Research Project Molecular Life Sciences:
Structure determination of a full-length glucansucrase

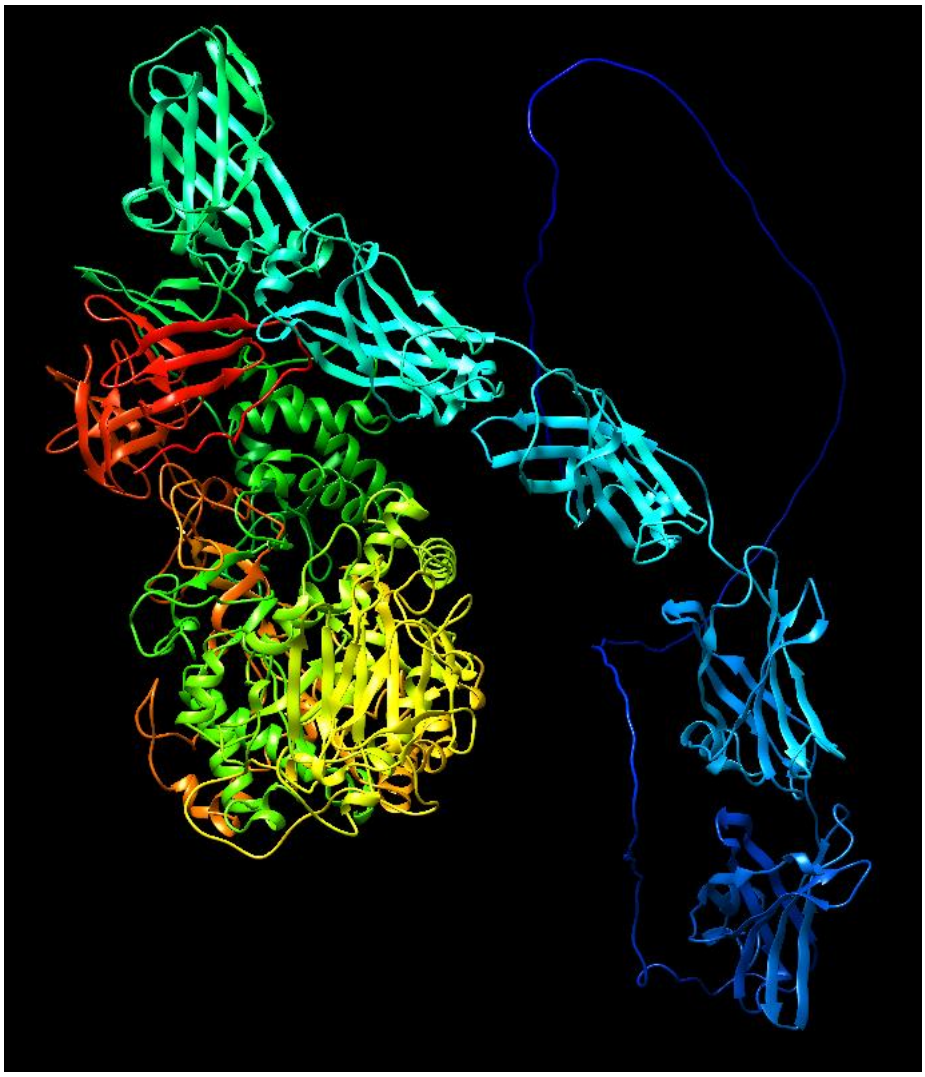


Figure 1: Alphafold model of *L. reuteri* Gtf180
(Jumper et al., 2021)

By: Tim Jonker
s4511964
Supervisors: dr. Pijning and prof. dr. Guskov
Date: 05/06/23

Abstract:

Glucansucrases (GH70) are a class of bacterial enzymes that break down sucrose to create α-linked glucose polymers. The polymers can be used by the bacteria to create biofilms like dental plaques. The products of such enzymes also have economic value. Either inhibiting or stimulating this enzyme would therefore be something of use to society. Crystallisation of full-length *Lactobacillus reuteri* Gtf180, a glucansucrase, has been successful before, but only truncated x-ray structures have been solved to date. This limits the understanding of the way the native enzyme works to the catalytic site. The termini (especially the N-terminus) have been hypothesized to bind the α-linked glucose polymers (α-glucans) as the chain still grows. It has also been theorized that one of the domains in truncated structures, the V-domain, might display a degree of flexibility. This might play a role in the functioning of the enzyme as well. We have attempted to solve the full-length structure of this enzyme, in order to better understand its mode of function. Furthermore, we attempted to observe flexibility of the V-domain in cryo-EM and attempted to determine which conditions are needed to crystalize Gtf180. In order to do this, we purified the protein, performed cryo-EM and crystallized the protein. The cryo-EM results were rather disappointing, with a mere 6.18 Å resolution. The crystallization experiments were promising with some interesting results. Crystals big enough for analysis were acquired, together with a type of crystal that has not been observed for this protein before. Data collection will take place in the near future.

Introduction:

Glucansucrases, also known by their class GH70, are a class of bacterial enzymes that catalyse the breakdown of sucrose into glucose and fructose (Leemhuis et al., 2013). According to the article the enzyme can then make polymers from these monomers that are called α-glucans and these α-glucans are reported to be excreted by the bacterium and then act as an extracellular polysaccharide or EPS for short. Many of these polymers together can form a sticky layer reminiscent of biofilm that surround the bacteria and protect them from environmental stresses, as well as helping with cell adhesion (Leemhuis et al., 2013). These α-glucans can be used for example in the food industry, as column chromatography matrices, prebiotics and potentially as blood plasma substitute and adjuvants (Wangpaiboon et al., 2018). Furthermore, the researchers write that the applications do not end with using the α-glucans themselves as product, but another one arises when one knows that bacteria often produce biofilms in areas where they give negative effects on human health like with dental plaques, caused by Streptococcus mutans (Leemhuis et al., 2013).

Studies have shown that gene knockouts of glucansucrases in *Lactobascillus reuteri (L. reuteri)*TMW1.106 lead to less effective formation of biofilms (Walter et al., 2008). So, this means even though biofilms are still formed (meaning formation of α-glucans is not the only culprit), the formation of biofilms is not as bad. So rather than producing the α-glucans as a possible application, the inhibition of the production of α-glucans might also be interesting. In order to inhibit an enzyme, one is greatly helped by knowing the structure of the enzyme, and especially knowing the dimensions of the catalytic site would be important for that, since this is the place where any competitive inhibitors would have to bind. In the *L. reuteri* Gtf180 glucansucrase this active site contains three important residues, aspartate as nucleophile, glutamate as general acid/base and aspartate as transition-state stabilizer, which is similar to GH13 class amylases (Vujičić-Žagar et al., 2010).

However, in the case of *L. reuteri* Gtf180 it has been hypothesised that the termini also play an important role in the proper functioning of the enzyme since it is believed that these domains, play a role in binding the α-glucan polymer while new monomers are still being added (Vujičić-Žagar et al., 2010). The article states that on the termini, especially on the long N-terminus which is predicted to contain a chain of β-solenoid domains, contain α-glucan binding sites. The main problem here is that available crystal structures up until this point have all been truncated at the N and C termini. This means that there is no crystal structure available of the complete protein in its full native form containing the termini, and so whether the N-terminus actually has the function of binding the α-glucan chain remains to be proven. As can be seen in figure 1, prediction models like Alphafold are not yet capable of predicting every all structures accurately. This can be seen by the blue appendage in the Alphafold model that is without secondary structure (Jumper et al., 2021). Obtaining a full-length

structure with x-ray crystallography and cryo-EM would therefore give insight in the structure of the termini (especially important being the N-terminus) and in that way improve our understanding of the modus operandi of the protein and provide more information on possible ways to inhibit or improve its function.

Furthermore, different x-ray crystalstructures of the Gtf180 protein of *L. reuteri* showed different positions of domain V relative to the rest of the enzyme (Pijning et al., 2014). The article also shows that the N-terminus, which might be important for binding the growing α-glucan chain as mentioned before, is bound to this V-domain. It has been theorized in the article that the multiple conformations of the V-domain might imply flexibility of conformation for the domain in solution. In the article this movement is hypothesized to possibly be up to 20 Å. If the V-domain moves, then logically the N-terminus moves as well. If this is where the α-glucan chain is bound, then any possible flexibility of the V-domain might have the function of bringing the α-glucan chain from and to the active site of the protein. It is in this active site that the α-glucan chain acts as an acceptor for any glucose monomers that are freed from sucrose (Vujičić-Žagar et al., 2010). With the current information it cannot be said with certainty whether the changes in conformation of domain V can be attributed to the crystals or to actual native intrinsic flexibility. In cryo-EM this problem is not present since the sample consists of a solution that is flash frozen. If different conformations exist in solution, these conformations should be preserved in the sample and can be detected in cryo-EM (Benjin & Ling, 2019). If this flexibility can be shown in solution, it would give more insight into the mechanism by which the enzyme works.

For this research project the specific glucansucrase, of which the structure was attempted to be solved, was the aforementioned *L. reuteri* Gtf180. This glucansucrase can make both α-1-3 and α-1-6 linkages. As mentioned before, crystallization of this protein in full-length has been successful before, but the crystals did not diffract sufficiently to determine a high-resolution structure of the protein. In this project the same thing was attempted using the findings from the previous time. Truncated structures are available with already gives an idea about the expected result. In order to be able to determine the structure, this protein first had to be obtained from *Escherichia coli (E. coli)* BL21 DE3 Star cells, then purified and then crystallised and analysed. The cells were taken from a 10-year-old glycerol stock, which was made after previous experiments. To elaborate a bit more on x-ray crystallography it would be noteworthy to mention the basic mechanism by which it works.

In order for crystallisation to occur, the protein sample must be high in concentration, but still be soluble and be homogenous (Smyth, 2000). Yet the article states that this does not mean that crystallisation will just start spontaneously if the basic conditions are met, since all proteins have different conditions in which they form crystals and with our current knowledge it is almost impossible to predict which conditions will work for which protein. Thus, they write, the only way is trial and error by varying different conditions like temperature, type of precipitant and its concentration, pH, protein concentration and type of buffer, but also some proteins may even never crystallise. With this knowledge we, in our case, may count ourselves lucky that crystallisation of *L. reuteri* Gtf180 has been successful before (unpublished). When the crystal is formed it is taken for analysis, x-rays are shot at the crystal and then these x-rays are scattered by the proteins in the crystal and can be caught by a detector (Smyth, 2000). The authors write that after each burst the crystal is turned around so in time a 3D structure can be determined. The raw data from the detector are then used to determine a visual 3D structure of the protein.

For cryo-Electron Microscopy (cryo-EM) the principle is a bit different. For this no crystals are needed and so it is a good alternative for x-ray crystallography in the unfortunate case when crystals cannot be formed (Benjin & Ling, 2019). According to the article its principle does not rely on x-rays, but the use of the wave properties of electrons; by varying the speed of the electrons their wavelength and thus the maximum resolution can be varied and when the electrons pass through the sample they too are scattered and a detector catches them. Just like the crystal in x-ray crystallography the sample is twisted around (Smyth, 2000), but only with a maximal tilt angle of 70 degrees in both directions (Benjin & Ling, 2019). The authors state that since cryo-EM works with material particles with a charge it is easy to focus the electrons into a lens and also because of this, the phase can be recovered. Both are not possible with x-ray crystallography since x-rays are photons and therefore have neither charge nor mass (Smyth, 2000). However, even though cryo-EM does not require crystallisation, the preparation protocol is challenging (Benjin & Ling, 2019). As the word cryo

already indicates, the researchers state that sample preparation requires flash freezing with liquid ethane on very small grids. The article however also shows that there are downsides to the eventual image which are mainly low contrast and low signal-to-noise ratio. Samples must furthermore be very thin since electrons are material particles and will not easily pass through the samples like x-rays (Benjin & Ling, 2019).
The downside of x-ray crystallography in this regard is that crystallisation forces the protein into a certain conformation that is both static and homogenous throughout the entire crystal (Smyth, 2000).

In order to perform these techniques, the protein first has to be acquired and purified. To acquire all the data the following steps were followed. Cells were cultured, expression was induced, and purification was performed. For the purification Immobilized Metal Affinity Chromatography (IMAC) was used first which relies on binding of a His-tag to metal-ions of the column (Chang et al., 2017). Then Anion-Exchange (AIEX) which is based on separation by binding to a column based on charge and the use of the iso-electric point (Lee, 1990). In the end size exclusion chromatography (SEC) was performed, which separated proteins based on size, in which bigger proteins elute first (Hong et al., 2012). Dynamic light scattering (DLS) was performed to confirm purity of the sample, resting on the principle that different size molecules scatter light in different ways based on speed of Brownian motion of the particles (Falke & Betzel, 2019). Eventually crystallisation experiments were performed, and cryo-EM was performed. Additionally, an activity assay was performed.

Eventually both x-ray crystallography and cryo-EM techniques were used to create a conclusion and answer the research questions. What are conditions to crystalize Gtf180? What is the full 3D structure of *L. reuteri* Gtf180? To what extent can conformational flexibility of the V-domain be observed in solution using cryo-EM?
The hypotheses to these questions were as follows. That the conditions to crystalize Gtf180 would be similar to those that were successful before, with possibly other conditions being discovered. That the 3D structure would be somewhat similar to the Alphafold model (figure 1) and the truncated structure (figure 15), with the fully solved termini included. Finally, the hypothesis of to what extent conformational flexibility of the V-domain can be observed in cryo-EM, the hypothesis that previous research strongly indicates that this flexibility does indeed exist (Pijning et al., 2014), led us to us believe that we would be able to observe this flexibility in cryo-EM.

Materials and methods:

Buffers and solutions:

Stock solutions and Luria Bertani (LB) medium were prepared. The stock solutions were: 50 mL 1M imidazole, 50 mL 1 M HAc/NaAc pH 5.5, 150 mL 1 M NaCl, 5 mL 1 M K-phosphate pH 8.0 and 10x 0.5 mL 1 M Dithiothreitol (DTT) (total volume 5 mL). The LB medium was divided in portions of 2x 1000 mL and 2x 20 mL and all portions were autoclaved. The tubes containing the DTT solutions were put in the freezer at -21 degrees Celsius. Furthermore, pipette tips, Eppendorf tubes, 2 Greiner tubes, 2 5L Erlenmeyer flasks and 2 100 mL Erlenmeyer flasks were autoclaved. The following buffers were prepared: Cell resuspension buffer (CRB), Elution buffer (EB), anion-exchange-A (AIEX) buffer, AIEX-B buffer, size exclusion chromatography (SEC) buffer and protein storage buffer. CRB contained 50 mM K-phosphate pH 8.0, 300 mM NaCl, 10 mM imidazole and 100 mL was made. EB contained 50 mM K-phosphate pH 8.0, 300 mM NaCl, 200 mM imidazole and 50 mL was made. AIEX-A buffer contained 25 mM Tris-HCl pH 7.5, 1 mM $CaCl_2$ and 500 mL was made. AIEX-B buffer contained 25 mM Tris-HCl pH 7.5, 1 mM $CaCl_2$, 1 M NaCl and 500 mL was made. SEC buffer contained 25 mM HAc/NaAc pH 5.5, 150 mM NaCl, 1 mM $CaCl_2$ and 500 mL was made. Protein storage buffer contained 25 mM Tris-HC pH 7.5, 1 mM $CaCl_2$, 150 mM NaCl and 10 mL. AIEX-A, AIEX-B, and protein storage buffer were all three brought to pH 7.5 using hydrochloric acid. AIEX-A, AIEX-B and SEC were filtered and put into new flasks. Furthermore, 4x 0.5 mL 0.25 M Isopropyl β-D-1-thiogalactopyranoside (IPTG) solution (total volume 2 mL) was prepared, as well as 5x 1 mL 35mg/mL carbenicillin (1000x) (total volume 5 mL). Both solutions were sterilized using a syringe with a filter and put in the freezer at -20 degrees Celsius.

Growth and expression:

A preculture was started. This was done by pouring 20 mL of sterile LB medium into sterile Erlenmeyer flasks (100 mL). 20 µL of 35 mg/mL carbenicillin was added to each of the Erlenmeyer flasks. Inoculation was done using sterile prickers with colonies from LB ampicillin plate that was streaked with *E. coli* BL21 DE3 Star strain. The precultures were grown in the shaking incubator at 30 degrees Celsius overnight. The cultures were taken out and were each put into a different 1 L fresh LB medium with 1 mL carbenicillin added in a sterile 5 L Erlenmeyer flask. They were put into the incubator at 37 degrees Celsius at 170 rpm. After 1 hour the OD600 was measured. The cultures were grown for a further 20 minutes at the same settings until OD600 = 0.6. The cultures were cooled and 400 µL IPTG solution was added to both cultures, in order to induce expression of the plasmid containing the target protein. The cultures were put back into the incubator at 17 degrees Celsius at 150 rpm. The 2x 1 L cultures were taken out of the incubator after 20 hours and put on ice. 2x 100 µL was stored in the freezer at -20 degrees Celsius one from each Erlenmeyer. The OD600 was measured of each Erlenmeyer. The cultures had an OD600 of 2.92 and 2.881. The cultures were divided over 3 centrifuge bottles. The cultures were centrifuged at 5150 rpm and at 4 degrees Celsius. A pellet was acquired in all three of the bottles. The supernatant was drained, and the pellet was scooped out and put into 50 mL tubes. The cells were weighed and stored at -20 degrees. Two Greiner tubes were stored this way and contained 7.849 and 4.441 grams of pellet coming to a total of 12.29 grams of cells.

Cell breaking:

70 µL β-mercaptoethanol was added to the CRB leading to a final concentration of 1 mM. Henceforth the CRB was kept cold, in either ice or in a cold room. The pellets were resuspended in CRB to which 50 mL was taken from the stock and to which a spatula tip of DNase was added. The amount of mL to be added to each pellet was calculated by multiplying the weight of the cells by 3. So, for the tube with 7.849 grams 23.5 mL was used and for the tube with 4.441 grams 13.323 mL was used. Resuspending was done using up and down pipetting. The cells were broken using the Maximator which uses pressure to lyse the cells. For this 2000 bar was used. Due to inexperience some of the resuspended pellet ended up going through the Maximator and into the waste bottle. The amount lost to this mistake was estimated at 5 mL. The samples were centrifuged at 4 degrees Celsius for 30 minutes at 15000 g. The supernatant was at this point a cell free extract. 10 µL of the CFE was stored at -20 degrees Celsius for later SDS-PAGE. The rest was stored on ice in two Greiner tubes.

IMAC:

Regenerated 2 mL Ni-NTA slurry was put into two columns and stored in 20% ethanol. 2 mL Ni-NTA slurry was washed in CRB. The Ni-NTA slurry turned brown when CRB was added. It was supposed that this was because the β-mercaptoethanol reduced the nickel. It was supposed that this was because of the use of regenerated Ni-NTA slurry and not fresh. 2 mL of fresh Ni-NTA slurry was put into 2 columns. CRB was added for washing. This time only light brown colouration occurred, but it was decided to carry on. The washed slurry was added to the two CFE tubes equally. It was left to incubate on the roller bank in the cold room (4 degrees Celsius) for 90 minutes. The contents of the tubes were poured into the columns. The contents of each tube into a different column. Flowthrough was collected, which contained all proteins that did not bind the Ni-NTA and thus are supposed to have no His-tag. Washing was done with 2 x 5 mL CRB for each column. Flowthrough was collected. Elution steps were done with Elution buffer to which 35 µL β-mercaptoethanol was added. The elution steps were done in steps of 0.5 mL. Each step was collected in separate Eppendorf tubes. The A280 was measured using a Nanodrop (results in result section) and based on these fractions 3 and 4 were selected and pooled. This led to a total volume of 2 mL divided over 2 Eppendorf tubes. The IMAC pools were centrifuged for 10 minutes at 13000 rpm. The IMAC pools were then added to a vivaspin 4 turbo 50 kDa column and concentrated at 3900 g. The column was subsequently with AIEX-A to remove NaCl from the pool, this was done 3 times and the final volume was 500 µL. The vivaspin

column was washed with 200 µL AIEX-A buffer and the volume was added to the final volume making it 0.700 µL. A280 was taken using Nanodrop. At this point the pool was very milky which indicates aggregation. The pool was centrifuged at 13000 rpm for 10 minutes. The A280 supernatant was taken again, and it was concluded that only a little bit of protein was lost. The supernatant was transferred to a new Eppendorf tube and the pellet was frozen for later SDS-PAGE analysis. 10 µL supernatant was taken and stored at -20 Celsius for the SDS-PAGE.

Anion Exchange:

AIEX was performed on a BioRad NGC system with a Resource Q column. The use of this system would make it easier to create a salt gradient, than if it were done by hand. Two buffers were used AIEX-A and AIEX-B to create a salt-gradient for elution. Fraction size of 1.5 mL was used and a flow rate of 3 mL/min. First a test run was done using 500 µL of 100x diluted IMAC pool. AIEX was performed three times with real sample with 300 µL and 195 µL. The A280 of the chosen fractions was measured (8, 9 and 10). The fractions (3-fold 8, 9 and 10) were combined in a vivaspin 4 turbo 50 kDa column and concentrated to a final volume of 457 µL. This volume was mixed using up and down pipetting and was transferred to a new Eppendorf tube. 20 µL SEC buffer was used to flush the vivaspin column after the final volume was transferred to a new Eppendorf tube. This led to a total volume of 477 µL. Nanodrop was used to take the A280. 1 µL was taken and diluted 10x for SDS-PAGE analysis.

Size Exclusion chromatography:

The rest of the fractions were kept on 4 degrees Celsius. SEC was performed with a BioRad NGC system with a Superdex 200 GL 10/300 Increase column. This system can provide more pressure that if this were done by hand, saving time. As a fraction size 0.25 mL was chosen with a flow rate of 0.45 mL/min. This procedure was performed three times. First a test run with a 50x diluted sample was done and next the sample was loaded in two parts and the procedure was run twice. The A280 of every pool was taken after centrifuging at 13000 rpm for 10 minutes. The SEC pools were concentrated with vivaspin 4 turbo 50 kDa. Protein was henceforth stored at 4 degrees Celsius. Two peaks were seen where only one was expected. A bit of protein C was diluted and reinserted into the BioRad NGC system and SEC was performed again. This was to test whether the second peak that is contained in protein C would also elute at the first peak. This would indicate that it is simply a different conformation of the same protein.

Dynamic light scattering:

Dynamic light scattering was performed on two 25x diluted fractions corresponding to the two peaks from the test run to see if these peaks were likely the same protein, but in different conformation states. Beforehand the protein solution was centrifuged at 13000 rpm for 10 minutes. Dynamic light scattering was performed at 20 degrees Celsius. Dynamic light scattering was also performed at 20 degrees Celsius for both protein A and C.

Cryo-EM:

Cryo-EM grids were frozen. For Cryo-EM two protein dilutions of protein solution A were used. The dilutions were 2 mg/mL and 4 mg/mL, and these were done in duplicates, so 4 grids were frozen. A blotting time of 4 seconds was used. Blob particle selection was used for the imaging. The imaging was done on a 200 kV electron microscope.

SDS-PAGE:

Sodium Dodecyl Sulfate Poly Acrylamide Gel Electrophoresis (SDS-PAGE) was performed on 8 samples: Cell-free extract (10x), IMAC 2, IMAC 3/4 (10x), IMAC 5, AIEX (100x), protein solution A, protein solution B, protein solution C. It should be noted that IMAC fraction 5 when taken out of the -

20 degrees storage contained a turbid film on the edges of the Eppendorf tube. Vortexing did not lead to resuspension and when pipetting was attempted the entire film stuck around the pipet tip. This was the case in both tubes that contained IMAC 5, and this was not the case in any other fractions. It is not unlikely that this film consisted of proteins that precipitated out of the solution for some reason. The gel was a standard SDS-PAGE 12% consisting of polyacrylamide with the addition of Ammonium Persulfate and SDS (both 10%) (Al-Tubuly, 2000). The stacking step was run on 140 volts. The actual running was performed at 180 volts. The running was performed in 10x diluted SDS buffer. After running the gel was stained using staining solution containing Coomassie blue for 90 minutes on a shaker. Afterwards the gel was de-stained using MilliQ overnight on the same shaker.

Activity assay:

An activity assay was performed to further confirm whether the two peaks indeed contained the same protein with the same activity (undamaged/uncompromised). The substrate for the protein is sucrose as mentioned in the introduction. For this SEC pools A, B and C were taken, as well as fractions 12 and 13. These were all diluted to the same concentration of 2.2 mg/mL. At t=0 20 µL protein dilution was added to 400 µL SAC solution (100 mM sucrose, NaAc pH 4.7, 1 mM $CaCl_2$) which was pre-heated to 37 degrees Celsius. While incubating the sample was continuously put on 37 degrees Celsius. At the timestamp when a measurement was to be taken 10 µL was taken out of the tube and put in a new tube which contained 10 µL DNS and was immediately put into a water bath of 100 degrees Celsius for 10 minutes. DNS would react with the reducing end of sugars to produce a red colour (wavelength 540 nm) and thus would indicate activity of the protein. Measurements were taken at t=2, t=5 and t=10. Afterwards the UV-Vis was measured at 540 nm (Gusakov et al., 2011).

Crystallization:

In total 6 screens were performed on 96 well plates with 2 drops per reservoir (different fractions). The reservoirs were filled with 50 µL pre-made screen solutions. The names of these screens were: LMB, BCS, JSCG+, Ligand friendly, PACT and SG2. The drops were made by a machine called mosquito. The drop size was 200 nL with 100 nL protein solution and 100 nL screen solution. For all conditions both a drop of protein solution A and C were made. All screens were left at room temperature. In the optimalisation plates different protein concentrations were used ranging from 3-6 mg/mL in increments of 1 and done in duplicates per plate for both protein solutions A and C. The conditions were based on the successful conditions from the screens (ones that formed crystals). These conditions were the use of polyethylene glycol (PEG) 3000 as precipitant and sodium acetate (NaAc) buffer (pH 4.6) and the other $NH_4H_2PO_4$ as precipitant and Sodium citrate as a buffer (pH 5.3). Horizontally on the plate precipitant concentration was increased to the right while the pH was kept constant at the original levels of 4.6 and 5.3. The reservoirs were filled by a machine called dragonfly, which took the original stock solutions concentrations and on command calculates the dilutions to create a gradient and the pH of the solution by giving it the desired gradient. The drops were made by the mosquito machine in exactly the same fashion as the previous plates. The plates were put away at room temperature. Based on the observations from the previous screens it was concluded that the best conditions were those with $NH_4H_2PO_4$ as precipitant and Sodium citrate as a buffer with an ideal protein concentration at 6 mg/mL. It was decided to make a plate in which the pH was varied horizontally (half of all wells were used) over 6 wells and the precipitant concentration was varied vertically over 8 wells. Using the conditions that were determined to be best ($NH_4H_2PO_4$ as precipitant and Sodium citrate as a buffer with a protein concentration at 6 mg/mL) hanging drop plates were made in which reservoirs were filled with 0.5 mL conditions. On top of the reservoirs glass plates were put and to make sure the reservoirs were airtight, adhesive oil was used to line the edges of the reservoir. On the glass plate a drop was placed which consisted of 1 µL conditions and 1 µL protein solution. The drop is thus on the inside of the glass and hangs hence the name hanging drop plates. This was done for protein A. The range of the precipitate was 0.8 to 1.5 M, and the pH was 5.0 in one plate and 6.0 in the other. The plates were left at room temperature. More hanging drop plates were made, this time with a different drop volume, a different pH 5.3, a different protein concentration and a slightly different range of precipitant concentration. For the drop volume 1.5 µL protein solution and

0.7 µL reservoir was used. For the protein concentration 5 mg/mL was used. The range for the precipitant concentration was 0.7 to 1.4 M. In total 4 plates were made, for protein A and C and this was done in duplicate, one duplicate at room temperature and one at 4 degrees Celsius. Again, from the results the best conditions were considered and based on this a new precipitate range was determined, this time between 1.3 and 1.65 M with increments of 0.05 M and a constant pH at 5.3. 4 plates were made in total, 2 with the original drop size and the other with a 1.5x increase drop size containing 2.25 µL protein solution and 1.05 µL reservoir solution. The plates were made with protein A and C and were left at room temperature. To prove that the crystals were not salt crystals 25x diluted Izit crystal dye was added to one of the drops with crystals. The results were that these are most likely protein crystals.

Results:

SDS-PAGE:

Despite the fact that this experiment was performed at the end of the project, it is important to report on the results from the SDS-PAGE first. This is because the results visualize the various purification steps that have been done during the project.

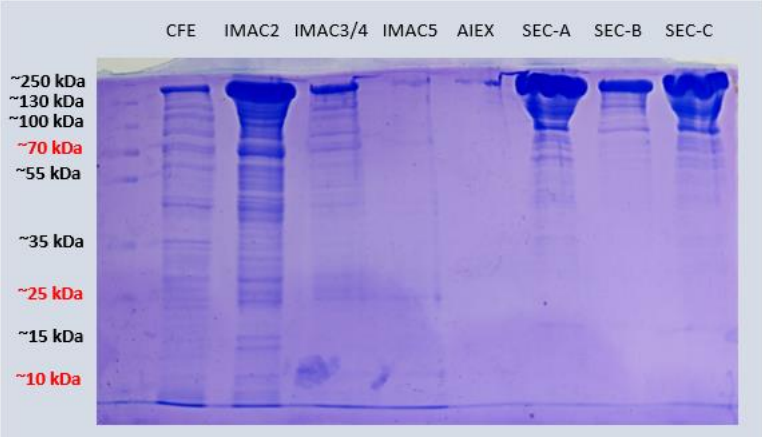


Figure 2: SDS-PAGE gel after de-staining.

The implications of these results shall be discussed in the respective sections of the samples. Something that should be mentioned is that the size of the protein is around 220 kDa and so the band containing the target will be on top of the gel.

Growth, expression and cell breaking:

This step yielded little results other than the physical starting components for the next steps. In total 12.29 gram of cells was acquired from two cultures of *E. coli* BL21 DE3 Star strain with an OD600 of 2.92 and 2.881 respectively. From this a cell free extract (CFE) was acquired by cell breaking and subsequent centrifuging. This CFE is seen in the SDS-PAGE gel. It has been diluted 10x and shows a concentrated band at the top of the gel and lighter blue bands all over the lane. This functions as the zero-point for the other lanes containing subsequent purification steps. It shows that the target protein (and/or other proteins of similar size) are present in elevated amount in the CFE and that for the rest the sample is very impure. This is to be expected since the CFE in principle contains all proteins that were present in the cytoplasm of the cells. Even when diluted 10x it shows reasonably high intensity at the expected band of the target, indicating most likely a good yield of target protein. This information is also important for indicating possible loss of target protein in the subsequent steps.

IMAC:

This was the first purification step, based on the His-tag that was added to the protein. This His-tag consists of a tail of multiple consecutive Histidine residues that bind to nickel in the nickel-NTA column (Chang et al., 2017). As seen in the SDS-PAGE gel the IMAC 2 fraction showed a large band at the expected band of the target protein and like the CFE it also shows protein spread out all over the lane indicating large impurity in the sample. The IMAC 3/4 fractions (the fractions that were pooled) were 10x diluted before being put on the gel. This means that this lane can be compared best to the CFE lane to indicate the effectivity of the purification step. It can be seen that there are still quite a lot of proteins in bands that do not correspond to the size of the target protein. This indicates that a lot of impurity remains, but there is less impurity than in the CFE which was what was expected. In the IMAC 5 fraction lane there are almost no bands to be seen. This could be due to the fact that most likely some non-target proteins remained bound to the column during the washing steps during IMAC, but were bound rather weakly so when the elution buffer was added, these proteins eluted first and that could be why they were not in the IMAC 5. However, the band where the target was expected was also much less intense than in IMAC 2. The possibility that the proteins had precipitated out of solution as described in materials and methods is also a probable explanation for the lack of proteins in this lane. In the tables below the A280 and the ratio of A260 and A280 can be found for the IMAC fractions. There were 2 Ni-NTA columns, and the fractions of each column were measured separately. For clarity: E3 and E4 were pooled.

E = IMAC

First column:

	A280	A260/A280
E1	1.98	1.71
E2	14.60	0.61
E3	45.58	0.56
E4	37.56	0.60
E5	13.34	0.72

Table 1: A280 of first column fractions

Second column:

	A280	A260/A280
E1	2.61	1.77
E2	14.51	0.60
E3	43.11	0.55
E4	39.27	0.58
E5	15.88	0.67
E6	7.20	0.81

Table 2: A280 of second column fractions

Vivaspin 4 turbo 50 kDa:

	A280	A260/A280
Before	34.06	0.51
After	51.74	0.56
After + centrifuge	50.95	0.50

Table 3: A280 of the IMAC pool before and after concentrating.

Before indicates the values of the pools before concentrating with vivaspin, after indicates after concentrating and after + centrifuge indicates the values after centrifuging out the aggregates/precipitate. As the numbers show, the centrifuging out of the aggregates does not lose a lot of protein, but it does improve the A260/A280. Vivaspin is primarily meant for concentrating and washing the fractions, but as an added function the column is permeable to particles of 50 kDa or below. This means that in this instance when done the first time it also has a purifying effect. To calculate the actual protein concentration of the sample the A280 should be divided by 1.58 for the target protein and after doing this it can be concluded that at this point after concentrating the protein concentration was 32.25 mg/mL with a volume of 0.700 mL protein solution present. This means there was 22.575 mg protein in total, but this mass does not necessarily consist of only the target protein like the SDS-PAGE also indicates.

AIEX:

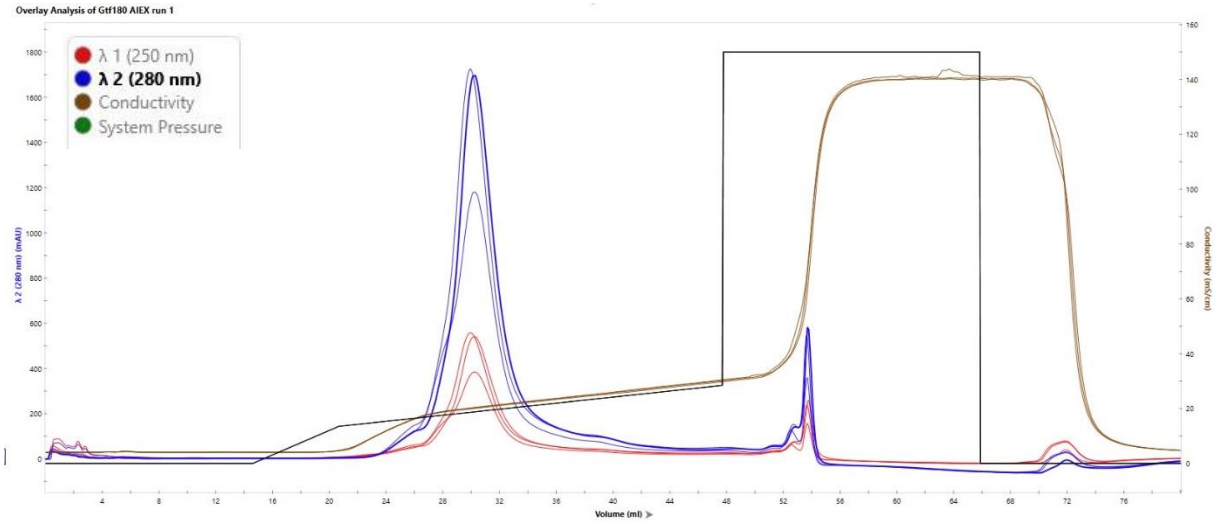


Figure 3:

AIEX runs 1, 2 and 3. Whole blue peak was pooled.

Nanodrop:

First run	A280	A260/A280	Second run	A280	A260/A280	Third run	A280	A260/A280
F6	1.61	0.49	F6	1.90	0.49	F6	1.26	0.49
F7	3.10	0.48	F7	3.05	0.49	F7	2.21	0.49
F8	1.74	0.49	F8	1.42	0.49	F8	1.24	0.49

Table 4: A280 of the AIEX fractions that were eventually pooled.

Vivaspin 4 turbo 50 kDa nanodrop:

Vivaspin	A280	A260/A280
100x dilution	0.58	0.50
AIEX-pool	47.68	0.49
Flow-through	0.26	1.13

Table 5: A280 of the AIEX pool together with the 100x dilution used for the SEC test and the flow-through of the Vivaspin 4 Turbo column.

In total 0.479 mL AIEX pool was obtained meaning that by estimation there was at this moment in time 14.45 mg of protein in the AIEX pool. The flow-through shows very little protein concentration and since the column has only limited flow-through size (50 kDa) it is virtually impossible that any intact Gtf180 was lost in the vivaspin concentration. The sizable difference between the undiluted pool and the 100x times dilution when compensated for the dilution might be, because of a pipetting mistake. The 100x dilution was made in order to do the test run for the SEC. In the SDS-PAGE gel the AIEX pool was diluted 10x twice leading to a 100x dilution. This means that the result is not representative compared to the other lanes. Despite the high dilution factor, it still shows a band where the target protein is expected to be and the rest of the lane is empty, which can be seen as a sign that in any case, there is a lot more target protein present than other non-target proteins. This is at the very least consistent with previous lanes. It can be concluded that the AIEX did not somehow cause impurity or extreme loss of target protein, even though some was lost as calculations indicate.

SEC and DLS:

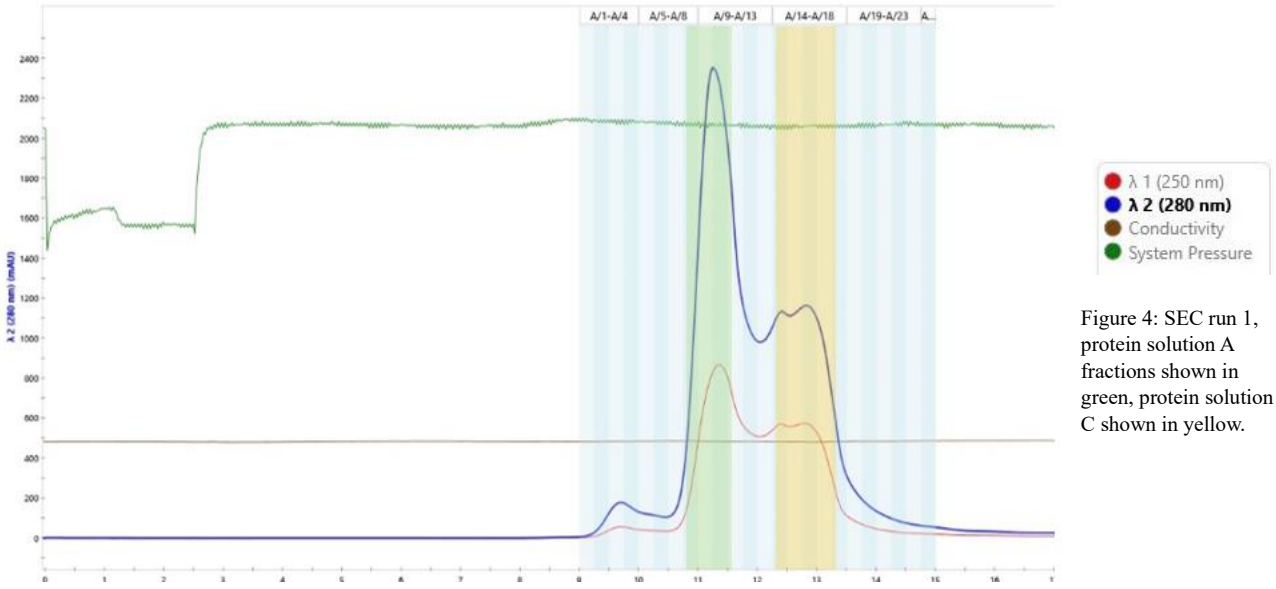


Figure 4: SEC run 1, protein solution A fractions shown in green, protein solution C shown in yellow.

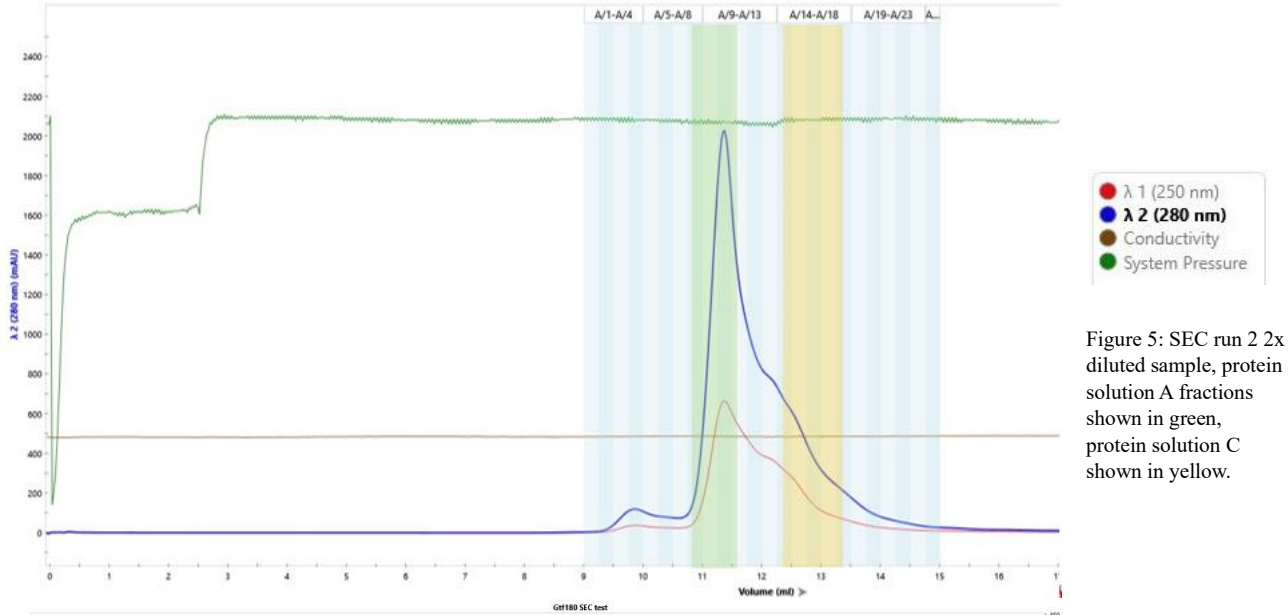


Figure 5: SEC run 2 2x diluted sample, protein solution A fractions shown in green, protein solution C shown in yellow.

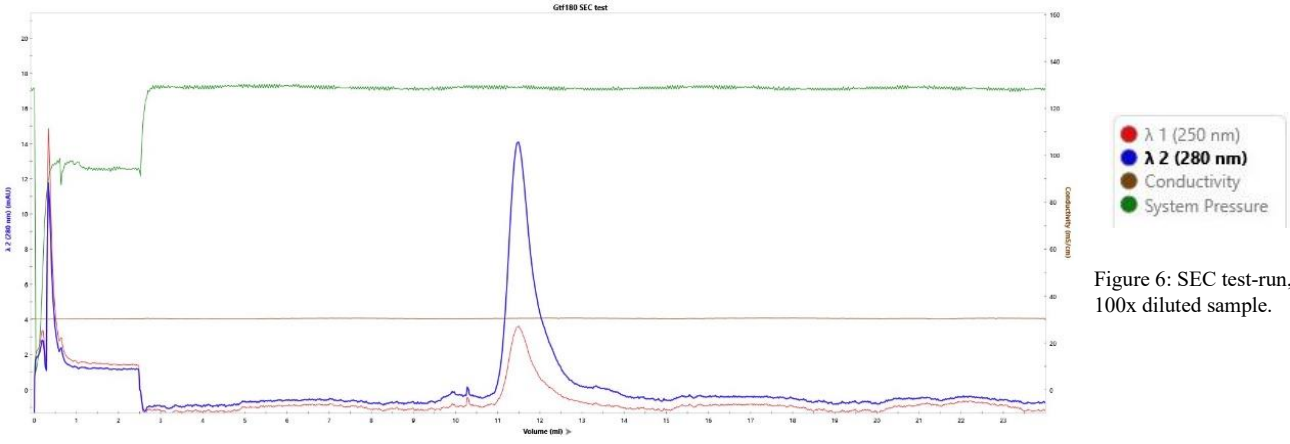


Figure 6: SEC test-run, 100x diluted sample.

	A280	A260/A280
A (S8-S10)	2.46	0.43
B (S11)	3.57	0.43
C (S14-S17)	3.48	0.43

Table 6: A280 of the SEC pools before concentrating.

The SEC being the last purification step it was expected that after this step a pure protein would be acquired. This would be indicated by good separation of any peaks that would occur in the chromatogram. This however was rather unfortunately not the case. The test run (figure 6) showed a relatively well-defined peak with only a little ‘’shoulder’’ on the right which indicated a very small impurity, but not enough to likely cause problems during crystallization experiments. The test run was diluted 100x, however, and so when the actual AIEX pool was loaded (2 runs) the results changed. The first run (figure 4) (undiluted, 0.3 mL) showed two largely unseparated peaks where one was expected, the second run (figure 5) (diluted approximately 2x, 0.3 mL) showed a more severe shoulder than the test run, but no second peak. This seems to indicate that depending on dilution factor the chromatogram shows a different shape. It was decided that 3 pools were to be made. One constituting to the first peak being fractions 8-10 (A), one to the dip in between 11 (B) and one for the second peak that could be seen in the first run 14-17 from run 1 and 14-16 from run 2 (C) this was determined based on figure 4 and figure 5.

In the final concentration only pool A and C were concentrated using vivaspin 4 turbo 50 kDa:

	A280	A260/A280	Concentration (mg/mL)	Volume (mL)
A	14.00	0.56	8.58	0.366
C	13.11	0.50	8.19	0.479

Table 7: A280 of the pools that were concentrated.

In total 7.06 mg of protein was acquired.

Dynamic light scattering (DLS) of SEC fractions

Before concentration

	Item	Time (s)	Temp (C)	Intensity (Cnt/s)	Radius (nm)	Polydispersity %Pd	MW Mw-R (kDa)	Amp	Baseline	SOS
1	SEC run 1 fraction 9	100.300	20	2096100.000	6.051	2.7	227	0.294	1.000	0.065
2	SEC run 1 fraction 15	358.500	20	1811420.000	6.233	3.3	244	0.292	1.000	0.111

After concentration

	Item	Time (s)	Temp (C)	Intensity (Cnt/s)	Radius (nm)	Polydispersity %Pd	MW Mw-R (kDa)	Amp	Baseline	SOS
1	Pool A	100.800	20	2096080.000	8.608	23.7	518	0.373	1.004	2.274
2	Pool C	1044.800	20	1897950.000	6.968	10.4	316	0.311	1.000	0.102
3	Pool A re-centrifuged	1839.700	20	1702020.000	7.468	14.2	372	0.315	1.000	0.689

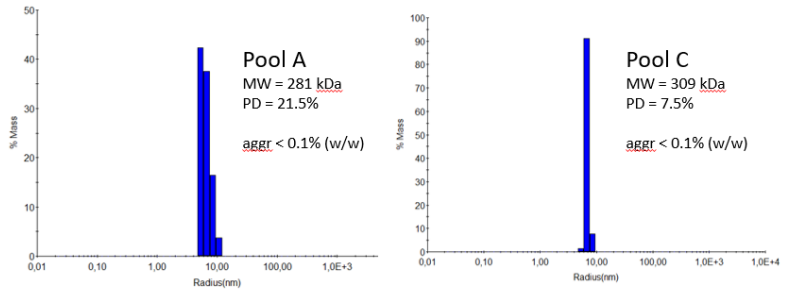


Figure 7: Dynamic light scattering results. Above the results from the separate fractions that were done to confirm that they were the sample protein. Below the results from the actual pools after concentrating, used to confirm purity.

For DLS both peaks show roughly the same results, which indicates that there is a high probability that though both peaks elute at different points, they are not different proteins. If they were different proteins the samples would show a higher polydispersity. Instead, it was most likely that something else caused part of the target protein population to elute at a different volume than expected. Since the SEC column separates on size, the most reasonable explanation for this unexpected result was that somehow a higher concentration of target protein forces part of the population in a different, more compact, conformation. This is consistent with DLS data which shows that after concentrating the polydispersity increases. This would explain why part of the protein elutes later. This was further confirmed when a 50x diluted volume of 0.5 mL of the second peak (protein pool C) was put on the

SEC column again. If this had been truly a different protein the protein should elute at exactly the same place as it did before and not at the expected volume of the target protein, but instead a similar shape was acquired as with the test run, in which there was a peak at the place where the target protein was expected to elute. This means that the protein first eluted later than expected and when reinserted eluted at the expected volume. These two findings practically confirmed that these peaks were simply the same protein. In principle this could have meant that all fractions that contained any peak material could have been safely combined, but to be sure the three pools were maintained: A, B, C as described before. SDS-PAGE gel shows that A and C contain the most protein where the target protein was expected on the gel compared to B, which is consistent with the fact that B was never concentrated. Some impurities could be seen, but compared to the target protein band this seemed negligible and would most likely be pure enough for the purposes of our experiments. So, it could be safely concluded that Gtf180 was purified successfully.

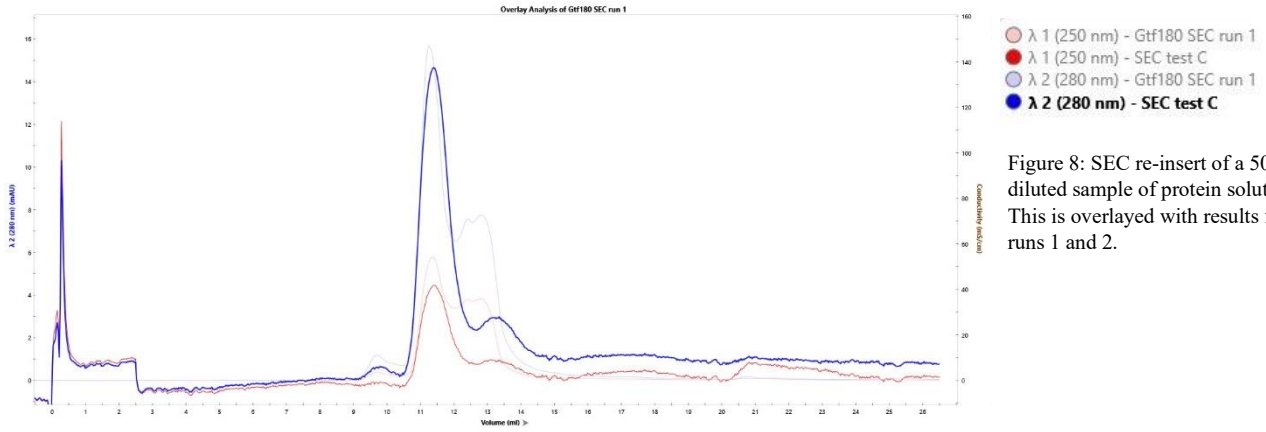


Figure 8: SEC re-insert of a 50x diluted sample of protein solution C. This is overlayed with results from runs 1 and 2.

Cryo-EM:

The results from the cryo-EM are not too promising. In the end a 3D density map with a resolution of 6.28 Å was acquired. This is a rather poor resolution. The map does show some resemblance to the central part of the enzyme, which has been well defined by the solution of truncated x-ray structures. It is possible that the N and C termini are badly visible, because they are intrinsically too flexible to get good electron densities for those parts of the protein. There is also no flexibility of the V domain visible in the data, but this might be because of a lack of good, isolated particles that could be picked for analysis. This made it rather difficult to answer the research question, because though we could not observe the flexibility of the V-domain, this might just be due to the poor quality of the data.

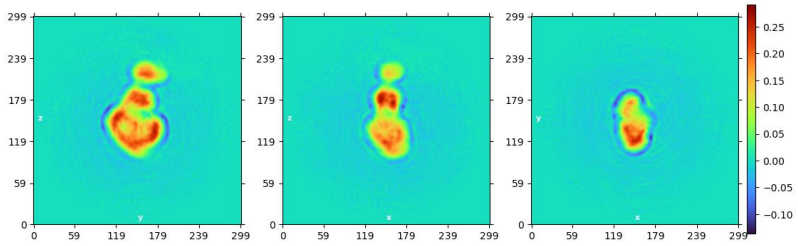


Figure 9: 3D density map of a single particle

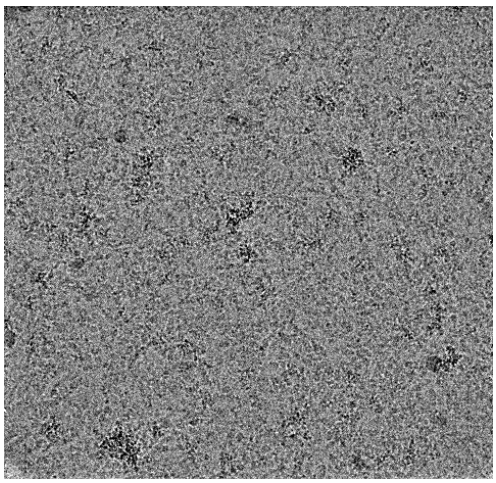
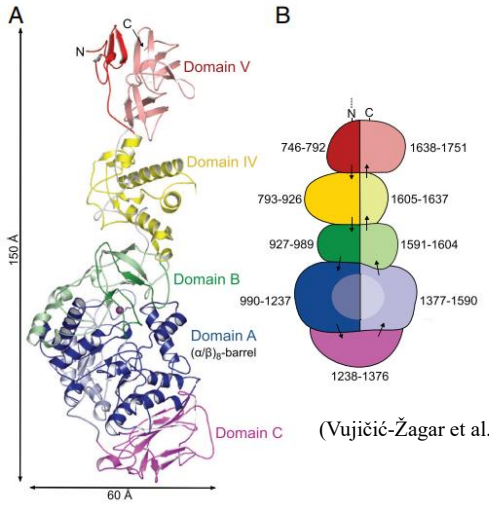


Figure 10: Cryo-EM micrograph of our sample



(Vujičić-Žagar et al., 2010)

Figure 15: For comparison with figure 9 a schematic representation of a truncated version of Gtf180 (Vujičić-Žagar et al., 2010).

Activity Assay:

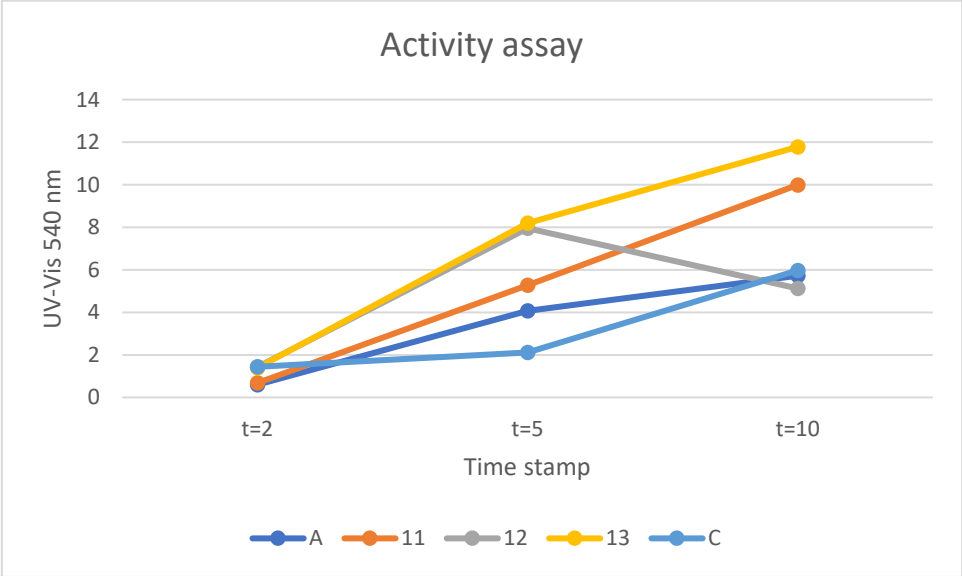


Figure 11: Absorbance at 540 nm for samples from the activity assay at different timestamps.

Absorption was expected at 540 nm, but the UV-Vis spectrum showed a peak at 370 nm, which was unexpected. Visually the results were as expected. The longer the protein solution was incubated with SAC, the redder the colour which indicates more binding to DNS as can be seen in the figure below. The UV-Vis results show this, except for fraction 12 which shows a decrease from t=5 to t=10, but this is most likely an artifact. It is however questionable how reliable the results from the UV-Vis are, since literature indicates that the peaks should be at 540 nm (Gusakov et al., 2011) where there is barely any absorption. In any case, the red colour still indicates that all samples contained an active protein that broke down sucrose into fructose and glucose. The results are however too unreliable to determine whether there is any statistically significant difference between the various samples.

Figure 12: Visual representation of red colour in samples. The Eppendorfs are grouped by sample and the time stamps increase from left to right.

First row, left to right: protein solution A, S11, S12.

Second row, left to right: S13, protein solution C

Crystallisation:

All protein screen conditions were observed under the microscope and every drop was documented. Most of these conditions either yielded an empty drop or precipitate, which indicates either too little precipitate and/or protein or too much precipitate and/or protein respectively. These results were not included in the results since they are not exactly relevant. The conditions that yielded crystals were documented. After one day at room temperature an LMB screen condition yielded crystals. The condition was NaAc pH 4.6, 9.5 % w/v PEG 3000 and 1.5 mM $MgCl_2$. As described in the materials and methods section another previously successful condition was also optimized as well as the newly discovered condition. The previously successful condition turned out to be better. These conditions were optimised on hanging drop plates. The results that were the best in quality are shown below. It can be concluded from this, that Gtf180 (our target protein) was successfully crystallized. To answer one of the research questions: The conditions needed for crystalizing Gtf180 from *L. reuteri* are 0.1 M sodium citrate/citric acid buffer pH 5.3, 1.5-1.6 M $NH_4H_2PO_4$. Something that was interesting was the crystal form displayed in figure 14c. This type of rod-like crystal was not seen before in these conditions for this protein in previous experiments. This means that most likely a different crystal has been discovered for this protein, which might diffract better than the other crystal form which has been analysed before.

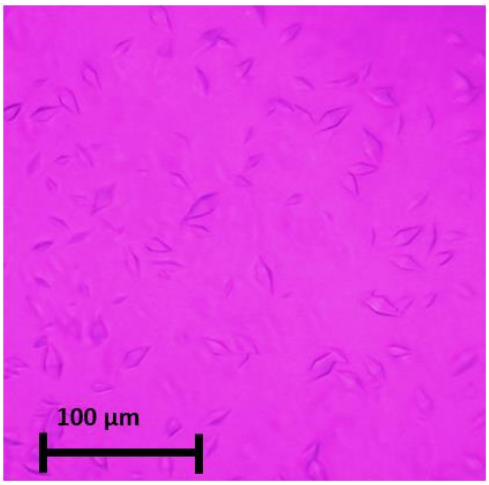


Figure 13:
Buffer: NaAc pH 4.6
Precipitant: 9.5% w/v PEG 3000, 0.1M
Salt: 1.5 mM MgCl2

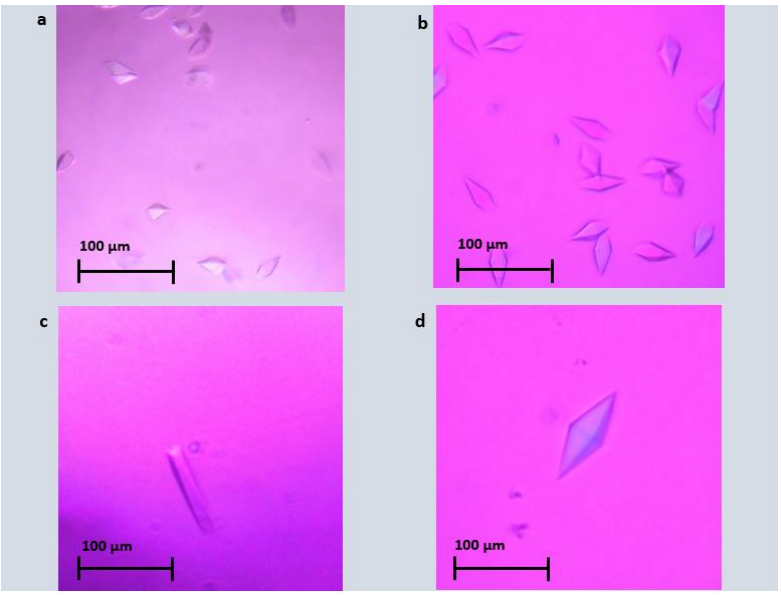


Figure 14:
Buffer: 0.1 M sodium citrate pH 5.6
Precipitant: 1M (NH4)2HPO4
a: 4 mg/ml; 1.35 M; pH 5.6 (96 well-plate)
b: 5 mg/ml; 1.65 M; pH 5.3 (hanging drop)
c: 5 mg/ml; 1.60 M; pH 5.3 (hanging drop)
d: 5 mg/ml; 1.50 M; pH 5.3 (large hanging drop)

Conclusions:

To summarize the results of all the purification experiments it can be concluded that though in the end not an entirely pure protein sample was acquired, the subsequent SDS-PAGE lanes still show incremental improvement in purity. This means it can be concluded that the steps were successful, and the success of the crystallization shows that for the purposes of our project the purification was sufficient. The activity assay confirmed that the peak from the SEC indeed contained a glucansucrase which can realistically only be *L. reuteri* Gtf180. Then to answer the research question of what the conditions necessary to crystallize Gtf180 from *L. reuteri.* The condition that was already known was sodium citrate/citric acid buffer pH 5.6 and as precipitate $NH_4H_2PO_4$ (1 M). This condition was optimised to citrate/citric acid buffer pH 5.3 and precipitate $NH_4H_2PO_4$ (1.5 - 1.6 M). The other condition that was found to work was NaAc pH 4.6, 9.5 % w/v PEG 3000 and 1.5 mM $MgCl_2$. This condition did however give poor results in the optimisation and so it was abandoned. The best conditions that were found by used were citrate/citric acid buffer pH 5.3 and precipitate $NH_4H_2PO_4$ (1.5 - 1.6 M). The answer to both other research questions is inconclusive. On the question ‘’ To what extent can conformational flexibility of the V domain be observed in solution using cryo-EM?’’ the answer is that this is not possible at all, but the inability to detect this might have something to do with the poor quality of our data. On the remaining research question ‘’ What is the full 3D structure of *L. reuteri* Gtf180?’’ we are also unable to give a conclusive answer apart from figure 9 from the cryo-EM data. The crystals will however if everything goes well be measured in a synchrotron in Grenoble, France at the end of June 2023. From the data acquired there, a structure may be solved if the crystals diffract well enough. In summary, we acquired *L. reuteri* Gtf180 in a pure enough solution for crystallisation and cryo-EM, cryo-EM failed to give satisfactory results, but crystallization was successful. From these crystals a 3D structure of a full-length *L. reuteri* Gtf180 could potentially be solved.

Discussion:

To reflect on the hypotheses made at the beginning it can be said that 1 was confirmed and 2 were inconclusive. Indeed, the conditions needed to crystallize Gtf180 were similar to the conditions that had been successful before. The hypothesis that the flexibility of the V-domain could be observed in cryo-EM turned out to be most likely false, but in principle inconclusive since the data was very poor. For future research the most obvious and most impactful change would be one to improve the results from the cryo-EM. This would be to decrease the protein concentration of the sample. This would improve the results, because it would reduce aggregates and it would make particle picking easier as there would be more isolated particles. Still, it might be so that the termini might be so intrinsically flexible, which would mean that it would be very difficult to get sufficient electron density for the termini to determine the structure for the entire protein. For the hypothesis that the 3D structure would resemble the truncated structures and the Alphafold model the results are mostly inconclusive. As mentioned before the cryo-EM 3D map in some ways resembles the truncated version of the protein (figure 15), but at very poor resolution. The crystals have not been measured yet. This is something that will be done in the future and so the results are simply not available yet.
Some unexpected findings arose during the project. One was that two shapes of crystals were discovered in the same conditions. The rod-like crystals had also not been found before in the same conditions. To speculate, this could be because of a different conformation of the protein or a different way of packing of the crystals, consistent with the unexpected of the SEC. However, since the SDS-PAGE also showed some impurity, it could be a different protein altogether. This is quite unlikely though, because the bands of the impurity were mostly weak and so there would almost certainly not be enough of the protein to crystalize. The idea that the new crystal form might be due to a different conformation is consistent with previous research finding different conformations for the V-domain (Pijning et al., 2014). Furthermore, we found weird behaviour of the protein during the SEC, where the protein seemed to behave differently at different concentrations. As described before SEC separates on size and so if both peaks (figure 4) are the same protein as DLS (figure 7) and reinsertion (figure 8) confirmed with some certainty, then a hypothesis could be that this was due to a more

compact conformation of the protein, possibly due to flexibility of the V-domain. More information will possibly arise if the protein crystals diffract sufficiently to determine the structure. For future research it could be worthwhile to measure the polydispersity at different conformations using DLS to test the hypothesis that any different populations of conformations come into existence under the influence of protein concentration.

Finally, we would like to look critically at the way the experiments were conducted. The weaknesses of the research were that no 3D structure was acquired, and that cryo-EM was done at a too high concentration. The 3D map acquired from cryo-EM was of poor quality. Something that was also a small concern was that considering getting a good protein yield it might be useful to reconsider the necessity of AIEX. SDS-PAGE, even though the sample was diluted a lot, showed not a lot of improvement after the AIEX, yet it did lose some milligrams of protein, which one can conclude also included target protein. Also, for whatever reason in the end the sample was not entirely pure, but at this point it cannot be said with certainty why. This might have been, because an artifact in the purification or thereafter. The affinity assay also gave confusing results and so in future work it might be better to use a different activity assay protocol, possibly TLC to avoid confusing UV-Vis results. The strength of the study was that in a matter of weeks protein crystals were acquired that were of sufficient size for analysis in a synchrotron. This could lead to some interesting results. Furthermore, our research gave some further implications about the conformational freedom of the V-domain and the termini of *L. reuteri* Gtf180 which might be a topic for further studies.

References:

Leemhuis, H., Pijning, T., Dobruchowska, J. M., van Leeuwen, S. S., Kralj, S., Dijkstra, B. W., & Dijkhuizen, L. (2013). Glucansucrases: Three-dimensional structures, reactions, mechanism, α-glucan analysis and their implications in biotechnology and Food Applications. *Journal of Biotechnology*, *163*(2), 250–272. https://doi.org/10.1016/j.jbiotec.2012.06.037

Koo, H., Xiao, J., Klein, M. I., & Jeon, J. G. (2010). Exopolysaccharides produced by *streptococcus mutans* glucosyltransferases modulate the establishment of microcolonies within multispecies biofilms. *Journal of Bacteriology*, *192*(12), 3024–3032. https://doi.org/10.1128/jb.01649-09

Walter, J., Schwab, C., Loach, D. M., Gänzle, M. G., & Tannock, G. W. (2008). Glucosyltransferase A (GTFA) and inulosucrase (INU) of lactobacillus reuteri TMW1.106 contribute to cell aggregation, in vitro biofilm formation, and colonization of the mouse gastrointestinal tract. *Microbiology*, *154*(1), 72–80. https://doi.org/10.1099/mic.0.2007/010637-0

Vujičić-Žagar, A., Pijning, T., Kralj, S., López, C. A., Eeuwema, W., Dijkhuizen, L., & Dijkstra, B. W. (2010). Crystal structure of a 117 KDA glucansucrase fragment provides insight into evolution and product specificity of GH70 enzymes. *Proceedings of the National Academy of Sciences*, *107*(50), 21406–21411. https://doi.org/10.1073/pnas.1007531107

Smyth, M. S. (2000). X ray crystallography. *Molecular Pathology*, *53*(1), 8–14. https://doi.org/10.1136/mp.53.1.8

Benjin, X., & Ling, L. (2019). Developments, applications, and prospects of cryo-electron microscopy. *Protein Science*, *29*(4), 872–882. https://doi.org/10.1002/pro.3805

Pijning, T., Vujičić-Žagar, A., Kralj, S., Dijkhuizen, L., & Dijkstra, B. W. (2014). Flexibility of truncated and full-length glucansucrase GTF180 enzymes from*lactobacillus reuteri*180. *FEBS Journal*, *281*(9), 2159–2171. https://doi.org/10.1111/febs.12769

Wangpaiboon, K., Padungros, P., Nakapong, S., Charoenwongpaiboon, T., Rejzek, M., Field, R. A., & Pichyangkura, R. (2018). An α-1,6-and α-1,3-linked glucan produced by Leuconostoc citreum ABK-1 alternansucrase with nanoparticle and film-forming properties. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-26721-w

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with alphafold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Chang, Y.-Y., Li, H., & Sun, H. (2017). Immobilized metal affinity chromatography (imac) for Metalloproteomics and phosphoproteomics. *Inorganic and Organometallic Transition Metal Complexes with Biological Molecules and Living Cells*, 329–353. https://doi.org/10.1016/b978-0-12-803814-7.00009-5

Lee, Y. C. (1990). High-performance anion-exchange chromatography for carbohydrate analysis. *Analytical Biochemistry*, *189*(2), 151–162. https://doi.org/10.1016/0003-2697(90)90099-u

Hong, P., Koza, S., & Bouvier, E. S. (2012). A review size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. *Journal of Liquid Chromatography & Related Technologies*, *35*(20), 2923–2950. https://doi.org/10.1080/10826076.2012.743724

Falke, S., & Betzel, C. (2019). Dynamic light scattering (DLS). *Radiation in Bioanalysis*, 173–193. https://doi.org/10.1007/978-3-030-28247-9_6

Gusakov, A. V., Kondratyeva, E. G., & Sinitsyn, A. P. (2011). Comparison of two methods for assaying reducing sugars in the determination of carbohydrase activities. *International Journal of Analytical Chemistry*, *2011*, 1–4. https://doi.org/10.1155/2011/283658

Al-Tubuly, A. A. (2000). SDS-PAGE and western blotting. *Diagnostic and Therapeutic Antibodies*, 391–405. https://doi.org/10.1385/1-59259-076-4:391