Recombinant production, purification and characterisation of the anti-tumour antibody 14F7 single chain Fv

Helene Mykland Hoås

Thesis submitted for the degree of
Master of Science in Molecular Bioscience
60 credits

Department of Bioscience
Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

October 2017

2017

Recombinant production, purification and
characterisation of the anti-tumour antibody 14F7 single chain Fv

Helene Mykland Hoås

http://www.duo.uio.no

Print: Reprosentralen, University of Oslo

Acknowledgements

Finally, my university education has come to an end, at least for the time being. It feels good to know that six years of education and hard work has given me more than a Master's degree. It has given me extreme amounts of joy and laughter, great experiences, friends for life, and love for life. Educational wise I never really knew what I wanted to study before I was right in the middle of it. I always knew that a wanted an education after I finished high school, but I think some of the reasons why I am here today being a clear coincident, but at the same time not.

First and foremost, I would like to give thanks to my main supervisor, Ute Krengel[1]. Thank you for inviting me to your research group with open arms. At our first meeting, I was already introduced to one of the nice and regular social events at the office, the afternoon coffee breaks, which have given me a lot of positive energy and refill this past year. I am grateful to you for believing in my capacity to take on such a big project as the Anti-tumour antibody project, guiding me with your knowledge and feedback and keeping your door open for questions whenever. I would also like to give thanks to my co-supervisor, Geir Åge Løset[2]. Your knowledge in immunology is somewhat breath taking and I am grateful for all the constructive and honest feedback you have given me on this project. Thank you for letting me use your lab at Nextera for ELISA and for finding the time to answer all of my long and confusing e-mails.

When I first started in Ute Krengel's group, Julie Elisabeth Heggelund[3] was working on the anti-tumour antibody project and was kind to give me a thorough introduction to the methods used in the project. You were always available whenever I had a question and I am thankful that you answered every one of my questions as they were equally important, even though I know they were not.
Before Christmas, Kaare Bjerregaard-Andersen[4] gradually took over as my lab supervisor as Julie was preparing to leave for Leeds on a research grant. Kaare, you deserve special thanks for taking on the anti-tumour antibody project alongside all your other projects so that I could have a lab supervisor when Julie left. Thank you for sharing your great knowledge in proteins when working in the lab, analysing results and helping me trouble shoot when things did not go as we would expect. I appreciate your thoughts and feedback concerning both my thesis and the project as a whole, and I am grateful for the time you have spent reading through this thing. I have learned a lot from you and I have enjoyed sharing an office with you this past 18 months.

[1] Professor, Department of Chemistry, University of Oslo, Norway
[2] Scientist, Centre for Immune Regulation and Department of Biosciences, University of Oslo, Norway
[3] Post. Doc, Department of Pharmaceutical Biosciences, University of Oslo, Norway
[4] Post. Doc, Department of Chemistry, University of Oslo, Norway

Thanks to Gertrudis Rojas[5] for the new and good batch of NeuGc GM3 and to Lene Støkken Høydahl[6] and Sebastian Berge-Seidl[7] for helping me with the ELISA experiments at Rikshospitalet and Nextera.

I would also like to thank the other members of Ute Krengel's group for the humorous and light conversations and lunch breaks throughout the year. A special thanks to Joel[8] for helping me with the ÄKTA purifier whenever something went wrong.

A special thanks to my parents for believing in me despite the frequent change in educational path, and for the economic support throughout the years. Thanks to all my friends I got to know from my Bachelor's at the Norwegian University of Lifesciences in Ås, and to my fairly new friends whom I've got to know during master courses here at the University of Oslo.

Last, but not least, a special thanks to my boyfriend, and now husband Even, for the emotional support and cheering during the toughest periods in the lab, and for cooking me dinner whenever I came home hungry and late from working on the thesis.

UiO, October 2017

Helene Mykland Hoås

[5] Doctor, Centre of Molecular Immunology, Havana, Cuba
[6] Doctor, Centre for Immune Regulation and Department of Biosciences, University of Oslo, Norway
[7] Doctor, Researcher at Nextera AS
[8] M.Sc, Department of Chemistry, University of Oslo

Sammendrag

Kreft er en sykdom som rammer millioner av mennesker hvert år. De ulike behandlingsmetodene brukt til å behandle kreft er dessverre fortsatt, stort sett svært skadelige for pasientene og det behøves derfor nye behandlingsmetoder som kan helbrede pasienten uten å forårsake skade. Immunterapi er et lovende felt innen behandling av kreft fordi behandlingen kan rettes mot kroppens eget immunsystem. Når en frisk celle utvikler seg til en kreftcelle, endrer cellens egenskaper seg til å kunne unngå kroppens kontrollerende immunforsvar, og kan derfor vokse og dele seg i all hemmelighet. Heldigvis klarer ikke kreftcellen å skjule seg helt da den ofte uttrykker molekyler på celleoverflaten som er spesifikke for kreftceller, og som kan oppdages av biologiske deteksjonsmetoder. Et slikt molekyl er gangliosidet NeuGc GM3 som sitter på overflaten av kreftceller funnet i brystkreft, hudkreft, lungekreft og fler, og som derfor kan kalles et antigen. Dette antigenet kan bli gjenkjent av det monoklonale antistoffet 14F7 som binder spesifikt til gangliosidet og initierer utvikling av hull i celleveggen på kreftcellene som av reaksjon til dette vil dø. Egenskapene til 14F7 og dens gjenkjenning av NeuGc GM3 er av spesiell interesse da denne interaksjonen kan gi viktig informasjon om hvordan 14F7 gjenkjenner og binder kreftceller.

En krystallstruktur av det antigenbindende fragmentet (Fab) til 14F7 ble løst i 2004, uten å gi noen flere reproduserbare resultater. Dette ledet til utviklingen av små enkeltkjedete fragmenter av antigenbindingssetet (scFv) til 14F7, det minste antistoff-fragmentet som fortsatt inneholder hele bindingssetet. Fire variasjoner av 14F7 scFv ble produsert, og demonstrerte affinitet til NeuGc GM3. Likevel ble ingen krystallstruktur løst. Etter nøye optimalisering av protokollene for proteinekspresjon i *Escherichia coli* og proteinekstraksjon fra periplasma, ble diffrakterende krystaller endelig et faktum. Flere utfordringer med proteinekspresjon, ekstraksjon, rensing og oppbevaring av rent, stabilt protein har likevel vært hovedfokus i denne oppgaven da det har vært vanskelig å reprodusere gode krystaller. Bindingsstudier av 14F7 scFv til NeuGc GM3 ble i tillegg til optimaliseringsutviklingen gjort i ELISA, og stabilitetsstudier ble gjort ved å måle smeltepunktet til 14F7 scFv ved bruk av fluorescens spektroskopi. Resultatene i denne oppgaven poengterer viktigheten av å jobbe ved lave temperaturer når målet er å få rent protein. Dette gir et godt grunnlag for videre krystalliseringsstudier med mål om å få strukturell innsikt i interaksjonen mellom 14F7 og NeuGc GM3 gangliosidet.

Abstract

Cancer is a disease that affects millions of people each year. The different approaches to treating cancer used to date have side effects that are harmful to the patient and the need for non-harmful treatments are therefore urgent. Immunotherapy is the promising field of which the treatment of cancer can be directed to utilise the patient's own immune system. When a healthy cell develops into a cancerous cell, the cell features are altered in such that the cell can hide from the immune system and decide for itself when to proliferate and grow. However, the cancerous cells are not entirely camouflaged as they can express molecules that are not present in healthy cells. One such molecule is the *N*-glycolyl GM3 (NeuGc GM3) ganglioside located on the cell surface of cancer cells found in breast cancer, melanoma, lung cancer and others. This molecule, known as an antigen, is recognised by the monoclonal antibody (mAb) 14F7 which with specific binding to the NeuGc GM3 will cause damage to the cell wall, causing the cell to die. This favourable feature of 14F7 needs further investigation to fully exploit the possibilities of developing cancer treatments. The interaction between the NeuGc GM3 and 14F7 is of specific interest as this interaction can give crucial information about how the 14F7 recognise and bind to the cancer cell.

A crystal structure of the antigen-binding fragment (Fab) of the 14F7 antibody was resolved in 2004, however without any reproducing results. This lead to the approach of developing single chain fragments of the variable domain (scFv) of 14F7, the smallest antibody fragment containing the entire antigen-binding site. Four different constructs of the 14F7 scFv developed in 2014 demonstrated binding affinity to the NeuGc GM3 ganglioside, but unfortunately no structural information was detected. After thorough optimisation of the protein expression protocol in *Escherichia coli* and the periplasmic extraction of protein from the cells, diffracting crystals of the 14F7 scFv was finally obtained. However, many obstacles in production, extraction, purification and storage of pure protein still needed further optimisation, and have been the focus in this thesis. Binding affinity of the 14F7 scFv to the NeuGc GM3 was tested by enzyme-linked immunosorbent assay (ELISA), and stability studies measuring the melting temperature of 14F7 scFv was done by Fluorescent spectroscopy. The results obtained in this thesis points out the importance of low working temperature when the goal is to obtain pure protein. This sets the stage for further crystallographic studies with the goal to obtain structural insight of the binding interaction between 14F7 and the NeuGc GM3 ganglioside.

Abbreviations

bp	Base pair
BSA	Bovine Serum Albumin
C-terminal	Carboxy-terminal
CDR	Complementary-determining region
C_H	Constant Heavy Chain
CIM	Centre of Molecular Immunology, Havana, Cuba
CIR	Centre of Immune Regulation and Department of Biosciences, University of Oslo, Norway
C_L	Constant Light Chain
Da	Dalton
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	deoxyribonucleotide thiphosphate
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediamine-tetraacetate
ELISA	Enzyme linked immunosorbent assay
ESRF	European Synchrotron and Radiation Facility
Fab	Fragment, antigen binding
Fc	Fragment, crystallisable
Fv	Fragment, variable
HAMA	Human anti-mouse antibody
HBS	Hepes Buffer Saline
His-tag	Polyhistidine tag
HRP	Horse Radish Peroxidase
Ig	Immunoglobulin
IMAC	Immobilised metal ion affinity chromatography
IPTG	Isopropyl betha-D-1-thiogalactopyranoside
KD	Equilibrium dissociation constant
Lac	Lactose
LB	Lysogeny broth
L_C	Elongated Cuba linker
L_R	Rikshospital linker

mAb	Monoclonal antibody
mAU	Milli absorption unit
MES	2-(N-monopholino)ethanesulfonate
MQ-H2O	Milli-q
N-terminal	Amino-terminal
NeuAc	N-acetyl neuraminic acid
NeuGc	N-glycolyl neuraminic acid
ON	Over night
P20	Polyoxyethylene sorbitan
PBS	Phosphate Buffer Saline
PBS-BT	Phosphate Buffer Saline with 2 % BSA and 0.05 % Tween-20
PBS-T	Phosphate Buffer Saline with 0.1% Tween-20
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
pL-HRP	Protein L linked with Horse Radish Peroxidase
RNA	Ribonucleic acid
RNase	Ribonuclease
scFv	Single chain Fv
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEC	Size exclusion chromatography
SPR	Surface Plasmon Resonance
TAE	Tris acetate EDTA
TEVp	Tobacco etch virus protease
TMB	Tetramethylbenzidine
V_H	Variable heavy chain
V_L	Variable light chain
$V_{L.A}$	Alternative Variable Light Chain
YT	Yeast extract + peptone (tryptone)

Table of Contents

1 Introduction

1.1 Cancer and cancer treatments

Cancer is the common designation for all diseases caused by uncontrolled cell proliferation, and will in most cases cause tumour formation, with the exception of leukaemia. Millions of people globally are each year diagnosed with cancer. In 2012, the number of cancer incidents worldwide was reported to be 14.1 million and approximately 8 million people died as a result of cancer [1]. The four cancer types that have the highest incident rate for both sexes combined are breast, prostate, lung and colorectal cancer. Of these four, lung cancer has the highest mortality rate. A high mortality rate is also seen in liver cancer, pancreas and oesophagus cancer, and the need for therapy that will prolong the life of patients with these diseases are therefore of much desire [2].

Cancer can be categorised as invasive (malignant) or non-invasive (benign) and the development is often a result of exposure to carcinogenic factors such as radiation, chemical exposure and viral infection. Tumour suppressor genes and oncogenes are specific genes involved in the initiation of tumour growth because their functions are involved in important stages of the cell cycle. Tumour suppressor genes are often involved in the regulation of cell division. Proteins encoded by tumour suppressor genes recognise abnormal cell conditions and order apoptosis or cell cycle arrest to prevent malignant cell formation. Oncogenes originate from proto-oncogenes that encode proteins involved in the signalling pathway of cell division. [3]

1.1.1 Traditional cancer therapy

Because cancer is a disease that causes millions of deaths, early detection, good diagnostics and effective cancer treatments are extremely important. Addressing the problem of rapid division and development of cancer cells, several pharmaceutical drugs have been developed that kill all fast-dividing cells in the body by targeting different stages in the cell cycle pathway. This type of treatment can be effective, but because the treatment targets all fast-dividing cells of the body, a patient treated with chemotherapy may suffer impaired immune system, intestinal distress and hair loss, among others. Radiation therapy is a treatment which targets the tumour directly by using high-energy radiation to destroy the DNA of the cell, which will lead to apoptosis. Compared to chemotherapy, radiation therapy is less stressful for the body because it can target the tumour directly, but because radiation is a known carcinogen, the radiation itself can as an effect lead to new disease. Surgery is often the initial treatment used before radiation- or chemotherapy because the removal of a malignant tumour will limit the cancer invasion to other tissues. Because a malignant tumour cell can invade other tissues, surgery is often accompanied with radiation- and/or chemotherapy afterwards to

limit the invasion. Early detection is therefore very important because it will affect the applicability of surgery and the outcome from tumour removal. [3, 4]

1.1.2 Immunotherapy

Immunotherapy is a rapidly developing method to treat cancer, using the knowledge about the immune system to develop treatments. Antibodies and other components of the immune system all have distinct roles when initiating an immune response and can thus be exploited. With better knowledge of immunological mechanisms, the goal is to use cancer-specific or cancer-associated molecules as targets for immunotherapy. These targets are called antigens. Tumour-specific antigens are molecules only expressed in cancer cells, while tumour-associated antigens are molecules that are present in higher amounts for cancerous cells than for healthy cells. Both types of antigen molecules can be targeted for immunotherapy diagnostics and treatment because they can be recognised by the immune system (Figure 1. 1 with citation [5] therein). [6]

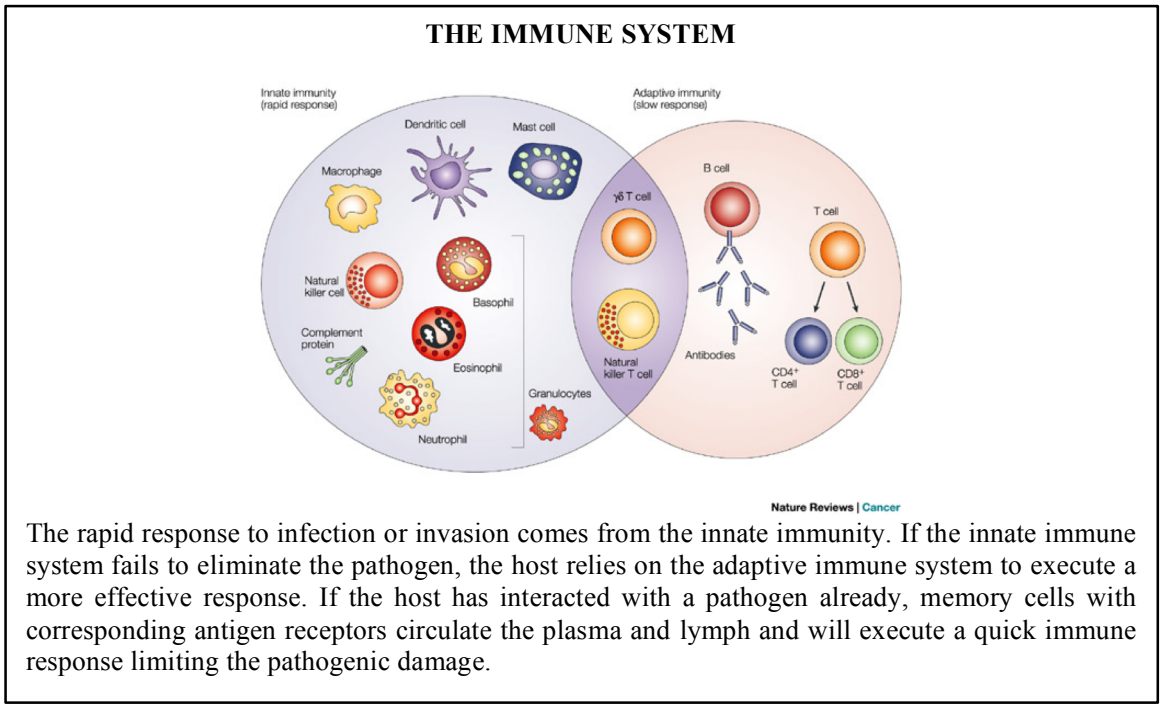


Figure 1. 1: Overview of the two main components of the immune system and the cells involved. Figure reproduced from [7].

Immunotherapy can be both active and passive. Cancer vaccines are one form of active immunotherapy that indirectly targets the cancer by triggering an immune response leading to the hosts own development of antibodies. A cancer vaccine can be used as both treatment and prevention of disease. The passive form of immunotherapy is, in example, the use of monoclonal antibodies to address the cancer in a more direct manner. The antibodies are manufactured to recognise tumour-associated or tumour-specific antigens to the cancer, and will enhance the host's immune response, but will not give any immunity to the host. The Human Papillomavirus (HPV) vaccine is an example

of active immunotherapy. HPV is shown to be the cause of cervical cancer and the vaccine is therefore a preventative against the development of this disease [8]. As treatment, anti-idiotypic monoclonal antibodies (Ab2) (Figure 1. 2) that are developed to mimic a tumour-specific antigen can be used. The immune system will be activated as a response to invasion and produce antibodies (Ab3) directed to kill cancer cells expressing the specific antigen.

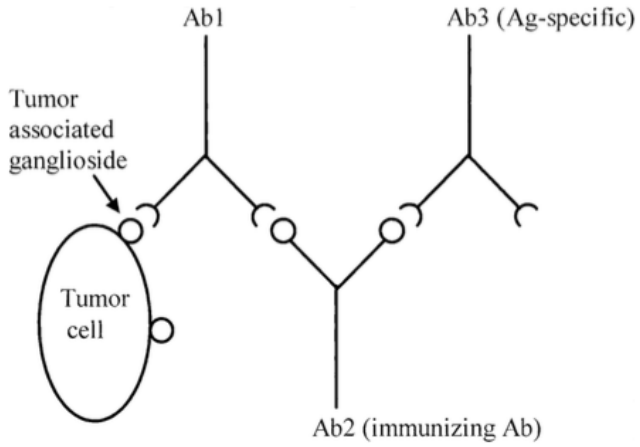


Figure 1. 2: The cascade mechanism of which the anti-idiotypic antibody is produced. Figure reproduced [9]

1.2 Tumour antigens

All cell surfaces display molecules that indicate the intracellular environment of the cell. Molecules that are recognised by the immune system, and thus trigger a response, are called antigens. Such extracellular antigens utilise the immune system to detect disease or irregularities in the cell by evaluating the extracellular signals. Tumour-specific antigens are molecules only expressed in cancer cells, while tumour-associated antigens are molecules that are present in higher amounts for cancerous cells, than for healthy cells [4]. Growth factor receptors and gangliosides are common antigens expressed on cell surfaces and are vastly known for their involvement in cancer [10-12]. Their presence on tumour surfaces can be targets for detection of disease and development of therapy.

1.2.1 Growth factor receptors

Growth factor receptors are tumour-associated antigens expressed on plasma membranes. They are involved in the signalling pathway of the cell by transferring signals through the plasma membrane when activated by a ligand. Genes that encode such cell receptors are often proto-oncogenes that can alter their function with one mutation and lead to uncontrolled cell proliferation. As a result of oncogenic expression, the cancer cells can have autocrine production of ligands and overexpression of growth factor receptors on the cell surface. Both are related to cancer because they cause dysfunctional signalling in such that they are overly activated. Overexpression of the genes encoding human epidermal growth factor receptors 1-4 (EGFR/HER) are well-known for their participation in

cancer disease [10, 13]. Alterations of these genes are found in breast, lung, colon, melanoma and gastric cancer, among others, and are associated with aggressive tumour progression and poor prognosis [14]. A recent study demonstrated that co-localization of EGFR and gangliosides can have an effect on ligand binding to the EGFR [15].

1.2.2 Gangliosides

Glycosphingolipids exist in all mammalian cells and consist of a ceramide with additional subgroups. Gangliosides are glycosphingolipids that have at least three sugar residues, where one must be a N- or O-substituted neuraminic acid residue, a sialic acid [16]. The ceramide part of the ganglioside is anchored in the lipid bilayer of the plasma membrane, exposing its sugars, with the sialic acid, to the extracellular environment. Gangliosides contain multiple important functions such as cell-to-cell interactions, development, inflammation, signal transduction and most importantly tumour progression [17]. There are three types of gangliosides, named GM1, GM2 and GM3. Whereas GM1 has five sugar residues, the GM2 has four, and GM3, being the simplest ganglioside, has only three (Figure 1. 3). One of these sugar residues are the N-acetylated neuraminic acid (NANA) which as a sialic acid residue is specific for the ganglioside family.

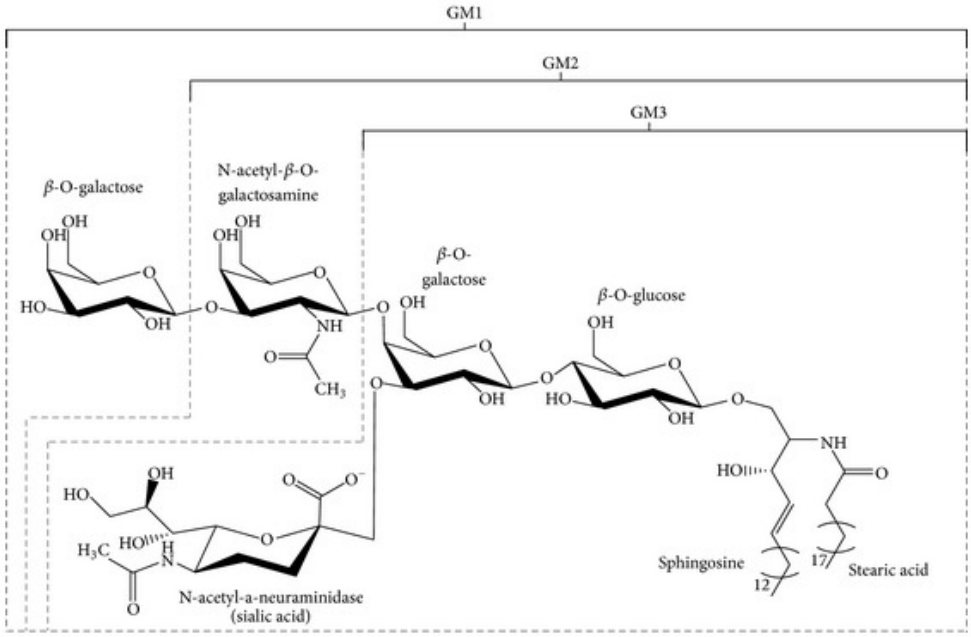


Figure 1. 3: Structural representation of the three gangliosides GM1, GM2 and GM3. Figured reproduced [18].

N-glycolyl GM3

Sialic acids are a common feature on cell membranes of lysosomes, on secreted proteins of the blood serum and mucus, among others [16]. Both *N*-acetylated neuraminic acid (NeuAc) and *N*-glycolated neuraminic acid (NeuGc) are widely expressed in vertebrates. Humans are an exception, only the

NeuAc is expressed, not the NeuGc [19]. The development of NeuGc comes from hydroxylation of NeuAc by the CMP-NeuAc hydroxylase enzyme, which means they only differ by one O-substitution [20]. (See Figure 1. 4 for structural presentation.) Studies show that the hydroxylase responsible for development of the NeuGc from NeuAc is inactive in humans, as a result from a 92 base pair deletion in the *cmah* gene encoding the CMP-NeuAc hydroxylase [21]. This deletion happened approximately 3 million years ago, and has most likely had a great impact on human development from the African great apes. This hypothesis relies on the fact that many pathogens recognise the NeuGc when invading the host, and that the lack of NeuGc in humans can have had a positive effect in natural selection [19, 20, 22, 23]. The small amount of NeuGc detected in healthy, human cells are free and most likely taken up through epithelial cells of the digestive system, arriving from red meat and milk in the diet [24].

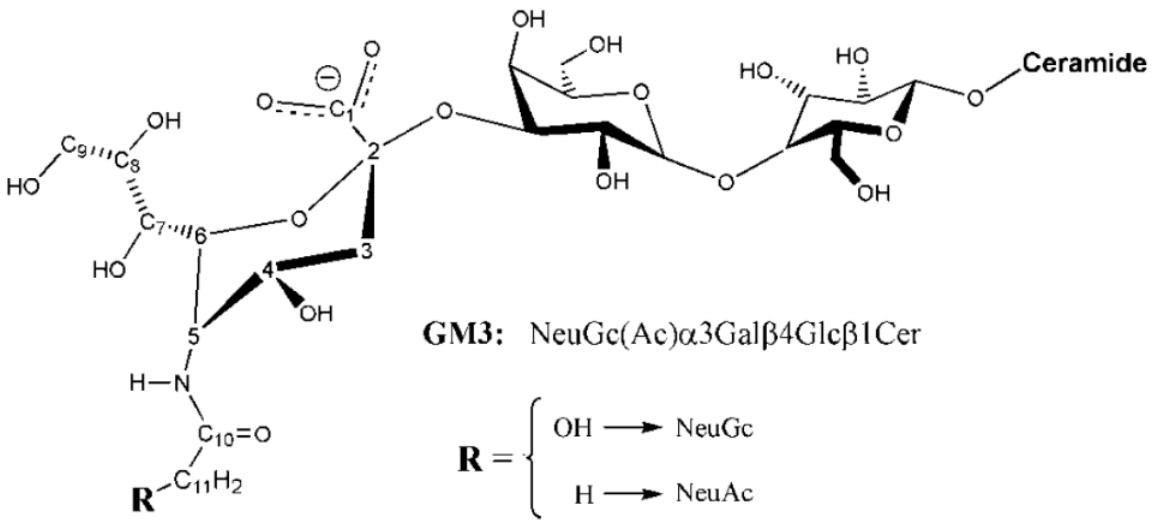


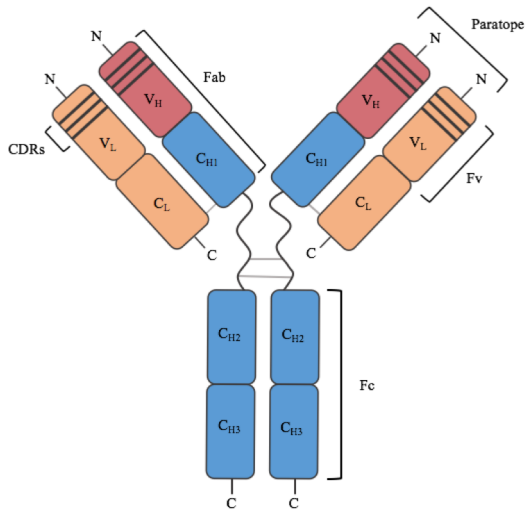
Figure 1. 4: The NeuGc(Ac) GM3 ganglioside. The difference between NeuGc GM3 and NeuAc GM3 is represented by the R-group and shows that only one O-atom separates the two forms. Figure reproduced [25]

Both *N*-acetyl (NeuAc) GM3 and *N*-glycolyl (NeuGc) GM3 is found in the plasma membrane of human cancer cells [26] in cancers such as breast [11, 27], colon [28], melanoma [29] among others. The difference is that the NeuGc GM3 is found in cancer cells only, and that the tumour-specific expression of this ganglioside gives a great advantage in cancer diagnostics and treatment.

The antibody 14F7 specifically recognise the NeuGc GM3 ganglioside [30]. This antibody has been used to produce an Ab2 that is thought to induce B-cell production of NeuGc specific antibodies in the host. [31]

1.3 Anti-tumour antibodies

ANTIBODIES



Antibodies are soluble immunoglobulins secreted from effector B- cells. The two antigen-binding regions (paratopes) in the N-terminus, the variable regions, have unique binding affinities, as the arrangement of gene segments are a result of somatic recombination in the complementarity-determining region (CDR) of the antigen binding site. There are three CDRs in each paratope. CDR3 of the heavy chain is important for antigen binding because diversity obtained from gene rearrangement is abundant in this region. Further diversity in the three CDR regions is activated when the B-cell interacts with the antigen and is obtained through somatic hyper mutation where single point mutations are introduced. As a result, some paratopes have obtained higher affinity to the antigen epitopes and will be selected for affinity maturation. The constant region can be categorised into five classes; G, E, A, M and D, depending on their structural differences. When an adaptive immune response is activated the subclass of immunoglobulin is chosen depending on the surrounding environment of infection, so called isotype switching. Of the five classes, IgG is the most abundant in blood and lymph.

Figure 1. 5: IgG model (to the left): The antibody consists of four peptide molecules; two heavy chains, and two light chains. The light chains are annotated C_L and V_L, and the heavy chains are annotated C_H and V_H, both representing the constant and variable regions of the antibody. The light grey lines connecting the light chain and heavy chain of the Fab unit are disulphide bonds, as so are the two lines connecting the two heavy chains in the Fc unit. The dark grey line connecting the Fc and Fab are flexible regions of the heavy chain, called the hinge region. To the right: A general repetition of antibodies.

The development of monoclonal antibodies with specific binding affinity to tumour-associated or tumour-specific antigens, has enabled targeted cancer treatment by enhancing an immune response to the cancer cells. There are four types of therapeutic monoclonal antibodies: Mouse (murine), chimeric, humanised and human. The chimeric antibody contains the variable part of a murine antibody, with a human constant domain. The humanised antibody contains only the murine complementarity-determining regions, the rest is of human origin. The development of humanised antibodies comes as a result of the human anti-mouse antibody (HAMA) response [32] found in humans treated with murine and chimeric monoclonal antibodies. The response is a consequence of the immune systems recognising the treatment as foreign and will react against it.

The use of monoclonal antibodies is a passive immunotherapy method used for patients with disease, with a non-activated or already activated immune response that will help in directly eliminating the disease. To date, the monoclonal antibodies developed target tumour-associated antigens, which are part of both cancerous and healthy cells. Rituximab is a monoclonal antibody which targets the B-cells, in this case the CD20 receptor expressed on B-cells. By targeting the CD20, the cell is ordered to undergo apoptosis. The treatment was first developed for use in non-Hodgkin's B-cell lymphoma, but is now also being used to treat autoimmune diseases [33, 34]. Trastuzumab is another receptor-

binding monoclonal antibody which targets the HER2 receptor found overexpressed in many patients with breast cancer [35]. The immune system usually produces antibodies against this receptor, but the overexpression of HER2 receptors on cancer cells makes the elimination of cancer cells very difficult to maintain and the cancer quickly develops. Because the monoclonal antibodies target receptors that can be found in healthy cells as well as cancerous cells, the healthy cells will also be affected by the treatment. Therefore, the development of antibodies that have tumour-specific antigen-receptors are a great advantage in immunotherapy because they will only affect the tumour cells. The antibody 14F7 [30] show specific recognition to the tumour-specific antigen N-glycolyl GM3 and can thus be of great potential for cancer treatment.

One monoclonal Ab2 antibody developed for active immunotherapy has shown activation of an immune response against the NeuGc GM3. Racotumomab-alum is a murine Ab2 monoclonal antibody which is shown to induce an antibody response against the NeuGc GM3 [36]. Adjuvated by aluminium hydroxide, the Ab2 demonstrate good antibody activation and a low level of toxicity. Good results in a phase II/III study launched the Racotumomab-alum as treatment for non-small cell lung cancer (NSCLC) in Argentina and Cuba in 2013 [37], and is to date in an ongoing clinical trial in the US [38]. The risk of developing a human anti-murine antibody response to the Racotumomab-alum is there, but no results implying this is presented in the study.

1.3.1 mAbs, Fabs and single chain Fvs

Monoclonal antibodies (mAbs) are complex to express due to their large size, requirement of disulphide formation and post-translational modifications such as glycosylation. To closer examine the binding affinity and other functions of antibodies, the use of biochemical or genetic modifications can be utilised to create smaller fragments of the antibody. Cleaving of an antibody in the hinge region with a protease, such as papain, results in two identical antigen binding fragments (Fab) and one crystallisable fragment (Fc) (Figure 1. 6). A genetic approach with recombinant engineering is used to obtain a single chain variable fragment (scFv) from the Fab. A synthetic linker is introduced between the two gene segments of the variable heavy chain and light chain subunits to create a single polypeptide.

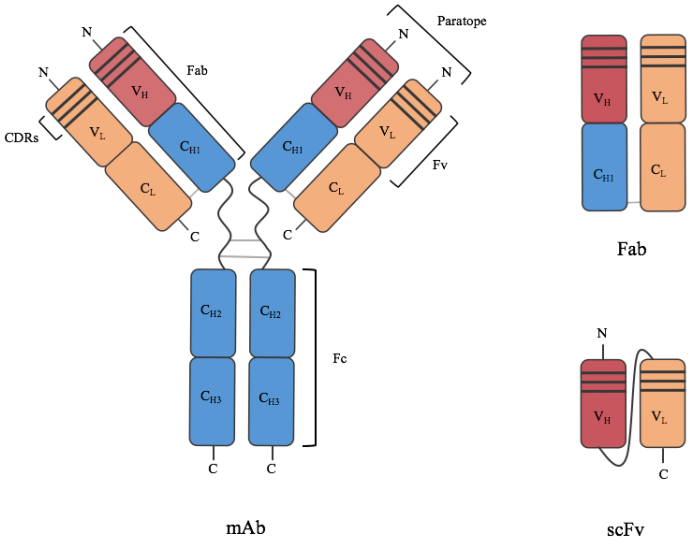


Figure 1. 6: A Schematic figure of the mAb, the Fab and the single chain Fv. The two Fab units on the mAb are identical. The constant heavy chains annotated C_{H1}, C_{H2} and C_{H3} are represented in blue. Both the constant and variable light chains annotated C_L and V_L are represented in light orange. The variable heavy chain annotated V_H is represented in red. The light grey lines connecting the light chain and heavy chain of the Fab unit are disulphide bonds, as so are the two lines connecting the two heavy chains in the Fc unit. The dark grey line connecting the Fc and Fab are flexible regions of the heavy chain, called the hinge region. The Fab unit represent the fragment that is a result of papain cleavage. The two subunits of the Fv are connected together through introduction of a flexible linker, giving the single chain variable fragment, scFv.

The introduction of prokaryote expression systems to limit the time and costs of antibody production reviled improper folding of mAb *in vivo* [39]. The lack of correct folding proteins in prokaryote expression systems makes the production of mAb difficult. Reducing the size of the complex down to one Fab fragment or a scFv with the same paratope, enables for recombinant production of an antibody fragment with the same binding affinity/antigen recognition. In addition, the need for glycosylation of the Fc for effector functions is removed [40, 41], which excludes the glycosylation problem of prokaryote expression systems [42]. Both Fabs and scFvs can be used in phage display, a prokaryote selection method for antigen affinity such as affinity maturation. This is a much more efficient method than screening in hybridomas, which is time consuming and more expensive, and the selection of stable antibody fragments using phage display has the favourable advantage in that they are already cloned [43, 44]. The small size of the scFv also makes the fragment a good candidate for crystallisation to closer examine the paratope structure [45].

1.3.2 The monoclonal antibody 14F7

The monoclonal antibody 14F7 was first reported in 1998 [46] as a murine monoclonal antibody generated from immunized Balb/c mice. By immunizing the Balb/c mice intramuscularly with a vaccine containing NeuGc GM3, Carr *et al.*[30] reported that this antibody was an IgG_1 antibody, specific for the NeuGc GM3 ganglioside. For immunization, the NeuGc GM3 was coupled hydrophobically to a human very low-density lipoprotein (VLDL) in presence of Freund's adjuvant. The monoclonal antibody 14F7 was the first antibody detected to have specific and selected affinity to the NeuGc GM3 ganglioside with a K_D of 25 nM [47], and showed effective killing of murine tumour cells, both *in vitro* and *in vivo*. [30, 48]. 14F7 mAb has been reported to show immunoreactivity to the NeuGc GM3 in breast cancer [27, 30], retinoblastoma [49], skin neoplasms [48, 50], lymphoid tumours [51], malignant epithelial tumours [52], colon cancer [28], lung cancer [53] and others [46, 54]. The variable domains of the murine 14F7 was cloned onto a human IgG_1 with κ constant domain to create a chimeric antibody [55], but because a chimeric antibody can provoke a human anti-mouse antibody (HAMA) response, the antibody was further humanised [55, 56]. The 14F7 does not show the typical cytotoxic mechanisms of cell death, but is shown to induce lesions in the plasma membrane resembling complement reactions such as oncosis [55]. The specific affinity to NeuGc GM3 and the selective binding over NeuAc GM3 makes the 14F7 mAb an important subject for cancer treatment and diagnostics. In 2006, a phase I clinical trial was reported with the murine 14F7 mAb as a passive treatment, in addition to general treatments, for patients with stage II breast cancer. The successful outcome resulted in advancement to phase II.

The Fab structure of 14F7 was solved by X-ray crystallography in 2004 [25] and indicated that three amino acids in the CDR H3 loop was specifically important for the binding of 14F7 mAb to NeuGc GM3. The Fab structure showed similarity to previous reported structures, with the exception of this CDR H3 loop with its 16 amino acids residues. To compare the paratope with other antibody fragments, the Kabat numbering scheme was used to keep the more constant residues in frame. Therefore, eight residues in the CDR H3 loop is annotated 100A to 100H. Tryptophan (Trp) H33 (H for the heavy chain), Tyrosine (Tyr) H50 and Aspartic acid (Asp) H52 showed in a docking model potential binding to the OH-methyl group of the NeuGc with nearby positions. Binding experiments with two samples of mutated Asp H52 to Isoleucine (Ile) and Valine (Val) showed unsuccessful binding of these mutants to the NeuGc GM3, and thus indicated the important role of Asp H52 [25]. Tyr H100D, Tyr H100E and Tyr H100F was proposed to have a stabilising effect on the supposedly flexible CDR H3 loop, making it rigid enough for crystallisation. Asp H98, H100 and H100A was said to have an impact on the packaging of the crystal in addition to the tyrosine residues. From engineering studies [57] such as site-directed mutagenesis, the Tyr H100D was once more indicated to have an important role in NeuGc GM3 recognition together with Trp33, however disconfirming the

Asp H52 and Tyr H50 of their important roles in binding. Arg H98 was introduced as a new and critically important amino acid for the binding affinity, intolerant of amino acid substitutions.

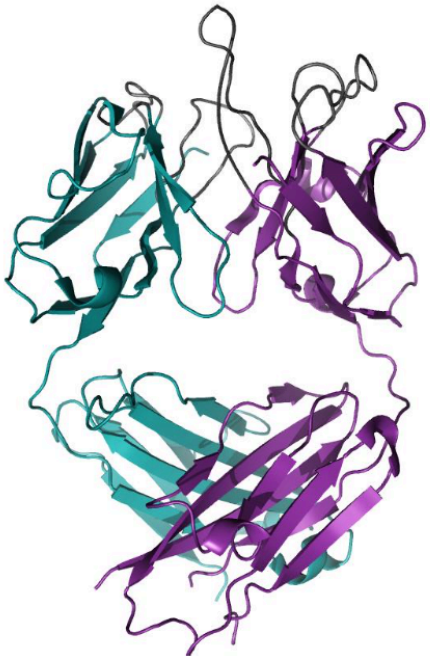


Figure 1. 7: The 14F7 Fab structure solved by X-ray crystallography in 2004 by Krengel *et al.*[25]. PDB code 1RIH. The V_H is coloured violet, and the V_L teal green. The CDR loops are coloured grey, with the CDR H3 loop in the top centre of the structure.

1.3.3 Experimental challenges

The benefits of creating scFvs are their ability to be used in phage display, their small sizes and thus their beneficial use in crystallisation experiments. The small sizes of scFvs has also shown better penetration through tumour cells than antibodies enabling for increased site-specific treatment [58]. ScFvs have also shown potential in toxic-coupled treatment because of the short circulating half-life, meaning that scFv can be rinsed from the host before affecting other parts of the body than the treatment area [43]. However, the benefits of using scFvs also come with many drawbacks, such as protein aggregation and precipitation, which again affect the quality and yield of protein [59]. The physical stability of scFvs, and their short half-life, has shown to be directed to specific sites in the protein sequence, such as the interface between the constant and variable regions of the Fv, replaced by the linker domain in the C-terminus for V_H and no replacement in the V_L. The composition of amino acid residues in the linker is responsible for correct and favourable folding of the two variable domains. To avoid interference with the hydrophobic interfaces of the variable domains, the linker usually consists of hydrophilic residues. However, since stability of the scFv has proven to increase with more hydrophobic residues in the constant/variable interface [60], engineering of the linker can be somewhat conflicted with the need for both correct folding and good stability of the protein [45].

1.4 Method related theory

1.4.1 Prokaryote expression of recombinant single chain Fvs

Expression of antibodies and antibody fragments in eukaryotes are time-consuming and expensive. In addition, the expression of functional monoclonal antibodies is dependent on glycosylation. With the development of Fabs and scFvs with no need for glycosylation, the utilisation of prokaryote expression strategies has been possible. By taking advantage of the secretion machinery in prokaryotes, the expression of antibody fragments can exceed eukaryote expression. *Escherichia coli* is by far the most used organism for prokaryote expression of recombinant antibody fragments [42, 61, 62]. This is mainly because the RNA polymerase in *E. coli* associates with the promoter in a higher rate than for eukaryotic cells, giving higher yield of scFv. In addition, the promoters in *E. coli* are often regulated by a repressor. This enables for manipulative expression in the way that the removal of repressor can initiate/enhance the expression of the desired scFv genes.

lac promoter control

The *lac* operon in *E. coli* is extensively used for prokaryote expression systems because the transcription of genes is controlled by the presence or non-presence of lactose. With no lactose present, the *lac* repressor will bind to the *lac* operator, just downstream of the *lac* promoter, and inhibit the binding of RNA polymerase to the *lac* promoter. With lactose present, the *lac* repressor will interact with allolactose, an isomer form of lactose, and the *lac* repressor will disassociate from the *lac* operator and transcription is activated. For *E. coli* cells, glucose is the most favourable energy source and will be used up first before other energy sources, such as lactose. Glucose is therefore a repressor of the *lac* operon because the uptake of lactose into the cell is repressed. In addition, for the XL1-Blue strain of *E. coli*, is the presence of lacIq in glucose-rich media, which also functions as a *lac* repressor. By removing glucose from the growth media, the uptake of lactose as an energy source will activate the transcription, and the genes coupled to the *lac* promoter will be transcribed. [3]

Periplasmic folding of scFvs

The reducing environment of the cytosol in prokaryotes [59] is troublesome for the folding of disulphide bonds in scFv, which requires a more oxidising environment [63]. Because disulphide bonds are of great importance for scFv stability, the inclusion of a signal sequence for translocation to periplasm (e.g. pelB) is necessary [59, 64]. The oxidising environment of the periplasm then enables for correct folding of scFv. The pelB signalling peptide enables translocation via the secB pathway for correct folding in the periplasm [65], for then to be extracted to the extracellular medium to avoid host toxicity in the cell [66]. This will also decrease the amount of insoluble aggregates in the periplasm due to high protein concentration [65].

Chaperone assistance

To help with folding of scFv, a chaperone can be introduced to be co-expressed with the scFv. In *E. coli,* the chaperone FkpA, a periplasmic peptidyl-propyl cis-trans isomerase, occurs naturally in the cell and can be utilised in scFv expression because it is localised in the periplasm [67-69]. Chaperones have been reported to not only help with folding of scFvs, but also prevent aggregation and protein degradation [68] [70, 71] and is thus very beneficial for scFv expression.

14F7 scFv

The periplasmic expression system used to create the 14F7 scFv of this project was developed in *E. coli XL1-Blue* cells with the vector pFKPEN specifically developed for this strain [72-74]. A cut-out of the vector pFKPEN in Figure 1. 8 shows the presence of the *lac* operon for transcription of recombinant scFv, a Shine-Dalgarno sequence important for ribosome binding to mRNA, and a pelB sequence coupled to the N-terminus of the scFv sequence for transmembrane translocation to the periplasm. In addition, the chaperone FkpA and its native promoter for folding-assistance and prevention of aggregation.

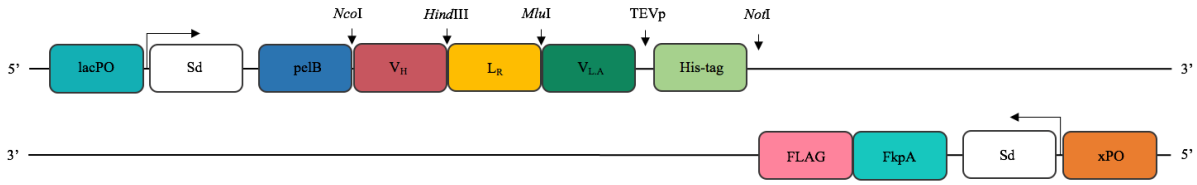


Figure 1. 8: A cut-out from the pFKPEN vector representing the *lac* operon with its promoter and operator (lacPO), the Shine-Dalgarno sequence (Sd) and the scFv sequence with its N-terminus pelB and C-terminus His-tag. On the opposite strand is the FkpA chaperone with its C-terminus FLAG-tag, controlled by its native promoter xPO. The design is inspired and adjusted from [73, 75].

1.4.2 Crystallisation

X-ray crystallography is a common technique used to determine protein structures. A protein crystal is exposed to X-rays, which will give a reflexion signal (scattering) known as the diffraction pattern. X-rays have the wavelength of the same magnitude as the distance between atoms which enables for the detection of detailed structures. However, one molecule is not enough in X-ray crystallography, because the diffraction power for one molecule is too low. This is the reason for the development of crystals, which are repeating patterns of a molecule. The lattice of the crystal will amplify the scattering signal and the given diffraction pattern can be used to calculate the structure. The crystallisation of proteins is the limiting factor in X-ray crystallography because the protein needs to grow into crystals before exposing them to X-rays. [76]

Vapour diffusion is a common crystallisation method used for gentle and gradual changes in the concentration of protein and precipitant, which will aid in crystal growth. The growth of crystals in vapour diffusion follows the phase diagram (Figure 1.9), where the increase in both protein concentration and adjustable parameter, such as precipitant, additive or buffer, will decrease the protein solubility. The vapour diffusion system is a sealed system to get equilibrium between the drop and the reservoir solution. The correct ratio between precipitant, additives and buffers will allow the proteins to enter the nucleation zone, where they can organize in a repeating pattern to create small crystals (nucleus) before they enter the metastable zone where the nucleus will grow into crystals. [76, 77]

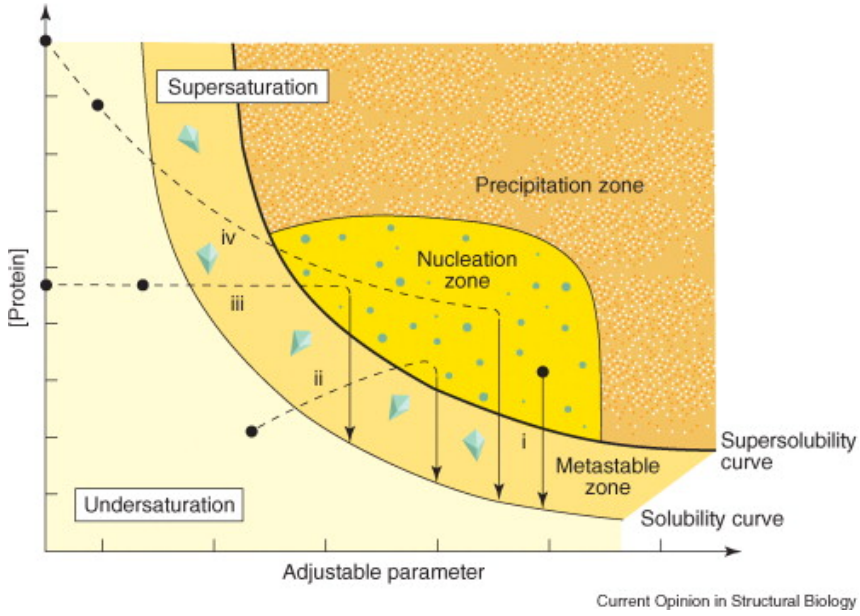


Figure 1. 9: The phase diagram representing the four phases of protein saturation. The protein must enter the nucleation zone to create nucleus that will grow into large crystals when entering the metastable zone. Both protein concentration and adjustable parameter can be altered to obtain the best crystal growth conditions. Figure reproduced from [77].

1.4.3 ELISA

Enzyme-linked immunosorbent assay (ELISA) is used for characterisation studies such as determination of antibody specificity. By coating an ELISA plate with an antigen, the antigen-binding fragment can be applied to the wells and bind to the antigen. The binding can be detected by an enzyme-linked detection protein with binding properties to the antigen-binding fragment. In the case of 14F7-derived scFv, protein L (pL), which has good binding properties to the variable kappa light chain [78], is linked with Horseradish Perioxidase (HRP). The addition of tetramethylbenzidine (TMB) substrate will activate the bound pL-HRP in the wells and give a colour that can be measured quantitatively by spectrophotometry. The mean absorbance can be plotted against the protein concentrations to obtain a standard curve of the binding affinity. A positive control must be included in the experiment, with known binding affinity to the antigen coated on the plate. In the case of 14F7, the 14F7 mAb was used as a positive control for scFv. Negative controls should also be included to

verify the results. Such controls can be scFvs and mAbs that are known to have other binding properties. [79]

1.4.4 Fluorescence spectroscopy

Fluorescence spectroscopy is a method used with biomolecules containing fluorescent moieties, such as aromatic benzene rings, that will absorb and re-emit visible light at different wavelengths. The re-emission of light will take place at a wavelength corresponding to a drop of energy. The difference in intensity between the absorption wavelength and the emission wavelength is known as the quantum yield which can change between two states of the fluorophore. The fluorophore is sensitive to its environment (dielectric potential), which will affect the wavelength of the emitted light. For a biomolecule such as the 14F7 scFv, the aromatic rings in tryptophan is a fluorophore that can be detected by spectroscopy. Knowing the localisation of tryptophan residues in the binding site and in the hydrophobic core in each immunoglobulin fold (Figure 1. 10) from the Fab structure [25], the change in emission can be detected. The change in dielectric properties of the environment of the fluorophore (tryptophan) will change the spectrum of emitted light. The change in emission is expressed as a change in the fraction between emission at 350 and 330 nm (F350/330) which can be used as a measurement of peak movement towards a longer wavelength. This change can be plotted against a rising temperature in the sample, which will reveal at which temperature the tryptophan is fully denatured. [80]

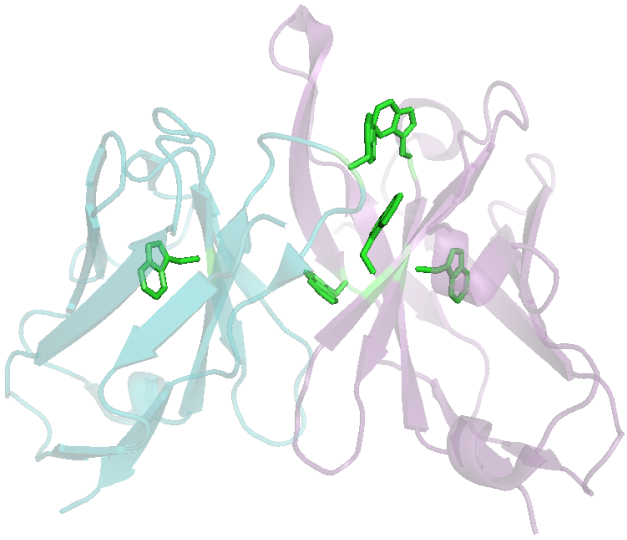


Figure 1. 10: The 14F7 paratope adjusted from the 14F7 Fab structure (PDB: 1RIH) in PyMol (Schrödiger). The V_H in violet have four tryptophan residues, aromatic rings shown in green, while the V_L in teal green have two tryptophan residues.

1.5 Project background

Because Krengel *et al.* [25] did not manage to obtain new crystallographic data of the 14F7 Fab, the scFv constructs were created to increase the possibility of crystals for collection of structural data. Previous studies indicated that the V_H have the most pivotal role in antigen binding, and that the V_L segment of the paratope could be more or less altered without affecting the binding affinity [57]. The creation of scFv constructs was performed with two different V_L, the original 14F7 V_L and a new, alternative V_L from the 3Fm scFv found by Rojas *et al.*[47] to have high affinity to the NeuGc GM3. This alternative light chain (named $V_{L.A}$) was introduced into 14F7 scFv to further investigate the binding affinity to NeuGc GM3. Hedda Johannesen [75] with supervisors hypothesized that the original linker produced in Havana [47] was too short for the folding of the scFv to be optimal and added four new amino acid residues to incorporate flanking restriction sites *Hind*III and *Mlu*I, giving the new linker (named L_C) a length of 21 amino acid residues (Figure 1. 11). An alternative linker was also introduced to investigate if the amino acid composition of the linker domain played an important role in folding and solubility. This linker (named L_R) was already integrated in the pFKPEN vector [73] (Figure 1. 11). The four alternatives of the 14F7 scFv were named: C1/C1*: V_H -L_R-$V_{L.A}$; C2/ C2*: V_H -L_R-V_L; C3/ C3*: V_H -L_C-$V_{L.A}$; C4/C4*: V_H-L_C-V_L. The * represents the four alternative constructs without a N-terminus His-tag, that was originally incorporated for purification and detection purposes. The His-tag was incorporated in the N-terminus of the V_H region of the 14F7 scFv, the region showed by both Krengel *et al.* [25] and Rojas *et al.* [57] to be important for binding to NeuGc GM3. The His-tag was therefore removed to exclude the possibility of interference with binding. New constructs with a C-terminus His-tag were instead suggested to be made with V_L/$V_{L.A,}$ not being of importance for NeuGc GM3 binding.

L_C: **KL**APQ**A**KSSGSGSESKVD**ARV**
L_R: **KL**SGS**A**SAPKLEEGEFSE**ARV**

Figure 1. 11: The two linker sequences LR and LC. The amino acid residues identical in the two linkers in bold.

2 Aims for the thesis

The main aim for the anti-tumour antibody project is to obtain a better knowledge and understanding about ganglioside recognition by the 14F7 mAb. Using the 14F7 scFv as a representative for the 14F7 mAb will give an understanding on how the various parts of the ganglioside contributes to the binding. In particular, 14F7 derived scFvs may enable the acquisition of structural data revealing the molecular basis of NeuGc GM3 recognition.

Previous studies on the 14F7 scFv found problems with aggregation and protease degradation, and indicated that the scFv is highly unstable. In order to increase the chances of crystallization, these issues would have to be addressed.

These sub-goals where targeted in order to reach the overall goal of the anti-tumour antibody project:

- Thorough optimisation of the 14F7 scFv expression protocol to avoid aggregation and protease degradation, and hence increase both quality and yield of protein
- Optimisation of previously used purification pathways to obtain good, homogenous protein
- Screening for crystallisation conditions
- Development of new 14F7 scFv constructs with C-terminus His-tags to use in an alternative purification pathway
- Quantitative analysis of binding affinity to the NeuGc GM3 ganglioside

3 Materials and Methods

3.1 Cloning

3.1.1 Restriction cleavage

For restriction cleavage reactions, the FastDigest restriction enzymes from Thermo Fisher Scientific were used. 1 µg plasmid DNA or 0.2 µg DNA fragment from PCR was mixed with MQ-H_2O, 10X FastDigest Buffer and the two restriction enzymes of choice to a total volume of 20 µL. The solution was mixed gently and incubated for 10 minutes at 37 °C to allow for best possible activation of the enzymes. To inactivate the enzyme activity, the mixture was heated to 65 °C for 15 min.

3.1.2 Ligation

For ligation of insert DNA and vector DNA, 20 – 100 ng vector DNA was needed for the ligase reaction and the molar ratio of insert DNA needed to be 4:1 over vector. This was mixed together with 10X T4 DNA Ligase buffer, Thermo Scientific T4 DNA Ligase and MQ-H_2O to a total volume of 20 µL. The ligation reaction was incubated overnight in room temperature, 22 °C, to ensure correct ligation. For a negative control, a ligation mixture without insert DNA was made.

The equations used to calculate the volume of insert and vector:

1) $$\frac{ng\ vector \times kb\ insert}{kb\ insert} \times \frac{molar\ ratio\ insert\ (4)}{molar\ ratio\ vector\ (1)} = ng\ insert$$

2) $$\frac{ng\ insert}{{ng}/{\mu L}\ insert} = \mu L\ insert\ to\ use\ per\ 1\ \mu L\ vector$$

3.1.3 Transformation

The ligation mix was divided in two and 10 µL were mixed with 100 µL XL1-Blue $CaCl_2$ competent *E. coli* cells and left on ice for 45 minutes. The cells were then heat shocked in a water bath for 3 minutes at 42 °C and immediately put on ice. 1 mL LB medium was added to each of the transformation solutions and then incubated with shaking at 37 °C for 30 minutes. 100 µL of each transformation solution was plated out on LB agar plates containing ampicillin before concentrating the transformation solution by centrifugation for 5 minutes at 4000 x *g* and discarding some of the supernatant. 100 µL of the concentrated transformation solution was then plated out and the plates were incubated over night at 37 °C. Colonies were picked the next day and incubated over night with 6 mL LB medium for each colony, 0.1 M glucose and 100 mg/L ampicillin at 37 °C. The cultures where then used for isolation of vector DNA, and preparation of glycerol stocks. The isolation of vector DNA was done by the SIGMA miniprep kit and final concentration was measured using the Nanodrop 2000c (Thermo Scientific) at 260 nm. The isolated DNA was then used for sequencing, test cleavage for size determination and further creation of the other constructs.

3.2 Expression protocols

3.2.1 Expression of 14F7 scFv in *E. coli* XL1-Blue cells

A pre-culture of 50 mL 2xYT medium containing transformed *E. coli* XL1-Blue cells (Stratagene), 0.1 M glucose (Sigma) and 100 mg/L ampicillin (AppliChem) was set up to incubate over night at 125 rpm, 37 °C for a gradual scale up. The pre-culture was then transferred to the pre-induction culture, 1:2000, and set to grow for 2-5 hours until OD_{600} reached 0.5-0.8. For transfer of the cells to the induction culture, the cells were spun down at 2100 x *g* in an Avanti J-26XP centrifuge with rotor JLA-8.100 (Beckmann) for 40 minutes and the cell pellet re-suspended in the induction culture. The induction was then run over night at 30 °C to an OD_{600} of at least 1.2. The yield of cells was then spun down at 3300 x *g* in the Avanti J-26XP centrifuge with rotor JLA-8.100 (Beckmann) for 40 minutes and used for periplasmic lysis.

3.2.2 Periplasmic lysis of *E. coli* XL1-Blue cells

For periplasmic lysis, the bacterial pellet and all pre-cooled buffers were kept on ice during the entire procedure. The bacterial pellet was re-suspended in ice-cold sucrose solution, 4 mL/ gram of cells. For easier re-suspension of the pellet, a magnetic stirrer with low intensity was used in a cold room (4 °C) for about 30 minutes, or until the pellet was dissolved, to avoid breaking the cell membrane. The sucrose sample was then spun down at 8500 x *g* for 30 minutes and the pellet was dissolved in $MgCl_2$ solution (BioChemica), 4-5 mL/gram of cells. Lysozyme (Sigma Aldrich) was added to the solution to a final concentration of 0.15 mg/mL and incubated on ice with stirring, or with stirring in the cold room, for 30 minutes. After the last centrifugation at 8500 x *g* for 20 minutes, the supernatant was filtered through a 0.2 µm filter (VWR) and stored at 4 °C if further analysis was not executed instantly. To avoid loss of protein yield, the first purification step was often executed on the same day to rinse the sample of possible degradation agents.

3.3 Protein purification

3.3.1 Affinity Chromatography

Protein L affinity chromatography

Both gravity column with 1 mL Protein L capto resin (GE Healthcare) and a 1 mL Pierce™ Protein L chromatography cartridge (Thermo Fisher Scientific) were used for protein purification. The latter was ordered for higher volumes of protein expression to minimize the purification time and was used with the ÄKTA Purifier-900 (Biosciences, GE Healthcare). Separate columns were used for the different constructs to avoid possible contamination.

Using the gravity column, the purification procedure was done in a cold room at 4 °C. The periplasmic fraction was prepared for the column by applying precooled 10X Binding buffer pH 7.8 (Appendix B: Solutions, buffers and gels) corresponding to 1/10 of the periplasmic volume, to adjust to the binding conditions of the column. The column was equilibrated with precooled binding buffer pH 7.8 and the protein sample was gently loaded onto the column. The column was then washed with 20 mL binding buffer pH 7.8 and the protein was eluted by applying 30 mL of Elution buffer with pH 2.5. The eluate was collected in 5 mL fractions and neutralized to approximately pH 7.8 by adding 1 mL 1 M Tris pH 9. The neutralized fractions were tested on pH paper (Millipore) to be certain of the pH value. After use, the column was neutralized by adding 20 mL 1 M Tris pH 9 and stored in 20% EtOH at 4 °C. All the samples were run on SDS-PAGE according to the protocol and the fractions containing the proteins of correct size (~28 kDa) were pooled and concentrated by spin columns (Millipore) before further purification.

For the purification procedure using the Pierce™ Protein L chromatography cartridge (Thermo Fisher Scientific) the attached protocol was used. Both the binding buffer and elution buffer was made according to the protocol (Appendix B: Solutions, buffers and gels) and cooled to approximately 4 °C. The column was first equilibrated with five times the column volume of binding buffer. Then the periplasmic fraction was applied to the column with a flow of 1 mL/min. Because the ÄKTA Purifier cabinet has a temperature of approximately 10 °C, the bottle containing the periplasmic fraction was held on ice inside the cabinet while applying to the column. The flow-through was collected and put on ice for further analysis before washing the column with binding buffer. When applying the elution buffer to the column the fractionation volume was set to 1 mL per fraction, with the same flow rate of 1 mL/min. One sample from the flow through, one from the wash and all fractions of the eluate were analysed on a SDS-PAGE. Fractions containing a band at 28 kDa, from the same peak, were collected and concentrated using spin columns (Millipore) before further purification. If the flow-through still contained a band at 28 kDa, this could be run once more on the column for a higher yield of protein.

Immobilised metal affinity chromatography (IMAC)

The periplasmic fraction was prepared for the column by applying Tris-HCl and Imidazole to the same concentration as the IMAC Binding buffer, 50 mM and 20 mM, respectively. For a final adjustment, IMAC Binding buffer was added to the periplasmic fraction correlating to 1/3 of the volume (Appendix B: Solutions, buffers and gels). A 5 mL HisTrap™ HP, prepacked with pre-charged Ni^{2+} Sepharose™ High Performance (GE Healthcare) was used to perform IMAC on the ÄKTA Purifier-900 (Biosciences, GE Healtcare). The column was equilibrated with five times the column volume of binding buffer before the adjusted sample was applied to the column. The column was then washed with five times the column volume of binding buffer to rinse off other proteins.

Imidazole was then used as elution buffer to outcompete the scFv bound to the column, and the protein was eluted into 1 mL fractions at a flow of 1 mL/min. The eluted material was then analysed on a SDS-PAGE and dialyzed into PBS for further purification.

3.3.2 Size exclusion chromatography (SEC)

Standard curve and determination of elution volume for the scFv

Both Superdex75 and Superdex200 (GE Healthcare) were used on the ÄKTA Purifier-900 (Biosciences, GE Healthcare) for size exclusion purification. In order to be certain at what volume, the 14F7 scFv would elute from the column, a Kit for Molecular Weights for Gel Filtration (Sigma) was used with a mixture of five proteins ranging from 12.4 kDa to 200 kDa in molecular weight. The mixture was applied to both columns and run at the same flow as for 14F7 scFv. The data from the five protein peaks were then used to make a standard curve for each column by calculating the logarithmic value of the molecular weights and plot this against the elution volumes in Microsoft Excel. When fitting a linear trend line to the single logarithmic values in the graph, a function is given which can then be used to calculate at what volume the 14F7 scFv will elute with the size of 28 kDa.

Size exclusion chromatography of the 14F7 scFv

Before applying protein to the Superdex75/Superdex200, the column was equilibrated with PBS by applying five times the column volume. Concentrated protein sample from the previous purification step was applied to the column by syringe through the injection loop in the ÄKTA Purifier. The column running flow was set to 1 mL/min and fraction size set to 0.5 mL. The fractions with corresponding peaks in the chromatogram were collected and analysed on a SDS-PAGE. The fractions with a band at 28 kDa were pooled and concentrated using the Amicon ultra filter 2 mL (Millipore). The concentration was then measured, and the purified protein sample was flash-freezed with liquid nitrogen and put in the -80 °C freezer.

3.4 Electrophoresis

3.4.1 SDS-PAGE

The samples were prepared for SDS-PAGE by mixing 15 µL sample with 5 µL 4x NuPAGE LDS Sample Buffer (Invitrogen) premixed with 0.1 M DTT (Bio-Rad Laboratories) for a reduced sample and heated to 95 °C for 10 minutes. For a non-reduced sample, the DTT was left out. SeeBlue® Plus2 pre-stained standard (Invitrogen) was used for comparison of protein sizes containing a 28 kDa standard.

3.4.2 Agarose gel

Depending on the desired percentage of agarose gel, the amount of agarose was measured and mixed with TAE buffer. For 1 % agarose gel, 0.5 g of agarose was mixed with 50 mL 1X buffer. The agarose was dissolved by swirling and heating the mixture until boiling in the microwave. When the mixture was approximately 50 °C, 5 µL of ethidium bromide (EtBr) was added and the agarose mixture was left to stiffen in a gel mould with the desired comb for correct number of wells. Stiffened agarose gel was taken out of the mould and placed in the gel running chamber with enough amounts of 1X buffer to cover the gel. Loading dye was added to ensure that the samples did not float out of the wells. 5 µL of DNA ladder was applied to the gel, and 20-25 µL sample was added in each well when using the 8-well comb. The gel was run on 80V for 30-60 minutes depending on the size of the DNA fragments.

3.5 ELISA

3.5.1 Indirect ELISA

The NeuGc GM3 and NeuAc GM3 gangliosides were each diluted in methanol to a final concentration of 10 µg/mL and coated on the Nunc-Immuno 96 MicroWell PolySorp solid plate (Sigma). 100 µL of ganglioside was distributed to each well following the ELISA scheme (Table 3. 1) and left to evaporate overnight. The next day the wells were washed three times with 100 µL of PBS-T (1 x PBS – 0.1 % Tween-20) before 200 µL of blocking buffer PBS-B (PBS – 2 % BSA) was added to each well and incubated for one hour under parafilm. Before adding 100 µL of the protein samples diluted in PBS-BT (PBS – 2 % BSA – 0.05 % Tween-20), the wells were again washed three times with 100 µL of PBS-T. After 1 hour incubation in room temperature (RT), the wells were again washed, and 100 µL of pL-HRP diluted in PBS (1:2000) were added to each well and incubated for one hour at room temperature. The wells were washed three times with PBS-T before adding 100 µL of TMB detection substrate (Chalbiocem). After 5-15 minutes of incubation in room temperature the absorbance was measured at 620 nm. 100 µL of 1M HCl was immediately added to the wells after the absorbance measurement to stop the detection reaction. The absorbance was then measured at 450 nm.

Table 3. 1: ELISA scheme showing which wells where coated with the two gangliosides (blue and red), the dilution series of the scFv and mAb, and the dilution series for the negative controls.

	Blank	14F7 scFv C1* (nM)						14F7 mAb (nM) Positive control				
Gc	0.0	0.7	2.1	6.2	18.5	55.6	166.7	2.5	7.4	22.2	66.7	200.0
Gc												
Gc												
Ac												
Ac												
Ac												
MetOH	0.0	0.0	0.0	-	166.7	166.7	166.7	-	-	-	-	-
	Irrelevant scFv (nM) Negative control						Irrelevant mAb (nM) Negative control					
Gc/Ac	0.0	5.0	200.0	0.0	2.0	200.0	0.0	2.0	200.0	0.0	2.0	200.0

3.6 Crystallisation

Prior to crystallisation setup, the protein solution needed a buffer exchange from PBS buffer, with a high salt and phosphate concentration, to a Tris-HCl buffer with a low salt concentration. ScFv C1* was either dialysed into Tris buffer by using the Dialysis-kit (Thermo Fisher Scientific) or simply prior to crystal tray setup.

In order to find the best crystallisation condition for your protein to reach the supersaturated state, a crystallisation screen is set up. For this experiment, the Oryx4 (Douglas Instruments) robot was used with the Wasp Run computer program (Douglas Instruments) for vapour diffusion in order to save time and protein for the setup. 36.72 µL of 16.5 mg/mL 14F7 scFv C1* in 20 mM Tris and 50 mM NaCl was taken out by the robot and distributed on a 96-well MRC sitting drop crystallisation plate (SWISSCI), with each reservoir containing two drops of a final volume of 0.5 µL consisting of 25 % and 50 % protein solution. The plate was then placed in a dark room with 20 °C to let the crystals grow.

Optimisation screens of the 16.5 mg/mL scFv C1* in 20 mM Tris and 50 mM NaCl were set up with the hanging drop method using different varieties of the conditions from the wells that gave crystals in the structure screen (Appendix I: Crystal screens). One TPP® tissue culture plate with 24 wells was prepared for each optimisation screen by placing a tiny strip of silicon grease on the edge of the wells. One drop with 25 % or 50 % protein solution, mixed with the correct reservoir solution (Appendix I: Crystal screens), was pipetted on to a siliconized cover slide (Hampton Research) and gently flipped upside down and placed over the corresponding well. The cover slide was very gently pressed down

on to the silicon grease to seal the reservoir from external evaporation. The plate was then placed in a dark room with 20 °C to grow. The conditions of the different solutions in the wells were checked after one hour, and then after one day, to exclude possible precipitation, or salt crystals, that could be mistaken for crystals at a later stage.

3.7 Fluorescence spectroscopy

The protein sample was diluted in PBS to a concentration of 1 µM and placed in a 200 µL cuvette. The cuvette was covered with parafilm to avoid evaporation when heating the sample. The cuvette was placed in the spectrofluorometer and the emission of light was detected continuously while the temperature of the sample was slowly raised to approximately 90 °C.

4 Results

4.1 Status at project start

When I started this project as a continuation of Hedda Johannesen's master's thesis, both Noha Abdel-Rahman[9] and Julie Elisabeth Heggelund[3] had made great progress in optimising the protein purification protocol and starting to design new constructs with a C-terminus His-tag (unpublished work). Johannesen [75] managed to express all four constructs, but C2/C2* did not express in the same amount as the three others and was therefore benched for further analysis. All four constructs where designed for expression with a C-terminus His-tag regardless of this finding. Because the results in Johannesen's thesis showed that the scFvs were prone to protease degradation the expression protocol was thoroughly investigated by Abdel-Rahman.

[9] Post. Doc, Department of Biochemistry, Mansoura University, Al Mansurah, Egypt

4.2 14F7 scFv constructs with a C-terminus His-tag

The four constructs were created by fragment PCR amplification and simple restriction cleavage and ligation reactions into the prepared vector pFKPEN (Appendix J: Vector). Figure 4. 1 shows an overview sketch of the DNA and domain compositions of each scFv construct. Figure 4. 2 and Figure 4. 3 show the results from PCR amplification, showing successful amplification of the fragments for C1#, C2# and C4#.

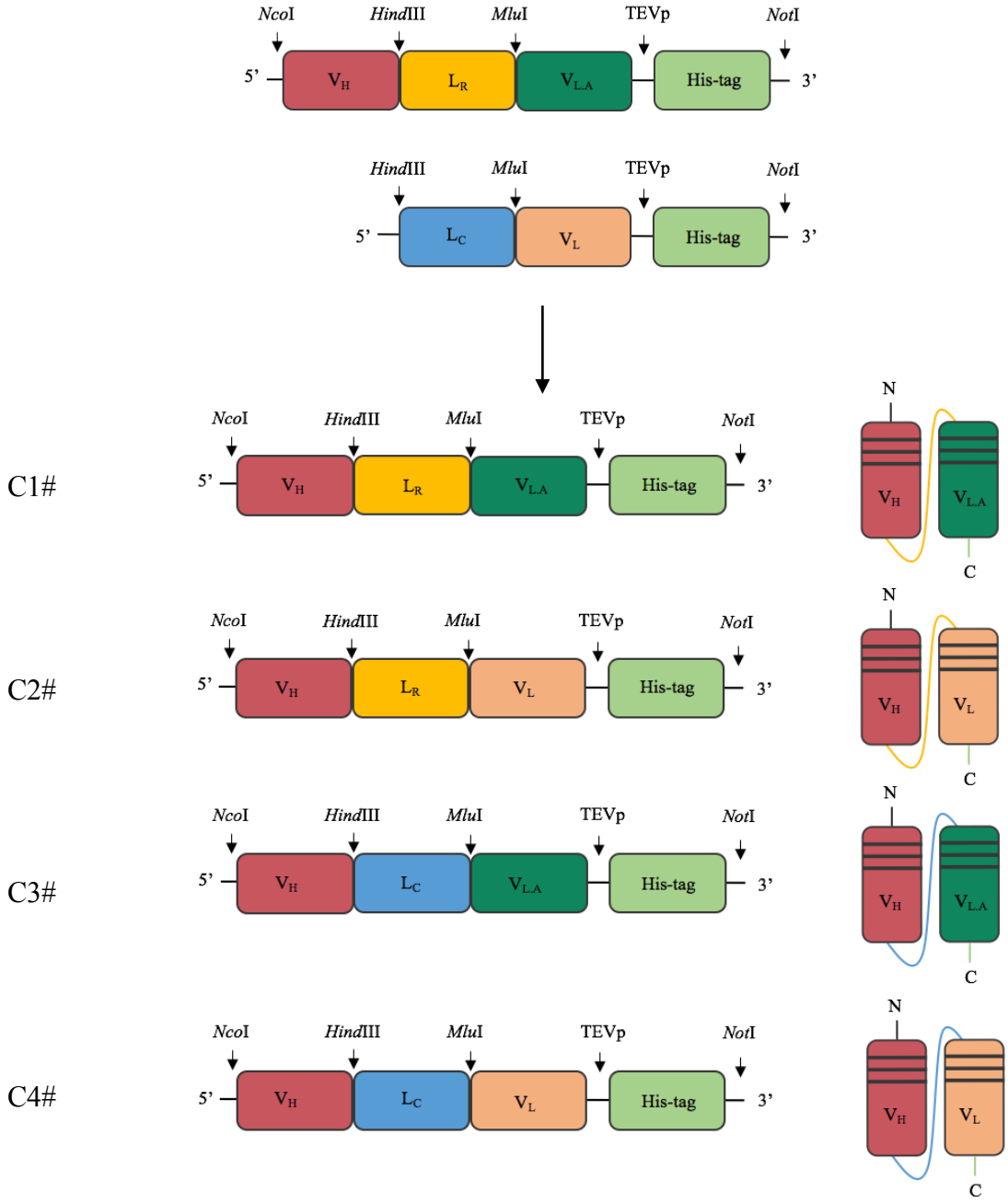


Figure 4. 1: An overview sketch of the two fragments C1# (V_H-L_R-$V_{L.A}$) and C4# (L_C- V_L) with integrated TEVp cleavage site and C-terminus His-tag, used for subcloning and cloning of the four constructs. To the right a sketch of the two domains of each scFv. $V_{L.A}$ in green and V_L in vague orange, the L_R in yellow and L_C in blue, all corresponding to the colours used in the overview sketch to the left. Figure inspired by [75].

4.2.1 Subcloning and cloning of new constructs

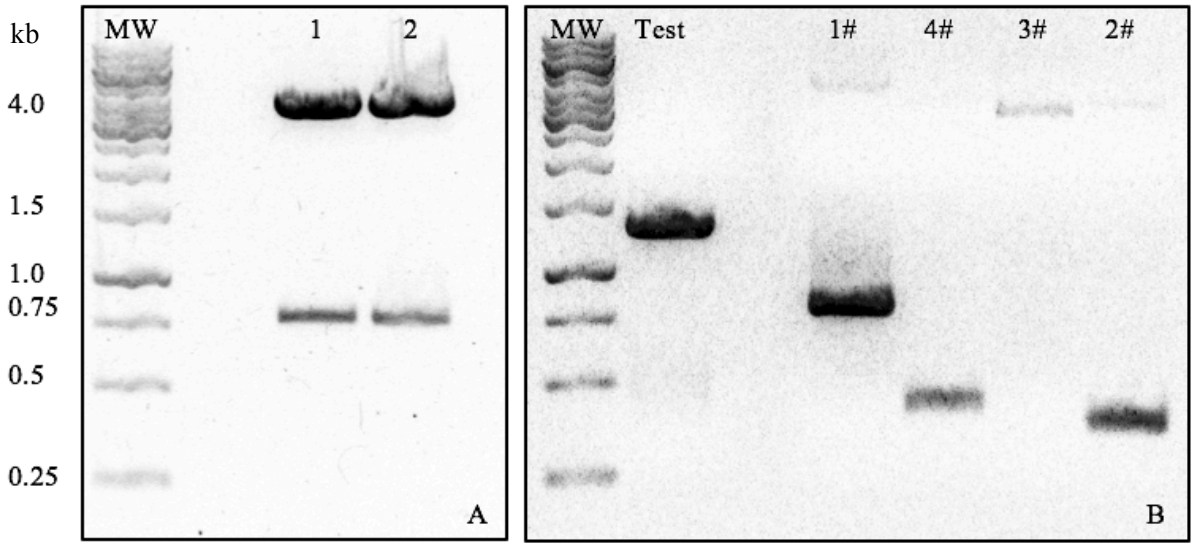


Figure 4. 2: Agarose gel 1 % showing; A: Successful double cleavage of pFKPEN-C1* with restriction enzymes *Not*I and *Nco*I to remove the C1* fragment. B: Results from the first PCR run of the four fragments; V_H-L_R-$V_{L.A}$ (1#)(826 bp), L_C-V_L (4#)(426 bp), L_C (3#)(63 bp), V_L (2#)(363 bp). The annotations on the gel represent which constructs the fragments will create when ligated into the vector pFKPEN. The fragment of L_C was not produced in this run.

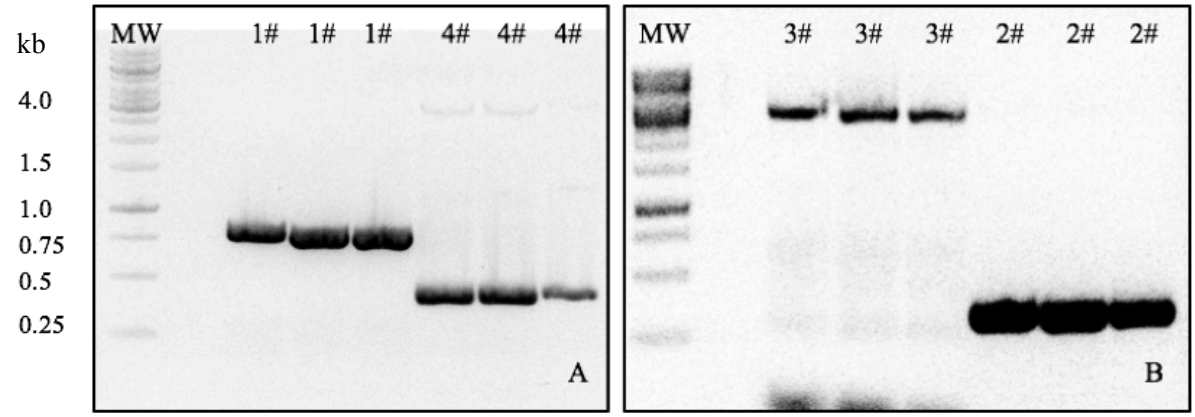


Figure 4. 3: Agarose gel 1 % showing; A: Results from PCR amplification of fragments V_H-L_R-$V_{L.A}$ (1#) and L_C-V_L (4#) with three PCR fragments each. B: Results from PCR amplification of fragments L_C (3#) and V_L (2#) with three PCR fragments each. The annotations represent which constructs the fragments will create when ligated into the vector pFKPEN. The L_C (3#) fragment was not produced in this run.

The amplification of fragment L_C showed to be unsuccessful in the first run as it did not show on the 1 % agarose gel. Because the L_C fragment was of such a small size, the possibility of the fragment to have run through the gel was there, but because of the relative large band appearing just below 4.0 kb in the same well, this was not thought to be the main issue. When a scale-up of the PCR amplification reactions were done to prepare the construct fragments for cloning experiments, the 1 % agarose gel with applied samples of the L_C fragment (3# in Figure 4. 3.B) was run for approximately 20 minutes less than the gel in Figure 4. 3.A to prevent the fragment from running out of the gel. This time a blurred fragment appeared in the bottom of the gel, but the same large fragment at approximately 4.0 kb was still there, which indicated that the production of C3# needed further optimisation. The

amplification of a new fragment, V_H-L_C, is shown in Figure 4. 4. The cloned vectors were transformated into *E. coli* XL1-Blue cells (Stratagene) and both test cleavage reactions results and (Figure 4. 5) sequencing results (Figure 4. 6) confirmed the creation of new constructs with a C-terminus His-tag.

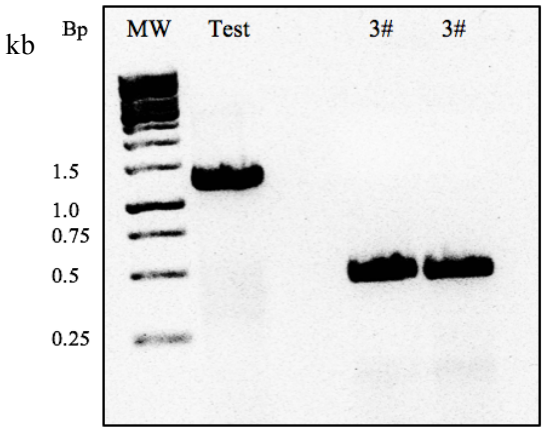


Figure 4. 4: Agarose gel 1 % showing results from PCR amplification of fragment V_H-L_C (3#) for creation of Construct 3#.

4.2.2 Conformational data

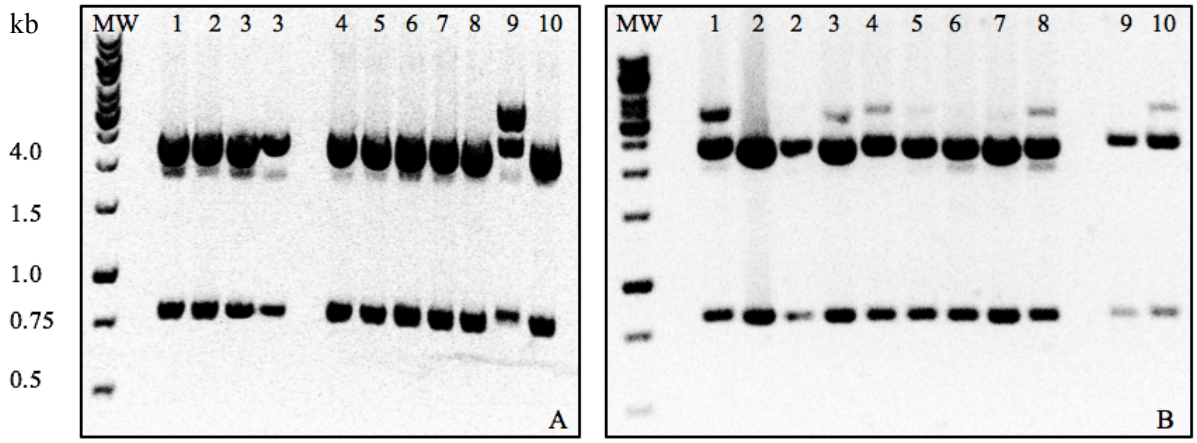


Figure 4. 5: Agarose gel 1% showing; A: Results from the test cleavage of 10 isolated colonies of pFKPEN-C1#. B: Results from the test cleavages of 8 isolated colonies of pFKPEN-C3# (1-8), and one of each of pFKPEN-C2# (9) and pFKPEN-C4# (10).

Isolated DNA from colonies number 2, 5, 6 and 7 for C3# (Figure 4. 5.B) were sent to sequencing at GATC and confirmed. The ligation and transformation reactions for C2# and C4# gave few colonies and were all sent to sequencing to be confirmed. An isolated DNA sample of each of C2# and C4# were cleaved with the NotI and NcoI and run on the same 1 % agarose gel as the C3# samples (Figure 4. 5.B). The sequences of the four constructs are shown in Figure 4. 6. The sequencing coverage is marked in grey for the reverse primer and with double underline for the forward primer. The V_H and V_L/$V_{L.A}$ are in bold, with the connecting linker L_R/L_C in normal. The TEV cleavage site is shown in red.

>C1#

MAQVQLQQSGNELAKPGASMKMSCRASGYSFTSYWIHWLKQRPDQGLEWIGYIDPATAYTESNQKFKDKAILTADRSSNTAF
MYLNSLTSEDSAVYYCARESPRLRRGIYYYAMDYWGQGTSVIVSSKLSGSASAPKLEEGEFSEARVDIQMTQTPSSLSASLG
DRVTISCRASQDISNYLNWYQQKPDGTVKLLIYYTSRLHSGVPSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPPTF
GAGTKLELLYFQSHHHHHH-AAA

>C2#

MAQVQLQQSGNELAKPGASMKMSCRASGYSFTSYWIHWLKQRPDQGLEWIGYIDPATAYTESNQKFKDKAILTADRSSNTAF
MYLNSLTSEDSAVYYCARESPRLRRGIYYYAMDYWGQGTSVIVSSKLSGSASAPKLEEGEFSEARVDLVLTQSPATLSVTPG
DSVSFSCRASQSISNNLHWYQQRTHESPRLLIKYASQSISGIPSRFSGSGSGTDFTLSISSVETEDFGMYFCQQSNRWPLTF
GAGTKLELLYFQSHHHHHH-AAA

>C3#

MAQVQLQQSGNELAKPGASMKMSCRASGYSFTSYWIHWLKQRPDQGLEWIGYIDPATAYTESNQKFKDKAILTADRSSNTAF
MYLNSLTSEDSAVYYCARESPRLRRGIYYYAMDYWGQGTSVIVSSKLAPQAKSSGSGSESKVDARVDIQMTQTPSSLSASLG
DRVTISCRASQDISNYLNWYQQKPDGTVKLLIYYTSRLHSGVPSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPPTF
GAGTKLELLYFQSHHHHHH-AAA

>C4#

MAQVQLQQSGNELAKPGASMKMSCRASGYSFTSYWIHWLKQRPDQGLEWIGYIDPATAYTESNQKFKDKAILTADRSSNTAF
MYLNSLTSEDSAVYYCARESPRLRRGIYYYAMDYWGQGTSVIVSSKLAPQAKSSGSGSESKVDARVDLVLTQSPATLSVTPG
DSVSFSCRASQSISNNLHWYQQRTHESPRLLIKYASQSISGIPSRFSGSGSGTDFTLSISSVETEDFGMYFCQQSNRWPLTF
GAGTKLELLYFQSHHHHHH-AAA

Figure 4. 6: Protein sequence coverage of the four constructs. Bold text are the V_H and $V_L/V_{L.A}$, connected by the linker L_R/L_C. Grey highlight represents the sequence coverage of the reverse primer (pHog) aligned in the $V_{L.A}$, but not in the V_L. Double underline represents the sequencing primer (Seq_Vh) which aligned in the V_H sequence of all four constructs confirming the integration of TEVp restriction site (red) and His-tag in the C-terminus.

Table 4. 1 shows calculations made of the four new constructs with their protein sequences showed in Figure 4. 6. The calculations are done in ProtParam ExPasy. C1# shows an instability index below 40, while C2#, C3# and C4# show an index above. Calculations of the protein sequences with cleaved off SHHHHHH (His-tag) are shown in the right lane for each construct. C3# have an instability index below 40 for the cleaved sequence.

Table 4. 1: Information on the new scFv constructs with a C-terminal His-tag calculated in ExPasy based on their amino acid sequences [81]. The calculations of the sequences without a His-tag are shown in the right lane for each construct.

	C1#		**C2#**		**C3#**		**C4#**	
Amino acids	265	258	265	258	265	258	265	258
Molecular weight (**kDa**)	29.7	28.7	29.7	28.7	29.6	28.7	29.6	28.7
Theoretical pI	6.70	6.15	7.12	6.83	8.24	8.22	8.58	8.57
Extinction coefficient **$M^{-1}cm^{-1}$**	54570	54570	55600	55600	54570	54570	55600	55600
Absorption coefficient	1.831	1.889	1.865	1.933	1.837	1.895	1.871	1.939
Instability index (Below 40 indicates stable protein)	39.47	38.57	52.08	51.52	40.14	39.26	52.75	52.20

4.2.3 Protein expression of constructs with His-tag

With the protocol for periplasmic expression of scFv in *E. coli* XL1-Blue cells, a small-scale expression experiment was done of the four new constructs, shown in Figure 4. 7. The cells were harvested after growing overnight (ON) without glucose and spun down before preparing both supernatant and pellet for SDS-PAGE. At approximately 30 kDa, a small band was detected for all four constructs, both in the supernatant and pellet.

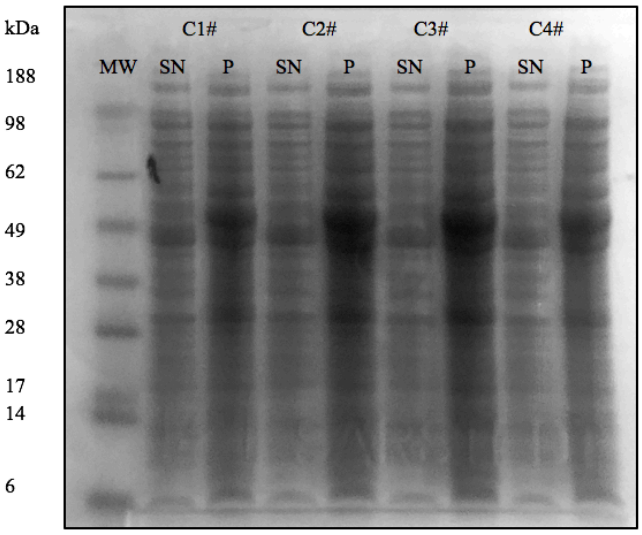


Figure 4. 7: Two samples (Supernatant (SN) and pellet (P)) of each construct from a small-size expression experiment showed only a small band of just above 28 kDa, not a thick band as expected.

4.3 Protein expression of C1*

Large scale protein expression of C1* was done in *E. coli* XL1-Blue cells according to the protocol. The cells were induced for periplasmic expression of scFv at OD_{600} = 0.6-0.8 by removing glucose in the medium. The cells were set to grow at 30 °C ON, and the OD_{600} was measured to approximately 2.5 in the morning. One sample prior to induction and after induction were collected and dissolved in SDS loading buffer, DTT and MQ. After periplasmic lysis of the cells, following the protocol, one sample of the sucrose fraction and one sample from the periplasmic fraction was collected and prepared for SDS-PAGE together with the dissolved fractions from the expression. All samples were run on the SDS-PAGE presented in Figure 4. 8 together with the SeeBlue® Plus2 Pre-Stained Standard for molecular weights. The distinct band at 28 kDa in the periplasmic fraction indicates the scFv C1*.

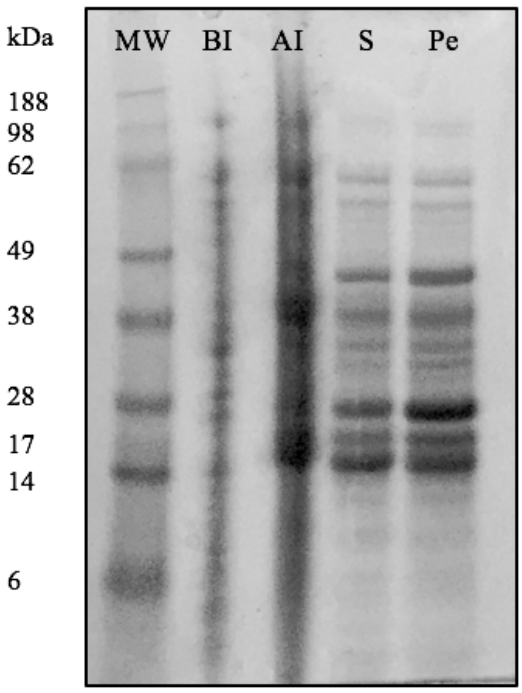


Figure 4. 8: SDS-PAGE with samples from protein expression. From the left: MW; SeeBlue® Plus2 Pre-Stained Standard, BI; Cell pellet from cell culture with glucose, AI; Cell pellet from cell culture without glucose, S; sucrose fraction, Pe; Fraction from periplasmic lysis.

4.4 Protein Purification

4.4.1 Protein L chromatography

The periplasmic fraction of approximately 80 mL were collected and filtered through a 0.2 µm syringe filter to remove insoluble fragments that could clog the column. The fraction was calibrated for the column by the addition of 10x Binding buffer. The column was calibrated with 1x Binding buffer before the periplasmic fraction was applied to the column with a flow of 1 mL/minute. The periplasmic fraction was at all times held on ice to keep the temperature low. When the periplasmic fraction was applied, the column was washed with 1x Binding buffer to remove unbound protein. When the absorbance reached the baseline, the elution buffer was applied to the column and the protein was eluted into 1 mL fractions. The chromatogram in Figure 4. 9 shows a distinct peak of eluted protein annotated with the fractions A1 to A12. A zoom-in of this peak is shown in Figure 4. 10.A and the SDS-PAGE with applied samples of the fractions A1-A12 are shown in Figure 4. 10.B together with a sample from the collected flow through (FT). The band at 28 kDa represents the purified scFv C1*. The fractions A2-A11 were picked for further purification by Size Exclusion Chromatography (SEC). The flow-through showed a distinct band at 28 kDa and was therefore collected, concentrated and run once more on the protein L column to collect as much protein as possible.

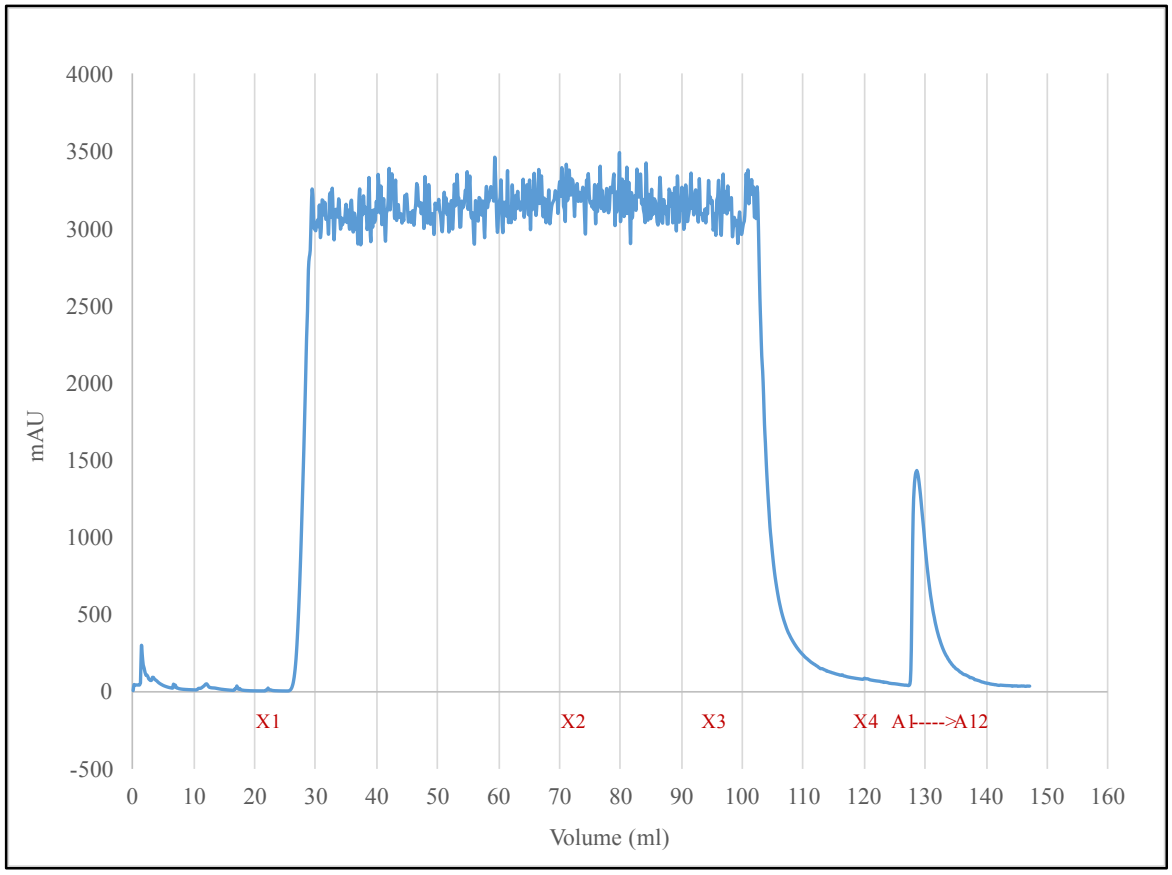


Figure 4. 9.: Chromatogram from the purification by protein L affinity column. Fractions (red) X1 to X3 represent the flow through of applied sample and X4 the washing of column with binding buffer. The sharp peak with fractions A1-A12 represent the eluted 14F7 scFv after changing the pH of the column to 2.5 by applying elution buffer.

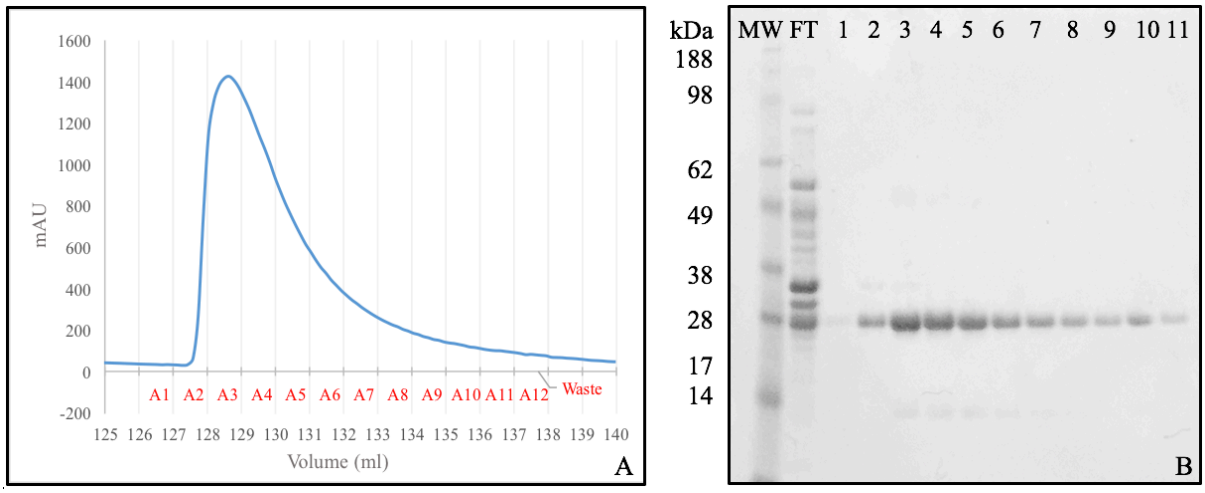


Figure 4. 10: A: The zoomed in curve of the peak from Figure 4. 9 representing the eluted protein 14F7 scFv. B: The SDS-PAGE run with fractions from the purification, showing a protein band at approximately 28 kDa. The fractions 1-11 represent the fractions A1-A11 in Figure 4.9.A.

4.4.2 Size Exclusion chromatography

Calculation of elution volume for the scFv

200 µL applied mixture of proteins from the Molecular Weights Kit for Gel Filtration (Sigma) was run with PBS on both Superdex75 10/300 GL (GE Healthcare) and Superdex200 10/300 GL (GE Healthcare) for creation of a standard curve for the two columns (Appendix E: Standard curves). The Superdex75 column is limited to proteins with a molecular weight below 75 kDa, hence two reference points with sizes of 200 kDa and 150 kDa were excluded. All reference points from the Superdex200 run were used for the standard curve. The standard curves obtained calculated the elution volume for the scFv to be approximately 16.7 mL for the Superdex200 and 11.5 mL for the Superdex75.

Purification of the 14F7 scFv

The fractions from the protein L purification run were collected and concentrated to approximately 500 µL before applied to the Superdex200 calibrated with PBS. The chromatogram in Figure 4. 11.A show two distinct peaks at approximately 14.8 mL and 16.7 mL. The first peak represents a dimeric form of the scFv according to the standard curve obtained for Superdex200. The second peak at 16.7 mL represent the monomeric form of scFv C1*. Fractions A7-A10 were collected and a sample from each fraction run on a SDS-PAGE shown in Figure 4. 11.B. The distinct band at 28 kDa represents the scFv C1*, while the bands at approximately 52 kDa and 14 kDa represent unknown proteins.

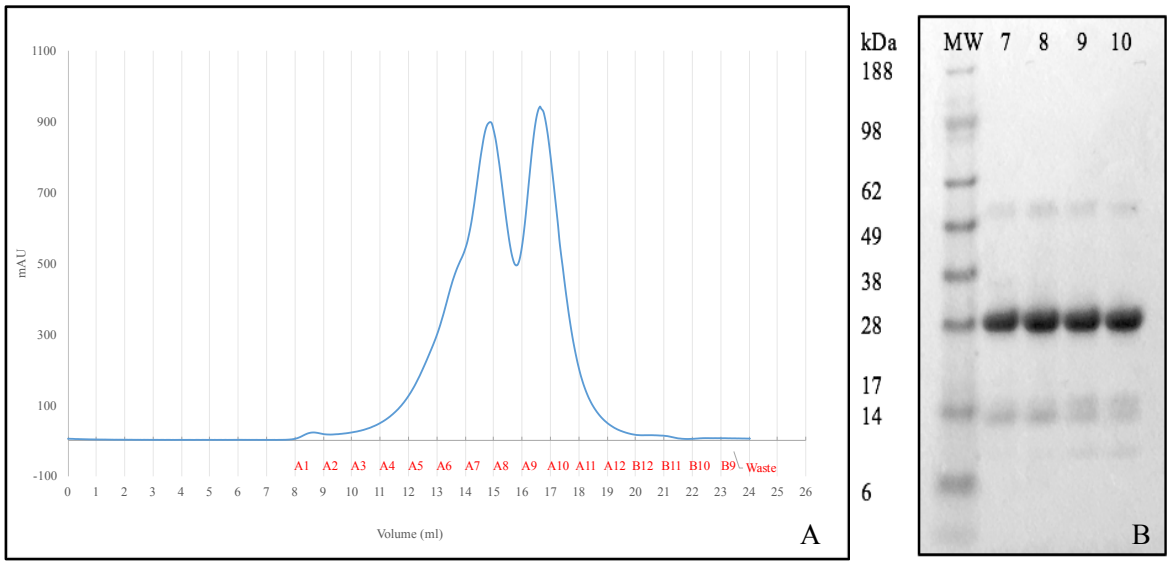


Figure 4. 11. A: Size exclusion chromatogram of the 14F7 scFv C1* run on the Superdex200 (GE Healthcare). The peak at approximately 14.8 mL represent a dimer of the scFv, while the peak at 16.7 mL represent the monomeric form. B: SDS-PAGE of the four fractions A7, A8, A9 and A10 from the chromatogram in A. The fractions A7 and A8 represent the dimeric form of the scFv, and the fractions A9 and A10 represent the monomeric form. The samples were run without DTT. All fractions show a large band at 28 kDa, a small band at approximately 52 kDa, and a small band at approximately 14 kDa.

Dimeric form of the 14F7 scFv

Fractions A7 and A8 from the first peak in Figure 4. 12 were collected and run a second time in the ÄKTA purifier with a higher salt concentration of the PBS buffer (300 mM) to see if the dimerization of scFv could easily be reversed. The strongest peak in Figure 4. 12 still represents the dimer form of the scFv, with a tiny signal for the monomeric scFv. Because of overlap of the two peaks seen in Figure 4. 12, it is possible that the signal from the monomeric scFv is from fraction A8.

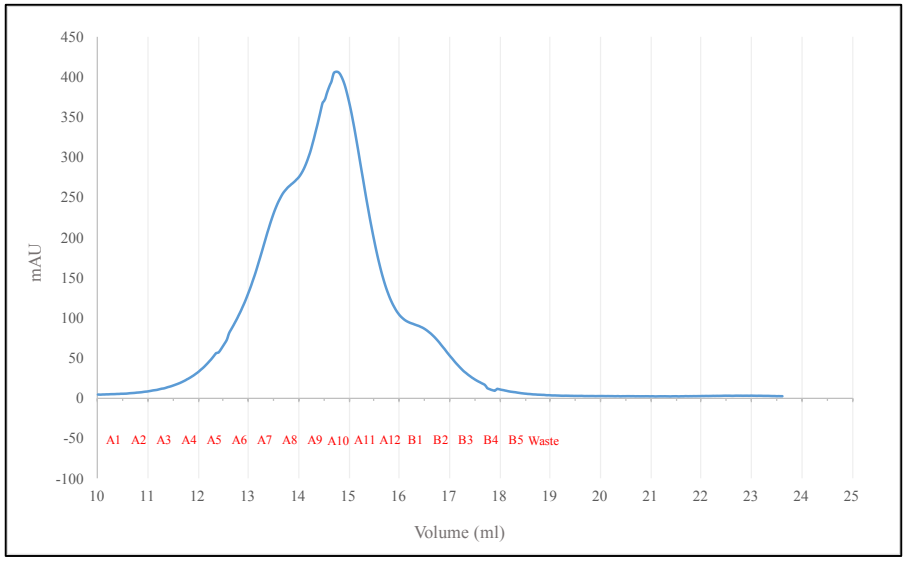


Figure 4. 12: Chromatogram of the size exclusion run of the collected fractions A7 and A8 from the purification shown in Figure 4. 11. The peak at 13.5 mL shows possible aggregation of the scFv, while there is a tiny signal at 16.5 mL representing the monomeric scFv. The tallest peak at 14.8 mL still represent the dimeric peak.

4.5 ELISA

The wells in the MicroWell PolySorp solid plate (Sigma) were incubated with 100 µL NeuGc GM3 and NeuAc GM3 diluted in methanol (Prolabo) to the concentration of 10 µg/mL. The plates were left to evaporate ON. Unspecific binding was blocked with PBS-B (1 x PBS and 2 % BSA), and the plates were washed with PBS-T (1 x PBS, 0. 1% Tween-20) three times between each step. The absorbance was normalised and plotted graphically in Microsoft Excel. The trend line was set to mimic the saturation curve with polynomic regression, as the option for a best fit curve was not available in this program. The results presented in Figure 4. 13 represent a trend in all executed ELISA experiments: The 14F7 scFv C1* discriminates between the NeuGc GM3 and NeuAc GM3, in favour of the NeuGc GM3.

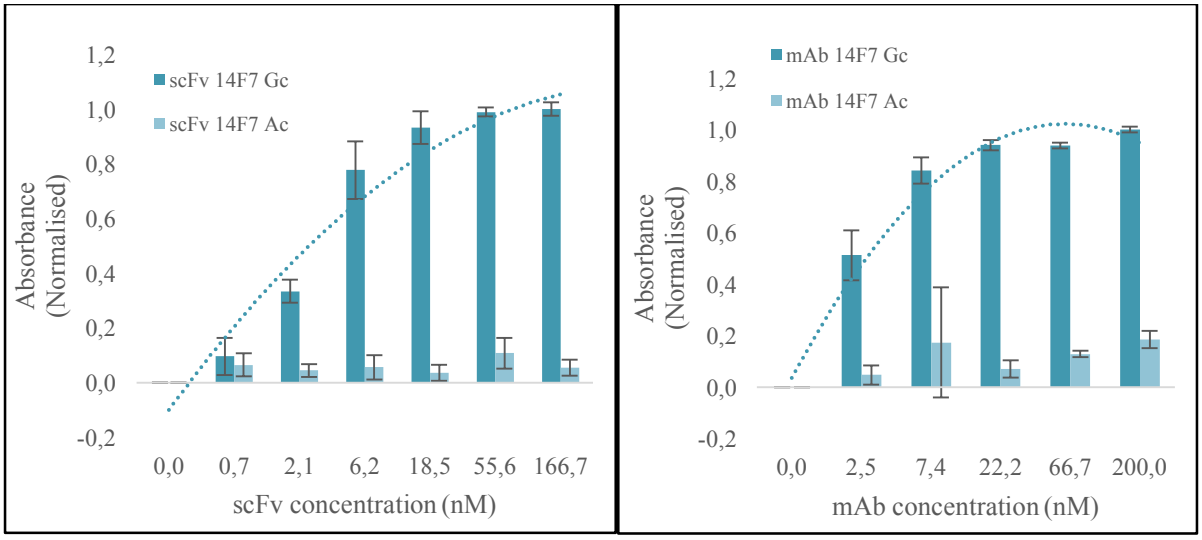


Figure 4. 13: To the left: The 14F7 scFv show favourable binding affinity to the NeuGc GM3 ganglioside (dark petrol) against the NeuAc GM3 ganglioside (light petrol). The figure represents the trend from all executed ELISA experiments with the scFv and mAb as a positive control. To the right: The 14F7 mAb used as a positive control for the scFv. The favourable binding affinity to the NeuGc GM3 is also shown here in dark grey. The figure represents the trend from all executed ELISA experiments with the scFv and mAb as a positive control.

4.6 Fluorescence detection

The 14F7 scFv C1* solution was run in the FP-8200 Spectrofluorometer (Jasco) in PBS, or with PBS and the trisaccharide part of the NeuGc GM3. This gave two melting curves that are shown in Figure 4. 14. The melting curve was obtained by plotting the change in fraction emission (F350/F330) against the rising temperature of the solution. The results indicate that the melting temperature for the 14F7 scFv C1* is 72 °C, both with and without the ligand. This result implies that the protein is stable.

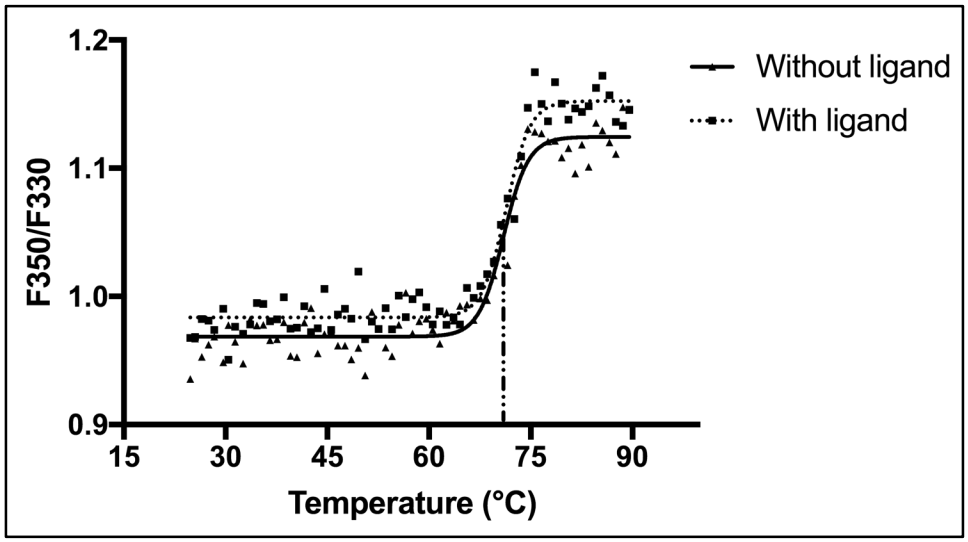


Figure 4. 14: The melting curve obtained by fluorescence detection of the tryptophan residues in the 14F7 scFv C1* structure. The melting temperature is the same both with and without ligand.

4.7 Crystallisation

Several crystallisation screens were sat up using both sitting drop and hanging drop vapour diffusion. A sitting drop robot tray with original structure screen solution gave crystals (Appendix I: Crystal screens). Crystals with nice, square shapes were thought to be good crystals, and were sent to be analysed at the European Synchrotron and Radiation Facility (ESRF) in Grenoble, France for diffraction experiments, but no good diffraction data was collected (Appendix I: Crystal screens).

Small crystals obtained from the crystallization condition G4 from Morpheus screen (Molecular Dimensions) (Figure 4. 15) diffracted at the synchrotron and gave several diffraction patterns used for solving the 14F7 scFv C1* structure. The crystal shape, with two flat surfaces pointing out from the centre, lead to many overlapping diffraction patterns.

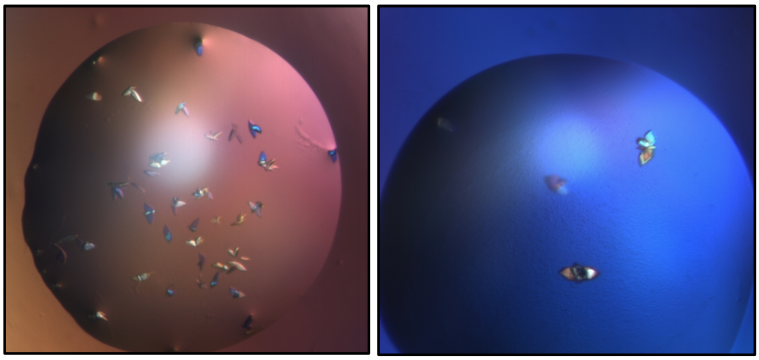


Figure 4. 15: Obtained crystals of 14F7 scFv C1*. Crystals were picked and sent to the ESRF for diffraction studies.

4.8 The structure of 14F7 scFv C1*

The structure of 14F7 scFv C1* was solved by Kaare Bjerregaard-Andersen[4] and is presented in Figure 4. 16.

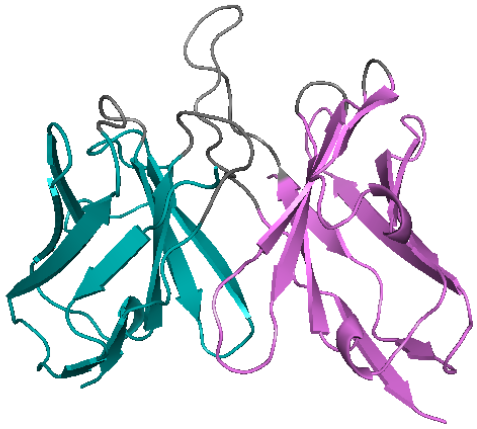


Figure 4. 16: The structure of 14F7 scFv C1* obtained by X-ray crystallography and visualised in PyMOL (Schrödinger). The V_H is presented in violet, and the $V_{L.A}$ in teal green, with the prominent CDR H3 loop in the top middle. The structure was resolved by Kaare Bjerregaard-Andersen[4].

5 Discussion

When Krengel *et al.* [25] solved the structure of the 14F7 Fab thirteen years ago, many of the challenges from then on have been to reproduce the obtained results to get further insight in the interaction between 14F7 and NeuGc GM3. The difficulty in reproducing the Fab crystal evolved to the development of scFv constructs with the same binding interaction to NeuGc GM3 as the Fab. Despite that scFvs are indicated to crystallise easier than Fabs because of their small size and low flexibility [45], the development of diffracting crystals has been unsuccessful until now. However, this is possibly due to the instability of the 14F7 scFv, and because of this, the focus in this thesis has been to optimise every step in the process that have shown to be of importance in the development of pure, stable protein for crystallisation.

5.1 Construct design

For the purpose of easier purification, the four scFv constructs produced by Johannesen [75], were further developed to contain a C-terminus His-tag (Figure 4.1 results). An N-terminus His-tag incorporated in the 14F7 scFv, together with a TEV cleavage site, indicated a disruption of binding to the NeuGc GM3 [75]. This confirmed the demonstrated importance of the V_H region in binding to NeuGc GM3 [25, 57] as one additional amino acid residue in the V_H region, after the TEV cleavage of the His-tag, disturbed the binding. It was therefore suggested to incorporate a tag in the C-terminus.

The introduction of a tag can be beneficial if a protein is difficult to purify [82]. However, an introduced tag alters the protein sequence which can have consequences for the original function of the protein such as binding affinity and solubility. In addition, tags are often flexible, which for the purpose of crystallisation is not beneficial. A TEV protease cleaving site was therefore included at the C-terminus of the variant light chain sequence, just before the tag, to easily cleave off the His-tag before crystallisation. The amino acid sequence recognised by pTEV is preferred to be ENLYFQ\S, where the cleavage site is annotated '\'. In the case of a C-terminus introduced His-tag with a pTEV recognition site, the first six amino acid residues of the restriction site will still be in the sequence after cleavage of the tag. However, reports show that the N can be switched to any residue of choice [83]. To avoid the introduction of many amino acid residues in the C-terminal light chain, we decided to mutate the last residue of the light chain sequence from a K (Lysine) to a L (Leucine), giving the three last residues of the light chain sequence compatible for the pTEV recognition sequence ELLYFQS. After TEV protease cleavage, the overlap will result in only three additional residues in the sequence, giving a less flexible C-terminus than with six amino acid residues. The variation in variable light chain residues have been reported to not have an impact on the binding affinity to the NeuGc GM3 [57], when in fact it is the V_H shown to be of great importance. This argues the decision

to mutate one residue in the C-terminus light chain to avoid any unnecessary additions of residues upon introducing a tag, when the main purpose is to obtain crystals for structural characterisation.

5.2 Optimising the protein expression protocol

The expression protocol used by Johannesen in 2014 [75] gave a rather low cell density after ON culturing without glucose, with OD_{600} of approximately 1-1.5. This is somewhat unexplainable because the growing conditions with 2 x YT medium, 0.1 M glucose and 100 mg/L ampicillin were used for 37 °C until OD_{600} reached 0.6-0.8, and glucose was removed for growing at 30 °C. The same approach is used today. The difference from then to now is that working with the periplasmic lysis is done in a cold room or on ice, instead of working at room temperature. This will not have an effect on the cell density as the protein expression comes first.

5.2.1 Growth conditions

Because the production of 14F7 scFv is controlled by the *lac* promoter, the cells were grown with glucose present until log phase was reached at OD_{600} 0.6-0.8, to maximise the possibility of high protein yield. Induction at OD_{600} higher than 0.8 will likely limit the protein expression to a short period before reaching the stationary phase of growth, resulting in reduced final OD600. Induction at OD600 lower than 0.6 will also likely give a reduced final OD600 because the stationary phase is reached at an earlier stage because of the logarithmic growth behaviour. Removing glucose activates the *lac* promoter, resulting in induced scFv production. Following the protocol optimised by Noha Abdel-Rahman (Manuscript in preparation), the induction culture was set to grow at 20 °C overnight with a resulting OD_{600} of approximately double the pre-induction measurement, 1.2-1.6. From experience, this was considered as poor cell density compared to other protein expression experiments (Julie Heggelund; personal communication). In addition, dead cells were present in a higher amount than expected. Talking to Geir Åge Løset, we were made aware of the fact that *E. coli* XL1-Blue cells do not grow below 22 °C, and that the preferred temperature for periplasmic expression in these cells is 30 °C [84]. The XL1-Blue strain is an unusual expression strain, but working with scFvs, this strain has shown to give the best yield [84]. After correcting the temperature, the OD_{600} exceeded to approximately 2.5, which again indicated high cell density. The protein yield could also be detected on a SDS-PAGE with a more distinct, thick band at 28 kDa (Figure 4.8) than for previous experiments.

5.2.2 Protein aggregation and degradation

After extracting the scFv from the periplasm, the fractions were either left in 4 °C ON or frozen at -20 °C over the weekend before further purification steps were utilised. This was done to limit the loss of

protein, as it shows rapid degradation in temperatures above 4 °C. However, a pattern was detected for the protein samples frozen over the weekend. The protein had high affinity to the protein-L-coated column which gave a wide, single peak when eluting with 0.1 M glycine with pH 2.5, but when purifying the samples with size exclusion chromatography, the result was rather miserable. The protein showed up in the chromatogram at a volume corresponding to a much larger size than the scFv of 28 kDa, indicating that the protein aggregated upon freezing, giving insoluble scFv.

The protein samples left at 4 °C overnight gave better results, but the yield was still quite low, giving a protein concentration in the down-limit for good crystallisation experiments. The temperature in the ÄKTA cabinet is standard 10 °C. We thought, in principle, this would not affect the stability in any way because the purification procedure was quite straight forward and the elution sample put on ice straight after. However, when purifying larger-scale purification batches, applying the protein samples to the column took 2-3 hours. In that time, the ice-cold sample is warmed up, which can then affect the protein stability. When applying big batches of protein sample to the column, the flask with protein sample was put on ice inside the cabinet to maintain the low temperature as long as possible. The continuous maintenance of low sample temperature resulted in a higher protein concentration of 5 mg/mL in PBS, compared to results down to 1-2 mg/mL. The same volume of periplasmic fraction was applied to the column.

5.2.3 Protease inhibitors

When expressing proteins in the periplasm of *E. coli* cells, the production of proteases can be troublesome when lysing the cells. By including protease-inhibitors in the lysis solution, the protein will be protected from degradation. The different protease inhibitors used in this project resulted in different outcomes. The c*O*mplete protease inhibitor (Roche) without EDTA is the most commonly used inhibitor in this project and is a cocktail of protease inhibitors for a broad spectre of serine and cysteine proteases. C*O*mplete protease inhibitor is soluble in aqueous solutions such as 50 mL 20 mM $MgCl_2$. This showed to be of great importance as we observed that with the use of PMSF (Phenylmethylsulfonyl fluoride), the protein was more prone to aggregation and degradation. PMSF (Sigma) is a serine protease inhibitor, and is only somewhat soluble in water and must be dissolved in isopropanol or 100 % ethanol. A low concentration of isopropanol introduced in the lysis solution is believed to in principal not affect the protein. However, the possibility of interference with the protein is present. There was no sign of degradation products in the gel, but the SEC demonstrated aggregated protein by detection of early elution volume when PMSF was used. The possible interference of isopropanol in the protein solution can be tested when expressing protein by dividing the cell culture in two, and use PMSF as a protease inhibitor in one half, and the c*O*mplete protease inhibitor in the other half.

It is a possibility that the storage of the periplasmic fraction overnight at 4 °C, which seemed to have the best outcome compared to storage at -20 °C, has an effect on the protein quality due to the presence of proteases in the solution, despite the supplemented protease inhibitors. One possibility to avoid protein degradation is to filter it directly after collecting the "periplasmic fraction" with a syringe filter, and then purify the fraction with Protein L chromatography to remove possible proteases. Because proteases in *E. coli* cells are known to have molecular weights of approximately 14 (lysozyme) – 60 kDa [85, 86], the syringe filter (Millipore) used to filter the fraction before applying to the Protein L column will not remove proteases from the solution, only clear the solution from insoluble products that will clog the column. To avoid protein degradation, it should thus be recommended to complete the first purification step by protein L chromatography the same day as periplasmic lysis of the cells.

5.2.4 Optimised protein L affinity

For 14F7 scFv, it was argued that creating a C-terminal His-tag would ease the purification procedure because the protein L affinity chromatography did not give satisfactory results. However, the affinity between a Kappa light chain and protein L is proven to be very good [78], and the probability that the resin in the gravity column was the problem, and not the protein L affinity per se, is quite high. This was confirmed by changing to a protein L column from another manufacturer (Thermo Fisher Scientific), which demonstrated strong affinity by a single slope of eluting protein. One advantage with the use of Protein L column is that all types of protease inhibitors can be used. For a His-tag column, the protease inhibitors cannot contain EDTA, because EDTA will interact with Ni^{2+} and release it from the column material, limiting the amount of protein binding. Because EDTA is used in periplasmic expression together with sucrose to make the membrane permeable, the exclusion of EDTA can have a negative effect on protein yield. In addition, EDTA is a good metalloprotease inhibitor because it outcompetes metalloproteases in their binding to divalent cations.

5.2.5 ScFv dimerization

The dimerization of 14F7 scFv demonstrated with SEC proved to be a repeating problem when purifying the scFv. This dimerization was not detected during protein L affinity chromatography, which indicates that the protein still retains protein L binding. The reason for dimerization was thought to be because of the salt concentration in the PBS buffer, and that this would disappear by increasing the concentration from 150 mM to 300 mM. As the results indicated, this did not help to reverse the dimerization. However, the possibility that the initial purification run with a PBS-buffer with 300 mM NaCl would decrease the amount of dimerized scFv is there and should definitely be tested. This could also have an effect if the dimerization is, as indicated, irreversible, to increase the yield of soluble, monomeric scFv.

In the SDS-PAGE run after both protein L and SEC, there are two additional bands at approximately 14 kDa and 52 kDa. These bands can represent misfolded protein which only appear in SDS-PAGE or as an additional peak in SEC. As these bands are not discovered in the chromatogram of protein L, it is likely that these misfolded scFv proteins still retain protein L binding. Because it has been well documented that the chaperone FkpA assists in correct folding of scFv in the periplasm [67, 68], the misfolding of protein can be a consequence of sporadic mutations in the *fkpA* gene, leading to improper functions of FkpA, or that its native promoter does not function properly in expressing the chaperone. The low level or lack of FkpA in the periplasm will affect the folding of protein and can be a cause of misfolded protein. There is also a possibility that the misfolding is a consequence of exposure to the harsh environment outside the cell after periplasmic lysis.

The additional peak in SEC is according to the standard curve corresponding to a protein size of 52 kDa, as demonstrated in the SDS-PAGE. This is representable for the size of a dimerized scFv. As this peak was never tested in ELISA, it is not known whether this dimerization causes disruption of binding to the NeuGc GM3, or if it retains the binding. In addition, if the binding is detectable by pL-HRP interacting with the kappa light chain of the scFv.

5.3 ELISA

5.3.1 Quality of the scFv

After completing several ELISA experiments the results indicated that the fresh state of protein was directly correlated to the quality of good binding affinity. Even though the scFv showed good binding affinity to the NeuGc GM3, and distinct discrimination against the NeuAc GM3, the binding somehow fluctuated between identical experiments. The same batch of NeuGc/Ac GM3 was each time diluted in methanol and left to evaporate overnight, however, the protein sample was either obtained from 1-4 days storage at 4 °C, or from storage at -80 °C. Protein used for ELISA the first day after the final purification step, or directly thawed from the -80 °C freezer, gave the best results. If the experiment was repeated the next day, and the protein was left in the fridge, the results implied a decrease in protein quality. Freezing protein samples with liquid nitrogen ensures that the protein does not precipitate out of solution. A consequence of slow freezing of protein samples is the formation of water crystals which will concentrate the protein in a smaller space. In PBS, the different components will not freeze at the same time, giving a fluctuation in pH, which can be harmful for the protein. The high concentration of protein and unstable pH can thus form insoluble aggregates [87], which then can have an effect on binding in ELISA.

5.3.2 Quality of the NeuGc GM3

It has been indicated in previous ELISA experiments (Heggelund - unpublished) that the quality of the NeuGc GM3 has had an effect on the obtained results. One batch of the NeuGc GM3 previously received from Cuba was difficult to dissolve in methanol to the concentration of 5 mg/mL, and needed further dilution to 0.5 mg/mL. This batch gave fluctuated results detected by uneven detection spots in each identical triplicate, indicating uneven immobilisation of the ganglioside. This was thought to be because of the formation of unilamellar cells of GM3 as a consequence of humid exposure [88], which then again indicated bad quality of the NeuGc GM3. For the ELISA experiments presented in Figure 4.14 and 4.15, a new batch with lyophilised NeuGc GM3 ganglioside was used. This batch was easy to dissolve in methanol, which indicated good quality. As ELISA is a quantitative method, the results are difficult to interpret in the direction of which part is responsible for the obtained results. Because the immobilised NeuGc/Ac GM3 used for ELISA was taken from the same batch each time, the probability of the ganglioside being responsible for the fluctuation in results is thought to be low.

5.3.3 ScFv affinity to NeuGc GM3

The ELISA results demonstrated a distinct discrimination between the NeuGc GM3 and NeuAc GM3 for the 14F7 scFv. The binding of scFv to NeuGc GM3 was expected to give a low detectable absorption due to previous results (Heggelund - unpublished) and the first ELISA experiment was completed with a 1:3 dilution series with 10 μM as the highest protein concentration. This gave a maximum absorption signal for almost all concentrations which made the analysation of data problematic as we wanted to see at which concentration in the dilution series saturation was reached. Diluting down into the nanomolar (nM) range we managed to obtain the binding curve presented in Figure 4. 13. The positive control used in an ELISA experiment should operate with the same binding specificities as the sample protein for the validation of binding affinity to be optimal. However, for the 14F7 scFv no other scFv have been reported with specific binding affinity to the NeuGc GM3, and the 14F7 mAb was therefore selected as a positive control. Because the 14F7 mAb contains two binding sites it also contains two identical variable light chains that will be bound by pL-HRP. This will give a higher signal per molecule than for the scFv, and it is thus difficult to validate if the scFv has the same binding affinity to NeuGc GM3 as the 14F7 mAb. However, the distinct discrimination between NeuGc GM3 and NeuAc GM3 is demonstrated for both the scFv and mAb. Based on the results it can be implied that the scFv can be of potential for use in therapeutics such as toxic-coupled treatment [43]. The small size of the scFv have great potential for plasma membrane penetration, and with the short half-life in serum the 14F7 scFv can carry toxic substances directly to the target cell, activating an immune response, and directly after be filtered out of the serum [58].

5.4 His-tag instability

The protein expression results of the new constructs indicated that something was not working the way we expected. By using the amino acid sequences of the new constructs, their characteristics were calculated in ExPasy. The calculations indicated that only C1# was a stabile protein, with an instability index of 39.47. The calculations in ExPasy would indicate that the introduction of His-tags in the C-terminus did not stabilize the constructs, but rather destabilised them. This result is the opposite of what was obtained with an N-terminus His-tag [75]. With the removal of residues SHHHHHH, mimicking the removal of the His-tag with TEV protease cleavage, the characteristics demonstrated an increase in stability, resulting in both C1# and C3# characterised as stable proteins.

Because the scFvs prove to be temperature sensitive and are prone to aggregation, the instability obtained by introducing a C-terminus His-tag would probably not help with obtaining pure, stable scFv. Because the purification by Protein L affinity chromatography became highly effective when changing the column manufacturer, the conclusion should be to continue working with the constructs without a His-tag.

5.5 Crystallisation

When crystals are obtained in screens such as the G4 Morpheus Screen (Molecular Dimensions) for the 14F7 scFv, an optimisation screen is created to further evaluate the parameters for good crystallisation of the protein. Parameters such as precipitant, additives, buffer pH, and protein concentration can be adjusted to see which condition gives the best crystal. Because the setup of crystal screens can be somewhat different between each time because of human errors, different conditions can give crystals at different times. New optimisation screens of crystallising conditions can therefore also result in no crystals forming. However, by setting up multiple new crystallisation screens from crystallising conditions, the probability of eventually finding the perfect crystallisation condition is present.

5.6 Structure of the 14F7 scFv

The main goal of the anti-tumour antibody project is to obtain important structural information about the binding interaction between the 14F7 and NeuGc GM3. The obtained crystals and diffraction data of the 14F7 scFv is therefore of great importance for future crystallisation experiments with the ligand as the crystallisation conditions sets the stage for obtaining new diffracting crystals with a ligand.

5.7 C1* stability

Working with the 14F7 scFv C1*, Johannesen [75] reported that the protein was prone to degradation and that the protein experienced cleavage, resulting in a band of 14 kDa in the SDS-PAGE. The size of this band could indicate a cleavage in the middle of the protein, somewhere close to the site between the linker and heavy chain. A band with the same size of 14 kDa was detected working with the optimised protocol, however, the band did not show up with each expression experiment, indicating that the frequency of cleavage was reduced. Due to the fact that Johannesen worked in room temperature during the periplasmic lysis and purification, the possibility is that this is the reason for protein cleavage. However, the cleavage remains to be analysed by isolating the 14 kDa fragment from the SDS-PAGE and send it to sequencing. Comparing the melting temperatures obtained by the ThermoFluor method [75] and Fluorescence spectroscopy, the temperatures of 71.5 °C and 72 °C respectively, indicate a rigid folded protein with good stability. The problem with the descending protein activity can therefore potentially exclude the stability of the protein as a cause.

5.8 Lessons learned

For stabile protein expression and purification, the results obtained in this thesis underlines the importance of low working temperature during periplasmic expression and every purification step thereafter, before flash freezing the pure protein sample with liquid nitrogen for storage at -80 °C. The activity of unknown proteases can be the cause of decreased protein activity when storing the protein at 4 °C. Flash freezing with nitrogen and storage at -80 °C will therefore exclude this factor. As the results indicated, this did not harm the protein and is preferable for to preserve good quality protein. It can be wise to divide the protein sample into samples of approximately 20 µL each, depending on the protein yield, to avoid thawing too much protein whenever a new experiment is on the agenda. As an additional step in avoiding possible protease activity, it can be worth underlining the importance of purification by protein L chromatography the same day as periplasmic lysis, as this has proven to give a higher yield of active protein.

.

6 Future perspectives

The successful structural finding of the 14F7 scFv has opened doors for future characterisation of the interaction with NeuGc GM3. The crystallisation conditions can be used to obtain new crystals, such as the 14F7 scFv in a complex with a ligand. This ligand can be the trisaccharide part of the NeuGc GM3, as the lipid part of the ganglioside is difficult to crystallise [76], or the 14F7 scFv in a complex with the scFv anti-idiotypic antibody 4G9 {Rodriguez, 2003 #87}. This to obtain a model of the interaction between 14F7 and NeuGc GM3 as the 4G9 is developed to mimic the NeuGc GM3 epitope. Because a scFv of this Ab2 has been made, the possibility of crystallisation in a complex with the 14F7 scFv is more likely to achieve than to crystallise the entire antibody as the flexible regions are difficult to crystallise. {Wörn, 2001 #86}{Ahmad, 2012 #82}

An approach to assess the specificity and affinity of the interactions between the 14F7 and the NeuGc GM3 in the plasma membrane is to incorporate the NeuGc GM3 ganglioside into proteoliposomes [89, 90] to study the native function of the NeuGc GM3 bound to the plasma membrane. For methods such as SPR this has been demonstrated to be of great potential with a developed chip that will capture the liposome with incorporated ganglioside [91]. A similar incorporation approach has been done with very small size proteoliposomes (VSSP) for the purpose of creating a cancer vaccine initiating an immune response [92-95].

For the continuation of obtaining good protein crystals of the 14F7 scFv the problem with aggregation and possible dimerization of the scFv needs to be addressed. As Johannesen [75] mentioned in her thesis, the possible change of vector from the pFKPEN with the FkpA expressed by its native promoter, to the pFKPEI where the expression of FkpA can be controlled by the lacPO [73] is a possibility. This can ensure the expression of the chaperone, which can result in the lower occurrence of misfolded protein. With the optimised purification protocol, this approach has good potential to obtain good quality protein as the protease activity should be limited when working at low degrees in addition to the use of protease inhibitors.

References

1. Ferlay J, S.I., Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. *Cancer incidence and mortality worldwide: Cancer fact sheets*. Globacan 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 2012 [cited 2017 07.09.17]; Available from: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.
2. Ferlay J, S.I., Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. *Cancer incidence and mortality worldwide: Population fact sheets*. Globacan 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 2012 [cited 2017 07.09.17]; Available from: http://globocan.iarc.fr/Pages/fact_sheets_population.aspx.
3. Watson, J.D., *Molecular biology of the gene*. 2014: Pearson.
4. Parham, P., *The immune system*. Fourth ed. 2014: Taylor & Francis Group.
5. Parkin, J. and B. Cohen, *An overview of the immune system.* Lancet, 2001. **357**(9270): p. 1777-89.
6. Banchereau, J. and K. Palucka, *Immunotherapy: Cancer vaccines on the move.* Nat Rev Clin Oncol, 2017.
7. Dranoff, G., *Cytokines in cancer pathogenesis and cancer therapy.* Nat Rev Cancer, 2004. **4**(1): p. 11-22.
8. Basu, P., et al., *Efficacy and safety of human papillomavirus vaccine for primary prevention of cervical cancer: A review of evidence from phase III trials and national programs.* South Asian Journal of Cancer, 2013. **2**(4): p. 187-192.
9. Díaz, A., et al., *Immune responses in breast cancer patients immunized with an anti-idiotype antibody mimicking NeuGc-containing gangliosides.* Clinical Immunology, 2003. **107**(2): p. 80-89.
10. Normanno, N., et al., *Epidermal growth factor receptor (EGFR) signaling in cancer.* Gene, 2006. **366**(1): p. 2-16.
11. Marquina, G., et al., *Gangliosides expressed in human breast cancer.* Cancer Res, 1996. **56**(22): p. 5165-71.
12. Krengel, U. and P.A. Bousquet, *Molecular recognition of gangliosides and their potential for cancer immunotherapies.* Frontiers in Immunology, 2014. **5**(325).
13. Nahta, R., G.N. Hortobágyi, and F.J. Esteva, *Growth factor receptors in breast cancer: Potential for therapeutic intervention.* The Oncologist, 2003. **8**(1): p. 5-17.
14. Witsch, E., M. Sela, and Y. Yarden, *Roles for growth factors in cancer progression.* Physiology (Bethesda), 2010. **25**(2): p. 85-101.
15. Wang, X., et al., *Ganglioside modulates ligand binding to the epidermal growth factor receptor.* Journal of Investigative Dermatology, 2001. **116**(1): p. 69-76.
16. Schnaar, R.L. and T. Kinoshita, *Glycosphingolipids and sialic acids*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2015, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).
17. Hakomori, S.-i., *The glycosynapse.* Proc Natl Acad Sci U S A, 2002. **99**(1): p. 225-232.
18. Salmaso, S. and P. Caliceti, *Stealth properties to improve therapeutic efficacy of drug nanocarriers*. Vol. 2013. 2013. 374252.
19. Irie, A., et al., *The molecular basis for the absence of N-glycolylneuraminic acid in humans.* Journal of Biological Chemistry, 1998. **273**(25): p. 15866-15871.

20. Chou, H.-H., et al., *Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11736-11741.
21. Varki, A., *N-glycolylneuraminic acid deficiency in humans.* Biochimie, 2001. **83**(7): p. 615-22.
22. Chou, H.H., et al., *A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence.* Proc Natl Acad Sci U S A, 1998. **95**(20): p. 11751-6.
23. Gagneux, P. and A. Varki, *Evolutionary considerations in relating oligosaccharide diversity to biological function.* Glycobiology, 1999. **9**(8): p. 747-55.
24. Bardor, M., et al., *Mechanism of uptake and incorporation of the non-human sialic acid N-glycolylneuraminic acid into human cells.* Journal of Biological Chemistry, 2005. **280**(6): p. 4228-4237.
25. Krengel, U., et al., *Structure and molecular interactions of a unique antitumor antibody specific for N-glycolyl GM3.* Journal of Biological Chemistry, 2004. **279**(7): p. 5597-5603.
26. Malykh, Y.N., R. Schauer, and L. Shaw, *N-glycolylneuraminic acid in human tumours.* Biochimie, 2001. **83**(7): p. 623-34.
27. Oliva, J.P., et al., *Clinical evidences of GM3 (NeuGc) ganglioside expression in human breast cancer using the 14F7 monoclonal antibody labelled with (99m)Tc.* Breast Cancer Res Treat, 2006. **96**(2): p. 115-21.
28. Lahera, T., et al., *Prognostic role of 14F7 mAb immunoreactivity against N-glycolyl GM3 ganglioside in colon cancer.* J Oncol, 2014. **2014**: p. 482301.
29. Ravindranath, M.H., et al., *Ganglioside GM3:GD3 ratio as an index for the management of melanoma.* Cancer, 1991. **67**(12): p. 3029-3035.
30. Carr, A., et al., *A mouse IgG1 monoclonal antibody specific for N-glycolyl GM3 ganglioside recognized breast and melanoma tumors.* Hybridoma, 2000. **19**(3): p. 241-7.
31. Rodriguez, M., et al., *Generation and characterization of an anti-idiotype monoclonal antibody related to GM3(NeuGc) ganglioside.* Hybrid Hybridomics, 2003. **22**(5): p. 307-14.
32. Tjandra, J.J., L. Ramadi, and I.F. McKenzie, *Development of human anti-murine antibody (HAMA) response in patients.* Immunol Cell Biol, 1990. **68 (Pt 6)**: p. 367-76.
33. Randall, K.L., *Rituximab in autoimmune diseases.* Australian Prescriber, 2016. **39**(4): p. 131-134.
34. McLaughlin, P., et al., *Rituximab chimeric anti-CD20 monoclonal antibody therapy for relapsed indolent lymphoma: half of patients respond to a four-dose treatment program.* J Clin Oncol, 1998. **16**(8): p. 2825-33.
35. Bartsch, R., C. Wenzel, and G.G. Steger, *Trastuzumab in the management of early and advanced stage breast cancer.* Biologics : Targets & Therapy, 2007. **1**(1): p. 19-31.
36. Gajdosik, Z., *Racotumomab - a novel anti-idiotype monoclonal antibody vaccine for the treatment of cancer.* Drugs Today (Barc), 2014. **50**(4): p. 301-7.
37. Alfonso, S., et al., *A randomized, multicenter, placebo-controlled clinical trial of racotumomab-alum vaccine as switch maintenance therapy in advanced non-small cell lung cancer patients.* Clin Cancer Res, 2014. **20**(14): p. 3660-71.
38. SL, R. *Immunotherapy with Racotumomab in advanced lung cancer*. 2016 Jul. [cited 2017; Phase 3]. Available from: https://clinicaltrials.gov/show/NCT01460472?link_type=CLINTRIALGOV&access_num=NCT01460472.

39. Cabilly, S., et al., *Generation of antibody activity from immunoglobulin polypeptide chains produced in Escherichia coli.* Proc Natl Acad Sci U S A, 1984. **81**(11): p. 3273-7.
40. Jefferis, R., *Glycosylation of recombinant antibody therapeutics.* Biotechnol Prog, 2005. **21**(1): p. 11-6.
41. Jefferis, R., *Antibody therapeutics: isotype and glycoform selection.* Expert Opin Biol Ther, 2007. **7**(9): p. 1401-13.
42. Sahdev, S., S.K. Khattar, and K.S. Saini, *Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies.* Mol Cell Biochem, 2008. **307**(1-2): p. 249-64.
43. Ahmad, Z.A., et al., *scFv antibody: principles and clinical application.* Clinical and Developmental Immunology, 2012. **2012**: p. 980250.
44. Chester, K.A., et al., *Clinical applications of phage-derived sFvs and sFv fusion proteins.* Disease Markers, 2000. **16**(1-2): p. 53-62.
45. Wörn, A. and A. Plückthun, *Stability engineering of antibody single-chain Fv fragments.* J Mol Biol, 2001. **305**(5): p. 989-1010.
46. Yoshimura, K., et al., *Monoclonal antibody 14F7, which recognizes a stage-specific immature oligodendrocyte surface molecule, inhibits oligodendrocyte differentiation mediated in co-culture with astrocytes.* J Neurosci Res, 1998. **54**(1): p. 79-96.
47. Rojas, G., et al., *Light-chain shuffling results in successful phage display selection of functional prokaryotic-expressed antibody fragments to N-glycolyl GM3 ganglioside.* J Immunol Methods, 2004. **293**(1-2): p. 71-83.
48. Carr, A., et al., *In vivo and in vitro anti-tumor effect of 14F7 monoclonal antibody.* Hybrid Hybridomics, 2002. **21**(6): p. 463-8.
49. Torbidoni, A.V., et al., *Immunoreactivity of the 14F7 mAb raised against N-glycolyl GM3 ganglioside in retinoblastoma tumours.* Acta Ophthalmol, 2015. **93**(4): p. e294-300.
50. Blanco, R., et al., *Immunohistochemical reactivity of the 14F7 monoclonal antibody raised against N-glycolyl GM3 ganglioside in some benign and malignant skin neoplasms.* ISRN Dermatol, 2011. **2011**: p. 848909.
51. Blanco, R., et al., *Immunoreactivity of the 14F7 mAb raised against N-glycolyl GM3 ganglioside in primary lymphoid tumors and lymph node metastasis.* Patholog Res Int, 2013. **2013**: p. 920972.
52. Blanco, R., et al., *Immunoreactivity of the 14F7 mAb raised against N-glycolyl GM3 ganglioside in epithelial malignant tumors from digestive system.* ISRN Gastroenterol, 2011. **2011**: p. 645641.
53. Blanco, R., et al., *Immunoreactivity of the 14F7 mAb (Raised against N-glycolyl GM3 ganglioside) as a positive prognostic factor in non-small-cell lung cancer.* Patholog Res Int, 2012. **2012**: p. 235418.
54. Blanco, R., et al., *Tissue reactivity of the 14F7 mAb raised against N-glycolyl GM3 ganglioside in tumors of neuroectodermal, mesodermal, and epithelial origin.* J Biomark, 2013. **2013**: p. 602417.
55. Roque-Navarro, L., et al., *Anti-ganglioside antibody-induced tumor cell death by loss of membrane integrity.* Molecular Cancer Therapeutics, 2008. **7**(7): p. 2033.
56. Fernandez-Marrero, Y., et al., *A cytotoxic humanized anti-ganglioside antibody produced in a murine cell line defective of N-glycolylated-glycoconjugates.* Immunobiology, 2011. **216**(12): p. 1239-47.
57. Rojas, G., et al., *Engineering the binding site of an antibody against N-glycolyl GM3: From functional mapping to novel anti-ganglioside specificities.* ACS Chemical Biology, 2013. **8**(2): p. 376-386.

58. Bird, R.E., et al., *Single-chain antigen-binding proteins.* Science, 1988. **242**(4877): p. 423-6.
59. Glockshuber, R., et al., *A comparison of strategies to stabilize immunoglobulin Fv-fragments.* Biochemistry, 1990. **29**(6): p. 1362-7.
60. Nieba, L., et al., *Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved in vivo folding and physical characterization of an engineered scFv fragment.* Protein Eng, 1997. **10**(4): p. 435-44.
61. Porowinska, D., et al., *Prokaryotic expression systems.* Postepy Hig Med Dosw (Online), 2013. **67**: p. 119-29.
62. Khow, O. and S. Suntrarachun, *Strategies for production of active eukaryotic proteins in bacterial expression system.* Asian Pacific Journal of Tropical Biomedicine, 2012. **2**(2): p. 159-162.
63. Gilbert, H.F., *Molecular and cellular aspects of thiol-disulfide exchange.* Adv Enzymol Relat Areas Mol Biol, 1990. **63**: p. 69-172.
64. Pugsley, A.P., *The complete general secretory pathway in gram-negative bacteria.* Microbiol Rev, 1993. **57**(1): p. 50-108.
65. Sapriel, G., C. Wandersman, and P. Delepelaire, *The N-terminus of the HasA protein and the SecB chaperone cooperate in the efficient targeting and secretion of HasA via the ATP-binding cassette transporter.* Journal of Biological Chemistry, 2002. **277**(8): p. 6726-6732.
66. Takkinen, K., et al., *An active single-chain antibody containing a cellulase linker domain is secreted by Escherichia coli.* Protein Engineering, Design and Selection, 1991. **4**(7): p. 837-841.
67. Bothmann, H. and A. Pluckthun, *The periplasmic Escherichia coli peptidylprolyl cis,trans-isomerase FkpA. I. Increased functional expression of antibody fragments with and without cis-prolines.* J Biol Chem, 2000. **275**(22): p. 17100-5.
68. Ramm, K. and A. Pluckthun, *The periplasmic Escherichia coli peptidylprolyl cis,trans-isomerase FkpA. II. Isomerase-independent chaperone activity in vitro.* J Biol Chem, 2000. **275**(22): p. 17106-13.
69. Ramm, K. and A. Plückthun, *High enzymatic activity and chaperone function are mechanistically related features of the dimeric E. coli peptidyl-prolyl-isomerase FkpA11. Edited by C. R. Matthews.* Journal of Molecular Biology, 2001. **310**(2): p. 485-498.
70. Ying, B.-W., et al., *Chaperone-assisted folding of a single-chain antibody in a reconstituted translation system.* Biochemical and Biophysical Research Communications, 2004. **320**(4): p. 1359-1364.
71. Miot, M. and J.-M. Betton, *Protein quality control in the bacterial periplasm.* Microbial Cell Factories, 2004. **3**: p. 4-4.
72. Loset, G.A., et al., *Construction, evaluation and refinement of a large human antibody phage library based on the IgD and IgM variable gene repertoire.* J Immunol Methods, 2005. **299**(1-2): p. 47-62.
73. Gunnarsen, K.S., et al., *Periplasmic expression of soluble single chain T cell receptors is rescued by the chaperone FkpA.* BMC Biotechnol, 2010. **10**: p. 8.
74. Loset, G.A., et al., *Functional phage display of two murine alpha/beta T-cell receptors is strongly dependent on fusion format, mode and periplasmic folding assistance.* Protein Eng Des Sel, 2007. **20**(9): p. 461-72.
75. Johannesen, H., *The assassin*, in *Department of Biosciences*. 2014, University of Oslo: https://www.duo.uio.no/bitstream/handle/10852/40619/Johannesen-Hedda_Thesis.pdf?sequence=9.

76. Blow, D., *Outline of crystallography for biologists*. 2003: Oxford University Press.
77. Chayen, N.E., *Turning protein crystallisation from an art into a science.* Curr Opin Struct Biol, 2004. **14**(5): p. 577-83.
78. Nilson, B.H., et al., *Protein L from Peptostreptococcus magnus binds to the kappa light chain variable domain.* J Biol Chem, 1992. **267**(4): p. 2234-9.
79. Gan, S.D. and K.R. Patel, *Enzyme Immunoassay and Enzyme-Linked Immunosorbent Assay.* Journal of Investigative Dermatology, 2013. **133**(9): p. 1-3.
80. Lakowicz, J.R., *Principles of fluorescence spectroscopy*. 2007: Springer US.
81. Artimo, P., et al., *ExPASy: SIB bioinformatics resource portal.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W597-603.
82. Wingfield, P.T., *Overview of the purification of recombinant proteins.* Current protocols in protein science / editorial board, John E. Coligan ... [et al.], 2015. **80**: p. 6.1.1-6.1.35.
83. Dougherty, W.G., S.M. Cary, and T.D. Parks, *Molecular genetic analysis of a plant virus polyprotein cleavage site: a model.* Virology, 1989. **171**(2): p. 356-64.
84. Løset, G.Å., *Echerichia coli XL1-Blue cells does not grow under 22 degrees*. 2017.
85. McCarter, J.D., et al., *Substrate specificity of the Escherichia coli outer membrane protease OmpT.* Journal of Bacteriology, 2004. **186**(17): p. 5919-5925.
86. Strongin, A.Y., D.I. Gorodetsky, and V.M. Stepanov, *The study of Escherichia coli proteases. Intracellular serine protease of E. coli-an analogue of bacillus proteases.* J Gen Microbiol, 1979. **110**(2): p. 443-51.
87. Kolhe, P., E. Amend, and S.K. Singh, *Impact of freezing on pH of buffered solutions and consequences for monoclonal antibody aggregation.* Biotechnol Prog, 2010. **26**(3): p. 727-33.
88. Sonnino, S., et al., *Aggregation properties of GM3 ganglioside (II3Neu5AcLacCer) in aqueous solutions.* Chem Phys Lipids, 1990. **52**(3-4): p. 231-41.
89. Banerjee, R.K. and A.G. Datta, *Proteoliposome as the model for the study of membrane-bound enzymes and transport proteins.* Mol Cell Biochem, 1983. **50**(1): p. 3-15.
90. Scalise, M., et al., *Proteoliposomes as tool for assaying membrane transporter functions and interactions with xenobiotics.* Pharmaceutics, 2013. **5**(3): p. 472-497.
91. Maynard, J.A., et al., *Next generation SPR technology of membrane-bound proteins for ligand screening and biomarker discovery.* Biotechnology journal, 2009. **4**(11): p. 1542-1558.
92. Segatori, V.I., et al., *Antitumor protection by NGcGM3/VSSP vaccine against transfected B16 mouse melanoma cells overexpressing N-glycolylated gangliosides.* In Vivo, 2012. **26**(4): p. 609-17.
93. Perez, K., et al., *NGcGM3/VSSP vaccine as treatment for melanoma patients.* Hum Vaccin Immunother, 2013. **9**(6): p. 1237-40.
94. Estevez, F., et al., *Enhancement of the immune response to poorly immunogenic gangliosides after incorporation into very small size proteoliposomes (VSSP).* Vaccine, 1999. **18**(1-2): p. 190-7.
95. de la Torre, A., et al., *NGlycolylGM3/VSSP vaccine in metastatic breast cancer patients: Results of phase I/IIa clinical trial.* Breast Cancer : Basic and Clinical Research, 2012. **6**: p. 151-157.

Appendix A: Materials

Table A. 1: Kits

FastDigest Primers	Invitrogen by Thermo Fisher Scientific
Dialysis kit	Thermo Fisher Scientific
Miniprep	SIGMA Aldrich
PCR kit	Thermo Fisher Scientific
PCR Purification Kit	QIAGEN
QIAprep Spin Midiprep Kit	QIAGEN
QIAquick gel extraction Kit	QIAGEN

Table A. 2: Reagents

Chemicals	**Vendor**
1 kbp DNA ladder for agarose gel	Thermo Scientific
14F7 mAb from collaboration partners	CIM
2-propanol	Sigma
3´,5,5´-tetramethylbenzidine	Chalbiocem
Acetic Acid	Merck
Agarose	Sigma
Ampicillin	AppliChem
Bovine Serum Albumin	Sigma
c*O*mplete inhibitor	Roche
Coomassie Brilliant Blue G250	VWR
D-(+)-glucose	Sigma
D-(+)-sucrose	Fluka
Dithiothreitol (DTT)	Bio-Rad Laboratories
DNA Loading Dye Solution (5x)	Thermo Scientific
DNAse	AppliChem
dNTP mix	Thermo Scientific
E. Coli XL1-Blue	Stratagene
EDTA	Fluka
Ethanol (Absolute)	Prolabo
Glycerol	Prolabo
Glycine	Sigma
HCl	Merck

HEPES	Sigma
*Hind*III restriction enzyme	Thermo Fisher Scientific
Imidazole	Sigma
Lysozyme	Sigma
Magnesium Chloride	BioChemica
Methanol	Prolabo
*Mlu*I restriction enzyme	Thermo Fisher Scientific
Na2HPO4	G-Biosciences
NaH2PO4	Fluka
*Nco*I restriction enzyme	Thermo Fisher Scientific
NeuGc GM3, from collaboration partners	CIM
*Not*I restriction enzyme	Thermo Fisher Scientific
NuPAGE® Loading buffer (4x)	Invitrogen for Thermo Fisher Scientific
NuPAGE® LDS Sample Buffer (4x)	Invitrogen for Thermo Fisher Scientific
PEG solutions	Fluka
Peptone (Tryptone)	Merck
Peptone from casein	Merck
Phosphate buffered saline (PBS)	Life technology for Thermo Fisher Scientific
Phusion DNA polymerase, 2U/µl	Thermo Scientific
Phusion reaction buffer, 5x	Thermo Scientific
PMSF	Sigma
Primers, PCR	Thermo Fisher Scientific
Primers, sequencing	Thermo Fisher Scientific
Protein L-HRP, 1mg/mL	Genscript
Rnase A	Sigma
SeeBlue® Plus2 standard, MW marker	Invitrogen for Thermo Fisher Scientific
Sodium chloride (NaCl)	Prolabo
Sodium citrate ($Na_3C_6H_5O_7$)	Sigma
Sodium hydroxide (NaOH)	Kebo Lab
Synthesized 14F7 scFv genes	Life technology for Thermo Fisher Scientific
T4 DNA ligase	Thermo Scientific
T4 DNA ligase buffer (10x)	Thermo Scientific
Tris Base	Chalbiocem
Tris-Hydrogenchloride (Tris-HCl)	Chalbiocem
Tween-20	Sigma

Urea	Merck
Yeast Extract	Merck

Table A. 3: Equipments

Equipment	**Manufacturer**
Amicon Ultra filter 15ml	Millipore
Amicon Ultra filter 2ml	Millipore
Crystal Clear Sealing Tape	Hampton
Crystallisation Plate, 24 MRC	SWISSCI
Crystallisation Plate, 96 MRC	SWISSCI
Cuvettes	Sarstedt
Eppendorf tupes, 1.5 ml and 2 ml	Eppendorf
Finnpipette	Thermo Fisher Scientific
Finnpipette multicanal	Thermo Fisher Scientific
Nunc-Immuno 96 MicroWell PolySorb solid plates	Sigma
PCR tubes	Eppendorf
Petman pipettes	Gilson
Syringe filter, 25 mm, 0.2 µm PES membrane	VWR
TPP® tissue culture plates, 24 well	Sigma
Vacuum filter, 500ml, 0.2 µm PES membrane	VWR

Table A. 4: Software

Software	**Vendor**
4 Peaks	Nucleobytes
EMBOSS Needle	EMBL-EBI
ProtParam, ExPASy	Swiss Institute of Bioinformatics
PyMOL	Schrödinger
SkanItTM Software	Thermo Fisher Scientific
Unicorn 5.11	GE Healthcare
Wasp Run	Douglas Instruments

Table A. 5: Machines

Machines	**Vendor**
ÄKTA purifier-900	GE Healthcare
Avanti Centrifuge J-26 XP	Beckman Coulter
Centrifuge 5810 R	Eppendorf
Electrophoresis power supply-EPS 601	GE Healthcare
FP-8200 Spectrofluorometer	Jasco
MQ-H2O, Direct Q	Millipore
Nanodrop 2000c	Thermo Fisher Scientific
NanoPhotometer	IMPLEN
Orbitrap-XL	Thermo Fisher
Oryz4 robot	Douglas Instruments
Thermo ScientificTM Varioskan LUXTM Multimode Microplate reader	Thermo Fisher Scientific
Tuttnauer 3870-ML, autoclave	Tuttnauer

Appendix B: Solutions, buffers and gels

Media

Table B. 1: Media

2xYT Medium 1L, pH 7.8		LB Medium 1L, pH 7.1	
10 g	Yeast	10 g	Yeast
16 g	Peptone	5 g	Peptone
5 g	NaCl	10 g	NaCl
Water up to 1 L, autoclaved and stored at 4 °C		*Water up to 1 L, autoclaved and stored at 4 °C*	

Protein extraction

Table B. 2: Solutions for protein extraction

Sucrose solution:		$MgCl_2$ protease inhibitor mix:	
Tris-HCl pH 8.0	20 mM	$MgCl_2$	5 mM
Sucrose	25 % (w/v)	Protease inhibitor cocktail (Roche)	1 tablet per 50 mL $MgCl_2$
EDTA	5 mM	or PMSF	1 mM

Gravity Protein L-column

Table B. 3: Buffers for Gravity Protein L-column

1xBinding buffer, pH 7.8		**10xBinding buffer, pH 7.8**	
20 mM	Na PO_4	200 mM	Na PO_4
150 mM	NaCl	1.5 M	NaCl

Elution buffer, pH 2-3			**Neutralisation buffer, pH9**	
0.1 M	NaCitrate	pH 2-3	1 M	Tris
300 mM	NaCl			

Pierce™ Protein L chromatography cartridge

Table B. 4: Buffers for Protein L chromatography cartridge

Binding buffer, 1 L			**Elution buffer, 1 L**		
100 mM		Sodium phosphate, NaH_2PO_4/Na_2HPO_4	0.1 M	7.5 g	Glycine
150 mM	30 mL of 5 M stock	Sodium chloride, NaCl			
Add water to 1 L, check pH 7.2			*Add water to 1 L, check pH 2-3. Add Sodium azid as a preservative.*		
Add sodium azid as preservative.			*Filtered and degassed*		
Filtered and degassed					

Neutralization buffer, 0.5 L	
1 M	Tris
Add water to 0.5 L, check pH 7.5-9	

Size exclusion chromatography

Table B. 5: Buffers for SEC

Phosphate Buffered Saline, PBS, 1L			**Tris buffer for SEC, 1L**		
137 mM	8 g	NaCl	20 mM	20 mL of 1 M stock	Tris-HCl pH 7.5
2.7 mM	0.2 g	KCl	100 mM	20 mL of 5M stock	NaCl
10.1 mM	1.44 g	Na_2HPO_4			
2 mM	0.78 g	NaH_2PO_4			
Add water to 1 L, check pH of 7.4			*Add water to 1 L, check pH of 7.5*		
Filtered and degassed			*Filtered and degassed*		

HiTrap His-column

Table B. 6: Buffers for His-column

IMAC Binding buffer pH 7.8		**IMAC Elution buffer pH 7.8**	
20 mM	Imidazole	500 mM	Imidazole
50 mM	Tris-HCl	50 mM	Tris-HCl
500 mM	NaCl	500 mM	NaCl
Filtered and degassed before use.		*Filtered and degassed before use.*	

ELISA

Table B. 7: Solutions for ELISA

Immobilised:					
NeuGc GM3	**10 µg/mL**	**5000 µL**	**NeuAc GM3**	**10 µg/mL**	**5000 µL**
NeuGc GM3	1000 µg/mL	50 µL	NeuAc GM3	5000 µg/mL	10 µL
Methanol	100 %	4950 µL	Methanol	100 %	4990 µL
NeuGc GM3	**5 µg/mL**	**5000 µL**	**NeuAc GM3**	**5 µg/mL**	**5000 µL**
NeuGc GM3	1000 µg/mL	25 µL	NeuAc GM3	5000 µg/mL	5 µL
Methanol	100 %	4975 µL	Methanol	100 %	4995 µL
Buffers:					
PBS-T 0.1 %			**PBS-T 0.05 %**		
PBS	250 mL		PBS	250 mL	
Tween-20	250 µL		Tween-20	250 µL	
PBS-B			**PBS-BT**		
PBS	50 mL		PBS	50 mL	
BSA	1.0 g		BSA	1.0 g	

Stock 14F7 scFv 3.3 mg/mL = 0.116 mM = 116 µM

6 µL 14F7 scFv stock in 1344 µL PBS-BT = 1350 µL of 500 nM 14F7 scFv

Table B. 8: 1:3 dilution series of 14F7 scFv C1*

nM	**PBS-BT**	**Stock solution**	**Total volume**
500.0	1344 µL	6 µL	1350 µL – 400 µL = 950 µL
nM	**PBS-BT**	**Well solution**	**Total volume**
166.7	800 µL	400 µL of 500.0 nM	800 µL

55.6	800 µL	400 µL of 166.7 nM	800 µL
18.5	800 µL	400 µL of 55.6 nM	800 µL
6.2	800 µL	400 µL of 18.5 nM	800 µL
2.1	800 µL	400 µL of 6.2 nM	1200 µL

Stock 14F7 mAb 5 mg/mL = 33.3 µM = 33333 nM

11 µL 14F7 mAb stock in 1189 µL PBS-BT = 1200 µL of 300 nM 14F7 mAb

Table B. 9: **1:3 dilution series of 14F7 mAb**

nM	**PBS-BT**	**Stock solution**	**Total volume**
300 nM	1189 µL	11 µL	1200 µL – 400 µL = 800 µL
nM	**PBS-BT**	**Well solution**	**Total volume**
100	800 µL	400 µL of 300 nM	800 µL
33.3	800 µL	400 µL of 100 nM	800 µL
11.1	800 µL	400 µL of 33.3 nM	800 µL
3.7	800 µL	400 µL of 11.1 nM	810 µL

Stock irrelevant mAb 5.35 mg/mL

2.7 µL in 297.3 µL PBS-BT = 300 µL of 300 nM irrelevant mAb

Table B. 10: 1:100 dilution series of the irrelevant mAb

nM	**PBS-BT**	**Stock solution**	**Total volume**
300 nM	297.3 µL	2.7 µL	300 µL – 3 µL = 297
nM	**PBS-BT**	**Well solution**	**Total volume**
3	297 µL	3 µL of 300 nM	300 µL

Stock irrelevant scFv 1.26 mg/ml

2.8 µL in 247.2 µL PBS-BT = 250 µL of 500 nM irrelevant scFv

Table B. 11: 1:100 dilution series of the irrelevant scFv

nM	**PBS-BT**	**Stock solution**	**Total volume**
500 µM	247.2 µL	2.8 µL	250 µL – 2.5 µL = 247.5 µL
nM	**PBS-BT**	**Well solution**	**Total volume**
5	247.5 µL	2.5 µL of 500 nM	250 µL

Appendix C: Cloning

Table C. 1: Digest and ligation reactions of pFKPEN and C1#

Double digest of pFKPEN-C1* to remove C1*		Double digest of V_H-L_R-$V_{L.A}$ for pFKPEN ligation creating pFKPEN-C1#	
1 µg –> 0.9 µL	1138 ng/µL pFKPEN-C1*	0.2 µg -> 15 µL	12.5 ng/µL V_H-L_R-$V_{L.A}$
15 µL	Water, nuclease-free	11 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Buffer
1 µL	*Not*I	1 µL	*Not*I
1 µL	*Nco*I	1 µL	*Nco*I
20 µL	Total	30 µL	Total
4:1 Ligation of pFKPEN (vector) and V_H-L_R-$V_{L.A}$ (insert) creating pFKPEN-C1#		Control ligation reaction of pFKPEN (vector) without insert	
27 µL	2,4 ng/µL V_H-L_R-$V_{L.A}$	0 µL	Insert
3 µL	7,4 ng/µL pFKPEN	3 µL	7,4 ng/µL pFKPEN
0 µL	Water, nuclease-free	27 µL	MQ-H_2O
2 µL	10X Ligase Buffer	2 µL	10X Ligase Buffer
1 µL	T4 DNA Ligase	1 µL	T4 DNA Ligase
33 µL	Total	33 µL	Total

Table C. 2: Digest and ligation reactions of pFKPEN and C4#.

Double digest of pFKPEN-C1.1# for C4#		Double digest of PCR L_C-V_L for C4#	
5 µL	240 ng/µL pFKPEN-C1.1#	2.5 µL	80 ng/µL LC-VL
11 µL	Water, nuclease-free	13.5 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Green Buffer
1 µL	*Not*I	1 µL	*Not*I
1 µL	*Hind*III	1 µL	*Hind*III
20 µL	Total	20 µL	Total
4:1 Ligation reaction of pFKPEN (vector) and L_C-V_L (insert) creating pFKPEN-C4#		Control ligation reaction of pFKPEN (vector) without insert	
11.4 µL	4.9 ng/µL L_C-V_L	0 µL	Insert
2.8 µL	17.5 ng/µL pFKPEN, C1#	2.8 µL	17.5 ng/µL pFKPEN, C1#
2.8 µL	Water, nuclease-free	14.2 µL	Water, nuclease-free
2 µL	10X Ligase Buffer	2 µL	10X Ligase Buffer
1 µL	T4 DNA ligase	1 µL	T4 DNA ligase

20 µL	Total	20 µL	Total

Table C. 3: Digest and ligation reactions of pFKPEN and C2#.

Double digest of pFKPEN-C1.1# for C2#		Double digest of PCR V_L for C2#	
5 µL	240 ng/µL pFKPEN-C1.1#	2.75 µL	72.5 ng/µL V_L
11 µL	Water, nuclease-free	13.25 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Green Buffer
1 µL	*Not*I	1 µL	*Not*I
1 µL	*Mlu*I	1 µL	*Mlu*I
20 µL	Total	20 µL	Total
4:1 Ligation reaction of pFKPEN (vector) and V_L (insert) creating pFKPEN-C2#		Control ligation reaction of pFKPEN (vector) without insert	
13 µL	5.2 ng/µL V_L	0 µL	Insert
3.3 µL	15 ng/µL pFKPEN, C1#	3.3 µL	15 ng/µL pFKPEN, C1#
0.7 µL	Water, nuclease-free	13.7 µL	Water, nuclease-free
2 µL	10X Ligase Buffer	2 µL	10X Ligase Buffer
1 µL	T4 DNA ligase	1 µL	T4 DNA ligase
20 µL	Total	20 µL	Total

Table C. 4: Digest and ligation reactions of pFKPEN and C3#.

Double digest of pFKPEN-C1# for C3#		Double digest of PCR V_H-L_C for C3#	
5 µL	240 ng/µL pFKPEN-C1.1#	5 µL	10.4 ng/µL V_H-L_C
11 µL	Water, nuclease-free	11 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Green Buffer
1 µL	*Nco*I	1 µL	*Nco*I
1 µL	*Mlu*I	1 µL	*Mlu*I
20 µL	Total	20 µL	Total
4:1 Ligation reaction of pFKPEN (vector) and V_H-L_C (insert) creating pFKPEN-C3#		Control ligation reaction of pFKPEN (vector) without insert	
14.5 µL	2.5 ng/µL V_H-L_C	0 µL	Insert
2.5 µL	37.5 ng/µL pFKPEN	10 µL	37.5 ng/µL pFKPEN
0 µL	Water, nuclease-free	7 µL	Water, nuclease-free
2 µL	10X Ligase Buffer	2 µL	10X Ligase Buffer
1 µL	T4 DNA ligase	1 µL	T4 DNA ligase

20 µL	Total	20 µL	Total

Control digestion

Table C. 5: Control digest reactions

Double digest of pFKPEN-C1.1#		Double digest of pFKPEN-C1.2#	
1 µg -> 5 µL	240 ng/µL pFKPEN-C1.1#	1 µg -> 5 µL	220 ng/µL pFKPEN-C1.2#
11 µL	Water, nuclease-free	11 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Green Buffer
1 µL	*Not*I	1 µL	*Not*I
1 µL	*Nco*I	1 µL	*Nco*I
20 µL	Total	20 µL	Total
Double digest of pFKPEN-C1.3#		Double digest of pFKPEN-C1.4#	
1 µg -> 5 µL	232 ng/µL pFKPEN-C1.3#	1 µg -> 6 µL	185 ng/µL pFKPEN-C1.4#
11 µL	Water, nuclease-free	10 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Green Buffer
1 µL	*Not*I	1 µL	*Not*I
1 µL	*Nco*I	1 µL	*Nco*I
20 µL	Total	20 µL	Total
Double digest of pFKPEN-C1.5#		Double digest of pFKPEN-C1.6#	
1 µg -> 6 µL	188 ng/µL pFKPEN-C1.5#	1 µg -> 5 µL	230 ng/µL pFKPEN-C1.6#
10 µL	Water, nuclease-free	11 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Green Buffer
1 µL	*Not*I	1 µL	*Not*I
1 µL	*Nco*I	1 µL	*Nco*I
20 µL	Total	20 µL	Total
Double digest of pFKPEN-C1.7#		Double digest of pFKPEN-C1.8#	
1 µg -> 5 µL	232 ng/µL pFKPEN-C1.7#	1 µg -> 5 µL	242 ng/µL pFKPEN-C1.8#
11 µL	Water, nuclease-free	11 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Green Buffer
1 µL	*Not*I	1 µL	*Not*I
1 µL	*Nco*I	1 µL	*Nco*I
20 µL	Total	20 µL	Total

Table C. 6: Control digest reactions

Double digest of pFKPEN-C1.9#		Double digest of pFKPEN-C1.10#	
1 µg -> 5 µL	218 ng/µL pFKPEN-C1.9#	1 µg -> 5 µL	228 ng/µL pFKPEN-C1.10#
11 µL	Water, nuclease-free	11 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Green Buffer
1 µL	*Not*I	1 µL	*Not*I
1 µL	*Nco*I	1 µL	*Nco*I
20 µL	Total	20 µL	Total
Double digest of pFKPEN-C3.1#		Double digest of pFKPEN- C3.2#	
1 µg -> 6 µL	188 ng/µL pFKPEN-C3.1#	1 µg -> 6 µL	163 ng/µL pFKPEN-C3.2#
10 µL	Water, nuclease-free	10 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Green Buffer
1 µL	*Not*I	1 µL	*Not*I
1 µL	*Nco*I	1 µL	*Nco*I
20 µL	Total	20 µL	Total
Double digest of pFKPEN-C3.3#		Double digest of pFKPEN- C3.4#	
1 µg -> 5 µL	208 ng/µL pFKPEN-C3.3#	1 µg -> 5 µL	205 ng/µL pFKPEN-C3.4#
11 µL	Water, nuclease-free	11 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Green Buffer
1 µL	*Not*I	1 µL	*Not*I
1 µL	*Nco*I	1 µL	*Nco*I
20 µL	Total	20 µL	Total
Double digest of pFKPEN-C3.5#		Double digest of pFKPEN- C3.6#	
1 µg -> 4 µL	255 ng/µL pFKPEN-C3.5#	1 µg -> 4 µL	245 ng/µL pFKPEN-C3.6#
12 µL	Water, nuclease-free	12 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Green Buffer
1 µL	*Not*I	1 µL	*Not*I
1 µL	*Nco*I	1 µL	*Nco*I
20 µL	Total	20 µL	Total

Table C. 7: Control digest reactions

Double digest of pFKPEN-C3.7#		Double digest of pFKPEN- C3.8#	
1 µg -> 6 µL	193 ng/µL pFKPEN-C3.7#	1 µg -> 6 µL	183 ng/µL pFKPEN-C3.8#
10 µL	Water, nuclease-free	10 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Green Buffer
1 µL	*Not*I	1 µL	*Not*I
1 µL	*Nco*I	1 µL	*Nco*I
20 µL	Total	20 µL	Total
Double digest of pFKPEN- C2.2#		Double digest of pFKPEN- C4.2#	
1 µg -> 3 µL	375 ng/µL pFKPEN-C2.2#	1 µg -> 8 µL	125 ng/µL pFKPEN-C4.2#
13 µL	Water, nuclease-free	8 µL	Water, nuclease-free
2 µL	10X FastDigest Green Buffer	2 µL	10X FastDigest Green Buffer
1 µL	*Not*I	1 µL	*Not*I
1 µL	*Nco*I	1 µL	*Nco*I
20 µL	Total	20 µL	Total

Appendix D: PCR; programs and mixtures

Mixtures

Table D. 1: First PCR reaction setup:

	V_H-L_R-$V_{L.A}$	L_C-V_L	L_C	V_L
2 µL	5 ng/µL C1#	5 ng/µL C4#	5 ng/µL C4#	5 ng/µL C4#
34 µL	dH_2O	dH_2O	dH_2O	dH_2O
10 µL	5X Phusion HF Buffer	5X Phusion HF Buffer	5X Phusion HF Buffer	5X Phusion HF Buffer
1 µL	10 mM dNTPs	10 mM dNTPs	10 mM dNTPs	10 mM dNTPs
1.25 µL	nVh_FP 4 µM	nLc_FP 4 µM	nLc_FP 4 µM	nVlo _FP 4 µM
1.25 µL	nVl_RP 4 µM	nVl_RP 4 µM	nLc_RP 4 µM	nVl_RP 4 µM
0.5 µL	Phusion DNA Polymerase	Phusion DNA Polymerase	Phusion DNA Polymerase	Phusion DNA Polymerase
50 µL				

Table D. 2: Optimised PCR reaction setup. Higher volume of primers.

	V_H-L_R-$V_{L.A}$	L_C-V_L	L_C	V_L
2 µL	5 ng/µL C1#	5 ng/µL C4#	5 ng/µL C4#	5 ng/µL C4#
24 µL	dH_2O	dH_2O	dH_2O	dH_2O
10 µL	5X Phusion HF Buffer	5X Phusion HF Buffer	5X Phusion HF Buffer	5X Phusion HF Buffer
1 µL	10 mM dNTPs	10 mM dNTPs	10 mM dNTPs	10 mM dNTPs
6.25 µL	nVh_FP 4 µM	nLc_FP 4 µM	nLc_FP 4 µM	nVlo _FP 4 µM
6.25 µL	nVl_RP 4 µM	nVl_RP 4 µM	nLc_RP 4 µM	nVl_RP 4 µM
0.5 µL	Phusion DNA Polymerase	Phusion DNA Polymerase	Phusion DNA Polymerase	Phusion DNA Polymerase
50 µL				

Table D. 3: Optimised PCR reaction setup for VH-LC

C3# Optimized	V_H-L_C
2 µL	5 ng/µL PFKPEN-C4#
10 µL	5X Phusion HF Buffer
1 µL	10 mM dNTPs
6.25 µL	nVh_FP 4 µM -> 0.5 µM
6.25 µL	nLc_RP 4 µM -> 0.5 µM
24 µL	dH_2O
0.5 µL	Phusion DNA Polymerase
50 µL	Total

Programs

Table D. 4: PCR programs

PCR of C1# and V_H-L_C for C3#	Time	Temp	Cycles
Initial Denaturation	1 min	98 °C	1
Denaturation Annealing Extension	10 s 20 s 20 s	98 °C 67.9 °C 72 °C	28
Final extension	10 min hold	72 °C 4 °C	1

PCR of C4#	Time	Temp	Cycles
Initial Denaturation	1 min	98 °C	1
Denaturation Annealing Extension	10 s 20 s 20 s	98 °C 68.6 °C 72 °C	28
Final extension	10 min hold	72 °C 4 °C	1

PCR of L_C	Time	Temp	Cycles
Initial Denaturation	1 min	98 °C	1
Denaturation Annealing Extension	10 s 20 s 20 s	98 °C 68.9 °C 72 °C	28
Final extension	10 min hold	72 °C 4 °C	1

PCR of V_L	Time	Temp	Cycles
Initial Denaturation	1 min	98 °C	1
Denaturation Annealing Extension	10 s 20 s 20 s	98 °C 66.2 °C 72 °C	28
Final extension	10 min hold	72 °C 4 °C	1

Appendix E: Standard curves

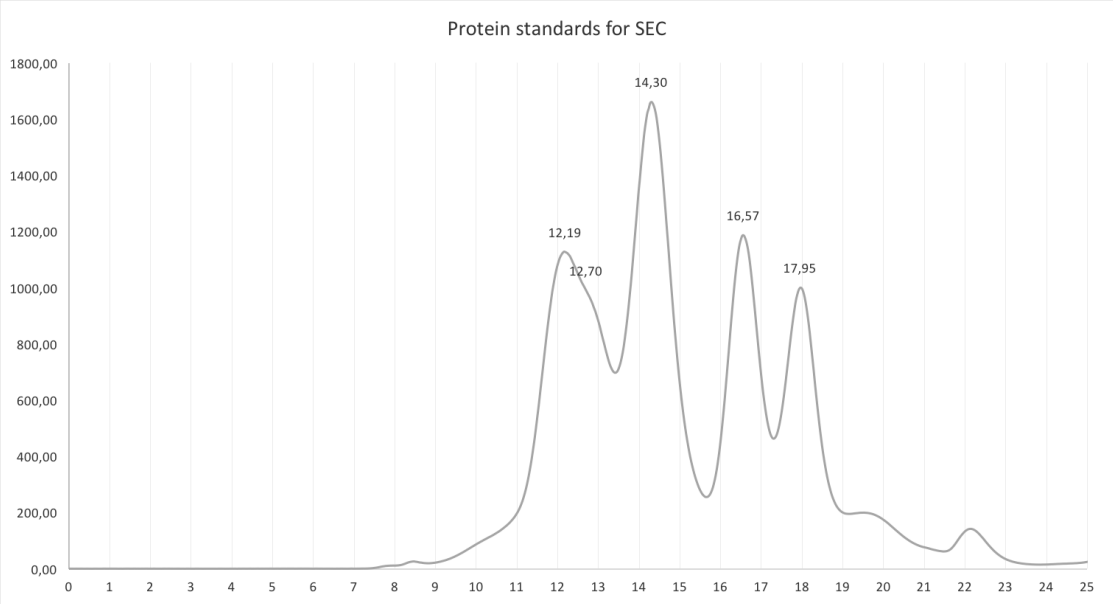


Figure E. 1: Protein standards run on Superdex 200 GL

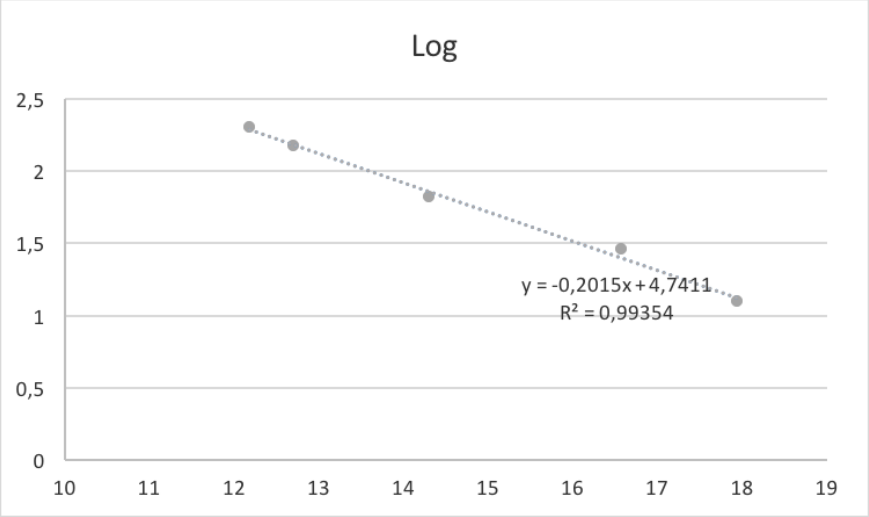


Figure E. 2: Standard curve obtained from plotting the values in Table E.1

Table E. 1: Elution volumes, molecular weights, and logarithmic values

mL	kDa	Log
12,19	200	2,301029996
12,7	150	2,176091259
14,3	66	1,819543936
16,57	29	1,462397998
17,95	12,4	1,093421685

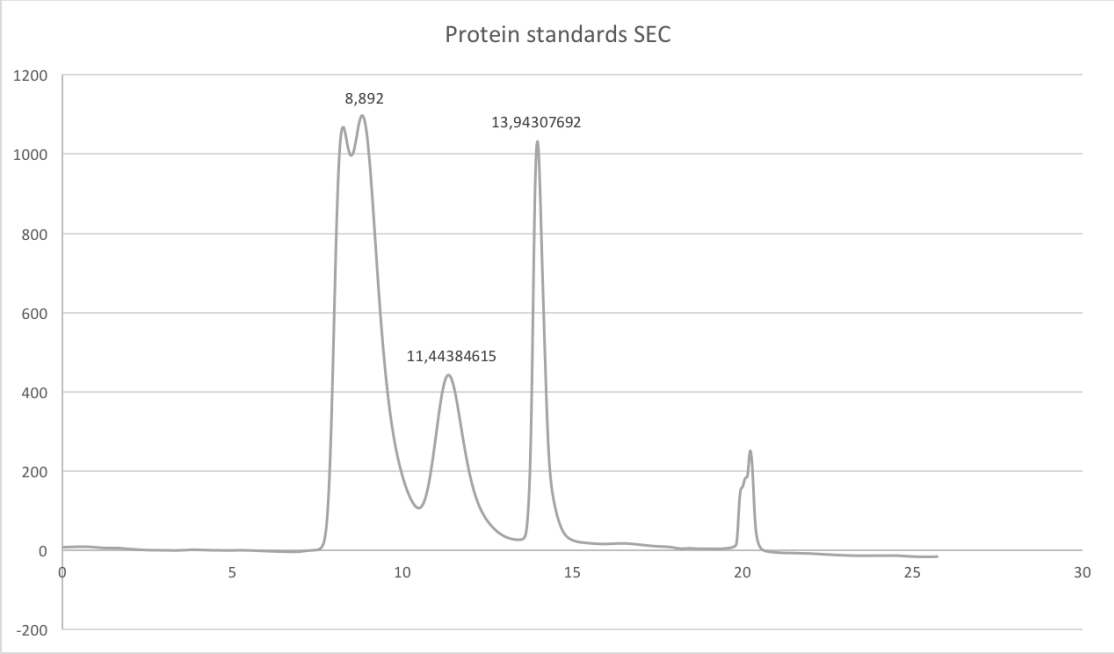


Figure E. 3: Protein standards run on Superdex75 GL

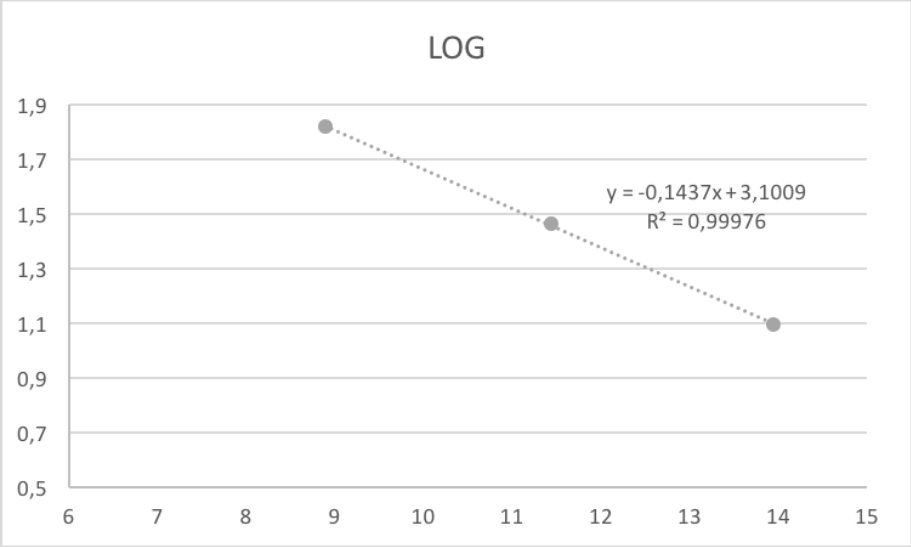


Figure E. 4: Standard curve obtained from plotting the logarithmic values in Table E.2.

Table E. 2: Elution volumes, molecular weights, and logarithmic values.

mL	kDa	Log
8,265	200	2,301029996
8,26	150	2,176091259
8,892	66	1,819543936
11,44384615	29	1,462397998
13,94307692	12,4	1,093421685

Appendix F: Nucleic acid sequences

>C1#

actt**ccatgg**cgcaggtgcagctgcagcagagcggcaacgaactggcgaaaccgggcgcgagcatgaaaatgagc
tgccgcgcgagcggctatagctttaccagctattggattcattggctgaaacagcgcccggatcagggcctggaa
tggattggctatattgatccggcgaccgcgtataccgaaagcaaccagaaatttaaagataaagcgattctgacc
gcggatcgcagcagcaacaccgcgtttatgtatctgaacagcctgaccagcgaagatagcgcggtgtattattgc
gcgcgcgaaagcccgcgcctgcgccgcggcatttattattatgcgatggattattggggccaggcaccagcgtg
attgtgagcagc**aagctt**agcggcagcgcgagcgcgccgaaactggaagaaggcgaatttagcgaagc**acgcgt**g
gatattcagatgacccagaccccgagcagcctgagcgcgagcctgggcgatcgcgtgaccattagctgccgcgcg
agccaggatattagcaactatctgaactggtatcagcagaaaccggatggcaccgtgaaactgctgatttattat
accagccgcctgcatagcggcgtgccgagccgctttagcggcagcggcagcggcaccgattatagcctgaccatt
agcaacctggaacaggaagatattgcgacctatttttgccagcagggcaacaccctgccgccgacctttggcgcg
ggcaccaaactggaactgctgtattttcagagccaccatcaccatcaccattaa**gcggccgc**t

>C2#

actt**ccatgg**cgcaggtgcagctgcagcagagcggcaacgaactggcgaaaccgggcgcgagcatgaaaatgagc
tgccgcgcgagcggctatagctttaccagctattggattcattggctgaaacagcgcccggatcagggcctggaa
tggattggctatattgatccggcgaccgcgtataccgaaagcaaccagaaatttaaagataaagcgattctgacc
gcggatcgcagcagcaacaccgcgtttatgtatctgaacagcctgaccagcgaagatagcgcggtgtattattgc
gcgcgcgaaagcccgcgcctgcgccgcggcatttattattatgcgatggattattggggccagggcaccagcgtg
attgtgagcagc**aagctt**agcggcagcgcgagcgcgccgaaactggaagaaggcgaatttagcgaagc**acgcgt**g
gatctggtgctgacccagagcccggcgaccctgagcgtgaccccgggcgatagcgtgagctttagctgccgcgcg
agccagagcattagcaacaacctgcattggtatcagcagcgcacccatgaaagcccgcgcctgctgattaaatat
gcgagccagagcattagcggcattccgagccgctttagcggcagcggcagcggcaccgatttaccctgagcatt
agcagcgtggaaaccgaagattttggcatgtatttttgccagcagagcaaccgctggccgctgacctttggcgcg
ggcaccaaactggaactgctgtattttcagagccaccatcaccatcaccattaa**gcggccgc**t

>C3#

actt**ccatgg**cgcaggtgcagctgcagcagagcggcaacgaactggcgaaaccgggcgcgagcatgaaaatgagc
tgccgcgcgagcggctatagctttaccagctattggattcattggctgaaacagcgcccggatcagggcctggaa
tggattggctatattgatccggcgaccgcgtataccgaaagcaaccagaaatttaaagataaagcgattctgacc
gcggatcgcagcagcaacaccgcgtttatgtatctgaacagcctgaccagcgaagatagcgcggtgtattattgc
gcgcgcgaaagcccgcgcctgcgccgcggcatttattattatgcgatggattattggggccagggcaccagcgtg
attgtgagcagc**aagctt**gcgccgcaggcgaaaagcagcggcagcggcagcgaaagcaaagtggatgc**acgcgt**g
gatattcagatgacccagaccccgagcagcctgagcgcgagcctgggcgatcgcgtgaccattagctgccgcgcg
agccaggatattagcaactatctgaactggtatcagcagaaaccggatggcaccgtgaaactgctgatttattat
accagccgcctgcatagcggcgtgccgagccgctttagcggcagcggcagcggcaccgattatagcctgaccatt
agcaacctggaacaggaagatattgcgacctatttttgccagcagggcaacaccctgccgccgacctttggcgcg
ggcaccaaactggaactgctgtattttcagagccaccatcaccatcaccattaa**gcggccgc**t

>C4#

actt**ccatgg**cgcaggtgcagctgcagcagagcggcaacgaactggcgaaaccgggcgcgagcatgaaaatgagc
tgccgcgcgagcggctatagctttaccagctattggattcattggctgaaacagcgcccggatcagggcctggaa
tggattggctatattgatccggcgaccgcgtataccgaaagcaaccagaaatttaaagataaagcgattctgacc
gcggatcgcagcagcaacaccgcgtttatgtatctgaacagcctgaccagcgaagatagcgcggtgtattattgc
gcgcgcgaaagcccgcgcctgcgccgcggcatttattattatgcgatggattattggggccagggcaccagcgtg
attgtgagcagc**aagctt**gcgccgcaggcgaaaagcagcggcagcggcagcgaaagcaaagtggatgc**acgcgt**g
gatctggtgctgacccagagcccggcgaccctgagcgtgaccccgggcgatagcgtgagctttagctgccgcgcg
agccagagcattagcaacaacctgcattggtatcagcagcgcacccatgaaagcccgcgcctgctgattaaatat
gcgagccagagcattagcggcattccgagccgctttagcggcagcggcagcggcaccgattttaccctgagcatt
agcagcgtggaaaccgaagattttggcatgtatttttgccagcagagcaaccgctggccgctgacctttggcgcg
ggcaccaaactggaactgctgtattttcagagccaccatcaccatcaccattaa**gcggccgc**t

>C1*

g**ccatgg**cccaggtgcagctgcagcagagcggcgcggaactggcgaaaccgggcgcgagcatgaaaatgagctgc
cgcgcgagcggctatagctttaccagctattggattcattggctgaaacagcgcccggatcagggcctggaatgg
attggctatattgatccggcgaccgcgtataccgaaagcaaccagaaatttaaagataaagcgattctgaccgcg
gatcgcagcagcaacaccgcgtttatgtatctgaacagcctgaccagcgaagatagcgcggtgtattattgcgcg
cgcgaaagcccgcgcctgcgccgcggcatttattattatgcgatggattattggggccagggcaccaccgtgacc
gtgagcagc**aagctt**tcagggagtgcatccgccccaaaacttgaagaaggtgaattttcagaagc**acgcgt**agac
atccagatgacccagaccccgtcttctctgtctgcttctctgggtgaccgtgttaccatctcttgccgtgcttct
caggacatctctaactacctgaactggtaccagcagaaaccggacggtaccgttaaactgctgatctactacacc
tctcgtctgcactctggtgttccgtctcgtttctctggttctggttctggtaccgactactctctgaccatctct
aacctggaacaggaagacatcgctacctacttctgccagcagggtaacaccctgccgccgaccttcggtgctggt
accaaactggaactgaaataa**gcggccgc**t

Appendix G: Primers

Primers used for Construct 1# amplification:

nVh_FP: `5′- atatacttccatggcgcaggtg -3′` Tm= **67.9** °C, GC= 50 % (22nt/18nt)

nVl_RP: `5′- atatagcggccgcttaatggtg -3′` Tm= 68.6 °C, GC= 50 % (22nt/18nt)

Primers used for Linker amplification:

nLc_FP: `5′- atatcaagcttgcgccgcag-3′` Tm= 68.9 °C, GC= 55 % (20nt/16nt)

nLc_RP: `5′- atatatccacgcgtgcatccac-3′` Tm= **68.9** °C, GC= 50 % (22nt/18nt)

Primers used for Vlo amplification:

nVlo_FP: `5′- atatgcacgcgtggatctgg -3′` Tm= 66.2 °C, GC= 55 % (20nt/16nt)

Primers used to make Construct 2#: Con1 → nVlo_FP + Vl_RP

nVlo_FP `5′- atatgcacgcgtggatctgg -3′` Tm= 66.2 °C, GC= 55 % (20nt/16nt)

nVl_RP `5′- atatagcggccgcttaatgatg -3′` Tm= 68.6 °C, GC= 56 % (22nt/18nt)

Primers used to make Construct 4#: Con1 → nLc_FP + nVl_RP

nLc_FP `5′- atatcaagcttgcgccgcag-3′` Tm= 68.9 °C, GC= 55 % (20nt/16nt)

nVl_RP `5′- atatagcggccgcttaatgatg -3′` Tm= 68.6 °C, GC= 56 % (22nt/18nt)

Primers used to make Construct 3# Con4 → nVh_FP + nLc_RP

nVh_FP `5′- atatcaagcttgcgccgcag-3′` Tm= 68.9 °C, GC= 55 % (20nt/16nt)

nLc_RP `5′- atatatccacgcgtgcatccac-3′` Tm= 68.9 °C, GC= 50 % (22nt/18nt)

Appendix H: Amino acid sequences

>Con1#
MAQVQLQQSGNELAKPGASMKMSCRASGYSFTSYWIHWLKQRPDQGLEWIGYIDPATAYTES
NQKFKDKAILTADRSSNTAFMYLNSLTSEDSAVYYCARESPRLRRGIYYYAMDYWGQGTSVI
VSSKLSGSASAPKLEEGEFSEARVDIQMTQTPSSLSASLGDRVTISCRASQDISNYLNWYQQ
KPDGTVKLLIYYTSRLHSGVPSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPPTFGA
GTKLELLYFQSHHHHHH-AAA

>Con2#
MAQVQLQQSGNELAKPGASMKMSCRASGYSFTSYWIHWLKQRPDQGLEWIGYIDPATAYTES
NQKFKDKAILTADRSSNTAFMYLNSLTSEDSAVYYCARESPRLRRGIYYYAMDYWGQGTSVI
VSSKLSGSASAPKLEEGEFSEARVDLVLTQSPATLSVTPGDSVSFSCRASQSISNNLHWYQQ
RTHESPRLLIKYASQSISGIPSRFSGSGSGTDFTLSISSVETEDFGMYFCQQSNRWPLTFGA
GTKLELLYFQSHHHHHH-AAA

>Con3#
MAQVQLQQSGNELAKPGASMKMSCRASGYSFTSYWIHWLKQRPDQGLEWIGYIDPATAYTES
NQKFKDKAILTADRSSNTAFMYLNSLTSEDSAVYYCARESPRLRRGIYYYAMDYWGQGTSVI
VSSKLAPQAKSSGSGSESKVDARVDIQMTQTPSSLSASLGDRVTISCRASQDISNYLNWYQQ
KPDGTVKLLIYYTSRLHSGVPSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPPTFGA
GTKLELLYFQSHHHHHH-AAA

>Con4#
MAQVQLQQSGNELAKPGASMKMSCRASGYSFTSYWIHWLKQRPDQGLEWIGYIDPATAYTES
NQKFKDKAILTADRSSNTAFMYLNSLTSEDSAVYYCARESPRLRRGIYYYAMDYWGQGTSVI
VSSKLAPQAKSSGSGSESKVDARVDLVLTQSPATLSVTPGDSVSFSCRASQSISNNLHWYQQ
RTHESPRLLIKYASQSISGIPSRFSGSGSGTDFTLSISSVETEDFGMYFCQQSNRWPLTFGA
GTKLELLYFQSHHHHHH-AAA

Appendix I: Crystal screens

Conditions from where crystals were obtained:
0.2 M Calcium acetate hydrate
0.1 M Na CaC pH 6.5 -> MES– Imidazole
18 % PEG 8000

Table I. 1: Optimisation screen

	PEG 8000			**MES-Imidazole**			
	15%	**18%**	20%	pH 6.3	**pH 6.5**	pH 6.7	
1 µL protein + 1 µL well solution							**0.2 M** CaAc
0.5 µL protein + 1.5 µL well solution							
1 µL protein + 1 µL well solution							0.4 M CaAc
0.5 µL protein + 1.5 µL well solution							

Crystallisation condition G4 from Morpheus Screen (Molecular Dimensions):

12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v MPD

0.02 M of each carboxylic acid

0.1 M MES/imidazole pH 6.5

Carboxylic acids

0.2 M sodium formate, 0.2 M ammonium acetate,

0.2 M trisodium citrate, 0.2 M sodium potassium

l-tartrate, 0.2 M sodium oxamate

Appendix J: Vector

Vector created by Geir Åge Løset (Centre for Immu

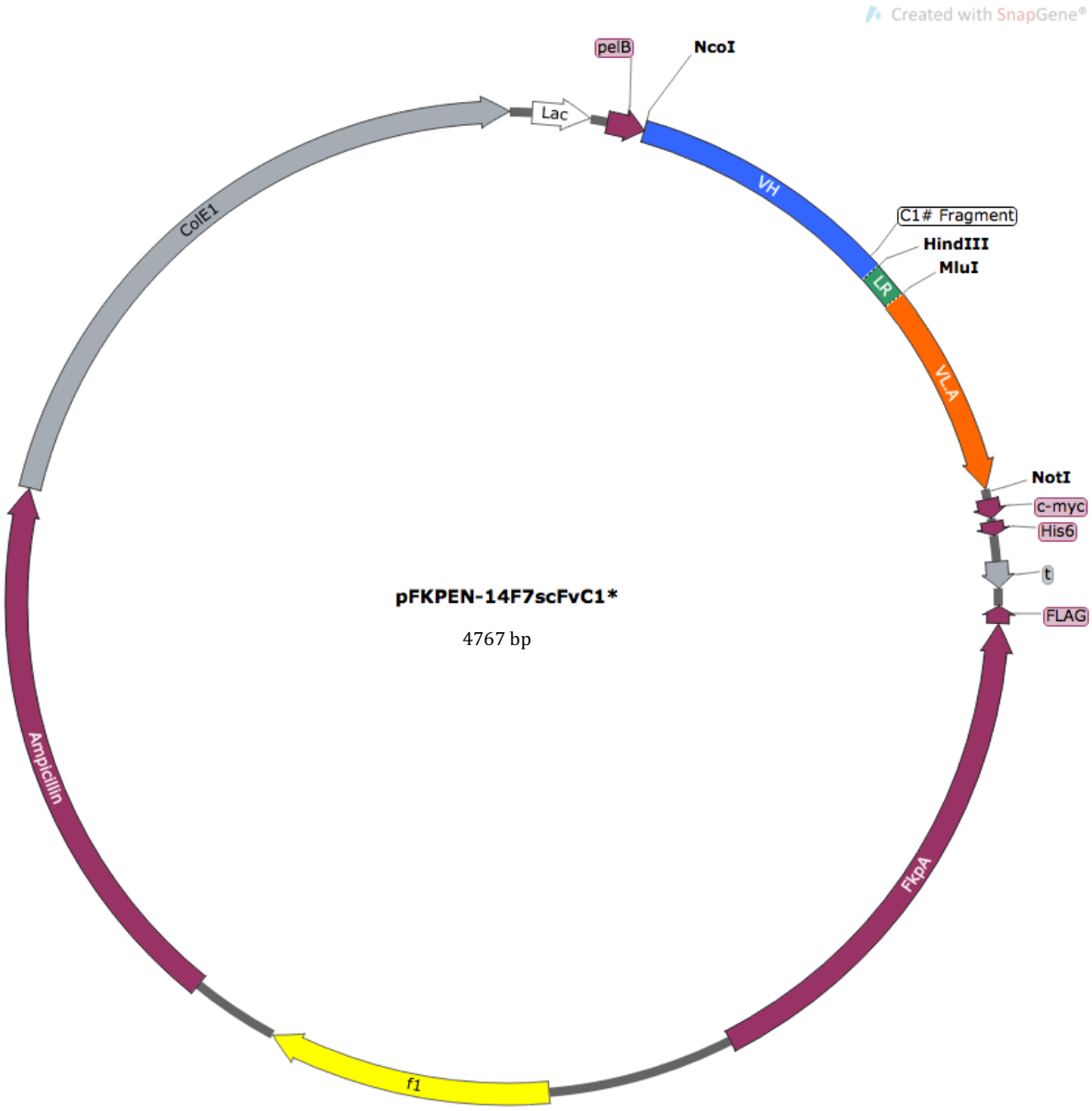


Figure J. 1: The vector pFKPEN-14F7scFvC1*