KTH ROYAL INSTITUTE
OF TECHNOLOGY

Engineered imaging scaffolds for cryo-EM of small proteins of interest

Degree project in Biotechnology

Second Cycle 30 credits

Author: Oscar Friberg

Supervisor: Johan Nilvebrant

Stockholm, Sweden 2022

Table of Contents

Abstract 3

Keywords 3

Sammanfattning 4

Introduction 5

Materials and methods 7

- Cloning 7
- Protein production 8
- Characterisation 8
  - SDS-PAGE 8
  - Size Exclusion Chromatography 8
  - Circular Dichroism 8
  - Surface Plasmon Resonance 9
- Stability study 9
- Negative Stain electron microscopy 9
- Cryogen Electron Microscopy 9

Results 10

- Cloning of new controls 10
- Protein production 11
- Characterisation 14
  - SEC purifications 14
  - Circular Dichroism 16
  - Surface Plasmon Resonance 18
- Stability study 20
- Negative stain electron microscopy 21
- Cryo-EM 23

Discussion 24

Future perspectives. 26

Acknowledgements 26

References 27

Appendix 28

Abstract

Determining structures of proteins is important to understand protein functions, and a rapidly evolving technique in this field is cryogen electron microscopy. However, size limitations are preventing wider applications of the technique because small proteins have poor signal to noise ratios and are not possible to distinguish in single-particle images. The hypothesis of this project is that it is possible to image very small proteins, bypassing the conventional size limitations of single-particle cryo-EM, by utilizing a carrier protein-scaffold (Putrescine Aminotransferase; YgjG) connected through helical fusion to an affibody (Zwt) that can bind to a small protein of interest. The complex provides a sufficient size, symmetry, and rigidity for successful electron microscopy also of the non-covalently bound small protein of interest. To characterise the proposed scaffold, thermal stability through CD, binding of target protein in SPR, purity through SEC and experiments towards proof-of-concept in cryo-EM will be performed. The small protein of interest to be imaged in the proof-of-concept setup is another affibody, called Z963, that would be the smallest protein ever solved with cryo-EM. The results show that the investigated tetrameric protein scaffold is a highly stable protein (T_m~85^{o}C) that can tolerate affibody fusion with retained binding function of multiple sites. The protein can be recombinantly expressed and purified in high yield and forms tetramers also when fused to affibody. The cryo-EM results are still pending, but promising grids have been created and in an initial particle selection clear 2-D classes that also reveal the small bound protein of interest have been generated. To conclude, biophysical characterization indicates that YgjG is a promising base structure for an imaging scaffold and preliminary single-particle cryo-EM analyses show that the proposed strategy to investigate structures of small proteins of interest is feasible.

Keywords

Protein structure determination, single-particle cryo-EM, imaging scaffold, size limitation problem, affibody, YgjG.

Sammanfattning

Strukturbestämning av proteiner är viktigt för att kunna förstå deras funktion och en snabbt utvecklande metod inom fältet är kryoelektronmikroskopi. Storleksbegränsningar förhindrar en bredare applikation av metoden eftersom små proteiner har för låg signal i förhållande till bakgrund för att kunna visualiseras som enstaka partiklar i elektronmiksoskopibilder. Hypotesen för projektet är att det är möjligt att avbilda väldigt små proteiner och kringgå den konventionella storleksbegränsningen genom att använda ett bärarprotein ((Putrescine Aminotransferase; YgjG) som kopplas till en affibody (Zwt) genom "helical fusion" och sedan binda ett litet målprotein till denna större struktur. Komplexet ska ge en tillräcklig storlek, symmetri och rigiditet för en lyckad elektronmikroskopi av bärare tillsammans med det lilla icke kovalent bundna målproteinet. För att karaktärisera den föreslagna bäraren, genomförs stabilitetstester genom CD, verifiering av inbindning av målproteinet i SPR, renhetsundersökning med SEC och slutligen kryoelektronmikroskopi för att testa konceptet. Det lilla målproteinet som kommer att avbildas i konceptstudien är en annan affibody (Z963), som i så fall skulle vara det minsta proteinet som någonsin har lösts med kryogenelektronmikroskopi. Resultaten visar att den undersökta tetramera-bäraren är väldigt stabil (T_m~ 85 ^oC) och kan tolerera en affibody-fusion med bibehållen bindning av multipla säten. Proteinet kan uttryckas rekombinant och renas till högt utbyte och bildar tetramerer även med fuserad affibody. De slutgiltiga resultaten från den kryoelektronmikroskopiska analysen inväntas fortfarande, men lovande griddar har skapats och en partikelselektion har gett klara 2-D klasser som också framhäver att det lilla målproteinet har bundit. Sammanfattningsvis har biofysikalisk karaktärisering indikerat att YgjG är en lovande bas för ett "imaging scaffold" och att preliminära enstaka-partikel kryoelektronmikroskopi analyser visar att den föreslagna strategin att undersöka små målproteiner är möjlig.

Introduction

Structural biology is an important discipline within biology and has helped us understand the function of proteins and other molecules at a molecular level. The developments in the field have led to significant insights in protein function. The field is rapidly developing but there are still many outstanding challenges, like structures having to be solved in isolation rather than in a natural context, and the time it takes to determining the structures, or problems with crystallization of samples (1).

X-ray crystallography has played a central role since the rise of the field, but there is currently a shift going on. A great increase in submitted structures solved with single-particle cryo-EM is happening and it is now possible to reach resolutions below 2 Å and a few structures have even been obtained with atomic resolution reconstructions of molecular features with this method (2)(3). In 2021 a total of 18056 structures were found in the Electron Microscopy Data Bank and 70 % of them had been submitted since 2017 (4). The popularity of single-particle cryo-EM is driven by the absence of the downsides present in X-ray crystallography, like requirement for crystallization of samples and demand for large sample quantities. The crystallization process can both be incredibly long, and some proteins are not even possible to crystallize. In cryo-EM this is not a problem since the solution containing the highly purified protein of interest is just flash frozen before microscopic imaging (1)(2). However, despite the benefits of cryo-EM, there is a limitation when it comes to the size of the imaged protein. The goal has for a long time been to break the barrier of ~100 kDa, and only a few structures have yet been solved under this threshold (5). One structure determination method that is currently used to solve structures of small proteins is nuclear magnetic resonance (NMR), but it suffers from other problems. The protein concentrations required for NMR are very high, and the required labelling of the samples can introduce structural changes (6), therefore it is still enticing to use cryo-EM.

The ionizing radiation of electron microscopes causes harm to the samples and becomes a limiting factor as the protein samples get damaged. To get a good image, the signal to noise ratio needs to be high, and for larger proteins it is higher, because they contain more atoms that can provide a signal (7). This contrast between large and small molecules becomes important because of the noisy environment in the vitrified samples (8).

The interest to use cryo-EM for smaller proteins and not having to keep relying on X-ray crystallography has inspired development of new strategies to study small proteins. In 2017 it was shown that by fusing a small protein (a designed ankyin repeat protein (DARPin)) to a larger protein carrier, it was possible to solve the structure of the smaller protein with cryo-EM (9). DARPins are small non-antibody binding proteins that can be engineered to bind target proteins of interest (10). In 2019 another paper was published where it was shown that a DARPin-aldolase fusion could be used as a scaffold-platform to resolve the structure of GFP noncovalently bound to the carrier scaffold via the DARPin. The entire complex was large enough for cryo-EM and the GFP could be structurally determined, however the resolution was poor in part due to flexibility within the DARPin and between the DARPin and the larger carrier protein (11). These advances have spawned my project where the possibility to use affibodies instead of DARPins will be investigated.

Affibody molecules are small non-antibody binding proteins made up of three helices (6,5 kDa) based on the Z domain derived from staphylococcal protein A (12). The binding site consist of a set of residues on helices 1 and 2 of the protein, which are randomized to create large libraries from which specific binders can be selected (13). Because of the large libraries it is possible to select affibodies against many different targets, these can then be used for various applications such as diagnostic tools or as therapeutics. An affibody can essentially be generated against any protein target, which may be used to provide a structure of the affibody:target protein complex using the method discussed in this thesis.

Additionally, there are already affibodies against more than 40 targets (as well as several unpublished), where most lack a solved structure (14). Insight into these structures could help shed light on their molecular interactions and prove useful in understanding and optimizing the function of existing affibodies.

In a proof-of-concept study, the affibody Zwt (12) is used as binding domain on the imaging scaffold. Zwt binds to Fc but also another (so-called anti-idiotypic) affibody, Z963, which has been selected using phage display to bind Zwt (15). In this project, Z963 will be used as a small model protein of interest (POI) and the aim is to determine its structure via single-particle cryo-EM and the "imaging platform" approach. If successful, this would represent the smallest structure ever resolved via cryo-EM (4).

One major concern regarding the utility of an imaging scaffold is the flexibility between the large carrier and the non-covalently bound protein of interest (POI), the flexibility reduces resolution or prevents accurate modelling of the protein (16)(17). One possible benefit of the small sized affibodies could be that the fusions would be more stable and, thus, reduce flexibility and potentially allow for higher resolution structure determination of bound POI. The shared helix technique (18)(19), which is an established procedure to connect two proteins with terminal alpha helices via an engineered, shared alpha-helix, was used to design the imaging scaffolds in this project and can hopefully provide sufficient rigidity between the carrier and affibody.

Desirable features of a protein carrier are sufficient size, rigidity, symmetry, existence of a terminal helix for rigid fusion, and a previously solved structure to facilitate future modelling of new variants (11). With these aspects in mind, previous master thesis students in the group have focused on both rabbit muscle aldolase and YgjG as candidate carrier proteins. Both proteins form symmetrical tetramers and contain terminal helices (11)(18). In this project four previously designed but never investigated variants of YgjG with C-terminally fused affibodies will be investigated. The fusion proteins His_6-YZw1 (previously designed with known structure from X-ray crystallography), His_6-YZw2, His_6-YZw3 and His_6-YZwIdeal and the wild type His_6-YgjG, which is included as a control, will be evaluated in this project. The difference between the designs is the sequence of the shared α-helical fusion where the relative orientation between YgjG carrier and Zwt could lead to possible sterical hindrances with a bound POI and therefore multiple designs are produced. Different orientations of the fusion could also be helpful to evaluate unknown POIs since the overall geometry will depend on the size and structure of the POI as well as on where the affibody binding epitope is located.

Once a scaffold is designed and successfully cloned, some characteristics should be studied to confirm suitability for cryo-EM. Surface Plasmon Resonance (SPR) can be used to make sure an actual POI-binding occurs to demonstrate that the affibody fused to the carrier is functional (20). The sample also needs to be monodispersed, if the protein lumps together and forms aggregates it is not possible to select single molecules in the microscope (21). The monodispersity can be studied by Dynamic light scattering (22), but it is also possible to use negative stain and look at the sample under the electron microscope (21). A promising result in negative stain EM does however not guarantee a good cryo-EM sample. Size exclusion chromatography can be used to assess homogeneity and evaluate if expected tetramer formation of novel fusion designs of the carrier proteins with affibodies occurs.

Since the initiation of this project reports the utilize nanobodies to capture proteins of interest (Legobodies, Megabodies) with different strategies to facilitate the use of imaging scaffolds have been published (23)(24). This further illuminates the potential of the idea and the validity of our method to address the cryo-EM size limitation problem.

The hypothesis of this project is that it is possible to image very small proteins, bypassing the conventional size limitations of single-particle cryo-EM, by utilizing a carrier protein-scaffold

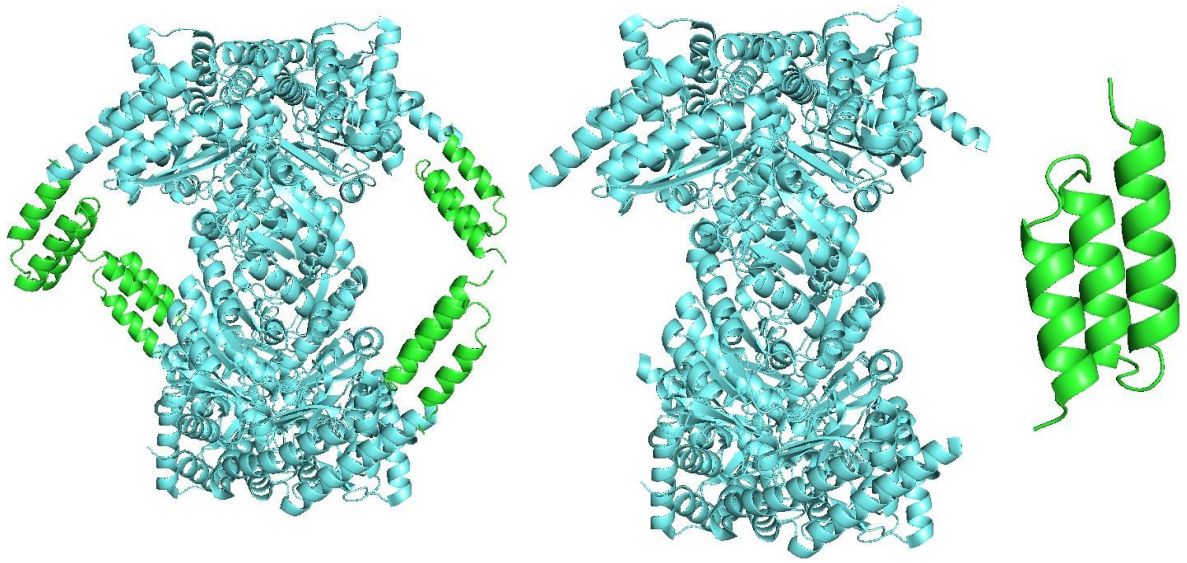
(Putrescine Aminotransferase; YgjG) connected through helical fusion to an affibody (Zwt) that can bind to a small protein of interest. The complex provides a sufficient size, symmetry, and rigidity for successful electron microscopy also of the non-covalently bound small protein of interest.

Materials and methods

Cloning

Plasmids containing the constructs His_6-YZw1, His_6-YZw2, His_6-YZw3, His_6-YZwIdeal and His_6-YgjG (control without Zwt fusion, **Fig. 1**) were already available at the beginning of the project, but five additional constructs were designed. These novel constructs were called His_6-YgjG-Zwt (direct end-to-end fusion), His_6-YgjG-(G_4S)-Zwt, His_6-YgjG-$(G_4S)_2$-Zwt, His_6-YgjG-$(G_4S)_3$-Zwt, and His_6-YZwIdeal-$(G_4S)_3$-Z963. The new constructs were cloned utilizing primers with overhangs containing linker sequences (appendix **Table A1**) Insert fragments (appendix **Table A2/A3**) and linearized expression vectors (appendix **Table A4/A5**) were generated by PCR. Fragments were purified using QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and assembled using In-Fusion seamless cloning (TaKaRa, Kusatsu, Japan) following the provided protocol scaled-down to 2.5 µl reactions. Cloning reactions were transformed to Stellar *E. coli* cells and grown Over Night (ON) on Kanamycin (Km) (50 µg/ml) infused agar plates. Colony PCR products using general primers confirmed expected lengths of the inserts by gel electrophoresis using GeneRuler DNA Ladder Mix (Thermo Scientific, Waltham, MA, USA; appendix **Fig. A1**). Colonies were picked and cultivated in 5 ml Tryptic Soy Broth + Yeast Extract (TSB+Y) and Km (25 µg/ml) ON and plasmids were purified using QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer's protocol. DNA-sequencing was used to verify sequences of the inserts.

Sequence-verified expression plasmids encoding the novel fusion proteins with N-terminal His_6 tags were transformed to *E. coli* BL21*(DE3) via heat shock (42°C, 1:30 min in a water bath). Immediately after 200 µl warm (37°C) TSB+Y was added. The bacteria were phenotyped (37°C, 1 hour) before plating on Km (50 µg/ml) infused agar plates.



His_6-YZw3 YgjG Affibody

Figure 1: *Model of the wild type YgjG in tetramer form, free affibody, and YgjG docked to an affibody to illustrate His_6-YZw3.*

Protein production

For each sample a single *E. coli* BL21*(DE3) colony was inoculated to an Erlenmeyer flask (E-flask) (200 ml) with 15 ml Tryptic Soy Broth + Yeast Extract (TSB+Y), Km (25 µg/ml) (37^{0}C, 150 RPM) and incubated ON. 200 ml TSB+Y (25 µg/ml Km) in 1 litre E-flasks were inoculated 1:100 with ON-culture and incubated (37^{0}C, 150 RPM) until OD600 was 0.8-1 before addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) (1 mM final concentration) to induce protein expression. Following ON incubation (37^{0}C, 150 RPM) cells were harvested by centrifugation (4^{0}C, 15 min, 2700 g), supernatant was discarded, and the pellets resuspended in 5 ml wash buffer (PBS 15 mM Imidazole, pH 7.4; WB). Samples were sonicated with a microtip (3 mm) for 3 min on ice with (1.0/1.0) pulsation, 30% amplitude. The lysed cells were centrifuged again (4^{0}C, 45 min, 13000 RPM), clarified lysates collected and purified via immobilized metal ion affinity chromatography (IMAC) using a batch protocol. 2 ml HisPur Cobalt Resin (Thermo Scientific) was added to each tube and washed once with 10 ml sterile water (StAq) and twice with 10 ml WB. In each step tubes were centrifuged (3 min, 3000 rpm) and the supernatant removed. Lysates were diluted with 5 ml WB and samples incubated with equilibrated resin on a rotating mixer at room temperature (RT) (30 min). Following three washes with 10 ml WB, samples were eluted with 2.5 ml elution buffer (PBS + 300 mM imidazole; EB) by incubation on a rotating mixer (RT, 20 min). The buffer was changed to phosphate buffered saline (150 mM NaCl, 0.08 M Na_2HPO_4, 0.02 M NaH_2PO, pH 7.4; PBS) on PD-10 desalting columns (Cytiva, Marlborough, MA, USA) using the manufacturer's instructions. Protein concentrations were estimated via A280 measurements both before and after buffer exchange using theoretical molecular weights and extinction coefficients calculated from amino acid sequences by Expasy ProtParam (https://web.expasy.org/protparam/) (see appendix **Table A6**).

Characterisation

SDS-PAGE

Purified protein samples in PBS were analysed by sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Reducing loading buffer mixed with 3 µg of protein sample was boiled (10 min, 95^{o}C) before loading on reducing 4-12% NuPAGE gels (Invitrogen, Waltham, MA, USA) and run at 4^{0}C (45 min, 200V) in MES buffer (50 mM MES, 50 mM Tris-Base, 3.5 mM SDS, 1 mM EDTA; MES buffer). Gels were stained by using Coomassie blue for 1 h and destained in water, before imaging in a Gel Doc EZ Imager (BioRad, Hercules, CA, USA). The ladder used in the experiments was Amersham Low Molecular Weight Calibration kit for SDS Electrophoresis (Cytiva), appendix **Fig. A2**.

Size Exclusion Chromatography

Size Exclusion Chromatography (SEC) was performed on a Cytvia HiLoad 16/600 200 pg column (Cytiva) connected to an ÄKTA Pure (Cytiva). The column was equilibrated with 1 CV of PBS and run at 1 ml/min. Samples were filtered (0.2 µm) and spun (10 min, 10 000 g) before loading. Sample volumes varied between 0.5-2 ml with ~5 mg/ml protein concentration (in run 1: 0.5 ml, in run 2: 1 ml, in run 3: 2 ml). Elution peaks were collected in fractions based on a defined A280 cut-off. Based on A280 values fractions containing protein were pooled and analysed by SDS-PAGE to verify the presence of protein of interest.

Circular Dichroism

Circular Dichroism (CD) spectra were recorded with Chirascan Circular Dichroism Spectrometer (Applied Biopphysics, Troy, NY, USA) in 1 mm cuvettes (282 QS 0.100), 195-260 nm at 20^{o}C and an average of five spectra was used. For melting point measurements spectra were recorded at 221 nm during sample heating from 20-95^{o}C with an increment of 5^{o}C per min. Spectra (195-260 nm at 20^{o}C)

were again recorded following sample cooling. For all measurements, the proteins were diluted to 0.5 mg/ml in PBS, and a blank PBS sample was subtracted.

Surface Plasmon Resonance

Samples were immobilized on a Series S Sensor Chip CM5 (Cytiva) on a Biacore T200 (Cytiva) using PBS + 0.005% (w/v) Tween 20, pH 7.4 (PBST) as running buffer and 10 mM HCl for regeneration. Proteins were diluted to 5 µg/ml in 10 mM Sodium acetate pH 4.5 and immobilized by amine coupling using N-Hydroxysuccinimide (NHS) and ethyl(dimethylaminopropyl) carbodiimide (EDC) followed by deactivation using 1 M ethanolamine. For each direct binding cycle sample was injected for 500 sec at 30 µl/min with a dissociation time of 1000 sec, before regeneration. The dual injection program was also used, after the first pulse of YgjG-variant (200 nM) a second pulse of PBST, Z963 or a mAb (Elotzumab) was injected (200 nM).

Stability study

To get a better understanding of long-term stability and the influence of storage conditions on sample stability proteins from the first batch were diluted to 1 ml aliquots of 1 mg/ml and stored frozen (-20°C), refrigerated (+4°C) or on ice in a cold room (+4°C). After 112 days, samples were brought out and stored on ice until the frozen samples thawed. The samples were briefly mixed by pipetting up and down three times (200 µl tip) and absorbance (280 nm) was measured, then the samples were centrifuged (10 min, 10000 g) and without mixing the absorbances of the top solutions were measured again. From each sample 3 µl was analysed with SDS-PAGE.

Negative Stain electron microscopy

Maxtaform 400 mesh copper grids (TAAB Laboratories Equipment Ltd, Aldermaston, UK) with a thin continuous carbon film added, were put in a glow discharger (Balzers SCD 004 Sputter coater; Oerlikon, Freienbach, Switzerland). 3 µl sample (40-90 µg/ml), in HEPES buffer (20 mM HEPES, 50 mM NaCl, pH 7.5; HEPES buffer) was stained in phosphotungstic acid hydrate (Sigma-Aldrich, St Louis, MO, USA) and incubated for 1 min, two drops of water were used to wash the grid before letting it air dry, excess water was removed with a tissue. The samples were observed in a JEM-2100f (JEOL Ltd., Akishima, Japan) with field emission gun, and pictured with TemCam-XF416 (TVIPS GmbH, Gauting, Germany).

Cryogen Electron Microscopy

Cryogen electron Microscopy (Cryo-EM) was performed two times in this project. The first time Cryo-EM was performed in KTH-Flemingsberg on the same setup described under negative Stain, but with Quantifoil R2/1 and R2/2 200 mesh copper (Quantifoil Micro Tools GmbH, Grosslöbichau, Germany) grids. After glow discharge, 3 µl sample (100-800 µg/ml) in HEPES buffer was added before blotting (2-4 sec) and plunged into liquid nitrogen. The second time Cryo-EM was performed at SciLifeLab national Cryo-EM facility and initiated in the final project week; therefore, experimental details are left out.

Results

Cloning of new controls

The five novel constructs, see **Fig. 2**, were successfully cloned, and their lengths corresponded to the expected sizes in PCR-screening, **Fig. 3**. All insert sequences were verified via DNA sequencing.

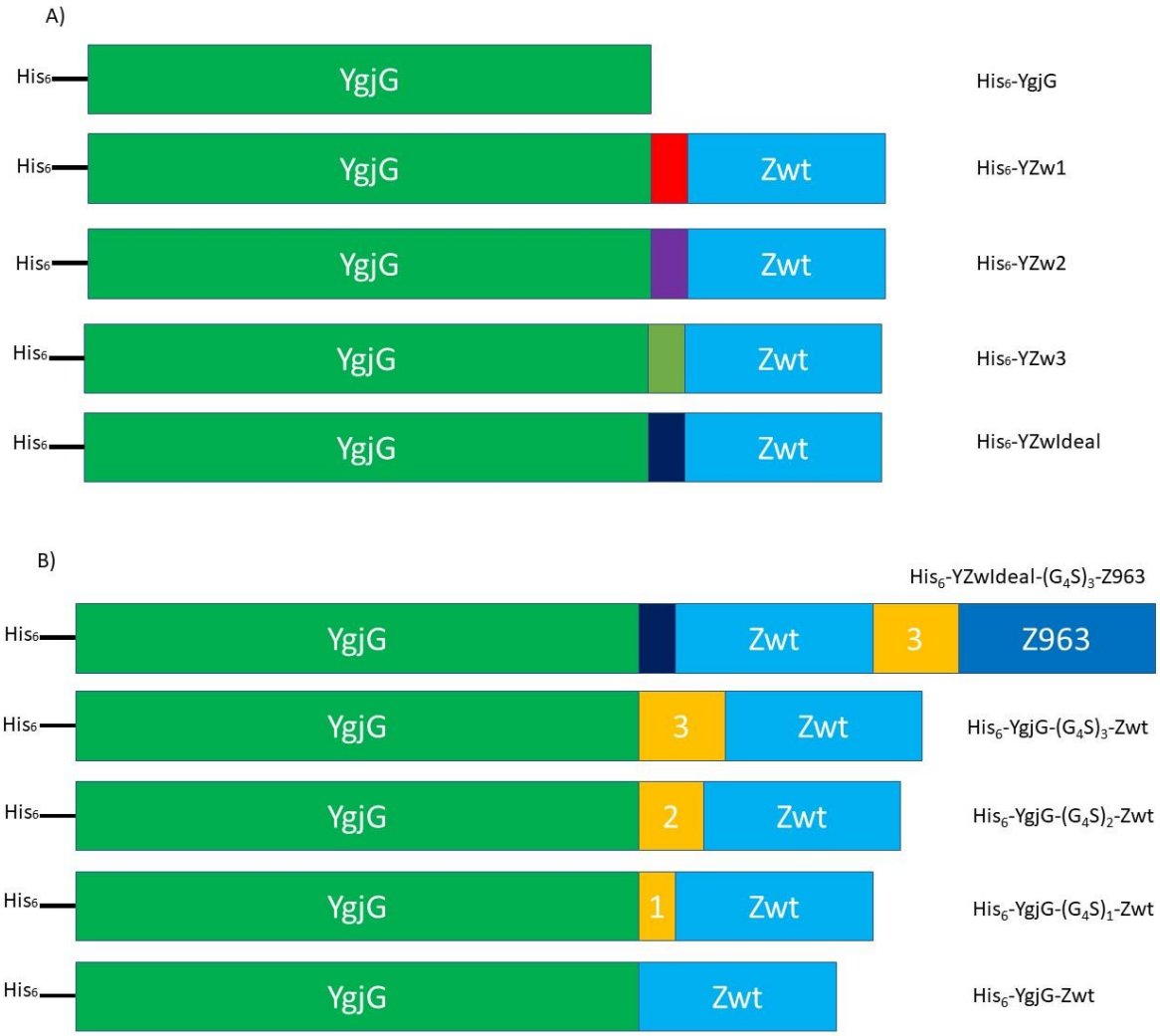


Figure 2: *Schematic illustrations of the constructs A: Previous constructs, all with a different rigid helical fusion, B: Newly cloned control constructs with flexible linkers. The green box is the carrier (YgjG), the yellow box is a flexible linker sequence (GGGGS) repeated 0-3 times, and in blue affibodies, either Zwt or the anti-idiotypic Z963.*

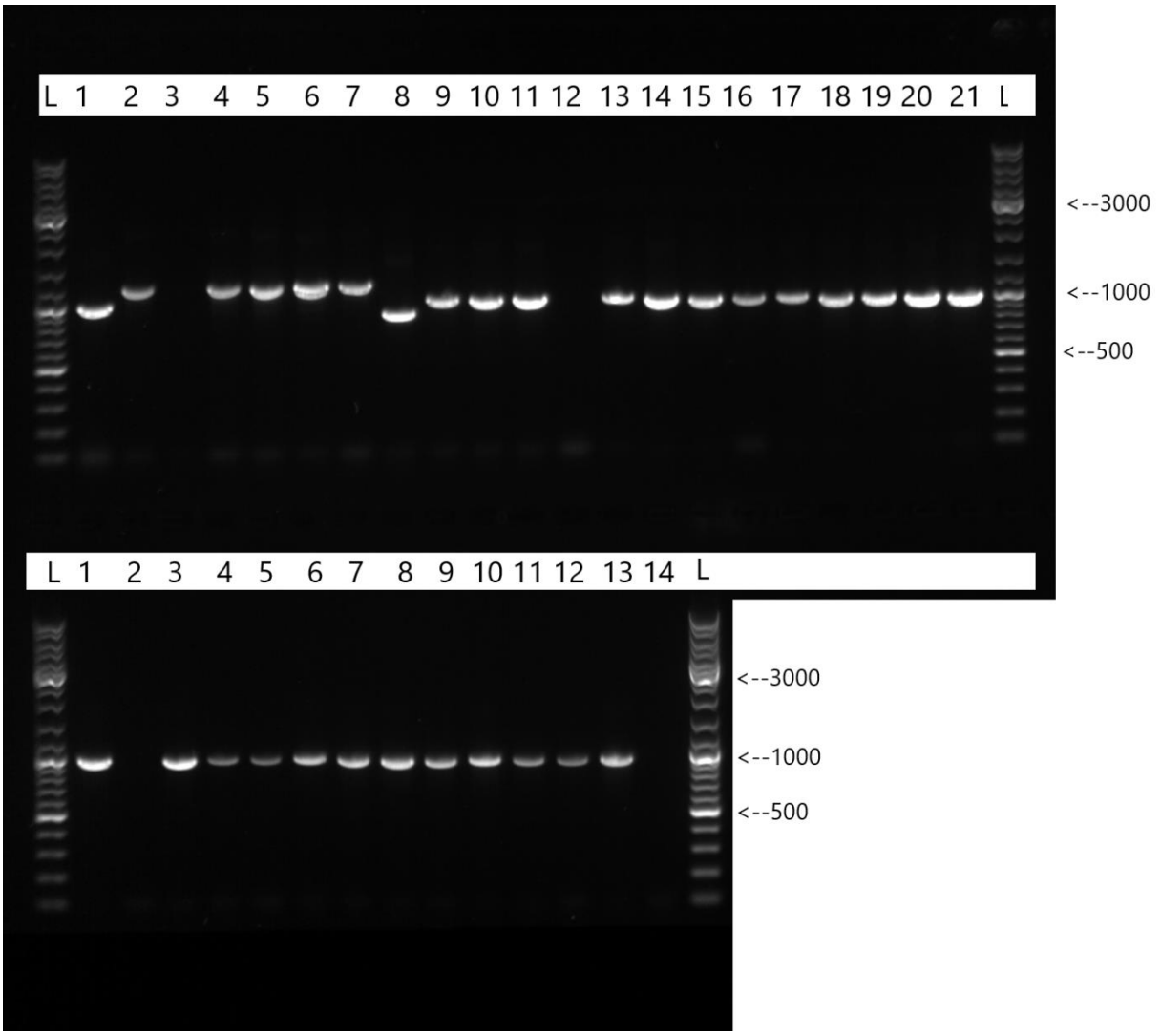


Figure 3: *Colony PCR, for ladders see appendix **Fig. A1**, the wells contain 1-5) random colonies for His_6-YZwIdeal-$(G_4S)_3$-Z963 8-12) His_6-YgjG-$(G_4S)_3$-Zwt, 15-19) His_6-YgjG-$(G_4S)_2$-Zwt, 22-26) His_6-YgjG-(G_4S)-Zwt, 29-33) His_6-YgjG-Zwt. Negative control colonies are included in two wells to the right of each sample where Takara Fusion Mix was left out (wells 6-7 for His_6-YZwIdeal-$(G_4S)_3$-Z963, 13-14, 20-21 and 27-28.*

Protein production

Protein production was performed three times during the project (See **Table 1**), the first batch, **Fig. 4**, was used for characterisation with CD and SEC, the second batch, **Fig. 5**, was freshly made and SEC-purified for cryo-EM in Flemingsberg and the third batch, **Fig. 6**, was freshly made and SEC-purified for cryo-EM at SciLifeLab, and investigated with SPR. Two of the newly cloned control samples were only included in the third purification batch, the remaining new constructs were only cloned and not expressed. The protein production yielded similar results each time as seen in Table 1, which shows the concentrations measured after IMAC purification and PD10 buffer exchange into PBS (total sample volume = 3.5 ml). Given the high protein yields, there was no need to increase the cultivation volume from 200 ml TSB+Y. During the elution step in the IMAC purification protein precipitation was observed in most cases, the precipitated fraction was presumably captured in the PD10 column filters during buffer exchange.

Table 1: *Protein amounts after IMAC purification and buffer exchange into PBS using P10-columns. Concentrations in mg/ml were calculated from absorbance at 280 nm using extinction coefficients and molecular weights calculated from sample amino acid sequences.*

Protein	Batch #1 (220127) (mg/ml)	Batch #2 (220310) (mg/ml)	Batch #3 (220510) (mg/ml)
His_6-YgjG	6,41	6,26	7,10
His_6-YZw1	5,25	5,55	6,62
His_6-YZw2	6,41	4,95	6,04
His_6-YZw3	7,16	7,22	7,73
His_6-Yzwideal	4,66	6,75	4,95
Z963-His_6	2,68	3,55	4,11
His_6-YgjG-$(G_4S)_3$-Zwt	-	-	6,40
His_6-YZwIdeal-$(G_4S)_3$-Z963	-	-	5,60

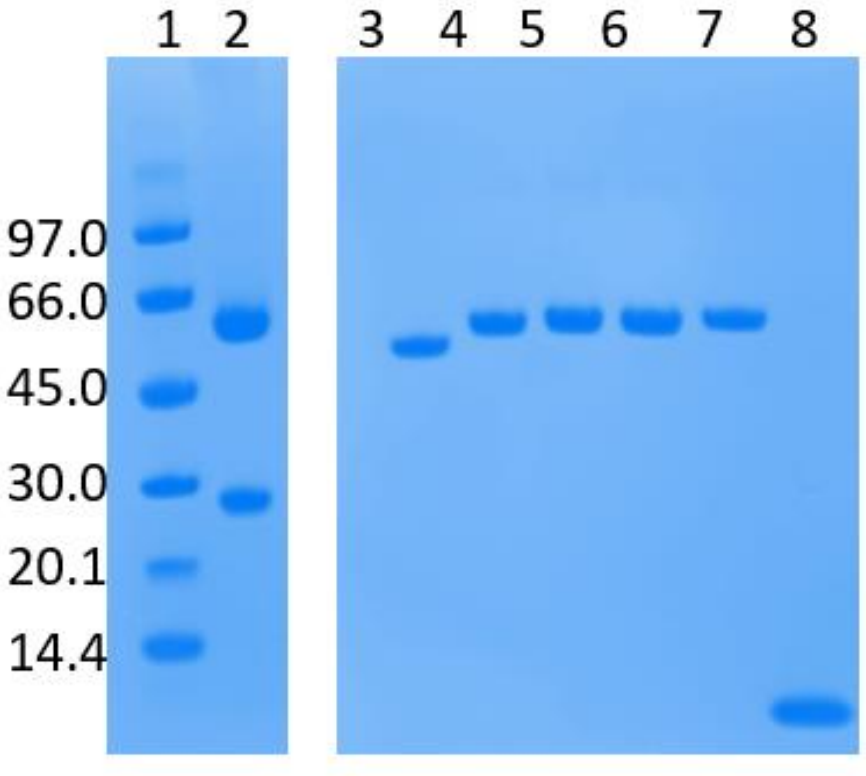


Figure 4: *SDS-PAGE analysis of proteins produced in batch #1 after IMAC purification and PD10 buffer exchange to PBS. Well 1: Ladder (indicated reference bands in kDa), 2: Cetuximab, 3: His_6-YgjG, 4: His_6-YZw1, 5: His_6-YZw2, 6: His_6-Yzw3, 7: His_6-YzwIdeal, 8: Z963-His_6. Cetuximab was used during the SPR experiments and was obtained in a pure form. Sample volumes corresponding to an expected amount of 3 µg sample were loaded in each well.*

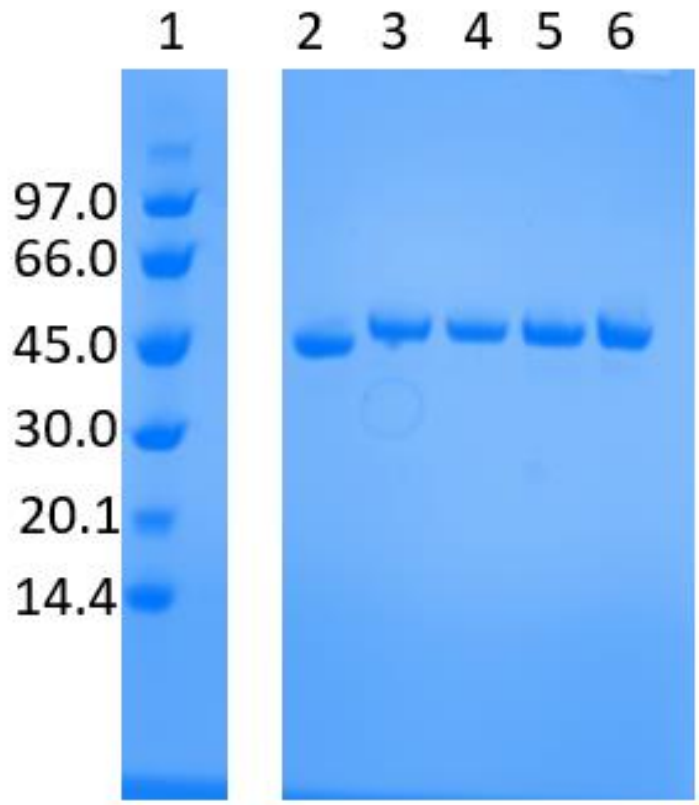


Figure 5: *SDS-PAGE analysis of proteins produced in batch #2 after SEC purification in PBS. Well 1: Ladder, 2: His_6-YgjG, 3: His_6-YZw1, 4: His_6-YZw2, 5: His_6-YZw3, 6: His_6-Yzwideal. Sample volumes corresponding to an expected amount of 3 µg sample were loaded in each well.*

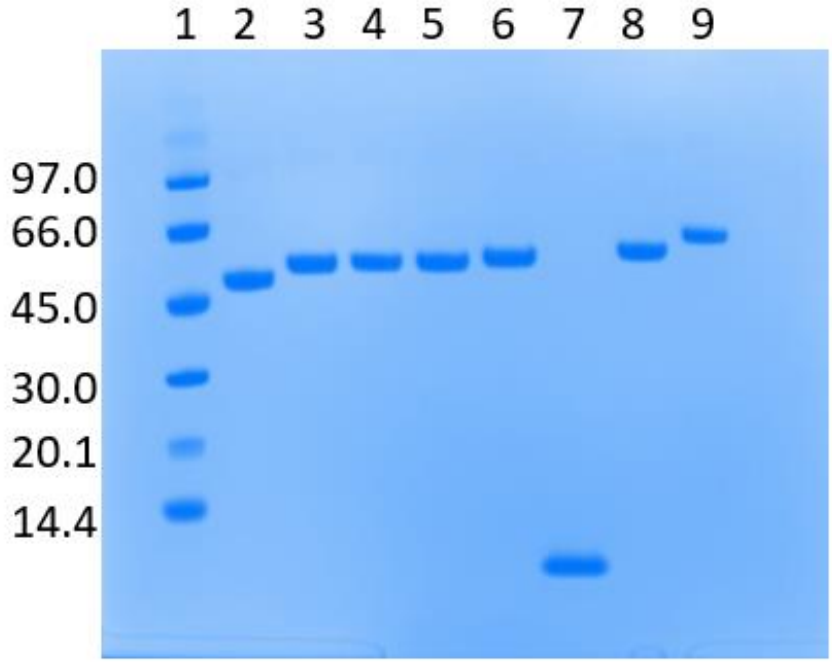


Figure 6: *SDS-PAGE analysis of proteins produced in batch #3 after SEC purification and buffer exchange to HEPES buffer. Well 1: Ladder, 2: His_6-YgjG, 3: His_6-YZw1, 4: His_6-YZw2, 5: His_6-YZw3, 6: His_6-Yzwideal, 7: Z963-His_6, 8: His_6-YgjG-$(G_4S)_3$-Zwt, 9: His_6-YZwIdeal-$(G_4S)_3$-Z963. Sample volumes corresponding to an expected amount of 3 µg sample were loaded in each well.*

Characterisation

SEC purifications

The first SEC purification (batch #1) aimed at characterising purity and oligomeric state of the samples in solution, it showed no detectable amount of monomer and most of the protein seemed to be part of a soluble tetramer that eluted in a symmetrical main peak ("tetramer peaks" **Fig. 7**) for all samples. A small earlier peak ("pre-peaks" **Fig. 7**) was also observed for all samples, it appeared at an elution time that correlated with the elution time of the respective main peak and was therefore expected to contain the same protein in a higher oligomeric state rather than being an impurity (**Fig. 7/8/9**). This smaller peak was collected and analysed by SDS-PAGE (appendix **Fig. A3**) and was indeed shown to contain the same protein species as found in the main peak. All observations in the first SEC were repeatedly observed in the subsequent purifications of samples from batch #2 and batch #3, the main difference between runs was the amount of protein loaded on to the column. This is illustrated in **Fig. 10**, a comparison where a sample volume of 0.5, 1 and 2 ml was loaded onto the column in batch #1, batch #2 and batch #3, respectively. The increased sample volumes yielded higher concentrations in the purified fractions from the main peak, and the time of elution for the main peak was essentially unaffected.

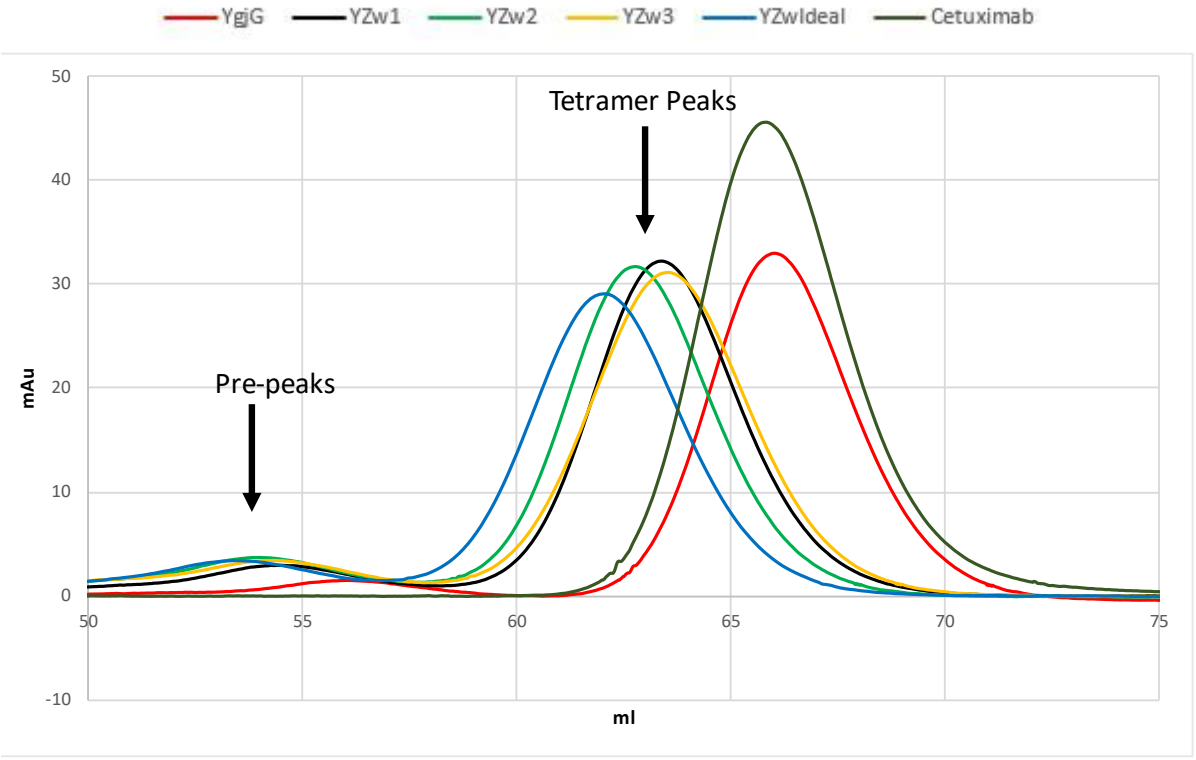


Figure 7: *Overlay of SEC chromatograms from purification of samples in Batch #1, absorbance at 280 nm on the y-axis and elution volume on the x-axis. Cetuximab was included in this run as a size reference, 0.5 ml sample was loaded. Each tetramer 220 kDa eluted at about the same time as IgG 150 kDa which indicates that a relatively compact tetrameric structure is formed by YgjG.*

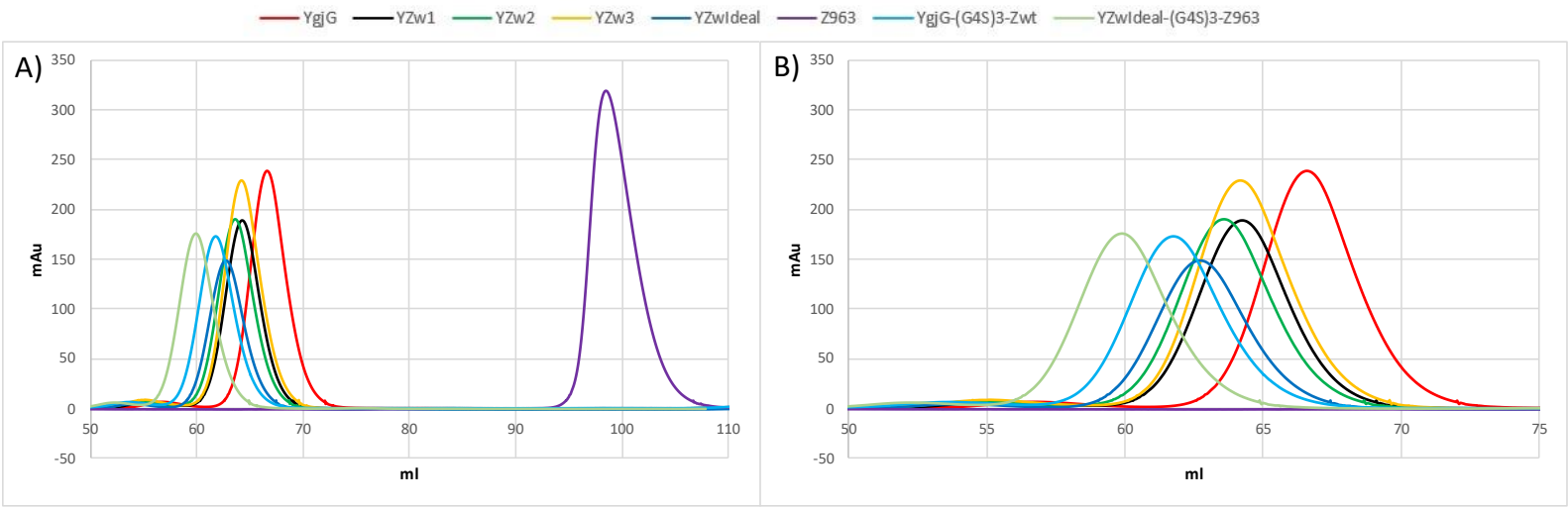


Figure 8: *Overlay of SEC chromatograms from purification of samples in Batch #2, 1 ml sample was loaded. Panel A: All eluted proteins, B: Zoomed-in view of main peaks for YgjG-based samples. The small Z963 (7,5 kDa) was purified for use in cryo-EM and eluted much later than the larger YgjG-based samples. No pre-peak was observed for Z963, which again illustrates that it is YgjG-dependent*

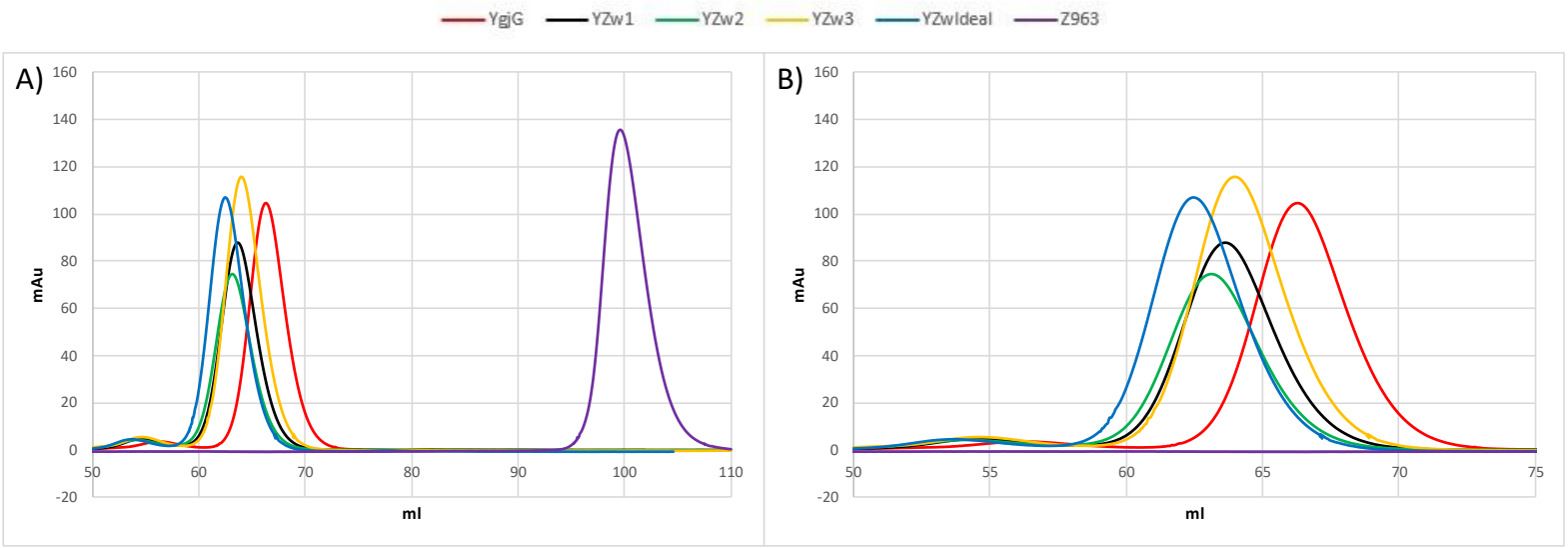


Figure 9: *Overlay of SEC chromatograms from purification of samples in Batch #3, 2 ml sample was loaded. Panel A: All eluted proteins, B: Zoomed-in view of main peaks for YgjG-based samples. In this purification two newly cloned variants are included, His6-YZwIdeal-(G4S)3-Z963 elutes first followed by His_6-YgjG-$(G_4S)_3$-Zwt, as expected because of their larger size given molecular weight and long linker flexibility.*

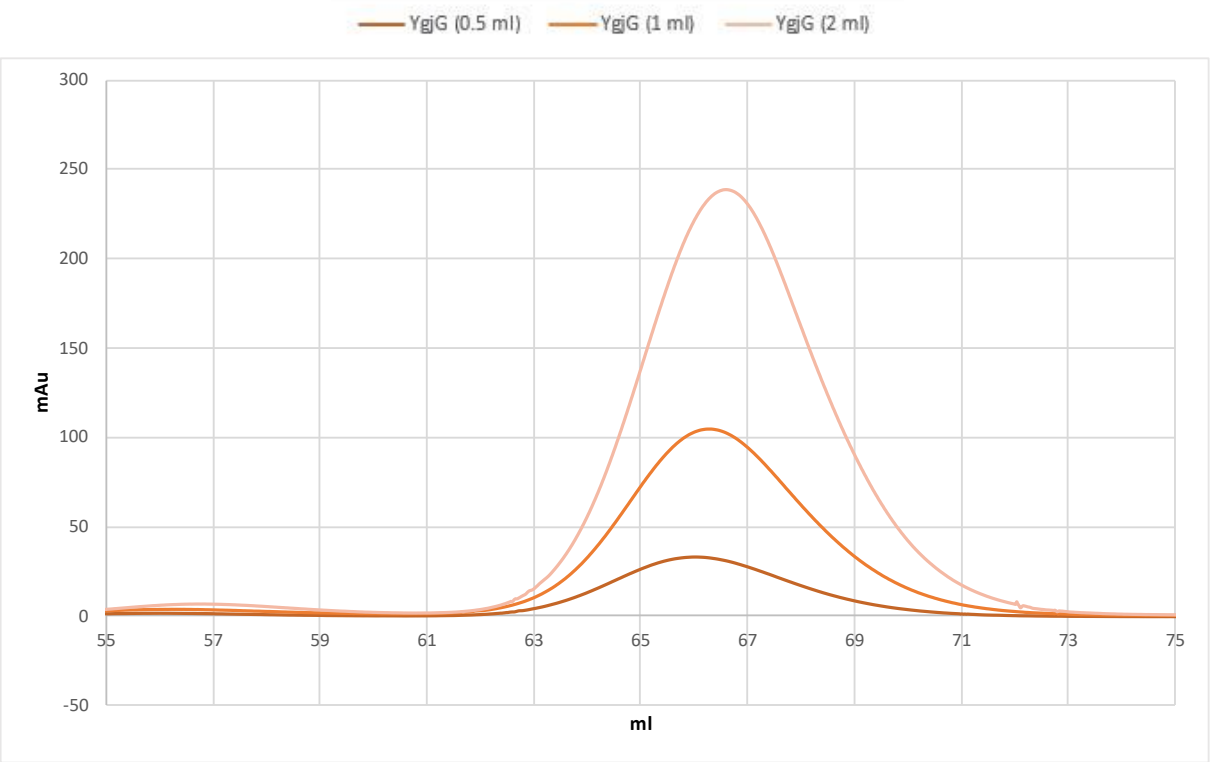


Figure 10: *Overlay of main elution peaks of SEC chromatograms from purification of YgjG between batch #1, #2 and #3, where sample volume was increased from 0,5 ml, 1 ml and 2 ml, respectively.*

Circular Dichroism

In the experiment the proteins used (Batch #1), were only purified with IMAC followed by buffer exchange into PBS. The measured proteins all yielded characteristic alpha helical spectra prior to heat denaturation, **Fig. 11**, which was lost after heating to 95°C, **Fig. 12**. All samples precipitated after thermal denaturation (see appendix **Fig. A4**), which resulted in lower signals in spectra after melting. The thermal melting transitions indicate protein melting points between 80-90°C and there were no obvious differences between the wild type and variants in any of the measurements, **Fig. 13**, which indicates that the C-terminal Zwt fusions are well tolerated.

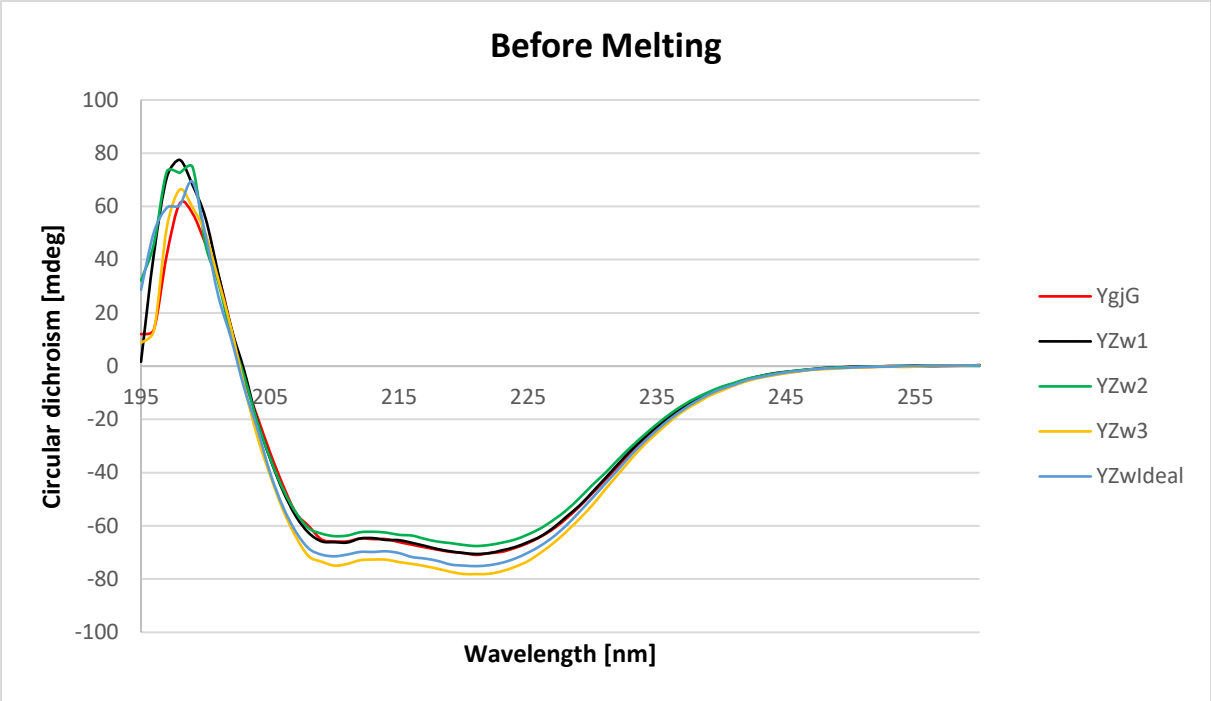


Figure 11: *Overlay of CD spectra (195-260 nm) of all carriers and the wild type 195-260 nm at 20°C. Average curves calculated from five spectra per sample are shown. The curve shapes suggest that all samples form α-helical structures.*

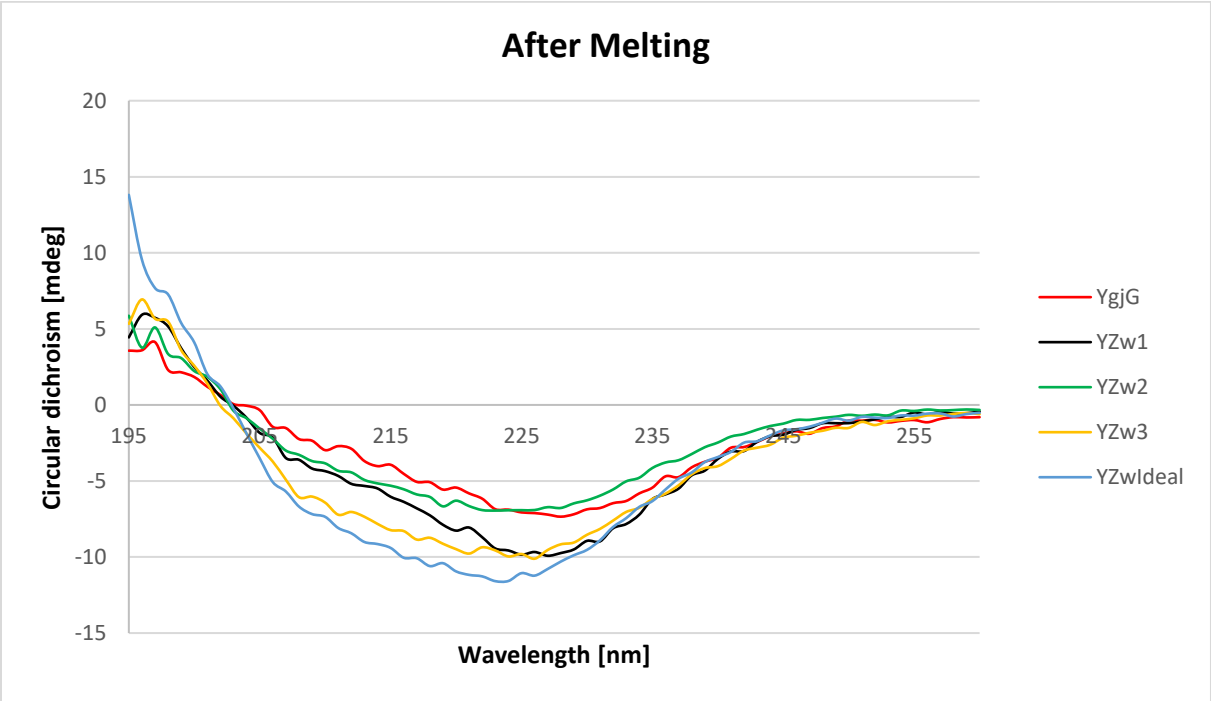


Figure 12: *Overlay of CD spectra (195-260 nm) of all carriers and the wild type 195-260 nm at 20°C. Average of five spectra per sample are shown. Major signal decreases were observed after thermal denaturation of samples, which is due to sample precipitation in the cuvette.*

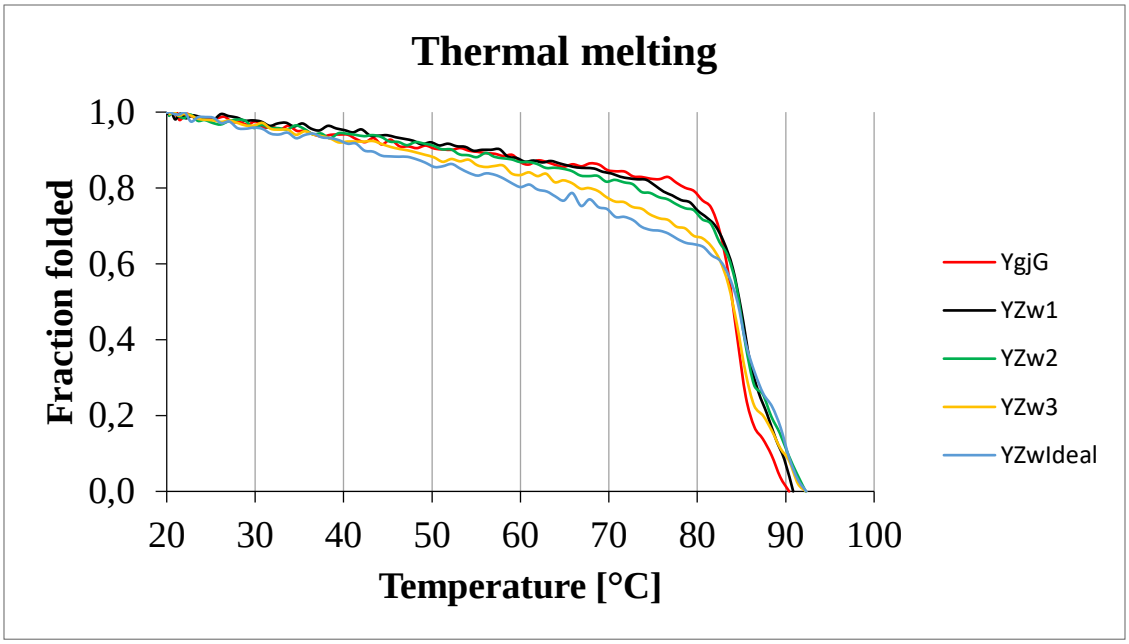


Figure 13: Overlay of *thermal melting curves, the fraction of folded protein recorded against rising temperature. Thermal melting points of about 85 ˚C for YgjG with and without affibody fusion were observed.*

Surface Plasmon Resonance

In the experiments the binding of the carrier variants varied, **Fig. 14** show that not all carriers bound Cetuximab but they could bind Z963, **Fig. 15**. The dual inject program resulting in additional response and indicate that more than two binding sites are functional, **Fig. 16**.

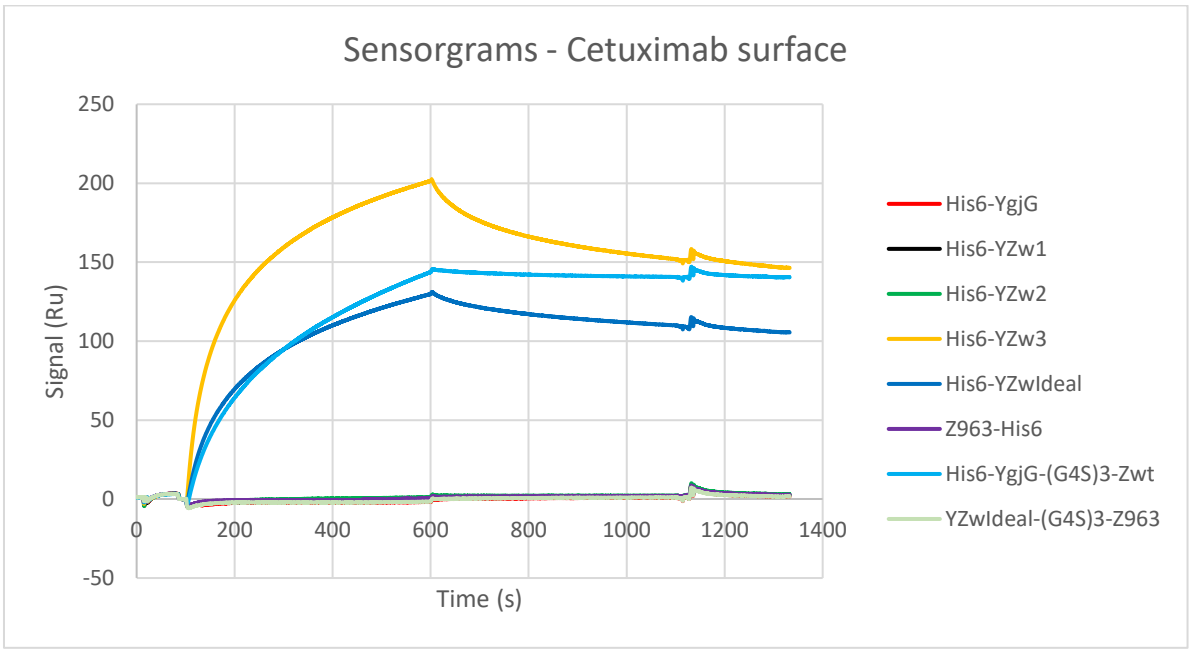


Figure 14: *Overlay of reference-subtracted sensorgrams for samples injected at 200 nM concentration over immobilized cetuximab (1600 RU). Two of the shared helix designs, YZw3 and YZwIdeal, and the control with flexibly linked Zwt (YgjG-(G$_4$S)$_3$-Zwt) display clear binding to Fc of the immobilized antibody. No binding was observed for two of the designs, YZw1 and YZw2, and no non-specific binding was observed for YgjG itself or Z963.*

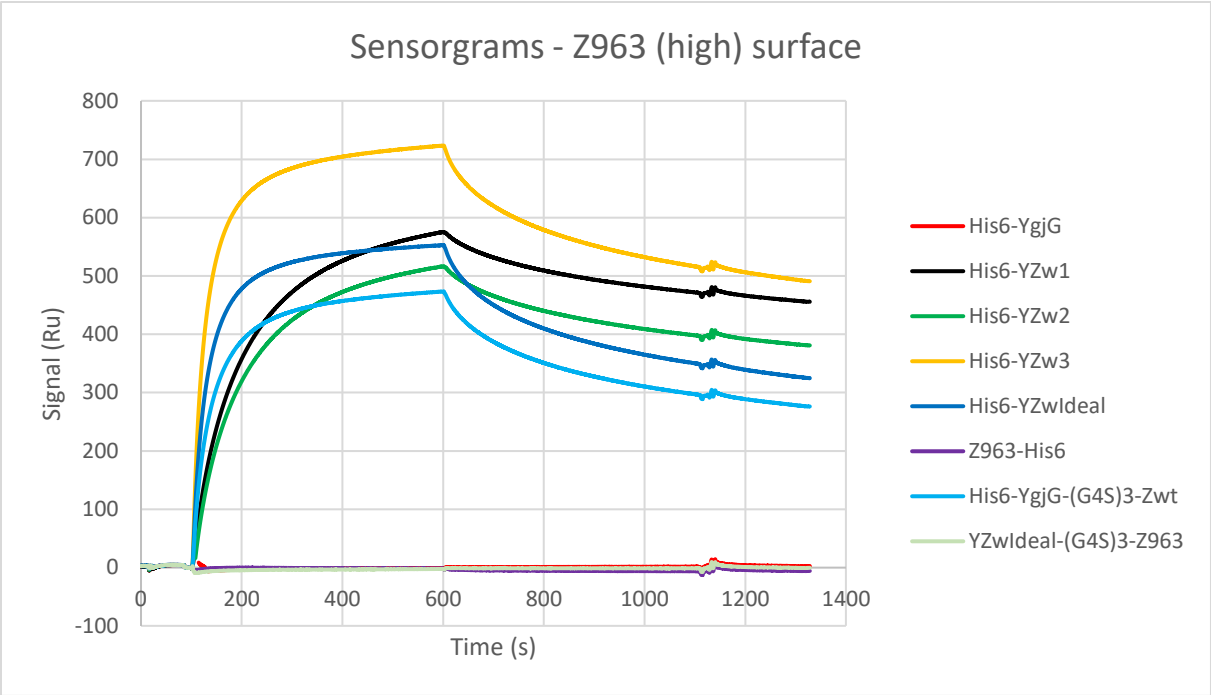


Figure 15: *Overlay of reference-subtracted sensorgrams for samples injected at 200 nM concentration over immobilized Z963 (900 RU). All YgjG variants containing Zwt bound immobilized Z963 except for the control where Z963 was included in the injected protein attached via a flexible linker. No non-specific binding to immobilized Z963 was observed for YgjG itself or for injected Z963. These results show that the Zwt moiety is functional in all rigid helix fusion constructs (e.g., YZw1, YZw2, YZw3 and YZwIdeal) and may indicate that binding to Fc is sterically hindered in the two designs that did not bind cetuximab.*

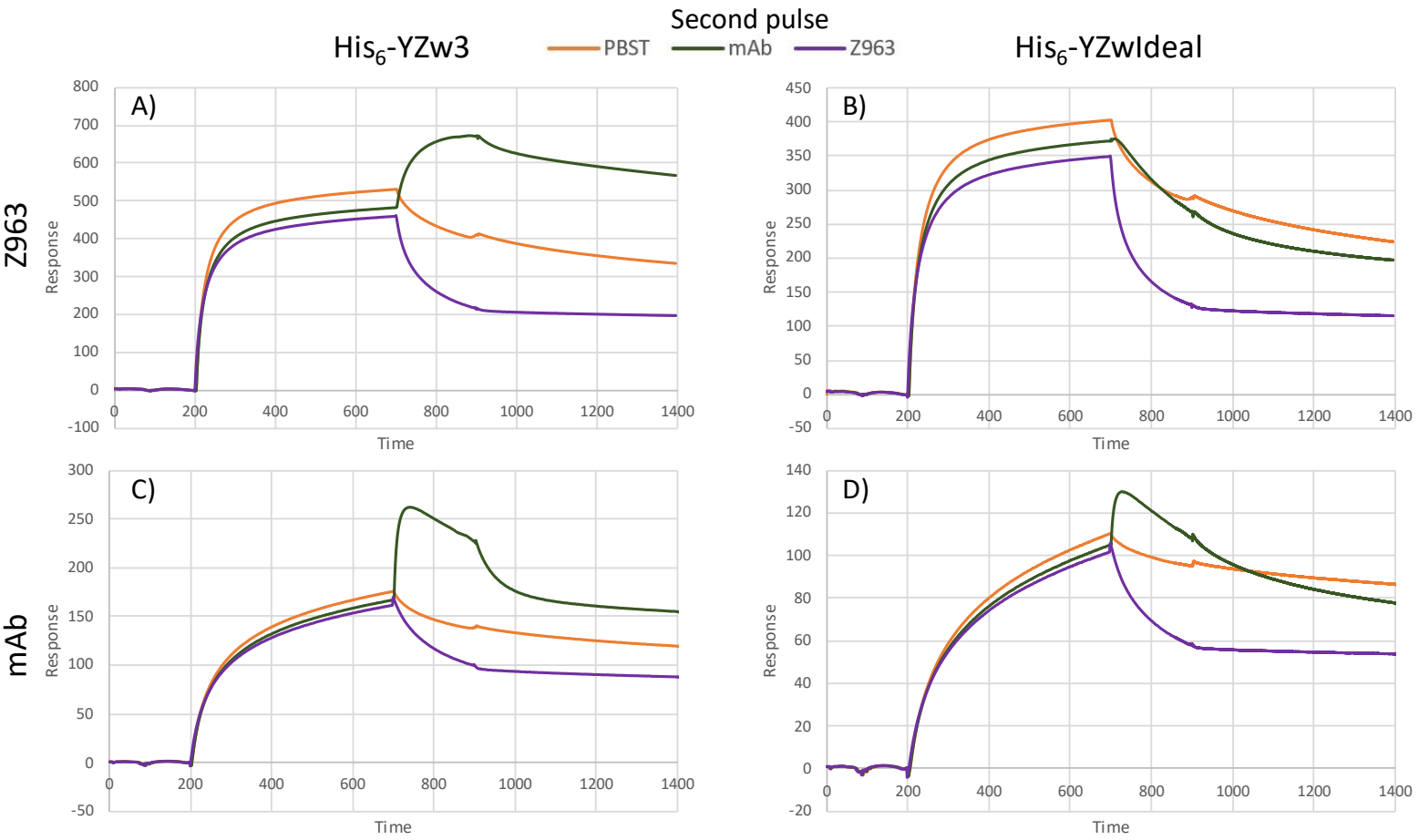


Figure 16: *Overlay of reference-subtracted sensorgrams for samples injected at 200 nM concentration over immobilized mAb or Z963. Sensogram A: Injected sample His$_6$-YZw3 on immobilized mAb, B: Injected sample His$_6$-YZwIdeal on immobilized mAb, C: Injected sample His$_6$-YZw3 on immobilized Z963, D: Injected sample His$_6$-YZwIdeal on immobilized Z963. The second pulse is either PBST, mAb or Z963. In all cases injection of Z963 increased dissociation of the bound protein, which can be interpreted as an interference with the avidity effect. PBST in the second pulse represents the dissociation baseline. Injection of mAb resulted in increased signal in A/C/D and indicate functional binding of more than two sites on the carriers. No signal increase upon injection of mAb was observed in B. Data for His$_6$-YZw1 and His$_6$-YZw2 are not shown since these designs did not bind Fc in the direct binding experiment.*

Stability study

The non-frozen samples had a smaller difference in concentration before and after centrifugation in comparison to the frozen samples, **Table 2**. On the SDS-PAGE gel **Fig. 17** clear bands are observed for all samples that look similar for all conditions with no visible protein degradation. Faint bands are visible further up on the gel, like what was observed for the pre-peaks (see appendix **Fig. A3**). These results indicate that the YgjG-Zwt fusion proteins are stable over time regardless of if they are stored refrigerated or frozen. The absorbance (280 nm) was higher before centrifugation than after which indicate precipitation of the sample, also there is a deviation upwards from 1 mg/ml of all measured samples.

Table 2: *Values are reported in mg/ml, Fridge +4˚C, Freezer -20˚C, on ice ~0˚C, (B) is before and (A) is after centrifugation of sample.*

Protein	Fridge (B)	Fridge (A)	Freezer (B)	Freezer (A)	Ice (B)	Ice (A)
His_6-YgjG	1,15	1,10	1,60	1,04	1,15	1,08
His_6-YZw1	1,07	1,10	1,36	0,94	1,05	1,00
His_6-YZw2	1,13	1,11	1,38	1,04	1,07	1,02
His_6-YZw3	1,20	1,12	1,36	1,05	1,25	1,04
His_6-Yzwideal	1,13	0,99	1,24	1,05	0,99	0,93
Z963-His_6	1,59	1,46	2,30	1,27	1,40	1,23

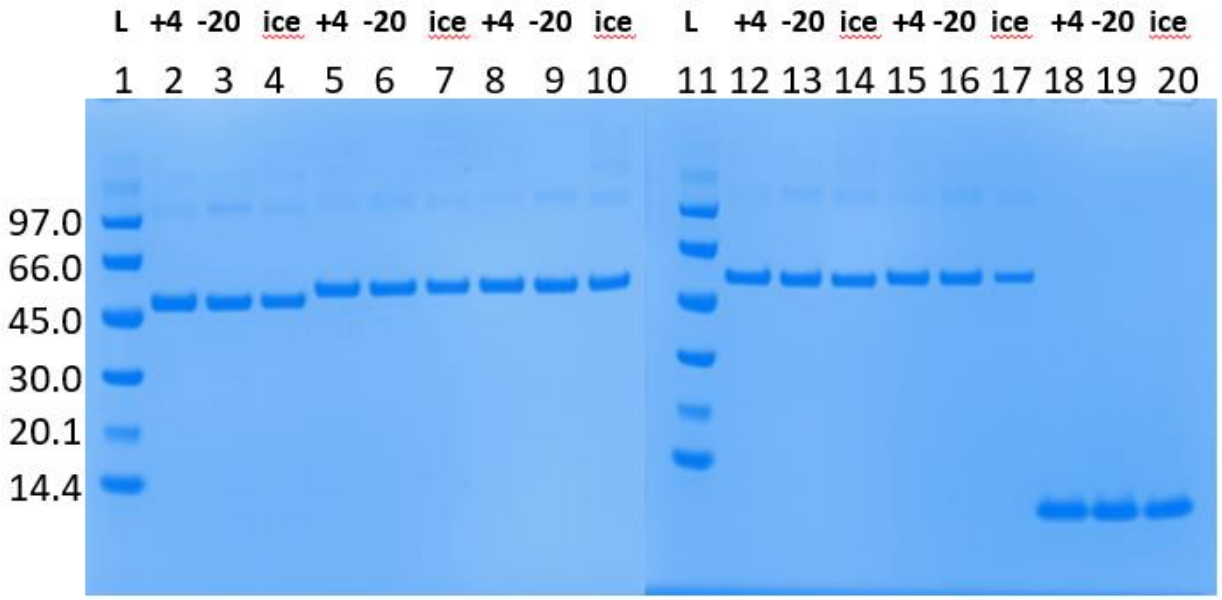


Figure 17: *To assess sample quality after long-term storage 3 µl of each sample was loaded in the order Fridge-Freezer-Ice and 1) Ladder, 2) His_6-YgjG, 3) His_6-YZw1, 4) His_6-YZw2, 5) Ladder, 6) His_6-YZw3, 7) His_6-Yzwideal, 8) Z963-His_6*

Negative stain electron microscopy

The negative stain EM performed in Flemingsberg resulted in multiple micrographs of the different variants, **Fig. 18**. All samples appear to be monodisperse with no visible aggregation, which suggest that they are all suitable for analysis also by single-particle cryo-EM.

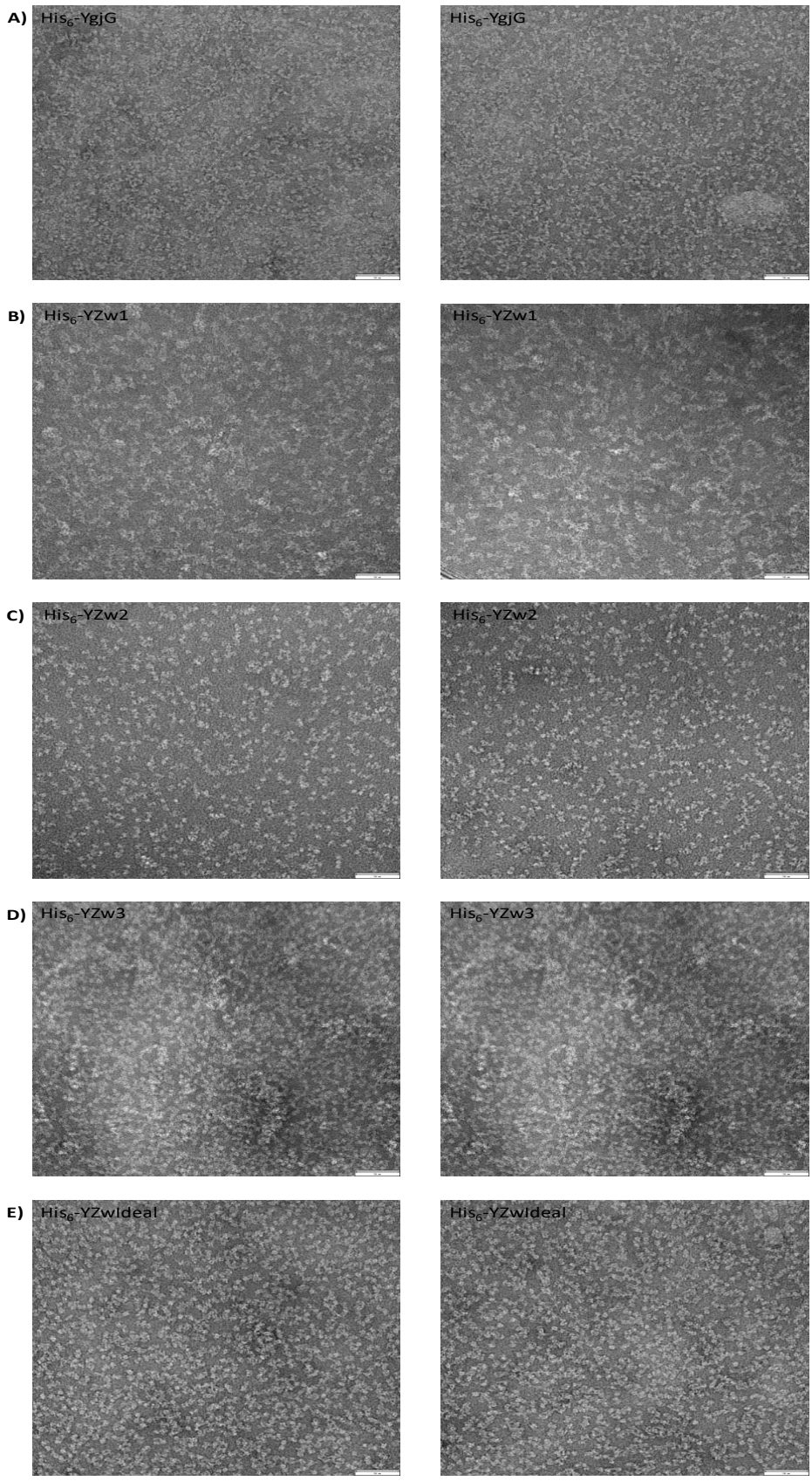


Figure 18: *The figure shows two representative pictures from the same grid of negatively stained samples A) His_6-YgjG (0,056 mg/ml) B) His_6-YZw1 (0,055 mg/ml) C) His_6-YZw2 (0,040 mg/ml) D) His_6-YZw3 (0,085 mg/ml) E) His_6-YZwIdeal (0,082 mg/ml). Scalebar is 50 nm.*

Cryo-EM

Representative micrographs of His_6-Yzwideal mixed with Z963-His_6 on different cryo-EM grids from both Flemingsberg and SciLifeLab shown in **Fig. 19** and **Fig. 20** respectively. Since data processing and modelling are very time intensive processes that are handled via external collaborators with experience in the specific softwares used, no final 2D particle alignments or three-dimensional cryo-EM reconstructions based on these data are currently available.

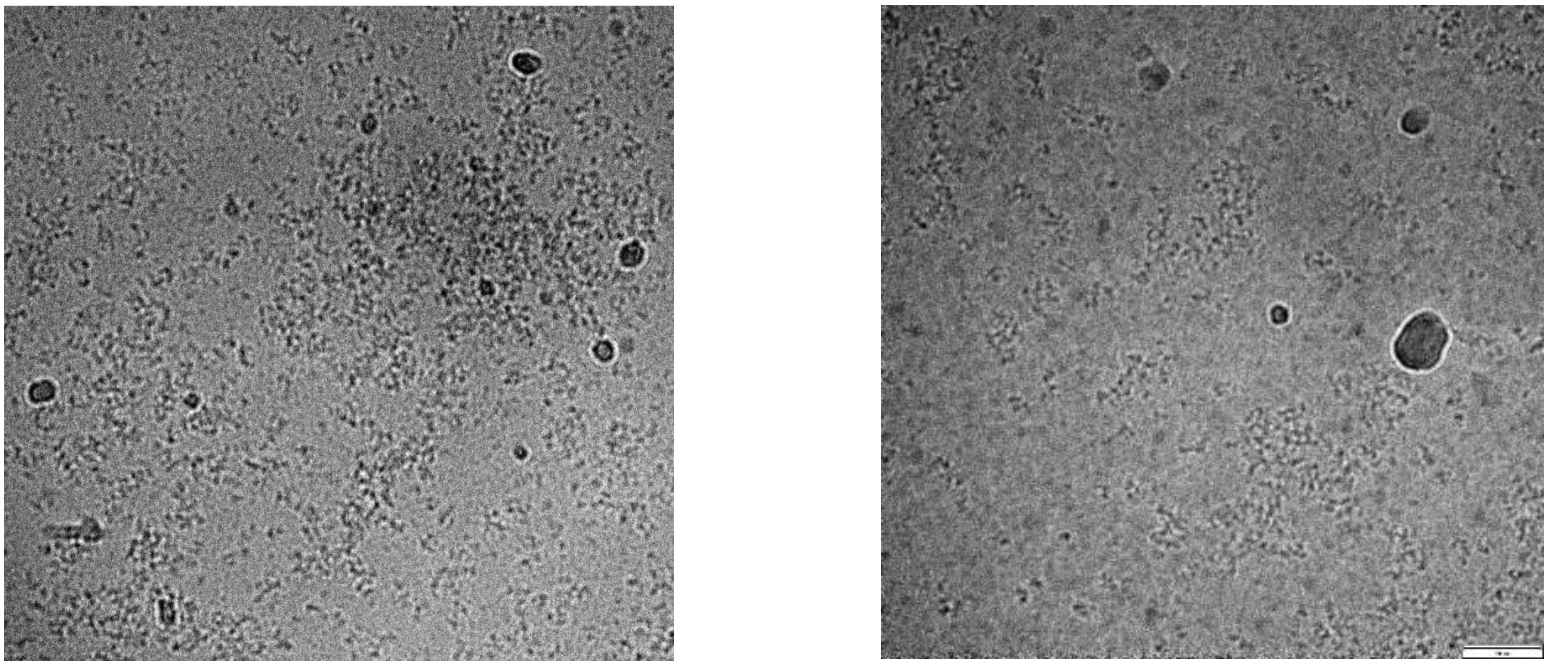


Figure 19: *Micrographs of His_6-Yzwideal (0.683 mg/ml) and Z963-His_6 mixed 1:1 (molar ratio based on YgjG monomer concentrations) cryo-EM coal grids 2 µl hole size, in Flemingsberg. Scale bar is 50 nm. The dark black spots show contamination of crystalline ice.*

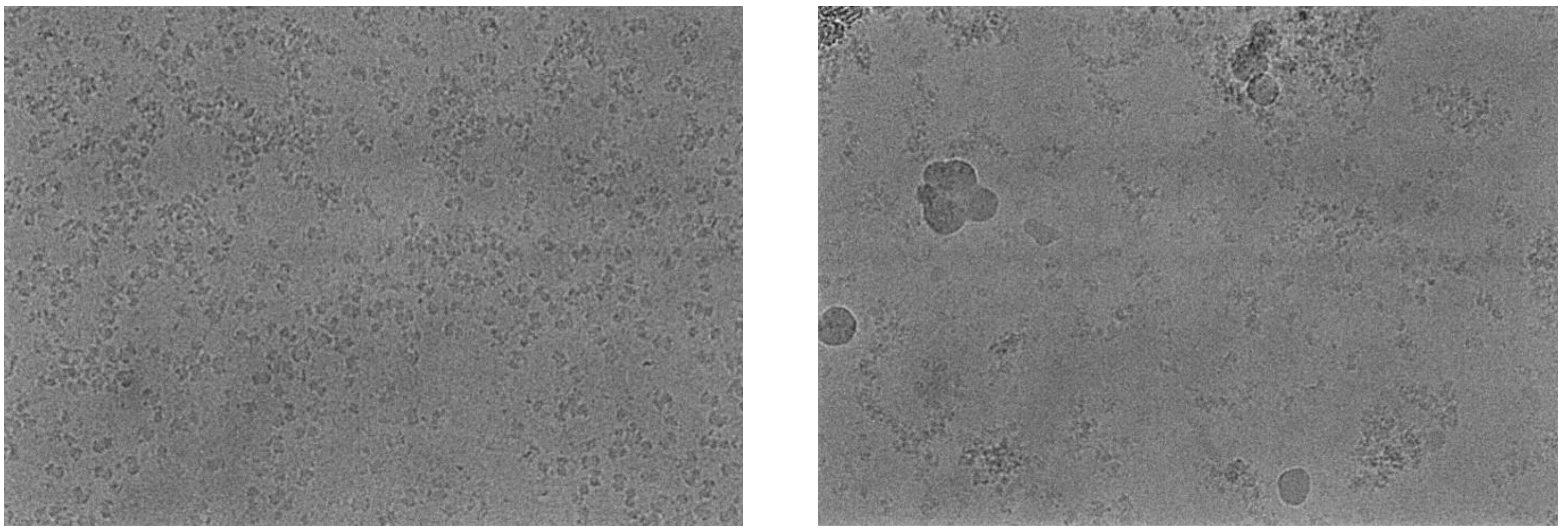


Figure 20: *Micrographs of His_6-YZwideal (1 mg/ml) and Z963-His_6 mixed 1:1 (molar ratio based on YgjG monomer concentration) on cryo-EM gold plated grids 1,2 µl hole size, at SciLifeLab.*

Discussion

The protein production gave high yield and was reproducible for all samples, IMAC purification resulted in very pure samples for all batches based on SDS-PAGE analysis. In future productions it could be beneficial to scale down and express protein in smaller volumes, because of both the high yield and the very small amount of protein used in cryo-EM. In SEC purification the majority of the purified protein eluted in a main peak with an elution time that suggest that samples exist in tetramer conformation, **Fig. 7-9**. No visible amounts of monomer could be seen in any of the purifications, but a small fraction of soluble higher order oligomers was visible and was confirmed by SDS-PAGE to have the same identity as the main sample, appendix **Fig. A3**. Before each SEC purification the sample was both spun and filtered 0.25 µm, therefore the possible presence of insoluble material in the samples is not reflected in the absorbance graphs. Another aspect to consider is the equilibrium between different oligomeric states. While SEC effectively separated the main peak (tetramer) from the soluble oligomers present in the pre-peak, it would be interesting to investigate if the species in the pre-peak would reform over time after purification of tetramers. In the SEC purifications the largest carrier His_6-YZwideal came out first and the smallest YgjG eluted last, however the other variants do not come out in the expected order according to molecular weight. This could perhaps be an indication of flexibility of helices, His_6-YZw3 is larger than His_6-YZw1 and His_6-YZw2 but does not always elute first, this could be favorable as it hints that the molecule is slightly more compact and could also therefore be more rigid. In the CD experiments proteins were shown to be extremely stable and melt at high temperatures, the spectra also confirmed the expectation of alpha helical structures. Protein solutions did precipitate at high temperature and did not show thermal refolding ability. In SPR analyses it was confirmed that binding of Zwt fused to YgjG-carrier to Z963 was functional in all explored designs **Fig. 14**, and the slow release from the chip surface is interpreted as an avidity effect arising from simultaneous binding of at least two functional sites of each carrier. The rate of dissociation is higher for monomer Zwt (data not shown). To assess if additional Zwt moieties on the same tetrameric carrier could be functional at the same time, the dual inject protocol was used. For samples His_6-YZwideal and His_6-YZw3 this experiment showed an increase of signal when antibody was injected after the carrier. Since the slow dissociation rate was retained, one interpretation of this result is that three binding sites on the same tetrameric carrier molecule were active simultaneously. An interesting difference in **Fig. 14** is that His_6-YZw1 and His_6-YZw2 do not bind to the Cetuximab coated surface, this could be caused by a structural hinderance not found in His_6-YZw3 and His_6-YZwideal because of the different helical extensions that result in different relative orientations of the binding interface of Zwt versus the YgjG carrier. However, since all four designs do bind Z963, all carriers are considered functional. There could be benefits with a different orientation of the binding site of the fused Zwt moiety on the carrier, and the different variants could be used for different POIs with different geometries. In the stability study the frozen samples seemed to precipitate the most and their concentrations deviated the most both before and after centrifugation. This could be a consequence of protein precipitation in the freeze-thawing process. It would be interesting to investigate whether the same results would be obtained if samples were frozen at -80˚C instead of -20˚C

The negative stain experiment shows some minor sample aggregation for all samples, but most of the proteins were disperse and looked promising in terms of ability to identify individual protein complexes (“particles”). It is important to note the ~10x dilution used in negative stain compared to the cryo-EM samples, and absence of Z963 could cause the sample to look entirely different on a cryo-EM grid. I believe that it is better to only screen cryo-EM grids instead, because both processes are time consuming. When looking at the sample in cryo-EM grids aggregation was more common than in negative stain. In Flemingsberg different grids with 2 µm holes were used in the optimization process, carrier proteins were mixed with different ratios of Z963, blot times were varied, and concentrations

of carriers changed, but the final grids all had aggregation present. The carrier that seemed to aggregate the least was YZwIdeal and it was also chosen for data collection.

When visiting SciLifeLab national cryo-EM facility, grids of different hole sizes were screened and all carriers were looked at, but because of the previous screening we focused on His_6-YZwideal and His_6-YZw3. The optimal concentration seemed to be 1 mg/ml, and for all samples a constant 1:1 ratio of Z963 to carrier binding site was used. All carriers showed aggregation to some degree, Z963-His_6 had the least aggregates present. Most aggregates were seen on grids with 2 µm holes, less aggregation was seen on grids with the smaller 1.2 µm holes, and much less when using gold plated grids with 1.2 µm holes. After screening multiple conditions, grid types and carriers the sample that stood out was once again His_6-YZwideal and it was chosen for data collection in complex with Z963-His_6 using a high-end Titan Krios microscope.

The HEPES buffer worked well in cryo-EM and gave good contrast. The entire harvesting and purification stage of protein production was done in PBS and PBS based buffers. But HEPES is preferred and therefore buffer exchange through PD-10 columns was a reoccurring procedure. In the preparative steps the protein went through two buffer exchange (PD10) columns and one SEC column, it is difficult to know if this process was harmful for the sample or not. In SciLifeLab it was suggested that considering the purity of the sample after IMAC purification, it could suffice, and PBS could also be tested, this could be a future endeavour, that would save a lot of time but may not effectively remove the oligomeric species observed in the pre-peak of SEC.

The benefit of the proposed method is future possibilities of a general platform and then being able to solve many structures with relative ease. There already exist many affibodies to targets that are interesting to model and solve. The ability and knowledge to select new affibodies to other important targets allow a nearly infinite number of possibilities, the limitations of protein not crystallizing or not precipitating in high concentrations is not a problem with this approach. This proposed method is not very quick, it would take time to both generate and process data but there is no obvious limitation, any desired target could be possible to address. The ongoing shift in the structure biology community towards cryo-EM only makes this method even more appealing.

A functional binding between a free affibody and POI do not guarantee function of the imaging scaffold, therefore different fusions between the affibody and carrier is needed. The SPR data show that only YZwIdeal/YZw3 could bind the mAb while all carriers could bind Z963. These differences exist within versions of one scaffold, therefore generating multiple scaffold-variants in parallel is a must. The varying degrees of aggregation in cryo-EM screening is another example of when multiple variants have been useful, because one has been better than the others.

The process of cryo-EM was initially thought to be very quick, however the data processing is very time consuming to produce a high-resolution model. However, the benefits are still overwhelming, and the method is very attractive compared to e.g, co-crystallization of a sample with a binder.

In this project controls with longer linkers between the carrier and the affibody were cloned and two of the five versions were produced. They were only analysed by SPR and showed some interesting results. The main idea behind them was to compare the flexibility in a method called SAXS, but after deciding that it might also require a lot of effort and data processing, it was not done within the thesis project. Later in the project it became apparent that the best and quickest way of assessing if the scaffold works was to try it directly in a single-particle cryo-EM experiment.

Conclusions from the microscopy occasions was, His_6-YZwIdeal behaves the best and does not aggregate to the same degree as the other samples. Buffer and other properties would be interesting to optimize. SEC purification might not be necessary to perform since the sample is so pure after just IMAC purification, this could reduce time a lot.

Future perspectives.

In the future more carriers should be investigated, but I also believe that there are benefits in developing a general workflow for finding them. It is easy to assess them in parallel, there is no drawback, the bottleneck is finding and designing them, which is not seemingly difficult. From the gained experience of characterising scaffolds, new ones can hopefully be investigated faster.

Acknowledgements

I want to thank my supervisor Johan Nilvebrant for letting me join this exciting project and for always being available when I needed help. Kim Anh Giang for giving great advice about any task I could ever ask about. Per-Åke Nygren for keeping his door open and for asking insightful questions that make me think more. All the nice people on floor 3 for always helping me when I am lost and my fellow thesis students for making the lab experience more fun. Pasi Purhonen for helping me learn more about cryo-EM, and letting me join in on the microscopy sessions, but also to the rest of the group, Hans Hebert, and Caroline Ingeborg Jegerschöld. I want to thank John Löfblom, Linnea Hjelm and Anna Mestre Borras for lending a column and helping me with the SEC purification process which has been a vital part of my project. Lastly, I would like to thank Karin Walldén and Marta Carroni at SciLifeLab for their help in screening, collecting data and providing much insight into the process.

References

1) Curry, S., 2015. Structural Biology: A Century-long Journey into an Unseen World. Interdisciplinary Science Reviews, 40(3), pp.308-328.
2) Callaway, E., 2020. Revolutionary cryo-EM is taking over structural biology. Nature, 578(7794), pp.201-201.
3) Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P., Grigoras, I., Malinauskaite, L., Malinauskas, T., Miehling, J., Uchański, T., Yu, L., Karia, D., Pechnikova, E., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S., Chirgadze, D., Murshudov, G., Aricescu, A. and Scheres, S., 2020. Single-particle cryo-EM at atomic resolution. Nature, 587(7832), pp.152-156.
4) Lawson, C., Patwardhan, A., Baker, M., Hryc, C., Garcia, E., Hudson, B., Lagerstedt, I., Ludtke, S., Pintilie, G., Sala, R., Westbrook, J., Berman, H., Kleywegt, G. and Chiu, W., 2015. EMDataBank unified data resource for 3DEM. Nucleic Acids Research, 44(D1), pp.D396-D403
5) Herzik, M., Wu, M. and Lander, G., 2019. High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. Nature Communications, 10(1).
6) Hu, Y., Cheng, K., He, L., Zhang, X., Jiang, B., Jiang, L., Li, C., Wang, G., Yang, Y. and Liu, M., 2021. NMR-Based Methods for Protein Analysis. Analytical Chemistry, 93(4), pp.1866-1879.
7) Wu, M. and Lander, G., 2020. How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. Current Opinion in Structural Biology, 64, pp.9-16.
8) Wu, X. and Rapoport, T., 2021. Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). *Proceedings of the National Academy of Sciences*, 118(41).
9) Liu, Y., Gonen, S., Gonen, T. and Yeates, T., 2022. Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system.
10) Plückthun, A., 2015. Designed Ankyrin Repeat Proteins (DARPins): Binding Proteins for Research, Diagnostics, and Therapy. Annual Review of Pharmacology and Toxicology, 55(1), pp.489-511.
11) Yao, Q., Weaver, S., Mock, J. and Jensen, G., 2019. Fusion of DARPin to Aldolase Enables Visualization of Small Protein by Cryo-EM. Structure, 27(7), pp.1148-1155.e3.
12) Nilsson, B., Moks, T., Jansson, B., Abrahmsén, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T. and Uhlén, M., 1987. A synthetic IgG-binding domain based on staphylococcal protein A. "Protein Engineering, Design and Selection", 1(2), pp.107-113.
13) Nygren, P., 2008. Alternative binding proteins: Affibody binding proteins developed from a small three-helix bundle scaffold. FEBS Journal, 275(11), pp.2668-2676.
14) Ståhl, S., Gräslund, T., Eriksson Karlström, A., Frejd, F., Nygren, P. and Löfblom, J., 2017. Affibody Molecules in Biotechnological and Medical Applications. Trends in Biotechnology, 35(8), pp.691-712.
15) Lindborg, M., Dubnovitsky, A., Olesen, K., Bjorkman, T., Abrahmsen, L., Feldwisch, J. and Hard, T., 2013. High-affinity binding to staphylococcal protein A by an engineered dimeric Affibody molecule. Protein Engineering Design and Selection, 26(10), pp.635-644.
16) Mori, T., Terashi, G., Matsuoka, D., Kihara, D. and Sugita, Y., 2021. Efficient Flexible Fitting Refinement with Automatic Error Fixing for De Novo Structure Modeling from Cryo-EM Density Maps. Journal of Chemical Information and Modeling, 61(7), pp.3516-3528.
17) Punjani, A. and Fleet, D., 2021. 3D Flexible Refinement: Structure and Motion of Flexible Proteins from Cryo-EM.

18) Youn, S., Kwon, N., Lee, J., Kim, J., Choi, J., Lee, H. and Lee, J., 2017. Construction of novel repeat proteins with rigid and predictable structures using a shared helix method. Scientific Reports, 7(1).
19) Padilla, J., Colovos, C. and Yeates, T., 2001. Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. Proceedings of the National Academy of Sciences, 98(5), pp.2217-2221.
20) Nguyen, H., Park, J., Kang, S. and Kim, M., 2015. Surface Plasmon Resonance: A Versatile Technique for Biosensor Applications. Sensors, 15(5), pp.10481-10510.
21) Weissenberger, G., Henderikx, R. and Peters, P., 2021. Understanding the invisible hands of sample preparation for cryo-EM. Nature Methods, 18(5), pp.463-471.
22) Lorber, B., Fischer, F., Bailly, M., Roy, H. and Kern, D., 2012. Protein analysis by dynamic light scattering: Methods and techniques for students. Biochemistry and Molecular Biology Education, 40(6), pp.372-382.
23) Uchański, T., Masiulis, S., Fischer, B., Kalichuk, V., López-Sánchez, U., Zarkadas, E., Weckener, M., Sente, A., Ward, P., Wohlkönig, A., Zögg, T., Remaut, H., Naismith, J., Nury, H., Vranken, W., Aricescu, A., Pardon, E. and Steyaert, J., 2021. Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. Nature Methods, 18(1), pp.60-68.
24) Wu, X. and Rapoport, T., 2021. Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). Proceedings of the National Academy of Sciences, 118(41).

Appendix

Table A1.A: *Primers used in cloning of flexible designs.*

Construct	Forward Fragment Primer	Reverse Fragment Primer	Forward Vector Primer	Reverse Vector Primer
His_6-YgjG-Zwt	YZ-0.fwd	OF2.rev	OF1.fwd	YZ-0.rev
His_6-YgjG-$(G_4S)_1$-Zwt	YZ-1.fwd	OF2.rev	OF1.fwd	YZ-1.rev
His_6-YgjG-$(G_4S)_2$-Zwt	Ztrimer.fwd	OF2.rev	OF1.fwd	YZ-2.rev
His_6-YgjG-$(G_4S)_3$-Zwt	Ztrimer.fwd	OF2.rev	OF1.fwd	YZ-3.rev
His_6-YZwIdeal-$(G_4S)_3$-Z963	Ztrimer.fwd	OF2.rev	OF1.fwd	Ztrimer.rev

Table A1.B: Sequences of the primers in A.

Primer	Sequence 5'-->3'
YZ-0.fwd	AGTGTCGAAGAAGCGGTAGACAACAAATTCAACAAAG
YZ-1.fwd	GGCGGCGGCGGCTCAGTAGACAACAAATTCAACAAAGAAC
Ztrimer.fwd	GGCGGCGGCGGCTCAGGGGGCGGTGGCAGCGTAGACAACAAATTCAACAAAGAA
OF2.rev	GTGGGGCGCGCCTTATTTCGGCGCCTGAGCATCATTTAG
OF1.fwd	TAAGGCGCGCCCCACCGCTGAGCAATAACTAGCATAAC
YZ-0.rev	TTTGTTGTCTACCGCTTCTTCGACACTTACTCGCATGG
YZ-1.rev	TGAGCCGCCGCCGCCCGCTTCTTCGACACTTACTCGCATG
YZ-2.rev	CCCCCTGAGCCGCCGCCGCCCGCTTCTTCGACACTTACTCGCATG
YZ-3.rev	CTGAGCCGCCGCCGCCGCTGCCACCGCCCCCCGCTTCTTCGACACTTACTCGCAT
Ztrimer.rev	CCCCCTGAGCCGCCGCCGCCGCTGCCACCGCCCCCTTTCGGCGCCTGAGCATCATT

Table A2: *PCR Master mix for fragment amplification*

Reagent	1x reaction	20x reaction
Nuclease Free Water	29.5 µl	590µl
5x Q5 Reaction Buffer	10 µl	200µl
10 uM Primer.rev	2,5 µl	50 µl
10 uM Primer.fwd	2,5 µl	50 µl
2 nM dNTPs (Chase)	5 µl	100µl
Z Template of interest	1 µl (1ng/µl)	-
Q5 High-fidelity DNA Polymerase (0,002 U/µl)	0.5 µl	10µl (Add last)
Total	51 µl	1020 µl

Table A3: *PCR program for amplification of fragments*

STEP	TEMP	TIME
Initial Denaturation	98 °C	30 seconds
35 cycles	98 °C	10 seconds
	60 °C	30 seconds
	72 °C	20 seconds
Final Extension	72 °C	2 minutes
Hold	4°C	∞

Table A4: *PCR mix for amplifying vector – different primers used for different vectors.*

Reagent	1x reaction	20x reaction
Nuclease Free Water	29,5 µl	885 µl
5x Q5 Reaction Buffer	10 µl	300 µl
10 uM Primer reverse	2,5 µl	-
10 uM Primer forward	2,5 µl	-
2 nM dNTPs (Chase)	5 µl	150 µl
YzwIdeal 1ng/µl	(1µl)	-
Q5 High-fidelity DNA Polymerase (0,02 U/µl)	0,5 µl	15 µl
Total	45 (51) µl	

Table A5: *PCR program for amplification of vectors*

STEP	TEMP	TIME
Initial Denaturation	98 °C	30 seconds
35 cycles	98 °C	10 seconds
	65 °C	30 seconds
	72 °C	3 min
Final Extension	72 °C	2 minutes
Hold	4 °C	∞

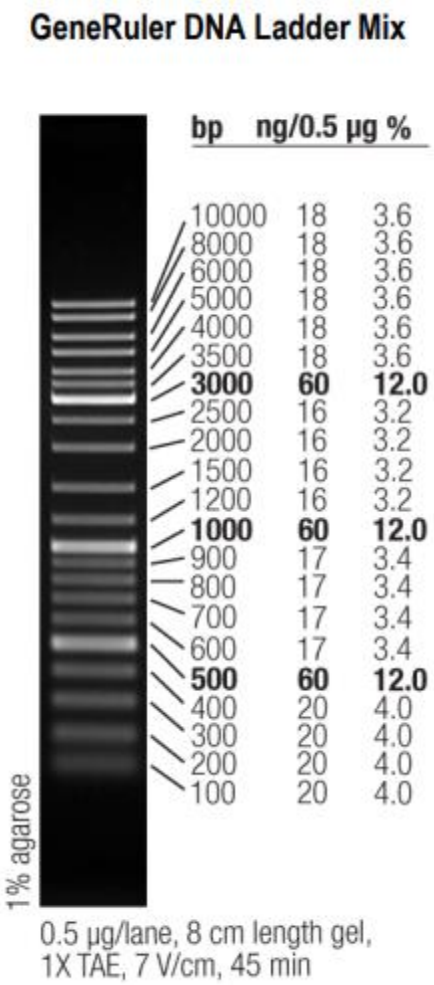


Figure A1: *The ladder used in the DNA gel electrophoresis experiment.*

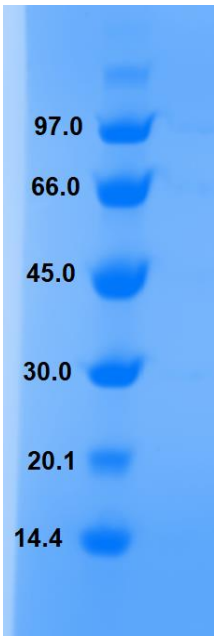


Figure A2: *An SDS-PAGE analysis of the Amersham ladder used in all gels with the program described in materials and methods. All annotated weights are in kDa.*

Table A6: *Protein parameters from (https://web.expasy.org/protparam/) used to calculate mg/ml from Absorbance at 280 nm.*

Name	Exctinction Coefficient	Molecular Weights (Da)
His_6-YgjG	20900	50585,21
His_6-YZw1	22390	55602,85
His_6-YZw2	22390	56032,28
His_6-YZw3	22390	56289,57
His_6-Yzwideal	22390	56989,32
Z963-His_6	8480	7537,31
His_6-YgjG-$(G_4S)_3$-Zwt	22390	57922
His_6-YZwIdeal-$(G_4S)_3$-Z963	22390	64305

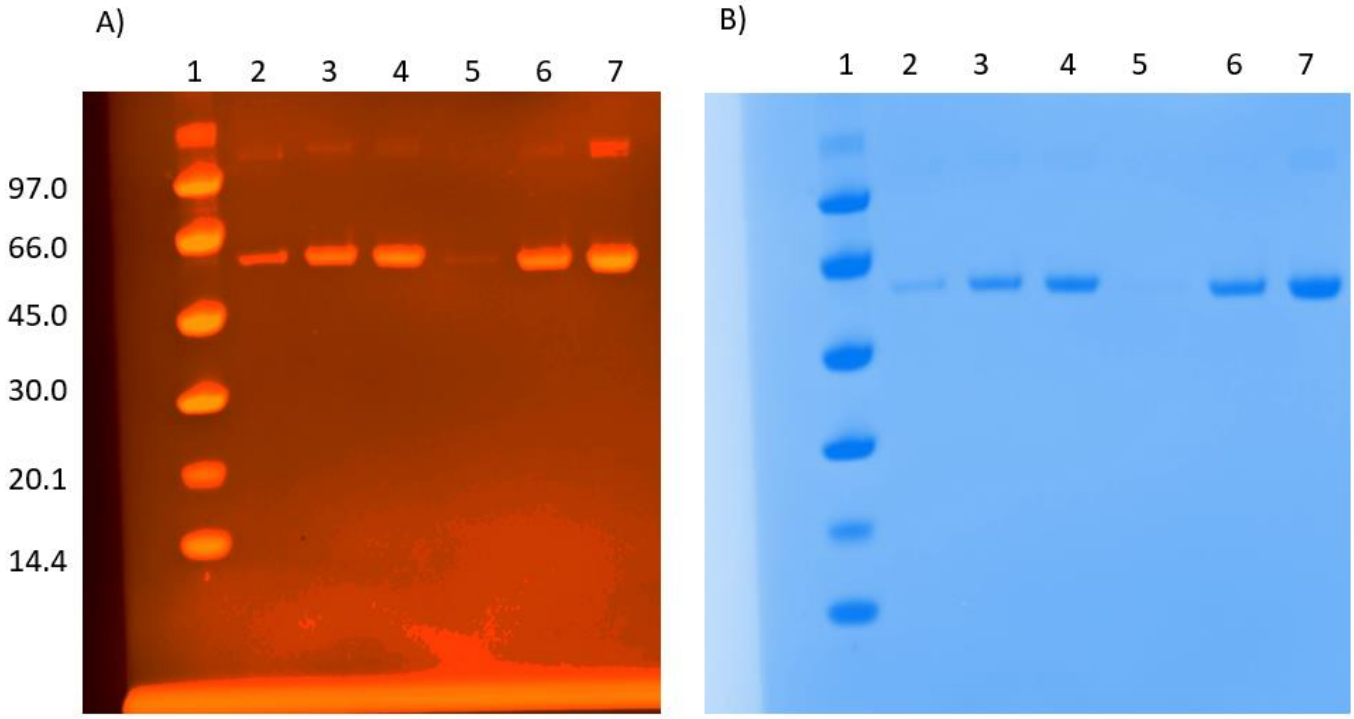


Figure A3: *An SDS-PAGE analysis of the "pre-peaks" in comparison to IMAC purified sample and the "tetramer-peak", A: Gel colors are changed to increase visibility of the bands above, B) The gel without any changes. Most protein appear as a monomer because of SDS-PAGE sample preparation. In well 1: Ladder, 2:"pre-peak" of His_6-YZwideal, 3:"tetramer-peak" of His_6-YZwideal, 4:IMAC purified His_6-YZwideal, 5:"pre-peak" of His_6-YZw3, 6: "tetramer-peak" of His_6-YZw3, 7: IMAC purified His_6-YZw3. For each well 1 µg protein was loaded.*

His_6-YgjG His_6-YZw1 His_6-YZw2 His_6-YZw3 His_6-YZwideal

Figure A4: *Images of the cuvettes after heating to 95˚C.*