KTH ROYAL INSTITUTE
OF TECHNOLOGY

Degree Project in Medical Biotechnology

Second cycle 30 credit

Sortase A coupling of the recombinant partial silk proteins 4Rep-Srt and G-/G5-CT to understand the structure of silk fiber

ELLEN ADLERZ

Department of Protein Science
School of Engineering Sciences in Chemistry, Biotechnology and Health (CBH)
KTH Royal Institute of Technology, Stockholm Sweden

Main supervisor: My Hedhammar
Supervisors: Ronnie Jansson and Danilo Kimio Hirabae De Oliveira
Examiner: Peter Savolainen

Stockholm, Sweden 2023

Table of contents

Abstract

Spider silk is a material of interest due its biocompatibility and therefore usage as a biomaterial. Its comparability to man-made materials in terms of strength and elasticity, along its biocompatibility, makes it desirable in the medical field. Dragline silk is one of seven types of silk made by orb-web-weaving spiders which is used as a lifeline in case of predators to escape and is therefore very strong and extensible. The dragline silk is composed of two proteins, called major ampullate spidroin silk protein 1 and 2 (MaSp1 & 2) since it is made in the major ampullate gland in the spider. These two proteins are in turn composed of three parts, a repetitive region in the middle, and two nonrepetitive regions at the terminals, N- and C-terminal domain. Having spiders as main producers of silk when conducting research come with difficulties and have made researchers turn to recombinant production for expression, mostly in *E. coli*. However, the protein size is a limitation when expressing in *E. coli* so researchers has come up with a smaller protein than MaSp made up of only the repetitive region and the C-terminal domain, called 4RepCT. 4RepCT is still able to self-assemble into fibers under physiological-like conditions and is biocompatible. 4RepCT can be functionalized with other biomolecules that can alter its function, e.g., a cell-adhesion motif from fibronectin, allowing 4RepCT to bind to cells. In this project we aim to produce 4Rep-Srt and G-/G5-CT separately before coupling them with enzyme Sortase A. Sortase A recognizes a Sortase tag (LPXTG) on the C-terminus of one protein (4Rep-Srt in this project) and connects it to another protein that has 1-5 glycine's on the N-terminus (G-/G5-CT in this project). When producing the proteins separately, we can express G-/G5-CT with ^{13}C and ^{15}N before coupling and then know which part of the 4Rep-G-/G5-CT protein is G-/G5-CT. By knowing the structure of the protein, protein features and functions can be discovered to better aid the engineering of proteins to get biomolecules with wanted functions.

Keywords

4RepCT, 4Rep-Srt, G-/G5-CT , Sortase A coupling, recombinant silk protein

Sammanfattning

Spindelsilke är ett intressant material på grund av dess biokompatibilitet och därav användning som biomaterial. Det är jämförbart med konstgjorda material när det gäller styrka och elasticitet, och i kombination med biokompatibiliteten, är spindelsilke eftertraktat inom det medicinska området. Sju olika spindeltrådar produceras av hjulspindeln och en kan användas som säkerhetslina vid flykt, och är därav mycket stark och töjbar. Säkerhetslinan består av två proteiner, så kallade stora ampullkörtel spidroiner 1 och 2 (MaSp1 & 2) eftersom de produceras i den stora ampullkörteln i spindeln. Dessa två proteiner är i sin tur sammansatta av tre delar, en repetitiv region i mitten och två icke-repetitiva regioner vid proteinets terminaler, N- och C-terminal domänerna. Att använda spindlar som huvudsakliga producenter av spindeltråd inom forskning är problematiskt och har fått forskare att vända sig till rekombinant produktion för uttryck av proteiner, mestadels i *E. coli*. Proteinstorleken är en begränsning vid produktion i *E. coli*, så forskare har uppfunnit ett mindre protein än MaSp som endast består av en mindre del av den repetitiva regionen och C-terminal domänen, så kallad 4RepCT. 4RepCT kan fortfarande bilda spindeltråd under fysiologiskt tillstånd och är biokompatibel. 4RepCT kan funktionaliseras med biomolekyler som kan ändra dess funktion, till exempel genom ett celladhesions motiv från fibronektin, som gör att 4RepCT kan binda till celler. I detta projekt siktar vi på att producera 4Rep-Srt och G-/G5-CT separat innan de kopplas med enzymet Sortase A. Sortase A känner igen en Sortase-tagg (LPXTG) på C-terminalen av ett protein (4Rep-Srt i detta projekt) och kopplar det till annat protein som har 1-5 glyciner på N-terminalen (G-/G5-CT i detta projekt). När proteinerna produceras separat kan G-/G5-CT uttryckas med ^{13}C och ^{15}N före kopplingen, och ge information om vilken del av 4Rep-G-/G5-CT-proteinet som är G-/G5-CT. Genom att känna till proteinets struktur, kan även dess egenskaper och funktioner förstås för att bättre bistå konstruktionen av proteiner för att erhålla biomolekyler med önskade funktioner.

Nyckelord

4RepCT, 4Rep-Srt, G-/G5-CT, Sortas A koppling, rekombinant silkesprotein

1. Introduction

1.1 Spider silk

Spider silk is a naturally strong, elastic, and biocompatible biomaterial made by spiders. In comparison to man-made materials, silk has a tensile strength similar to steel and is almost as elastic as rubber on a weight-to-weight basis. It's also tougher than Nylon and Kevlar, which are synthetic fibers. [1] Its biocompatibility makes it possible to use as biomaterial in medical applications [2]. One study on open wounds on pigs showed that compared to already known wound dressings, spider silks performed equally as good, with visual rapid healing of the wound and no difference between the dressings in the histopathological images taken after 14 days. The study did however show that polyurethane films was most rapid in the wound healing process. [3]

Silk produced by silkworms, mostly *Bombyx mori*, is a biomaterial that has been known and used for centuries as suture materials. The production of silkworm silk is highly cost-efficient because it has been optimized through its many years of usage. The disadvantage with silkworm silk is a sericin-protein that coats the fibers which can cause hypersensitive reactions due to its high immunogenicity. Fibers from spiders on the other hand don't have this coating and don't cause inflammation or allergic reaction in humans and it is known to be compatible as implantation in humans. On a mechanical level, spider silk is stronger and more extendable, and while spider silk combines strongness and elasticity, silkworm silk can only be one or the other. [1][2]

The orb-web-weaving araneid spiders can produce up to seven different types of silk fibers depending on its function [4][5]. Dragline silk is one type of silk that is used as the spider's lifeline in case of predators, and it's called dragline silk since it's always dragged. The fiber is very strong and extensible which makes its toughness comparable to known commercial polyaramid filaments used in e.g., radial tires and bulletproof clothing. The dragline silk is made in the major ampullate (MA) gland, along with the web frame (**Figure 1**), which has a long tail, a sac and a duct. The tail stores the protein solution for fiber formation and making the fiber dissolvable in water. The sac is the main storage of spidroins that leads to the tapering duct through a funnel. The fiber formation mostly takes place in the duct even tough further processing is needed before it's completely done. [1][5-7] As the spidroin go through the narrowing duct, the sheer force in combination with pH gradient and ion concentration gradients makes the proteins align and fibers form [7].

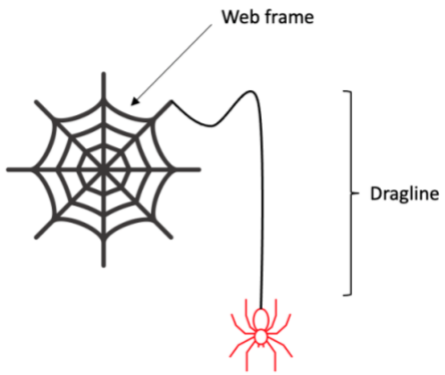


Figure 1. MA silk in the form of web frame and dragline silk.

The dragline fiber is composed of two silk proteins of the molecular mass around 300 kDa, called major ampullate spidroin silk protein 1 and 2 (MaSp1 respectively MaSp2). They are composed of three parts; first a nonrepetitive N-terminal (NT) domain, second a repetitive region that consists of poly-alanine's and glycine rich blocks, and thirdly a nonrepetitive C-terminal (CT) domain. The alanine's in the spidroins give the fiber strength while the glycine's give elasticity. The N-terminal domain is highly conserved between spiders, with an important role of conferring solubility at neutral pH and facilitating connecting of spidroins at low pH. [8] The N-terminal domain will dimerize at low pH [9] while the C-terminal domain form intermolecular disulfide bonds and under oxidizing conditions, can stabilize dimers and multimers. Hence, the terminals are key for the silk protein to transform into fibers [1][10].

1.2 Recombinant 4RepCT silk

One problem with spiders is that they are cannibalistic and territorial by nature which makes it difficult and expensive to rely on spiders to produce high amounts of fibers [11]. The solution to this problem is to use recombinant technology for expressing silk proteins. *Escherichia coli* is the common choice of host for recombinant proteins due their inexpensiveness, fast growth kinetics and the fact that cultures with high cell density is easily attained. *E. coli* can grow in rich complex mediums that are inexpensive and readily available to make, and they are easily genetically modified. [12] The full-length spidroins (MaSp1 and MaSp2) are too large to be efficiently recombinantly expressed so combinations of the N-terminal and/or C-terminal domain and parts of the repetitive regions are often combined to be expressed recombinantly [7]. Another advantage to recombinantly producing silk is the opportunity to functionalize the silk, meaning adding other molecules that can alter the protein [13].

4RepCT (repetitive region of 4 poly-Ala and Gly rich blocks together with CT), **Figure 2**, is one of those combinations that has been developed to create fibers [7]. 4RepCT in combination with solubility tag, allows for high yield after purification and for the protein to be soluble in *E. coli* and during purification [14]. The protein forms macroscopic fibers spontaneously under physiological conditions (ambient temperature and pressure) [2][14]. 4RepC is therefore a biocompatible as implantable biomaterial, e.g., in tissue engineering.

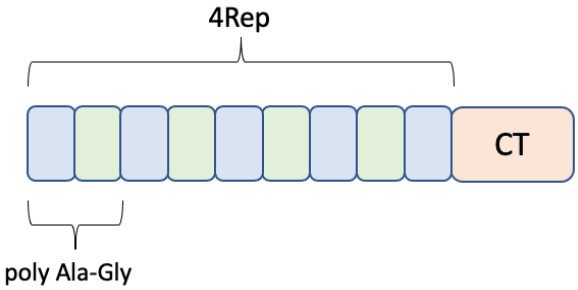


Figure 2. 4RepCT. Recombinantly produced protein 4RepCT with the ability to form silk fibers. The protein consists of one repetitive region with alanine's and glycine's, and one C-terminal domain.

One study on rats showed that subcutaneously implantation of fibers from 4RepCT showed a tissue response in the form of macrophages and multinucleated cells, similar to the response of commercialized Mersilk™. The 4RepCT also aided ingrowth of fibroblasts and vasculogenesis. [2] The silk fibers were able to support growth of HEK cells and attach to them in another study, showing once again its ability to be biocompatible [14].

1.3 Functionalization of silk materials

As mentioned, silk materials are as of right now highly interesting due to its ability for biomedical applications. There has been extensive work on silk to develop different kinds of materials up to 3 dimensions, particles, fibers, films, foams, gels, capsules, and microspheres. Microspheres are produced by mixing silk proteins with high concentrations of salts and films are produces by spin/dip-coating/casting. [13]

One way that 4RepCT has been functionalized is by adding a motif from the extracellular matrix (ECM) protein fibronectin (FN), which allows for cells adhesion and results in FN-silk. Through FN-silk, one can get 3D scaffolds with high biological relevance. Cell cultures have been studied since the 1940's but they are mostly executed on hard plastics or glass surfaces which results in half of the cells working on surface adhesion, which in the body is not occurring. By not being able to receive signals from neighboring cells in all 3 dimensions, the cells will not behave as they would in their in vivo environment in terms of metabolism and functionality, and the results are therefore not very relevant in terms of the cells natural behavior. By adding cells to silk prior to assembly, the cells get spread out and integrated in the microfibers which allows them to be highly proliferative and viable for long term culture. The fibers that are formed with cells have similar mechanical properties to artery walls. [15]

Another way to functionalize silk is by the addition of affinity domains. The recombinantly 4RepCT protein has been fused together with affinity domains on a gene level. The four domains used in the study were two IgG-binding domains, an albumin-binding domain ABD and a biotin-binding domain and they all have biotechnological use. The functionalized silk was produced successfully, and the resulting fibers and films were chemically and thermally stable. Affinity domains needs to be folded correctly and accessible to their target for the binding to work in complex samples and all four fusion proteins achieved this. [16]

1.4 Sortase A

Sortase A (SrtA) is an enzyme of the Sortase transpeptidase family, and this isoform is preserved across all gram-positive bacteria. SrtA, so called "the housekeeping sortase", recognizes the sequence LPXTG located on the C-terminal of a protein and anchors it to lipid II through peptide bonds. SrtA can also anchor the donor substrate to the acceptor if it in the N-terminus contains Glycine(s), even as little as one G, but remarkable specificity towards G5, see **Figure 3**. [17] Furthermore, SrtA has also been shown to be able to conjugate substrates, e.g., aminoglycine-derivatized small molecules, to the LPXTG of a recombinant protein, which can be used for conjugation of peptides with different biological properties [18]. SrtA has also been used to conjugate FN-silk to biofilm-dispersal enzymes that has the Srt-tag to inhibit adhesion of bacteria to coatings. This application can be used be on implantation where biofilms are concerning. [19]

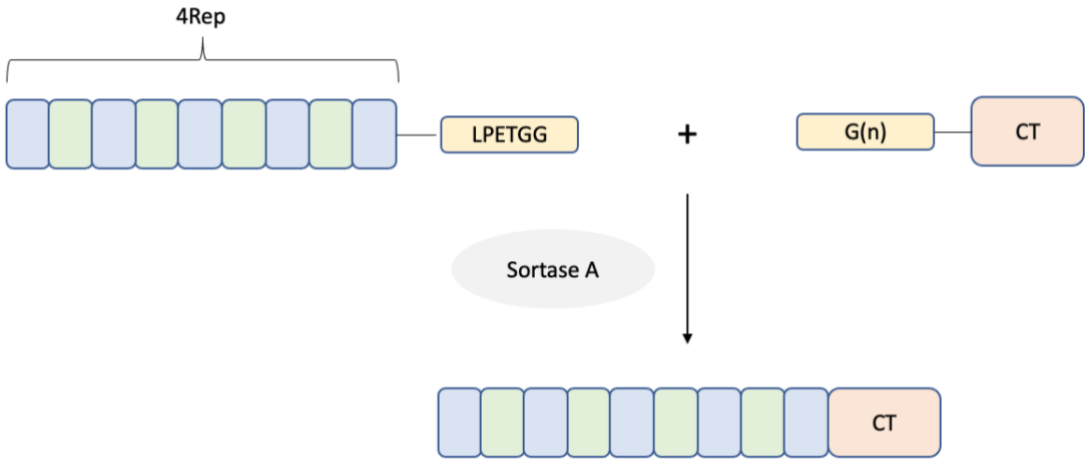


Figure 3. Sortase A coupling of 4Rep-LPETGG and G(n)-CT, where n can be 1-5 glycine's. LPETGG is the sortase-tag that Sortase has affinity towards.

1.5 Aim of project

This project is a part of a bigger project that is trying to understand the structure of the 4RepCT fibers. Determination of protein structure is important when wanting to learn about a protein's features and functions. It allows scientists to know details on molecular level to understand how proteins interact with each other and other molecules, such as inhibitors, substrates, etc., which could be important for instance in drug research. [20, 21] Nuclear magnetic resonance (NMR) spectroscopy is one of many structural determination methods that allows proteins to be in solution during analysis [22]. By labelling 4RepCT with ^{13}C and ^{15}N during expression, prior to NMR analysis, the resulting structure will show the protein as whole without knowing which part is which. The working hypothesis of this project is that 4Rep-Srt and G-/G5-CT will be expressed separately and then coupled by Sortase A which allows labelling of only G-/G5-CT before NMR and therefore information on which part of the protein is CT and which is 4Rep-Srt. This will overall give information on the structure of CT and possible solutions to making engineered fibers in the future.

2. Materials and Methods

2.1 Expression of G-CT

2.1.1 Overnight pre-culture

An overnight (O/N) pre-culture was prepared by transferring 100 ml TSB+Y (30 g Tryptic soy broth and 5 g Yeast extract in deionized water) and 100 µl Kanamycin (C_{Final} = 50 µg/ml) into a sterile 500 ml shake flask before inoculating with a scrape of cells from a glycerol stock (*E. coli* BL21(DE3)). The cells were incubated O/N (16±1h) at 30°C on a shaking table, 200 rpm. A glycerol stock (Gly-stock) was prepared from the O/N pre-culture by transferring 35 µl 85% Sterile Glycerol (C_{Final} = 15%) and 165 µl pre-culture to a sterile Eppendorf tube. The Gly-stock was stored in -80°C.

2.1.2 G-CT cultivation and expression, and glycerol stock preparation

Expression 1 of G-CT was performed with 4 sterile 5 L flasks, with 600 ml LB media (20 g LB broth; NaCl 5 g/L, Tryptone 10 g/L, Yeast Extract 5 g/L, in 1 L Milli-Q) and 600 µl Kanamycin (C_{Final} = 50 µg/ml) was transferred. 6 ml O/N pre-culture, from section **2.1.1**, was inoculated to each flask. The cells were incubated at 30°C on a shaking table, 200 rpm. Optical density (OD) was measured at 600 nm at start of cultivation and throughout cultivation until OD_{600} reached 0.8-1. The cells were cooled to room temperature (RT), 200 rpm, for 15 min to slow down cell growth. Protein expression was induced by adding 600 µl Isopropyl β-D-1-thiogalactopyranoside (IPTG) (C_{Final} = 0.5 mM) to each flask. Cells were incubated at RT, 200 rpm for 3 h. Cells were harvested through centrifugation (4700 rpm, +4°C, 20 min) with rotor JLA 10.500. Each pellet was resuspended in 16 ml 20 mM Tris pH 8 (96 ml 20 mM Tris in total) by VX-2500 Multi Tube Vortexer (VMR Scientific Products). The resuspended pellets were pooled and divided in 6 50 ml Falcon tubes with ~16 ml in each tube and stored in -20°C.

Expression 2-5 of G-CT and preparations of Gly-stocks can be seen in **Appendix A1**.

2.2 Purification 1 of G-CT

Immobilized Metal Ion Affinity Chromatography (IMAC) was used to purify G-CT from **2.1**. Two HiTrap Chelating Sepharose Fast Flow columns (GE Healthcare, Cytiva) charged with Zn^{2+} was used, 1 ml column for IMAC 1, and 5 ml column for IMAC 2. The preparation of columns is described in **Appendix A2**.

2.2.1 Lysis and centrifugation

Cells were thawed in water bath before 600 µl 50 mg/ml Lysozyme was added to each tube and incubated for 30 min. 20 µl 14.6 U/ml DNase I and 60 µl 1M $MgCl_2$ was added before another incubation for 30 min. The cells were pooled before 10 ml of 4 M NaCl was added (C_{Final} = 400 mM) and put on ice. Lysed cells were centrifuged (11 500 rpm, +4°C, 30 min) with rotor JA17 and filtered (∅0.45nm) with syringe before 2 ml 1 M Imidazole (C_{Final} = 20 mM) was added and then stored on ice during the entire purification. One gel sample, 40 µl sample and 10 µl 5x gel loading dye (C_{Final} = 1x, 60 µl 85% glycerol, 25 µl 20% SDS, 10 µl 2-mercaptoethanol and 5 µl 1% Bromophenol blue), marked Lysate was prepared and stored in +4°C.

2.2.2 Purification IMAC 1 and dialysis

The 5 ml column prepared in **Appendix A2.1** was run through with 2 CV Milli-Q (2 ml/min) and equilibrated with 10 CV Binding buffer (A; 20 mM Imidazole, 20 mM Tris and 400 mM NaCl, pH 8) (2 ml/min). pH of columns was checked to 8 before starting the purification with sample. Lysed cells from **2.2.1** was loaded onto column with flow rate 1 ml/min and flow through (FT) was collected on ice. When ~ 5 ml sample was left, some Binding buffer (A) was added to sample to collect all protein from sample tube. A gel sample marked FT was prepared and stored on ice. Column was washed with stepwise increase of Imidazole from 20 mM to 75 mM with Binding buffer (A) and Elution buffer (B; 500 mM Imidazole/20 mM Tris/400 mM NaCl, pH 8) with flow rate 1 ml/min and 2 CV on each gradient and flow throughs was collected on ice. One gel sample was prepared for each gradient step and stored on ice. Fusion protein was eluted with 300 mM Imidazole (60% Elution buffer (B)) at 1 ml/min and collected as 1 ml fractions. Gel sample marked Elution 1 was prepared and stored on ice. Tightly bound proteins were eluted with 2 CV 500 mM Imidazole (100% Elution buffer (B)) at 2 ml/min and collected in falcon tube on ice before washing column with 6 CV Milli-Q at 2 ml/min. Gel sample marked Elution 2 was prepared and stored on ice. All gel samples prepared were heated for ~5-10 min at +95°C. A SDS-PAGE (Biorad) was run with gel samples and 1x TGX buffer (100 ml 10x TGX SDS and 900 ml Milli-Q) for 45 min, at 180 volt and +4°C. The PageRuler (Thermo Scientific PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa) was loaded with 20% of max volume and samples were loaded with 67% of max volume.

2.3 Purification 2 of G-CT

Immobilized Metal Ion Affinity Chromatography (IMAC) was used to purify G-CT from the fifth expression, see **Appendix A1.6**. Two 5 ml HiTrap Chelating Sepharose Fast Flow columns (GE Healthcare, Cytiva) charged with Zn^{2+} for IMAC 1 and Ni^{2+} for IMAC 2 was used. The preparation of 5 ml column for IMAC 1 with Zn^{2+} was repeated as in **Appendix A2.1** but the column was stored in 20% EtOH. The preparation of a 5 ml column for IMAC 2 with Ni^{2+} was also repeated as in **Appendix A2.1** but the column was charged with 0.1 M $NiSO_4$ and stored in 20% EtOH.

2.3.1 Lysis and centrifugation

The lysis and centrifugation process from **2.2.1** was repeated but with 18.4 ml 4 M NaCl (C_{Final} = 400 mM) and 4 ml 1 M Imidazole (C_{Final} = 20 mM).

2.3.2 Purification IMAC 1 and dialysis

Purification with IMAC 1 was repeated as in **2.2.2** but the column (Zn^{2+}) was run with 6 CV Milli-Q (2 ml/min) and equilibrated with 20 CV Binding buffer (A). The stepwise washing with Imidazole was done with 3 CV on each step and the elution of fusion protein was collected as 1.5 ml fractions. Tightly bound proteins were eluted with 4 CV Elution buffer (B) before washing the column with 6 CV Milli-Q and 20% EtOH at 2 ml/min and storing it at +4°C.

Suitable fractions with fusion protein from elution 1 was pooled and absorbance was measured at 280 nm ($A_{280\ nm}$) with spectrophotometer (Eppendorf BioPhotometer)to calculate protein concentration (mg/ml) in pool with Beer-Lambert law, $A_{280\,nm} = \varepsilon_{0.1\%} * c * l$, where $\varepsilon_{0.1\%}$ is the percent extinction coefficient of protein ($\varepsilon_{0.1\%}$ (HTH-CT) = 0.616) and l is the length of the cuvette used (l = 1 cm). If the concentration was above 1 mg/ml, the pool was diluted with Third buffer (20 mM Tris/400 mM NaCl, pH 8). 0.5 mg/ml protease Thrombin ($V_{Thrombin}(\mu l) = \frac{Amount\ of\ fusion\ protein\ (mg)}{0.44}$) and 1 M DTT ($C_{Final}$ = 1 mM) was added to pool before preparing a gel sample. The pool was dialyzed O/N (16-20 h) in Dialysis buffer (20 mM Tris and 400 mM NaCl, pH 8) at +4°C with Spectra/Por®3 Dialysis Membrane Standard RC Tubing MWCO: 3.5 kDa. If the dialyzed pool had precipitated O/N, the sample was

centrifuged (4000 rcf 10 min, +4°C) before continuing process with supernatant and resuspending pellet in Third buffer. Gel samples was prepared of dialysis, supernatant and pellet. Gel samples and SDS-PAGE were performed as in **2.2.2** but with gel (Invitrogen) and 1x MES buffer (50 ml 20x MES and 950 ml Milli-Q).

2.3.3 Purification IMAC 2 and sample concentration

The 5 ml column (Ni^{2+}) was run through with 3 CV Milli-Q (2 ml/min) and equilibrated with 20 CV Third buffer at 2 ml/min before pH of column was checked to be 8. Supernatant from **2.3.2** was loaded onto column with flow rate 0.5 ml/min and flow through (FT) was collected on ice. When ~ 5 ml sample was left, some Third buffer was added to sample to collect all target protein from sample tube. A gel sample marked FT was prepared and stored on ice. Unspecifically bound target protein was eluted from column with 5 CV Binding buffer (A) at 2 ml/min. Gel sample marked Elution 1 was prepared and stored on ice. Tag and thrombin protease was eluted with 6 CV Elution buffer (B) at 2 ml/min. Gel sample marked Elution 2 was prepared and stored on ice.

The protein concentration was determined in the same way as in **2.3.2** but with $\varepsilon_{0.1\%}$ (G-CT) = 0.149 for target protein. Sample was concentrated with Amicon® - 4/15 Centrifugal Filters (3,000 NMWL) (4200 rcf, 10-20 min, +4°C) until concentration reached 1.5-3 mg/ml. Gel sample marked Concentrated 4Rep-Srt was prepared and stored on ice Sample was stored in -80°C. Gel samples and SDS-PAGE were performed as in **2.3.2**.

2.4 Expression 1 of 4Rep-Srt

An O/N pre-culture was prepared as in **2.1.1** with a scrape of cells from Gly-stock (*E. coli* BL21 4Rep-Srt). The expression was repeated as in **2.1.2**, with 6 sterile 2 L flasks instead of 4 5 L flasks. Each 2 L flasks contained 500 ml LB media, 500 µl Kanamycin (C_{Final} = 50 µg/ml), was inoculated with 10 ml O/N culture, and was induced with 500 µl IPTG (C_{Final} = 0.5 mM). The protein expression was carried out for 4 h before harvesting (4750 rcf, +4°C, 20 min) and resuspending cells in 144 ml 20 mM Tris (pH 8) in total.

2.5 Expression 2 of 4Rep-Srt

The O/N pre-culture and expression of 4Rep-Srt was repeated as in **2.4** with Gly-stock prepared **2.4**. with 4 sterile 5 L flasks with 850 ml LB media and 850 µl Kanamycin (C_{Final} = 50 µg/ml) used in the cultivation and 850 µl IPTG (C_{Final} = 0.5 mM) to induce.

2.6 Purification 1 of 4Rep-Srt

2.6.1 Preparing columns, lysis, and centrifugation

Two 5 ml columns, one with Zn^{2+} and with Ni^{2+}, was prepared in the same way as **2.3**. Cells from **2.4** was lysed and centrifuged in the same way as **2.2.1** but with 15 ml of 4 M NaCl (C_{Final} = 400 mM) and 3 ml 1 M Imidazole (C_{Final} = 20 mM).

2.6.2 Purification IMAC 1 and dialysis

The 5 ml column (Zn^{2+}) was run through with 2 CV Milli-Q (2 ml/min) and equilibrated with 20 CV Binding buffer (A) (2 ml/min). pH of columns was checked to 8 before starting the purification with sample. Lysed cells from **2.6.1** was loaded onto column with flow rate 1-1.5 ml/min and flow through (FT) was collected on ice. When ~ 5 ml sample was left, some Binding buffer (A) was added to sample to

collect all protein from sample tube. A gel sample marked FT was prepared and stored on ice. Wash 1 was conducted with 20 mM Imidazole with 6 CV Binding buffer (A) at flow rate 1 ml/min. A gel sample marked Wash 1 was prepared and stored on ice. Wash 2 was conducted with 40 mM Imidazole with 6 CV Binding buffer (A) and Elution buffer (B) at flow rate 1 ml/min. A gel sample marked Wash 2 was prepared and stored on ice. Fusion protein was eluted with 250 mM Imidazole (50% Elution buffer (B)) at 1 ml/min and collected as 1.5 ml fractions. Gel samples from fractions interest was prepared and stored on ice. Tightly bound proteins were eluted with 4 CV 500 mM Imidazole (100% Elution buffer (B)) at 2 ml/min and collected in falcon tube on ice before washing column with 6 CV Milli-Q and 20% EtOH at 2 ml/min. Gel sample marked Elution 2 was prepared and stored on ice.

Suitable fractions with fusion protein from elution 1 was pooled and protein concentration was determined as in **2.3.2** with percent extinction coefficient $\varepsilon_{0.1\%}$ (HTH-4Rep-Srt) = 0.848. If the concentration was above 1 mg/ml, the pool was diluted with 20 mM Tris pH 8. For the cleavage of HTH-tag, 1.8 mg/ml protease 3C (From Spiber Technologies AB) ($Amount\ of\ 3C\ needed\ (\mu g) = \frac{100\ \mu g\ 3C}{1\ mg\ fusion\ protein} * Measured\ protein\ amount \rightarrow V_{3C}(\mu l) = \frac{1\ ml * \frac{Amount\ of\ 3C\ needed}{1000}}{1.8\ mg}$) and 1 M DTT ($C_{Final}$ = 1 mM) was added to pool before preparing a gel sample. The pool was dialyzed O/N (12 h) in Dialysis buffer (20 mM Tris, pH 8) at +4°C with Spectra/Por®1 Membrane Standard MWCO: 6-8 kDa. If the dialyzed pool had precipitated O/N, the sample was centrifuged (4000 rcf 2 min, +4°C) before continuing process with supernatant and resuspending pellet in 20 mM Tris pH 8. Gel samples was prepared of dialysis, supernatant and pellet. Gel samples and SDS-PAGE were performed as in **2.3.2**.

2.6.3 Purification IMAC 2 and sample concentration

The 5 ml column (Ni^{2+}) was run through with 6 CV Milli-Q (2 ml/min) and equilibrated with 20 CV 20 mM Tris (pH 8) at 2 ml/min before pH of column was checked to be 8. Supernatant from **2.6.2** was loaded onto column with flow rate 0.5 ml/min and flow through (FT) was collected on ice. When ~ 5 ml sample was left, some 20 mM Tris was added to sample to collect all target protein. A gel sample marked FT was prepared and stored on ice. Unspecifically bound target protein was eluted from column with 5 CV Binding buffer (A) at 1-2 ml/min. Gel sample marked Elution 1 was prepared and stored on ice. Tag and 3C protease were eluted with 10 CV Elution buffer (B) at 2 ml/min before washing column with 6 CV Milli-Q and 20% EtOH at 2 ml/min. Gel sample marked Elution 2 was prepared and stored on ice.

The protein concentration was determined in the same way as in **2.6.2** but with $\varepsilon_{0.1\%}$ (4Rep-Srt) = 0.774 for target protein. Sample was concentrated with Amicon® - 15 Centrifugal Filters (3,000 NMWL) (4200 rcf 20 min, +4°C) with target concentration 3 mg/ml in mind. Gel sample marked Concentrated 4Rep-Srt was prepared and stored on ice. Sample was stored in -80°C. Gel samples and SDS-PAGE were performed as in **2.3.2**.

Second purification 4Rep-Srt is described in **Appendix A3**.

2.7 Cloning, expression and purification of G5-CT

Cloning of G5-CT is described in **Appendix A4**.

O/N culture and expression of G5-CT was repeated as **A1.6** but the O/N culture was inoculated with one colony from second transformation of clone 3 from **A4.11**.

Purification (including Strip, CiP and Charging columns Zn^{2+}/Ni^{2+}) was repeated as in **2.3** with few alterations. Protein concentration was measured with $\varepsilon_{0.1\%}$ (HTH-G5-CT) = 0.677 and the target concentration before cleavage was 0.3 mg/ml to avoid precipitation. Protease TEV ($V_{TEV}(\mu l) =$

$\frac{Amount\ of\ fusion\ protein\ (mg)}{50*Concentration\ of\ TEV\ (mg/ml)} * 1000$) was used instead of Thrombin and when estimating target protein concentration, $\varepsilon_{0.1\%}$ (G5-CT) = 0.147 was used. Due to large volume of cleaved pool sample, final sample FT was concentrated in two rounds. Gel samples and SDS-PAGE were performed as in **2.3.2**.

2.8 Sortase assay and fiber formation

2.8.1 Sortase coupling 1: G-CT and 4Rep-Srt

G-CT (**2.3**) and 4Rep-Srt (**2.6**) was coupled with Sortase A to obtain 4Rep-CT. G-CT (2.09 mg/ml) and 4Rep-Srt (0.486 mg/ml) was thawed in RT while keeping tubes still. Samples were transferred to 2 new tubes without introducing air bubbles. Tubes were centrifuged with table-top centrifuge (30-60 sec) before transferring sample to 2 new tubes while avoiding pellet and air bubbles. Samples was kept on ice from this point. 500 µM Sortase A 7M was thawed in RT and diluted to 5 µM with 369 µl 1x SrtLigBuffer (Ca^{2+} free) to a total volume of 400 µl. 28 µl 10x SrtLigBuffer (Ca^{2+} free) (C_{Final} = 1x) was mixed with 140 µl 4Rep-Srt (C_{Final} = 5 µM), 28 µl G-CT (C_{Final} = 10 µM), 16.8 µl 5 µM Sortase A 7M (C_{Final} = 0.3 µM) and filled up to 280 µl with StH_2O. Coupling mix was incubated for 1h with gel samples taken at 6 timepoints (0 min, 15 min, 30 min, 45 min, 60 min and 90 min). Gel samples and SDS-PAGE were performed as in **2.3.2**.

2.8.2 Sortase coupling 2: G-CT, G5-CT and 4Rep-Srt

Two different coupling reactions were performed, one with G-CT (2.09 mg/ml) and 4Rep-Srt (0.486 mg/ml) and one with G5-CT (0.850 mg/ml) and 4Rep-Srt (0.486 mg/ml). G-CT, G5-CT and 4Rep-Srt were thawed in RT and handled as in **2.8.1** and the concentration of G5-CT was measured as in **2.6.3**. The coupling reaction with G-CT and 4Rep-Srt was repeated as in 2.16.1 while the one with G5-CT and 4Rep-Srt included 67 µl G5-CT (C_{Final} = 10 µM) and not G-CT. Gel samples of G-CT, G5-CT and 4Rep-Srt (20 µl sample and 4 µl gel loading dye) was also prepared, as well as one for 0.3 µM Sortase A 7M (40 µl sample and 10 µl gel loading dye). Gel samples and SDS-PAGE were performed as in **2.3.2** but with loading 6 µl PageRuler and 95% of max volume for samples.

2.8.3 Sortase coupling 3: G-CT, G5-CT and 4Rep-Srt and fiber formation

Two tubes with 4Rep-Srt (4Rep-Srt 0.486 mg/ml 221021 EA) were thawed and concentrated with Amicon® - 4 Centrifugal Filters (3,000 NMWL) (4200 rcf 5 min, +4°C) before measuring protein concentration as in **2.3.2** but with $\varepsilon_{0.1\%}$ (4Rep-Srt) = 0.774. Two tubes with G5-CT (0.632 mg/ml), 1 tube with G-CT (2.09 mg/ml) and 1 tube with FN-silk (3.3 mg/ml) were thawed in RT and together with the concentrated 4Rep-Srt sample, they were handled as in **2.8.1**. 500 µM Sortase A 7M was diluted to 5 µM as in 2.16.1.

In two different tubes, 100 µl 10x SrtLigBuffer (Ca^{2+} free) (C_{Final} = 1x) was added, along with 325 µl of G-CT or G5-CT, 325 µl 4Rep-Srt, 60 µl 5 µM Sortase A 7M (C_{Final} = 0.3 µM) and filled up to 1 ml with StH_2O. Gel samples (20 µl sample and 5 µl gel loading dye) were taken at timepoint 0 min and tubes were left to incubate for 30 min in RT. 333 µl of processed FN-silk (C_{Final} = 1 mg/ml) were mixed with 667 µl 20 mM Tris buffer. The three solutions (1 ml in each) were transferred to separate glass vials and fiber formation was executed with machine provided by Spiber Technologies AB. Gel samples were taken at the start of fiber formation, after 30 min and 3 h. After 3 h, the fiber formed by FN-silk was transferred to plate filled halfway with 20 mM Tris (pH 8) for washing O/N before transferring the fiber to Falcon tube filled with 20 mM Tris (pH 8) and storing it in +4°C. The vials with 4Rep-G-CT and 4Rep-G5-CT were

left to form fiber O/N before trying to transfer the fiber to plate with 20 mM Tris (pH 8). Gel samples and SDS-PAGE were performed as in **2.8.2**.

2.8.4 Sortase coupling 4: G-CT and 4Rep-Srt and fiber formation

Procedure was repeated as in **2.8.3** with only G-CT and 4Rep-Srt. 1 tube 4Rep-Srt (1.725 mg/ml) and 2 tubes G-CT (2.09 mg/ml) was used and no concentrating samples was executed. After handling the solutions, 400 µl of 4Rep-Srt and 280 µl G-CT was transferred to tube with 100 µl 10x SrtLigBuffer (Ca^{2+} free) and 160 µl StH_2O before adding 60 µl 5 µM Sortase A 7M. 5 gel samples were prepared (10 µl sample and 5 µl gel loading dye), at start and after 30 min of coupling and after 30 min, 3 h and 7 h 20 min of fiber formations. 20 mM Tris (pH 8) was continuously added to vial due to evaporation of liquid. The fiber was stored with parafilm over vial in +4°C. Gel samples and SDS-PAGE were performed as in **2.8.2**. Scanning electron microscope (SEM) was performed on fiber. The glass was washed with Milli-Q to remove salt before drying sample. The glass was broken before covering the sample with 10 nm gold (metal evaporation, Provac PAK 600 Coating System, Germany) and taking pictures with SEM (Gemini Ultra 55, Zeiss, Germany).

3. Results

This master project was executed in order to be able to obtain the structure of CT in fiber form through NMR analysis. The proteins G-/G5-CT and 4Rep-Srt were produced separately so that only G-/G5-CT can be labelled in the future before performing NMR. After being expressed and purified (see **Figure 4** for purification process), the proteins were coupled with Sortase A before doing fiber formation.

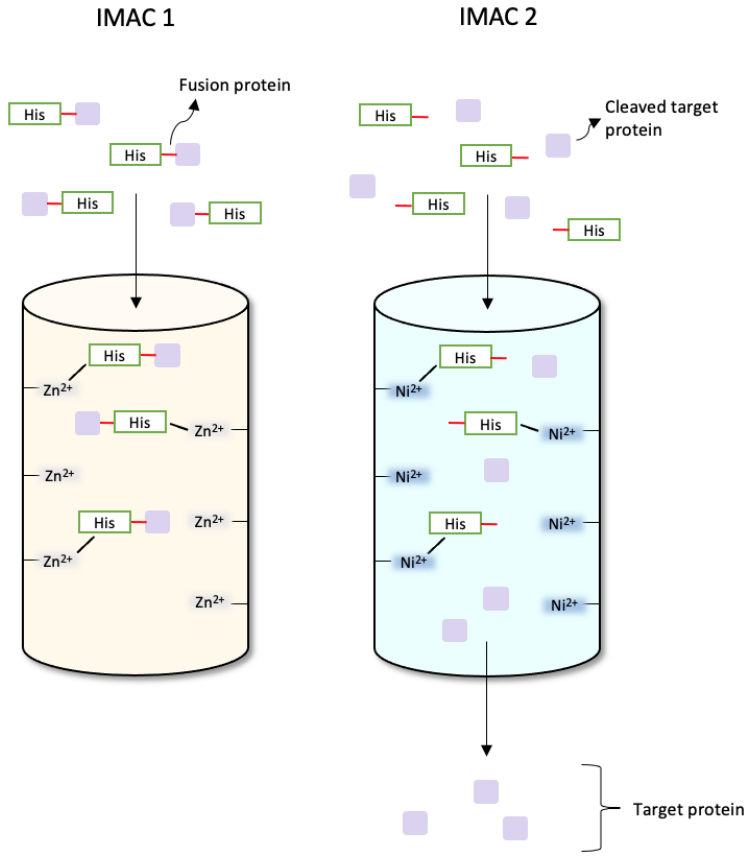


Figure 4. IMAC 1 and 2. Visualization of how target fusion protein (HTH-G-/G5-CT and HTH-4Rep-Srt) binds to column in IMAC 1 due to affinity between metal ion on the matrix and histidine (His) on the protein. After cleavage, in IMAC 2, target protein will flow through the column while the His-tag will stick to resin. The tag that is removed between IMAC 1 and 2 is Histidine$_5$-Thioredoxin-Histidine$_5$ (HTH) but is only referred to as His in this Figure.

3.1 Production of G-CT

The production of G-CT started with expression in *E. coli* (BL21), followed by purification through immobilized metal affinity chromatography.

3.1.1 Expression of G-CT

The expression of G-CT was performed five times in total due to difficulties getting the cells to reach OD_{600} = 0.8-1 within reasonable time, **Table 1**. Since the first expression reached OD_{600} only after 8 h, the second expression was performed in the same way only to rule out that it wasn't an accidental mistake causing expression 1. Even though a transformation with G-CT was performed to get a new Gly-stock before expression 3 and 4, the expression still didn't seem to work optimally. A transformation and expression were performed by Ronnie Jansson after expression 4, to try to figure out why the first four

didn't work. The fifth expression that was performed with Ronnies Gly-stock and another medium than the first four, did reach the wanted OD_{600} within 3.8 h.

The transformation that occurred between expression 2 and 3 did result in few colonies, **Figure 5**, which is consistent with how BL21 cells should grow.

Table 1. Expressions of G-CT. The absorbance at 600 nm measured for expression and their respectively times.

	OD_{600}	**Time (h)**
Expression 1	0.88	8
Expression 2	0.76	7.5
Expression 3	1.02	8
Expression 4	0.99	7.5
Expression 5	0.81	3.8

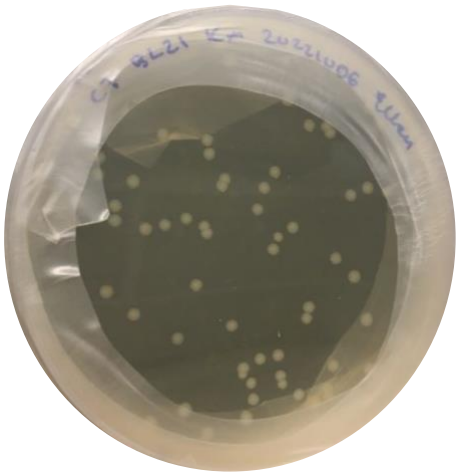
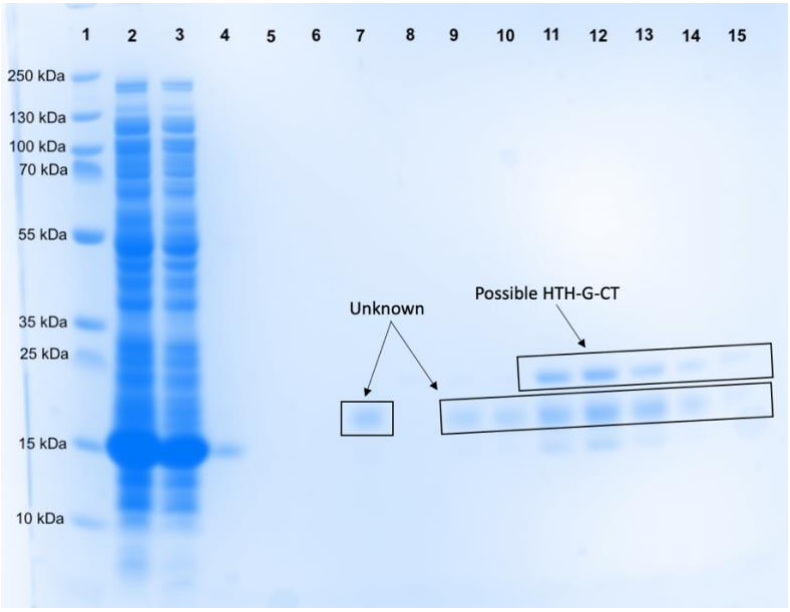


Figure 5. Transformation of G-CT plasmid with *E. coli* BL21. Few colonies which is known for BL21 cells.

3.1.2 Purification of G-CT

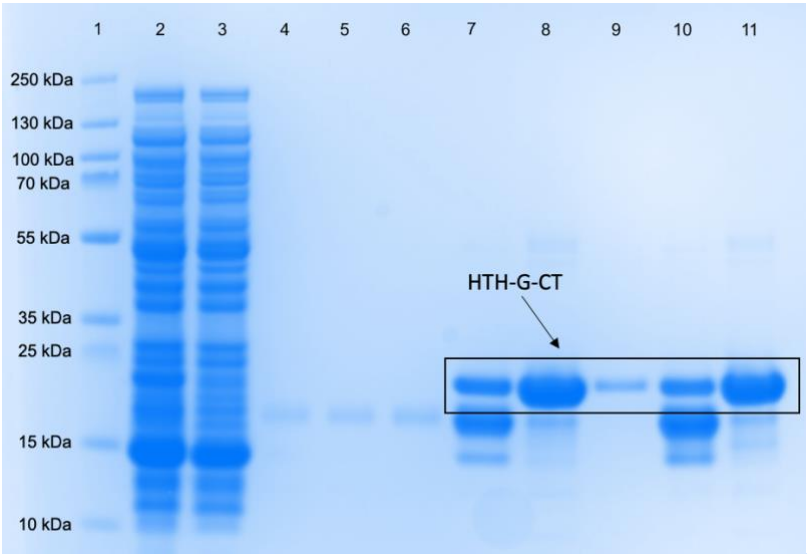
Even though the first four expressions didn't perform optimally, an attempt to purify a batch was done. There was no peak during elution 1 of fusion protein from column, so the purification was deemed unsuccessful and IMAC 2 wasn't carried out. Looking at the protein gel, **Figure 6**, there isn't a distinctive band for the target fusion protein HTH-G-CT (25.1 kDa) in the lysate, indicating that there wasn't a good expression in the first place. However, there's very faint bands for fusion protein during elution 1.



Well	Sample	Volume (µl)
1	PageRuler	3
2	Lysate	10
3	FT	10
4	Wash 1	10
5	Wash 2	10
6	Wash 3	10
7	Wash 4	10
8	Elution 2	10
9	Elution 1 Fraction 1	10
10	Elution 1 Fraction 2	10
11	Elution 1 Fraction 5	10
12	Elution 1 Fraction 6	10
13	Elution 1 Fraction 8	10
14	Elution 1 Fraction 10	10
15	Elution 1 Fraction 11	10

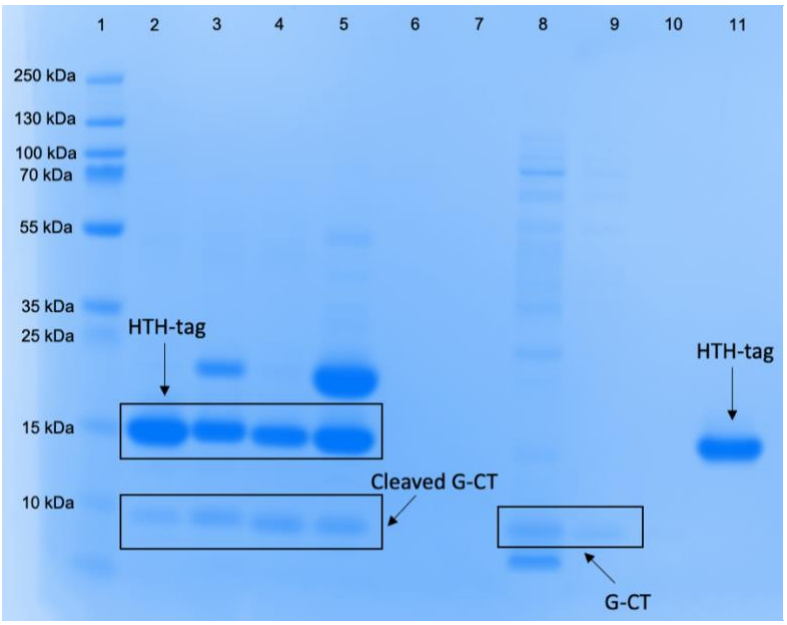
Figure 6. Gel IMAC 1 of first G-CT purification. No clear band for fusion protein HTH-G-CT (25.1 kDa) in lysate. Very faint bands in well 11-15 which could possibly be the target fusion protein being diluted. There is also an unknown protein being eluted in wash 4 and elution 1.

The second purification was more successful than the first one. There were two peaks in the chromatogram for IMAC 1, see appendix **Figure A1**. The peak in wash 4 indicates that some fusion protein HTG-G-CT is starting to elute from column already at 75 mM Imidazole, along with some other proteins which can be seen in the gel, well 7 **Figure 7**. The peak for elution 1 was sharp and 6 fractions were pooled, which shows a thick band for HTH-G-CT in well 8, **Figure 7**. Prior to cleavage and dialysis, the concentration of each sample was determined to $C_{Wash\ 4}$ = 0.644 mg/ml and C_{Pool} = 0.955 mg/ml. After dialysis, there was aggregation in Pool, because the high concentration prior to dialysis. Sample was centrifuged and IMAC 2 was proceeded with supernatant of Pool (along with dialyzed Wash 4), but as can be seen in **Figure 8**, there are some cleaved G-CT lost in the resuspended pellet. During IMAC 2 there where two small peaks for target protein, one for dialyzed Wash 4 and one for Pool. However, there's no band for cleaved G-CT in these samples, **Figure 8**, due to them being too diluted for proteins to be visible in gel. There are bands for target G-CT in final samples, although very faint, most likely due to few basic amino acids in the protein. Final concentrations were $C_{Final\ Wash\ 4}$ = 2.09 mg/ml in 800 µl and $C_{Final\ Pool}$ = 0.611 mg/ml in 600 µl. The concentrated Wash 4 sample is contaminated with other proteins, so the high concentration of 2.09 mg/ml is probably because a mix of G-CT and contaminants.



Well	Sample	Volume (µl)
1	PageRuler	4
2	Lysate	14
3	FT	14
4	Wash 1	14
5	Wash 2	14
6	Wash 3	14
7	Wash 4	14
8	Elution 1 pool	14
9	Elution 2	14
10	Wash 4 + Thrombin	14
11	Elution 1 pool + Thrombin	14

Figure 7. Gel 1 of second G-CT purification. Clear and thick bands for target fusion protein HTH-G-CT (25.1 kDa) in well 7, 8, 10 and 11, as well as a small band in Elution 2.



Well	Sample	Volume (µl)
1	PageRuler	4
2	Wash 4 after dialysis	14
3	Elution 1 pool after dialysis	14
4	Supernatant Elution 1 pool	14
5	Pellet Elution 1 pool	14
6	FT 1 with Wash 4	14
7	FT 2 with Elution 1 pool	14
8	Concentrated Wash 4 sample	14
9	Concentrated Pool sample	14
10	Elution 1	14
11	Elution 2	14

Figure 8. Gel 2 of second G-CT purification. The cleavage with Thrombin clearly worked since there are bands for HTH-tag (15.1 kDa) and G-CT (10 kDa) after dialysis. Faint bands for target protein G-CT in concentrated Wash 4 and Pool. Concentrated Wash 4 sample contains many bands other than the one for target protein, while concentrated Pool sample seems pure.

3.2 Production of 4Rep-Srt

3.2.1 Expression of 4Rep-Srt

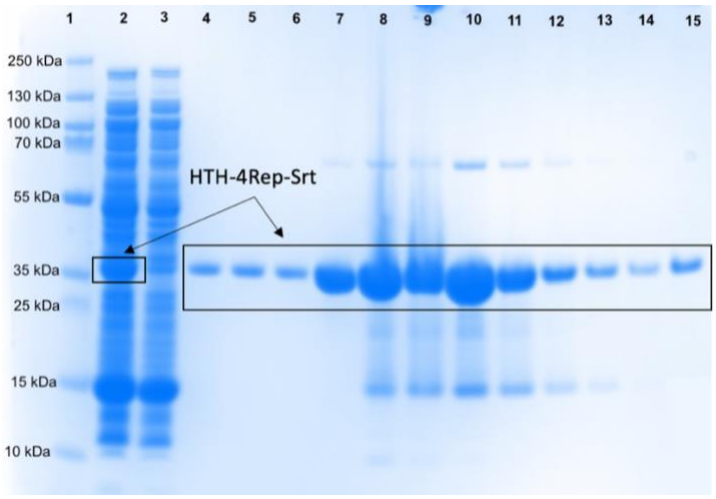
Both performed expressions of 4Rep-Srt reached wanted OD_{600} within reasonable time, 3.15 respectively 3 h, **Table 2**.

Table 2. Expressions of 4Rep-Srt. The absorbance at 600 nm measured for each expression and their respectively times.

	OD_{600}	Time (h)
Expression 1	0.89	3.15
Expression 2	0.76	3

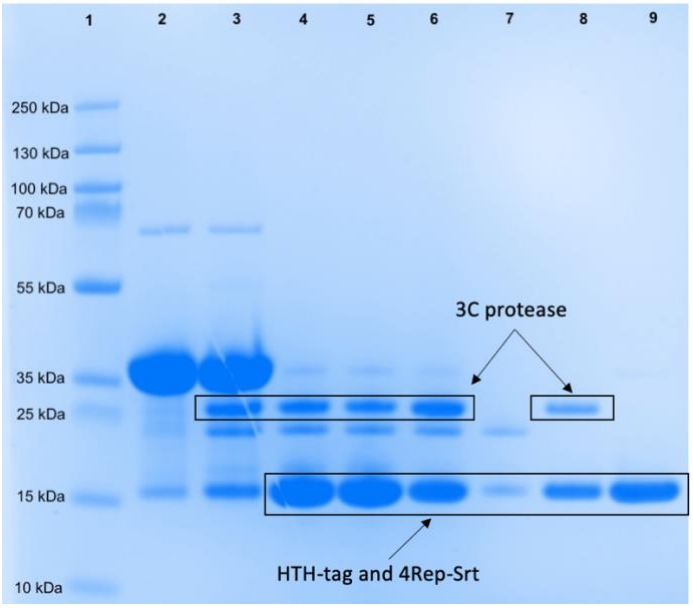
3.2.1 Purification of 4Rep-Srt

IMAC 1 of first 4Rep-Srt purification resulted in a sharp peak when eluting target fusion protein. Fractions 4-8 (**Figure 9**) were pooled before cleavage and dialysis, and as can be seen in **Figure 9** there are thick bands for fusion protein (28.79 kDa) in those fractions. However, there's also HTH-4Rep-Srt in other fractions, the washes, and elution 2. In gel 2 (**Figure 10**) almost all fusion protein has been cleaved by 3C, when comparing well 3 and 4. There was a low but noticeable peak for FT, with target protein, during IMAC 2 which is confirmed by gel 2 where the band for 4Rep-Srt is very faint in well 7. It is difficult to determine if the bands around 15 kDa is cleaved 4Rep-Srt or HTH-tag in well 7-9 since they are similar in size. The final concentrated sample had a concentration of 1.725 mg/ml in ~ 500 µl and is not pure although the band for 4Rep-Srt is thick, see **Figure 11**.



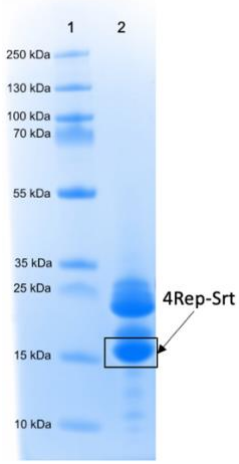
Well	Sample	Volume (µl)
1	PageRuler	3
2	Lysate	10
3	FT	10
4	Wash 1	10
5	Wash 2	10
6	Elution 1 Fraction 3	10
7	Elution 1 Fraction 4	10
8	Elution 1 Fraction 5	10
9	Elution 1 Fraction 6	10
10	Elution 1 Fraction 7	10
11	Elution 1 Fraction 8	10
12	Elution 1 Fraction 9	10
13	Elution 1 Fraction 10	10
14	Elution 1 Fraction 11	10
15	Elution 2	10

Figure 9. Gel 1 of first 4Rep-Srt purification. Clear and thick bands for target fusion protein HTH-4Rep-Srt (28.79 kDa) in well 2 and 4-15. Fractions in well 7-11 were pooled prior to cleavage and dialysis.



Well	Sample	Volume (µl)
1	PageRuler	5
2	Elution 1 pool	17
3	Elution 1 pool + 3C	17
4	Elution 1 pool after dialysis	17
5	Supernatant Elution 1 pool	17
6	Pellet Elution 1 pool	17
7	FT	17
8	Elution 1	17
9	Elution 2	17

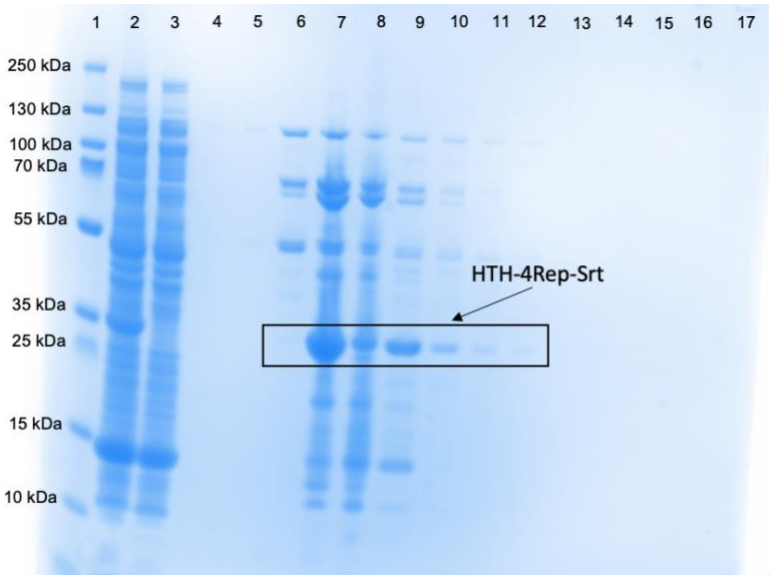
Figure 10. Gel 2 of first 4Rep-Srt purification. The cleavage with 3C clearly worked since there are bands for HTH-tag (15.32 kDa) and 4Rep-Srt (13.47 kDa) after dialysis. Faint bands for target protein G-CT in concentrated Wash 4 and Pool. Concentrated Wash 4 sample contains many bands other than the one for target protein, while concentrated Pool sample seems pure.



Well	Sample	Volume (µl)
1	PageRuler	5
2	Concentrated 4Rep-Srt sample	17

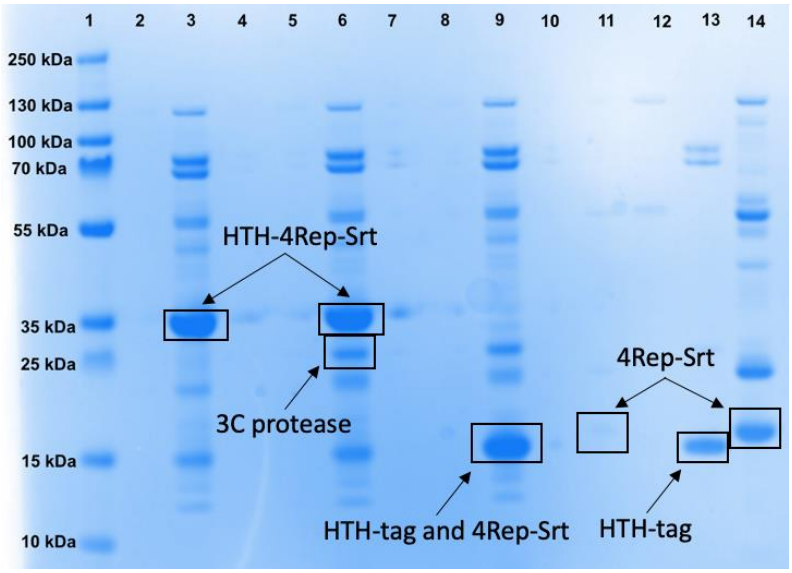
Figure 11. Final gel of first 4Rep-Srt purification containing concentrated sample. The sample is not pure but have a thick band for target protein 4Rep-Srt (13.47 kDa).

In the second purification of 4Rep-Srt, there was a sharp peak in elution 1 with target fusion protein during IMAC 1. Due to the contaminants, in well 6-11, **Figure 12**, the sharp peak is probably due to a combination of HTH-4Rep-Srt and contaminants. Fractions 6-12 were pooled and all of them contain fusion protein, although it is hard to see in well 6 and 12. The concentration of pooled sample prior to cleavage and dialysis was 0.659 mg/ml in 10.5 ml and there was no precipitation in sample after dialysis. IMAC 2 resulted in a peak for FT, with 4Rep-Srt present (Well 11, **Figure 13**). The final concentration of sample was 0.486 mg/ml in 1.5 ml, although not pure (Well 14, **Figure 13**).



Well	Sample	Volume (µl)
1	PageRuler	3
2	Lysate	10
3	FT	10
4	Wash 1	10
5	Elution 1 Fraction 5	10
6	Elution 1 Fraction 6	10
7	Elution 1 Fraction 7	10
8	Elution 1 Fraction 8	10
9	Elution 1 Fraction 9	10
10	Elution 1 Fraction 10	10
11	Elution 1 Fraction 11	10
12	Elution 1 Fraction 12	10
13	Elution 1 Fraction 17	10
14	Elution 1 Fraction 19	10
15	Elution 1 Fraction 21	10
16	End Elution 1	10
17	Elution 2	10

Figure 12. Gel 1 of second 4Rep-Srt purification. Visible bands for target fusion protein HTH-4Rep-Srt (28.79 kDa) in well 7-11, along a lot of contaminants. Fractions in well 6-12 were pooled prior to cleavage and dialysis. No fusion protein in wash 1 or elution 2.



Well	Sample	Volume (µl)
1	PageRuler	3
2	Wash 1 w/o 3C	10
3	Pool w/o 3C	10
4	Elution 2 w/o 3C	10
5	Wash 1 w/ 3C	10
6	Pool w/ 3C	10
7	Elution 2 w/ 3C	10
8	Wash 1 after dialysis	10
9	Pool after dialysis	10
10	Elution 2 after dialysis	10
11	FT	10
12	Elution 1	10
13	Elution 2	10
14	Concentrated 4Rep-Srt	10

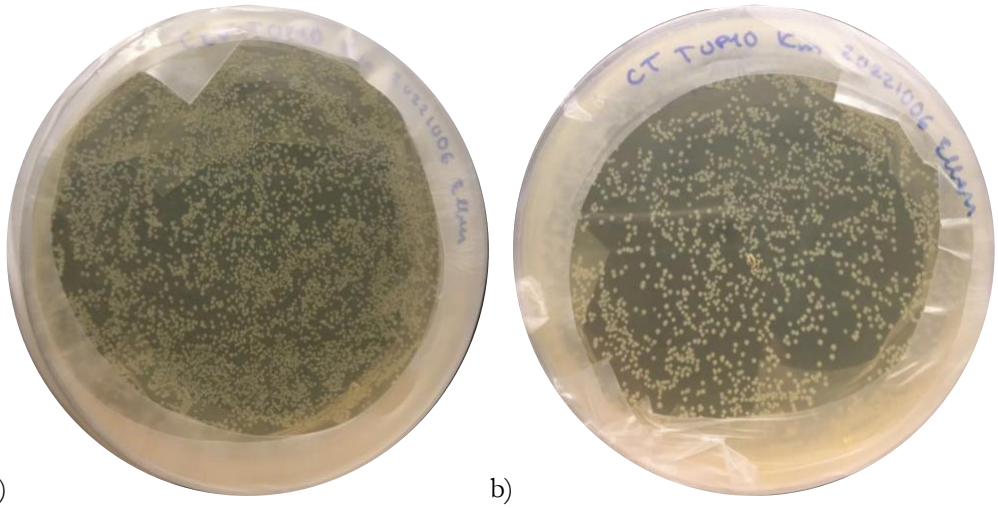
Figure 13. Gel 2 of second 4Rep-Srt purification. The target fusion protein HTH-4Rep-Srt (28.79 kDa) in well 3 and 6 has been cleaved, which is clear by lack of band in well 9 and a new lower band which is a mix of cleaved 4Rep-Srt (13.47 kDa) and HTH-tag (15.32 kDa). Target protein is present in final sample, well 14, along with contaminants.

3.3 Production of 5G-CT

The production of 5G-CT started with transformation of two plasmids, one containing the target vector and one with target insert. The plasmids were cleaved, ligated, and sent for sequencing where only one clone came back as a hit for G5-CT, which was expressed and purified.

3.3.1 Cloning of 5G-CT

The transformation of plasmids G4-FN-4RepCT (target vector) and G-CT (insert) was successful due to the many colonies on Km plates, **Figure 14**. There're more colonies for G4-FN-4RepCT plasmid, than for G-CT. After isolating the plasmid DNA's, the concentration of each O/N culture can be seen in **Table 3**.



a) b)

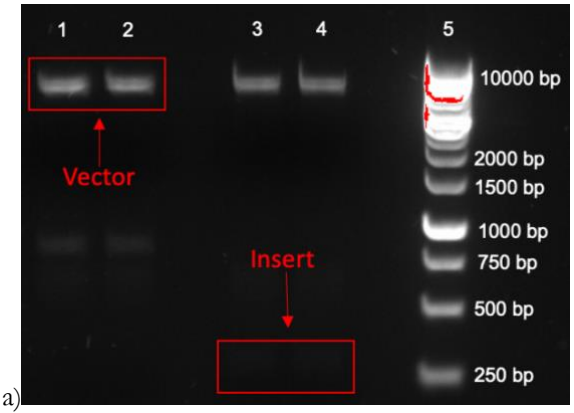
Figure 14. Transformation of plasmids a) G4-FN-4RepCT and b) G-CT, with *E. coli* TOP10 cells. Both transformations resulted in a lot of colonies, which is consistent with TOP10 cells. Two clones from each plate were chosen for O/N culturing.

Table 3. DNA concentration of transformed G4-FN-4RepCT and G-CT plasmids.

Sample Clone #	A_{260} (ng/µl)	A_{260}/ A_{280}	A_{260}/ A_{230}
G4-FN-4RepCT Clone 1	91.9	1.89	1.66
G4-FN-4RepCT Clone 2	140.6	1.90	1.64
G-CT Clone 1	124.0	1.87	1.52
G-CT Clone 2	142.1	1.90	1.47

After cleavage with all four clones with HindIII-HF and EcoRI, a DNA gel was run first with G4 Clone 1 (target vector from plasmid G4-FN-4RepCT) and G-CT Clone 1 (insert from plasmid G-CT) which resulted in **Figure 15**. The target vector (5643 kb) was clearly visible while the insert (299 bp) was hard to see even with longer exposure in **b)**. Bands of cut-out piece from G4 plasmid (829 bp) and remaining vector from G-CT (5618 bp) are also visible in the gel, indicating that the cleavage has worked. The two cutout gel pieces of target vector (G4 1 and 2) weighed 185.5 mg respectively 146.8 mg. Bands for insert were not visible while cutting with the UV lamp, so one gel piece was cut out (383.9 mg). Another DNA gel was repeated with G-CT Clone 2, resulted in **Figure 16**, where the insert, along remaining G-CT plasmid, are seen clearly in **b)**. Bands were still not clearly seen with UV lamp, so the bands were cut out blindly, resulted in two gel pieces G-CT 1 and 2 (217.5 mg respectively 219.5 mg). The DNA concentration of extracted gel DNA from cleaved target vector and insert can be seen in **Table 4**.

a)



Well	Sample	Volume (µl)
1	Vector	30
2	Vector	30
3	Insert	30
4	Insert	30
5	Gene Ruler 1 kb DNA ladder	15

b)

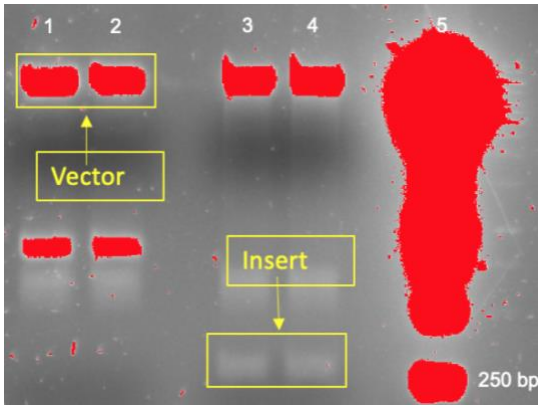
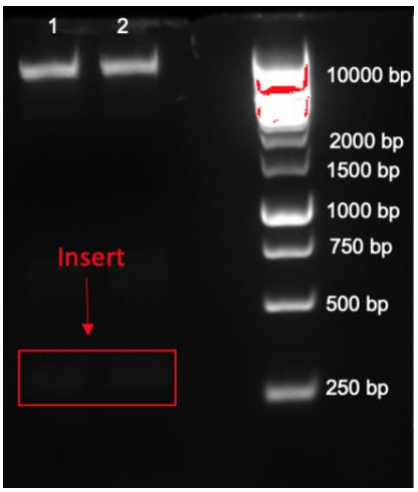


Figure 15. DNA gel of target vector (G4) and insert (G-CT). a) The two bands of target vector (5643 bp) are visible, while bands for insert (299 bp) are hard to see. Two bands of cut-out piece from G4 plasmid (829 bp) and two bands of remaining vector from G-CT (5618 bp) are also visible, indicating that the cleavage has worked as planned. b) Same gel as a), but with longer exposure allows for visualization of the insert bands.

a)



Well	Sample	Volume (µl)
1	Insert	30
2	Insert	30
3	Gene Ruler 1 kb DNA ladder	15

b)

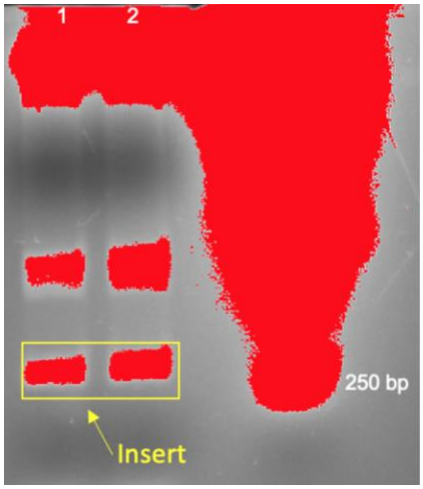


Figure 16. DNA gel of insert (G-CT). a) The bands for insert (299 bp) are hard to see. Two bands of the remaining vector from G-CT (5618 bp) are also visible, indicating that the cleavage has worked as planned. b) Same gel as a), but with longer exposure allows for visualization of insert.

Table 4. DNA concentration of cleaved target vector G4 and insert G-CT, extracted from DNA gel pieces.

Sample Clone #	A_{260} (ng/µl)	A_{260}/ A_{280}	A_{260}/ A_{230}
G4 1	2.9	1.60	0.03
G4 2	13.5	1.63	0.08
G-CT 1	8.4	2.02	0.03
G-CT 2	13.9	1.64	0.04

Ligation was performed between G4 1 and G-CT 1 (ligation 1), and G4 2 and G-CT 2 (ligation 2). The resulting Km plates for ligation 1 showed ~460 colonies and ~130 colonies on its negative control and for ligation 2 ~84 colonies and ~63 colonies on its negative control. The first ligation therefore seemed more promising, and 8 colonies were picked for O/N cultures (clone 1-8) and 2 colonies from ligation 2 (clone 9&10). The isolated DNA from each clone resulted in concentrations in **Table 5**. Clones 1-7 and 10 were sent for sequencing and clone 3 came back as a hit for target plasmid G5-CT (See Appendix **Figure A2**).

Table 5. DNA concentration of ligated clones.

Clone	A_{260} (ng/µl)	A_{260}/ A_{280}	A_{260}/ A_{230}
1	115.7	1.92	1.51
2	123.7	1.92	1.63
3	188.2	1.90	1.51
4	166.2	1.91	1.58
5	126.7	1.84	1.43
6	152.4	1.86	1.44
7	181.2	1.92	1.66
8	141.6	1.93	1.70
9	86.5	1.85	1.42
10	179.4	1.87	1.47

3.3.2 Expression of G5-CT

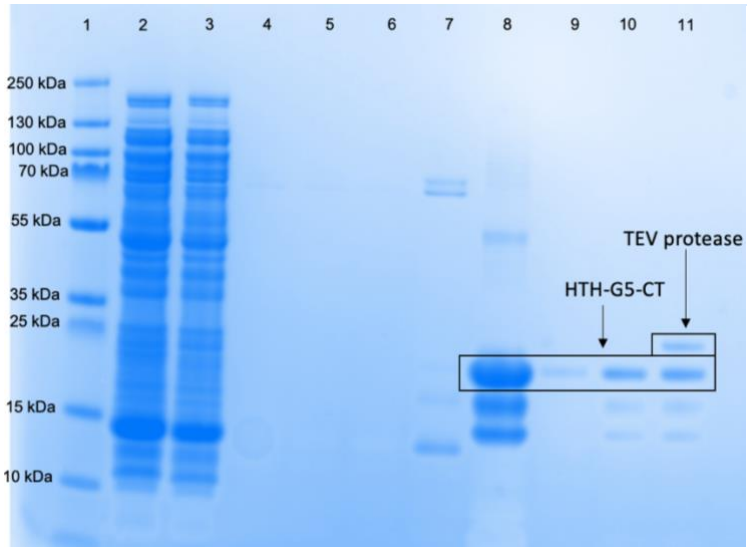
Only one expression was performed with G5-CT, and it reached wanted OD_{600} fast (**Table 6**).

Table 6. Expression of G5-CT The absorbance at 600 nm measured for each expression and their respectively times.

	OD_{600}	Time (h)
Expression 1	0.84	2.67

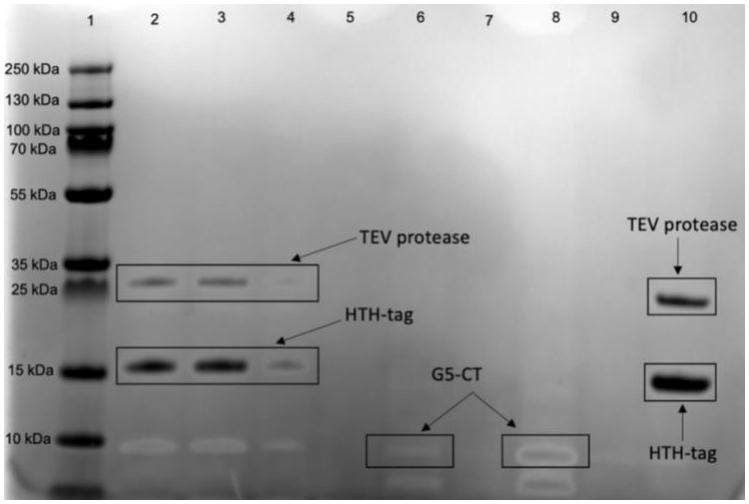
3.3.3 Purification of G5-CT

There was a sharp peak for elution 1 with target fusion protein during IMAC 1. 7 fractions were pooled and there is HTH-G5-CT present in sample, **Figure 17**. Before cleavage and dialysis, the sample started to aggregate a lot even at a low concentration of 0.481 mg/ml in 30 ml. The dilution proceeded to 200 ml sample with final concentration 0.697 mg/ml to prevent further aggregation of sample. After dialysis there were a few aggregates left. In gel 1 (**Figure 17**), it is clear that the band for fusion protein has fainted during the dilution process, when comparing well 8 and 10. Due to the large, dialyzed Pool volume, the loading of sample took a long time during IMAC 2, resulting in a low and stretched peak with target protein. As a result, the FT was concentrated in two rounds, FT 1 and 2, resulting in concentration 1.612 mg/ml in 1.7 ml respectively 0.653 mg/ml in 500 µl. Both samples have G5-CT present, even though the bands (Well 6 and 8, **Figure 18**) are almost white instead of black.



Well	Sample	Volume (µl)
1	PageRuler	3
2	Lysate	10
3	FT	10
4	Wash 1	10
5	Wash 2	10
6	Wash 3	10
7	Wash 4	10
8	Elution 1 Pool	10
9	Elution 2	10
10	Diluted Elution 1 pool	10
11	Diluted Elution 1 pool + TEV	10

Figure 17. Gel 1 of G5-CT purification. Fusion protein not visible in lysate, due to diluted sample. HTH-G5-CT band in Pool sample is very thick, while after dilution to 200 ml volume, the band is faint.



Well	Sample	Volume (µl)
1	PageRuler	3
2	Elution 1 pool after dialysis	10
3	Supernatant Elution 1 pool	10
4	Pellet Elution 1 pool	10
5	FT 1	10
6	Concentrated sample 1	10
7	FT 2	10
8	Concentrated sample 2	10
9	Elution 1	10
10	Elution 2	10

Figure 18. Gel 2 of G5-CT purification. Cleavage successful since there is no band for fusion protein left after dialysis, well 2-4. Target protein G5-CT (10.2 kDa) is present in both concentrated samples.

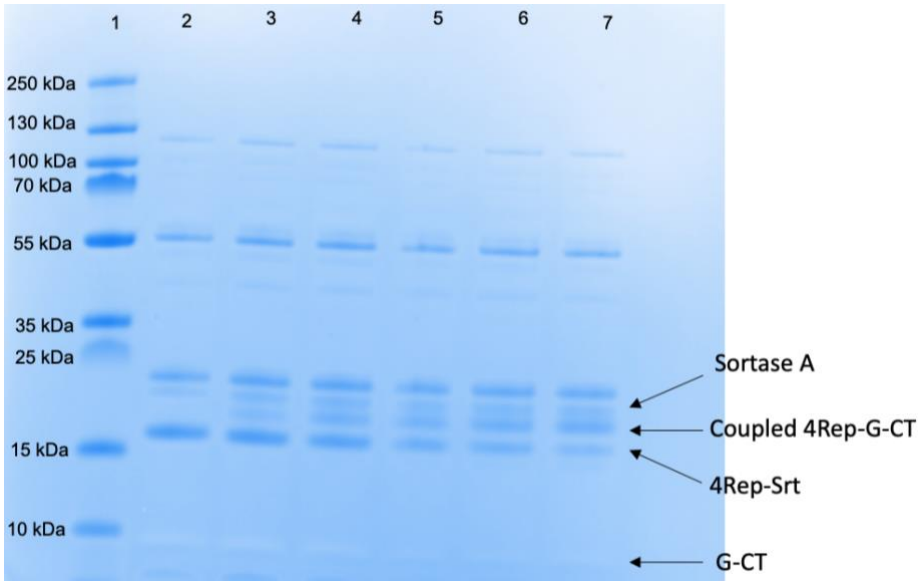
3.4 Sortase coupling and fiber formation

To test if produced G-/G5-CT and 4Rep-Srt are the wanted target proteins and can be coupled before making fibers, four Sortase coupling experiments were performed, along with 2 fiber formations.

3.4.1 Sortase coupling 4Rep-Srt and G-CT

The first Sortase experiment was done with 4Rep-Srt and G-CT.

In **Figure 19** the bands for G-CT (10 kDa) and 4Rep-Srt (13.47 kDa) gets less visible along the different time points, indicating that they are being coupled. A new band for coupled 4Rep-G-CT (23.47 kDa) is forming along the time points, already visible after 15 min (well 3).



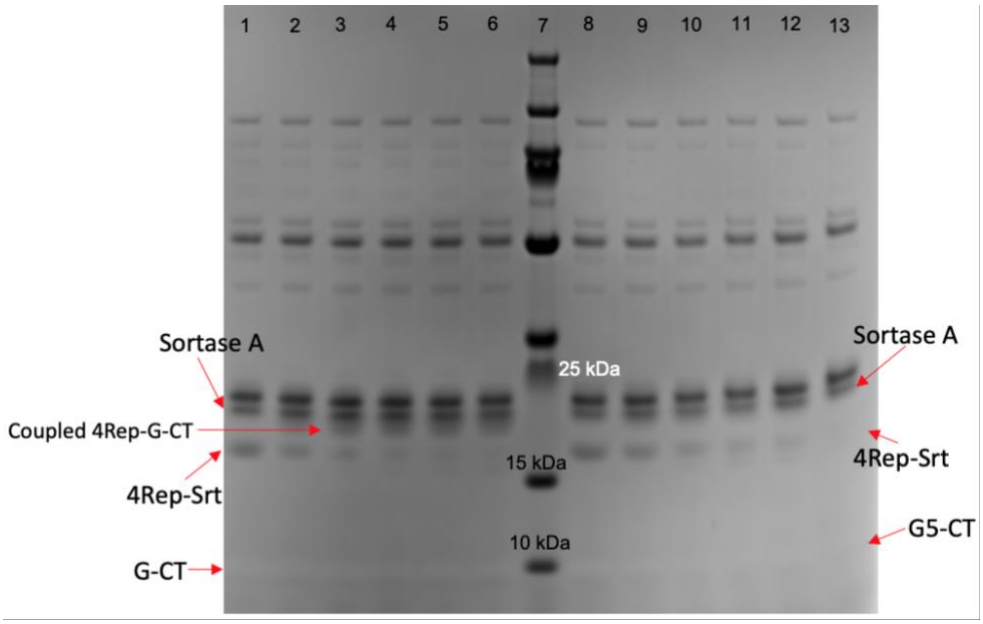
Well	Sample	Volume (µl)
1	PageRuler	5
2	0 min	16
3	15 min	16
4	30 min	16
5	45 min	16
6	60 min	16
7	90 min	16

Figure 19. Sortase coupling of 4Rep-Srt (13.47 kDa) and G-CT (10 kDa). Well 2-7 shows that the bands for 4Rep-Srt and G-CT gets fainter along the time points. In well 3-7, there's is a band forming which is coupled 4Rep-G-CT (23.47 kDa).

3.4.2 Sortase coupling 4Rep-Srt, G-CT and G5-CT

A second Sortase experiment with G-/G5-CT and 4Rep-Srt was conducted to investigate if G-CT or G5-CT is most efficient coupled with 4Rep-Srt.

Well 1-6 (**Figure 20**) shows the same result as in **3.4.1**, that the coupling between 4Rep-Srt and G-CT is working. In well 8-13, the bands for 4Rep-Srt and G5-CT (10.2 kDa) are disappearing along the time points, but no new band is forming where the coupled product (23.47 kDa) is supposed to be.



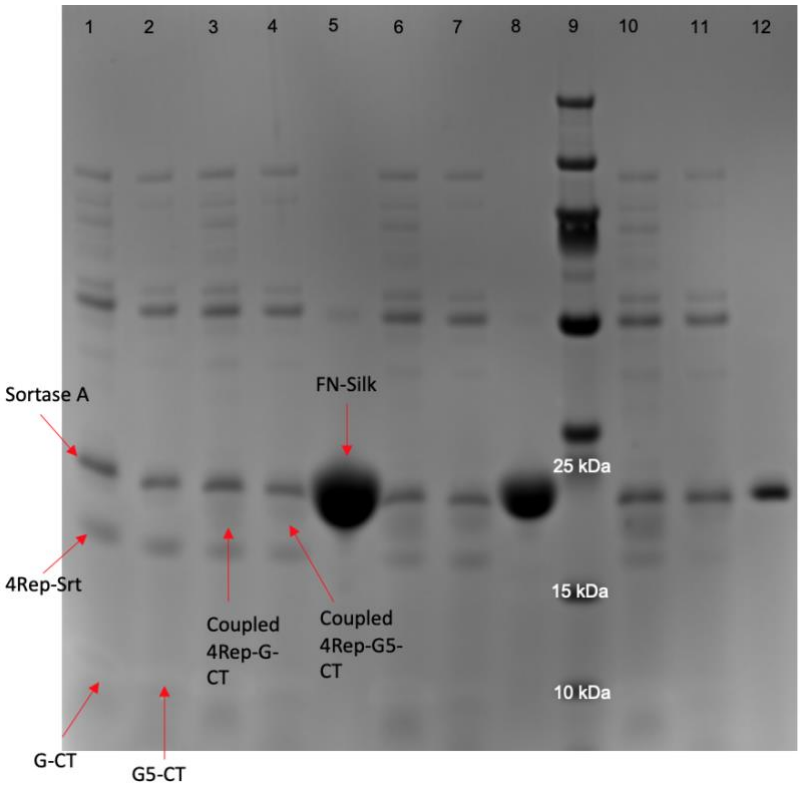
Well	Sample	Volume (µl)
1	4Rep-Srt and G-CT 0 min	14
2	4Rep-Srt and G-CT 15 min	14
3	4Rep-Srt and G-CT 30 min	14
4	4Rep-Srt and G-CT 45 min	14
5	4Rep-Srt and G-CT 60 min	14
6	4Rep-Srt and G-CT 90 min	14
7	PageRuler	6
8	4Rep-Srt and G5-CT 0 min	14
9	4Rep-Srt and G5-CT 15 min	14
10	4Rep-Srt and G5-CT 30 min	14
11	4Rep-Srt and G5-CT 45 min	14
12	4Rep-Srt and G5-CT 60 min	14
13	4Rep-Srt and G5-CT 90 min	14

Figure 20. Sortase coupling of 4Rep-Srt (13.47 kDa) and G-CT (10 kDa), and 4Rep-Srt and G5-CT (10.2 kDa). Well 1-6 shows that the bands for 4Rep-Srt and G-CT gets fainter along the time points. In well 3-6, there's is a band forming which is coupled 4Rep-G-CT (23.47 kDa). It seems like the bands 4Rep-Srt, and G5-CT gets fainter in well 8-13, but there is no new band forming where coupled 4Rep-G5-CT (23.67 kDa) is supposed be.

3.4.3 Fiber formation of coupled 4Rep-Srt-G-CT, 4Rep-Srt-G5-CT and FN-silk

A first experiment was conducted to test if coupled 4Rep-G-CT and 4Rep-G5-CT was able to form fibers, along with FN-silk that is already known to work.

Figure 21 shows that there is some coupled product being formed of both 4Rep-G-CT (23.47 kDa) and 4Rep-G5-CT (23.67 kDa) in well 3&4. After 30 min of fiber formation, soluble protein of FN-silk (24 kDa) is clearly forming fiber since the band is getting smaller. Coupled 4Rep-G-/G5-CT is still being coupled since the bands are getting thicker. There was no visible fiber in the glass cylinders after 30 min of fiber formation for 4Rep-G-CT or 4Rep-G5-CT. After 3 hours, band for FN-silk is almost completely gone and the fiber (**Figure 22**) was removed from glass cylinder. The band for 4Rep-G5-CT is also seemingly gone after 3 h but the fiber isn't very prominent. The band for 4Rep-G-CT seems to only have gotten thicker compared to the one from 30 min fiber formation, and the fibers itself was small compared to FN-silk.



Well	Sample	Volume (µl)
1	4Rep-Srt & G-CT 0 min Sortase coupling	14
2	4Rep-Srt &G5-CT 0 min Sortase coupling	14
3	4Rep-G-CT 30 min Sortase coupling/0 min Fiber formation	14
4	4Rep-G5-CT 30 min Sortase coupling/0 min Fiber formation	14
5	FN-silk 0 min Fiber formation	14
6	4Rep-G-CT 30 min Fiber formation	14
7	4Rep-G5-CT 30 min Fiber formation	14
8	FN-silk 30 min Fiber formation	14
9	PageRuler	6
10	4Rep-G-CT 3 h Fiber formation	14
11	4Rep-G5-CT 3 h Fiber formation	14
12	FN-silk 3 h Fiber formation	14

Figure 21. Fiber formation of coupled 4Rep-Srt-G-CT, 4Rep-Srt-G5-CT, and FN-silk. Visual bands forming in well 3 & 4, indicating that the coupling of 4Rep-Srt with G-CT resp. G5-CT is working. After 30 min fiber formation (well 6-8), the band of FN-silk (24 kDa) is already starting to become smaller due to soluble protein disappearing. The bands for 4Rep-Srt-G-CT (23.45 kDa), 4Rep-Srt-G5-CT (23.67 kDa) after 30 min of fiber formation are only getting thicker compared to FN-silks fiber formation ability. After 3 h, band of FN-silk is almost completely gone, along with coupled 4Rep-G5-CT. Coupled 4Rep-G-CT is seemingly still being coupled and band is ticker compared to the one of 30 min fiber formation.

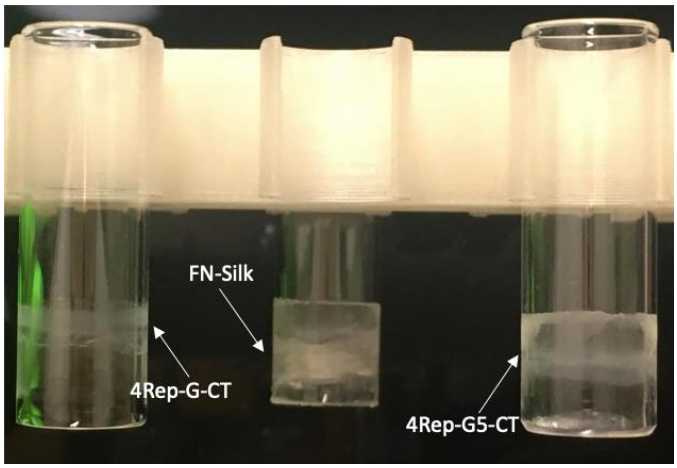
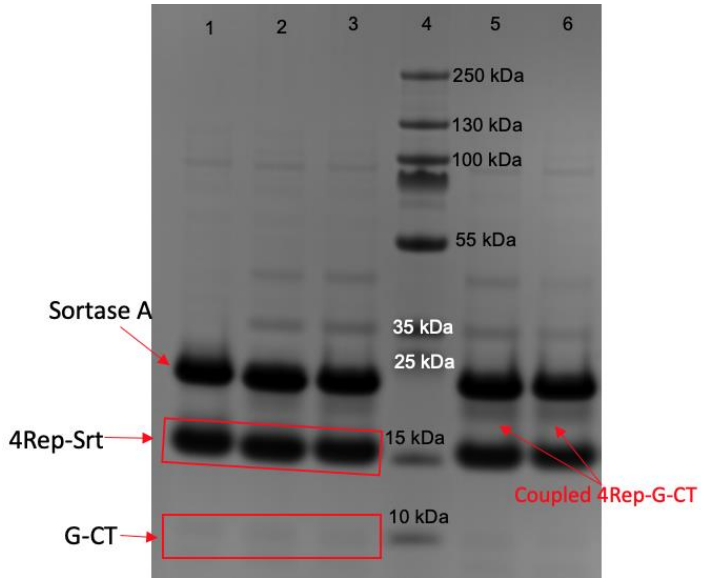


Figure 22. Fibers of coupled 4Rep-Srt-G-CT, 4Rep-Srt-G5-CT, and FN-silk after 3 h of fiber formation. The FN-silk is thick and was able to be removed from the glass cylinder. The 4Rep-G5-CT fiber is more visible than the one of 4Rep-G-CT. None of the 4Rep-G-/G5-CT fibers were thick enough to be removed from glass cylinder as a whole fiber.

3.4.4 Fiber formation of coupled 4Rep-Srt-G-CT

A last attempt to make one fiber was done with 4Rep-Srt and G-CT. After 30 min of coupling some band is seemingly forming where coupled 4Rep-G-CT (23.45 kDa) is supposed to be, **Figure 23**. There are also two bands forming between 35 and 55 kDa. Band for coupled protein in well 6 is fainter than in well 5 (3 h) so some soluble proteins are disappearing. The fiber formed after 7 h 20 min, **Figure 24**, is visible but still not thick enough to be removed from cylinder. SEM was conducted on the formed fiber, seen in **Figure 25**, though it's impossible to determine if the fiber is a product of 4Rep-G-CT, 4Rep-Srt or G-CT.



Well	Sample	Volume (µl)
1	4Rep-Srt & G-CT 0 min Sortase coupling	19
2	4Rep-G-CT 30 min Sortase coupling/0 min Fiber formation	19
3	4Rep-G-CT 30 min Fiber formation	19
4	PageRuler	6
5	4Rep-G-CT 3 h Fiber formation	19
6	4Rep-G-CT 7 h 20 min Fiber formation	19

Figure 23. Fiber formation of coupled 4Rep-Srt-G-CT (23.45 kDa). Very faint band for coupled protein being formed after 30 min fiber formation, along two bands between 35 and 55 kDa. Seems like some soluble protein is forming fiber since the band for 4Rep-G-CT is fainter in well 6 than in well 5.

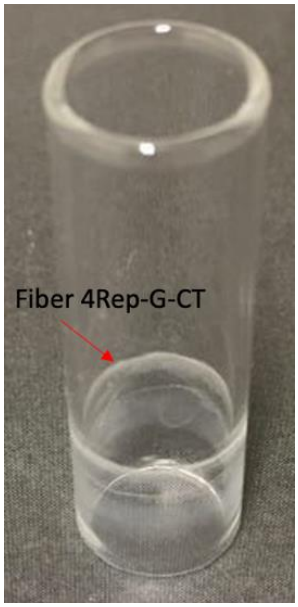


Figure 24. Fibers of coupled 4Rep-Srt-G-CT after 7 h 20 min of fiber formation. The fiber is very thin and stuck to the glass cylinder.

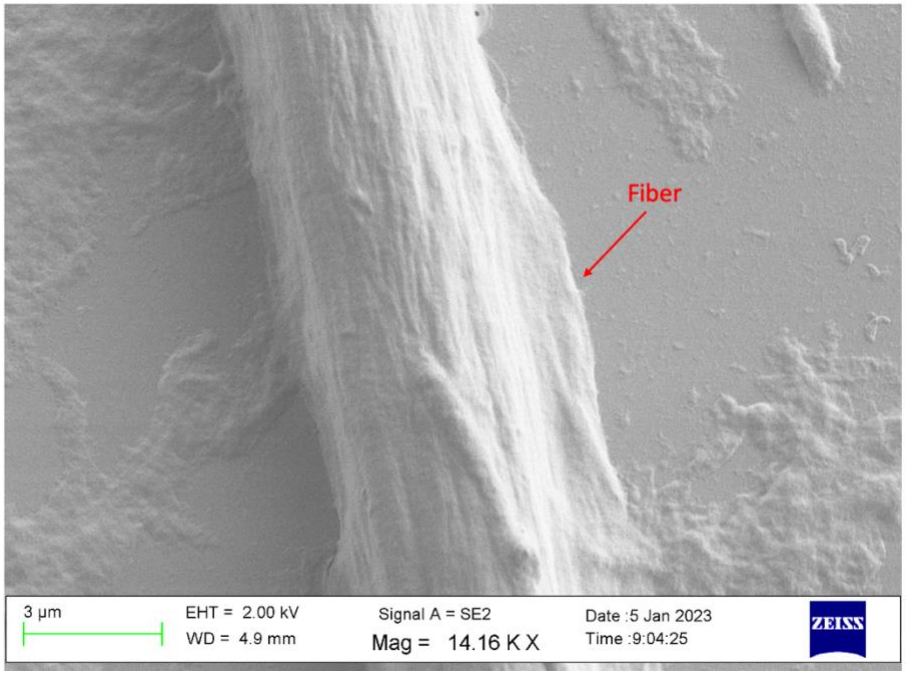


Figure 25. Fiber of coupled 4Rep-G-CT taken with SEM (Gemini Ultra 55, Zeiss, Germany). The fiber is approximately 3 µm in width.

4. Discussion

4.1 Production of G-CT

The expression of G-CT was the very first experiment executed in this master project. The first two weren't very successful since the cells took so long to reach the optimal OD_{600}. The idea was that the Gly-stock used had gone bad if the tube had been taken in and out of the freezer too many times and causing the cells to not grow well anymore. Therefore, a transformation with plasmid G-CT was performed, however the expression wasn't improved, and the performance was equal prior to transformation. There is a possibility that condensation in bottom of Km plate, formed when taking the plate from +37°C to +4°C, went into the colonies and reduced the cells' ability to grow. There was a second transformation performed by supervisor Ronnie Jansson, where he was extra careful not to flip the Km plate when picking colonies for O/N culture so condensation could go into agar-side of plate. It was with this Gly-stock that the final expression was performed with, and it worked well. Another difference between the first four and fifth expression is that TSB+Y media was used during cultivation and expression, while LB media was used the other times. Both mediums are rich [23] but since TSB+Y is also used for the O/N pre-culture in all 5 cases, the cells don't have to get used to the new media in the fifth expression which could have made the cells grow better during the exponential phase. It would be interesting to perform another expression with the Gly-stock from the first transformation of G-CT but with TSB+Y media during main cultivation to see if the OD_{600} is reached after 8 hours or if the time would decrease. But nevertheless, a protocol was developed that successfully express G-CT.

The first purification of G-CT didn't go as planned. Probably due to the poor expression but it could also be because the purification (IMAC 1 & 2) was performed in two different departments, so the samples had to be transferred ~15 min one way. Silk proteins are very sensitive and prone to aggregation so any shaking movements could cause them to behave in unwanted ways. However, the reason for the non-existing peak in IMAC 1 is most likely because of the poor expression, but the transportation probably didn't help.

The second purification resulted in two peaks during IMAC 1 which both contained target fusion protein. Wash 4 was done with 75 mM Imidazole which was too high to only elute unspecific bound protein, because HTH-G-CT also eluted. Elution of unspecific bound proteins from the column to get pure final sample is dependent on each protein and because G-CT hasn't been purified many times, this purification is a part of the process to optimize the purification, so until next time wash 4 should only be conducted with 60 mM Imidazole. The aggregation that occurred in Pool during dialysis resulted in the conclusion that one should always dilute the sample to approximately 0.5 mg/ml (also applicable on 4Rep-Srt and G5-CT) to avoid aggregation. The only downside to this is that loading of sample during IMAC 2 will take longer because of the larger volume. The final concentrated sample from Wash 4 (2.09 mg/ml) was not pure in the end. This isn't surprising since the washes during IMAC 1 is to elute unspecific bound protein, so the peak that occurred was not only caused by fusion protein, and those contaminants followed along the purification until final sample. The final sample from Pool (0.611 mg/ml and 0.367 mg) didn't reach the target concentration because the volume became too small to concentrate further. Spider silk proteins are known to have low yield after purification, so this yield isn't very surprising. Researchers have tried to fix the low yield by introducing a Km-resistance gene to increase protein expression and solubility tags, such as Thioredoxin that is being used in this project, but even though the vector contains both of these, the yield wasn't high. [4][24]

4.2 Production of 4Rep-Srt

Both expressions of 4Rep-Srt went smoothly, but the purification had some issues. The first one resulted in a lot of fusion protein during elution 1 (IMAC 1), so that was a good sign to get high yield. There was however also fusion protein in the washes and elution 2 so until the second purification, it was decided to omit the second wash (IMAC 1). It is also worth mentioning that the volume for wash 1 and wash 2 are 30 ml each so the bands that are very faint in gel is probably due to them being diluted. It would have been beneficial to proceed with those two samples in IMAC 2 to get more protein in the end. As the same as for G-CT, the sample wasn't diluted enough prior to dialysis so the sample aggregated. The concentrated sample actually ended up with a decent concentration but it was not pure so the concentration can't be assumed to be a result from only 4Rep-Srt. There is one other prominent band in the sample which clearly wasn't removed at all. Its presence could be because of not enough washing but it was never determined, and it's also present in the second purification. In the end, only proteins with the Srt-tag (i.e., 4Rep-Srt) and glycine's at the N-terminal (i.e., G-/G5-CT) will be coupled so it's not a major problem that the sample isn't pure.

An improvement that occurred until the second purification was that there wasn't any fusion protein in the wash (IMAC 1), because there wasn't any imidazole in the buffer. The unwanted proteins in elution 1 was even more the second time around so it would have been beneficial to do a short wash with imidazole just to remove some protein, even though it could decrease the yield in the end. The initial FT (IMAC 2) did show a very faint band for 4Rep-Srt, and once it was concentrated it was much thicker, along with many other proteins. It was hard to determine if the target protein had ended up in elution 2 or in the FT, since the HTH-tag and 4Rep-Srt are similar in size, but since the sample is able to form fiber (will be discussed further down), its presence in the sample was proven. It is still undetermined why there is a band in well 13, next to 4Rep-Srt in well 14, it was believed to be the tag but it's larger than the target protein so it shouldn't have travelled as far down the gel. Comparing the second and first purification, the yield (mg) of protein was less in the second purification.

4.3 Production of G5-CT

The production of G5-CT including many steps that overall went well. During the transformation of the ligated product, one of the transformations didn't result in a huge difference in colonies between the plate and its negative control (i.e., without insert). The negative control was a part of the transformation to estimate how many colonies from the real plate contained the right plasmid. Fortunately, the other transformation did generate a larger difference in colonies and one of the clones came back as a hit for G5-CT.

During the purification, after IMAC 1 and prior to dialysis, the sample aggregated a lot even at the low concentration of 0.481 mg/ml. It's known from experiments conducted by Spiber Technologies AB on G5-FN-4RepCT that uncleaved protein can precipitate even at 0.4 mg/ml due the construct being unstable. It seems like G5-CT is as unstable and to prevent further aggregation of the sample, dilution proceeded up to a volume of 200 ml. It seems to have worked since there were no more aggregation after dialysis. Due to the large volume and flow rate 0.5 ml/min during loading of sample (IMAC 2), the sample had to be concentrated in two rounds. Keeping the sample in a beaker on ice for too long could have resulted in aggregation, since it's known from G5-FN-4RepCT that the cleaved protein is even more prone to aggregate than the uncleaved one. One of the concentrated samples did reach a decent concentration of 1.612 mg/ml in 1 while the other only reached 0.653 mg/ml.

4.4 Sortase assay and Fiber formation

The first coupling of G-CT and 4Rep-Srt was successful where it was clear in the gel that they were being coupled by Sortase A. It was harder to determine if the coupling of 4Rep-Srt and G5-CT worked since there wasn't any band for coupled product, only removal of the separate bands. It was believed that G5-CT would be more efficient coupled to 4Rep-Srt since the protein has more glycine's but that wasn't the case, so it probably is something else causing the protein not being coupled in the second experiment. However, the concentrations of all the proteins needs to be higher before any final conclusions can be made about the coupling efficiency, since the bands for G-CT and G5-CT are barely visible at current concentrations.

What was weird was that the fiber was more visible in the glass cylinder for 4Rep-G5-CT, than for 4Rep-G-CT, even though the gel says the opposite. However, the fibers were so thin that they couldn't be removed from cylinder and therefore, couldn't determine what actually was causing the fiber formation. It is known that both G-CT and 4Rep-Srt can form fibers individually, so the fiber formed for 4Rep-G5-CT can be 4Rep-Srt or G5-CT individually. The protein samples that were being used for the fibers needs to have higher concentration from the beginning, in order to be able to detach the fibers from cylinder and perform NMR or other analyzes.

The last attempt to make fibers were done with 4Rep-Srt and G-CT, since G-CT sample had higher concentration than all available G5-CT samples. Still, the fiber was too thin for detachment even though it was visible. As mentioned before, concentrations of G-CT and 4Rep-Srt needs to be higher in future experiments in order to make a thick fiber. In order to determine if the formed fiber seen by SEM is in fact coupled 4Rep-G-CT, and not 4Rep-Srt or G-CT, one would have to produce one fiber of each construct to see their difference.

5. Conclusion and future perspectives

The aim of this project was to produce two proteins, G-CT and 4Rep-Srt, that would be coupled with Sortase A before fiber formation. G-CT is in future experiments supposed to be produced with ^{13}C and ^{15}N to label only one of the proteins being coupled. With this method, future experiments would be able to determine the structure of the CT domain in a fiber of 4Rep-G-CT.

Throughout the project, it was decided that G5-CT should also be produced since it was known from previous master projects that 5 glycine's increased the coupling efficiency by Sortase A. The cloning of G5-CT had some obstacles, but one clone did manage to come back as a match for target plasmid. The production of all the proteins was successful and as much as three Sortase experiments was performed, along two fiber formation experiments. The problem was that these proteins hadn't been expressed and purified much prior to this project, so optimization of the protocols occurred during the project. Because of this, the final samples of each protein weren't that pure, and the yield and concentration weren't high enough to easily interpreted results. Normally, 1 mg/ml of silk protein in fiber formation sample is needed to get a thick fiber, however none of the samples produced here had high enough concentration to reach 1 mg/ml.

Before attempting to label G-CT or G5-CT prior to NMR analysis, more optimization of the purification procedures is necessary for all the proteins to increase the purity while still getting high protein yield. It is crucial to get high concentration of proteins so that they can be diluted to 1 mg/ml in the fiber formation solution. If the fiber produced isn't thick enough, detachment from the glass cylinder won't be possible and therefore neither NMR analysis. Further optimization of Sortase assay (e.g., concentration of enzyme, time of coupling) is also needed in order to get good coupling while still making fibers since now the coupling seems to proceed throughout the whole fiber formation. To conclude this project, the experiments were successful, it just needs more optimization before the structure of CT in fiber form can be completely investigated.

6. Acknowledgements

A very special thank you to my supervisors for guiding me through this project. Thank you Ronnie Jansson for always helping me when I'm in need, for making sure I understand and for always encouraging me. I appreciate all our meetings and you taking time to help me even if you're busy. Thank you My Hedhammar for having me in your wonderful research group and for trusting me with this project. I have learned so much from you and your group, and I will always remember this exciting time. Thank you Danilo Kimio Hirabae De Oliveira for you guidance and for helping me when in trouble. Thanks for staying late with me in the lab and even then, staying positive and encouraging.

Thanks My Hedhammars research group and Spiber Technologies AB for the very interesting group meetings and for welcoming me. Special thanks to Simon Cannava for providing TEV protease and your insight into tricky silk proteins, and to Linnea Gustafsson for helping me with my final experiment and providing pretty figures for my report.

Thanks to all the people on floor 3 for fun conversations at lunches and for the joyful Christmas time.

And last but not least, thanks to my sambo Olof for always making me food after late evenings in the lab, giving me comfort when feeling stressed and for your support.

7. References

[1] Römer L, Scheibel T. The elaborate structure of spider silk: structure and function of a natural high performance fiber. Prion. 2008 Oct-Dec;2(4):154-61. doi: 10.4161/pri.2.4.7490. Epub 2008 Oct 20. PMID: 19221522; PMCID: PMC2658765.

[2] Fredriksson C, Hedhammar M, Feinstein R, Nordling K, Kratz G, Johansson J, Huss F, Rising A. Tissue Response to Subcutaneously Implanted Recombinant Spider Silk: An *in Vivo* Study. Materials (Basel). 2009 Nov 20;2(4):1908–22. doi: 10.3390/ma2041908. PMCID: PMC5513568.

[3] Vollrath F, Barth P, Basedow A, Engström W, List H. Local tolerance to spider silks and protein polymers in vivo. In Vivo. 2002 Jul-Aug;16(4):229-34. PMID: 12224131.

[4] Hinman MB, Jones JA, Lewis RV. Synthetic spider silk: a modular fiber. Trends Biotechnol. 2000 Sep;18(9):374-9. doi: 10.1016/s0167-7799(00)01481-5. PMID: 10942961.

[5] Gosline JM, Guerette PA, Ortlepp CS, Savage KN. The mechanical design of spider silks: from fibroin sequence to mechanical function. J Exp Biol. 1999 Dec;202(Pt 23):3295-303. doi: 10.1242/jeb.202.23.3295. PMID: 10562512.

[6] Vollrath F, Knight DP. Liquid crystalline spinning of spider silk. Nature. 2001 Mar 29;410(6828):541-8. doi: 10.1038/35069000. PMID: 11279484.

[7] Hedhammar M, Rising A, Grip S, Martinez AS, Nordling K, Casals C, Stark M, Johansson J. Structural properties of recombinant nonrepetitive and repetitive parts of major ampullate spidroin 1 from Euprosthenops australis: implications for fiber formation. Biochemistry. 2008 Mar 18;47(11):3407-17. doi: 10.1021/bi702432y. Epub 2008 Feb 23. PMID: 18293938.

[8] Andersson M, Johansson J, Rising A. Silk Spinning in Silkworms and Spiders. Int J Mol Sci. 2016 Aug 9;17(8):1290. doi: 10.3390/ijms17081290. PMID: 27517908; PMCID: PMC5000687.

[9] Sarr M, Kitoka K, Walsh-White KA, Kaldmäe M, Metlāns R, Tārs K, Mantese A, Shah D, Landreh M, Rising A, Johansson J, Jaudzems K, Kronqvist N. The dimerization mechanism of the N-terminal domain of spider silk proteins is conserved despite extensive sequence divergence. J Biol Chem. 2022 May;298(5):101913. doi: 10.1016/j.jbc.2022.101913. Epub 2022 Apr 7. PMID: 35398358; PMCID: PMC9097459.

[10] Sponner A, Unger E, Grosse F, Weisshart K. Conserved C-termini of Spidroins are secreted by the major ampullate glands and retained in the silk thread. Biomacromolecules. 2004 May-Jun;5(3):840-5. doi: 10.1021/bm034378b. PMID: 15132670.

[11] Tucker CL, Jones JA, Bringhurst HN, Copeland CG, Addison JB, Weber WS, Mou Q, Yarger JL, Lewis RV. Mechanical and physical properties of recombinant spider silk films using organic and aqueous solvents. Biomacromolecules. 2014 Aug 11;15(8):3158-70. doi: 10.1021/bm5007823. Epub 2014 Jul 29. PMID: 25030809; PMCID: PMC4130237.

[12] Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol. 2014 Apr 17;5:172. doi: 10.3389/fmicb.2014.00172. PMID: 24860555; PMCID: PMC4029002.

[13] Gustafsson L, Jansson R, Hedhammar M, van der Wijngaart W. Structuring of Functional Spider Silk Wires, Coatings, and Sheets by Self-Assembly on Superhydrophobic Pillar Surfaces. Adv Mater. 2018 Jan;30(3). doi: 10.1002/adma.201704325. Epub 2017 Dec 4. PMID: 29205540.

[14] Stark M, Grip S, Rising A, Hedhammar M, Engström W, Hjälm G, Johansson J. Macroscopic fibers self-assembled from recombinant miniature spider silk proteins. Biomacromolecules. 2007 May;8(5):1695-701. doi: 10.1021/bm070049y. Epub 2007 Apr 3. PMID: 17402782.

[15] Johansson U, Widhe M, Shalaly ND, Arregui IL, Nilebäck L, Tasiopoulos CP, Åstrand C, Berggren PO, Gasser C, Hedhammar M. Assembly of functionalized silk together with cells to obtain proliferative 3D cultures integrated in a network of ECM-like microfibers. Sci Rep. 2019 Apr 18;9(1):6291. doi: 10.1038/s41598-019-42541-y. PMID: 31000733; PMCID: PMC6472362.

[16] Jansson R, Thatikonda N, Lindberg D, Rising A, Johansson J, Nygren PÅ, Hedhammar M. Recombinant spider silk genetically functionalized with affinity domains. Biomacromolecules. 2014 May 12;15(5):1696-706. doi: 10.1021/bm500114e. Epub 2014 Apr 8. PMID: 24678858.

[17] Clancy KW, Melvin JA, McCafferty DG. Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. Biopolymers. 2010;94(4):385-96. doi: 10.1002/bip.21472. Erratum in: Biopolymers. 2010;94(5):681. PMID: 20593474; PMCID: PMC4648256.

[18] Mao H, Hart SA, Schink A, Pollok BA. Sortase-mediated protein ligation: a new method for protein engineering. J Am Chem Soc. 2004 Mar 10;126(9):2670-1. doi: 10.1021/ja039915e. PMID: 14995162.

[19] Nilebäck L, Widhe M, Seijsing J, Bysell H, Sharma PK, Hedhammar M. Bioactive Silk Coatings Reduce the Adhesion of *Staphylococcus aureus* while Supporting Growth of Osteoblast-like Cells. ACS Appl Mater Interfaces. 2019 Jul 17;11(28):24999-25007. doi: 10.1021/acsami.9b05531. Epub 2019 Jul 5. PMID: 31241302.

[20] Slabinski L, Jaroszewski L, Rodrigues AP, Rychlewski L, Wilson IA, Lesley SA, Godzik A. The challenge of protein structure determination--lessons from structural genomics. Protein Sci. 2007 Nov;16(11):2472-82. doi: 10.1110/ps.073037907. PMID: 17962404; PMCID: PMC2211687.

[21] Petrotchenko EV, Borchers CH. Protein Chemistry Combined with Mass Spectrometry for Protein Structure Determination. Chem Rev. 2022 Apr 27;122(8):7488-7499. doi: 10.1021/acs.chemrev.1c00302. Epub 2021 Dec 30. PMID: 34968047.

[22] Puthenveetil R, Vinogradova O. Solution NMR: A powerful tool for structural and functional studies of membrane proteins in reconstituted environments. J Biol Chem. 2019 Nov 1;294(44):15914-15931. doi: 10.1074/jbc.REV119.009178. Epub 2019 Sep 24. PMID: 31551353; PMCID: PMC6827292.

[23] Kram KE, Finkel SE. Rich Medium Composition Affects Escherichia coli Survival, Glycation, and Mutation Frequency during Long-Term Batch Culture. Appl Environ Microbiol. 2015 Jul;81(13):4442-50. doi: 10.1128/AEM.00722-15. Epub 2015 Apr 24. PMID: 25911475; PMCID: PMC4475895.

[24] Costa S, Almeida A, Castro A, Domingues L. Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. Front Microbiol. 2014 Feb 19;5:63. doi: 10.3389/fmicb.2014.00063. PMID: 24600443; PMCID: PMC3928792.

Appendix

A1. Expression 2-5 of G-CT and preparation of Gly-stock

A1.1 Expression 2 of G-CT

The O/N pre-culture and expression of G-CT was repeated as in **2.1** with Gly-stock prepared **2.1.1** and 10 ml of O/N pre-culture was used to inoculate each 5 L flask. Each pellet was resuspended in 24 ml 20 mM Tris pH 8 (144 ml 20 mM Tris in total), pooled and divided in 6 50 ml Falcon tubes with ~24 ml in each tube and stored in -20°C.

A1.2 Preparation of glycerol stock (G-CT *E. coli* BL21)

A1.2.1 Transformation of *E. coli* by heat shock with plasmid G-CT plasmid

In a sterile Eppendorf tube, 7.5 µl sterile H_2O (StH_2O) and 0.5 µl plasmid DNA (G-CT plasmid 48.2 µg/ml) was transferred and spun down. 2 µl 5x KCM (500 mM KCl, 150 mM $CaCl_2$ and 250 mM $MgCl_2$) was transferred to the tube and kept on ice for minimum 3 min to chill the sample before adding 10 µl of thawed cells (*E. coli* BL21). Sample was incubated 20 min on ice before incubation at +42°C for 1 min. Sample was kept on ice for 3 min before 200 µl TSB+Y was added. Sample was incubated 1 h at +37°C with shaking, 650 rpm. ~220 µl of sample was plated on agar plate (40 g Blood Agar Base, 1 L deionized water and 1 ml 50 mg/ml Kanamycin) with Kanamycin (Km) and incubated 16±1h, +37°C.

A1.2.2 O/N culture of target vector G-CT plasmid and preparing Gly-stock

One colony of transformed target plasmid from Km agar plate from **A1.2.1** was inoculated into a sterile 50 ml shake flask with 5 ml TSB+Y and 5 µl Kanamycin (C_{Final} = 50 µg/ml). Culture was incubated O/N (16±1h), +37°C, 150 rpm. A Gly-stock was prepared from the O/N pre-culture by transferring 35 µl 85% Sterile glycerol (C_{Final} = 15%) and 165 µl pre-culture to a sterile Eppendorf tube. The Gly-stock was stored in -80°C.

A1.3 Expression 3 of G-CT

The O/N pre-culture and expression of G-CT was repeated as in **2.1** with Gly-stock prepared **A1.2.2** and 10 ml of O/N pre-culture was used to inoculate each 5 L flask. Each pellet was resuspended in 20 ml 20 mM Tris pH 8 (120 ml 20 mM Tris in total), pooled and divided in 6 50 ml Falcon tubes with ~20 ml in each tube and stored in -20°C

A1.4 O/N culture of target vector G-CT plasmid and preparing Gly-stock

The making of O/N culture and Gly-stock was repeated as in **A1.2.2** with incubating O/N at 30°C, 150 rpm the first 30 min and then 200 rpm. 2 agar plates with Km were scraped with Gly-stock and incubated over weekend in +4°C.

A1.5 Expression 4 of G-CT

From the two Km plates made in **A1.4**, two O/N pre-cultures were made as in **2.1.1** with one colony from each plate and one O/N pre-culture was made from Gly-stock made in **A1.4**. The expression was repeated as in **2.1.2** with one of the O/N pre-cultures made from Km plate, with 6 sterile 2 L flasks instead of 4 5

L flasks. Each 2 L flasks contained 500 ml LB media, 500 µl Kanamycin (C_{Final} = 50 µg/ml) and was induced with 500 µl IPTG (C_{Final} = 0.5 mM). The flasks were shaken at 180 rpm the first 30 min and then increased to 220 rpm. Two Gly-stocks were made, one from each O/N pre-culture made from colonies from Km plate.

A1.6 Expression 5 of G-CT

The making of O/N pre-culture and Gly-stock was repeated as in **2.1.1** with Gly-stock (G-CT plasmid) provided by Ronnie Jansson. The culture was incubated in a sterile 1 L shake flask, O/N at 30°C, 180 rpm. The expression was repeated as in **2.1.2**, with 6 sterile 2 L flasks instead of 4 5 L flasks. Each 2 L flasks contained 500 ml LB media, 500 µl Kanamycin (C_{Final} = 50 µg/ml), was inoculated with 8 ml O/N culture, and was induced with 500 µl IPTG (C_{Final} = 0.5 mM). The flasks were shaken at 180 rpm the first 30 min and then increased to 200 rpm. The cultures were cooled down for 30 min to slow down cell growth and protein expression was carried out for 3.5 h before harvesting (4750 rcf, +4°C, 20 min) and resuspending cells in 180 ml 20 mM Tris (pH 8) in total.

A2. Strip, CiP and charge of columns

A2.1 Strip, CiP and charge 5 ml column for IMAC 1

The column was attached to ÄKTA upside-down when Strip, CiP and charging it. Between each step, column was washed with 6 column volumes (CV) Milli-Q, 2 ml/min. The column was stripped of metal ions (Zn^{2+}/Ni^{2+}) with 1.5 CV 0.2 M EDTA&0.5 M NaCl (pH 7.5) with flow rate 2 ml/min. The column was clean-in placed (CiP) with 10.5 CV 2 M NaCl at 3.5 ml/min and 6 CV 1 M NaOH at 2 ml/min followed by incubation for 1 h with 1 M NaOH, 1 ml/min. Column washed with 0-100% gradient of 70% ethanol (EtOH) with 2 CV on each step and flow rate 2 ml/min. Column charged with 0.5 CV 0.1 M $ZnCl_2$ (pH 5.0-5.5) at 0.5 ml/min and with 3.5 CV 0.02 NaAc/0.5 M NaCl (Acetate buffer, pH 8) at 1 ml/min, the loosely bound metal ions was removed. The column was equilibrated with 20 CV 20 mM Tris (Binding buffer (A), pH 8) at 2 ml/min and pH in column was assured to be 8. As a final step, the column run with 6 CV Milli-Q at 2 ml/min before finally storing column in +4°C.

A2.2 Strip, CiP and charge 1 ml column for IMAC 2

1 ml column was prepared as in **A2.1** but the flow rate when stripping metal ions were 1 ml/min and the flow rate for 2 M NaCl was 2 ml/min.

A3. Purification 2 of 4Rep-Srt

A3.1 Preparing columns, lysis and centrifugation

Two 5 ml columns, one with Zn^{2+} and with Ni^{2+}, was prepared in the same way as **2.3**. Cells from **2.5** was lysed and centrifuged in the same way as **2.2.1** but with 14.8 ml of 4 M NaCl (C_{Final} = 400 mM) and no added Imidazole.

A3.2 Purification IMAC 1 and dialysis

IMAC 1 was performed with lysed cells from **A3.1** as in **2.6.2** but the column (Zn^{2+}) was equilibrated with Third buffer and when collecting all protein in sample tube, Third buffer was used. Wash 1 was performed at 2 ml/min, no second wash was performed and elution 1 was performed with 260 mM Imidazole (52% Elution buffer (B)). The cleavage and dialysis were performed similarly as in **2.6.2** with

pooled fractions from elution 1, along wash 1 and elution 2. The cleavage was carried out with half the amount of calculated protease 3C ($V_{3C}(\mu l) = \frac{1\ ml * \frac{Amount\ of\ 3C\ needed}{1000}}{1.8\ mg}/2$) and the dialysis was done with Dialysis buffer (20 mM Tris/150 mM NaCl, pH 8) and Spectra/Por®3 Dialysis Membrane Standard RC Tubing MWCO: 3.5 kDa. Gel samples and SDS-PAGE were performed as in **2.3.2**.

A3.3 Purification IMAC 2 and sample concentration

IMAC 2 (Ni^{2+}) was performed as in **2.6.3** with dialyzed pool sample from **A3.2** and collecting all target protein in sample tube was done with equilibrated Buffer D (20 mM Tris and 150 mM NaCl, pH 8). The protein concentration was determined in the same way as in **2.6.3**.

A4. Cloning of G5-CT

Dry cloning was performed prior to wet cloning in Geneious (Geneious Prime® 2022.2.2) with G-CT plasmid and G4-FN-4RepCT plasmid to make sure that the target plasmid would be obtained.

A4.1 Transformation of *E. coli* by heat shock with plasmids G-CT and G4-FN-4RepCT and O/N cultures

Two transformations were performed as in **A1.2.1** with plasmids G-CT (48.2 µg/ml) and G4-FN-4RepCT (109 ng/µl) and TOP10 *E. coli* cells. O/N cultures was prepared in similar manners as in **A1.2.2**, with two colonies from each transformation in total four flasks.

A4.2 Isolation of plasmid DNA

The four O/N cultures from **A4.1** were centrifuged (3237 rcf, 15 min, RT) and supernatants was discarded while pellets were put on ice. The plasmid DNA in pellets were obtained by performing QIAprep® Spin Miniprep Kit Protocol with Kit (250), with point 7 omitted. 35 µl StH_2O was used to elute DNA from each spin column.

The DNA plasmid concentrations was determined with NanoDrop (Thermo Scientific) with StH_2O as blank before storing the samples in -20°C.

A4.3 Restriction cleavage

Four restriction cleavages were performed on G-CT and G4-FN-4RepCT from **A4.2** to obtain the insert (299 bp) and target vector (5643 bp), respectively. The two restriction cleavage enzymes used were EcoRI and HindIII-HF. 2 µg of each plasmid was mixed with 5 µl 10x rCutSmart Buffer, 1 µl EcoRI and filled up to 49 µl with StH_2O before incubation at 37°C for 1 h. 1 µl of HindIII-HF was added to each tube before incubating again (37°C, 1 h). The cleavage was inactivated by putting samples in +80°C for 20 min before storing them in -20°C.

A4.4 DNA gel of cleaved plasmid DNA

A 2% agarose gel was prepared by mixing 1 g agarose (SeaKem GTG Agarose, Lonza BioScience) with 50 ml 1x TAE before microwaving until the agarose was dissolved. Due to evaporation during microwaving, gel was filled up to 50 ml with 1x TAE buffer, the sample was cooled down before adding 5 µl GelRed (Biotium). Gel was casted with tray (B1A-12, 1.5 mm, 1 well ≈ 30 µl) while 2 cleaved plasmids from **A4.3** (one with insert and one with target plasmid) was mixed with 10 µl 6x DNA Gel Loading Dye (Thermo Scientific). Each dyed sample (Total 60 µl in each) was divided into two wells and GeneRuler 1 kb DNA

Ladder (Thermo Scientific) was added to one well before running gel (140 V, 40 min). With an UV lamp, the band for target vector and insert was cut out from gel, each piece was weighed before being stored in -20°C.

The procedure was repeated but with only the cleaved insert (G-CT) as the sample.

A4.5 Gel extraction of cleaved DNA

Cleaved DNA from four pieces of DNA gel (**A4.4**) was extracted using QIAquick Gel Extraction Kit (250), following protocol QIAquick Gel Extraction Kit, with point 6 omitted. 30 µl StH_2O was used to elute cleaved DNA from each spin column.

The extracted DNA concentrations was determined as in **A4.2**. Samples stored in -20°C.

A4.6 DNA ligation of cleaved DNA

DNA ligation was performed with two reactions (Target vector and insert in each) and two negative controls (No insert). The insert was added in 16x access compared to target vector which was 25 ng in each reaction. Target vector and insert (Not in controls) was mixed with 2 µl 10x DNA Ligase Buffer (NEB) (C_{Final} = 1x), 1 µl T4 DNA Ligase (NEB) and filled up to 20 µl with StH_2O. Each reaction was incubated O/N (~ 20h) in +4°C before being stored in -20°C.

A4.7 Transformation of *E. coli* by heat shock with ligated DNA

Transformation was repeated as in **A1.2.1** with the two ligated plasmids and two negative controls from **A4.6** with TOP10 cells. Dilution before transformation was carried out by mixing 7.5 µl of ligation mixture with 0.5 µl StH_2O.

A4.8 O/N cultures of target vector

O/N cultures were prepared as in **A1.2.2** but with 25 ml sterile flasks, 8 colonies from first transformed ligated DNA and 2 from the second (**A4.7**).

A4.9 Isolation of target plasmid DNA

To isolate target plasmid DNA and measure DNA concentration from the 10 O/N cultures (**A4.8**), **A4.2** was repeated with 5 colonies at a time, while the remaining samples were kept in -20°C.

A4.10 Sequencing of target plasmid DNA

Eight colonies (Clone 1-7 & 10) from **A4.9** were sent for sequencing at Eurofins Genomics. 7.5 µl of each of clone were diluted with 7.5 µl StH_2O to C_{Final} = 50-100 ng/µl. 2 µl of a forward primer (WiMa33s 10 µM) was added to each tube before sending them for DNA sequencing. Sequencing results were processed with CLUSTAL O(1.2.4) multiple sequence alignment tool and sequences for target plasmid (G5-CT) and G-CT (i.e., insert).

A4.11 2.14.12 Transformation of *E. coli* by heat shock with Clone 3 of target plasmid

Transformation was repeated twice as in **A1.2.1** with clone 3 from **A4.9**.

A5. Chromatogram IMAC 1 of second G-CT purification

Figure A1 shows two peaks during IMAC 1, second G-CT purification, one in wash 4 (75 mM Imidazole) and one expected during Elution 1 (300 mM Imidazole). The peak in wash 4 indicates that some target fusion protein is being diluted prematurely in wash 4.

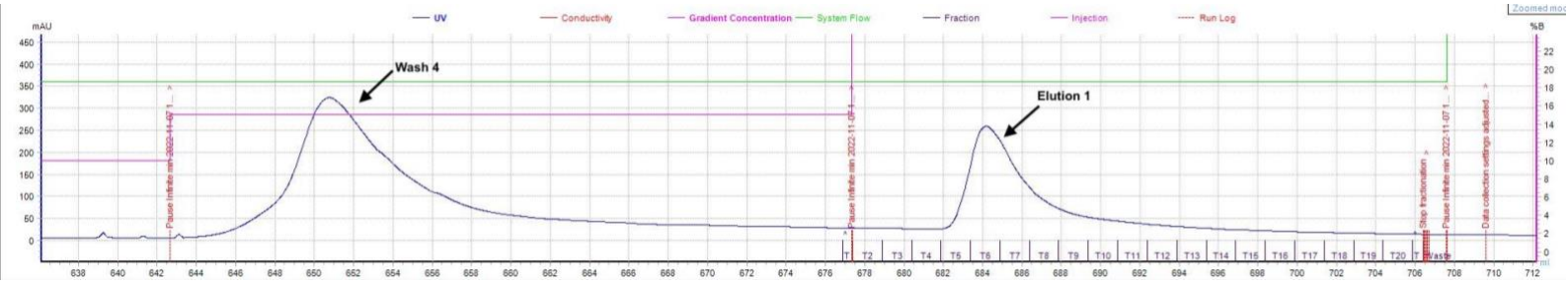


Figure A1. IMAC 1 chromatogram of second G-CT purification. Two distinctive peaks in wash 4 (75 mM Imidazole) and elution 1 (300 mM Imidazole), which probably contain target fusion protein HTH-G-CT. Fractions T5-T10 were pooled before cleavage and dialysis.

A6. Resulting sequence of Clone 3 of G5-CT

Resulting sequence of clone 3 of G5-CT from Eurofins. The sequence of clone 3 has been aligned with sequence of G-CT and theoretical sequence of G5-CT plasmid. All three sequences are perfectly aligned.

```
pT7HisTrxHis-TEVp-G5-CT_EA_20220929      CATCATCATCATCATCATATGGCTAGCAGCGATAAAATTATTCACCTGACTGACGACAGT     60
Clone3                                   CATCATCATCATCATCATATGGCTAGCAGCGATAAAATTATTCACCTGACTGACGACAGT     60
G-CT                                     ------------------------------------------------------------     0

pT7HisTrxHis-TEVp-G5-CT_EA_20220929      TTTGACACGGATGTACTCAAAGCGGACGGGGGCGATCCTCGTCGATTTCTGGGCAGAGTGG    120
Clone3                                   TTTGACACGGATGTACTCAAAGCGGACGGGGGCGATCCTCGTCGATTTCTGGGCAGAGTGG    120
G-CT                                     ------------------------------------------------------------     0

pT7HisTrxHis-TEVp-G5-CT_EA_20220929      TGCGGTCCGTGCAAAATGATCGCCCCGATTCTGGATGAAATCGCTGACGAATATCAGGGC    180
Clone3                                   TGCGGTCCGTGCAAAATGATCGCCCCGATTCTGGATGAAATCGCTGACGAATATCAGGGC    180
G-CT                                     ------------------------------------------------------------     0

pT7HisTrxHis-TEVp-G5-CT_EA_20220929      AAACTGACCGTTGCAAAACTGAACATCGATCAAAACCCTGGCACTGCGCCGAAATATGGC    240
Clone3                                   AAACTGACCGTTGCAAAACTGAACATCGATCAAAACCCTGGCACTGCGCCGAAATATGGC    240
G-CT                                     ------------------------------------------------------------     0

pT7HisTrxHis-TEVp-G5-CT_EA_20220929      ATCCGTGGTATCCCGACTCTGCTGCTGTTCAAAAACGGTGAAGTGGCGGCAACCAAAGTG    300
Clone3                                   ATCCGTGGTATCCCGACTCTGCTGCTGTTCAAAAACGGTGAAGTGGCGGCAACCAAAGTG    300
G-CT                                     ------------------------------------------------------------     0

pT7HisTrxHis-TEVp-G5-CT_EA_20220929      GGTGCACTGTCTAAAGGTCAGTTGAAAGAGTTCCTCGACGCTAACCTGGCCGGTTCTGGT    360
Clone3                                   GGTGCACTGTCTAAAGGTCAGTTGAAAGAGTTCCTCGACGCTAACCTGGCCGGTTCTGGT    360
G-CT                                     ------------------------------------------------------------     0

pT7HisTrxHis-TEVp-G5-CT_EA_20220929      TCTGGCCATGCGGCCGCGCACCATCATCATCATCATGAAAACCTGTACTTCCAGGGTGGT    420
Clone3                                   TCTGGCCATGCGGCCGCGCACCATCATCATCATCATGAAAACCTGTACTTCCAGGGTGGT    420
G-CT                                     ------------------------------------------------------------     0

pT7HisTrxHis-TEVp-G5-CT_EA_20220929      GGCGGCGGCGCGAATTCACGCCTCTCATCGCCTTCCGCAGTATCTCGAGTTTCTTCAGCA    480
Clone3                                   GGCGGCGGCGCGAATTCACGCCTCTCATCGCCTTCCGCAGTATCTCGAGTTTCTTCAGCA    480
G-CT                                     ---------------TCACGCCTCTCATCGCCTTCCGCAGTATCTCGAGTTTCTTCAGCA     45
                                                        *******************************************

pT7HisTrxHis-TEVp-G5-CT_EA_20220929      GTTTCTAGCTTGGTTTCAAATGGTCAAGTGAATATGGCAGCGTTACCTAATATCATTTCC    540
Clone3                                   GTTTCTAGCTTGGTTTCAAATGGTCAAGTGAATATGGCAGCGTTACCTAATATCATTTCC    540
G-CT                                     GTTTCTAGCTTGGTTTCAAATGGTCAAGTGAATATGGCAGCGTTACCTAATATCATTTCC    105
                                         ************************************************************

pT7HisTrxHis-TEVp-G5-CT_EA_20220929      AACATTTCTTCTTCTGTCAGTGCATCTGCTCCTGGTGCTTCTGGATGTGAGGTCATAGTG    600
Clone3                                   AACATTTCTTCTTCTGTCAGTGCATCTGCTCCTGGTGCTTCTGGATGTGAGGTCATAGTG    600
G-CT                                     AACATTTCTTCTTCTGTCAGTGCATCTGCTCCTGGTGCTTCTGGATGTGAGGTCATAGTG    165
                                         ************************************************************

pT7HisTrxHis-TEVp-G5-CT_EA_20220929      CAAGCTCTACTCGAAGTCATCACTGCTCTTGTTCAAATCGTTAGTTCTTCTAGTGTTGGA    660
Clone3                                   CAAGCTCTACTCGAAGTCATCACTGCTCTTGTTCAAATCGTTAGTTCTTCTAGTGTTGGA    660
G-CT                                     CAAGCTCTACTCGAAGTCATCACTGCTCTTGTTCAAATCGTTAGTTCTTCTAGTGTTGGA    225
                                         ************************************************************

pT7HisTrxHis-TEVp-G5-CT_EA_20220929      TATATTAATCCATCTGCTGTGAACCAAATTACTAATGTTGTTGCTAATGCCATGGCTCAA    720
Clone3                                   TATATTAATCCATCTGCTGTGAACCAAATTACTAATGTTGTTGCTAATGCCATGGCTCAA    720
G-CT                                     TATATTAATCCATCTGCTGTGAACCAAATTACTAATGTTGTTGCTAATGCCATGGCTCAA    285
                                         ************************************************************

pT7HisTrxHis-TEVp-G5-CT_EA_20220929      GTAATGGGCTAAGCTTCTCGAGGGCCTCCCAGGCCGGCGCGCCCCACCGCTGAGCAATAA    780
Clone3                                   GTAATGGGCTAAGCTTCTCGAGGGCCTCCCAGGCCGGCGCGCCCCACCGCTGAGCAATAA    780
G-CT                                     GTAATGGGC---------------------------------------------------    294
                                         *********

pT7HisTrxHis-TEVp-G5-CT_EA_20220929      CTAGCAT          787
Clone3                                   CTAGCAT          787
G-CT                                     -------          294
```

Figure A2. Resulting sequence of Clone 3, aligned with sequences for G-CT (insert) and theoretical sequence for G5-CT, named pT7HisTrxHis-TEVp-G5-CT_EA_20220929 in figure.