Finding a stabilized sequence for reconstructed Icosahedra Lumazine Synthase by random mutagenesis and Fluorescence-Activated Cell Sorting

Master Thesis

Caroline Olsson

LUND
UNIVERSITY

Center for Molecular Protein Science
Lund University
Sweden
June 2019

Finding a stabilized sequence for reconstructed Icosahedra Lumazine Synthase by random mutagenesis and Fluorescence-Activated Cell Sorting

Master Thesis

Caroline Olsson

Abstract

Lumazine Synthase forms pentamers, C_5 in some organism and icodahedra, I_h in other. The aim of this project is to investigate which evolutionary steps could have led the pentamers to assembly into icosahedral capsids. Therefore prevoius work has been performed to reconstruct ancestral sequences for the pentameric and icosahedral forms of Lumazine Synthase. The icosahedral sequence had been expressed and shown to give insoluble expression. The work in this thesis has therefore been done to, firstly, try to purify and refold the protein from inclusion bodies and secondly, try to improve the sequence to give soluble expression.

Expression of the protein resulted in the protein in inclusion bodies. The protein could be purified from the inclusion bodies by guanidine·HCl extraction and size exclusion chromatography. However it was not possible to refold the protein into native structure by dialysis to remove the denturant, guanidine·HCl.

To improve the sequence to give stable expression, a stability assay system was used in combination with random mutagenesis of the sequence. The stability assay system consist of two reporter proteins, red fluorescent *tag*RFP and green fluorescent *sf*-GFP. Expression of the protein is monitored by red fluorescence by *tag*RFP fused to the protein and the stability of the expressed protein is monitored by *sf*-GFP expressed under the control of stress activated DnaK promoter. The random mutagenesis was performed by error prone PCR to produce fragments spanning the sequence of Lumazine Synthase. The created fragments were then used as megaprimers to amplify the whole plasmid containing the protein with *tag*RFP and the stability assay system.

Contents

1 Populärvetenskaplig sammanfattning

Proteiner är de små byggstenar som bygger upp nästan allt omkring oss. Proteiner bygger upp våra kroppar och utför olika funktioner som är livsviktiga för alla djur, växter och andra organismer. Byggstenarna kan ha olika former och utseende och de har olika funktioner. Det finns proteiner som hjälper till att bryta ner maten vi äter, proteiner som bygger upp våra muskler och massor av andra proteiner som hjälper till med olika kemiska reaktioner. Proteinerna består av långa kedjor av aminosyror som sitter ihop i en viss ordning, en viss sekvens. Genom evolutionens gång har proteinernas sekvenser förändrats och utvecklats till de varianter som finns i olika organismer idag. Många proteiner är lite annorlunda beroende på vilken organism man undersöker.

Det protein som detta arbete undersöker heter Lumazine Synthase och det bildar tre olika varianter i olika organsimer. I svampar och arkéer bildar 5 proteinenheter, som sitter i en cirkel tillsammans, så kallade pentamerer av proteinet. I de flesta bakterier bildar 12 stycken pentamerer tillsammans en sfärisk struktur. Ungefär som rutorna på en fotboll sitter de 12 pentamererna sida vid sida och bildar en rund kapsel. I vissa organismer sitter två pentamerer ihop som en sandwich. Det är intressant att det bara är i vissa organismer som enheterna bildar kapslar medan de i andra organismer bara bildar pentamerer eller dubbla pentamerer. Därför vill man undersöka hur det kan ha gått till när evolutionen har gjort så att det finns dessa varianter. Det måste ha skett några förändringar i strukturen som gjorde att pentamererna kunde sitta ihop i varandra och bilda kapslar. Därför har man räknat ut vilken sekvens den första Lumazine Synthase som bildade en kapsel kan ha haft.

Det som ställde till med ett problem var att sekvensen man räknat fram inte lyckades bilda ett stabilt protein. Därför ville man försöka ändra på sekvensen litegrann för att få ett lite mer stabilt protein. Detta kan man göra genom att driva en slags evolution på liten skala genom att göra slumpmässiga förändringar i aminosyrasekvensen och sen välja ut varianter som fungerar bättre. Tanken är då att någon av de slumpmässiga förändringarna ska ge en variant som bildar ett mer stabilt protein vilket man sen kan använda för att undersöka vidare.

2 Introduction

A lot of research has been conducted to understand how evolution resulted in the variations of protein sequences that we find in different organisms today. The variants that exist today are the evolutionary end products and the intermediate states of how they emerged are of interest to study[1]. The information gained can be used for in depth understanding of how certain protein structures have emerged, which could also be used for protein design. This can for example be interesting to study when it comes to structural features such as oligomerization. One example of such oligomerisations is the formation of capsids, out of which homomeric icosahedra capsids are among the most simple to study because of the self assembly into capsids from only one sequence[2]. The aim of the project is to study the evolutional intermediate states between pentamer and icosahedra capsids in the model system of Lumazine Synthase.

2.1 Lumazine Synthase

In order to study evolutionary intermediates of icosahedra capsids a model system was chosen that could illustrate the evolutionary process from lower to higher homooligomeric symetries. The model system that was chosen is the protein Lumazine Synthase, LS. LS exists both in a pentameric, C_5, form and as an icosahedra. The pentamer is found in fungi and archaea while the icosahedra is found in bacteria. Examples of both forms have previously been structurally determined by x-ray crystallography, see Figure 1. The LS icosahedra consists of 60 identical subunits, which is the smallest of the icosahedral symmetries. The monomer units that form pentamers and icosahedra have very small structural differences and the C-terminal helix has been identified as important for the oligomerization state[3]. This project aims to investigate which amino acid differences between the sequences for pentameric and icosahedra Lumazine Synthase are important for the oligomerisation state. The aim is to find which paths of mutations of amino acid sequence that can lead from pentameric to icosahedral symmetry.

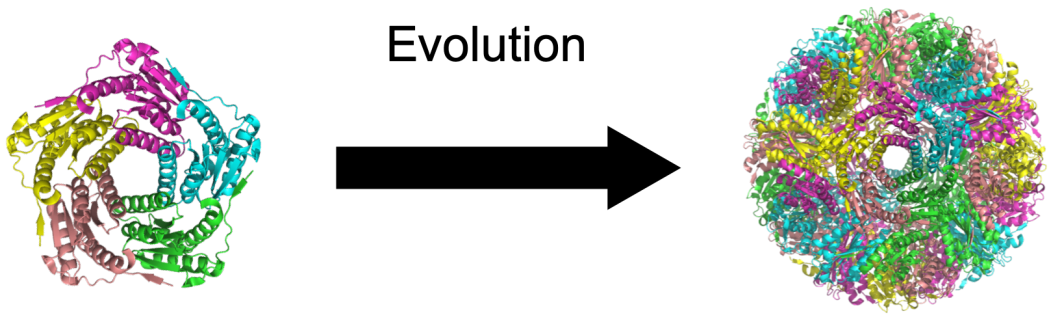


Figure 1: Pentamer Lumazine Synthase Saccharomyces Cerevisiae, PDB ID: 1EJB [4]. Icosahedra Lumazine Synthase Aquifex Aeolicus, PDB ID: 1HQK [5].

Lumazine Synthase is an enzyme, EC 2.5.1.78, that is part of the riboflavin

biosynthesis in bacteria, fungi, archaea and eubacteria. Lumazine Synthase is a transferase that catalyses the formation of 6,7-dimethyl-8-ribityllumazine, see reaction Figure 2 [6]. Two of these molecules are then further used by the transferase Riboflavin Synthase to form one molecule of riboflavin (Vitamin B2), see reaction Figure 3 [6].

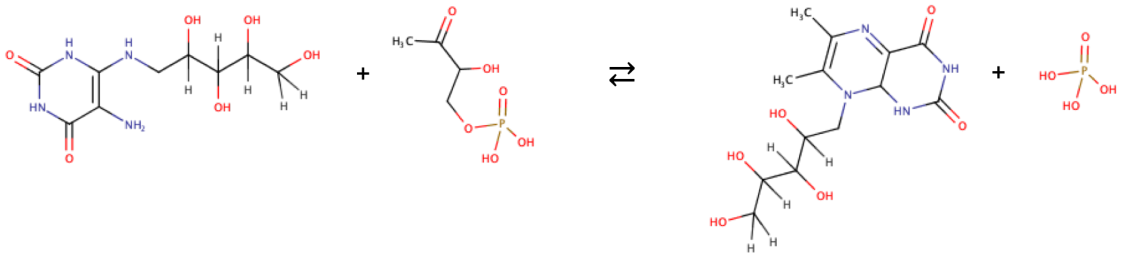


Figure 2: Reaction catalysed by Lumaine Synthase. One molecule of 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedion and one molecule of 3,4-dihydroxy-2-butanone 4-phosphate is combined to give one molecule of 6,7-dimethyl-8-ribityllumazine and a phosphate[6].

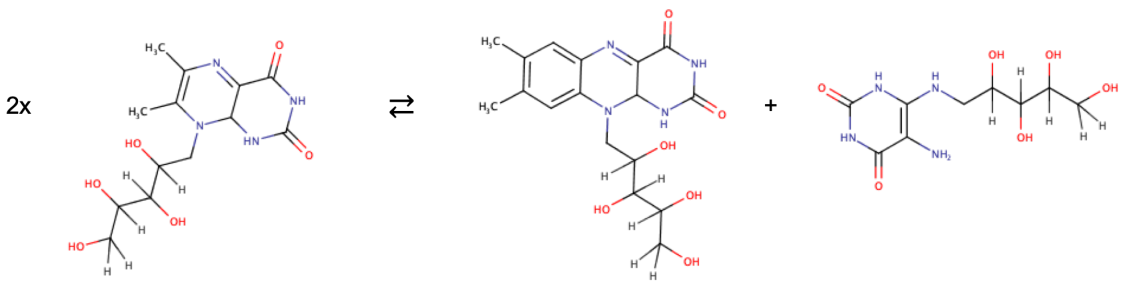


Figure 3: Reaction catalysed by Riboflavin Synthase. Two molecules of 6,7-dimethyl-8-ribityllumazine are combined to give one molecule of riboflavin and one molecule of 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedion that can go back to be used by Lumazine Synthase[6].

Riboflavin synthesis is essential for bacteria and plants since riboflavin, or vitamin B_2 is further used to synthezise flavin mononucleotide, FMN and flavin adenin dinucleotide, FAD used for energy production in the electron transport chain and other reactions. Animals do not have LS and RS and therefore has to take up vitamin B_2 in their diet.

2.2 Ancient Sequence Reconstruction

The variants of Lumazine Synthase that are found today in different organisms can be arranged in a phylogenetic tree based on their evolutional relationships, see Figure 4. The pentameric forms in fungi and archaea are more closely related while the icosahedra forms in bacteria are more related. Between them are some differences that let to the icosahedra form develop from the pentameric form. Some mutations

caused the structure to change in such way that it made assembly into capsids possible.

Ancient Sequence Reconstruction is a method to find a sequence that could be a last common ancestor of the variants that are found today [7]. Last common ancestor for the pentamer and for the icosahedra is marked with * in Figure 4. The reconstructed sequence is not a true sequence that existed during evolution but based on probability of which amino acid is present at each location. In each location the probability of certain amino acid residues are found. For example it can be 60% probability for Alanine and 40% for Serine at one position, then it is guessed that Alanine is the correct residue but it could also be Serine. This means that there is uncertainty in the sequence that is reconstructed. It is not a true ancestor sequence but a guess based on probability.

Previous work has been done to find such sequences for both pentameric and icosahedra Lumasine Synthase. In this project the ancestral sequence of icosahedra Lumazine Synthase is used.

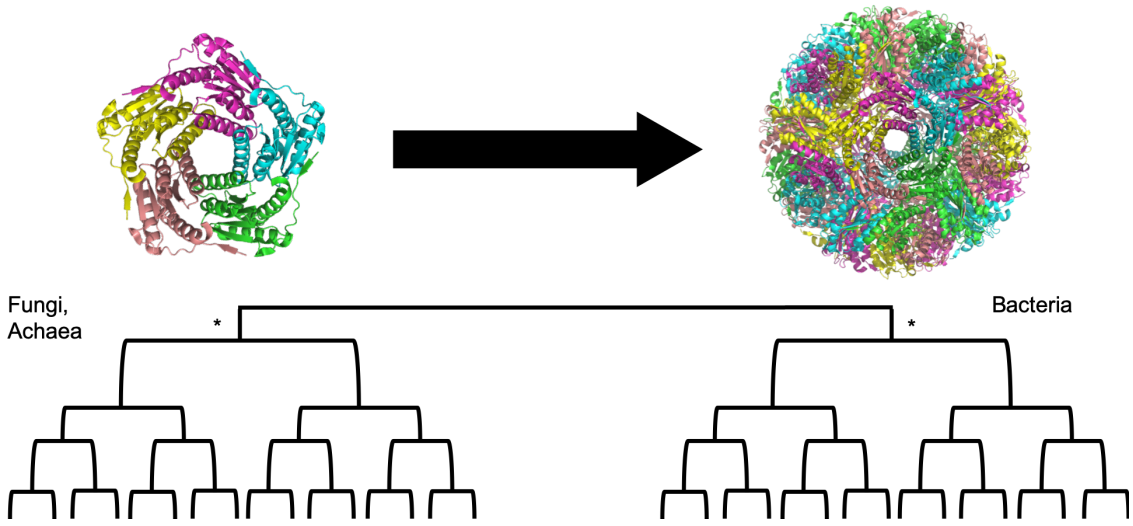


Figure 4: Schematic figure of the presence of pentameric and icosahedra Lumazine Synthase. * marks the two sequences that have been reconstructed by Ancient sequence reconstruction.

2.3 Prevoius work

Icosahedra Lumazine Synthase, with the reconstructed ancestral sequence, have previously been expressed in E. coli and attempted to purify using N-terminal histag and Ni^{2+} affinity chromatography. This has not been successful why the work of this thesis was performed to, firstly, try to purify the protein from inclusion bodies, and secondly, find a more stably expressing variant of the sequence. Since the ancient sequence is based on probabilities of which amino acids occupy which position, it is possible to find another variant of the sequence that can be more stably expressed in E. coli and stil represent the last common ancestor of icosahedra Lumazine Synthase.

2.4 Specific aim of this work

In the first part of this work Lumazine Synthase with N-terminal histag, LShis, was expressed and experiments were performed to purify LShis from inclusion bodies by guanidine·HCl extraction followed by refolding. Protein was expressed via induction in BL21star cells, cells lysed and protein was determined to be in the insoluble fraction after centrifugation. The protein was purified from the insoluble fraction by guanidine·HCl extraction and size exclusion chromatography. Protein refolding was tried using dialysis into buffer without guanidine·HCl, however this was not successfull.

In the second part of this work the aim was to find a more stable variant of the ancestral sequence for icosahedral Lumazine Synthase. This was done using directed evolution and a stability assay based on the heat shock chaperone stress response of Escherichia coli. The assay consists of two fluorescent reporter proteins, *tag*RFP and *sf*-GFP. Red fluorescent protein tag, *tag*RFP, as an expression reporter fused to the protein via a linker and *Superfolder* Green fluorescent protein, *sf*-GFP, as a reporter for the heat shock stress response of the cell. The red fluorescence is a signal of how much of the protein is expressed while the green fluorescence is a signal of how stressed the cells are. This system can be used to select cells that express more stable protein by directed evolution. Combination of the protein stability assay and introduction of random mutations to the sequence gives the basis for directed evolution to find more stable variants of the sequence [8]. Introduction of mutations was done by random mutagenesis of the sequence by production of megaprimers by Error Prone PCR and cloning into the plasmid by Megaprimer Whole Plasmid PCR[9]. A more stable variant of the sequence is then to be found using Fluorescence-Activated Cell Sorting flow cytometry.

3 Introduction to the methods used in this work

3.1 Lumazine Synthase with N-terminal Histag

Icosahedra Lumazine Synthase, using the reconstructed ancient sequence, with 6 x Histidine tag (LShis) at the N-terminal in pET 28c(+) vector had been previously constructed. It was determined that purification of this construct was not possible due to low expression of soluble protein, why a new construct was designed that could be used to find a more stable variant of the sequence.

3.2 Refolding

Protein that is in a solution of high concentration denaturant is in a non-native conformation [10]. In order for the protein to gain its native conformation the conditions has to be changes to non-denaturating. This can for example be done by dialyzing into a buffer that does not contain the denaturant. As the denaturant is diluted, the protein can start to fold and possibly reach its native structure. However, if the denaturant concentration decreases very rapidly protein might aggregate faster than it folds [11].

3.3 Lumazine Synthase with N-terminal *tag*RFP

In order to find a more solubly expressing variant of the Lumazine Synthase sequence, a new construct was created with the same ancestral sequence but fused at the N-terminal with *tag*RFP (Red Fluorescence protein tag), see Appendix I. The construct was cloned into a plasmid containing the gene for *sf*-GFP under the control of the stress activated DnaK promoter, see 3.4. Plasmid map of the whole construction is found in Appendix II

3.4 Escherichia coli Heat Shock response

One of the components of the directed evolution used to find a more stably expressed variant of the sequence is the protein stability selection system. The protein stability assay is based on the heat shock stress response of Escherichia coli. E. coli, among other organism have systems for coping with changing external and internal conditions. Factors as heat shock and protein aggregation in the cell trigger the Heat Shock response including induction of more than 20 Heat Shock Proteins (HSPs). HSPs, such as DnaK folding chaperon, DnaJ co-chaperone, GrpE regulator, GroEL and GroES are up to 15-fold increased during stress response compared to basal levels [12][13]. In this project the E. coli stress response system is used in that the promoter controlling DnaK production is used to express a reporter protein, *sf*-GFP.

DnaK chaperone production in the cell is regulated by σ^{32} transcription factor which bind to RNA polymerase and induce transcription at the DnaK promoter. Basal transcription of the rpoH gene, which codes for σ^{32}, is induced by σ^{70} at promoters P1, P4 and P5. The rpoH gene is translated to σ^{32} transcription factor, which promotes production of DnaK folding chaperone by induction of the DnaK promoter [12].The C-terminal domain of DnaK binds to exposed hydrophobic residues on misfolded protein together with DnaJ co-chaperone and, via ATP hydrolysis promotes folding of the protein [13][14]. The regulator GrpE increases the release of ADP from DnaK which helps to drive the hydrolysis of ATP and the protein folding [15][16]. DnaK can also bind to σ^{32} which targets it for FtsH protease degradation, resulting in a half-life of less than 1 minute and a steady-state concentration of 10-30 copies σ^{32} per cell in the absence of misfolded protein [12].

In response to misfolded protein in the cell, rpoH transcription is further induced via σ^{E} at the P3 promoter, leading to higher production of σ^{32} and hence DnaK [12]. When DnaK binds to misfolded protein, it no longer binds to and promote degradation of σ^{32} by FtsH and the half-life of σ^{32} is increased by a factor 8 [12]. This means that the presence of misfolded or aggregated protein in the cell increases the presence and availability of σ^{32} which also increases the induction of the DnaK promoter. When the amount of DnaK in the cell exceeds the need by misfolded protein, it again binds to σ^{32}, which prevents it from further inducing DnaK expression and increases its degradation by FtsH. This creates an off-switch for the stress response caused by misfolded protein.

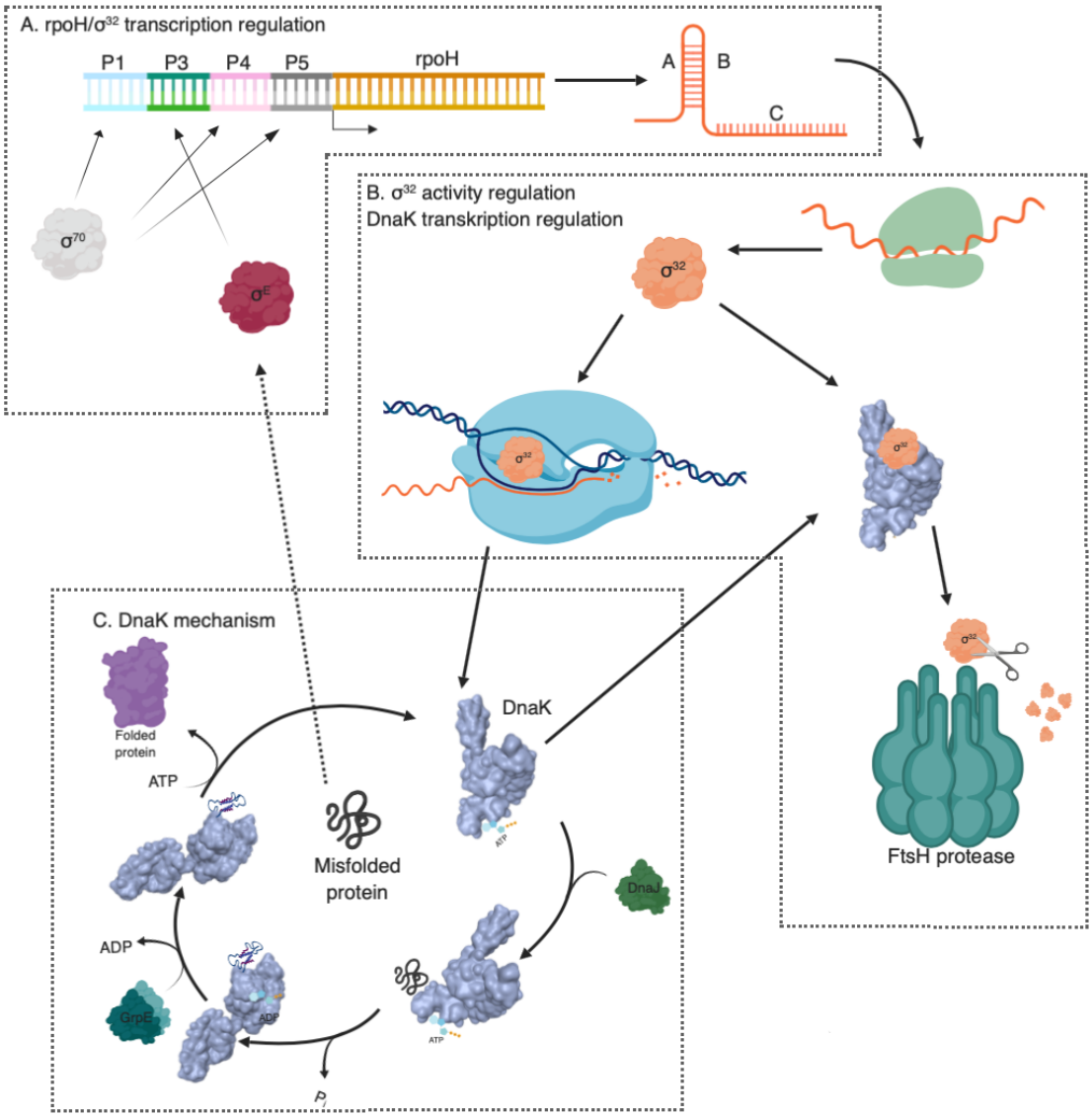


Figure 5: A. Basal transcription of the rpoH gene is induced by σ^{70} at promoters P1, P4 and P5. B. The rpoH gene is translated to σ^{32} which can bind to RNA polymerase and induce production of DnaK folding chaperone. C. DnaK binds to misfolded protein together with DnaJ co-chaperone and via ATP hydrolysis promoting folding of the protein. GrpE removes the bound ADP from DnaK which drives the hydrolysis of ATP [15]. In response to misfolded protein in the cell, rpoH transcription is further induced via σ^E at the P3 promoter, leading to higher production of σ^{32} and hence DnaK. DnaK can also bind to σ^{32} which targets it for FtsH protease degradation. Figure created with biorender.com

3.4.1 Stress induced DnaK promoter

In the protein stability assay used in this work for directed evolution of the protein sequence, see 2.4, the cellular stress response to misfolded protein is used to monitor

protein stability by coupling of a reporter protein, *superfolder* Green Fluorescent Protein, *sf*-GFP, to the DnaK promoter. By this it is possible to monitor the amount of stress that the cell is experiencing due to misfolded or aggregated protein by monitoring the green fluorescence signal. A high amount of misfolded protein in the cell give rise to induced expression of *sf*-GFP and an increased green fluorescence signal.

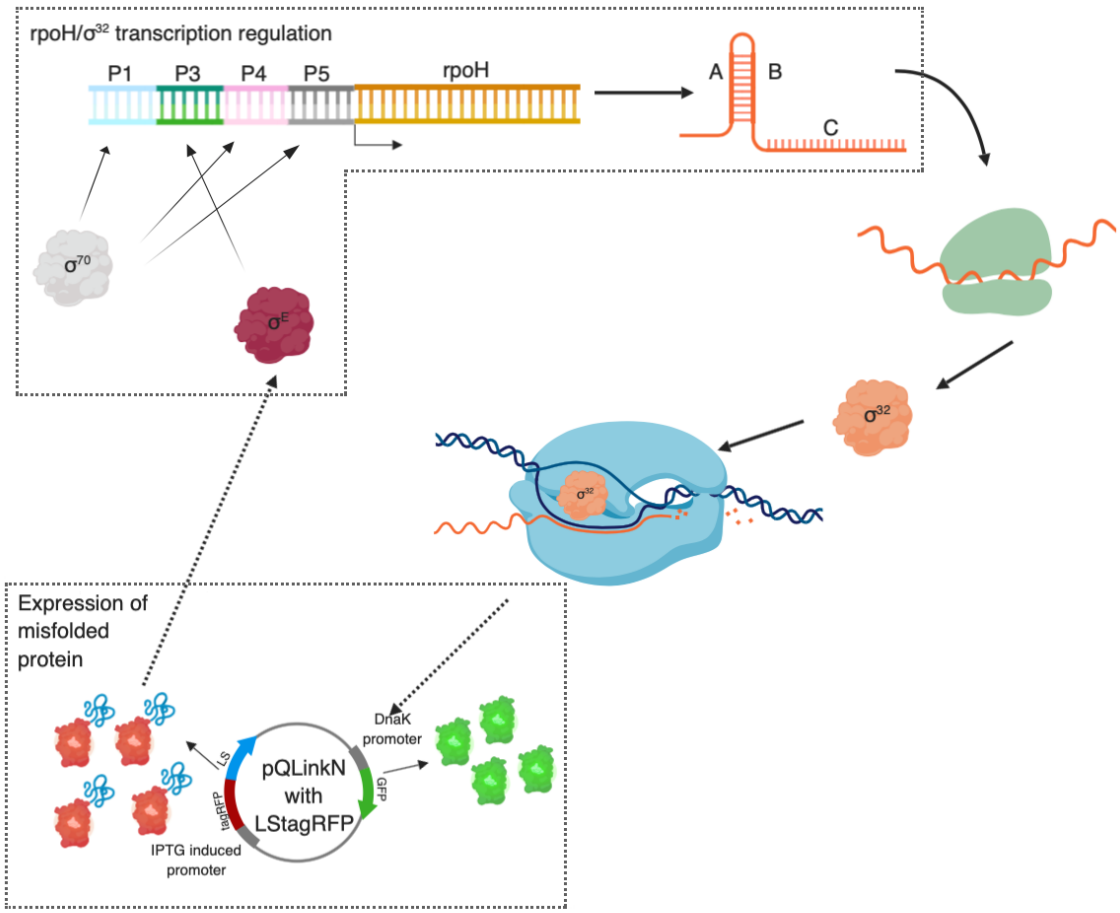


Figure 6: Expression of misfolded protein leads to induction of rpoH transcription via σ^E at promoter P3. The increased σ^{32} level increase the induction of the DnaK promoter and expression of *sf*-GFP and increased green fluorescence signal. Figure created with biorender.com

3.5 Random mutagenesis of protein sequence

In order to use directed evolution by the protein stability assay to find a more stably expressed variant of the sequence, mutations are introduced to the sequence by random mutagenesis. Random mutagenesis is performed on a sequence to produce a library containing many variants with different mutations in the sequence. This library can then be used to select variants that show higher stability when expressed in bacterial cells. In this work random mutagenesis was done by Error prone PCR combined with Megaprimer whole plasmid PCR to introduce mutations

to the protein sequence in the plasmid containing the stability assay system.

3.5.1 Library design

For this work, single amino acid mutations at all amino acid positions are desired. In order to cover all possible single amino acid mutations in a library, 20 times the length of the mutated fragment variants are needed. In the case for this work, the mutated fragment is 565 bp or 188 amino acid residues, which means that around 4000 variants are needed to cover only all of the single mutations. However, it is not possible to create only single mutations using error prone PCR. Since the number of mutations are Poisson distributed, also no mutations and multiple mutations will be created. In order to create a library with the desired number of variants it is crucial that a high enough transformation efficiency is achieved when transforming the plasmid library into the host bacteria. At least as many transformants as desired variants in the library has to be obtained in order for the library to cover all the desired mutations.

3.5.2 Error Prone PCR

Error Prone PCR was performed using the lower fidelity of Taq DNA polmerase to create DNA fragments that contain the sequence of the protein with incorporated mutations. The mutation rate of the Taq DNA Polymerase can be increased by increasing concentration of $MnCl_2$ and the number of PCR cycles. In order to create a library of single amino acid mutations, 1-3 nucleotide changes in each variant are needed. This means that an error rate of 1.8 - 5.3 per kbp is desired. However, the error rate of the Taq DNA polymerase is also dependent on the specific sequence which makes it difficult to predict the actual error rate that will be obtained under certain conditions [10]. The error prone PCR, epPCR, generates linear fragments that are then used as Megaprimers to amplify the plasmid containing the protein sequence to incorporate the mutations generated by Taq into the plasmid sequence.

3.5.3 Megaprimer Whole Plasmid PCR

Megaprimer Whole Plasmid PCR, Megawhop, is a method for incorporating a DNA fragment into a vector by amplification of the whole vector using a template and so called megaprimers, amplified fragments containing the sequence to be cloned [17]. In this work fragments generated by epPCR are used as megaprimers in a PCR reaction to amplify the plasmid that contains the protein sequence. The mutations introduced in the megaprimers by Taq DNA polymerase are then incorporated into the plasmid, generating a library of variants of the plasmid with mutations in the protein sequence.

3.6 Fluorescence-activated cell sorting (FACS)

Fluorescence-Activated Cell Sorting uses lasers and detectors to gain information on the fluorescence and scatter signals of single cells. The signals are then used to sort the cells based on their fluorescence.

A cell suspension is passed through the nozzle which separates the suspension into single cell droplets. Lasers are aimed at the single cell in the stream and

signals are registered by detectors for Forward Scatter (FSC), Side Scatter (SSC) and Fluorescence. Based on obtained signals the stream is charged and when the single cell droplet breaks of the stream it is guided by the deflection plates to a specified sorting bin. After the droplets break from the stream, the stream is neutralized and ready to be charged again for the next droplet to be sorted. See scheme in Figure 7.

3.6.1 How FACS was used in this work

The constructed plasmid used for creating a library, see 3.3 contain two reporter proteins that can be used to evaluate the stability of the generated variants. The *tag*RFP linked to Lumazine Synthase acts as a reporter for the expression of the protein which can be monitored by the red fluorescence signal. The stress response of the cell in response to misfolded or aggregated over expressed protein induces the DnaK promoter and the production of *sf*GFP which can be monitored by the green fluorescence signal.

Red fluorescence is a signal for how much of the tagged, overexpressed protein is present in the cells. Green fluorescence is a signal for how stressed the cells are and hence how stable folded overexpressed protein is. Cells that show a high Red:Green ratio is then said to express variants of the protein that show increased stability. By selecting cells with high Red:Green a variant of the sequence can be picked that gives more stable expression of the protein. After sorting the cells, variants are sequenced to determine sequences that improve the stability. These variants should also be checked for premature termination of the protein sequence.

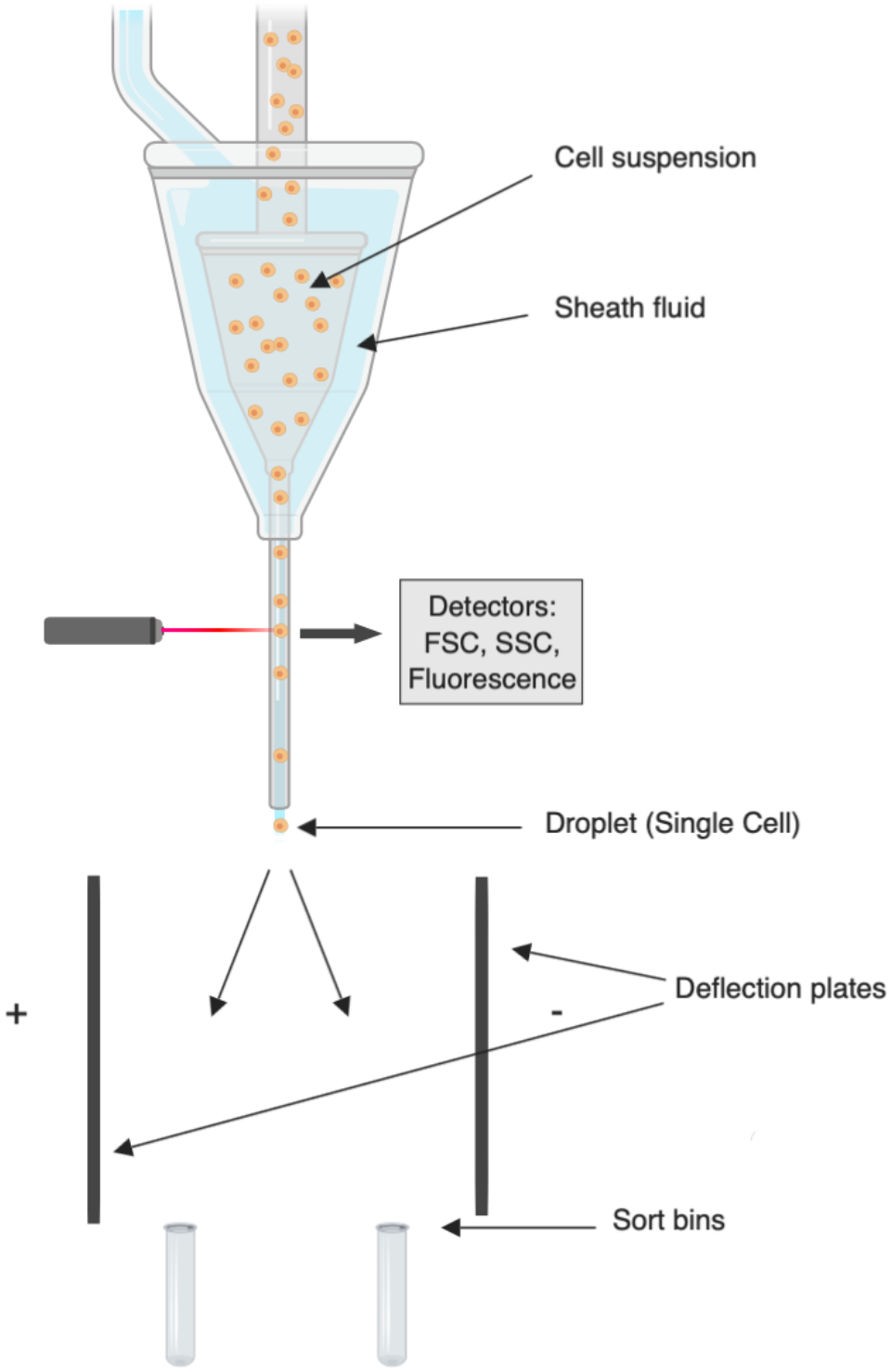


Figure 7: Scheme describing the principle of Fluorescence-Activated Cell Sorting. The Cell suspension is passed through the nozzle which divides the suspension into single cell droplets. Lasers are aimed at the single cell in the stream and signals are registered by detectors for Forward Scatter (FSC), Side Scatter (SSC) and Fluorescence. Based on obtained signals cells can be sorted. Figure created with biorender.com

3.7 Summary of methods used for directed evolution

Directed evolution is used to find more stably expressing variant of the ancient sequence of icosahedral Lumazine Synthase, see Figure 8 . Firstly, mutated fragments spanning the gene for LS are generated by Error prone PCR. The mutations are incorporated into the expression vector by Megaprimer Whole plasmid PCR. Fluorescence-Activated cell sorting is then performed using the RFP and GFP signals as part of the stability assay system. More stable variants, with higher RFP:GFP ratio.

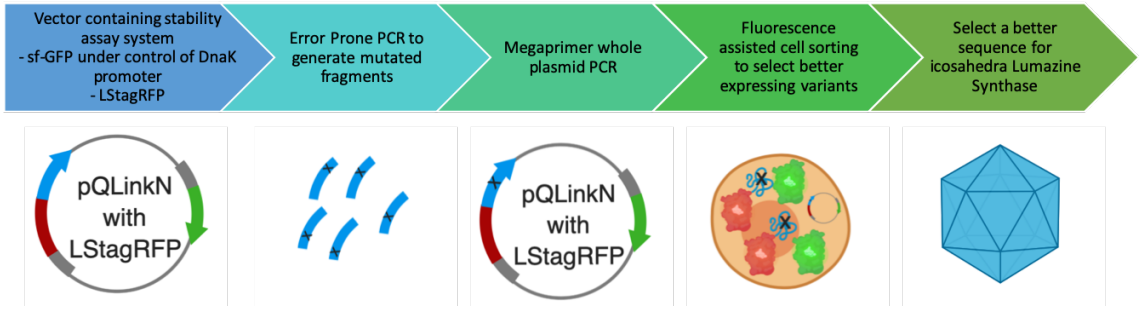


Figure 8: Workflow for directed evolution of icosahedra Lumazine Synthase. Figure created with biorender.com

4 Materials

4.1 Enzymes

Listed in Tables 1 and 2 are the enzymes and enzyme buffers used.

Enzyme	Bought from
SwaI FastDigest	Thermo Scientific
PacI FastDigest	Thermo Scientific
T4 DNA polymerase	Thermo Scientific
BamHI FastDigest	Thermo Scientific
DreamTaq DNA polymerase	Thermo Scientific
Phusion HS polymerase	New England Biolabs
Taq DNA Ligase	New England Biolabs

Table 1: Enzymes

Buffer	Bought from
FastDigest Green buffer	Thermo Scientific
T4 polymerase buffer	Thermo Scientific
Dream Taq buffer	Thermo Scientific
GC buffer	New England Biolabs
HF buffer	New England Biolabs
Taq Ligase Buffer	New England Biolabs

Table 2: Enzyme buffers

4.2 DNA purification kits

Purification of DNA from agarose gel was performed with QIAquick Gel Extraction Kit and of PCR products with QIAquick PCR Purification Kit from Qiagen. Purification of plasmid DNA from cell cultures was performed with Thermo Scientific GeneJET Plasmid Miniprep Kit.

4.3 Buffers

Protein purification buffers	
Lyis buffer:	HEPES 100 mM NaCl 500 mM DTT 2 mM cOmplete protease inhibitor cocktail (Roche) pH 7,6
Urea wash buffer:	HEPES 100 mM EDTA 5 mM DTT 5 mM Urea 2 M Triton X-100 2% pH 7,6
Wash buffer without urea:	HEPES 100 mM EDTA 5 mM DTT 5 mM pH 7,6
Guanidine-HCl extraction buffer:	HEPES 100 mM EDTA 5 mM NaCl 100 mM Guanidine·HCl 8 M pH 7,6
Size exclusion buffer	HEPES 100 mM NaCl 100 mM EDTA 5 mM Guanidine·HCl 6 M pH 7,6
Dialysis buffer:	Sodium Phosphate 150 mM pH 7
Phosphate buffer:	Sodium Phosphate 50 mM Sodium chloride 150 mM pH 7

Table 3: Buffers used for protein purification

Electrophoresis buffers	
DNA Loading dye:	Bought from Thermo Fischer
TAE buffer:	Tris base 40 mM Acetic acid 20 mM EDTA 1 mM
SDS loading buffer:	Tris-HCl pH 6,8 125 mM Glycerol 20 % SDS 10 % w/v bromophenol blue 0,01 % DTT 100 mM
SDS page running buffer:	Glycine 192 mM Tris 25 mM SDS 0,1 %
SDS page staining solution	Coomassie Brilliant Blue 60 mg/mL HCl 0,3 %

Table 4: Buffers used for DNA gels and SDS page

4.4 Growth media

E. coli growth media	
LB (autoclaved):	Yeast extract 5g/L Tryptone 10 g/L NaCl 10 g/L
SOC (autoclaved):	Tryptone 20 g/L Yeast extract 5 g/L NaCl 0,5 g/L KCl 2,5 mM $MgCl_2$ 10 mM $MgSO_4$ 20 mM (Seperately autoclaved) Glucose 20 M (Seperately autoclaved)

Table 5: Media used for transformation of bacteria and cultures.

5 Methods

5.1 General methods for electrophoresis

5.1.1 SDS-Page

Samples for SDS page was mixed with SDS Loading buffer and gels, Biorad, was run for 25-30 minutes and 200 V. After electrophoresis the gel was put in water, heated for 1 minutes and left shaking for 5 minutes. The wash was repeated 3 times and SDS staining solution, see Table 4, was added to gel. Gel with staining solution was heated for 30 seconds and left shaking for 30 minutes. Staining solution was poured of and replaced with water to destain the gel.

5.1.2 Preparation of samples containing guanidine for SDS page

450 μL 95 % cold ethanol was added to 25 μL protein sample and sample vortexed and frozen at -20°C 10 minutes. Sample was centrifuged 5 minutes at 20000 xg and ethanol removed. Pellet was dissolved in 250 μL 95 % ethanol, vortexed and centrifuged as before. Ethanol was removed and residual ethanol evaporated. The protein pellet was suspended in SDS loading buffer, see Table 4, and used for SDS page.

5.1.3 Agarose Gel Electrophoresis

DNA samples were mixed with DNA Loading dye (or digested in fast digestion green buffer) and loaded to 1 % Agarose gel (1 % agarose in TAE buffer, see Table 4) with 0.5x gelstar. Electrophoresis was run for 60 minutes and 100 V. Protein bands were visualized by black light.

5.2 Transformation and competent cells

5.2.1 Ca^{2+} competent cells preparation

BL21star and XL1blue from glycerol stocks were inoculated in 5 mL LB and grown overnight at 37 °C in shaking incubator. OD_{600} of cultures were measured and cultures diluted in 100 mL LB to OD_{600} =0.1 and grown for 3h at 37 °C in shaking incubator.

For future preparations of competent cells, overnight cultures were used to make new glycerol stocks by mixing 500 μL culture with 500 μL sterile 50 % glycerol in cryotubes and stored at -80 °C. Ca^{2+} competent Tuner DE3 cells had been previously prepared.

Cells were collected by centrifugation for 10 minutes at 3000xg after 10 minutes on ice. The media was removed and cells re-suspended in 10 mL cold sterile 0.1 M $CaCl_2$ and incubated on ice 20 minutes. Cells were again collected by centrifugation 10 minutes at 3000xg, supernatant discarded and cells re-suspended in 5 mL cold sterile 0.1 M $CaCl_2$ /15 % Glycerol. Dissolved cells were frozen with liquid nitrogen in 50 μL aliquotes and stored at -80 °C.

5.2.2 Transformation of Ca^{2+} competent cells

One aliquote of 50 μL thawed Ca^{2+} competent cells were carefully mixed with 1 μL plasmid or Megawhop product, see 5.9.2, and incubated 30 minutes on ice. Cells were heat-shocked 45 seconds in water bath at 42 °C and incubated 2 minutes on ice. 450 μL pre-heated (37 °C) SOC medium was added and sample incubated 1 hour at 37 °C in shaking incubator. 50 - 200 μL transformed cells were plated on LB agar plates containing suitable antibiotics.
When highly competent BL21Gold-pLysS cells (Agilent) was used, the same procedure was carried out except that the heat shock was limited to 20 seconds.

5.3 Expression of Lumazine Synthase with N-terminal His

Tuner DE3 was transformed with Lumazine Synthase with N-terminal 6xHistidine tag in pET28c(+), LShis. Single colonies were inoculated in 5 mL LB (50 μg/mL Kanamycin) and grown at 37 °C for 17 hours. Culture was diluted to OD_{600}=0.1 in 50 mL LB (50 μg/mL Kanamycin) in 250 mL baffled flasks. Cells were grown at 37 °C until OD_{600}=0.4-0.6, IPTG was added to a final concentration of 1mM and cells were induced for 3 hours at the same temperature. All cultures were grown with 150-200 rpm shaking. Cells were collected by 15 minutes centrifugation at 3000 xg and media removed. Cell pellets were stored at -20 °C until used for purification.

5.4 Purification of Lumazine Synthase with N-teminal histag

5.4.1 Lysis

Cell pellets were suspended in cold Lysis buffer, see Table 3. While kept on ice cellsuspension was sonicated at 40%amplitude, 6 sec pulses (20s on, 20 sec off repeated 8 times). Sonicated cell suspension was centrifuged 15 minutes at 15000 xg. Samples of supernatant and pellet were analyzed with SDS page.

5.4.2 Inclusion body extraction

The insoluble pellet from lysis, 5.4.1, was suspended in urea containing wash buffer, see Table 3 using a homogenizer. Suspension was centrifuged 30 minutes at 20000 xg and supernatant discarded. The was procedure was repeated 4 times with the urea containing wash buffer and once with wash buffer without urea.
The washed pellet was suspended in Guanidine·HCl extraction buffer, see Table 3 . Suspension was centrifuged 1 h at 50000 xg and supernatant collected.

5.4.3 Size exclusion chromatography

Superdex 75 10/300 column was equilibrated with Size exclusion buffer, see Table 3. 100 μL of extracted sample was loaded to the column and eluted at 0.3 mL/min, 500 μL fractions were collected. SDS page analysis was performed to determine which fractions contained protein of interest and these were pooled.

5.4.4 Dialysis

Pooled fractions from size exclusion chromatography, 5.4.3, was loaded to a Slide-A-Lyser Dialysis Cassette, 3 mL with 10 kDa cut-off. The cassette was placed in 1 L Dialysis buffer, see Table 3, for 3.5 h. The cassette was then shifted to 5L dialysis buffer over night.

5.5 Ligation Independent Cloning (LIC)

5.5.1 Linearization

Recipient vector (pQLinkN_SC2) and insert (LStagRFP_pQLinkN) were linearized with SwaI and PacI respectively. 1 μg DNA was mixed with 1 μL enzyme in 1X Fast Digest Green buffer and incubated at 37 °C for 1 hour. Enzymes were heat inactivated at 65 °C for 20 minutes.

Samples including controls with no added enzyme were run on agarose gel electrophoresis, see 5.1.3. Bands corresponding to linearized vector (5.6 kb) and Lumazine Synthase fragment (1.6 kb) were purified from gel.

5.5.2 3'-overhang production

Purified linearized vector and insert were combined with T4 DNA polymerase and dGTP or dCTP respectively, see Table 6 and 7 . Samples were incubated at room temperature 30 minutes followed by 20 minutes heat inactivation 75 °C.

Reaction mix 3'-overhang vector

Amount	
20 μL	Linearized vector
2.5 mM	dGTP
10 μg/L	BSA
6 U	T4 DNA Polymerase
1X	T4 polymerase buffer
Up to 40 μL	H_2O

Table 6: 3'-overhang production reaction mixture for linearized vector

Reaction mix 3'-overhang insert

Amount	
20 μL	insert fragment
2.5 mM	dCTP
10 μg/L	BSA
6 U	T4 DNA Polymerase
1X	T4 polymerase buffer
Up to 40 μL	H_2O

Table 7: 3'-overhang production reaction mixture for insert fragment

5.5.3 Annealing

T4 DNA polymerase treated vector and insert fragment was combined with vector:insert ratios 1:7 and 1:10 in total volume 20 μL and incubated 20 seconds at 65 °C. Controls with only vector and only insert fragment were also performed. After cooling to room temperature 1 μL 25 mM EDTA was added, samples carefully mixed and incubated 5 minutes at room temperature. 2 μL of each sample was transformed into XL1blue.

5.5.4 Confirmation of LIC products

Single colonies from transformation plates were grown in 10 mL LB (100 μg/mL ampicillin) at 37 °C and 200 rpm 17 hours. Plasmid purification was performed with plasmid miniprep kit. 2 μL purified plasmid were digested with 1 μL fast digest BamHI in 20 μL Fast Digestion Green buffer and analysed with agarose gel electophoresis. 4 clones were also sent for sequencing with eurofinsgenomics. One clone that was determined correct was then used for further work.

5.6 Expression of LStagRFP_SC2

Overnight culture was inoculated with a single colony or from a glycerol stock in LB (50 μg/ml Ampicillin and 34 μg/ml cloramphenicol). OD_{600} was measured and 250 mL LB (ampicillin and cloramphenicol) was inoculated to OD_{600}=0.1. Cultures were grown at 37 °C in shaking incubator for 2-3h until OD_{600}=0.5. IPTG was added to a final concentration of 2 mM and cultures were grown at 20 °C in shaking incubator for 21-22 h.

Cells were collected by centrifugation 30 minutes at 5000 xg. 5.5 g cells (from 500 mL expression culture) were dissolved in 50 mL Lysis buffer, see Table 3. The solution was sonicated 4 x 90 cycles (1 s on, 1s off, 50%) and centrifuged 30 minutes at 15000 xg.

5.7 Purification of LStag_RFP

To supernatant after cell lysis $(NH_4)_2SO_4$ was step wise added to a saturation of 25 %. After addition the suspension was stirred for at least 30 minutes at 4 °C and centrifuged 30 minutes at 20000 xg. The pellet was frozen and to the supernatant $(NH_4)_2SO_4$ was added to a final saturation of 30 %. Stirring and centrifugation was repeated. The pellet was dissolved in 5 mL Phosphate buffer, see Table 3. Samples were centrifuged 5 minutes at 3000 xg, pellet discarded and supernatant filtered (0.2 μm wattmann filter) before it was further used.

5.8 Verification of oligomeric state

5.8.1 Size exclusion chromatography

500 μL 4.4mg/mL filtered supernatant after $(NH_4)_2SO_4$ precipitation was applied to a Superose 6 10/300 column equilibrated and ran with Phosphate buffer, see Table 3, at 0.3 ml/min flow rate. 500 μL fractions were collected and fractions containing protein was determined by absorption at 280 nm.

5.8.2 Dynamic Light Scattering (DLS)

Fractions form size exclusion chromatography containing protein and filtered supernatant after Ammonium sulphate precipitation was analyzed with DLS using Malvern Panalytical Zetasizer.

5.8.3 Analytical Ultra Centrifuge (AUC)

Ammonium sulphate precipitated sample was used in three different concentrations (2.3, 0.6 and 0.096 mg/mL) for velocity sedimentation with Beckman Coulter Optima AUC, at 25000 xg for 8.5 h. The rotor used was AN 60 Ti at 20 °C with 30 minutes scan delay and 409 scans. Absorption was measured at 230, 280 and 300 nm. Data was analyzed using UltraScan III Analysis Software.

5.9 Random mutagenesis

Random mutagenesis was performed to incorporate mutations to the Lumazine Synthase sequence. The mutations were indroduced by error prone PCR to produce mutated megaprimers combined with Megaprimer whole plasmid PCR to incorporate the mutations into the vector.

5.9.1 Error-prone PCR

Primers for error-prone PCR, epPCR were designed to flank Lumazine Synthase in pQLinkN(SC2) vector, see Appendix II. The forward primer, named CO_FWD_1, covers the last 15 bases of the sequence for TagRFP and the first 5 bases of the GC-linker. The reverse primer, named CO_REV_1, covers the last 16 bases of the sequence for Lumaziune Synthase including the stop codon and 4 bases of the following sequence in the plasmid. Sequences for the designed primers are found in Appendix III.

EpPCR was performed using the reaction mixture listed in Table 8. $MnCl_2$ concentrations 200 and 500 μM were used. The PCR program used is described in Table 9 with 21 and 29 cycles.

Reaction mix error-prone PCR

Amount		Work name
100 ng	Template Plasmid DNA	LStagRFP_SC2
0.25 μM	Forward primer	CO_FWD1
0.25 μM	Reverse primer	CO_REV1
0.4 mM	each dNTP	
1X	Dream taq buffer	
200 and 500 μM	$MnCl_2$	
2.5 U	Dream taq polymerase	
Up to 50 μL	H_2O	

Table 8: epPCR reaction mixture

Error-prone PCR protocol

Step	Temperature [C °]	Time [min:sec]	
Initial denaturation	95	2:00	
Denaturation	95	0:30	× 21/29 cycles
Annealing	52	0:30	
Elongation	72	1:00	
Final elongation	72	5:00	

Table 9: PCR program used for epPCR. 21 and 29 cycles were performed.

Resulting PCR products were purified using PCR clean up kit and analyzed with agarose gel electophoresis.

5.9.2 Megaprimer Whole Plasmid PCR, Megawhop

Megaprimer whole plasmid PCR, Megawhop, was performed to incorporate the sequences mutated by epPCR into the vector. The products from the previous epPCR was used as megaprimers in the Megawhop reaction that amplifies the whole plasmid containing containing the mutated gene. Megawhop was performed with the reaction mix in Table 10 according to the protocol in Table 11. The Megawhop was performed both with and without the Taq DNA ligase and NAD^+ added and ligation steps included in the PCR protocol.

Reaction mix Megawhop

Amount		Work name
50 ng	Template Plasmid DNA	LStagRFP_SC2
30/120 ng	Megaprimer	epPCR product
0.2 mM	each dNTP	
1X	HF/GC buffer	
1 U	Phusion Hot Start II	
Up to 25 μL	H_2O or double the reaction	

Table 10: Reaction mix for Megawhop. The reaction was performed both with and without NAD^+ and Taq DNA Ligase, marked with *.

Megawhop PCR protocol

Step	Temperature [C °]	Time [h:min:sec]	
Initial elongation	72	0:30	
Initial denaturation	98	2:00	
Denaturation	98	0:30	× 20 cycles
Annealing	58	0:30	
Elongation	72	3:33	
Final extension	72	5:00	

Table 11: PCR protocol used for Megawhop.

The products from the Megawhop reaction was loaded on 1 % agarose gel electrophoresis and bands corresponding to the plasmid was purified. Purified linear plasmid created with 120 ng megaprimer from epPCR conditions 200 μ M and 21 cycles was transformed into BL21Gold-pLysS competent cells, from Agilent.

5.10 Fluorescence-Activted Cell Sorting

5.10.1 Preparation of cell cultures for FACS

LS_tagRFP in BL21STAR or LS_tagRFP library in BL21Gold-pLysS inoculated from overnight culture in 50 mL LB (50 μg/mL ampicillin and 34 μg/mL cloramphenicol) with OD_{600}=0.1. Cultures were grown at 37 °C for 2h in shaking incubator until OD_{600}=0.4. IPTG was added to a final concentration of 1 mM and incubated at 37, 30 and 20 °C. Cultures incubated at 37 and 30 °C were grown for 4 h and then left shaking at 4 °C overnight. Cultures incubated at 20 °C were grown for 21-22 h. The cells were collected by centrifugation 15 minutes at 3000 xg and re-suspended in PBS, see Table 3. Centrifugation and washing procedure was repeated and cells were finally suspended in PBS to 10^5-10^7 cells/mL.

5.10.2 Analysis of cells with FACS

Washed and PBS suspended cells were analysed on Biorad S3e Cell Sorter with excitation lasers at 488 and 561 nm, red fluorescence detection with 615/25nm filter and green fluorescence detection with 525/30 nm filter.

5.10.3 Sorting of cells with FACS

LS_tagRFP library in BL21Gold-pLysS was sorted based on gates applied to select cells that show a higher RFP:GFP ratio. 16×10^6 cells were sorted into 500 μL LB and after the first sort a sample of 500 μL was taken out, centrifuged at 200 xg 5 minutes, media was removed and the cells were suspended in PBS. The cells were then sorted once more using the same sorting gate and 5000 cells were collected. The twice sorted cells were incubated at 37 °C for 30 minutes, centrifuged at 200 xg and plated on LB agar plates (100 μg/mL ampicillin). The once sorted cells were diluted with LB (50 μg/mL ampicillin and 34 μg/ml cloramphenicol) to 10 mL and incubated overnight at 37 °C shaking. 100 μL of the overnight culture was plated on LB agar plates (100 μg/mL ampicillin) and incubated at room temperature over the weekend. New plated were then streaked from the resulting mat of bacteria and incubated at 37 °C. Single colonies were used to inoculate over night cultures, from which plasmids were purified.

6 Results

In the first part of this work, Lumazine Synthase with N-terminal histag was expressed in E. coli and purified by extraction from inclusion bodies. Refolding of protein was tried by dialysis into buffer without denaturant. However refolding of the protein was not successful.

In the second part, work was performed to try to find a more stably expressing variant of the sequence of Lumazine Synthase. LS was fused to *tag*RFP and cloned into a plasmid containing a stability assay system that can be used to monitor the stress response of the cell in response to misfolded protein. The sequence for Lumazine Synthase was mutated by random mutagenesis and fluorescence-activated cell sorting was used to select for more stable protein expression. LS without mutations was analysed with FACS to investigate stability dependent on expression temperature. LS was also purified and the oligomeric state of the protein was investigated by size exclusion chromatography, dynamic light scattering and analytical ultra centrifugation.

6.1 Purification of Lumazine Synthase with N-terminal histag

Lumazine Synthase with N-terminal histag had previously been determined to express into inclusion bodies. Purification through extraction from exclusion bodies and refolding into native structure was tried. However, it was determined that refolding was not possible with the tried method.

Lumazine Synthase with N-terminal histag was expressed in E. coli BL21Star. Protein was purified from the cell cultures by extraction from inclusion bodies with guanidine·HCl and size exclusion chromatography. Protein re-folding was tried by dialysis into phosphate buffer.

6.1.1 Expression and extraction

Lumazine Synthase was expressed in E. Coli BL21star by induction with IPTG at 37 °C . Cells were lysed by sonication. SDS page of induced cells and lysed cell pellet and supernatant shows that the major fraction of the overexpressed protein is in the insoluble fraction, lane D in Figure 9.

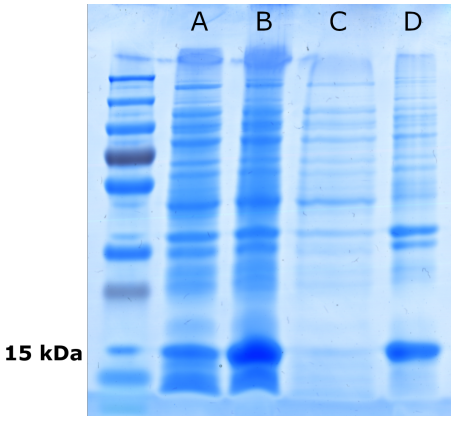


Figure 9: SDSpage electrophoresis of BL21star expressing LShis, 17 kDa. a) Uninduced cells b) Induced cells c) Lysate supernatant d) Insoluble pellet fraction.

Insoluble fraction was washed with urea buffer and protein was extracted by guanidine·HCl. Wash fractions, lanes B-D in Figure 10, contains small amounts of LShis while most of the protein is extracted with guanidine·HCl, lane A.

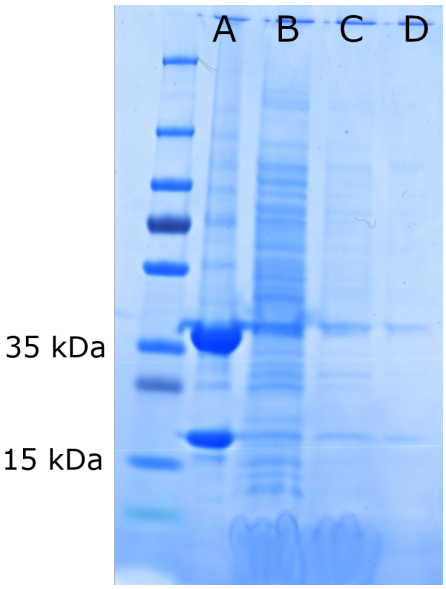


Figure 10: SDSpage electrophoresis of wash and extract fractions of insoluble fraction of LShis, 17 kDa. a) Guanidine· HCl extracted sample b-d) Urea wash fractions

6.1.2 Size exclusion chromatography

Protein extracted from inclusion bodies with guanidine·HCl was ran on Superdex 75 10/300 column in 6 M guanidine·HCl. Protein eluted as a double peak at around 8 mL, see Figure 11. Extracted sample and fractions from size exclusion chromatography were analysed with SDS page. Two protein bands at 15 and 35 kDa was found in both the extracted sample, lane A, and the fractions, lane B in Figure 12

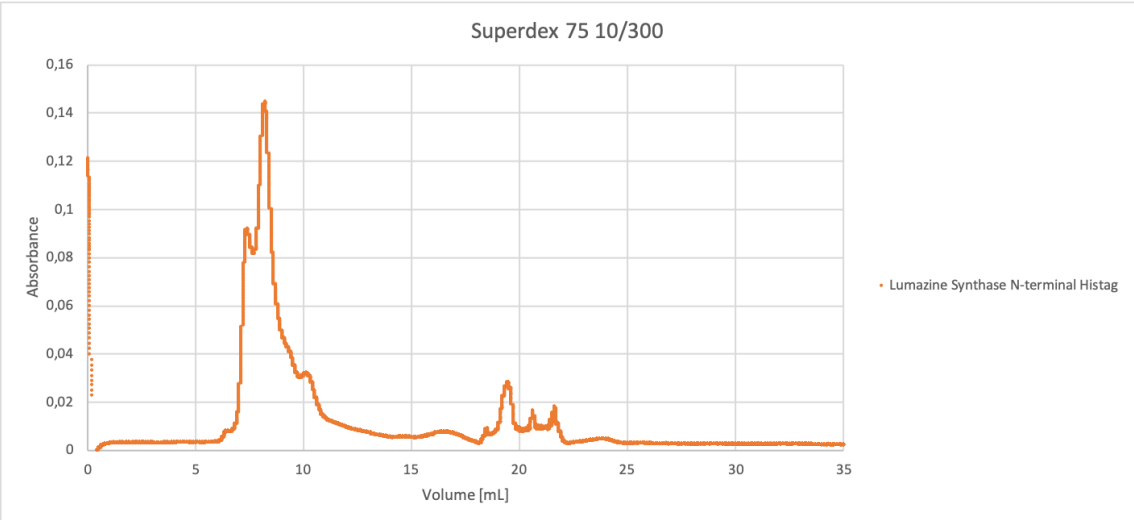


Figure 11: 100 μL LShis extracted from inclusion bodies ran on Superdex 75 10/300 column run with Size exclusion buffer, see Table 3

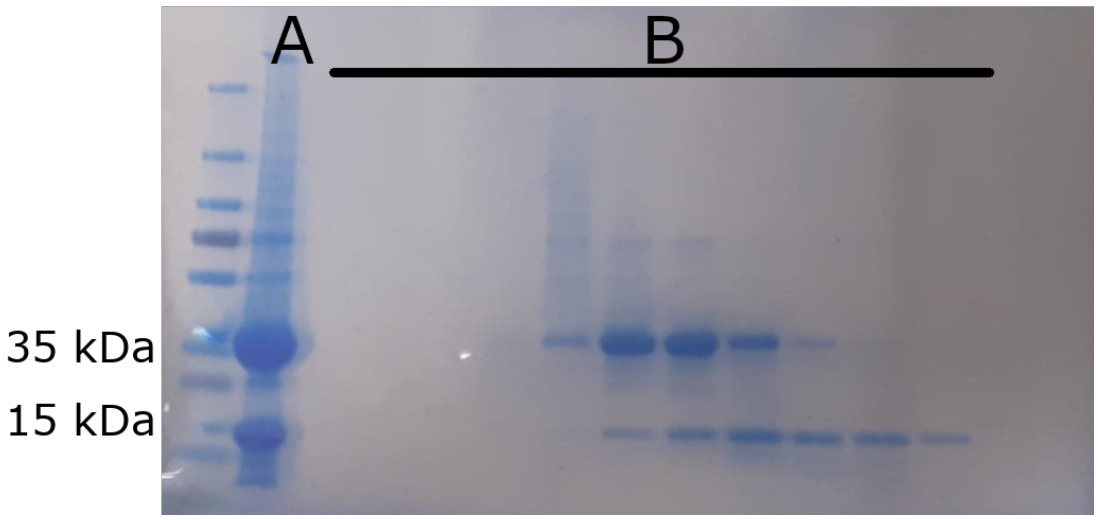


Figure 12: SDSpage electrophoresis of fractions from size exclusion chromatography of LShis, 18.6 kDa a) Guanidine· HCl extracted sample b) Size exclusion chromatography fractions 3.5-8.5 mL.

6.1.3 Dialysis

Size exclusion fractions containing protein were pooled and dialysed into phosphate buffer to refold the protein into native structure by removing the denaturant from the sample. During dialysis aggregation of the protein occurred and it was not possible to recover soluble protein.

6.2 Ligation independent cloning, LIC

In order to improve the stability of expression of Lumazine Synthase, the protein was fused to *tag*RFP. Ligation independent cloning was performed to incorporate the gene for LStagRFP into pQLinkN_SC2 vector containing the stability assay system.

Digestion of pQLinkN_SC2 vector with SwaI and the insert fragment, containing the gene for LStagRFP, with PacI was successful. In Figure 13 lane A contains digested vector, lane D contains the digested insert and lanes b-c contains non-digested controls. Insert fragment and vector annealing with vector:insert ratios 1:7 and 1:10 resulted in 3 and 5 colonies respectively. Control with only vector gave no colonies while control with only insert gave one colony. Agarose gel electrophoresis of BamHI digested purified plasmid from the resulting colonies showed that the plasmids had the right size except for the colony from only insert control which was of smaller size, see Figure 14. One of the plasmids was determined as correct through sanger sequencing. The correct plasmid was then further used for expression of LStagRFP and as template for random mutagenesis.

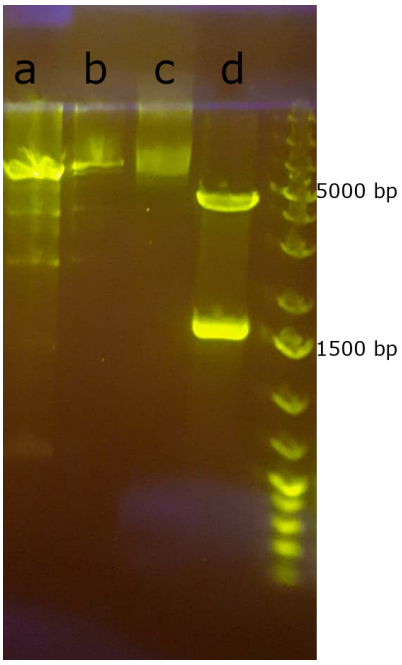


Figure 13: Agarose gel electrophoresis of linearized insert and vector. a) SwaI digested pQLinkN_SC2 vector at above 5000 bp. b) Vector control c) Insert control d) PacI digested Insert fragment at around 1500 bp and plasmid backbone at 5000bp.

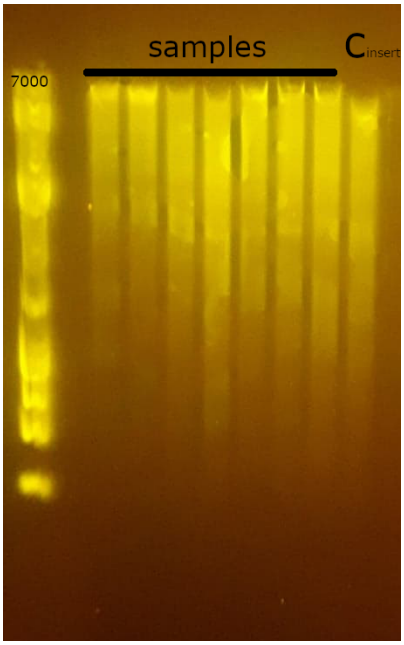


Figure 14: Agarose gel electrophoresis of purified plasmid digested with BamHI. Samples, corresponding to 8 colony forming units from cloning, have a correct size at 7000 bp. while the plasmid from the only insert control has a smaller size.

6.3 LStagRFP_SC2

The construct with Lumazine Synthase fused to *tag*RFP was expressed and purified by ammonium sulphate precipitation. Ammonium sulphate purification was chosen

since the construct did not contain any affinity tag and it was a simple way to get fairly clean protein that could be used for investigating the oligomeric state of the expressed fusion protein.

LStagRFP in pQLinkN_SC2 vector was expressed in E. coli strain BL21star. Protein was purified by step wise ammonium sulphate precipitation and size exclusion chromatography. Oligomeric state of purified protein was examined by DLS and AUC.

6.3.1 Size exclusion chromatography

LStagRFP purified by ammonium sulphate was ran on Superose 6 10/300 column. Protein eluted at around 7 mL, see Figure 15.

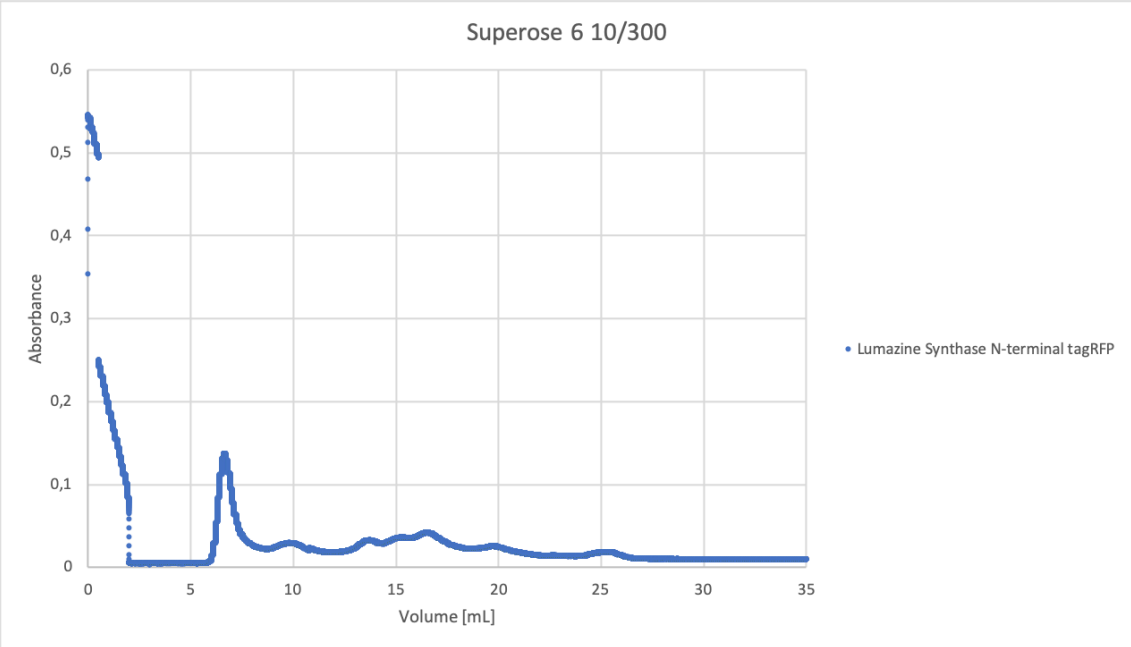


Figure 15: 2.2 mg LStagRFP ran on Superose 6 10/300 column in phosphate buffer.

6.3.2 Dynamic Light Scattering (DLS)

Dynamic Light Scattering was performed to gain information on the size distribution of the purified protein particles. LStagRFP purified by ammonium sulphate precipitation and dissolved in Phosphate buffer, gave two populations with 253.2 (79%) and 60.09 (21%) nm diameter in size distribution by intensity. LStagRFP was then loaded to a size exclusion chromatography column. The eluted protein analyzed with DLS, see Figure 15, gave a single population with 110.6 nm diameter. This suggests that the protein sample is not aggregated since this should give a broad distribution of particle sizes. The size exclusion chromatography eliminates some polydispersity of the sample leaving a single size of particles. However the obtained size of the particle is larger than expected for Lumazine Synthase capsids, the diameter of Aquifex Aeolicus Lumazine Synthase is reported to be 15.4 nm [5] and *tag*RFP is a 27 kDa protein. Perhaps the apparent bigger size is caused by the

flexibility of the linker and *tag*RFP units on the outside of the capsids. See also Size Distribution Reports in Appendix IV.

6.3.3 Analytical Ultra Centrifugation (AUC)

Lumazine Synthase with *tag*RFP purified with ammonium sulphate precipitation was analysed with analytical ultra centrifugation to gain information on the oligomeric state of the protein. Three different concentrations of the sample was used in order to screen for optimal sample concentration.

Sedimentation velocity experiment was run with LStagRFP at 0.096, 0.6 and 2.3 mg/mL. Figures 16, 17 and 18 shows Van Holde-Weischet analysis of the scans. At 0.096 mg/mL, Figure 16, three separate sedimentation coefficients was obtained with 40%, 25% and 20% of the sample in each. At 0.6 mg/mL, Figure 17, one sedimentation coefficient could be distinguished with 60 % of the sample. At 2.3 mg/mL, Figure 18, no sedimentation coefficient could be distinguished. For a sample with only one oligomeric state, only one sedimentation coefficient is obtained and the Van Holde-Weischet analysis gives a vertical alignment of the points. Also the obtained sedimentation coefficients are supposed to be on the positive scale, which is not the case for the samples at 0.096 mg/mL and 2.3 mg/mL. Extrapolation plot of 0.6 mg/mL data, see Figure 19 shows a shift of the extrapolation point from 0 and also an irregular distribution of the extrapolations. In optimal experiment conditions and with only one oligomeric state all extrapolations meet at origo. In this extrapolation plot the lines meet at approximately 0.01 and not a single point.

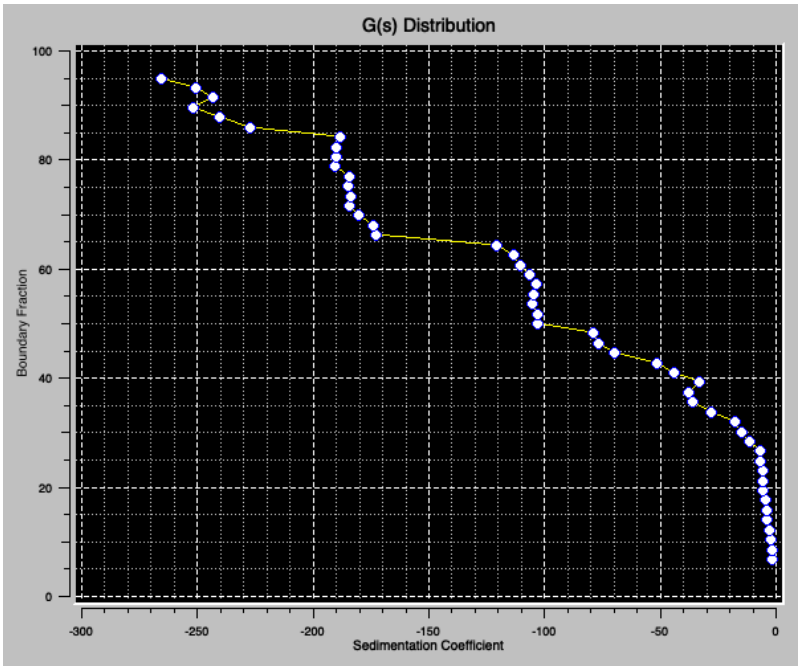


Figure 16: Analytical Ultra Centrifuge data of 0.096 mg/mL LStagRFP at 230 nm Van Holde-Weischet analysis

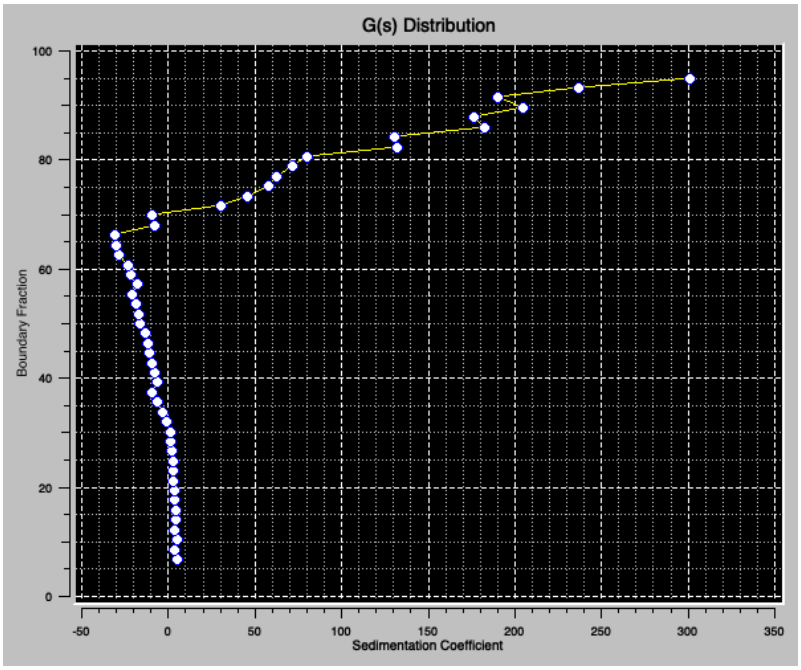


Figure 17: Analytical Ultra Centrifuge data of 0.6 mg/mL LStagRFP at 280 nm Van Holde-Weischet analysis

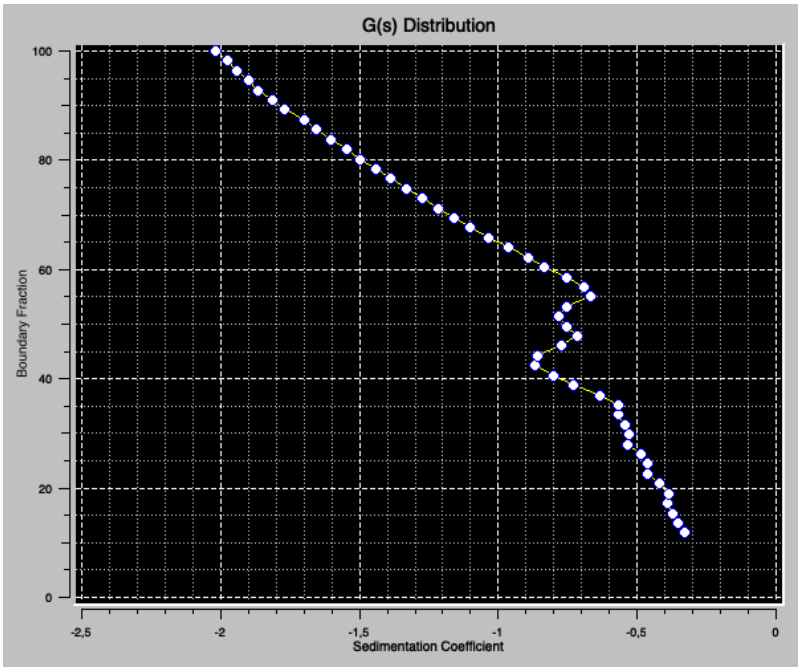


Figure 18: Analytical Ultra Centrifuge data of 2.3 mg/mL LStagRFP at 280 nm Van Holde-Weischet analysis

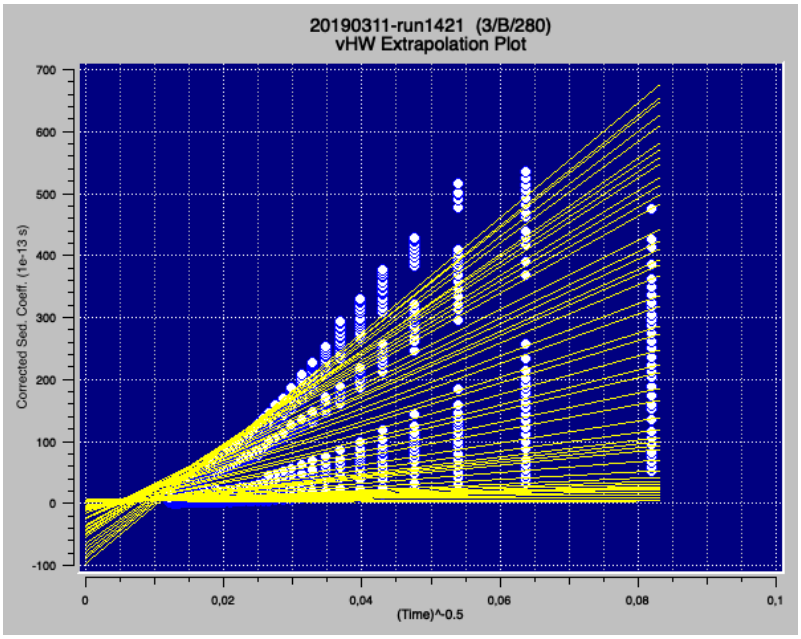


Figure 19: Analytical Ultra Centrifuge data of 0.6 mg/mL LStagRFP at 280 nm extrapolation plot

6.4 Random mutagenesis

Random mutagenesis of the Lumazine Synthase sequence was performed in order to use directed evolution to find a better sequence. Mutations were introduced to the sequence by error prone PCR in combination with Megawhop PCR. Error prone PCR produces megaprimers that are DNA fragments spanning the sequence of Lumazine Synthase containing mutations. The megaprimers are then used in the Megawhop PCR to amplify the whole plasmid and introduce the mutations into the sequence.

6.4.1 Error Prone PCR

Error Prone PCR was performed to produce megaprimers, DNA fragments spanning the Lumazine Synthase sequence, with randomly distributed mutations of the sequence. In order to obtain different ratios of mutations $MnCl_2$ concentrations of 200 and 500 μM and PCR cycle numbers of 21 and 29 cycles was used.

Purified epPCR products ran on agarose gel showed correct size of 565 bp, see Figure 20. Total amount of 3.5 μg was obtained for each PCR condition.

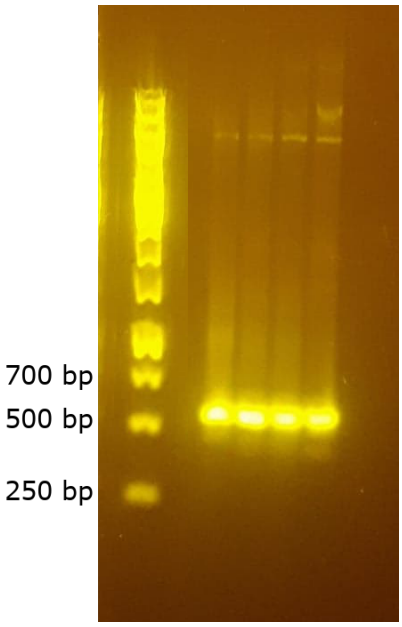


Figure 20: Agarose gel electrophoresis of purified Error Prone PCR products produced with 200/500 μM $MnCl_2$ and 21/29 PCR cycles.

6.4.2 Megawhop

Megaprimers produced by error prone PCR was used to amplify the whole LStagRFP_SC2 plasmid and incorporate the mutations in the megaprimers into the plasmid. This creates a library of plasmids with the stability assay system and variations of the Lumazine Synthase sequence fused to *tag*RFP.

With 120 ng megaprimer in the megawhop reaction, the resulting agarose gel electrophoresis shows a lot of smear and the most prominent band corresponds to linear plasmid fragments, see Figure 21.

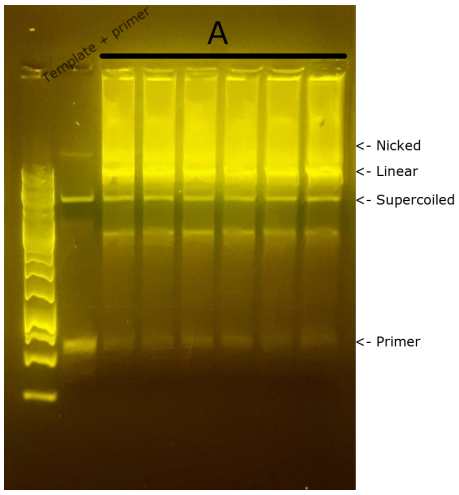


Figure 21: Agarose gel electrophoresis of megawhop product, 120 ng megaprimer in reaction mix. A. Megaprimer epPCR conditions: 500 μM $MnCl_2$ 29 cycles.

When 30 ng megaprimer was used for the Megawhop reaction mostly super coiled and nicked plasmid was obtained, see Figure 22. No DpnI treatment was performed

and the lower band corresponds to supercoiled template plasmid. Lanes A-D contain megawhop products produced with four different megaprimers produced with the four different epPCR conditions, 200/500 μM $MnCl_2$ and 21/29 cycles.

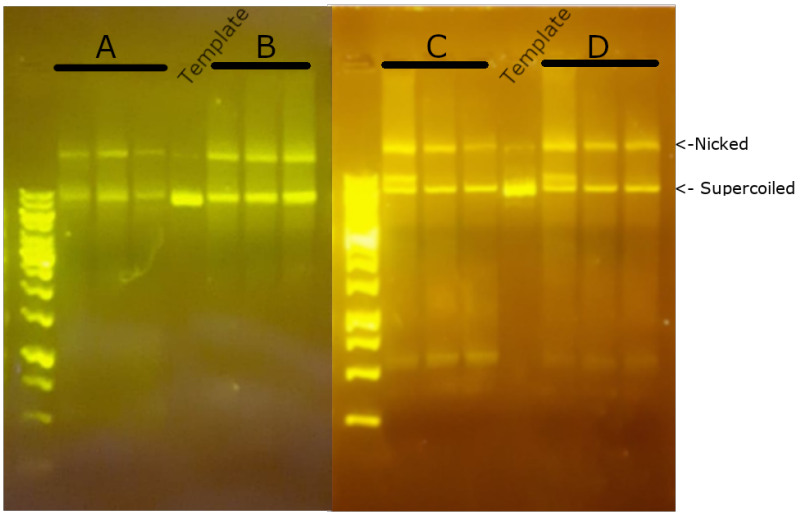


Figure 22: Agarose gel electrophoresis of megawhop product, 30 ng megaprimer in reactionmix. Different epPCR conditions megaprimers used. A 200 μM $MnCl_2$ 21 cycles, B 500 μM $MnCl_2$ 21 cycles, C 200 μM $MnCl_2$ 29 cycles and D 500 μM $MnCl_2$ 29 cycles.

Comparison HF and GC buffers from New England Biolabs, see Table 2, in the megawhop reaction mix gives a somewhat stronger band on agarose gel electrophoresis for GC buffer, B in Figure 23.

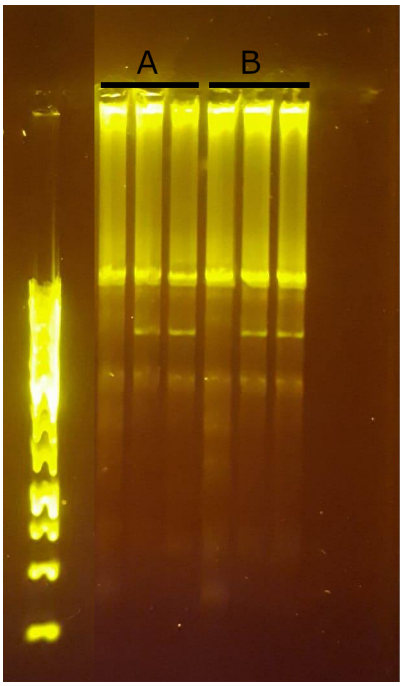


Figure 23: Agarose gel electrophoresis of megawhop product, 60 ng megaprimer in reactionmix. Comparison of HF and GC buffer in reaction mix. Megaprimer epPCR conditions: 500 μM $MnCl_2$ 29 cycles. Comparison of HF, A, and GC, B, buffer in reaction mix.

6.4.3 Transformation of library into BL21Gold-pLysS

In order to use the library of plasmids created by megawhop PCR they have to be transformed into E. coli. It is crucial that enough transformants are obtained in order to have the library cover all of the variants created by error prone PCR and megawhop PCR. For the purpose of this work, the aim is that all single amino acid mutations of the protein sequence are represented in the library. Therefore at least 4000 transformants are desired for the plasmids created by megawhop.

Transformation of Megawhop product, linear fragment purified from gel Figure 21, into BL21Gold-pLysS resulted in 840 transformants for 100 μL of competent cells. The lower number of transformants might be because linear fragment of plasmid was use. Nicked plasmid from gel Figure 22 has not yet been transformed into BL21Gold-pLysS.

6.5 FACS

Fluorescence-activated cell sorting was used to analyze the over expression of LStagRFP and stress response of the cells. Red fluorescence was used as a signal for protein expression and green fluorescence was used as a signal for the cellular stress response. Firstly, non-mutated LStagRFP_SC2 in BL21Star was induced at three different temperatures to investigate the effect on expression temperature on the stability of the expression. Secondly, LStagRFP_SC2 library in BL21Gold-pLysS was analysed and sorted for high RFP:GFP ratio.

6.5.1 Stability depending on expression temperature

LStagRFP_SC2 was induced at three different temperatures and analysed with FACS to investigate the effect of expression temperature on the stability of the over expressed protein and the stress response.

When BL21Star cells with LStagRFP_SC2 plasmid was induced at 20 °C o/n they showed higher Red fluorescence and lower Green fluorescence, and hence higher RFP:GFP ratio, than cells induced at 30 and 37 °C for 4 hours and incubated o/n in cold, see Table 12. The green fluorescence at 20 °C, Figure 24 a, gives a single peak with a shoulder towards higher fluorescence centered at about 3.8, at 30 °C, Figure 25a, a single peak centered about 4.3 and at 37 °C, Figure 26 a, a single peak centered at about 4.8. The red fluorescence at 20 °C, Figure 24 b, gives two combined peaks centered about 5.0 and 5.5, at 30 °C, Figure 25 b, three combined peaks centered about 3.7, 4.7 and 5.1 and at 37 °C a single peak with a shoulder towards lower fluorescence centered about 4.3. The resulting RFP:GFP plot of cells induced at 20 °C is shifted towards higher RFP and lower GFP compared to the plots of cells induced at 30 and 37 °C.

Red and green fluorescence at different expression temperature

Expression temperature [°C]	Mean RFP	Mean GFP	RFP:GFP ratio
20	5.099	3.707	1.376
30	4.577	4.202	1.089
37	4.189	4.588	0.913

Table 12: Mean RFP, GFP and RFP:GFP ratios for cells analyzed with FACS. Protein expression at 20, 30 and 37 °C

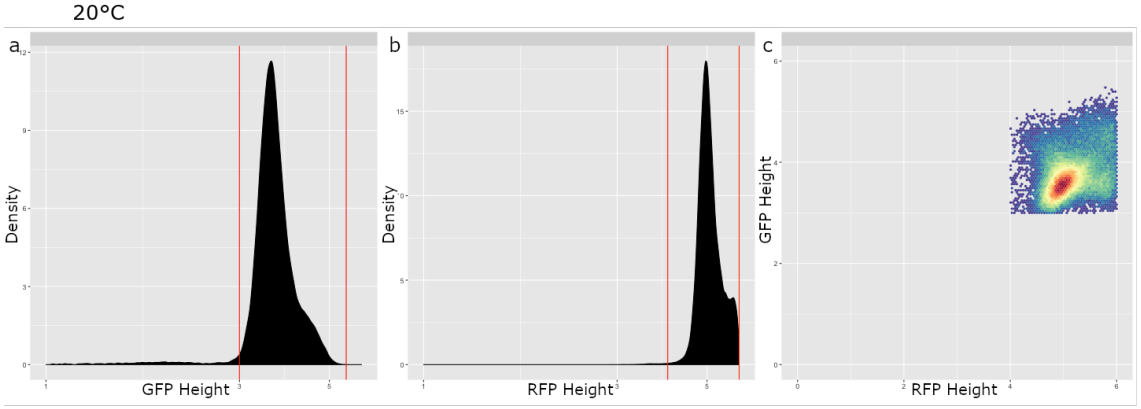


Figure 24: Data from FACS of LStagRFP expressed at 20 °C. a) Green fluorescence signal b) Red fluorescence signal c) RFP vs GFP plot

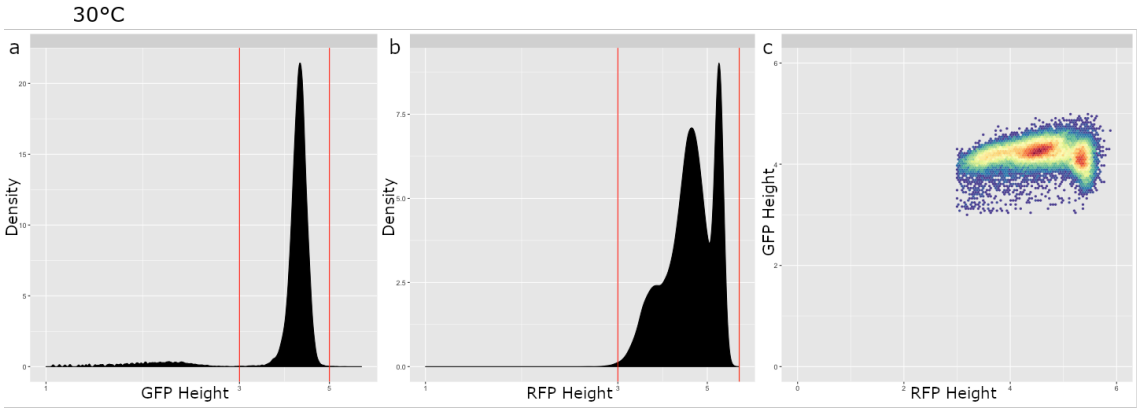


Figure 25: Data from FACS of LStagRFP expressed at 30 °C. a) Green fluorescence signal b) Red fluorescence signal c) RFP vs GFP plot

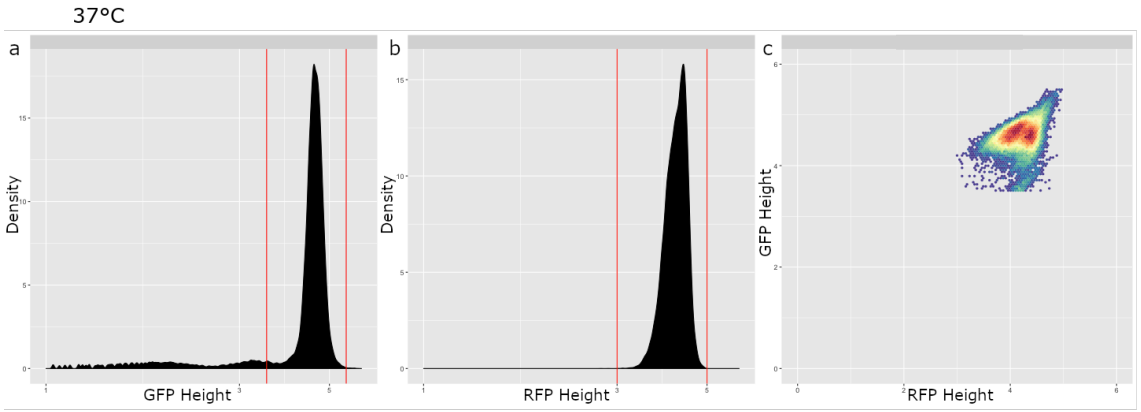


Figure 26: Data from FACS of LStagRFP expressed at 37 °C. a) Green fluorescence signal b) Red fluorescence signal c) RFP vs GFP plot

6.5.2 Library sort

LStagRFP_SC2 library was analysed and sorted with FACS to select cells with higher RFP:GFP ratio. This is done to select variants of the Lumazine Synthase sequence that are more stably expressed in E. coli.

BL21Gold pLysS with LStagRFP_SC2 library was grown and induced at 37 °C. The library showed a mean RFP signal of 4.39, mean GFP signal of 4.62 and a RFP:GFP ratio of 0.949, see Table 13. Gates were set up to sort the 3%, gate A, and 3-10%, gate B, of the cells that had the highest RFP:GFP ratio, see Figure 27 c. The cells within the sort gates gave mean RFP of 4.583 and 4.442, Mean GFP of 3.920 and 4.178, and RFP:GFP ratios 1.169 and 1.090, see Table 13. When the sorted cells in gate A were re-sorted they gave Mean RFP 4.562, Mean GFP 4.246 and RFP:GFP ratio 1.074. For the cells in gate B, the re-sort gave Mean RFP 4.585, Mean GFP 4.344 and RFP:GFP ratio 1.055. The RFP:GFP ratio for the library increased after the sorting, from 0.949 to 1.074 and 1.056 for gate A and B respectively.

21 variants of the cells sorted into gate A were sequenced, where 16 were wild-type, 3 had one nucleotide mutation resulting in C85R in the Lumazine Synthase sequence. Two sequences had 5 nucleotide mutations out of which 3 were common for the two, resulting in S to G in $(GS)_1$0-linker, P19R and L38P. The last two mutations were D44V and A45S, and Y80H and I84T respectively.

Red and green fluorescence of sorted and re-sorted library

Sort	Mean RFP	Mean GFP	RFP:GFP ratio
First sort	4.386	4.620	0.949
3 % highest RFP:GFP GateA	4.583	3.920	1.169
3-10 % highest RFP:GFP GateB	4.552	4.178	1.090
Gate A re-sort	4.562	4.246	1.074
Gate B re-sort	4.585	4.344	1.055

Table 13: Mean RFP, GFP and RFP:GFP ratios for cells analyzed and sorted with FACS. Protein expression at 37 °C

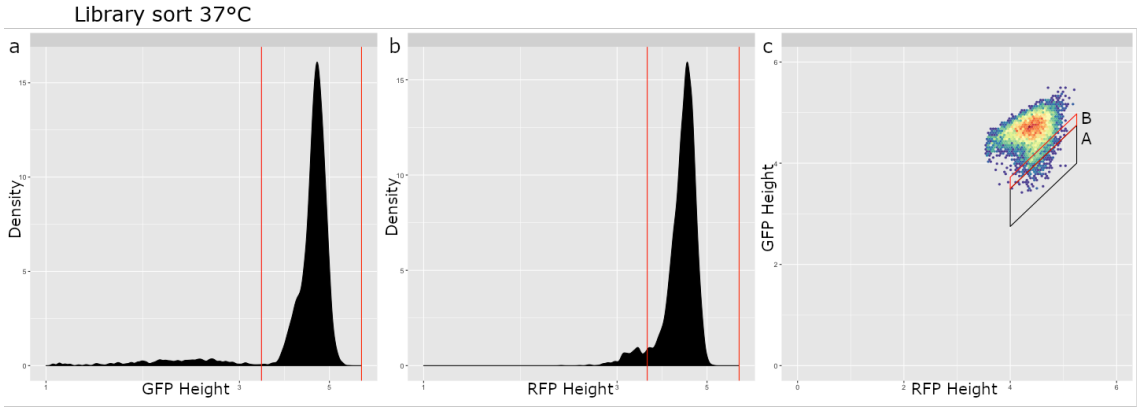


Figure 27: Data from FACS of LStagRFP library expressed at 37 °C. a) Green fluorescence signal b) Red fluorescence signal c) RFP vs GFP plot

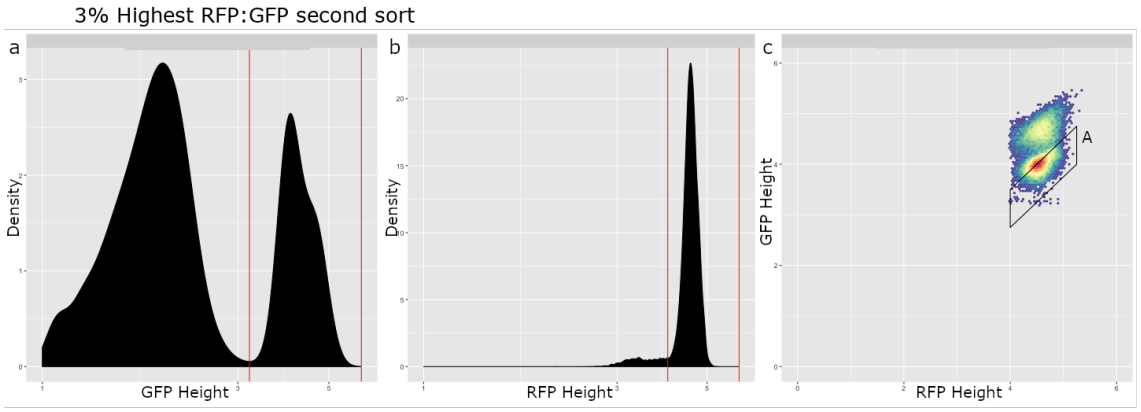


Figure 28: Data from FACS of LStagRFP library resort 3% highest RFP:GFP. a) Green fluorescence signal b) Red fluorescence signal c) RFP vs GFP plot

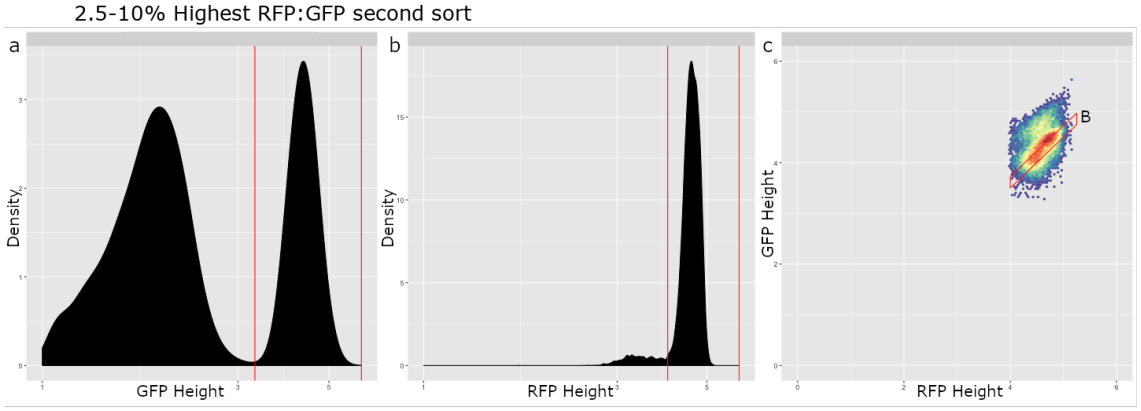


Figure 29: Data from FACS of LStagRFP library resort 3-10% highest RFP:GFP. a) Green fluorescence signal b) Red fluorescence signal c) RFP vs GFP plot

7 Discussion

An ancient sequence of icosahedra Lumazine Synthase had previously been reconstructed. LS from the reconstructed sequence had been expressed and determined to express into inclusion bodies.

The first part of this work purified LS with a histag from inclusion bodies and tried to refold the protein into native conformation. The refolding tries were not successful and therefore directed evolution was to be used to improve the stability of the protein expression

In the second part of this work directed evolution was used to improve the reconstructed sequence of Lumazine Synthase. This was done by random mutagenesis of the sequence in combination with a stability assay system, consisting of *tag*RFP fused to the protein sequence and *sf*-GFP expressed under the control of the stress activated DnaK promoter.

7.1 Lumazine Synthase with N-terminal histag

Lumazine Synthase with N-terminal histag had previously been expressed and it was determined that no soluble expression could be achieved. Therefore work in this thesis was performed to purify the protein from inclusion bodies. Extraction from inclusion bodies was successful but the protein could not be re-folded into native structure. It was therefore determined that the reconstructed sequence of LS did not express stably enough and finding a better expressing sequence is tried.

Lumazine Synthase with N-terminal histag was possible to express in E. coli BL21Star and extract from inclusion bodies with 8 M guanidine·HCl. Extracted LShis, diluted to 6 M guanidine·HCl, ran on Superdex 75 10/300 column eluted as two peaks very close to the void volume indicating weights above 75 kDa. Although, in high concentration of denaturant protein should be in random coil [18] and in solution as monomers. For LShis the monomers have weight 18.6 kDa. This means that the protein in 6 M guanidine·HCl forms some kind of aggregated of total weight of above 75 kDa. Also in SDS the extracted protein forms two populations of sizes around 15 and 35 kDa, potentially corresponding to a monomer and a dimer of LShis.

Dialysis into phosphate buffer caused the protein to precipitate. This might be caused by too swift decrease of guanidine·HCl concentration or due to low stability of the protein. The low stability of the expressed protein lead to the decision to make a new construct that could be used with directed evolution to improve the protein sequence to get more stable soluble expression.

In the new constuct Lumazine Synthase is fused to *tag*RFP, which can be used as a reporter for protein expression. The fusion protein was then cloned into a plasmid that contains *sf*-GFP under the control of stress activated Dnak promoter. This acts as a reporter for the stability of the overexpressed protein. This system allows for monitoring of the expression and stability of the over expressed protein. When the sequence of the protein is randomly mutated, the stability assay system can be used for variants that show higher expression and lower stress reaction.

7.2 LIC

Ligation Independent Cloning, LIC, was performed to clone the LStagRFP construct into the vector containing the stability assay system. LIC was determined successful and LStagRFP was incorporated into the pQLinkN_SC2 vector with the correct sequence. This plasmid was then used for further expression and purification of LStagRFP and directed evolution of the LS sequence.

7.3 Lumazine Synthase with N-terminal *tag*RFP

Lumazine Synthase with N-terminal *tag*RFP was expressed and purified by ammonium sulphate precipitation to investigate the stability and the oligomerization of the protein fused to *tag*RFP. The new construct, LStagRFP, is expressing soluble protein which indicates that the RFP helps the protein to fold. However, the protein is still aggregating which suggest that the protein still has low stability. The fusion to RFP helps with folding but the fold is not stable enough.

LStagRFP ran on Superose 6 column eluted at 7 mL indicating that the protein is forming some larger particles. It could be either that the proteins form capsids or that it is caused by aggregation. However, there was an issue of the column being compressed during running which could affect the outcome. A major fraction of the loaded sample was lost when the sample was injected to the column. Therefore it is not known whether the protein would give rise to more than the one peak that was obtained had all the protein been properly ran on the column.

Dynamic Light Scattering measurements indicate that LStagRFP purified by size exclusion chromatography is monodisperse which suggest that is not aggregated. However the appearent size is bigger than expected with 110 nm compared to 16 nm for Aquifex Aeolicus Lumazine Synthase. LStagRFP not ran on size exclusion column gave two separate populations indicating that some impurity or aggregation was eliminated by the chromatography step.

Analytical ultra centrifuge data of LStagRFP at 0.096 mg/mL processed by the Van Holde-Weischet analysis, see Figure 16, indicated that three separate populations of different sizes were present in the sample. However, the data is not of high enough quality for making conclusions about the distribution of size or oliogomerisation state. At 0.6, see Figure 17, the Van Holde-Weischet analysis indicates only one separate population but also a continued stretch towards higher sedimentation values, indicating a broader distribution of oligomerisation states. For the highest sample concentration 2.3 mg/mL, see Figure 18, no single population was distinguished. The decreasing distribution of sedimentation coefficients indicates concentration dependent oligomerisation or aggregation. For future experiments with LStagRFP, protein concentration in the lower range of the concentrations tried should be used in order to avoid aggregation of the protein. At 0.096 mg/mL the analysis showed indications of three separate populations which should be further investigated.

In the extrapolation plot showed for 0.6 mg/mL LStagRFP, see Figure 19, the extrapolations does not meet a sedimentation coeffiecient 0 but are more wider distributed. This indicates that back diffusion has affected the sedimentation. This could possibly be limited by higher centrifugation speed and shorter spin time.

7.4 Directed evolution

To apply directed evolution to improve the stability of expression of Lumazine Synthase the sequence of LS in the stability assay plasmid was randomly mutated. The library of mutated sequences is then used with Fluorescence-Activated Cell Sorting to select sequences that could be more stable. Random mutagenesis was performed by error prone PCR and the mutations was introduced into the plasmid by megaprimer whole plasmid PCR.

Error prone PCR produced DNA fragments of the correct size, 565 bp, indicating that the reaction was successful.

Higher concentration, 120 ng/50μL see Figure 21, of megaprimer in the Megawhop reaction mix caused more smear on agarose gel electrophoresis and also most of linear plasmid. This might be caused by non-specific binding of the primers to the template. In order to clearly see what bands are formed on the agarose gel, this smear should be avoided.

When lower concentration, 30 ng/50μL Figure 22, of megaprimer was used more distinct bands were obtained on agarose gel electrophoresis. With the lower megaprimer concentration mostly nicked and super coiled plasmid was obtained. The super coiled corresponds to the template plasmid while the nicked is the newly synthesised plasmid. This shows that the lower concentration was more successful for producing the desired nicked plasmid.

A comparison of megawhop PCR using HF and GC buffer from New England Biolabs, see Figure 23, shows slightly stronger bands when GC buffer is used than when HF buffer is used. This indicates that GC buffer works better for this PCR reaction.

When purified megawhop product was transformed into 100 μL BL21Gold-pLysS 840 transformants were obtained. This is a bit less than expected, in order to have all single mutations of the sequence represented in the library, at least 4000 transformants are needed. Also more transformants are expected for the competent cells, 10^8 cfu/μg. The lower than expected transformation efficiency might be due to that linear plasmid was used. Also the concentration of DNA could have had an effect on the transformation.

7.5 FACS

The increase of RFP:GFP ratio, see Table 12, when the cells were induced at 20 °C o/n compared to 30 and 37 °C for 4 hours, indicates that the cells are able to over-express more stable protein at the lower temperature. The same relation is seen between 30 °C and 37 °C. The lower induction temperature slows down the production of over expressed protein which improves the stability of the expression.

The RFP:GFP ratio increased from 0.949 to 1.074 and 1.055 when cells were sorted for highest ratio, indicating that the sort managed to select variants for the protein that has more stable sequences. From the RFP vs GFP plots of the re-sorted cells, see Figure 28 C and 29 C, it looks as though the sorted cells fall quite well in the same RFP interval but are spread out in GFP towards higher intensity. This suggests that the distribution of the cells are not solely caused by uncertainty of the intensity reads but are caused by specifically increased GFP signal. This indicates the process of the sorting causes the cells to increase the GFP signal, perhaps caused by increased stress response and expression of *sf*-GFP.

Out of the 21 sequenced variants, there was no mutations that caused premature termination of the protein, most of the sequences were wild-type. Except the wild-type 3 different variants were found. A contributing reason why so few variants were found might be that when the sorted cells were plated on LB agar, there was so much growth that single colonies were not distinguishable and the bacteria had to be re-streaked why the variation might have decreased. It could also that the gate used to sort the cells is too generous. Sorting for example the 1% highest RFP:GFP, or even less, instead of 3% could increase the chance of finding mutated variants of the sequence. It could also be useful to analyze the non-mutated cell at the same time to distinguish the change that the mutations has caused. That would make it easier to create gates that selects only the variantion that is caused by mutations.

However, it could be interesting to investigate the stability of the found variants. The single mutation C85R that was present in three of the sequenced variants is of most interest to further investigate since multiple mutations are more likely to cause conflicting changes to the structure why single mutations are preferable. The C85 in the reconstructed sequence used in this work correspond to A77 in the sequence of icosahedra Aquifex Aeolicus Lumazine Synthase which has been structure determined by x-ray crystalization[5]. In Figure 30 A77 is marked in red. The marked amino acid is placed in the center of the monomeric unit and it is possible that a cystein in that position could cause the protein to misfold by disulfide bonding.

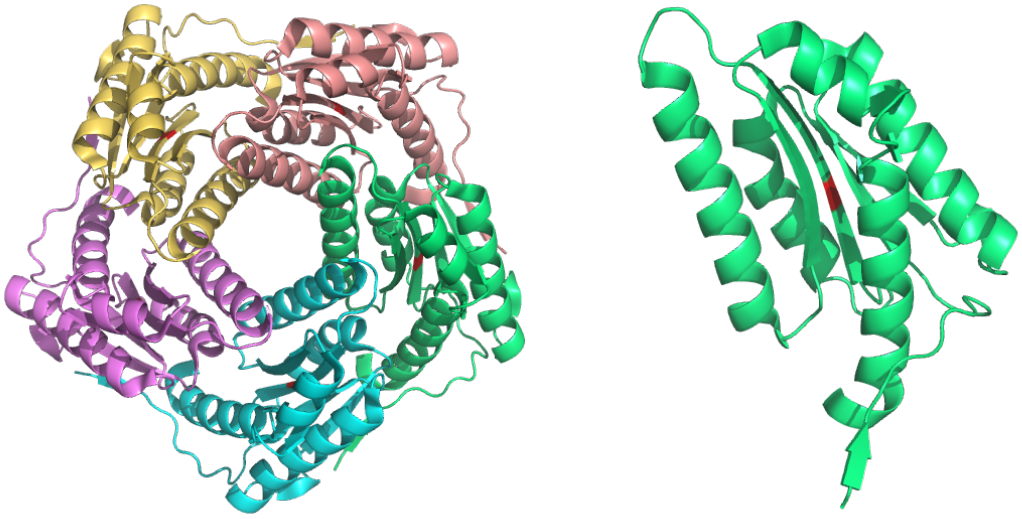


Figure 30: Structure of Aquifex Aeolicus Lumazine Synthase, PDB ID: 1HQK. A Pentameric and monomeric unit of LS with A77 marked in red.

Libraries should also be created with the 3 epPCR conditions that was not yet transformed. Analysis of these libraries should be similarly performed and sorted for increased RFP:GFP ratio to further improve the likeliness of finding stabilizing mutations to the protein sequence.

8 Conclusion

Icosahedra Lumazine Synthase from reconstructed sequence was not possible to express solubly with N-terminal histag. The protein could be extracted from inclusion bodies but refolding into native structure was not successful. Lumazine Synthase with N-terminal *tag*RFP was successfully cloned into the plasmid containing the stability assay system. Error prone PCR generated correctly sized fragments of the Lumazine Synthase sequence and one of the epPCR conditions products was used to incorporate the mutations into the plasmid containing the stability assay system.

LStagRFP was also expressed and purified by ammonium sulphate precipitation and size exclusion chromatography. The purified protein was used to investigate the oligomeric state. However, it seemed that the protein was not stable enough to determine the oliogmerization.

8.1 Further investigation

Further work should be performed to optimize and use the Megaprimer Whole Plasmid PCR with the additional epPCR products to generate plasmids containing the mutations introduced by epPCR. The resulting plasmids should also be transformed into E. coli to generate libraries of variants of the Lumazine Synthase sequence. Obtaining nicked plasmid rather than linear plasmid by using lower concentration of megaprimer to avoid non-specific amplification should increase the yield of the transformation. It should then be possible to generate libraries of more than 4000 variants, which is the aim. The obtained libraries should then be sorted with FACS similarly to what was done with the library that was transformed in this work. Further investigation of the sequences of the sorted cells should give indications of mutations that are acting stabilizing on the structure of the protein. This should be possible to use to find a more stably expressing variant of the Lumazine Synthase.

Once a solubly expressing variant of the protein is found it should be expressed and the oligomeric state investigated.

Acknowledgements

Firstly, I would like to thank Ingemar André for giving me the opportunity to work with this project and for helpful supervision throughout the project.

I would also like to give a big thanks to Signe Christensen for a lot of help in the lab and with the project. I want to also thank the rest of the André group for a lot of interesting discussions and all the people at the Center for Molecular Protein Science.

References

[1] Frank J. Poelwijk, Daniel J. Kiviet, Daniel M. Weinreich, and Sander J. Tans. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature*, 445(7126):383–386, jan 2007.

[2] Emmanuel D. Levy, Elisabetta Boeri Erba, Carol V. Robinson, and Sarah A. Teichmann. Assembly reflects evolution of protein complexes. *Nature*, 453(7199):1262–1265, jun 2008.

[3] María Silvina Fornasari, Diego A Laplagne, Nicolás Frankel, Ana A Cauerhff, Fernando A Goldbaum, and Julián Echave. Sequence Determinants of Quaternary Structure in Lumazine Synthase. *Molecular Biology and Evolution*, 21(1):97–107, 2004.

[4] Fischer M. Cushman M. Bacher A. Ladenstein R. Meining W., Mörtl S. The atomic structure of pentameric lumazine synthase from saccharomyces cerevisiae at 1.85 Å resolution reveals the binding mode of a phosphonate intermediate analogue11edited by r. huber. *Journal of Molecular Biology*, 299(1):181 – 197, 2000.

[5] Fischer M. Bacher A. Ladenstein R. Zhang X., Meining W. X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile aquifex aeolicus at 1.6 Å resolution: determinants of thermostability revealed from structural comparisons11edited by r. huber. *Journal of Molecular Biology*, 306(5):1099 – 1114, 2001.

[6] Bacher A. Ladenstein R., Fischer M. The lumazine synthase/riboflavin synthase complex: shapes and functions of a highly variable enzyme system. *FEBS Journal*, 280(11):2537 – 2563, 2013.

[7] Michael J Harms and Joseph W Thornton. Analyzing protein structure and function using ancestral gene reconstruction. *Current Opinion in Structural Biology*, 20(3):360–366, jun 2010.

[8] Philip A. Romero and Frances H. Arnold. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology*, 10(12):866–876, dec 2009.

[9] Kentaro Miyazaki. Creating Random Mutagenesis Libraries by Megaprimer PCR of Whole Plasmid (MEGAWHOP). In *Directed Evolution Library Creation*, pages 23–28. Humana Press, New Jersey, 2003.

[10] David S. Wilson and Anthony D. Keefe. Random Mutagenesis by PCR. In *Current Protocols in Molecular Biology*, volume 51, pages 8.3.1–8.3.9. John Wiley & Sons, Inc., Hoboken, NJ, USA, may 2001.

[11] Ira Palmer and Paul T Wingfield. Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from Escherichia coli.

[12] Bukau B. Arsène F., Tomoyasu T. The heat shock response of Escherichia coli. *International Journal of Food Microbiology*, 55(1-3):3–9, 2000.

[13] Giulia Calloni, Taotao Chen, Sonya M Schermann, Hung Chun Chang, Pierre Genevaux, Federico Agostini, Gian Gaetano Tartaglia, Manajit Hayer-Hartl, and F Ulrich Hartl. DnaK Functions as a Central Hub in the E. coli Chaperone Network. *Cell Reports*, 1(3):251–264, 2012.

[14] Matthias P. Mayer. Intra-molecular pathways of allosteric control in Hsp70s, jun 2018.

[15] François Baneyx and Mirna Mujacic. Recombinant protein folding and misfolding in Escherichia coli. *Nature Biotechnology*, 22(11):1399–1408, nov 2004.

[16] Thomas Hesterkamp and Bernd Bukau. Role of the DnaK and HscA homologs of Hsp70 chaperones in protein folding in E.coli. Technical Report 16, 1998.

[17] K. Miyazaki. MEGAWHOP cloning: A method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids. *Methods in Enzymology*, 498:399–406, 2011.

[18] Charles Tanford. Protein Denaturation. pages 121–282. 1968.

Appendix I

Sequence for LStagRFP construct

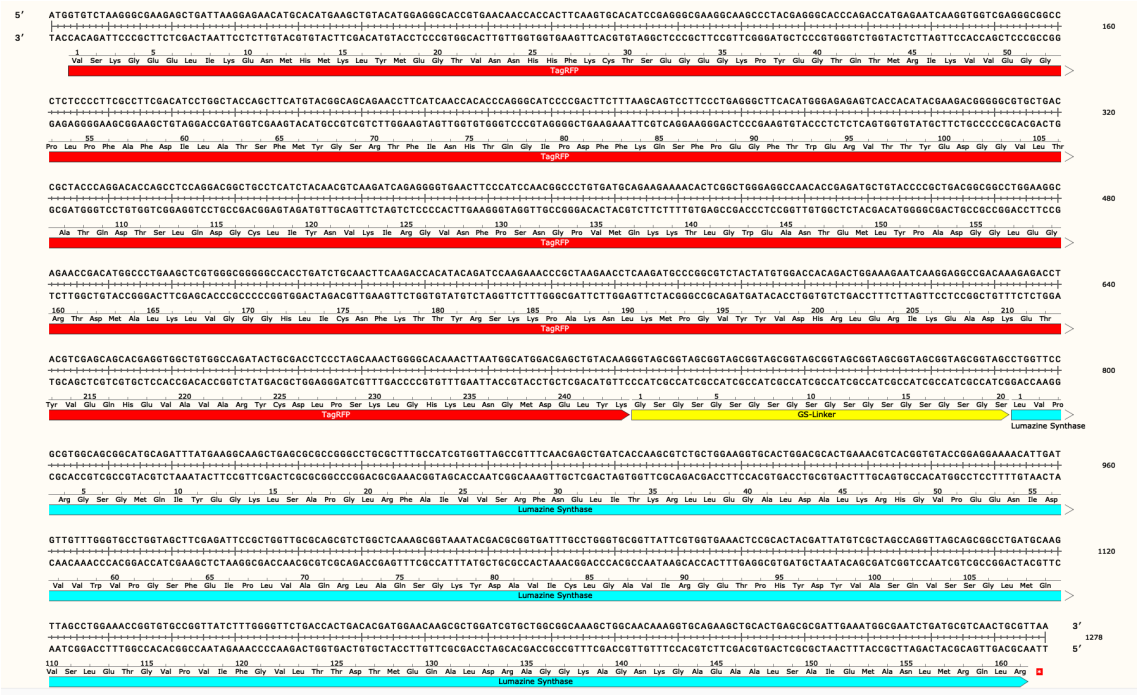


Figure 31: Sequence of construct of Lumazine Synthase fused with reporter *tag*RFP. Figure created with SnapGene Viewer.

Appendix II

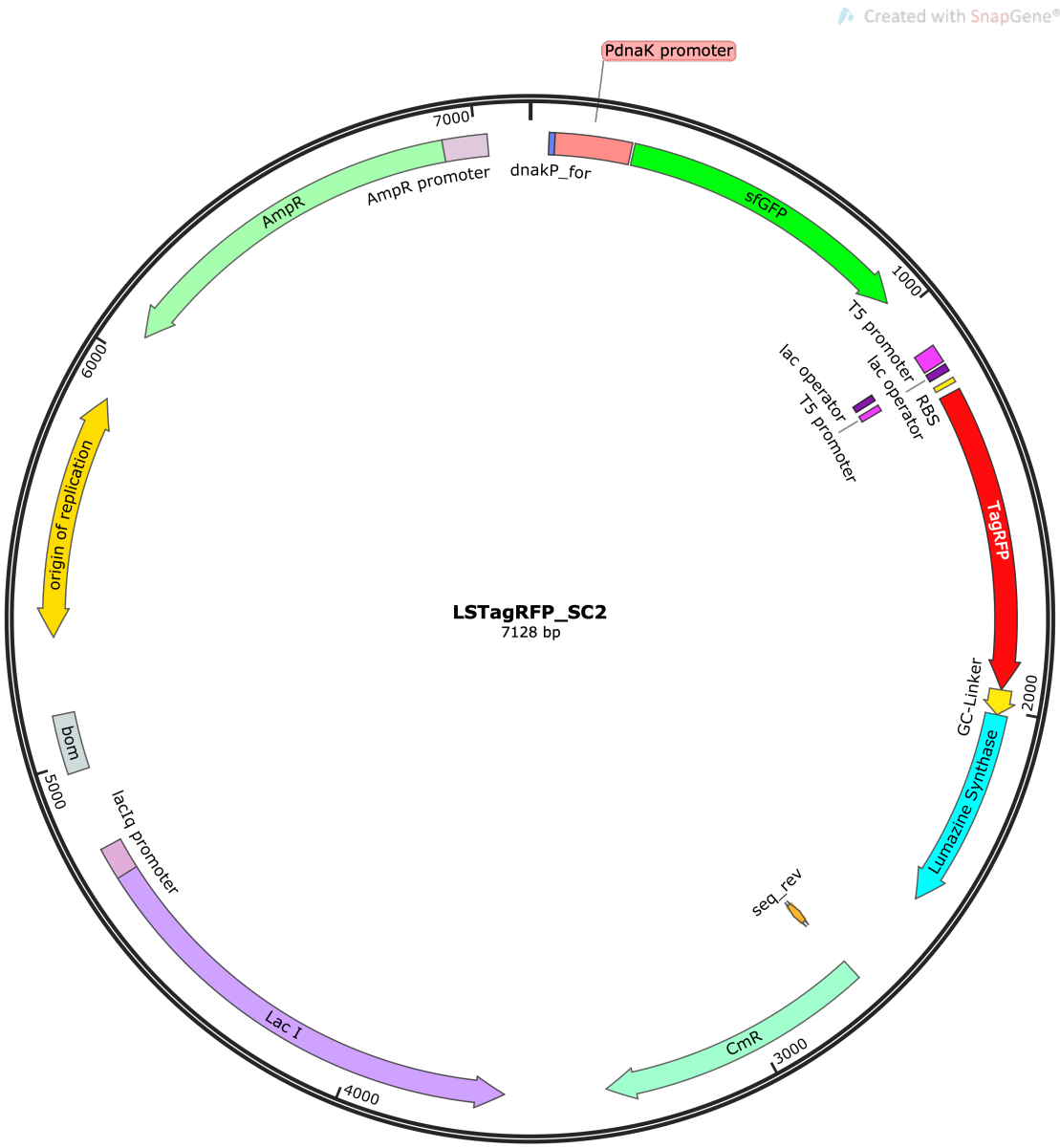


Figure 32: Complete plasmid map of vector with inserted TagRFP-Lumazine Synthase. Figure created with SnapGene Viewer.

Appendix III

Sequence for epPCR forward primer, CO_FWD_1:
GACGAGCTGTACAAGGGTAG
Sequence for epPCR reverse primer, CO_REV_1:
GCTTTTAACGCAGTTGACGC

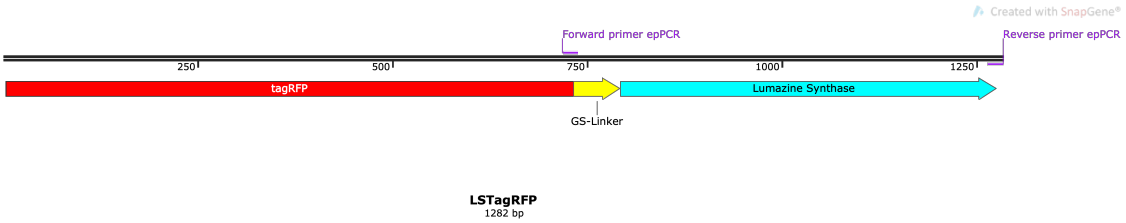


Figure 33: DNA segment containing TagRFP-Lumazine Synthase. Primers used for error-prone PCR are indicated in purple labels. Figure created with SnapGene Viewer.

Appendix IV

DLS measurement Size Distribution Report by Intensity of LStagRFP purified by ammonium sulphate precipitation, Figure 34, and by size exclusion chromatography, Figure 35.

Size Distribution Report by Intensity
v2.2

Malvern

Sample Details

Sample Name: 20190311_LS_300 1
SOP Name: CCMV_measurment.sop
General Notes:

File Name:	Lumasine synthase.dts	**Dispersant Name:**	Water
Record Number:	3	**Dispersant RI:**	1,330
Material RI:	1,45	**Viscosity (cP):**	1,0031
Material Absorbtion:	0,001	**Measurement Date and Time:**	Monday, 11 March 2019 15:3...

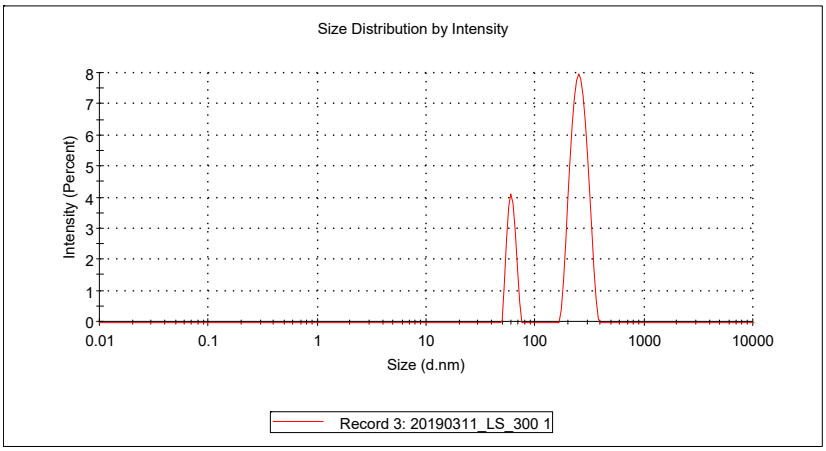
System

Temperature (°C):	20,0	**Duration Used (s):**	30
Count Rate (kcps):	213,0	**Measurement Position (mm):**	0,85
Cell Description:	Disposable sizing cuvette	**Attenuator:**	5

Results

			Size (d.nm):	% Intensity:	St Dev (d.n...
Z-Average (d.nm):	169,9	**Peak 1:**	253,2	79,0	41,40
PdI:	0,409	**Peak 2:**	60,09	21,0	4,989
Intercept:	0,947	**Peak 3:**	0,000	0,0	0,000

Result quality : Good



Malvern Panalytical
www.malvernpanalytical.com

Zetasizer Ver. 7.13
Serial Number : MAL1015162

File name: Lumasine synthase
Record Number: 3
12 Mar 2019 16:10:05

Figure 34: DLS Size Distribution Report by Intensity measurement of LStagRFP purified with ammonium sulphate precipitation by Malvern Panalytical Zetasizer

Size Distribution Report by Intensity

v2.2

Malvern

Sample Details

Sample Name: Lumasine synthase fraction 14 1

SOP Name: CCMV_measurment.sop

General Notes:

File Name:	Lumasine synthase.dts	**Dispersant Name:**	Water
Record Number:	1	**Dispersant RI:**	1,330
Material RI:	1,45	**Viscosity (cP):**	1,0031
Material Absorbtion:	0,001	**Measurement Date and Time:**	Tuesday, 5 March 2019 14:19...

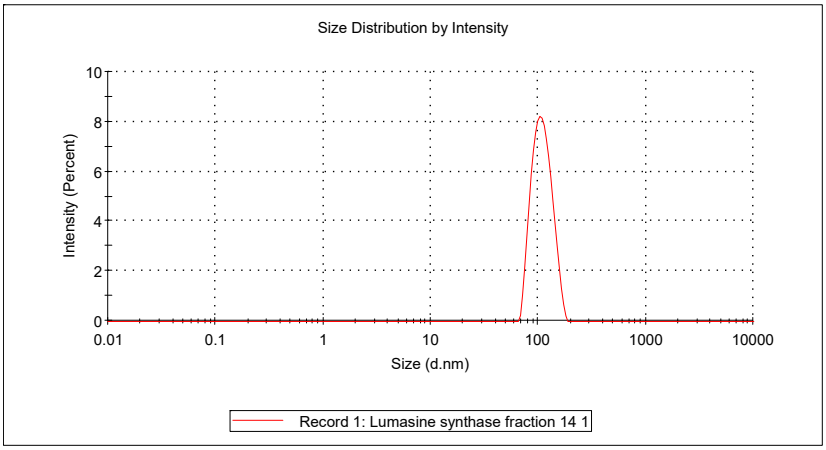
System

Temperature (°C):	20,0	**Duration Used (s):**	30
Count Rate (kcps):	176,2	**Measurement Position (mm):**	4,65
Cell Description:	Disposable sizing cuvette	**Attenuator:**	7

Results

			Size (d.nm):	% Intensity:	St Dev (d.n...
Z-Average (d.nm):	110,6	**Peak 1:**	109,5	100,0	22,33
PdI:	0,120	**Peak 2:**	0,000	0,0	0,000
Intercept:	0,957	**Peak 3:**	0,000	0,0	0,000

Result quality : Good



Malvern Panalytical
www.malvernpanalytical.com

Zetasizer Ver. 7.13
Serial Number : MAL1015162

File name: Lumasine synthase
Record Number: 1
12 Mar 2019 16:04:23

Figure 35: DLS Size Distribution Report by Intensity measurement of LStagRFP purified with size exclusion chromatography by Malvern Panalytical Zetasizer