313051 were all mainly based on nsHb from mainly rice but also from other monocots, e.g barley and maize. In addition, the characteristic 3-on-3 sandwich fold of seven α-helices was found in this simulation. Even though these results were expected and did not provide any significant new insights, it is still a confirmation that the mentioned scaffolds are probably belonging to class 1 nsHb and share characteristics with Hbs from this group. However, the complete structure has to be determined using e.g X-ray diffraction which could also provide information if the oat Hbs consist as homodimers, as in the case of nsHb2 in rice and Hb from barley, or not. Due to the sequence similarities, it is likely that the oat Hbs also appears as dimers. In addition, mass spectrometry may be an important tool to determine the exact molecular mass and concentrations of all scaffolds, confirming these simulated results.

4.2 Cloning and Transformation

Before expression of nsHbs from oat, the synthesized scaffolds had to be cloned and transformed into different strains of *E. coli* (**Materials and Methods 2.2 and 2.3**). Both the cloning and transformation steps were successful for all scaffolds (**Results 3.2** and **Appendix 7.2 and 7.3**). This was based on the fact that suitable *E. coli* strains showed significant growth in form of isolated colonies when grown on plates containing the appropriate selection marker in comparison with the negative controls, which did not show any growth. These results was evident in both BP and LR reactions, where kanamycin and ampicillin were used, respectively. In addition, DNA sequencing of the entry vector and expression vector were used after the BP and LR reactions which confirmed that the correct vector had been obtained in both cases. All the important segments to create the correct clone were found, including the forward and reverse primers as well as the gene itself. Although, most of the bases in the sequencing result had distinct peaks, which provides confidence that the correct base has been registered, there was uncertainty at a small number of bases as well (**Appendix 7.3**). Despite this uncertainty, the succeeding results show active Hb for all scaffolds, pointing to the fact that cloning were indeed successful.

However, the non-induced cells also showed hemoglobin expression but to a lesser extent compared to the induced ones. This may indicate that even though cells are expressing Hb, the expression vector was leaking to a certain point, thus allowing expression despite the absence of IPTG and ALA. The reason for this is unclear since the expression vector seemed to be constructed in a correct way. Possibly, the lac repressor protein might not be expressed to a sufficient degree to block transcription of the Hb genes, even though this seems unlikely. On the other hand, since active Hb was found for all scaffolds and the amount of Hb increased in the induced samples, the reasons for the plasmid leakage might be of less importance.

4.3 Expression conditions for oat Hbs

After successfully obtaining the correct expression vectors for every scaffold, these were transformed into *E. coli* BL21-DE3 and Hb production was induced (**Materials and Methods 2.4-2.6)**. After dissatisfying results using the standard protocol, condition optimization was used for successive rounds in order to optimize Hb production, thus generating more red/pink cell pellets. During this phase, the parameters OD_{600} at induction, concentration of ALA and IPTG were looked into in detail. More distinctive bands around 18.3 kDa were observed in the induced samples in comparison with negative control in the initial SDS-PAGE, even though a weaker band could be distinguished also here. As previously touched upon, this is in line with the observation that non-induced cells still produce Hb but to lesser extent (**Results 3.3)**. In addition, the impact of ALA and IPTG concentration seem to be less significant in producing reduced Hb, as long as the two are added in moderate concentrations (~0.3 mM ALA and ~0.8 mM IPTG). In contrast, the OD_{600} were of greater importance, where cultures induced at OD_{600}~3 or higher showed a more characteristic pink color in comparison to the negative control (**Results 3.3** and **Appendix 7.4**).

Therefore, a high OD_{600} seems to be one of the most important condition parameters for oat Hb expression. This has also been the case in expression of other plant Hbs (Leiva-Eriksson *et al.*, 2018), thus seen as a reasonable result.

The supernatants that were used in the SDS-PAGE analysis were also used to approximate the Hb concentration by the spectrophotometric assay (**Materials and Methods 2.5** and **Results 3.4)**. Here, two absorbance peaks were observed which is characteristic for the absorption by the heme group (Karnaukhova *et al*,. 2014)). In this case, the untreated Hbs had the greatest absorbance ~424 nm, while the smaller peak occurred at ~560 nm. Therefore, due to sequence and peak similarity, it seemed reasonable to assume the same extinction coefficient as for Hb from barley for now, before the constant has been experimentally determined for oat Hb. In addition, when the reducing agent NaD and CO were added, a small shift (~1 nm) was observed, indicating the presence of active Hb. This was also the case when analyzing the absorbance after attempted purification, to an even higher extent in that case, therefore reinforcing this result.

When comparing the expressed Hb concentration for the scaffolds, Sc485 showed the highest (0.00241 mM), while Sc313051 showed the lowest (0.00137 mM). Since all scaffolds show similar structure (~95%), it is possible but unlikely that the amino acid sequences affect this difference in expression. Instead, the reason might be a more efficient cloning and transformation for Sc485 than for Sc313051. However, it should be pointed out that the Hb concentrations obtained by the 3-point drop method are estimations, which could cause variations in these results. More importantly, all scaffold show Hb expression which could be considered to be the main result for now.

Altogether, the most important parameter to obtain higher expression of reduced Hb was OD_{600}, where OD_{600}~3 would be preferable. The generated Hb also showed activity when NaD and CO were added.

4.4 Purification using cation and anion exchange

In order to separate the Hbs from the rest of the proteins in the supernatants, both cation and anion exchange chromatography were attempted (**Materials and Methods 2.4.5)**. This is an important phase due to the requisite of purified and highly concentrated Hb before attempts to characterize them can be performed. In the case of the cation exchange, ~54% of the Sc485 did bind to the column but did not elute during the elution phase. Instead, this amount were probably flushed out during the cleaning phase of the same column (**Results 3.4.2** and **Appendix 7.5)**. This number for the other scaffolds was lower, 1.3% and 15% for Sc1780 and Sc313051, respectively. The remaining Hb simply passed through the column and were fractionated with the rest of the proteins that did not bind. Thus, these results indicate that the net positive charge of the Hb was not strong enough to be caught by the negatively charged resins in the column. Furthermore, the elution solution seemed to be insufficient enough for the Hbs bound to the column to be eluted. In addition, since oxidized Hb tends to bind stronger to the column by exposing other amino acids on the surface in comparison with its reduced form, the oxidized form of Hb might be contributing factor as well. Based on this, it

is evident that the purification protocol has to be optimized in order to obtain purified samples of Hb.

With this in mind, one possible solution would be to lower the pH of the starting buffer used. Since all scaffold has an estimated pI~8.65, this would increase the amount of positive charges present on the surface of the Hb, increasing the probability for the negatively charged column to attract the Hb. In addition, when more and stronger attached Hb is present in the column, the elution solution needs to be more potent than before. This can be done by increasing the pH and/or increasing the ionic strength of the elution solution. This would be the recommended first step of optimization towards more purified samples of all oat scaffolds moving forward.

In the case of the anion exchange, the results differ compared with the cationic version. Here, Hb was detected in the eluted fraction with varying amounts for all scaffolds (**Results 3.4.3** and **Appendix 7.5)**. Around 25% was eluted for Sc485, ~15% and 25% for Sc1780 and Sc313051, respectively. However, a large majority of the other proteins present in the supernatant were also detected in this eluted fraction, thus not resulting in more purified sample of Hb per se. Instead, this indicates that the anion exchange is more prone to bind many different proteins than in the case of the cation, which is not optimal. This observation is also confirmed by the peaks of the eluted samples, which does not have a more defined absorbance maximum in comparison with the unpurified supernatant. This would be the expected result if the purification was successful. In addition, some of the Hb was stuck in the column (~7% for Sc485 and Sc313051), as in the case for the cationic exchange, and eluted during the cleaning phase. The majority of the Hb simply passed through the column.

When examining the results for the anion exchange, a small difference between the weak and strong anion exchangers were observed, which indicates that neither of them were sufficient in the purification. Surprisingly, Hb was more prone to stick to the weak anion exchanger. The reason for this might be that the strong exchanger bound a larger amount of other proteins than Hb, leading to less available area for the Hb to be bound to.

An attempt to optimize the protocol could be done here as well by increasing the pH of the starting solution in this case. This provide more anionic surface charges on the Hb and increase the chance for the positively change column to attract the Hb. However, due to the pI of the scaffolds, the resulting solution needed might be too alkaline and therefore causing other problems instead, such as denaturation of proteins and damaging the purification system.

In summary, both cation and anion exchange chromatography were unsuccessful in purifying the Hb at the initially set conditions. However, due to more and better opportunities for optimization in addition to more specific binding of Hb, the cationic separation show more promise and should be focused on in future attempts when purifying oat Hb.

4.5 Reflections and future considerations

Even though a lot of effort has been put into this project and many different steps have been done, all the way from determining the genomic sequence of the oat Hbs to attempts to purifying these proteins, a considerable amount of work is still at hand for future projects within this area. The first aspect would be to obtain reduced Hb directly from the expression phase. A lot of attempts to reach this result have been done in this thesis. In condition optimization phase, appropriate concentrations of ALA and IPTG were determined, as well as suitable OD_{600}. In addition, temperature (17-22ºC) and shaking speed (100-150 rpm) were also tested during this period, with little observed effect. With this said, a wider range of these parameters may be tested to obtain reduced Hb. In addition, the sequence of the Hbs may be looked into and altered to a similar sequence as in rice Hb. These Hbs has been successfully expressed in reduced form, appearing as red cell pellets. However, this might alter the oat Hbs to extensively, losing their unique characteristics and resemble rice Hb too much.

An alternative to this method would be to reduce the Hb outside of the cells. In this thesis, NaD has been used to reduce the Hb, making it possible for the protein to bind CO instead. In a paper submitted by Sainz *et al*, class 1 nsHbs have reported to be rapidly reduced by electron scavenging flavins, where FAD was the most prominent one (Sainz *et al*, 2013). This could be an interesting approach generating reduced Hb from the expressed oxidized version of the protein.

Hb purification using protein tags can also be an option. Ascher and Bren has reported of a heme fusion tag that has shown to bind strong yet reversibly to L-histidine-immobilized sepharose resins (Ascher and Bren, 2010). It has to be evaluated if the heme already present in the Hb is sufficient for attachment to the resins or not. In their method, the authors use a heme complex containing five heme groups in close contact to achieve the desired binding. A more conventional way of creating fusion protein would be to add a histidine tag to the N-terminal of the Hb and separate these by applying immobilized metal affinity chromatography (IMAC). However, as described, these methods includes adding a tag to the N- or C-terminal of the existing protein. When doing this, there is always a risk of reducing the activity of the protein, which may also be the case here.

Whatever the choice of method, it is crucial to obtain purified Hbs in the end in order to be properly analyzed and characterized. Once this has been achieved, important constants such as oxygen affinity and hexacoordination equilibrium constant (K_H) can be determined and compared with other Hbs form monocots but also human Hb. A thorough characterization of oat Hb is required before more elaborate studies can be done, such as oxygen transporting capacity *in vivo* using mice.

In conclusion, a lot of effort has to be put in to future studies of oat Hb before these proteins would be available as a HBOC or food additive. However, this is a new and interesting type of Hb that might be of great use in modern medicinal and/or nutritional areas.

5. Conclusions

In summary, a couple of conclusions can be drawn from this thesis. Firstly, three nsHbs and two trHbs has been found in the oat genome, where the nsHbs is the most interesting ones for future studies. These Hbs show high similarity with other monocotic Hbs, primarily from barley and rice. Secondly, expression of the oat Hbs were successful and show activity but the proteins were expressed mainly in oxidized state. The most important parameter to achieve proper expression was induction at high OD_{600} (~3), while precise ALA and IPTG concentration seemed to be of less importance. Lastly, optimization of the purification protocol has to be made for sufficient separation of Hb by e.g use lower pH for the starting buffer to increase the presence of positively charges on the surface of the proteins. Only when this has been done, the oat Hbs can be fully characterized and compared to other Hbs.

6. References

Alayash, A I. (2012). Blood Substitutes: Working to fulfill a dream. *U.S Food and Drug Administration.* [online] Available at: https://blogs.fda.gov/fdavoice/index.php/2012/06/blood-substitutes-working-to-fulfill-a-dream/ [Accessed 2018-09-05]

Asher, W.B and Bren, K.L. (2010). A heme fusion tag for protein affinity purification and quantification. *Protein Science.* **19**(10). 1830-1839

Chapman, Ashley. (2016). Presentation of Protein Chemistry. [online] Available at: https://slideplayer.com/slide/7753891/ [Accessed 2018-09-05]

Chukwubueze-Okeke, D. (2017). Development of Plant Hemoglobin-Based Blood Substitute: Fed-Batch Fermentation Optimization and Stability Characterization: Master Thesis. Lund University and University of Stavanger.

Cooper, C E. Silkstone, G.A. Simons, M.Rajagopal, B. Syrett, N. Shaik, T. Gretton, S. Welbourn, E. Bülow, L. Leiva Eriksson, N. Ronda, L. Mozzarelli, A. Eke, A. Mathe, D and Reeder, B J. (2018). Engineering tyrosine residues into hemoglobin enhances heme reduction, decreases oxidative stress and increases vascular retention of a hemoglobin based blood substitute. (Manuscript)

European Bioinformatics Institute. (2018). Clustal Omega: Multiple Sequence Alignments. [online] Available at: https://www.ebi.ac.uk/Tools/msa/clustalo/ [Accessed 2018-08-24]

European Bioinformatics Institute. (2018). Genewise: Pairwise Sequence Alignment. [online] Available at: https://www.ebi.ac.uk/Tools/psa/genewise/ [Accessed 2018-08-24]

European Molecular Biology Laboratory. (2018). Cloning Methods: Gateway cloning technology. [online] Available at: https://www.embl.de/pepcore/pepcore_services/cloning/cloning_methods/recombination/gateway/ [Accessed 2018-08-21]

Flora, Swaran J.S. (2014). Biomarkers in Toxicology: Chapter 29- Metals. pp 485-519. Available at: https://doi.org/10.1016/B978-0-12-404630-6.00029-4

Gupta, Kapuganti J. Hebelstrup, Kim H. Mur, Luis A.J and Igamberdiev, Abir U. (2011). Plant Hemoglobin: Important players at the crossroads between oxygen and nitric oxide. *FEBS Letters.* **585,** 3843-3849

Hoy, Julie A and Hargrove, Mark S. (2008). The structure and function of plant hemoglobin. *Plant Physiology and Biochemistry.* **46.** 371-379

Institute for Agriculture and Trade Policy. (2007). The botany of gluten: How closely are wheat, oat, corn, rice and sugar cane related? [online]. Available at: https://www.iatp.org/news/the-botany-of-gluten-how-closely-are-wheat-oats-corn-rice-and-sugar-cane-related [Accessed 2018-08-21]

Invitrogen. (2012). Gateway® Technology with Clonase® II: A universal technology to clone DNA sequences for functional analysis and expression in multiple systems. [online]. Available at: https://tools.thermofisher.com/content/sfs/manuals/gateway_clonaseii_man.pdf [Accessed 2018-09-24]

Jensen, Frank.B. Fago, Angela and Weber, Roy. E. (1998). Hemoglobin Structure and Function. *Fish Physiology.* **17,** 1-40

Karnaukhova, E. Rutardottir, S. Rajabi, M. Wester-Rosenlöf, L. Alayash, A I and Åkerström, B. (2014). Cheracterization of heme binding to recombinant α1-microglobulin. *Frontiers in Physiology.* **5.** 465

Kelley, L.A. Mezulis, S. Yates, C.M. Wass, M.N and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols.* **10**, 845-858

Leiva-Eriksson, N. (2014). Biochemical and Physiological Characterization of Nonsymbiotic Plant Hemoglobin. Doctor, Lund University

Leiva-Eriksson, N. Pin, P.A. Kraft, T. Dohm, J.C. Minoche, A.E. Himmelbauer, H. and Bülow, L. (2014). Differential expression patterns of non-symbiotic hemoglobin in sugar beet (Beta vulgaris ssp. Vulgaris). *Plant and Cell Physiology*. **0**(0), 1-11

Leiva-Eriksson, N. Reeder, B.J. Wilson, M.T and Bülow, L. (2018). Sugar beet hemoglobin: reactions with nitric oxide and nitrite reveal differential roles for nitrogen metabolism. (Manuscript submitted).

Macherey-Nagel. (2012). Plasmid DNA Purification. Protocol, Rev. 08.

Massachusetts Institute of Technology. (2005). GenScan: Identification of complete gene structures in genomic DNA. [online] Available at: http://genes.mit.edu/GENSCAN.html [Accessed 2018-08-25]

Mozzarelli, Andrea. Ronda, Luca. Faggiano, Serena. Bettati, Stefano and Bruno, Stefano. (2010). Hemoglobin-based oxygen carriers: research and reality towards an alternative to blood transfusion. *Blood Transfus.* **8**(3), 59-68

Rasane, Prasad. Jha, Alok. Sabikhi, Latha. Kumar, Arvind and Unnikrishnan, V.S. (2015). Nutritional advantages of oats and opportunities for its processing as value added foods. *Journal of Food Science and Technology*. **52**(2). 662-675

Riquelme, A. and Hinrichsen, P. (2015). Non-symbiotic hemoglobin and its relation with hypoxic stress. *Chilean Journal of Agricultural Research.* **75** (1)**.** 80-89

Rodriguez-Alonso, G and Arredondo-Peter, R. (2013). Variability of non-symbiotic and truncated hemoglobin genes from the genome of cultivated monocots. *Communicative and Integrative Biology.* **6**(6). pp. e27496 1-3

Roy, A. Kucukural, A and Zhang, Y (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, **5**, 725-738

Sainz,M. Perez-Rontom, C. Ramos, J. Mulet, J.M. James, E.K. Bhattacharjee, U. Petrich, J.W and Becana, M. (2013). Plant hemoglobin may be maintained in functional form by reduced flavins in the nuclei, and confer differential tolerance to nitro-oxidative stress. *The Plant Journal.* **76**. 875-887

ScanOats. (2018). The oat genome. Faculty of Engineering, Lund University. [online]. Available at: http://www.scanoats.se/scanoats-rd/1-the-oat-genome/ [Accessed 2019-01-04]

Silkstone, G. G. A., Silkstone, R. S., Wilson, M. T., Simons, M., Bülow, L., Kallberg, K., … Cooper, C. E. (2016). Engineering tyrosine electron transfer pathways decreases oxidative toxicity in hemoglobin: implications for blood substitute design. *Biochemical Journal*, **473**(19). 3371–3383

Softberry. (2018). Gene Finding: Gene models construction, splice sites, protein-coding exons. [online] Available at: http://www.softberry.com/berry.phtml?topic=index&group=programs&subgroup=gfind&gclid=Cj0K

CQjwz93cBRCrARIsAEFbWsgmuYdZxza3S2Y2uee5wtv-bpqmfVZ4aJKTYE2Qv02EHIt4DpwNa24aAvbiEALw_wcB [Accessed 2018-08-25]

Stryer, L. Tymoczko, J L. and Berg, J M. (2012). *Biochemistry*. New York: W.H Freeman and Company. 203-215

Swiss Institute of Bioinformatics. (2018). ExPaSy: Translate tool. [online] Available at: https://web.expasy.org/translate/ [Accessed 2018-08-24]

Swiss Institute of Bioinfomatics. (2018). ExPaSy: ProtParam. [online] Available at: https://web.expasy.org/cgi-bin/protparam/protparam [Accessed 2018-11-22]

Vazquez-Limon, C. Hoogewijs, D. Vinogradov, S.N. and Arredondo-Peter, R. (2012). The evolution of land plant hemoglobin. *Plant Science.* **191-192,** 71-81

University of Greifswald. (NA). Augustus: Gene Prediction. [online] Available at: http://bioinf.uni-greifswald.de/augustus/ [Accessed 2018-08-25]

7. Appendix

7.1 DNA cloning of hemoglobin genes

In order to express the oat hemoglobin most efficiently, the correct plasmids must be created and transformed into a suitable host. In total, three transformation steps were required to obtain the final expression vector in *E. coli* BL21-DE3 cells. In this section, the two first steps are presented.

7.1.1 BP cloning in competent OmniMax® *E. coli* cells

In the first cloning step, the synthesized hemoglobin genes were inserted into an entry vector by Gateway cloning (**Materials and Methods** section 2.2). Since the entry vector contained a kanamycin resistance gene, colonies were screened on plates containing the antibiotic. The result can be seen in **Figure A1**, where the left plate is the negative control and the right plates corresponds to Sc780. As shown in the picture, the sample plate displayed a large amount of separated colonies, indicating that the correct plasmid has been cloned and transformed in the OmniMax® cells. In addition, the negative control did not show any growth, meaning that cells without or containing the incorrect plasmid cannot form colonies. Thus, the first cloning step was considered to be successful.

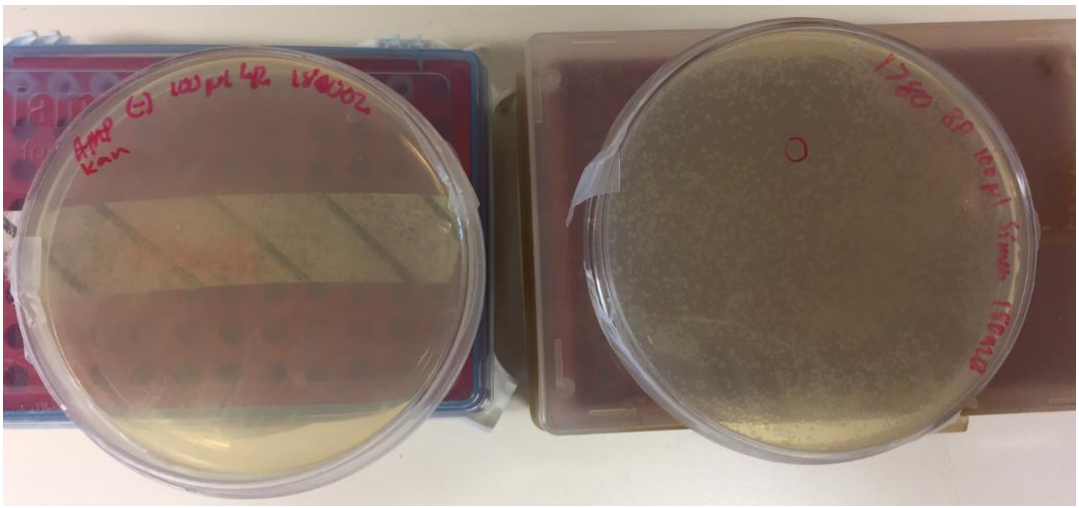


Figure A1. First cloning step and transformation into OmniMax® cells. The resulting entry vector of the cloning was transformed into the cells and grown on kanamycin plates. The negative control (left plate) did not show any growth, while the sample plates (right plate) all showed distinctive colonies.

7.1.2 LR cloning in competent OmniMax® *E. coli* cells

In the second step of cloning, the hemoglobin genes from the entry vector (BP reaction) were cloned into the final plasmid, creating the expression vector. Since the expression vector contained an ampicillin resistance gene, transformed OmniMax® cells were screened on

ampicillin containing plates. The result can be seen in **Figure A2**, where the negative control (left plate) and Sc485 (right plate) are shown. Several distinctive colonies on the sample plate could be seen, where the negative plate showed no growth at all. Thus, the second step of cloning was also successful. The purified expression vectors from this step were transformed into *E. coli* BL21-DE3 cells for expression (**Results 3.2.1**).

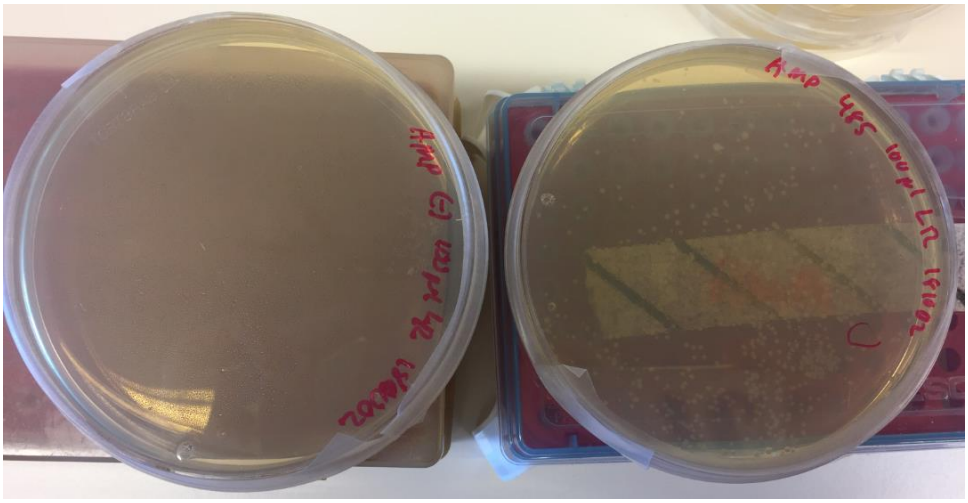


Figure A2. Second cloning step and transformation into OmniMax® cells. The resulting expression vector of the cloning was transformed into the cells and grown on ampicillin plates. The negative control (left plate) did not show any growth, while the sample plates (right plate) all showed distinctive colonies.

7.2 DNA sequencing of vectors

7.2.1 Entry vector from BP reaction

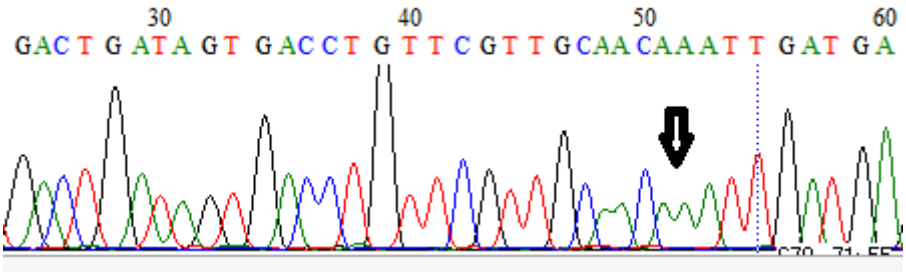
The entry vector, generated from the BP reaction and screened by kanamycin resistance, was isolated using the Plasmid DNA purification kit from Macherey-Nagel. The result from the concentration and absorbance measurements can be seen in Table A1. The samples indicated in yellow were chosen for sequencing due to appropriate concentration, high A260/A280 and A260/A230 ratios (indicating appropriate DNA purity and less contamination of salts and solvents, respectively).

Table A1. Concentration and absorbance measurements of the isolated plasmids from BP reaction. The measurements were done in duplicated and the samples chosen for sequencing are indicated in yellow, one from each hemoglobin gene.

Sample	Concentration (ng/µl)	A260/A280	A260/A230
485 (100 µl)	35.8	1.821	2.448
485 (20 µl)	65.6	1.808	2.357
1780 (100 µl)	51.7	1.825	2.364
1780 (20 µl)	27.8	1.806	2.545
313051 (100 µl)	106	1.821	2.536
313051 (20 µl)	115	1.841	2.522

The selected samples were sequenced by GATC© and one example of the result can be seen below. The most important aspects to be found in the sequenced plasmid is the presence of the *att*L1 and *att*L2 sites, which must be present for a successful LR recombination reaction to create the expression vector, and the hemoglobin genes themselves. Below, the results for the scaffold 313051 is shown, where the *att*L-sequences are highlighted in yellow and the hemoglobin gene region is underlined. Both sites were present except for a single nucleotide substitution (marked as a red A instead of a C in *att*L1).

However, this altered sequence is uncertain in this area when compared to the raw data. This can be seen in the raw data below, where the fluorescent peaks indicated by the arrow are less defined in comparison to the others (see arrow), adding a higher degree of uncertainty. Thus, it is possible the complete sequence is present even if the sequencing does not display this completely. The outcome from the LR reaction would determined if the correct sequences was present or not, which indicated that the reaction was indeed successful.



In addition, when comparing the sequencing result with the original scaffold sequences (on amino acid level), the complete hemoglobin gene is also present but with a small overhang in the sequencing result of 313051. It is unclear if this should cause a problem or if it is just the translating program that have added this additional amino acids. This is also present in the other scaffolds as well. This will be revealed when the expression plasmid is generated and the hemoglobin gene are expressed. Due to the observed results, showing active Hb, it is more likely that the added overhang came from the simulation than present in the vector.

7.2.2 Expression vector from LR reaction

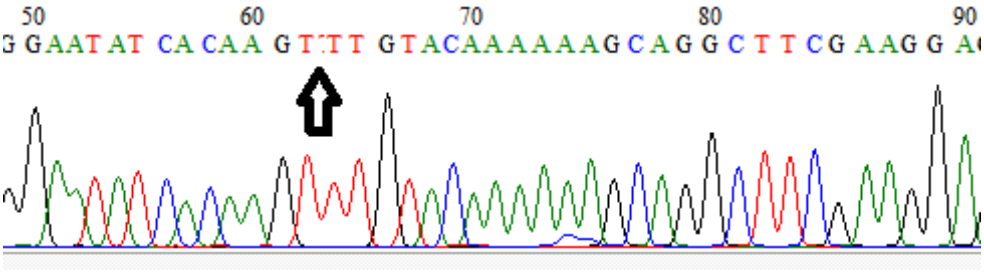
The expression vector, generated from the LR reaction and screened by ampicillin resistance, was isolated using the Plasmid DNA purification kit from Macherey-Nagel. The result from the concentration and absorbance measurements can be seen in **Table A2**. The samples indicated in yellow were chosen for sequencing due to appropriate concentration, high A260/A280 and A260/A230 ratios (indicating appropriate DNA purity and less contamination of salts and solvents, respectively).

Table A2. Concentration and absorbance measurements of the isolated plasmids from LR reaction. The measurements were done in duplicated and the samples chosen for sequencing are indicated in yellow, one from each hemoglobin gene.

Sample	Concentration (ng/µl)	A260/A280	A260/A230
485 (100 µl)	60.1	1.862	2.283
485 (20 µl)	76.0	1.800	2.186
1780 (100 µl)	62.6	1.853	2.681
1780 (20 µl)	67.6	1.838	2.061
313051 (100 µl)	51.7	1.815	2.722
313051 (20 µl)	91.9	1.832	2.034

The selected samples were sequenced by GATC© and one example of the result can be seen below. The most important aspects to be found in the sequenced plasmid is the presence of the *att*B1 and *att*B2 sites, which must be present after a successful LR recombination reaction to create a functional expression vector, and the hemoglobin genes themselves. Below, the results for the scaffold 485 is shown, where the *att*B-sequences are highlighted in yellow and the hemoglobin gene region is underlined. Both sites were present and correct except for one additional T in the *att*B1 sequence of the forward primer (marked red).

However, the raw data shows uncertainty in this area and this error was not found in the reverse complement. The sequencing result in this area can be seen below, where the additional T is marked by the arrow. Thus, this is considered to be a sequencing error and that the correct plasmid is present. The other scaffold showed perfect *att*B-sites and hemoglobin gene.



In addition, as can be seen by the amino acid sequence alignment below, the complete hemoglobin gene is present in between the *att*B-sites. Thus, it is concluded that the LR reaction and transformation was successful and the expression plasmids can be used to express the desired oat hemoglobin.

7.3 Results of expression in initial test cultures

7.3.1 Expression of all scaffolds

In the first set of hemoglobin expression, all genes were used and the original protocol was deployed (**Materials and Methods 2.6).** The cells were first cultivated and induced at OD_{600}~2. The cell weight after harvesting can be seen in **Table A3**. Cells that contained scaffold 313051 showed the most growth, where the other two showed similar growth with each other. This is also shown in **Figure A3**. Due to the white/brown color of the cell pellets, it was believed that the hemoglobin had not been produced to a desired extent. Instead, it was decided that the conditions had to be optimized in order to increase the expression. In addition, the higher cell mass of scaffold 313051 might indicate that the cells did induce hemoglobin expression to a lower extent in comparison with the other scaffolds. Thus, condition optimization was first performed on scaffold 485 and later 1780.

Table A3. Measured OD_{600} and cell weight after 180 min cultivation and induction at OD_{600}~2.

Sample	OD_{600} (180min)	Cell weight (g/L)
1. 485	~2.0	4.6390
2. 485	~2.0	6.5971
3. 1780	~2.0	4.1912
4. 1780	~2.0	4.5112
5. 313051	~2.0	26.8322
6. 313051	~2.0	14.1493

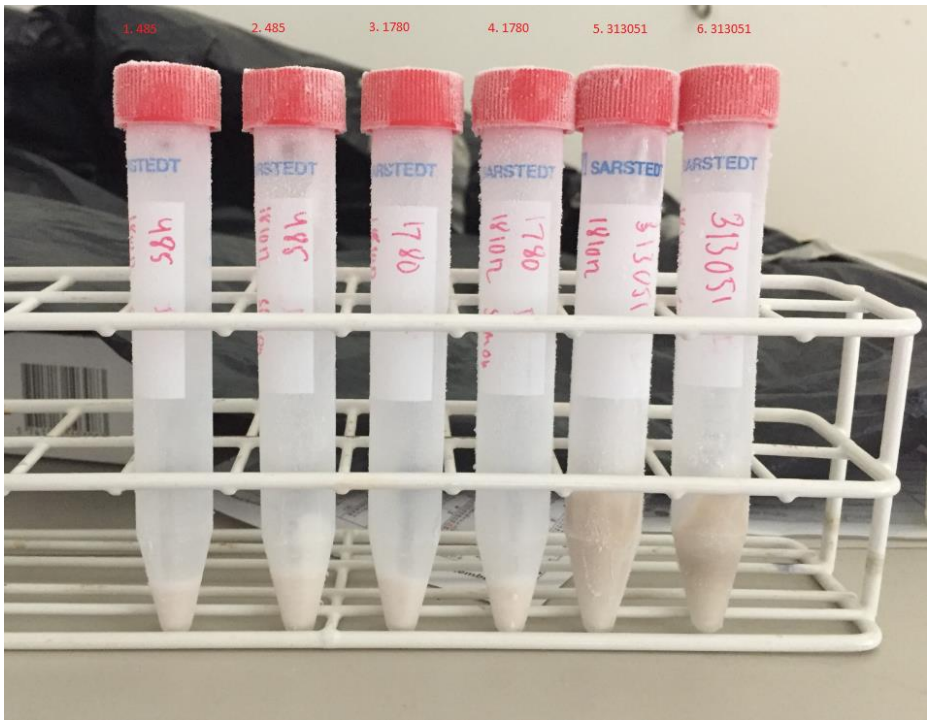


Figure A3. Cell pellets after harvesting of initial test cultures using all scaffolds. Sample 1 and 2 correspond to scaffold 485, 3 and 4 to scaffold 1780 and 5 and 6 to scaffold 313051. Due to the higher cell weight of 313051, hemoglobin expression might be less efficient here. Thus, scaffold 485 and 1780 was used for condition optimization.

All scaffolds were induced at OD_{600}~2 with 0.3 and 0.5 mM ALA and IPTG, respectively. The resulting SDS-PAGE result can be seen in **Figure A4**.

Hemoglobin expression can be seen for all scaffold, including samples 15 of the second round of optimization for scaffold 485 and 1780. However, no negative control was included in the picture, making it impossible to compare the degree of expression for the scaffolds with no expected expression. In addition, the samples for scaffold 313051 were overestimated due to poor normalized amount of cells used. Although, the results for scaffolds 485 and 1780 were normalized and therefore decided to run these results again with less volume of the samples for 313051.

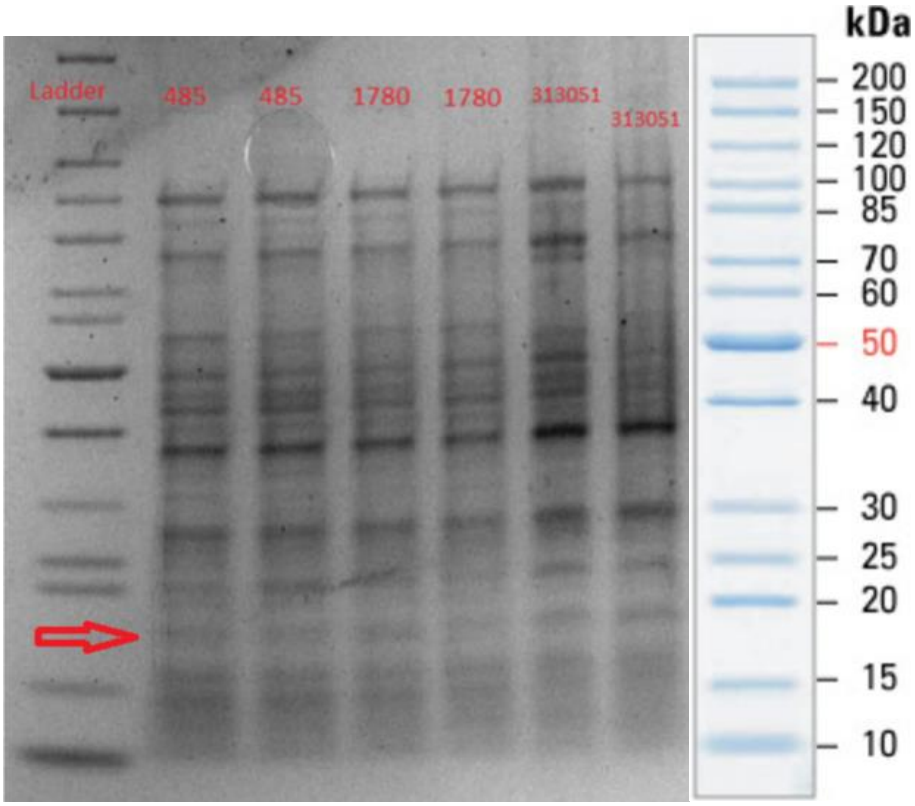


Figure A4. First around of condition optimization with all scaffolds. The cultures were induced at OD_{600}~2 with 0.3 mM of ALA and 0.5 mM IPTG. The arrow indicates the location of the projected band of hemoglobin, around 18 kDa.

7.3.2 Condition optimization for scaffold 485

7.3.2.1 Combination of 0.1 and 0.5 mM ALA with 0.1, 0.5 and 1 mM IPTG at different OD

In the first round of condition optimization, scaffold 485 was used with varying concentrations of ALA and IPTG. The set-up of the optimization, where 0.1 and 0.5 mM ALA were used in combinations with 0.1, 0.5 and 1 mM IPTG (**Materials and Methods 2.6**). In addition, the expression temperature was lowered to 17ºC. Samples 1-7 were induced at OD_{600}~1, while 8-13 were induced at OD_{600}~1.5. Also, sample 1 was not induced at all.

The measured OD_{600} and cell weights can be seen in **Table A4**. Pictures of the results can be seen in **Figure A5**. The brown color increased for all cell pellets in comparison with the previous testing, indicating higher expression of hemoglobin but in oxidized form. Due to the oxidized for of the iron in heme, the cells were more brown than red. As previously mentioned, decreased cell growth might indicate increased protein expression. In this case, the concentration of ALA seemed less important than the concentration of IPTG. The lowest amount of cell mass was found with 0.1 or 1 mM IPTG (e.g sample 2 and 7, respectively). Thus, it was decided to keep the ALA concentration constant at 0.3 mM but vary the IPTG concentration between 0.2 or 0.8 mM. Note, the negative control (sample 1) was still bubbled with carbon monoxide and therefore not a valid control.

Table A4. Results from OD_{600} and cell weight measurements with combinations of ALA and IPTG concentration. Sample 1 was not induced, while 2-7 were induced at ~0.1 and 8-13 at ~1.5.

Sample	OD_{600} (30 min)	OD_{600} (130 min)	Cell weight (g/L)
*1.	0.105	-	1.9198
2.	0.125	-	1.1718
3.	0.095	-	5.8821
4.	0.105	-	1.7542
5.	0.100	-	3.3704
6.	0.090	-	10.7002
7.	0.110	-	1.8963
8.	-	1.670	3.8545
9.	-	1.830	3.4257
10.	-	1.885	4.5188
11.	-	1.570	4.5696
12.	-	1.645	3.5510
13.	-	1.595	3.9515

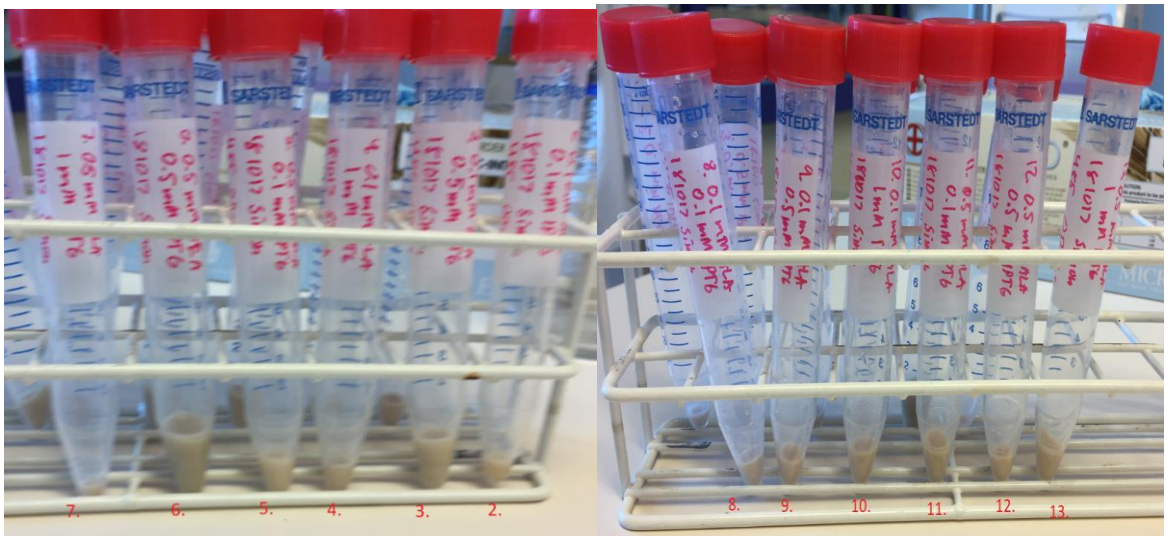


Figure A5. Cell pellets for samples induced at OD_{600}~0.1 (samples 2-7, left side) and samples induced at OD_{600}~1.5 (samples 8-13, right side).

7.4.2.2 Combination of 0.3 mM ALA with 0.2 and 0.8 mM IPTG at different OD

In the second round of condition optimization, the ALA concentration was constant at 0.3 mM and combine with either 0.2 or 0.8 mM IPTG at different OD_{600}. The measured OD_{600} at different time points and the final cell weight can be seen in **Table A5**, while pictures of the cell pellets after harvesting is shown in **Figure A6**.

In this case, the negative control (sample 1) was not bubbled with carbon monoxide, thus providing a more reliable control. A difference in color can be distinguished between the negative control and the induced samples, where the negative control displays a predominant white color while the induced samples has a brownish character as before. This might indicate that hemoglobin are produced in the induced samples to a higher extent but in an oxidized form. Sample 11 (0.3 mM ALA, 0.8 mM IPTG) seemed to generated a small cell mass but maintain the brown color and could thus be a candidate for optimal conditions. However, the cells seemed to have no problem generating the brown color while still growing, which point to the fact that the cells can manage to both grow and produced hemoglobin to a higher extent after induction.

Table A5. Measured OD_{600} at different time points and cell weights for all samples. Sample 1 was the negative control (no induction), samples 2,4,6,8,10,12 and 14 were induced with 0.2 mM IPTG at different OD_{600}, while 3,5,7,9,11,13,15 were induced with 0.8 mM IPTG.

Sample	OD_{600} (0min)	OD_{600} (30min)	OD_{600} (60min)	OD_{600} (90min)	OD_{600} (120min)	OD_{600} (150min)	OD_{600} (180min)	Cell weight (g/L)
*1.	0	-	-	-	-	-	-	1.6662
2.	0	-	-	-	-	-	-	1.5338
3.	0	-	-	-	-	-	-	1.5110
4.	-	0.095	-	-	-	-	-	1.2319
5.	-	0.090	-	-	-	-	-	1.5971
6.	-	-	0.170	-	-	-	-	1.2704
7.	-	-	0.130	-	-	-	-	3.5349
8.	-	-	-	0.210	-	-	-	1.8624
9.	-	-	-	0.165	-	-	-	2.5725
10.	-	-	-	-	0.485	-	-	7.2986
11.	-	-	-	-	0.540	-	-	1.3515
12.	-	-	-	-	-	0.980	-	9.6937
13.	-	-	-	-	-	1.050	-	12.0662
14.	-	-	-	-	-	-	1.970	5.4439
15.	-	-	-	-	-	-	2.190	6.2815

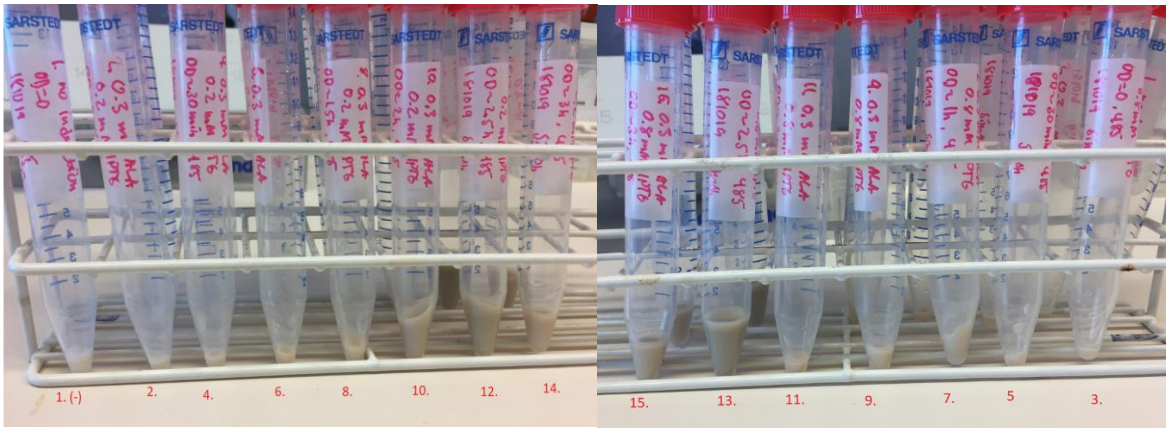


Figure A6. Cell pellets from second round of condition optimization. Sample 1 was the negative control (no induction), samples 2,4,6,8,10,12 and 14 were induced with 0.2 mM IPTG at different OD_{600}, while 3,5,7,9,11,13 and 15 were induced with 0.8 mM IPTG. The ALA concentration was kept constant at 0.3 mM.

The resulting SDS-PAGE result is shown in **Figure A7**. As before, the expected hemoglobin band is located at approximately 18 kDa (indicated in the figure). Higher OD_{600} (samples 8-13) still tends to show a more distinctive band at this location in comparison with the samples induced at lower OD_{600} (samples 2-7). However, the gel was loaded with samples from different concentrations of cells, which is evident in the case of samples 2 and 3. Thus, it is not valid to compare these results with each other since the amount of cells were not properly normalized. This might be due to incorrect OD_{600} measurements or suspension in SDS loading buffer. Although the samples are incomparable, the results are in line with previous observations presented in the **Results 3.3.1.**

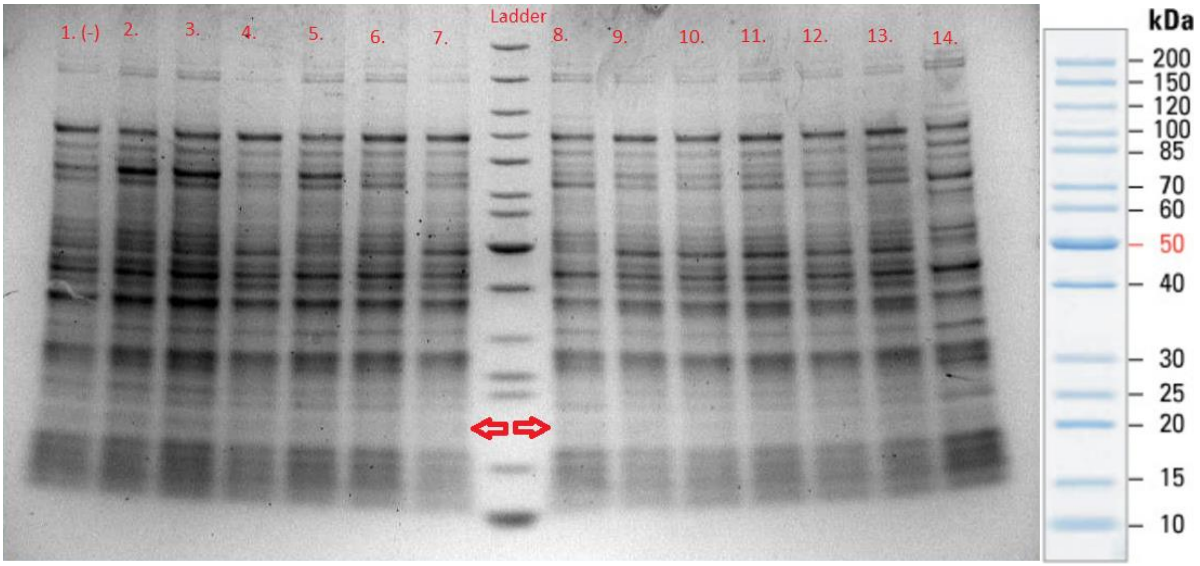


Figure A7. Second round of condition optimization with scaffold 485. The cultures were induced at different OD_{600} with 0.3 mM ALA and 0.2 or 0.8 mM IPTG. The negative control (sample 1) was not induced. For complete set-up, see Materials and Methods section 2.5.

7.3.3 Condition optimization for scaffold 1780

7.3.3.1 Combination of 0.1 and 0.5 mM ALA with 0.1, 0.5 and 1 mM IPTG at different OD

In the first round of condition optimization with scaffold 1780, the same procedure was done as in the case of scaffold 485, except for the extra negative control at OD_{600} ~1 (sample 8). The measured OD_{600} and cell weights can be seen in **Table A6**, while the cell pellets after harvesting can be seen in **Figure A8**. The general feature is that all induced samples has a brown color as before, but the negative controls (sample 1 and 8) also seem to have a slightly brownish color. This might indicate that the non-induced samples still produce oxidize hemoglobin or that no sample is producing hemoglobin to a high extent. In addition, samples 2 and 4 had the lowest growth at the OD_{600}~0.1, while 9 and 12 had the lowest growth at OD_{600}~1. Thus, it was decided to keep the ALA concentration constant at 0.2 mM in combination with 0.2 or 0.8 mM IPTG in the next round of optimization. The SDS-PAGE results can be seen in the **Results 3.3.2.**

Table A6. Results from OD_{600} and cell weight measurements with combinations of ALA and IPTG concentration. Sample 1 and 8 was not induced, while 2-7 were induced at ~0.1 and 9-14 at ~1.0.

Sample	OD_{600} (30 min)	OD_{600} (100min)	OD_{600} (130min)	Cell weight (g/L)
*1.	0.090	-	-	7.4940
2.	0.100	-	-	0.7709
3.	0.060	-	-	6.9110
4.	0.080	-	-	2.4526
5.	0.090	-	-	7.2142
6.	0.075	-	-	6.5167
7.	0.100	-	-	7.5591
*8.	-	0.405	0.845	17.9316
9.	-	0.460	0.960	1.7841
10.	-	0.505	1.040	7.8758
11.	-	0.495	0.970	12.4990
12.	-	0.545	1.035	4.2315
13.	-	0.455	0.980	9.3040
14.	-	0.480	1.065	11.5250

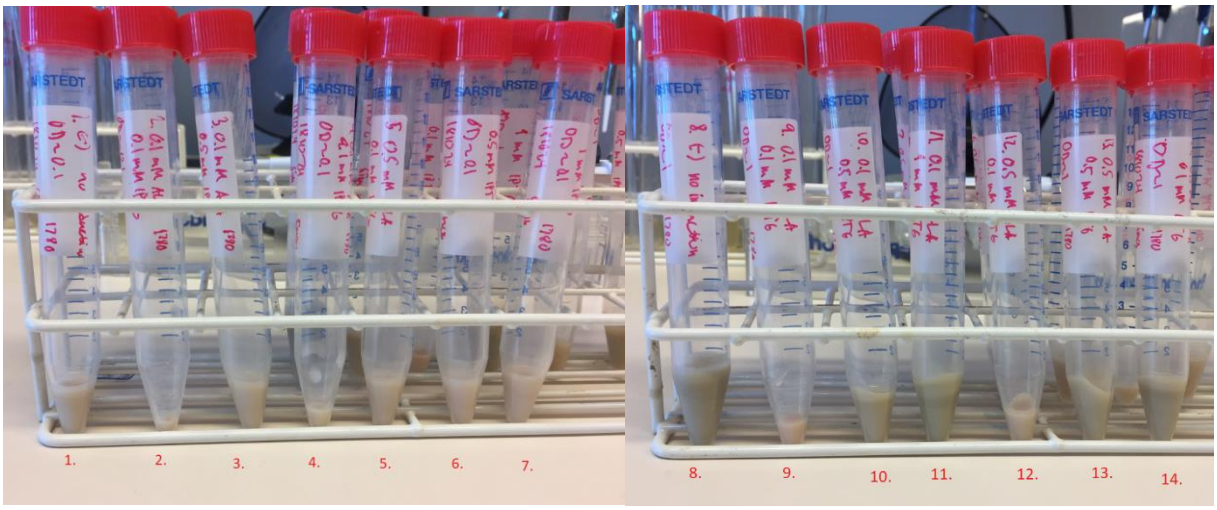


Figure A8. Cell pellets for samples induced at OD_{600}~0.1 (samples 2-7, left side) and samples induced at OD_{600}~1.0 (samples 9-14, right side). Samples 1 and 8 are negative controls (no induction).

In the first round of condition optimization of scaffold 1780, the same conditions as for scaffold 485 were tried. The resulting SDS-PAGE result is shown in **Figure A9**. The expected hemoglobin band is located at approximately 18 kDa (indicated in the figure).

As can be seen, higher OD_{600} (samples 9-14) tends to show a more distinctive band at this location in comparison with the samples induced at low OD_{600} (samples 2-7) and the negative controls (samples 1 and 8). In addition, this in combination with higher concentration of IPTG and ALA (samples 12 and 13 in particular) show the most distinctive band around 18 kDa. In alignment with the results of scaffold 485, these result in combination with previous seen results (**Appendix 7.4.2**), the most important parameters to take into account are a high OD_{600} (>2) before induction and IPTG concentration between 0.5-1 mM to increase the hemoglobin expression. The concentration of ALA did not seem to have any significant impact on the expression. In addition, the cells containing scaffold 1780 show higher growth in comparison with 485 under the same conditions.

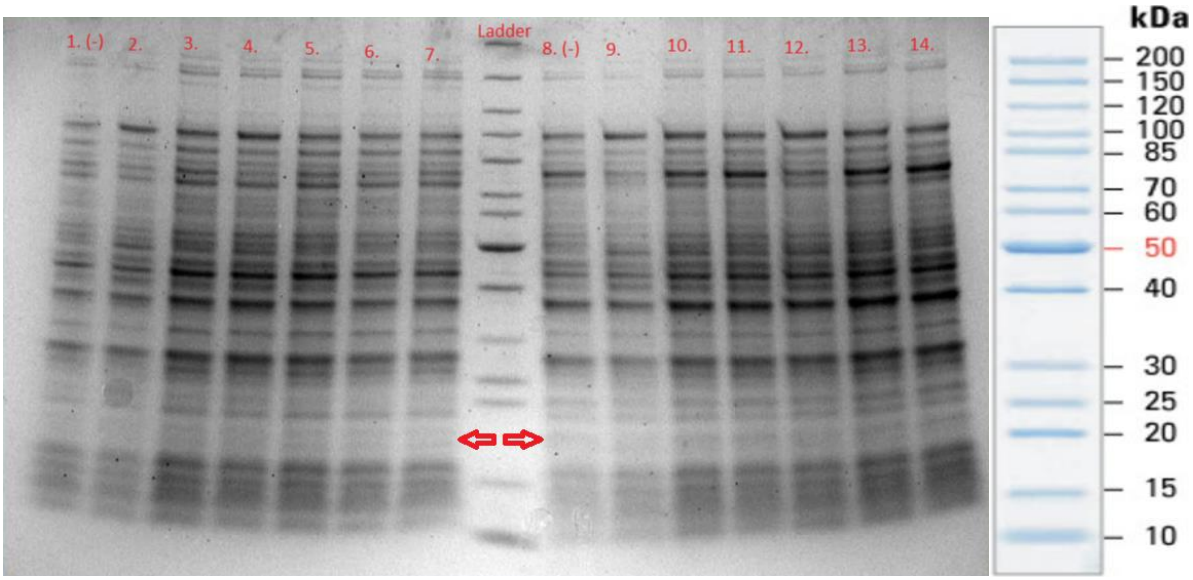


Figure A9. First around of condition optimization with scaffold 1780. The cultures were induced at different OD_{600} with varying concentrations of ALA and IPTG. Apart from the negative control (samples 1 and 8), the remaining samples were induced with 0.1 or 0.5 mM ALA in combination with 0.1, 0.5 or 1 mM IPTG (see Materials and Methods section 2.5 for complete set-up).

7.3.3.2 Combination of 0.2 mM ALA with 0.2 and 0.8 mM IPTG at different OD

In the second round of condition optimization with scaffold 1780, the ALA concentration was constant at 0.2 mM and combine with either 0.2 or 0.8 mM IPTG at different OD_{600}. The measured OD_{600} at different time points and the final cell weight can be seen in **Table A7**, while pictures of the cell pellets after harvesting is shown in **Figure A10.**

One general feature is that 0.2 mM IPTG seems to decrease the growth of the cells, which can be seen samples 2 and 4 in the left side of **Figure A10**. In addition, a more predominant brown color can be seen in the cells with higher OD (from sample 6) where the most brownish color can be seen in the cells induced at highest OD. This points to the fact that high OD before induction of hemoglobin expression may be the preferred as well as an ALA and IPTG concentration between 0.2-0.3 and 0.8-1 mM, respectively.

Table A7. Measured OD_{600} at different time points and cell weights for all samples. Sample 1 was the negative control (no induction), samples 2,4,6,8,10,12 and 14 were induced with 0.2 mM IPTG at different OD_{600}, while 3,5,7,9,11,13 and 15 were induced with 0.8 mM IPTG. The ALA concertation was kept constant at 0.2 mM.

Sample	OD_{600} (0min)	OD_{600} (30min)	OD_{600} (60min)	OD_{600} (90min)	OD_{600} (120min)	OD_{600} (150min)	OD_{600} (180min)	Cell weight (g/L)
*1.	0	-	-	-	-	-	-	7.5193
2.	0	-	-	-	-	-	-	4.9880
3.	0	-	-	-	-	-	-	5.7874
4.	-	0.090	-	-	-	-	-	2.7498
5.	-	0.105	-	-	-	-	-	6.2701
6.	-	-	0.155	-	-	-	-	10.9058
7.	-	-	0.155	-	-	-	-	9.7969
8.	-	-	-	0.290	-	-	-	11.7831
9.	-	-	-	0.285	-	-	-	11.5240
10.	-	-	-	-	0.675	-	-	10.6619
11.	-	-	-	-	0.570	-	-	10.0052
12.	-	-	-	-	-	0.965	-	11.8709
13.	-	-	-	-	-	1.125	-	11.7295
14.	-	-	-	-	-	-	2.295	9.9347
15.	-	-	-	-	-	-	1.975	10.6428

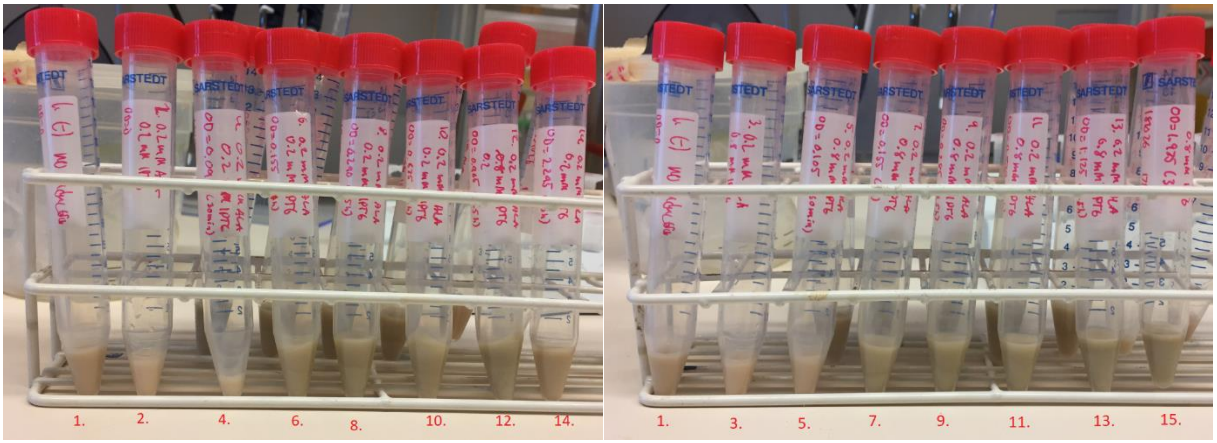


Figure A10. Cell pellets from second round of condition optimization with scaffold 1780. Sample 1 was the negative control (no induction), samples 2,4,6,8,10,12 and 14 were induced with 0.2 mM IPTG at different OD_{600}, while 3,5,7,9,11,13 and 15 were induced with 0.8 mM IPTG. The ALA concentration was kept constant at 0.2 mM.

The resulting SDS-PAGE result is shown in **Figure A11**. As for the results of the second round of optimization with scaffold 485, the gel was loaded with samples from different concentrations of cells. Therefore, it is not valid to compare these results with each other since the amount of cells were not properly normalized. This is especially evident for samples 5 and 6, showing more distinctive bands for every protein expressed. In addition, the samples with high OD_{600} (samples 12-14) seemed to have lower amounts of cells. Thus, this result can be used to strengthen the results presented before (**Results 3.3.2**).

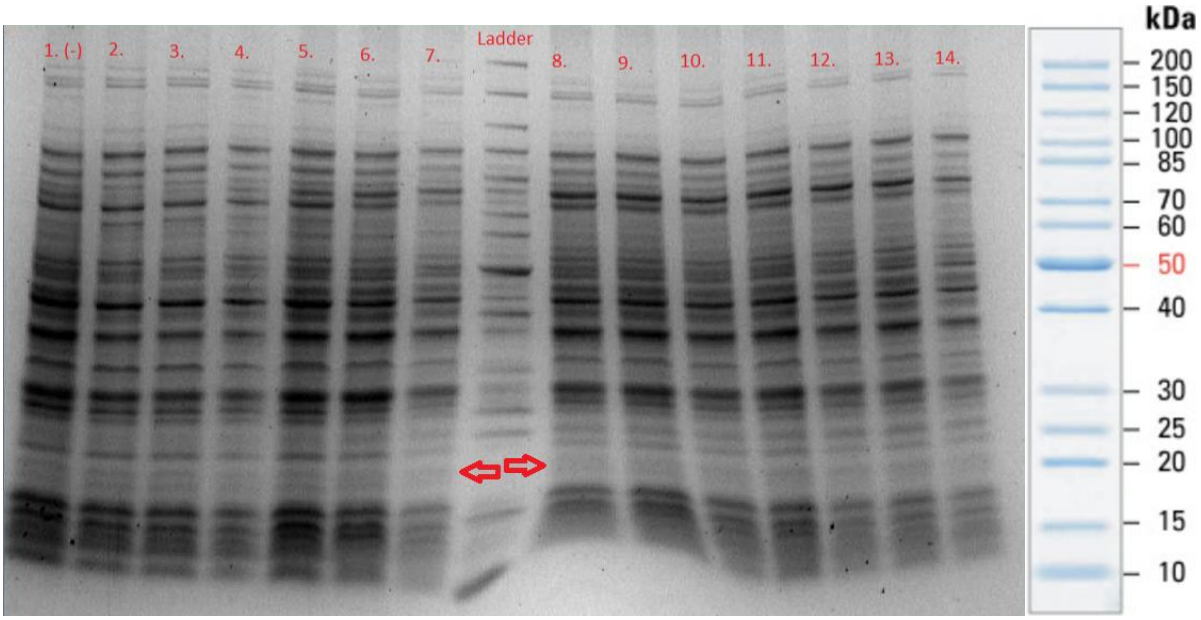


Figure A11. Second round of condition optimization with scaffold 485. The cultures were induced at different OD_{600} with 0.2 mM ALA and 0.2 or 0.8 mM IPTG. The negative control (sample 1) was not induced. For complete set-up, see Materials and Methods section 2.5.

7.3.4 Expression with all scaffolds

Apart from optimization of scaffolds 485 and 1780 with expression overnight, an attempt in catching the hemoglobin before oxidizing for all scaffolds were also done. In this attempt, the optimal condition chosen based on previous results was induction with 0.3 mM ALA and 0.8 mM IPTG at high OD_{600} (<2). In total, three flasks of each scaffold were cultivated and the cells were harvested after 4, 6 and 8 hours of induction, one of each scaffold. In addition, the cultures were bubbled with carbon monoxide for 20 seconds in order to saturate the solution and avoid oxidation. The obtained OD_{600} and cell weights can be seen in **Table A8**, while the cells pellets can be seen in **Figure A12**.

Unfortunately, the cells had to be induced at an earlier stage than the desired OD_{600} due to shortage of time. Similar results as before were observed, where the cells pellets showed a brown color instead of the desired pink/red, indicating oxidized hemoglobin for all scaffolds.

Thus, the decreased time of expression did not rescue the hemoglobin from being oxidized. However, different layers of cells in the pellets showed a darker brown color, which points to that some cells are expressing the protein to a higher extent than others.

Table A8. Inducing conditions, measured OD_{600} after 3 h and cell weights for all samples. Sample 1 was the negative control (no induction), samples 2, 5 and 8 were from scaffold 485. Samples 3, 6 and 9 were from scaffold 1780 and samples 4, 7 and 10 were from scaffold 313051. The samples were harvested after different times of expression. In addition, a 2 liter culture was induced at left for expression overnight.

	0.3 mM δ-ALA	0.8 mM IPTG	Time after induction (h)	OD_{600}~ 3h	Cell Weight (g/L)
*1. 1780	-	-	-	1.345	5.3361
2. 485	x	x	4	1.155	2.3544
3. 1780	x	x	4	1.200	2.6543
4. 313051	x	x	4	1.230	2.7154
5. 485	x	x	6	1.170	3.7735
6.1780	x	x	6	1.040	3.8068
7.313051	x	x	6	1.020	2.8891
8.485	x	x	8	1.075	3.1105
9.1780	x	x	8	1.275	3.3899
10.313051	x	x	8	1.530	3.3494
485 (2L)	x	x	overnight	0.935	3.8566

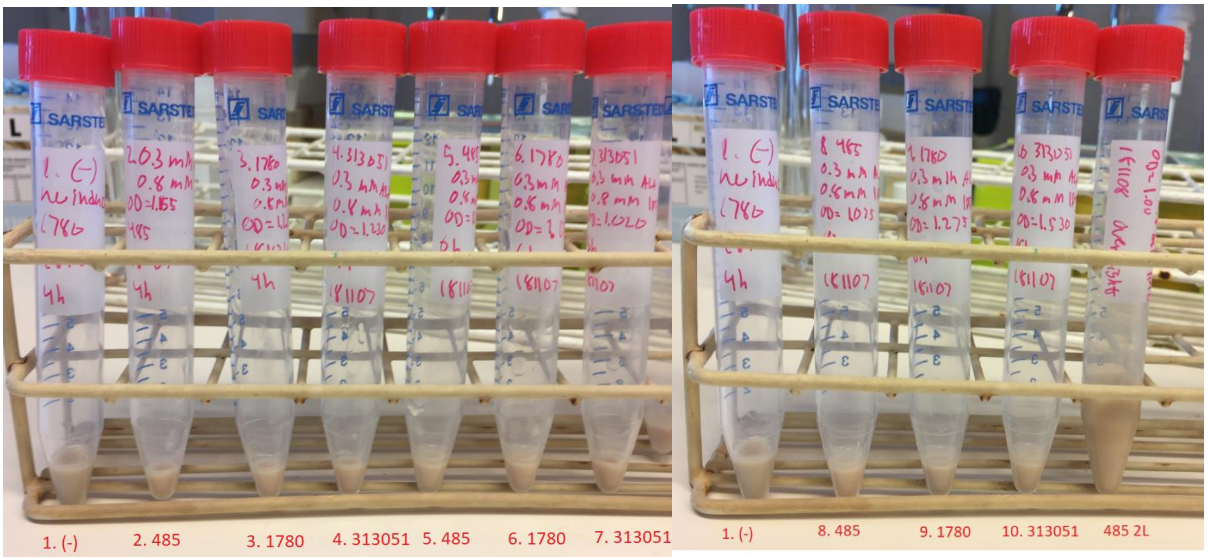


Figure A12. Cell pellets of the samples in Table Aq. Sample 1 was the negative control (no induction), samples 2, 5 and 8 were from scaffold 485. Samples 3, 6 and 9 were from scaffold 1780 and samples 4, 7 and 10 were from scaffold 313051. The samples were harvested after different times of expression. In addition, a 2 liter culture was induced at left for expression overnight.

7.3.4.1 Expression with all scaffold grown in 2L cultures

In this cultivation, a total volume of 0.67 L in 2L-flasks was used for all the scaffolds. As before, the cell cultures were induced with 0.3 mM ALA and 0.8 mM IPTG but at higher OD_{600} because of the tendency of the cells to express more Hb in more dense cultures and this was done twice. All conditions can be seen in **Table A9** and the cell pellets in **Figure A13**.

All scaffolds were induced at an OD_{600} well above 2 (but not the negative control). The induced cultures showed a pinkish color while the negative control displayed a greener pellet. The results was the same for both runs. Therefore, it seems that reduced oat Hb has been produced in all cultures, pointing to the fact that the Hb expression is facilitated by higher OD_{600} before induction. Thus, this might be the kind of condition that is best suited for expression of oat Hb. As a side note, the 0.6 M ALA stock was remade before this batch, which might have influenced the outcome of the experiment in a positive way.

Table A9. Conditions, OD_{600} and cell weight for all scaffold grown in 2L-flasks. Duplicates of the negative control and induced samples were made. The cultures were induced at OD_{600}>>2.

	0.3 mM δ-ALA	**0.8 mM IPTG**	**OD_{600}~ 3h**	**OD_{600}~ 4h**	**OD_{600}~ 4.5h**	**OD_{600}~ 5h**	**Cell Weight (g/L)**
*1.1 1780	-	-	0.820	-	5.450	-	10.2630
*1.2 1780	-	-	0.505	2.245	4.105	-	8.0701
2.1 485	x	x	0.435	-	4.245	-	4.7709
2.2 485	x	x	0.300	1.695	2.910		3.7881
3.1 1780	x	x	0.735	-	5.045	-	6.2918
3.2 1780	x	x	0.345	1.465	2.810	-	3.7121
4.1 313051	x	x	0.165	-	1.855	3.540	4.5513
4.2 313051	x	x	0.245	1.535	2.940	-	3.7940

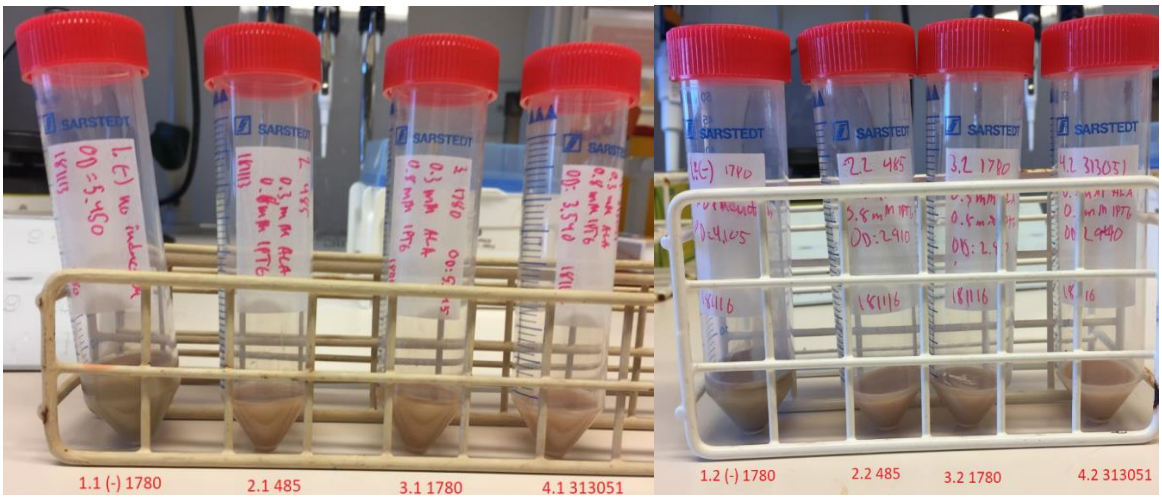


Figure A13. Cell pellets of all scaffolds including negative controls. The cultures grown the first round can be seen in the left picture, while the duplicate is shown in the right one. All induced samples displayed a pinkish color while the negative controls had a green character.

7.4 Quantification of expressed oat Hb

7.4.1 Quantification of unpurified Hb in 2L-cultures

In order to determine Hb concentration and activity, a spectrophotometric assay was used measuring the absorbance at different wavelengths. The results can be seen in **Table A10** and were obtained using the different graphs presented below for each scaffold. All induced cultures show similar results for max peak location, obtained real absorbance and Hb concentration. In addition, even though the negative control also showed Hb expression, the induced cultures showed higher Hb amounts for all scaffolds (485 and 1780 based on +NaD curve and 313051 based on +CO curve.

Also, based on this, the amount of cells needed to obtain at least 2 mg Hb (suitable for purification) was calculated. The results can be seen in **Table A11**. The data were used to cultivate enough cells that were later used for Hb purification.

Table A10. Determined data from spectrophotometric assay for all scaffold and negative control.

Scaffold	Max peak +NaD (nm)	Max peak +CO (nm)	Real Abs (+NaD)	Real Abs +CO	Shift Nad-CO (nm)	Shift NaD-Standard (nm)	Conc. (mM, based on +NaD)	Conc. (mM based on +CO)	Amount (mg) (based on +CO)
(-) (1780)	424	421	0.0654	0.909	3	1	0.00138	0.0015	0.405 (0.461)
485	424	423	0.114	0.105	1	1	0.00241	0.0020	0.573 (0.496)
1780	424	423	0.0807	0.111	1	1	0.00170	0.0022	0.500 (0.645)
313051	424	421	0.0652	0.122	3	0	0.00137	0.0024	0.332 (0.583)

Table A11. Calculated amount of cells needed to get a least 2 mg Hb to be used for purification.

Scaffold	Amount of cells (g)	Volume (L)	Concentration (g/L)	Cells needed for 2 mg Hb (g)
(-) (1780)	6.876	0.67	10.23	-
485	5.735	1.34	4.280	20.03
1780	6.073	1.34	5.002	26.82
313051	5.637	1.34	4.21	33.98

7.4.1.1 Quantification of Hb in scaffold 1780

The resulting graph from the spectrophotometric assay for scaffold 1780, showing absorbance as a function of wavelength, can be seen in **Figure A14**. As before, three different curves were generated, standard Hb (nothing added, blue), +NaD (red) and +CO (gray) were plotted. Similar results as the other scaffolds were obtained, even though the peak at 560 nm was less defined. The graph where used to determine the data presented in **Table A10.**

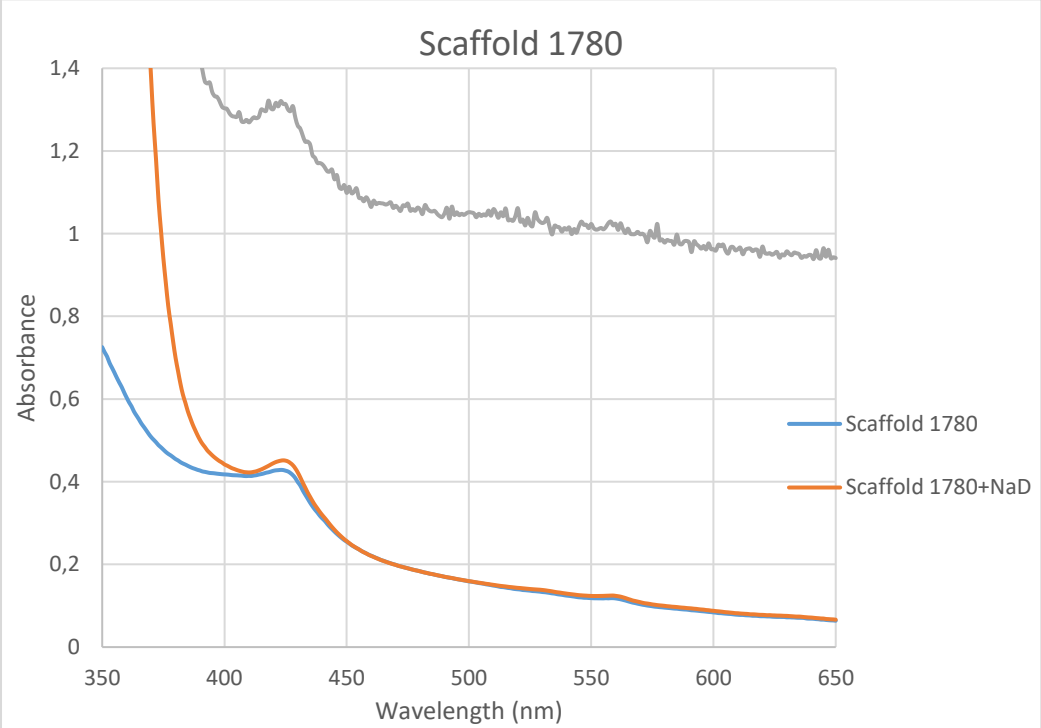


Figure A14. Absorbance as a function of wavelength for scaffold 1780. The three curves represent standard 1780 (nothing added, blue) +NaD (red) and +CO (gray).

7.4.1.2 Quantification of Hb in scaffold 313051

The resulting graph from the spectrophotometric assay for scaffold 313051, showing absorbance as a function of wavelength, can be seen in **Figure A15**. As before, three different curves were generated, standard Hb (nothing added, blue), +NaD (red) and +CO (gray) were plotted. Similar results as the other scaffolds were obtained, even though the peak at 560 nm was less defined. In addition, the absorbance for the +CO curve was the same as the other curves, which was not the case for the other scaffolds. The graph where used to determine the data presented in **Table A10.**

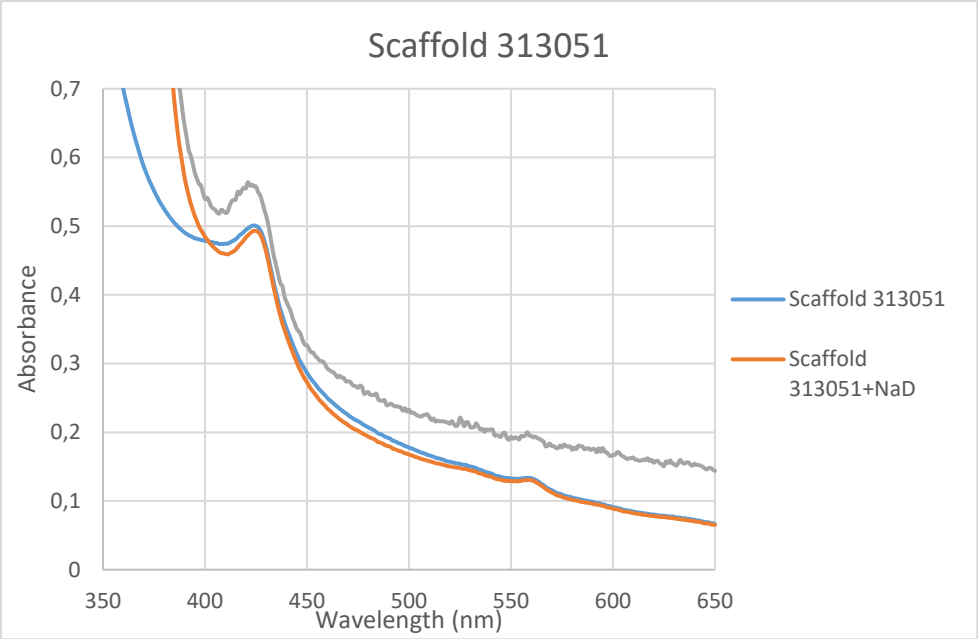


Figure A15. Absorbance as a function of wavelength for scaffold 313051. The three curves represent standard 313051 (nothing added, blue) +NaD (red) and +CO (gray).

7.4.1.3 Quantification of Hb in negative control

The resulting graph from the spectrophotometric assay for the negative control (non-induced scaffold 1780), showing absorbance as a function of wavelength, can be seen in **Figure A16**. As before, three different curves were generated, standard Hb (nothing added, blue), +NaD (red) and +CO (gray) were plotted. Similar results as the other scaffolds were obtained, even though the peak at 560 nm was less defined. This shows that Hb is produced even though without induction, suggesting tht the plasmid is leaking. The graph where used to determine the data presented in **Table A10.** It can be seen here that the induced scaffolds express more Hb than the negative control.

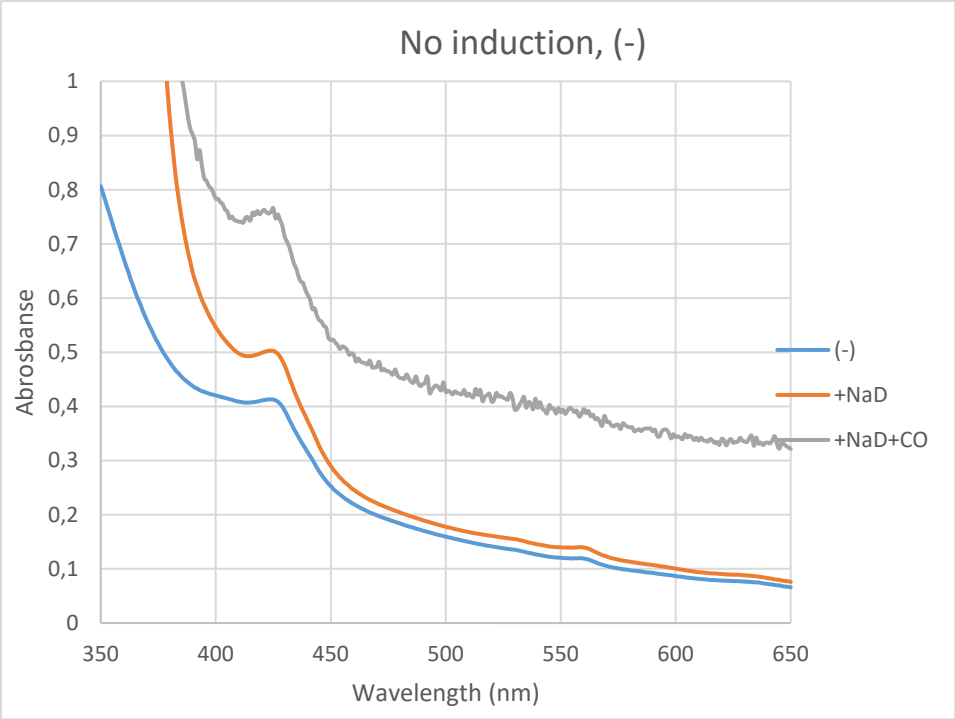


Figure A16. Absorbance as a function of wavelength for the negative control (non-induced scaffold 1780). The three curves represent standard 1780 (nothing added, blue) +NaD (red) and +CO (gray).

7.4.2 Quantification of purified Hb through cation exchange

In order to purify the Hbs, 25-30 grams of cells for every scaffold were cultivated. The conditions and total cell weight for these cultivations can be seen in **Table A12**. The cultivation involved eight 2L-flasks containing 750 mL total volume. In the case of the Sc485, almost the doubled amount of cells were produced in comparison with the other scaffolds. It should be pointed out that a portion of these cells displayed a darker color than the rest of the cells and gave rise to a higher cells mass. The reason for this is unclear but these cells might not have been induced in the same way as the rest of the cells e.g no CO, even though this seems unlikely. In any case, only cells that displayed a pink/light brown color was used in both cation and anion purifications.

Table A12. Cultivated cells used in Hb purification. Induction amounts and OD_{600} for induction can be seen here, as well as total cell weight. The cells were used in both cation and anion purifications.

	0.3 mM δ-ALA	**0.8 mM IPTG**	**OD_{600}~ 3h**	**OD_{600}~ 4h**	**OD_{600}~ 4.5h**	**OD_{600}~ 5h**	**OD_{600}~ 6h**	**Total Cell Weight (g/L)**
1.485	x	x	0.895	-	4.000	-	-	8.338
2.485	x	x	0.545	-	2.635	-	-	
3.485	x	x	0.590	-	2.950	-	-	
4.485	x	x	0.715	-	3.420	-	-	
5.485	x	x	0.473	-	3.805	-	-	
6.485	x	x	0.288	-	3.565	-	-	
7.485	x	x	0.580	-	4.591	-	-	
8.485	x	x	0.603	-	5.695	-	-	
1.1780	x	x	0.390	-	2.275	-	-	4.837
2.1780	x	x	0.445	-	4.470	-	-	
3.1780	x	x	0.450	-	4.420	-	-	
4.1780	x	x	0.725	-	4.320	-	-	
5.1780	x	x	0.860	-	4.310	-	-	
6.1780	x	x	0.805	-	4.735	-	-	
7.1780	x	x	0.530	-	2.260	-	-	
8.1780	x	x	0.560	-	5.050	-	-	
1.313051	x	x	-	0.225	-	-	4.085	4.981
2.313051	x	x	-	0.250	-	-	3.915	
3.313051	x	x	-	0.170	-	-	3.230	
4.313051	x	x	-	0.265	-	-	4.710	
5.313051	x	x	-	0.165	-	-	2.890	
6.313051	x	x	-	0.155	-	-	3.215	
7.313051	x	x	-	0.270	-	-	4.105	
8.313051	x	x	-	0.350	-	-	4.665	

7.4.2.1 Quantification of fractionated Hb for scaffold 1780

In the first attempt to purify the oat Hbs, cation exchange by the CaptoS HiScreen™ was used. The most important data for the attempted purification of Sc1780 can be seen in **Table A13** and **Figure A17**, where the untreated and fractionated Sc1780 can be found. Both fractionated and unfractionated Hb shifted the location of the maximum peak, as was seen in previous results. In addition, the calculated Hb concentration and amount differs substantially between these two samples. The measured amount of the fractionated Hb was 98.66% in comparison with the untreated sample. However, this fraction was not collected during the elution phase of the purification and the remaining 1.34% were still bound to the column. This indicates the Hb was not efficiently bound to the column and different measures have to be taken to increase the level of bound Hb.

Table A13. Analyzed data from spectrophotometric assay for Sc1780 using cation exchange. Here, calculated real absorbance, concentrations and amounts of Hb for fractionated and unfractionated Sc1780 can be found.

Scaffold	Max peak +NaD (nm)	Max peak +CO (nm)	Real Abs (+NaD)	Real Abs +CO	Conc. (mM, based on +NaD)	Conc. (mM based on +CO)	Amount (mg) (based on +CO)	Difference Abs (for NaD)	Percentage in column (%,for NaD)
1780	424	428	0.0325	0.0644	0.00086	0.00163	0.448 (0.836)	0.00045	1.34
Fraction 1780	424	425	0.0321	0.0478	0.00084	0.00118	0.442 (0.619)		

This small difference in amount of Hb can also be seen in **Figure A17**. It is evident that the curves for with and without NaD are located close to each other, meaning that most of the Hb has run-through the column and collected into this fraction. Thus, Sc1780 was the Hb that bound to the least extent in comparison with the other scaffolds.

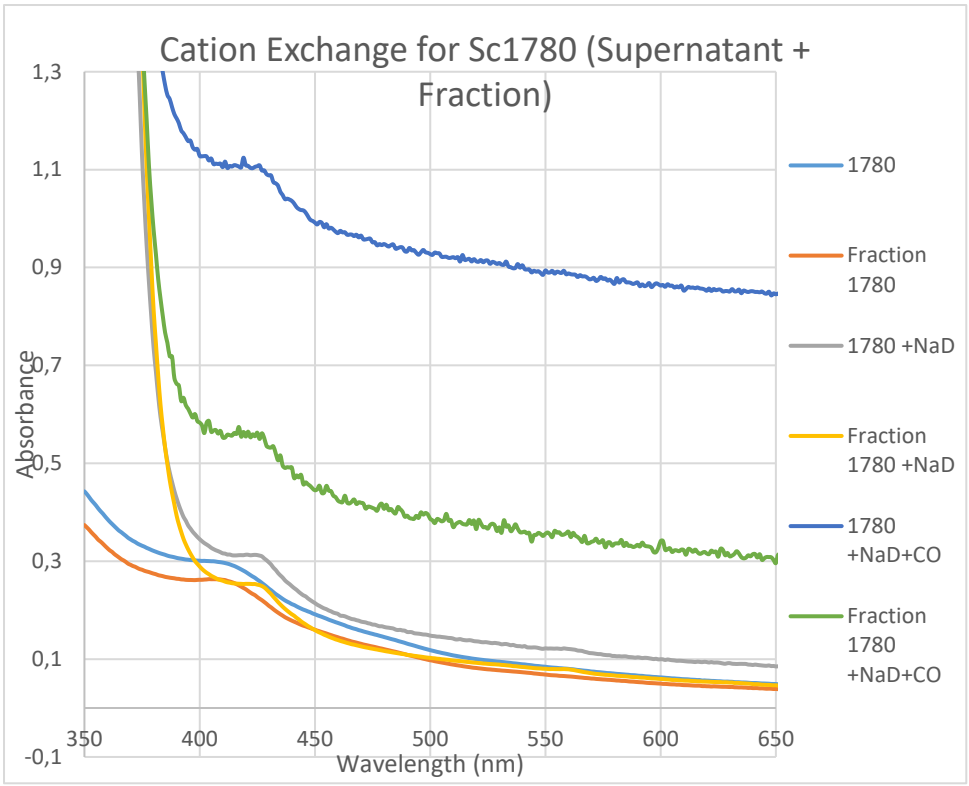


Figure A17. Absorbance as a function of wavelength for Sc1780 using cation exchange. The curves for untreated supernatant (light blue, grey and dark blue) and fractionated Sc1780 (red, yellow and green) can be seen here.

7.4.2.2 Quantification of fractionated Hb for scaffold 313051

In the first attempt to purify the oat Hbs, cation exchange by the CaptoS HiScreen™ was used. The most important data for the attempted purification of Sc313051 can be seen in **Table A14** and **Figure A18**, where the untreated and fractionated Sc313051 can be found. Both fractionated and unfractionated Hb did not shift the location of the maximum peak, not between the NaD and CO curves. In addition, the calculated Hb concentration and amount differs substantially between these two samples. The measured amount of the fractionated Hb was 85.5% in comparison with the untreated sample. However, this fraction was not collected during the elution phase of the purification and the remaining 14.5% were still bound to the column. This portion was probably eluted during the cleaning of the column.

Table A14. Analyzed data from spectrophotometric assay for Sc313051 using cation exchange. Here, calculated real absorbance, concentrations and amounts of Hb for fractionated and unfractionated Sc313051 can be found.

Scaffold	Max peak +NaD (nm)	Max peak +CO (nm)	Real Abs (+NaD)	Real Abs +CO	Conc. (mM, based on +NaD)	Conc. (mM based on +CO)	Amount (mg) (based on +CO)	Difference Abs (for NaD)	Percentage in column (%,for NaD)
313051	424	424	0.0334	0.0277	0.00088	0.00069	0.564 (0.440)	0.00013	14.5
Fraction 313051	424	424	0.0285	0.0356	0.00075	0.00081	0.482 (0.523)		

This small difference in amount of Hb can also be seen in **Figure A18**. It is evident that the curves for with and without NaD are located close to each other, meaning that most of the Hb has run-through the column and collected into this fraction. This is also the case for the +CO curves. Thus, more Hb from Sc313051 was bound compared with Sc1780, but less then for Sc485.

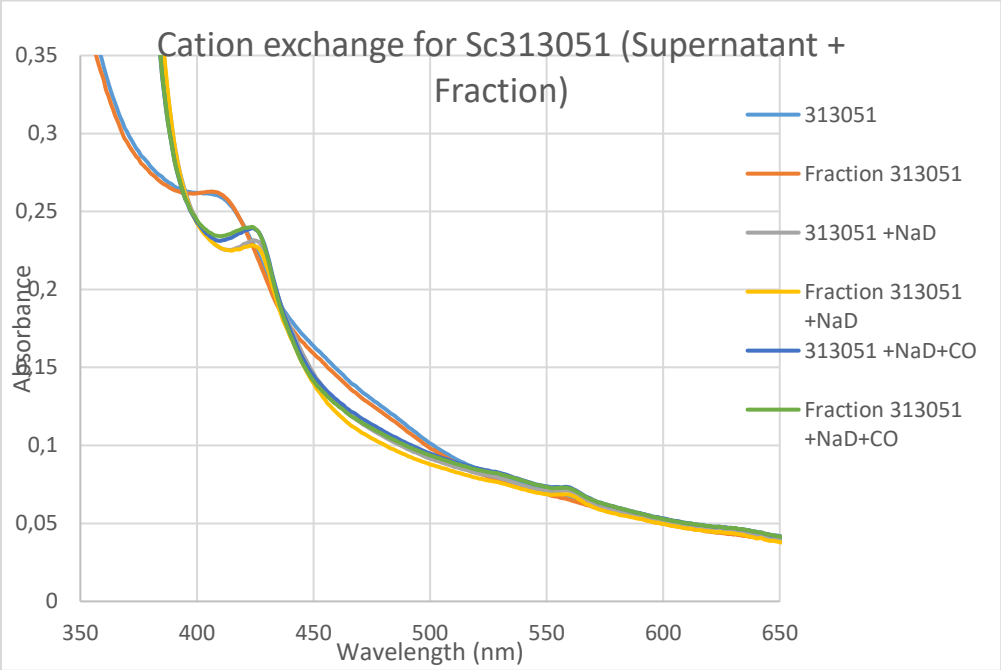


Figure A18. Absorbance as a function of wavelength for Sc313051 using cation exchange. The curves for untreated supernatant (light blue, grey and dark blue) and fractionated Sc313051 (red, yellow and green) can be seen here.

7.4.3 Quantification of purified Hb using anion exchange

7.4.3.1 Quantification of fractionated Hb from Sc485

As before, the untreated supernatant was compared to the fractionated Sc485. The fractionated samples were divided into eluted Sc485 (eluted during the elution phase) and run-through sample (not collected during the elution phase). The analyzed data was shown in the **Results 3.4** and the remaining plots from these measurements can be found in **Figure A19** and **Figure A20**.

The plot with added NaD can be seen in **Figure A19**. The sample not used in the anion exchange (485 Supernatant, blue) showed higher absorbance than both run-through fractions as before, but not the Q elute. In addition, this curve show very oscillating data points with a higher degree of uncertainty. Thus, this plot was not included in the main results.

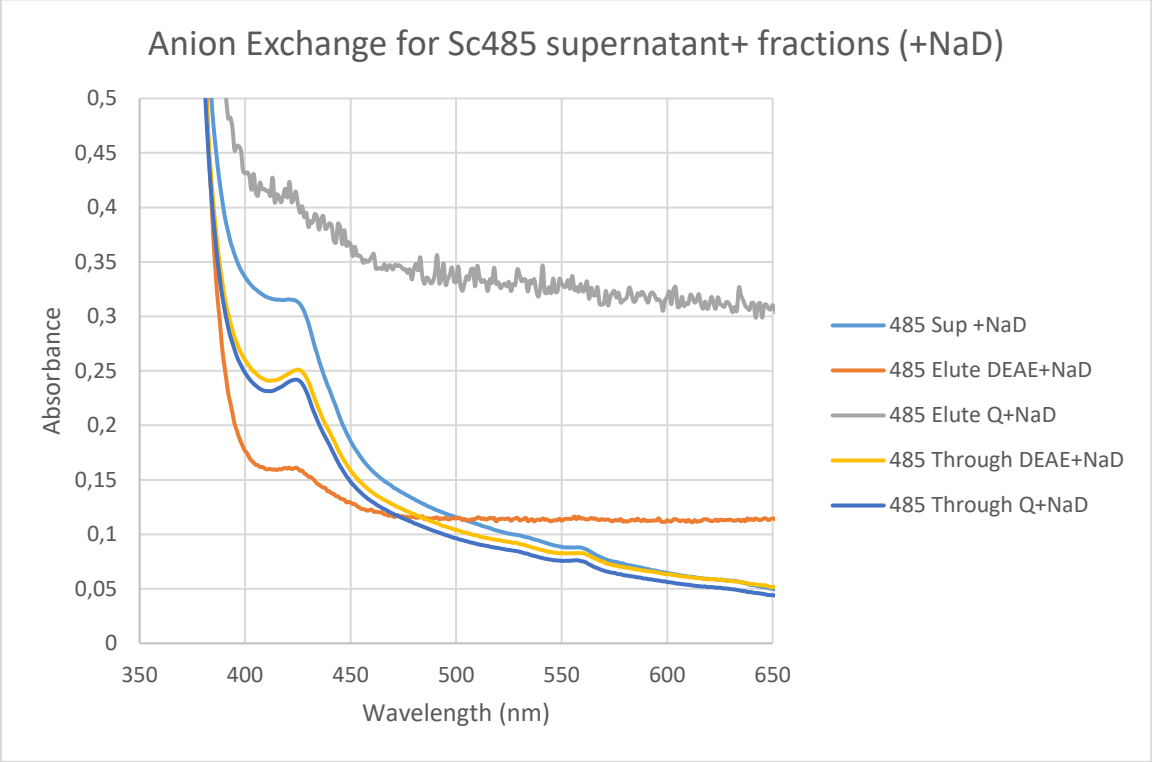


Figure A19. Absorbance as a function of wavelength for Sc485 +NaD using anion exchange. The curves for the eluted fractions (red and grey), run-through fractions (yellow and dark blue) and the untreated supernatant (light blue) can be seen in this figure.

The plot when CO was added is shown in **Figure A20.** As can be seen here, when CO, the absorbance is increased. Even here, the uncertainty of these results are higher in comparison with the data presented in the **Results 3.4**. Also, the elute from Q cannot be seen here due to the high absorbance. Thus, this plot was not used in the results.

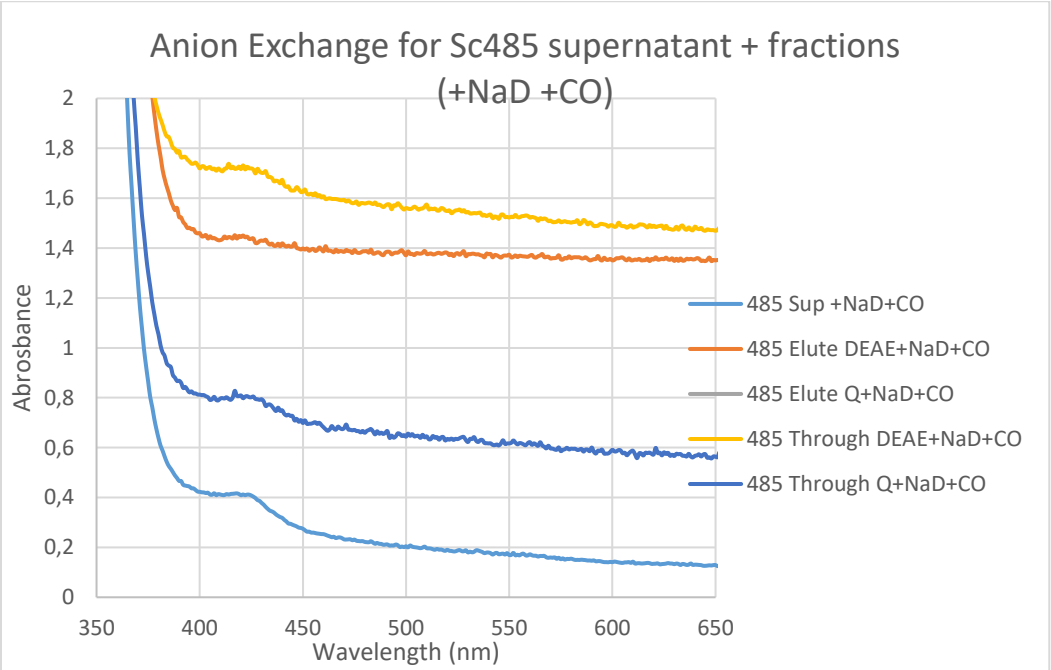


Figure A20. Absorbance as a function of wavelength for Sc485 +NaD +CO using anion exchange. The curves for the eluted fraction (red), run-through fractions (yellow and dark blue) and the untreated supernatant (light blue) can be seen in this figure. The eluted sample from Q are not shown due to too high absorbance values.

7.4.3.2 Quantification of fractionated Hb from Sc1780

The calculated data from the spectrophotometric assay for Sc1780 as well as the resulting plot (+NaD) can be seen in **Table A15** and **Figure A21**, respectively. Here, less Hb was found in the eluted fractions (14.7%) in comparison with same results for Sc485, while the majority of Hb was detected in the run-through fractions for both columns. In addition, no significant amount of Hb eluted during the cleaning phase (% in column) could be found.

Table A15. Analyzed data from spectrophotometric assay for Sc1780 using anion exchange. Real absorbance, concentrations and amounts for untreated supernatant and resulting fractions are presented here, as well as different percentages of the location of detected Hb.

Sample	Max peak (nm)	Real Abs (+NaD)	Conc (mM)	Amount (mg)	% Eluted (DEAE)	% Eluted (Q)	% Through (DEAE)	% Through (Q)	% In column (DEAE)
1780	424	0.0414	0.00109	0.960	14.7	14.7	85.3	85.3	-
1780 Elute DEAE	421	0.00944	0.00025	0.142					
1780 Elute Q	421	0.00944	0.00025	0.142					
1780 Through DEAE	424	0.0353	0.00093	0.819					
1780 Through Q	424	0.0353	0.00093	0.819					

The resulting plot of these data can be seen in **Figure A21,** where untreated supernatant and fractions were treated with NaD. The untreated supernatant (light blue) show distinctively higher absorbance in comparison with the eluted fractions (red and grey), pointing to that most of the Hb was found in the run-through fractions (yellow and dark blue). It should be noted that both curves for the eluted and run-through fractions were closely intertwined, resulting in three clearly visible curves.

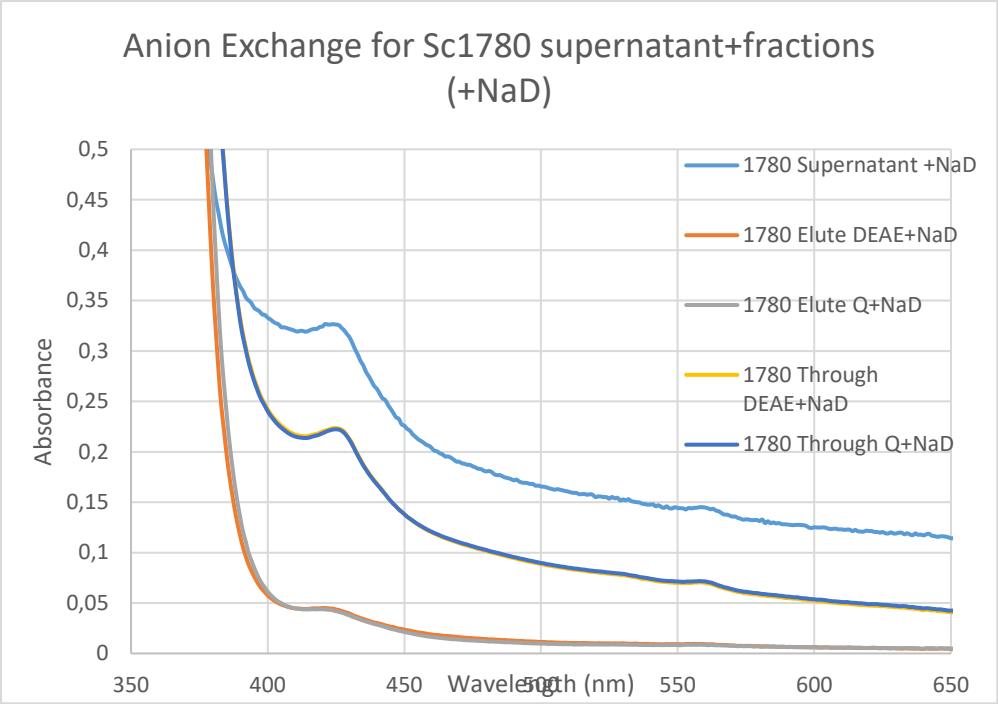


Figure A21. Absorbance as a function of wavelength for Sc1780 using anion exchange. The curves for the eluted fractions (red and grey), run-through fractions (yellow and dark blue) and the untreated supernatant (light blue) can be seen in this figure. It should be noted that both curves for the eluted and run-through fractions were closely intertwined, resulting in three clearly visible curves.

The plot with no added NaD or CO can be seen in **Figure A22**. The sample not used in the anion exchange (1780 Supernatant, blue) showed higher absorbance than both run-through fraction and the eluted fractions. These results are in line with the previous presented results.

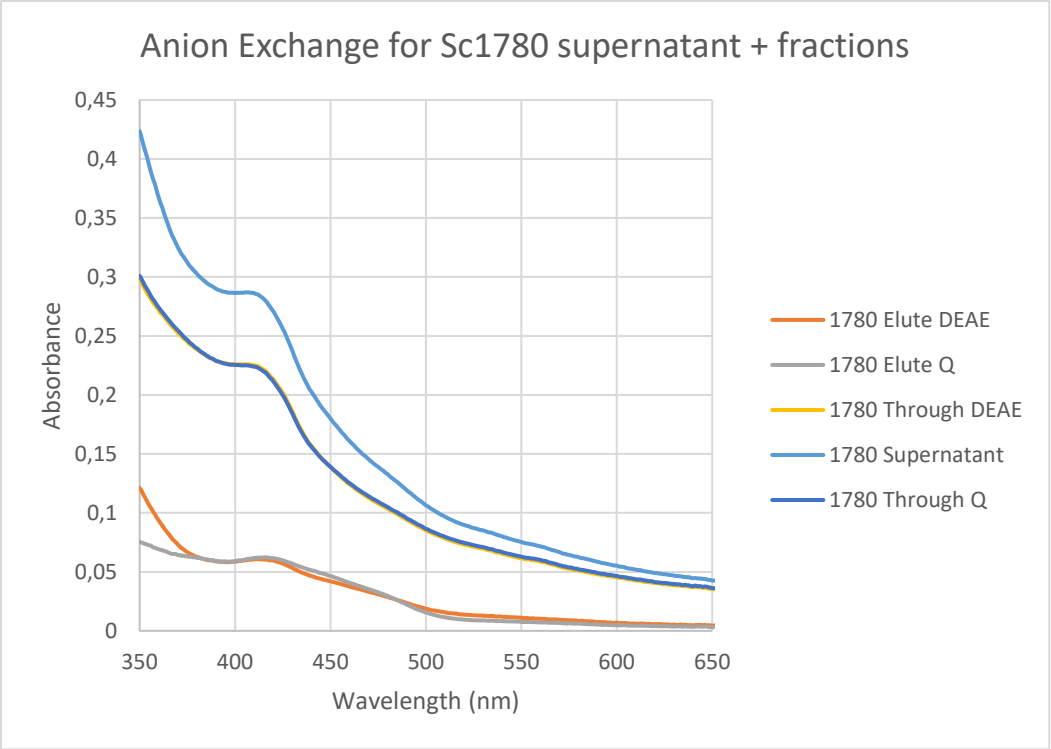


Figure A22. Absorbance as a function of wavelength for Sc1780 using anion exchange. The curves for the eluted fractions (red and grey), run-through fractions (yellow and dark blue) and the untreated supernatant (light blue) can be seen in this figure. Note that the run-through fractions are closely intertwined.

The plot when CO was added is shown in **Figure A23.** As can be seen here, when CO is added, the absorbance is increased. Even here, the uncertainty of these results are higher in comparison with the previous presented data. The results are still in line with previous results but less clear. Thus, this plot was not included in the results.

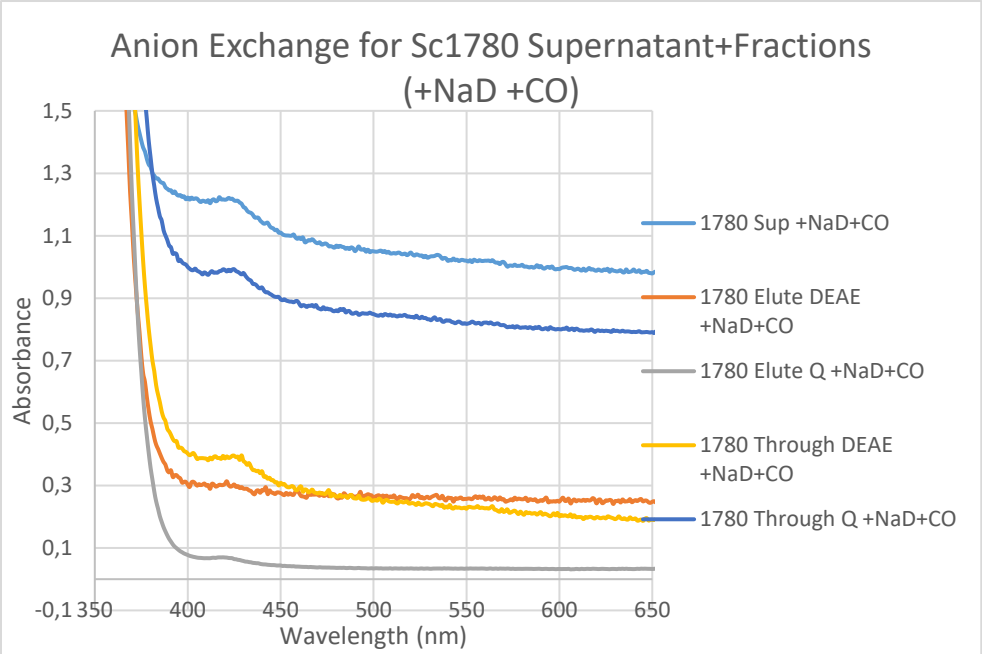


Figure A23. Absorbance as a function of wavelength for Sc1780 +NaD +CO using anion exchange. The curves for the eluted fractions (red and grey), run-through fractions (yellow and dark blue) and the untreated supernatant (light blue) can be seen in this figure.

7.4.3.3 Quantification of fractionated Hb from Sc313051

The calculated data from the spectrophotometric assay for Sc313051 can be seen in **Table A16**. Here, less Hb was found in the eluted fractions (28.7 and 21.7%) in comparison with same results for Sc485, while the majority of Hb was detected in the run-through fractions for both columns. In addition. 6.8% of the total Hb content was not found in elute or run-through for DEAE, suggesting this percentage was stuck in the column and released during cleaning of the column.

Table A15. Analyzed data from spectrophotometric assay for Sc313051 using anion exchange. Real absorbance, concentrations and amounts for untreated supernatant and resulting fractions are presented here, as well as different percentages of the location of detected Hb.

Sample	Max peak (nm)	Real Abs (+NaD)	Conc (mM)	Amount (mg)	% Eluted (DEAE)	% Eluted (Q1)	% Through (DEAE)	% Through (Q1)	% In column (DEAE)
313051	414	0.0367	0.00102	0.626	28.7	21.7	64.5	78.3	6.8
313051 Elute DEAE	414	0.0112	0.00029	0.180					
313051 Elute Q1	413	0.0131	0.00034	0.136					
313051 Through DEAE	414	0.0250	0.00066	0.404					
313051 Through Q1	414	0.0304	0.00080	0.490					

These results are also evident in the spectrophotometric plot for the different fractions, shown in **Figure A24**. The sample not used in the anion exchange (313051 Supernatant, blue) show substantially higher absorbance in comparison to the eluted fractions (red and grey curves), indicating that the majority of the Hb is not collected during the elution phase. Instead, this majority was detected in the run-through fractions (yellow and dark blue curves) for both columns.

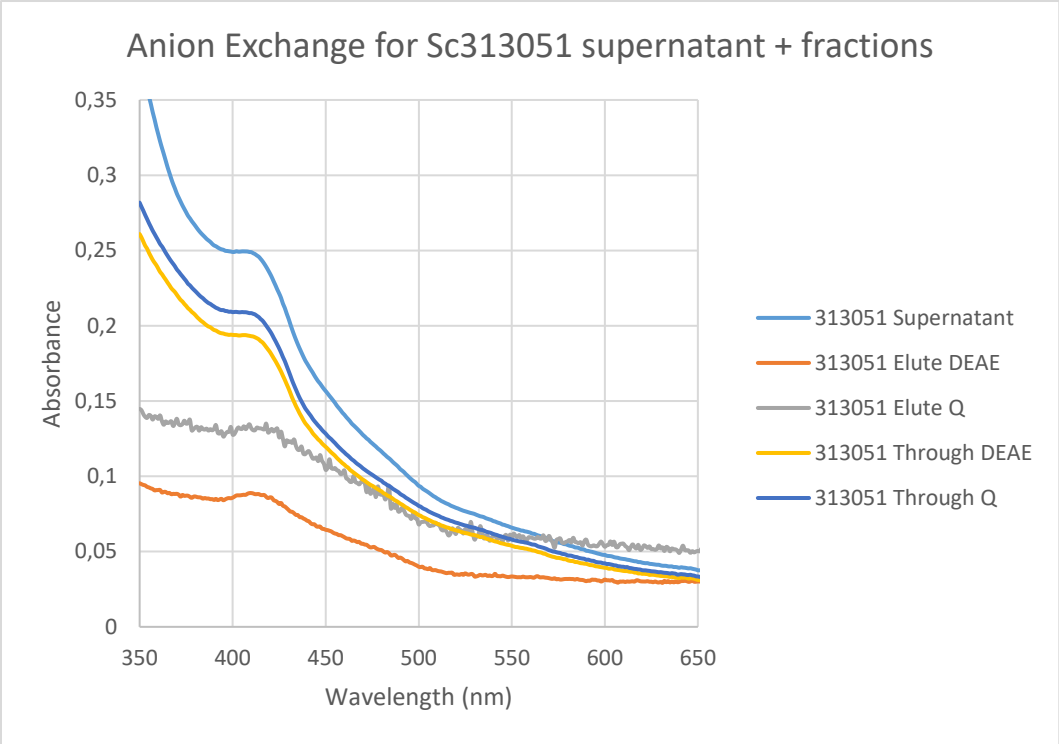


Figure A24. Absorbance as a function of wavelength for Sc313051 using anion exchange. The curves for the eluted fractions (red and grey), run-through fractions (yellow and dark blue) and the untreated supernatant (light blue) can be seen in this figure.

The plot with added NaD can be seen in **Figure A25**. The sample not used in the anion exchange (313051 Supernatant, blue) showed higher absorbance than both run-through fractions as before, but not the Q elute. In addition, this curve show very oscillating data points with a higher degree of uncertainty. Thus, this plot was not included in the main results.

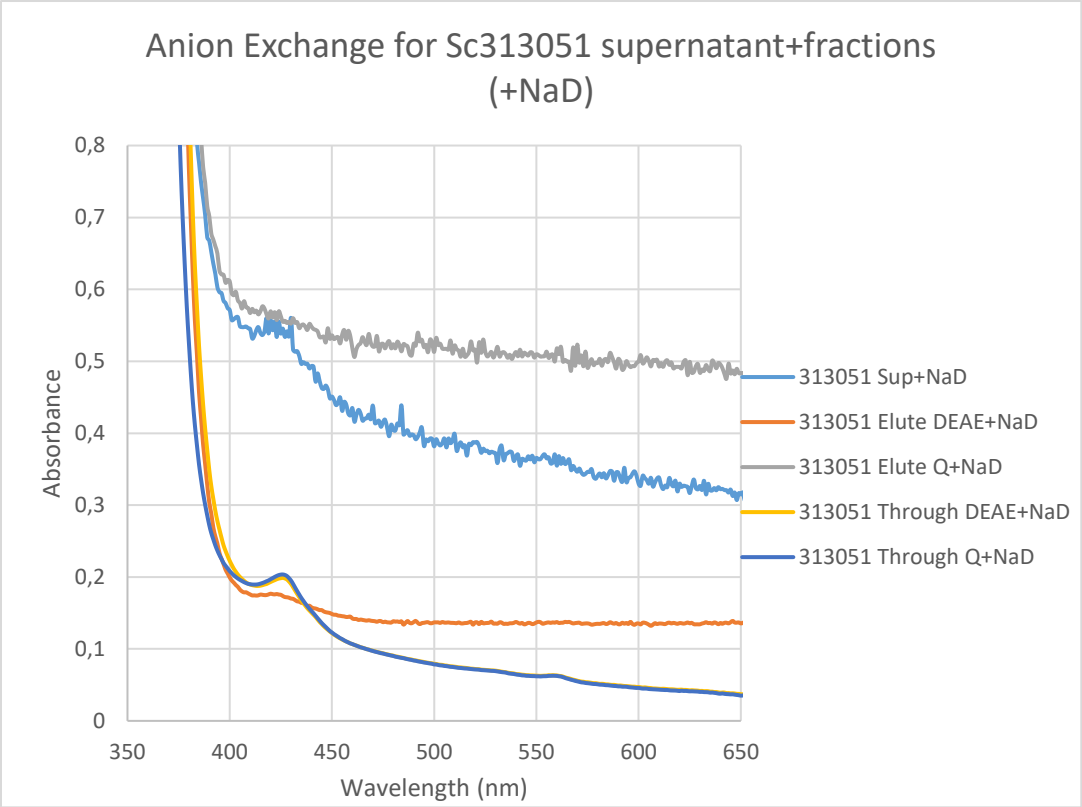


Figure A25. Absorbance as a function of wavelength for Sc1780 +NaD using anion exchange. The curves for the eluted fractions (red and grey), run-through fractions (yellow and dark blue) and the untreated supernatant (light blue) can be seen in this figure.

The plot when CO was added is shown in **Figure A26.** As can be seen here, when CO is added, the absorbance is increased. Even here, the uncertainty of these results are higher in comparison with the previous presented data. No additional information could be obtained from these results, thus not included in the main results.

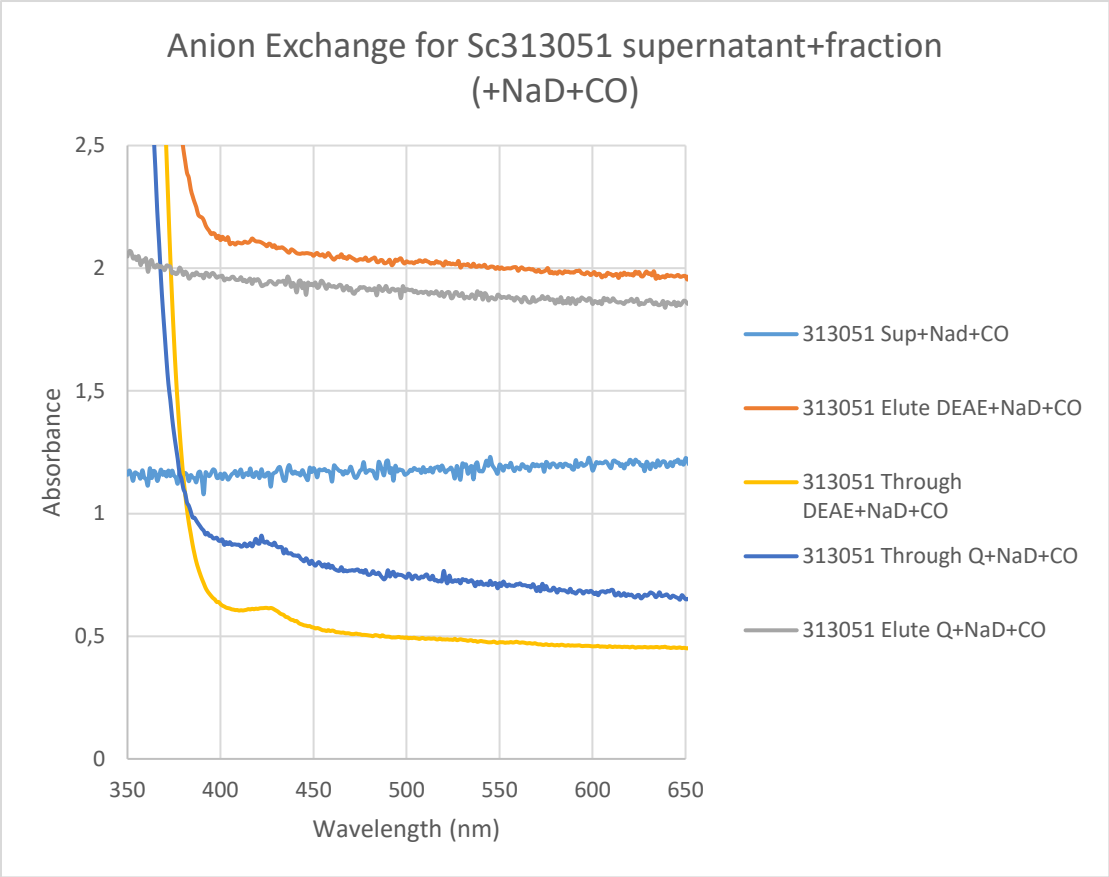


Figure A26. Absorbance as a function of wavelength for Sc313051 +NaD +CO using anion exchange. The curves for the eluted fractions (red and grey), run-through fractions (yellow and dark blue) and the untreated supernatant (light blue) can be seen in this figure.

7.4.3.4 SDS-PAGE of eluted and run-through fractions from anion exchange

The eluted and run-trough fractions for all scaffolds from the anion exchange were also analyzed using SDS-PAGE. The resulting gel can be seen in **Figure A27**, where the labelling of the lanes are shown in **Table A17**. As can be seen here, a large number of different proteins are still present after the usage of the attempted purification. In addition, no significant difference in the predicted location of the oat Hb (indicated by arrows) was observed in this figure, suggesting that the Hbs are present in every sample. However, it is more difficult to distinguish the difference between the eluted and run-through fractions. Thus, no additional information was gained from this data.

Table A17. Labelling of the lanes in the SDS-PAGE from the anion exchange. The resulting gel can be seen in Figure A27.

1.	2.	3.	4.	5.	6.	7.
Ladder	Sc485 Elute DEAE	Sc485 Elute Q1	Sc1780 Elute DEAE	Sc1780 Elute Q1	Sc313051 Elute DEAE	Sc313051 Elute Q1
8.	9.	10.	11.	12.	13.	14.
Sc485 Through DEAE	Sc485 Through Q1	Sc1780 Through DEAE	Sc1780 Through Q1	Sc313051 Through DEAE	Sc313051 Through Q1	Ladder

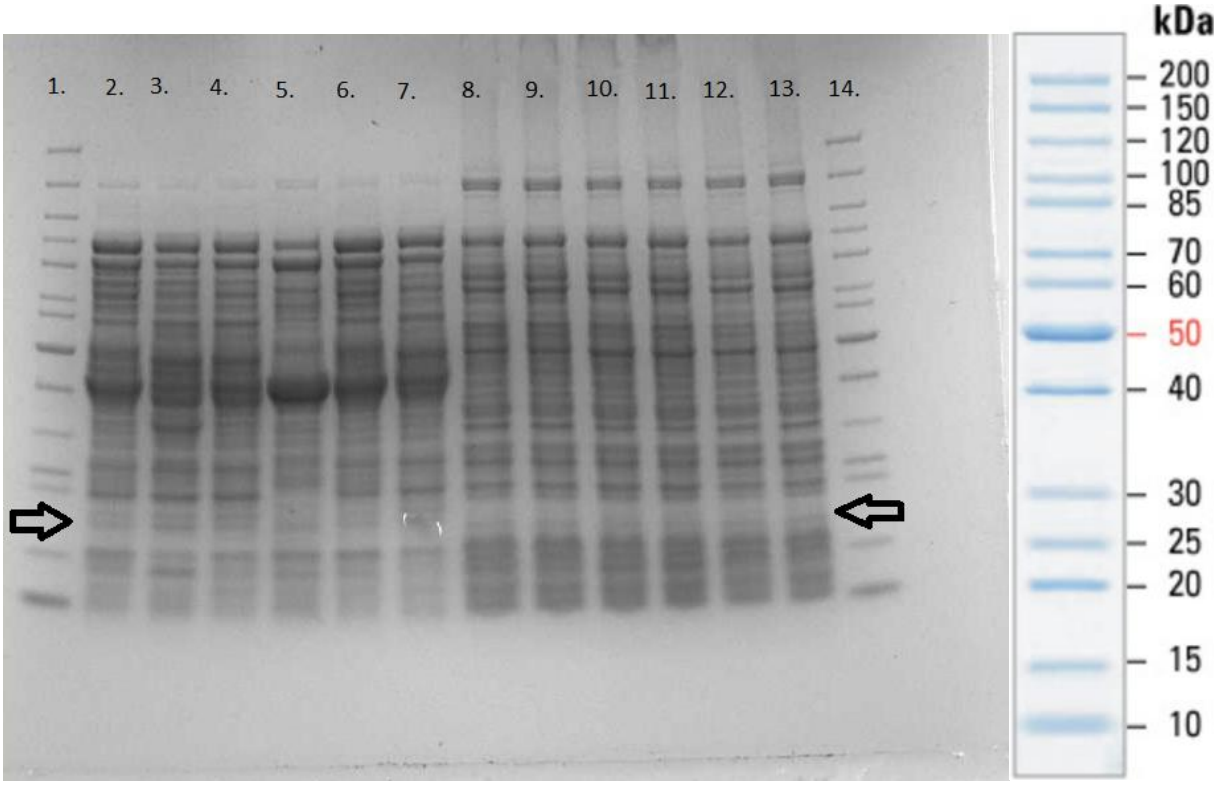


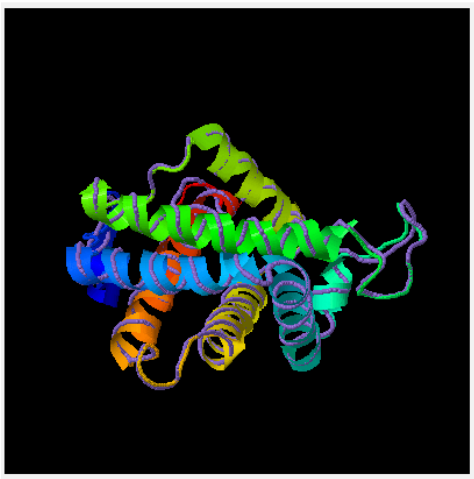
Figure A27. SDS-PAGE from the eluted and run-trough fractions generated from the anion exchange. The labelling of the lanes can be seen in Table A17.

7.5 Predicted 3D structure of oat hemoglobin

In order to get an overview of he estimated 3D-structure of the different oat hemoglobin, online services for this kind of estimate was used. The most frequently used one was Phyre[2] (Kelley L.A *et al*, 2015) and I-Tasser (Roy *et al*, 2010), where the amino acids sequences were used to create an estimation of the 3D-model with comparisons with close-related proteins from different origins. In this section, the estimated structures from the three main scaffolds (485, 1780 and 313051) and also the truncated hemoglobin (445 and 233).

7.5.1 3D-structure of scaffold 1780

Apart from the final model, secondary structure and accessibility predictions presented in the **Results 3.1**, I-Tasser also provided top candidates and suggested ligand binding sites. The top candidates represents the best matching structural analogs of which the prediction has been based on (**Figure A28**). The three top candidates were non-symbiotic hemoglobin from rice, Trema tomentosa and corn, respectively. All the remaining candidates were also hemoglobin or hemoglobin-like proteins from different plants.



Top 10 Identified stuctural analogs in PDB

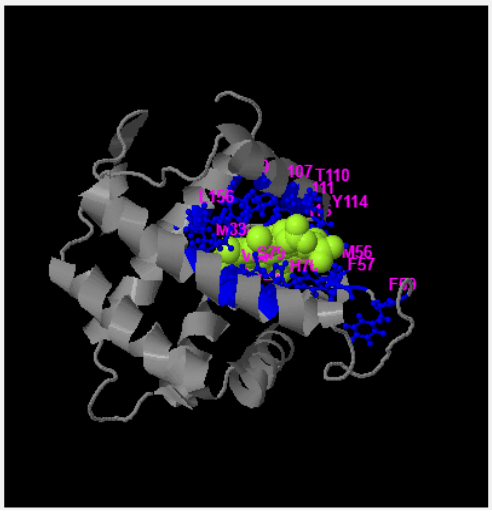
Click to view	Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov	Alignment
○	1	2gnwA	0.969	0.62	0.782	0.982	Download
○	2	3qqqA	0.845	1.47	0.780	0.893	Download
○	3	2r50D	0.843	0.77	0.848	0.863	Download
○	4	1gdm_	0.812	2.05	0.430	0.893	Download
○	5	2oifA	0.786	2.21	0.935	0.905	Download
○	6	1binB	0.772	1.85	0.408	0.845	Download
○	7	1oj6D	0.740	2.25	0.192	0.869	Download
○	8	106m_	0.728	2.50	0.183	0.845	Download
○	9	3ba2A	0.724	2.46	0.149	0.839	Download
○	10	1gvha1	0.722	2.18	0.158	0.827	Download

(a) Query structure is shown in cartoon, while the structural analog is displayed using backbone trace.
(b) Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library.
(c) RMSD[a] is the RMSD between residues that are structurally aligned by TM-align.
(d) IDEN[a] is the percentage sequence identity in the structurally aligned region.
(e) Cov represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

Figure A28. Identified top candidates of structural analogs to oat hemoglobin from scaffold 1780. The three top candidates were non-symbiotic hemoglobin from rice, Trema tomentosa and corn, respectively. All candidates were hemoglobin or hemoglobin-like proteins from the plant kingdom.

Suggested ligand binding sites were also found in the I-Tasser prediction. This result can be seen in **Figure A29,** where the heme group got the highest condifence score of every ligand. The figure also displays where the heme group is located and which residues that are predicted to be invloved in the ligand binding site.

Ligand binding sites



Click to view	Rank	C-score	Cluster size	PDB Hit	Lig Name	Download Complex	Ligand Binding Site Residues
◉	1	0.97	1016	3k8bA	HEM	Rep, Mult	56,57,59,76,79,80,83,107,110,111,114,116,120,121,124,156,160
○	2	0.04	48	1fslB	NIO	Rep, Mult	43,57,76,80,124
○	3	0.01	23	4g1bA	ECN	Rep, Mult	38,39,42,43,57,76,77,80,128,153,156
○	4	0.01	8	1qxdA	FU2	Rep, Mult	14,15,151,154,155,158
○	5	0.01	20	3g53A	LT1	Rep, Mult	27,28,31,38,42,78,81,131,132,149

Download the residue-specific ligand binding probability, which is estimated by SVM.
Download the all possible binding ligands and detailed prediction summary.
Download the templates clustering results.

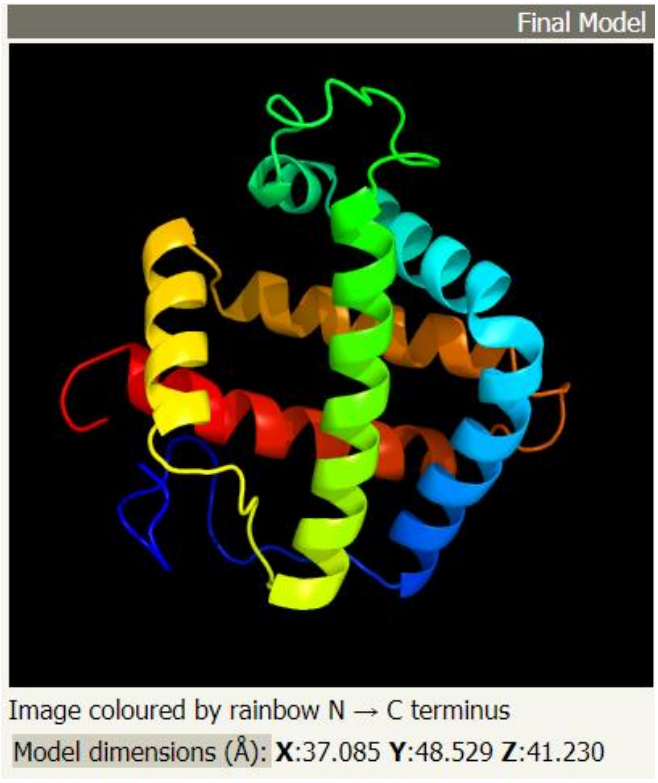
(a) **C-score** is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction.
(b) **Cluster size** is the total number of templates in a cluster.
(c) **Lig Name** is name of possible binding ligand. Click the name to view its information in the BioLiP database.
(d) **Rep** is a single complex structure with the most representative ligand in the cluster, i.e., the one listed in the **Lig Name** column. **Mult** is the complex structures with all potential binding ligands in the cluster.

Figure A29. Predicted ligand binding site in hemoglobin form scaffold 1780. One ligand site be found in the protein, where heme had the highest confidence score. The figure also shows which residues that is involved in ligand binding.

7.5.2 3D structure of scaffold 485

Based on the online tools Phyre[2] (Kelley *et al*, 2015) and I-Tasser (Roy *et al*, 2010), the three-dimensional structure of the oat hemoglobin could be predicted. These model use the amino acid sequence of the hemoglobin and compare it with known proteins with similar characteristics. The predicted model of scaffold 485 from Phyre[2] can be seen in **Figure A30**. Here, the characteristic α- helix motif consisting of 7 helices in close proximity can be found. The confidence key (**Figure A30b**) indicates that the majority of the amino acids can be placed with high confidence, except for the N- and C-terminal. **Figure A30c** displays the predicted secondary structure, where every amino acid is placed in helices, strands or coils, and the predicted accessibility of the different amino acids.

(a)



(b)

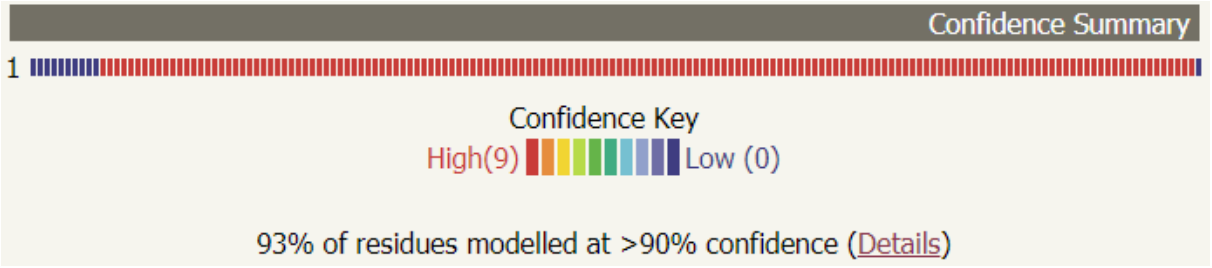
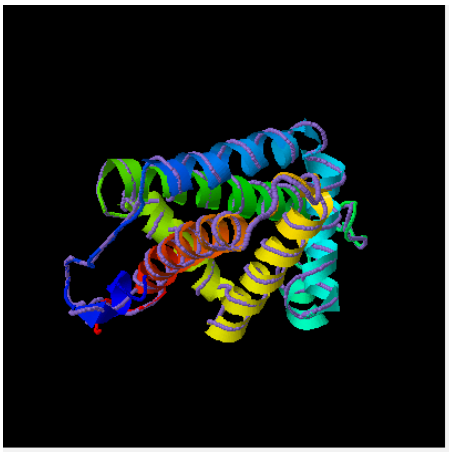


Figure A30. Predicted model of the oat hemoglobin (a) and the confidence summary of this prediction (b) from scaffold 485 (Kelley *et al*, 2015). In addition, the predicted secondary structure and accessibility of the amino acids (c) was generated by I-Tasser (Roy *et al*, 2010). These are not presented here.

I-Tasser also provided top candidates and suggested ligand binding sites. The top candidates represents the best matching structural analogs of which the prediction has been based on (**Figure A31**). The three top candidates were non-symbiotic hemoglobin from rice, Arabidopsis Thaliana and Trema tomentosa, respectively. All the remaining candidates were also hemoglobin or hemoglobin-like proteins from different plants.



Top 10 Identified stuctural analogs in PDB

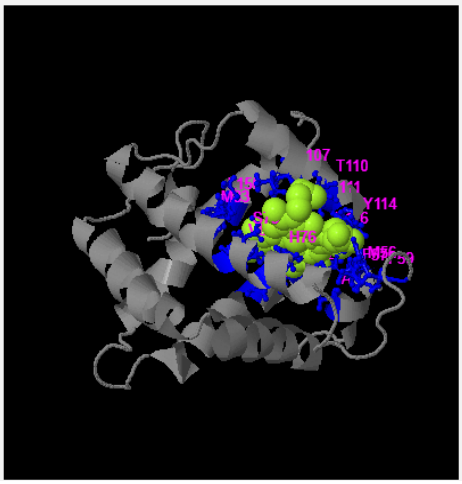
Click to view	Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov	Alignment
○	1	2gnwA	0.971	0.55	0.776	0.982	Download
○	2	3zhwA	0.866	1.38	0.686	0.911	Download
○	3	3qqqA	0.846	1.45	0.767	0.893	Download
○	4	2r50D	0.844	0.76	0.848	0.863	Download
○	5	1gdi_	0.815	2.16	0.437	0.899	Download
○	6	2oifA	0.787	2.20	0.935	0.905	Download
○	7	1binB	0.773	1.82	0.408	0.845	Download
○	8	1oj6D	0.742	2.21	0.192	0.869	Download
○	9	106m_	0.728	2.46	0.183	0.845	Download
○	10	3ba2A	0.724	2.44	0.149	0.839	Download

(a) Query structure is shown in cartoon, while the structural analog is displayed using backbone trace.
(b) Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library.
(c) RMSD[a] is the RMSD between residues that are structurally aligned by TM-align.
(d) IDEN[a] is the percentage sequence identity in the structurally aligned region.
(e) Cov represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

Figure A31. Identified top candidates of structural analogs to oat hemoglobin from scaffold 485. The three top candidates were non-symbiotic hemoglobin from rice, Arabidopsis Thliana andTrema tomentosa, respectively. All candidates were hemoglobin or hemoglobin-like proteins from the plant kingdom.

Suggested ligand binding sites were also found in the I-Tasser prediction. This result can be seen in **Figure A32,** where the heme group got the highest condifence score of every ligand. The figure also displays where the heme group is located and which residues that are predicted to be invloved in the ligand binding site. Unlike scaffold 1780, residue 53 is included here while residue 160 is not.

Ligand binding sites



Click to view	Rank	C-score	Cluster size	PDB Hit	Lig Name	Download Complex	Ligand Binding Site Residues
◉	1	0.97	998	1a6gA	HEM	Rep, Mult	53,56,57,59,76,79,80,83,107,110,111,114,116,120,121,124,156
○	2	0.04	49	1tesA	ETN	Rep, Mult	43,57,76,80,124
○	3	0.02	36	4g1bA	ECN	Rep, Mult	38,39,42,43,57,76,77,80,128,153,156
○	4	0.01	11	1qxdA	FU2	Rep, Mult	14,15,151,154,155,158
○	5	0.01	21	3g53A	LT1	Rep, Mult	27,28,31,38,42,131,132,149

Download the residue-specific ligand binding probability, which is estimated by SVM.
Download the all possible binding ligands and detailed prediction summary.
Download the templates clustering results.

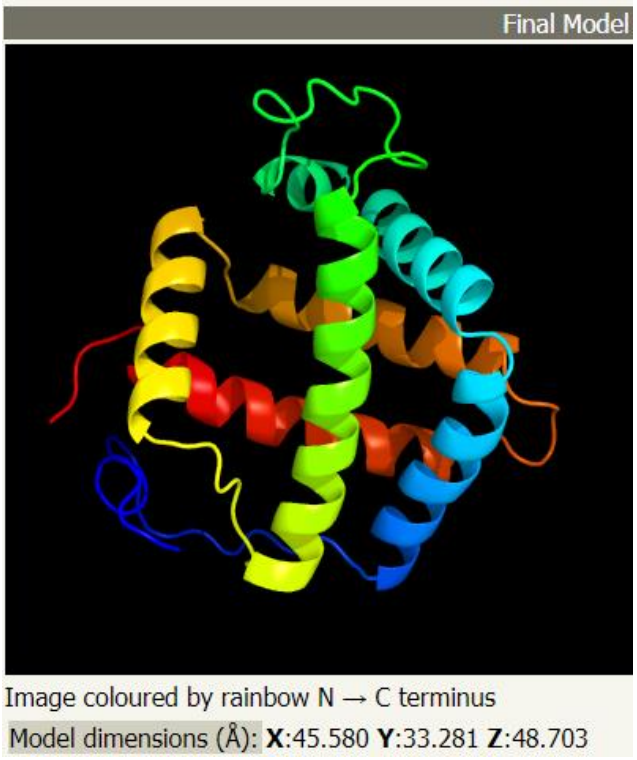
(a) **C-score** is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction.
(b) **Cluster size** is the total number of templates in a cluster.
(c) **Lig Name** is name of possible binding ligand. Click the name to view its information in the BioLiP database.
(d) **Rep** is a single complex structure with the most representative ligand in the cluster, i.e., the one listed in the **Lig Name** column. **Mult** is the complex structures with all potential binding ligands in the cluster.

Figure A32. Predicted ligand binding site in hemoglobin form scaffold 485. One ligand site be found in the protein, where heme had the highest confidence score. The figure also shows which residues that is involved in ligand binding.

7.5.3 3D structure of scaffold 313051

Based on the online tools Phyre[2] (Kelley *et al*, 2015) and I-Tasser (Roy *et al*, 2010), the three-dimensional structure of the oat hemoglobin could be predicted. These model use the amino acid sequence of the hemoglobin and compare it with known proteins with similar characteristics. The predicted model of scaffold 313051 from Phyre[2] can be seen in **Figure A33**. Here, the characteristic α- helix motif consisting of 7 helices in close proximity can be found. The confidence key (**Figure A33b**) indicates that the majority of the amino acids can be placed with high confidence, except for the N- and C-terminal. **Figure A33c** displays the predicted secondary structure, where every amino acid is placed in helices, strands or coils, and the predicted accessibility of the different amino acids

(a)



(b)

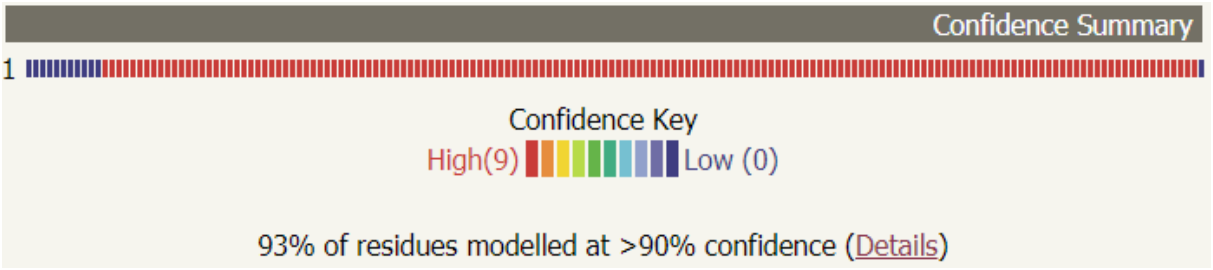


Figure A33. Predicted model of the oat hemoglobin (a) and the confidence summary of this prediction (b) from scaffold 313051 (Kelley *et al*, 2015). In addition, the predicted secondary structure and accessibility of the amino acids (c) was generated by I-Tasser (Roy *et al*, 2010). These are not presented here.

I-Tasser also provided top candidates and suggested ligand binding sites. The top candidates represents the best matching structural analogs of which the prediction has been based on (**Figure A34**). The three top candidates were non-symbiotic hemoglobin from rice, Trema tomentosa and corn, respectively. All the remaining candidates were also hemoglobin or hemoglobin-like proteins from different plants.

Top 10 Identified stuctural analogs in PDB

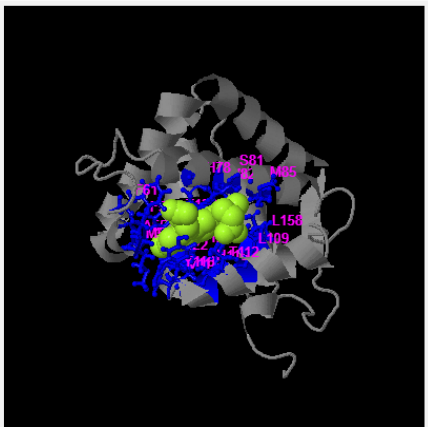
Click to view	Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov	Alignment
○	1	2gnwA	0.963	0.44	0.782	0.971	Download
○	2	3qqqA	0.836	1.48	0.780	0.882	Download
○	3	2r50D	0.836	0.72	0.862	0.853	Download
○	4	1gdl_	0.805	2.01	0.430	0.882	Download
○	5	2oifA	0.777	2.23	0.961	0.894	Download
○	6	1binB	0.765	1.82	0.408	0.835	Download
○	7	1oj6D	0.734	2.23	0.199	0.859	Download
○	8	102m_	0.721	2.47	0.183	0.835	Download
○	9	1gvha1	0.715	2.18	0.158	0.818	Download

(a) Query structure is shown in cartoon, while the structural analog is displayed using backbone trace.
(b) Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library.
(c) RMSD[a] is the RMSD between residues that are structurally aligned by TM-align.
(d) IDEN[a] is the percentage sequence identity in the structurally aligned region.
(e) Cov represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

Figure A34. Identified top candidates of structural analogs to oat hemoglobin from scaffold 313051. The three top candidates were non-symbiotic hemoglobin from rice, Trema tomentosa and corn, respectively. All candidates were hemoglobin or hemoglobin-like proteins from the plant kingdom.

Suggested ligand binding sites were also found in the I-Tasser prediction. This result can be seen in **Figure A35,** where the heme group got the highest condifence score of every ligand. The figure also displays where the heme group is located and which residues that are predicted to be invloved in the ligand binding site. Unlike scaffold 1780, residue 160 is not included here and neither is residue 53, as in 485.

Ligand binding sites



Click to view	Rank	C-score	Cluster size	PDB Hit	Lig Name	Download Complex	Ligand Binding Site Residues
◉	1	0.97	1047	4ns2A	HEM	Rep, Mult	55,58,59,61,78,81,82,85,109,112,113,116,118,122,123,126,158
○	2	0.03	44	107mA	NBN	Rep, Mult	45,48,59,78,82,126
○	3	0.02	31	4g1bA	ECN	Rep, Mult	40,41,44,45,59,78,79,82,130,155,158
○	4	0.01	13	3a0gA	OXY	Rep, Mult	78,82
○	5	0.01	20	3g4wA	8CL	Rep, Mult	33,40,44,80,83,130,133,134,151

Download the residue-specific ligand binding probability, which is estimated by SVM.
Download the all possible binding ligands and detailed prediction summary.
Download the templates clustering results.
(a) **C-score** is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction.
(b) **Cluster size** is the total number of templates in a cluster.
(c) **Lig Name** is name of possible binding ligand. Click the name to view its information in the BioLiP database.
(d) **Rep** is a single complex structure with the most representative ligand in the cluster, i.e., the one listed in the **Lig Name** column. **Mult** is the complex structures with all potential binding ligands in the cluster.

Figure A35. Predicted ligand binding site in hemoglobin form scaffold 313051. One ligand site be found in the protein, where heme had the highest confidence score. The figure also shows which residues that is involved in ligand binding.

7.5.4 3D structure of truncated scaffold 233

Based on the online tools Phyre[2] (Kelley *et al*, 2015) and I-Tasser (Roy *et al*, 2010), the three-dimensional structure of the oat hemoglobin could be predicted. These model use the amino acid sequence of the hemoglobin and compare it with known proteins with similar characteristics. The predicted model of scaffold 233 (truncated hemoglobin) from Phyre[2] can be seen in **Figure A36**. Here, the characteristic α- helix motif consisting of 7 helices in close proximity can be found. The confidence key (**Figure A36b**) indicates that the majority of the amino acids can be placed with high confidence, except for the N- and C-terminal. **Figure A36c** displays the predicted secondary structure, where every amino acid is placed in helices, strands or coils, and the predicted accessibility of the different amino acids

(a)

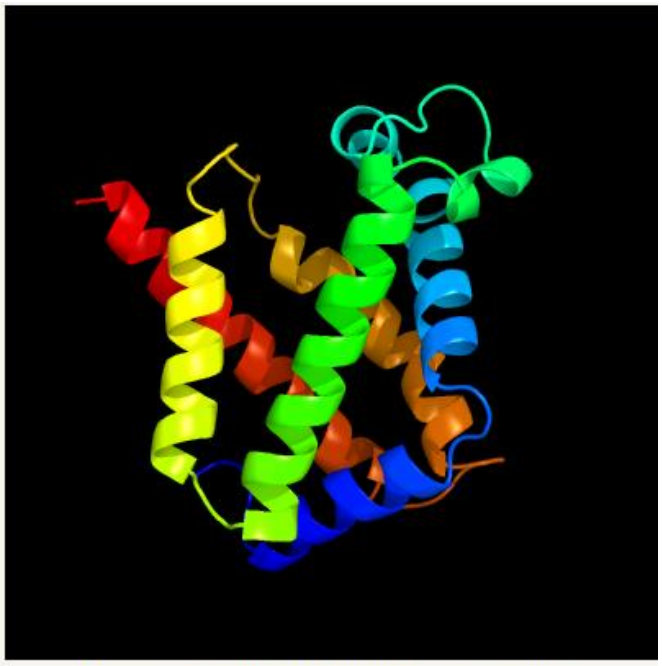


Image coloured by rainbow N → C terminus

Model dimensions (Å): **X**:44.938 **Y**:36.132 **Z**:43.645

(b)

Top template information

PDB header:transport protein
Chain: A: **PDB Molecule:**protein glb-12;
PDBTitle: globin-like protein glb-12 from c.elegans

Confidence and coverage

Confidence:	100.0%	Coverage:	99%

155 residues (99% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

(c)

Predicted Secondary Structure

```
                  20                  40                  60                  80                  100                 120                 140
                   |                   |                   |                   |                   |                   |                   |
Sequence    MGFSEAQEELVLRSWKAMKPDSESIALNFFLRIFEIAPAAKPMFPFLREDAPLESHPKLRAHAVTVFVMACESATQLRKTGDVKVREATLRRLGATHVKAGVADAHFEVVKTALLDTIEGAVTEMWTPEMKGAWEEAYDQLAAAIKEEMKIAASASA
Prediction  CCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCHHHHHCCCCCCCCHHHCCHHHHHHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHHHCHHCCHHHHHHHHHHHHHHHHHHHHHCCCC
Conf.Score  988999999999999999855899999999999989889976276567752212999999999999999998602558567899999999999885998999999999999999816008999999999999999999988653589
            H:Helix; S:Strand; C:Coil
```

Predicted Solvent Accessibility

```
                  20                  40                  60                  80                  100                 120                 140
                   |                   |                   |                   |                   |                   |                   |
Sequence    MGFSEAQEELVLRSWKAMKPDSESIALNFFLRIFEIAPAAKPMFPFLREDAPLESHPKLRAHAVTVFVMACESATQLRKTGDVKVREATLRRLGATHVKAGVADAHFEVVKTALLDTIEGAVTEMWTPEMKGAWEEAYDQLAAAIKEEMKIAASASA
Prediction  771376225103500530373145003300330065234036214424546435524403410230031023104305727416303410440043036360446205302400130037204740253035004300310041035104535747 8
            Values range from 0 (buried residue) to 9 (highly exposed residue)
```

Figure A36. Predicted model of the oat hemoglobin (a) and the confidence summary of this prediction (b) from scaffold 233 (Kelley *et al*, 2015). In addition, the predicted secondary structure and accessibility of the amino acids (c) was generated by I-Tasser (Roy *et al*, 2010). Here, the individual amino acids are depicted as being a part of helices, strands or coil, as well as a confidence score ranging from 0-9 as in (b).

I-Tasser also provided top candidates and suggested ligand binding sites. The top candidates represents the best matching structural analogs of which the prediction has been based on (**Figure A37**). The three top candidates were non-symbiotic hemoglobin from rice, Trema tomentosa and Arabidopsis, respectively. All the remaining candidates were also hemoglobin or hemoglobin-like proteins from different plants

Top 10 Identified stuctural analogs in PDB

Click to view	Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov	Alignment
○	1	1d8uA	0.931	1.40	0.654	0.968	Download
○	2	3qqqA	0.928	0.88	0.613	0.955	Download
○	3	3zhwA	0.921	1.26	0.589	0.962	Download
○	4	1gdL_	0.891	1.68	0.454	0.968	Download
○	5	2r50D	0.889	0.95	0.611	0.911	Download
○	6	2oifA	0.835	2.09	0.636	0.955	Download
○	7	1binB	0.830	1.59	0.430	0.904	Download
○	8	101m_	0.819	2.10	0.151	0.930	Download
○	9	1lhsA	0.808	2.03	0.138	0.924	Download
○	10	3ba2A	0.807	2.09	0.132	0.917	Download

(a) Query structure is shown in cartoon, while the structural analog is displayed using backbone trace.
(b) Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library.
(c) RMSD[a] is the RMSD between residues that are structurally aligned by TM-align.
(d) IDEN[a] is the percentage sequence identity in the structurally aligned region.
(e) Cov represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

Figure A37. Identified top candidates of structural analogs to oat hemoglobin from scaffold 233. The three top candidates were non-symbiotic hemoglobin from rice, Trema tomentosa and corn, respectively. All candidates were hemoglobin or hemoglobin-like proteins from the plant kingdom.

Suggested ligand binding sites were also found in the I-Tasser prediction. This result can be seen in **Figure A38,** where the heme group got the highest condifence score of every ligand. The figure also displays where the heme group is located and which residues that are predicted to be invloved in the ligand binding site. Since the truncated hemoglobin differes in

structure from conventional hemoglobin, the suggested binding of heme is different from the other scaffolds.

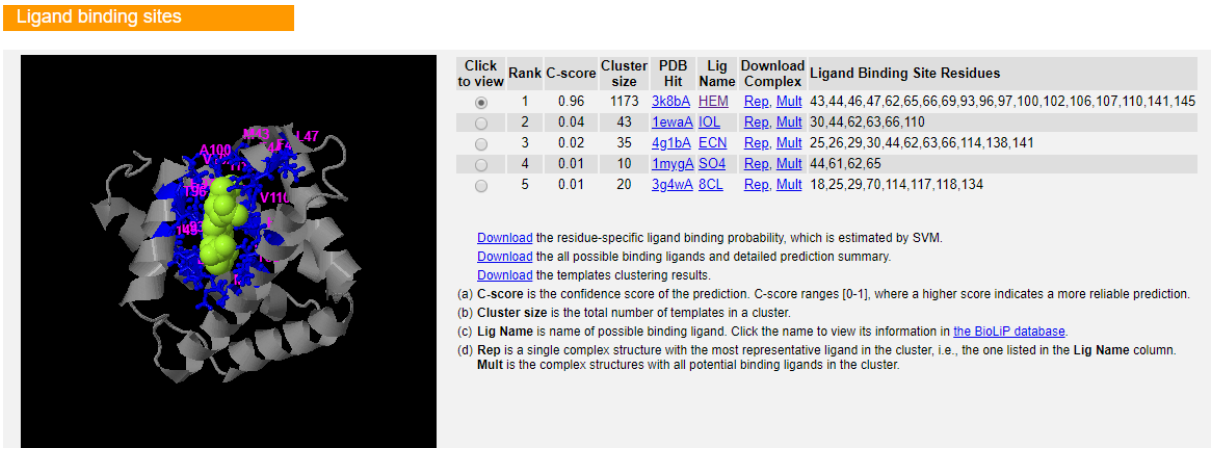


Figure A38. Predicted ligand binding site in hemoglobin form scaffold 233. One ligand site can be found in the protein, where heme had the highest confidence score. The figure also shows which residues that is involved in ligand binding.

7.5.5 3D structure of truncated scaffold 445

Based on the online tools Phyre[2] (Kelley *et al*, 2015) and I-Tasser (Roy *et al*, 2010), the three-dimensional structure of the oat hemoglobin could be predicted. These model use the amino acid sequence of the hemoglobin and compare it with known proteins with similar characteristics. The predicted model of scaffold 445 (truncated hemoglobin) from Phyre[2] can be seen in **Figure A39**. Here, the characteristic α- helix motif consisting of 7 helices in close proximity can be found. The confidence key (**Figure A39b**) indicates that the majority of the amino acids can be placed with high confidence, except for the N- and C-terminal. **Figure A39c** displays the predicted secondary structure, where every amino acid is placed in helices, strands or coils, and the predicted accessibility of the different amino acids

(a)

Image coloured by rainbow N → C terminus

Model dimensions (Å): **X**:45.169 **Y**:35.373 **Z**:47.012

(b)

Top template information

PDB header:transport protein
Chain: A: **PDB Molecule:**protein glb-12;
PDBTitle: globin-like protein glb-12 from c.elegans

Confidence and coverage

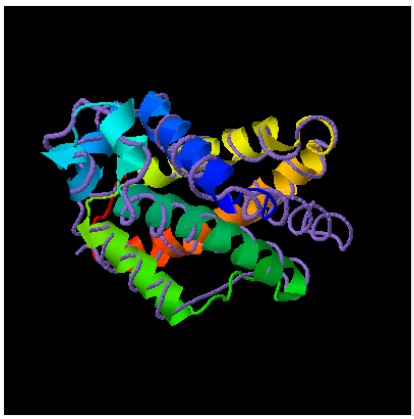
Confidence:	**100.0%**	Coverage:	**97%**

139 residues (97% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

(c)

Figure A39. Predicted model of the oat hemoglobin (a) and the confidence summary of this prediction (b) from scaffold 445 (Kelley *et al*, 2015). In addition, the predicted secondary structure and accessibility of the amino acids (c) was generated by I-Tasser (Roy *et al*, 2010). These are not presented here.

I-Tasser also provided top candidates and suggested ligand binding sites. The top candidates represents the best matching structural analogs of which the prediction has been based on (**Figure A40**). The three top candidates were non-symbiotic hemoglobin from leghemoglobin VI, Trema tomentosa and Arabidopsis, respectively. All the remaining candidates were also hemoglobin or hemoglobin-like proteins from different plants.



Top 10 Identified stuctural analogs in PDB

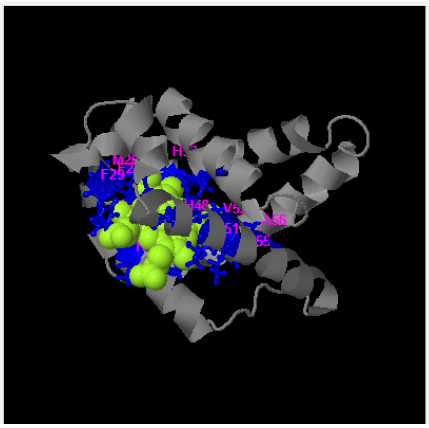
Click to view	Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov	Alignment
○	1	2lh7A	0.879	1.35	0.471	0.951	Download
○	2	3qqqA	0.871	1.33	0.634	0.937	Download
○	3	3zhwA	0.861	1.68	0.634	0.937	Download
○	4	1d8uB	0.856	1.64	0.642	0.930	Download
○	5	2r50D	0.837	1.31	0.633	0.888	Download
○	6	2oifA	0.834	1.95	0.654	0.944	Download
○	7	1oj6B	0.830	1.81	0.222	0.937	Download
○	8	1lhsA	0.826	1.87	0.169	0.951	Download
○	9	1ajgA	0.817	1.91	0.207	0.944	Download
○	10	2vlxA	0.810	2.00	0.193	0.944	Download

(a) Query structure is shown in cartoon, while the structural analog is displayed using backbone trace.
(b) Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library.
(c) RMSD[a] is the RMSD between residues that are structurally aligned by TM-align.
(d) IDEN[a] is the percentage sequence identity in the structurally aligned region.
(e) Cov represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

Figure A40. Identified top candidates of structural analogs to oat hemoglobin from scaffold 445. The three top candidates were non-symbiotic hemoglobin from rice, Trema tomentosa and corn, respectively. All candidates were hemoglobin or hemoglobin-like proteins from the plant kingdom.

Suggested ligand binding sites were also found in the I-Tasser prediction. This result can be seen in **Figure A41,** where the heme group got the highest condifence score of every ligand. The figure also displays where the heme group is located and which residues that are predicted to be invloved in the ligand binding site. Since the truncated hemoglobin differes in structure from conventional hemoglobin, the suggested binding of heme is different from the other scaffolds.

Ligand binding sites



Click to view	Rank	C-score	Cluster size	PDB Hit	Lig Name	Download Complex	Ligand Binding Site Residues
◉	1	0.94	1351	1fsxA	HEM	Rep, Mult	26,27,29,48,51,52,55,56,79,82,83,86,88,92,93,96,127,131
○	2	0.08	96	2d2mA	OXY	Rep, Mult	13,27,48,52,96
○	3	0.04	59	2w72A	XE	Rep, Mult	9,12,13,52,56,96,100
○	4	0.01	9	1j52A	XE	Rep, Mult	93,96,97,124,127,128
○	5	0.01	14	3u3eA	IPH	Rep, Mult	79,80,83,93,127,131,135

Download the residue-specific ligand binding probability, which is estimated by SVM.
Download the all possible binding ligands and detailed prediction summary.
Download the templates clustering results.

(a) **C-score** is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction.
(b) **Cluster size** is the total number of templates in a cluster.
(c) **Lig Name** is name of possible binding ligand. Click the name to view its information in the BioLiP database.
(d) **Rep** is a single complex structure with the most representative ligand in the cluster, i.e., the one listed in the **Lig Name** column. **Mult** is the complex structures with all potential binding ligands in the cluster.

Figure A41. Predicted ligand binding site in hemoglobin form scaffold 445. One ligand site can be found in the protein, where heme had the highest confidence score. The figure also shows which residues that is involved in ligand binding.