The Assassin

Recombinant single-chain Fv production of the anti-tumour antibody 14F7

Hedda Johannesen

Master Thesis in Biochemistry
Department of Biosciences
Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

01.06.14

Title: The assassin

http://www.duo.uio.no/

Print: Reprosentralen, University of Oslo

Acknowledgements

Six years of education at the University of Oslo are coming to an end. I started on a bachelor to become a science teacher, but decided to continue with a master. After finishing a bachelor in biochemistry and a year of pedagogics, I was ready to work more independently in the lab. A variety of projects were available, and after visiting over ten research groups, I chose to work with Ute Krengel's group. The group was well structured with monthly literature seminars, had regular meeting times, and provided a very interesting assignment, packed with biochemistry and immunology.

First and foremost, I would like to express my gratitude to my main supervisor Ute Krengel[1]. You always had your door open, and I really liked the coffee breaks where we could talk more informal. By being genuinely interested in both my academic education and the project, you were always updated and gave me much appreciated feedback. I would also like to thank my co-supervisor Geir Åge Løset[2]. You have a brilliant mind. Every time we meet, I am filled with new ideas. By showing such an enthusiasm every time we met, and asking difficult questions where it was needed, you insipid me to push it a bit further. Geir Åge Løset also synthesised the synthetic genes, provided me with the pFKPEN, and the idea of using Protein L to purify and detect the scFvs.

In Ute Krengel's group every master student is assigned to a lab-supervisor for the everyday work. When starting with my thesis, I was assigned to Paula Bousquet[3], since she was working with 14F7 mAb. Paula: you have taught me much about working in the lab. Thank you so much for being there in the morning, answering questions, supervising, eating lunch with me every day and being my friend. In the end of my thesis, Paula was done working in the laboratory and left. In the beginning it felt a bit frustrating, eating lunch alone. But after a while I appreciated the opportunity to become more independent. The unfortunate guy who stepped in for Paula was Daniel Burschowsky[4]. Dani: Thank you so much for investing the time to get familiar with an entirely new project, for your patience, good humour and most important: helping me to write this thing.

---

[1] Professor, Department of Chemistry, University of Oslo, Norway
[2] Scientist, Centre for Immune Regulation and Department of Biosciences, University of Oslo, Norway
[3] M.Sc., Department of Chemistry, University of Oslo, Norway
[4] Doctor, Department of Chemistry, University of Oslo, Norway

I would also like to thank the other contributors for making this thesis possible: Gertrudis Rojas[5], for providing me with 3Fm and the original 14F7 scFv sequence, as well as 14F7 mAb protein. Lena Støkken[6], for basic training in cloning technique and periplasmic *E coli* expression. Ernesto Moreno[7], for providing me with the NeuGc GM3 ganglioside. Stefan Oscarson[8], for the kind gift of Tricer synthesised by Fana Abraha[9]. Rune Johansen, Forstrøm[10], for teaching me how to use the SPR machine and for kinetic evaluation of the SPR results. Gabriele Cordara[11], for teaching me and helping me interpret the ThermoFluor results. Preben Morth[12], for using the high-pressure homogeniser at his lab.

Special thanks go to all the members of Ute Krengel's group. Thank you guys for making me feel like a part of the group and for all the fun. I especially want to thank Julie helping me decipher the differences between British and American writing, and Øyvind for sharing my frustration when writing the thesis.

Thank you Ellen, Andreas, Lena and Marte, I would not have survived the bachelor without you guys. Especially you, Victoria Teigland Holck, for being my lab and study partner throughout the bachelor. At the very end, I would like to thank my parents and Knut Magnus Solbakken, for moral support and providing me with a hug when I needed it.

UiO, May 2014

Hedda Johannesen

---

[5] Doctor, Centre of Molecular Immunology, Havana, Cuba
[6] M.Sc., Centre for Immune Regulation and Department of Biosciences, University of Oslo, Norway
[7] Senior scientist, Centre of Molecular Immunology, Havana, Cuba
[8] Professor, Centre for Synthesis and Chemical Biology, UCD School of Chemistry and Chemical Biology University College Dublin, Dublin
[9] Doctor, Centre for Synthesis and Chemical Biology, UCD School of Chemistry and Chemical Biology University College Dublin, Dublin
[10] Engineer, Oslo university Hospital, Oslo, Norway
[11] Doctor, Department of Chemistry, University of Oslo, Norway
[12] Doctor, Centre for Molecular Medicine Norway

Sammendrag

Kreft er en sykdom som oppstår når en mutert celle vokser og deler seg, uavhengig av plassrestriksjoner eller tilgang på næring. Dette lager et indre press på andre organer i kroppen, hvis det ikke blir behandlet. I dag har vi ingen behandlingsform som dreper kreft uten at vi selv blir syke, derfor er det viktig å utarbeide nye kreftmedisiner for å fjerne kreftcellene mer effektivt I tillegg til effektiv celledeling, produserer kreftceller molekyler av andre sammensetninger og konsentrasjoner sammenlignet med friske celler. Denne kunnskapen utnyttes i immunterapi, hvor kroppens eget immunforsvar blir stimulert til å angripe kreftcellene. Et molekyl som kun eksisterer på celleoverflaten til kreft hos voksene mennesker, er gangliosidet NeuGc GM3. Dette gangliosidet blir gjenkjent av et antistoff med navnet 14F7, med kallenavnet Snikmorderen fordi det kun dreper kreftceller. Når 14F7 binder til gangliosidet, starter cellen å svulme, sprekke opp, og dø. Dette gjør 14F7 spesielt attraktiv for fremtidig kreftbehandling, siden den spesifikt dreper kreft celler. For å videreutvikle 14F7 og øke kunnskapen for hvordan antistoffer gjenkjenner karbohydratbaserte molekyler, må vi først kartlegges hvordan 14F7 interagerer med NeuGc GM3 gangliosidet. Strukturen til 14F7s fragment, antigen-bindende (Fab) ble kjent i 2004 ved hjelp av røntgenkrystallografi. På tross av iherdig innsats var det ikke mulig å reprodusere Fab krystallen, derfor finnes det i dag ikke et eksperimentelt svar på hvor og hvordan gangliosidet assosieres med 14F7. For å løse dette mysteriet prøves en ny vinkling, hvor 14F7 single-chain variable fragmenter (scFvs) blir uttrykt i periplasma til *E. coli*. scFv er det minste fragmentet som fortsatt inneholder det komplette bindingssete til gangliosidet. Siden scFv er mindre komplekst, sammenlignet med Fab og det komplette antistoffet, er det en lovende kandidat for å karakterisere interaksjonene mellom 14F7 og NeuGc GM3 gangliosidet ved bruk av røntgenkrystallografi. Fire forskjellige varianter av scFv 14F7 ble klonet og transformert inn i *E. coli*, hvor de ble isolert fra periplasma og renset med affinitetskromatografi. Bindingen til gangliosidet ble testet, siden konstruktene ikke gjenkjente gangliosidet ble scFv konstruktene optimalisert. Resultatet ble funksjonelle scFv proteiner med affinitet for NeuGc GM3. 14F7 scFv proteinene ble brukt til å utføre ELISA, ThermoFluor, SPR og krystalliseringsstudier. Vi klarte å gro krystaller, men fikk ikke brukbare diffraksjonsdata. Systemet trenger derfor å bli optimalisert videre. Denne masteroppgaven danner et grunnlag for videre arbeid både for krystalliseringsstudier, men også for hvordan ulike scFv design påvirker proteinets stabilitet og uttrykking i prokaryote celler.

Abstract

Cancer is a disease that can develop from any cell within our body. Today, there is no commensally available cancer medicine that kills the cancer effectively, without harming healthy human cells. When a healthy cell mutates into a cancerous cell, it exhibits uncontrolled proliferation and expresses molecules in a different ratio. This information can be used in cancer immunotherapy for diagnostics and therapeutic purposes. Immunotherapy exploits the tumour's unique cytogenetics and stimulates the immune system to destroy it by targeting biomarkers produced by the tumour. One such marker is the *N*-glycolyl GM3 (NeuGc GM3) ganglioside located on the cell surface. The ganglioside is recognised by a monoclonal antibody (mAb) named 14F7, nicknamed The Assassin, because it specifically recognises and kills tumours. When the antibody binds, it results in cell-swelling and lesion formation, thereby killing the cell. To fully exploit the favourable properties of 14F7, it is important to know exactly how the antibody interacts with the NeuGc GM3 ganglioside. This information can be used to produce the next-generation of 14F7 and give general insight into antibody-ganglioside recognition.

The structure of the 14F7 fragment, antigen-binding (Fab) was solved by X-ray crystallography in 2004, and the ganglioside binding was predicted in a docking model. Despite significant effort, it was not possible to reproduce the Fab crystals. Consequently, the mystery of how 14F7 binds to the NeuGc GM3 ganglioside is still not solved experimentally. Therefore, a new approach is tested using the single-chain Fv (scFv), the smallest antibody fragment containing the whole binding site. Because a scFv is less complex, and easier to obtain in larger amounts, the scFv is an excellent candidate for studying the binding characteristics by X-ray crystallography. Four different scFv versions of the 14F7 were cloned and transformed into *E. coli*. They were isolated from the periplasm and purified using affinity chromatography. The binding activity to the NeuGc GM3 ganglioside was tested by enzyme-linked immunosorbent assay (ELISA). Since no activity was detected, the constructs were optimised. The new scFv constructs displayed affinity towards NeuGc GM3 ganglioside and were subjected to crystallographic studies, SPR and ThermoFluor. We were able to obtain crystals, but since they did not diffract, further optimisation is needed. To date, we have achieved to produce three functional 14F7 scFvs versions that crystallise. These can be used as tools for further study, both with respect to crystallographic studies and to obtain insight into how scFv designs affect expression yields and molecular stability.

Abbreviations

ANS	8-Anilino-1-naphtalalenesuffonic acid
AP	alkaline phosphate
bp	base pair
BSA	bovine serum albumin
C-terminus	carboxy-terminus
CIM	Centre of Molecular Immunology, Havana, Cuba
CIR	Centre for Immune Regulation and Department of Biosciences, University of Oslo, Norway
CAP	cAMP receptor protein
CD	circular dichroism
CDR	complementarity-determining region
C_H	constant heavy chain
CIP	alkaline Phosphatase, calf Intestinal
C_L	constant light chain
cmah	CMP-*N*-acetylneuraminic acid hydroxylase
Da	Dalton
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
EDTA	ethylenediamine-tetraacetate
ELISA	enzyme-linked immunosorbent assay
ESRF	European Synchrotron and Radiation Facility
Fab	fragment, antigen-binding

Fc	fragment, crystallisable
FcRn	neonatal Fc receptor
Fv	fragments of the variable domains
HAMA	human anti-mouse antibody
His-tag	Polyhistidine-tag (6x)
HRP	horseradish peroxidase
IEC	ion-exchange chromatography
Ig	immunoglobulin
IMAC	immobilised metal affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
K_D	equilibrium dissociation constant
Lac	lactose
LB	lysogeny broth
LB-A	LB containing 100 mg/l ampicillin.
LB-AG	LB containing 100 mg/l ampicillin and 0.1 M glucose
L_C	elongated Cuba linker
$L_{C.o}$	original Cuba linker
L_R	Rikshospital linker
mAb	monoclonal antibody
mAU	Milli absorption unit
MES	2-(*N*-monopholino)ethanesulfonate
MOPS	3-(*N*-morfolino)propanesulfonate
MS	mass spectrometry
MQ-H_2O	Milli-Q filtered and ion-exchanged water
m/z	mass to-charge ratio

N-terminus	amino-terminus
NEB	New England Biolabs
NeuAc	*N*-acetylneuraminic acid
NeuGc	*N*-glycolyl neuraminic acid
NMR	nuclear magnetic resonance
ON	over night
P20	polyoxyethylene sorbitan
PCR	polymerase chain reaction
pNPP	*p*-nitrophenylphosphate substrate
PBS	phosphate buffered saline
PBS-T	PBS containing 0.1% Tween 20
PBS-TM	PBS containing 0.1% Tween 20 and 5% skimmed milk
PEG	polyethylene glycol
RNase A	ribonuclease
RS	recognition site
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	3,3′,5,5′-Tetramethylbenzidine
V_H	variable heavy chain
V_L	variable light chain
$V_{L.A}$	alternative variable light chain

YT	yeast extract + tryptone
YT-A	YT medium containing 100 mg/l ampicillin.
YT-AG	YT medium containing 100 mg/l ampicillin and 0.1 M glucose.

Table of Contents

1 Introduction 1
1.1 Cancer disease 1
1.1.1 Cancer treatment 1
1.2 Gangliosides as tumour antigens 4
1.3 Anti-tumour Antibodies 6
1.3.1 The immune system 6
1.3.2 Immunoglobulin subclass G 7
1.3.3 The monoclonal antibody 14F7 10
1.4 Method-related theory 14
1.4.1 Recombinant antibody expression 14
1.4.2 Prokaryote expression of recombinant scFvs 15
2 Aims of the thesis 18
3 Materials & procedures 19
3.1 Cloning 19
3.1.1 Preparing the scFv constructs 19
3.1.2 Oligonucleotide-directed mutagenesis 23
3.2 Expression protocols 24
3.2.1 Growth conditions 24
3.2.2 Isolation of soluble periplasmic proteins 24
3.3 Protein purification 25
3.3.1 Affinity chromatography 25
3.4 Concentration measurements 27
3.4.1 DNA 27
3.4.2 Protein 27
3.5 Electrophoresis 27
3.5.1 Agarose gel 27
3.5.2 SDS-PAGE 29
3.5.3 Coomassie staining 30
3.5.4 Silver staining 30
3.6 Analysis by Mass Spectrometry 30
3.7 ELISA 31

3.7.1 Direct ELISA ... 31
3.7.2 Indirect ELISA ... 31
3.7.3 Detecting protein binding in ELISA ... 32
3.8 Western blotting ... 32
3.8.1 Blotting ... 32
3.8.2 Immunodetection ... 33
3.9 Surface plasmon resonance ... 33
3.9.1 Immobilising NeuGc GM3 on the CM5 chip ... 34
3.9.2 scFv affinity measured by SPR ... 34
3.10 ThermoFluor ... 34
3.11 Crystallisation ... 36
4 Results and Discussion ... 37
4.1 Status at project start ... 37
4.1.1 Synthesising the genes ... 37
4.2 scFv constructs containing a His-tag ... 39
4.2.1 Cloning, subcloning and transformation ... 39
4.2.2 Protein expression and isolation I ... 42
4.2.3 Purification using immobilised metal affinity chromatography ... 43
4.2.4 Estimating the initial scFv constructs by ELISA ... 46
4.3 Removing the His-tag ... 48
4.4 scFv constructs without a His-tag ... 51
4.4.1 Protein expression II ... 51
4.4.2 Purification ... 53
4.5 Additional band at ~14 kDa ... 57
4.6 Western blotting ... 62
4.7 scFv's affinity for NeuGc GM3 ... 65
4.7.1 Testing the scFv-ganglioside affinity with ELISA ... 65
4.7.2 Surface plasmon resonance (SPR) ... 66
4.7.3 ThermoFluor ... 67
4.8 Protein Crystallisation ... 70
4.9 Comparison of the scFv constructs ... 73
5 Methodological considerations and future perspectives ... 75
5.1 Optimising the scFv constructs ... 75

5.2 How the presence of Tricer effects the experiments 76
5.3 Ganglioside affinity estimated by ELISA 76
5.4 General thoughts for future experimental procedures 77
6 Summary and conclusion 79
References 81
7 Appendix 1

1 Introduction

1.1 Cancer disease

Tumour formation involves abnormal cell growth that may either be restricted within its originating tissue (benign tumour), or as malignant tumour, capable of spreading throughout the body and invading other tissues (metastasis). A malignant tumour is often referred to as cancer, but cancer cannot always be classified as tumours, as in the case of early stage blood cancer. Cancer exhibit defects in the signal transduction pathways, causing uncontrolled proliferation. While healthy cells only divide 20 to 60 times in culture before they die, cancer cell are immortal (Voet and Voet, 2011). Several somatic events, or a serial of germ line mutations, can lead to alteration of cancer suppressor genes, oncogenes or microRNA (Croce, 2008, Stoler et al., 1999). Such genetic alterations are often a result of chemical carcinogenesis, radiation or certain viruses (Voet and Voet, 2011). When a healthy cell mutates into a cancerous cell it is not only the growth regulation that is affected. Cancer cells also express a high diversity of molecules in other ratios and compositions compared to healthy cells. Cancer can develop in almost any organ or tissue. The treatments are diverse, depending on the type of tumour and the stage of progression.

1.1.1 Cancer treatment

There are four main methods used today to treat cancer: surgery, radioactive therapy, chemotherapy and immunotherapy. In order to produce a synergetic effect surgery is often combined with one or a subset of different therapeutic methods. The most common method to treat cancer is radioactive therapy, using high-energy radiation to damage the DNA leading to cell death. Radiation damage to the genome is most effective during cell division. Since cancer cells divide more rapidly, radiation damage accumulates in their DNA faster compared to healthy cells, killing the cancer cells. Chemotherapy uses pharmaceutically produced drugs to target molecular processes during cell division, thereby killing the cancer. Since the therapy affects the whole body, all cells are exposed, and especially the fast dividing cells in the digestive tract, bone marrow and hair follicles.

Immunotherapy

Another diagnostic and therapeutic method acknowledged as one of the four major cancer treatments is immunotherapy. Cancer immunotherapy is developing rapidly and has become one of the major fields within cancer treatment, immunotherapy exploits the tumour's unique cytogenetics and directly or indirectly stimulates the immune system to destroy the cancer by targeting biomarkers exclusively produced by cancers (tumour-specific antigens) or molecules found in higher abundance compared to regular cells (tumour-associated antigens). The innate and adaptive immune systems in humans have the ability to recognise and kill cancers (Parish, 2003). Through immunoselective processes, tumours become highly resistant against the host's immune system and the body can no longer fight the tumours on its own (Fenton and Longo, 1995). The heterogeneity of tumours makes it difficult to treat cancer even within the same diagnostic type. Information about general epitopes and treatments targeting tumour antigens are therefore at high demand. Since these molecules are often intracellular, they are not very accessible for the immune system (Fenton and Longo, 1995). Fortunately, some tumour antigens are located on the cell surface, for example the *N*-glycolyl GM3 ganglioside. This tumour-specific antigen is identified in several different types of human tumours and is therefore a prime target for diagnostic and therapeutic treatments (Blanco et al., 2011b, Blanco et al., 2011a, Blanco et al., 2012, Blanco et al., 2013, Scursoni et al., 2011).

Modern immunotherapy includes a diversity of methods. Generally, cancer immunotherapy can be divided into two classes: active and passive immunotherapy. The active form of immunotherapy often targets the cancer indirectly, using an active vaccine, or using adaptive cell transfer, where isolated tumour-reactive lymphocytes are stimulated *ex vivo* before being re-introduced into the cancer-bearing host, giving prolonged immunity (Dudley and Rosenberg, 2003). Passive immunotherapy, of which antibody-based immunotherapy is one example, targets the cancer in a more direct manner. In antibody-mediated immunotherapy, antibodies manufactured outside the body are injected into the patient. The antibodies recognise cancer-associated or tumour-specific antigens, but do not give any prolonged immunity. The natural effector mechanisms of the antibodies used in cancer therapy may trigger the body's own immune system to attack the cancer cells. Another alternative is to couple the antibody to a radioactive or toxic molecule, inducing cell death. Due to very high costs and side effects in the range of mild to severe, combined with the increased awareness

in the difficulties in identifying responders from non-responders, very few antibody-based cancer drugs are approved as first-line treatments. Moreover, despite the fact that many of these antibodies offer significant clinical advantages, their influence is difficult to assess and in most cases, the treatment only slightly affects the overall survival. Thus, advanced cancer treatment today involves a careful stepwise process in which the anti-cancer weapon arsenal is gradually exploited within the course of treatment to prolong the therapeutic window.

Most antibodies approved for cancer treatment today do not target the cancer cells directly, instead they recognise tumour-associated antigens, one example being the antibody bevacizumab (Avastin™), targeting a vascular endothelial growth factor signal. Bevacizumab was approved by the Food and Drug Administration in USA, 2013 for metastatic colorectal cancer treatment in combination with chemotherapy. All antibodies recognising tumour-associated molecules also target the host's healthy cells in some extent. The consequence is both dampening the therapeutic effect due to sink effects, as well as resulting in potentially side effects due to off-target effector functions. Another example is the FDA approved antibody rituximab (MabThera™), a cell-lineage specific antibody targeting the CD20 molecule found on virally all B cells, and thus is used to treat B cell lymphomas. Here, complete compartment depletion resulting in a constitutive immune-compromised state of the patient is a direct result of the intervention. Antibodies recognising cancer-specific antigens have major advantages, as they do no damage the host's healthy cells, but since cancer cells originate from the host itself there are a limited number of unique epitopes available. Therefore, antibodies recognising these tumour-specific antigens are of special importance. One such antibody with a promising potential in therapeutic treatments is 14F7 recognising a tumour-specific antigen, namely the ganglioside NeuGc GM3. Since the ganglioside can be defined as non-self the therapeutic effect will not be reduced by the hosts own immune system, and the negative side effects are reduced, since the antibody can separate between host cells and cancer cells. This gives hope for the production of an effective cancer medicine and optimally a cancer vaccine, quenching the cancer in its early phase.

1.2 Gangliosides as tumour antigens

The molecular composition covering the cell surface is partially tissue specific and can be used to identify cell function. Gangliosides cover the cell surface of all animal cells and have multiple functions. They are important for cell-cell interaction, cell-adhesion, development, differentiation, inflammation, signal transduction, tumour progression and more (Hakomori, 2002, Stults et al., 1989). Gangliosides consist of a lipid and a carbohydrate moiety: a sphingolipid and a carbohydrate containing one or more sialic acid residues. The hydrophilic carbohydrate region is exposed to the outer environment of the cell, while the sphingolipid portion anchors the ganglioside to the membrane. With the negative charge of the sialic acids, the glycosphingolipids contribute to the outer negative charge of the cells.

N-acetylneuraminic acid (NeuAc) and *N*-glycolyl neuraminic acid (NeuGc) are the most common types of sialic acids in vertebrates. While most vertebrates express both NeuAc and NeuGc gangliosides, healthy chicken and human tissues express minimal levels of NeuGc gangliosides (Malykh et al., 2001). The small traces of NeuGc ganglioside found in human tissue might be explained as exogenous incorporation from the diet (Tangvoranuntakul et al., 2003, Bardor et al., 2005). The exception is human foetal tissue and several different tumour tissues where NeuGc gangliosides are localised in large amounts. How these cells acquire the NeuGc gangliosides is still a mystery but this makes NeuGc gangliosides a molecular marker and an excellent target for cancer treatment (Malykh et al., 2001, Marquina et al., 1996). NeuGc differs from NeuAc by an extremely subtle chemical modification, the addition of a single oxygen atom at the C5 position of the NeuGc (see Figure 1). The NeuAc gangliosides can be converted to NeuGc gangliosides by a hydroxylase enzyme present in most vertebrates, including chimpanzees (Muchmore et al., 1998). Healthy human tissues are genetically deficient in producing NeuGc due to a homozygous deletion of 92 base pairs within the active site of the CMP-*N*-acetylneuraminic acid hydroxylase (*cmah*) enzyme (Varki, 2001). This Alu-mediated inactivation happened between 2.5 and 3 million years ago, leading to a frame shift in the in the human *cmah* gene, thereby cancelling NeuGc production in humans (Chou et al., 1998, Irie et al., 1998, Chou et al., 2002). The loss of NeuGc in humans was followed by a correspondent increase of NeuAc ganglioside expression (Varki, 2001). Many pathogens use sialic acids as binding sites, so removal of NeuGc gangliosides from the cell surface has probably played an important role for the evolution of human interacting pathogens. The deletion is detected in all human populations studied to date and

might be due to a negative selection by a lethal pathogen that selectively recognised NeuGc (Chou et al., 1998, Gagneux and Varki, 1999). Today, several antibodies are developed for recognising NeuGc gangliosides as a method to diagnose tumours. One such NeuGc ganglioside is NeuGc GM3. This ganglioside is specifically recognised by a monoclonal antibody (mAb) named 14F7. The antibody is today used in diagnostic and therapeutic methods to fight tumour growth in humans.

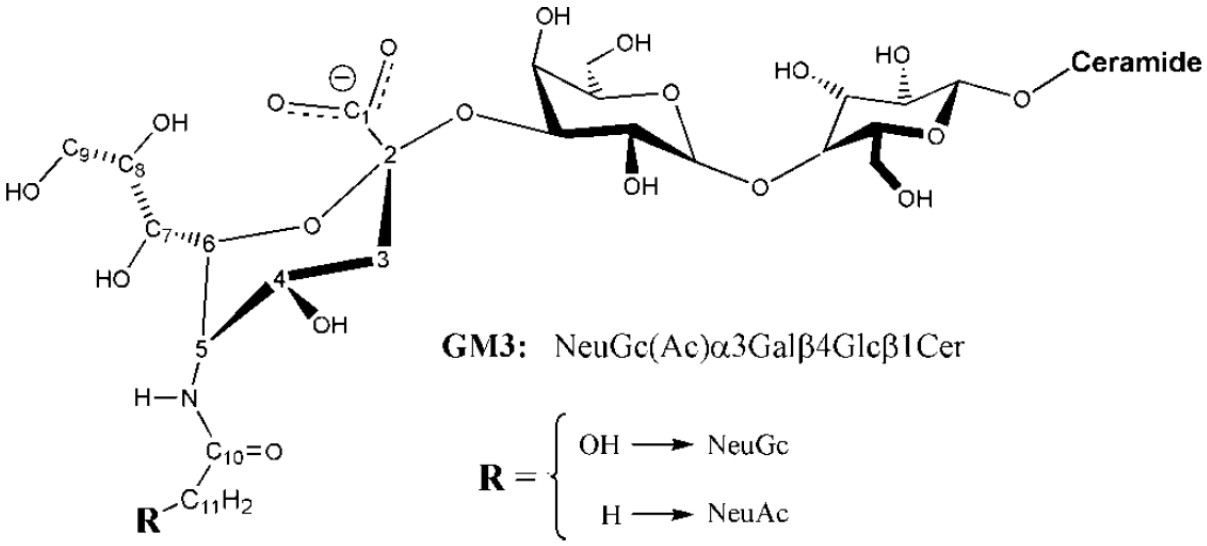

Figure 1: NeuGc(Ac) GM3 ganglioside. The figure was published by Ute Krengel et al. (Krengel et al., 2004) The R-group represents the difference between NeuAc GM3 and NeuGc GM3 gangliosides where NeuGc GM3 gangliosides are produced mainly in human cancer cells, while NeuAc GM3 gangliosides can be found in all human cells.

1.3 Anti-tumour Antibodies

1.3.1 The immune system

The immune system is a collection of defence systems protecting an organism against diseases and harmful transformations. In vertebrates, it can be separated into two main categories. The innate system is the organism's primary defence. It is an assembly of physical barriers as for example skin and mucus membranes, the complement system and a variety of defence cells: macrophages, monocytes, phagocytes, NK cells, mast cells and more. The second category of defence is the adaptive immune system, primarily involving T and B lymphocytes (B cells). There is a delicate interplay and inter-dependency between the innate and adaptive immune system and many immune cells may be assigned to both arms. The adaptive immune system targets pathogens by recognising the molecular differences between self and non-self, making it possible for vertebrates with slow evolutionary development to keep up with the fast-evolving pathogens. The chief molecules responsible are the T-cell receptors produced by the T cells, and antibodies produced by B cells. Every antibody contains a unique non-self paratope that might help the vertebrate defend itself against invading pathogens. Antibodies evolve by somatic rearrangement resulting in a very high diversity of unique binding specificities. The variable light chain DNA undertakes one rearrangement of the variable (V) and joining (J) gene segment, while the variable heavy chain DNA is formed via two rearrangements. First the J-segment and the diverse (D) gene segment are rearranged, then the resulting JD-segment goes through a second rearrangement with the V-segment. In humans, this results in over 1.5×10^7 antibodies with different specificity, and this number does not include the extra diversity obtained by somatic mutations (Nelson, 2008). Thus, the actual diversity of antibody specificities present at a given time point is thought to be in a range exceeding 10^{10}. Native B cells produce membrane-bound antibodies of subclass IgM and IgD. In a later stage, B cells can secrete IgM pentamers known for their high avidity, but low affinity for the antigen (Ehrenstein and Notley, 2010). In order to get a high-affinity antibody, the antibody-producing B cells must undergo somatic hypermutation where new mutations are introduced in the antibody's DNA. This process take place in specialised structures called germinal centres, developing in the secondary lymphoid organs and is dependent on the assistance of T-helper cells and follicular dendritic cells. Somatic hypermutation primarily occurs within three complementarity-determining regions (CDR). The complementarity-determining region number 3 (CDR3) of

the heavy chain has been shown to be of special importance for antigen binding. It consists of the C-terminus part of the J-segment, the N-terminus fragment of the V-segment, together with the whole D-segment. The second round of selection is followed by isotypic switching. The IgM and IgD subclass DNA is deleted, driving the B cell into producing antibodies with the same antigen specificity, but coupled to a different subclass. Depending on the site of the infection, the high-affinity antibodies can be of subclass IgA, IgE or IgG, where IgG is the most abundant in serum, and IgA dominates the mucosal surfaces. After activating an adaptive immune response, some B cells differentiate into memory cells. These cells have a prolonged lifetime, providing the host with immunity against the specific epitope.

1.3.2 Immunoglobulin subclass G

IgG is a heterotetrameric molecule consisting of two identical heavy and light chains shaped like a Y (Figure 2, and a 3 dimensional model in Figure 3). The heavy chain folds into four structural domains: V_H, C_{H1}, C_{H2} and C_{H3}, while the light chain folds into two structural domains: V_L, and C_L, as reviewed by Davies and Chacko (Davies and Chacko, 1993). The constant (C) domains have the characteristic immunoglobulin fold consisting of 3+4 antiparallel β-sheets forming a β-sandwich, while the variable (V) domains consist of an immunoglobulin fold composed of a 4+5 antiparallel β-sheet sandwich. The four chains fold into three structural units, defined by proteolytic cleavage by papain. There are two identical units called the antigen-binding fragments (Fab), each consisting of the C_{H1}, C_L and the V_H, V_L subunits. The two Fab units are both connected to the third unit through a flexible hinge region. The third unit is called the crystallisable fragment (Fc) because it crystallises more easily. Fc consists of two C_{H2} and C_{H3} subunits. The C_{H3}/C_{H3}, C_{H1}/C_L and V_H/V_L subunits are connected with numerous non-covalent hydrophobic interactions, while the two C_{H2} subunits are associated through hydrophilic interactions, due to *N*-glycosylation of Asp297 of both the C_{H2} domains (Jefferis, 1990). Every domain also contains one internal disulphide bridge, increasing its stability. Disulphide bridges also help holding the different chains together. The heavy chains are covalently connected by disulphide bridges localised in the hinge region between C_{H1} and C_{H2}, and the light and heavy chains are connected with a disulphide bridge at the C-terminus of C_{H1} and C_L.

Fabs can be further divided into a constant and a variable fragment. The V_H and V_L are together called the variable fragment (Fv). It contains the complete antigen binding site, primarily localised in six complementarity-determining regions (CDRs), three in each domain (Poljak, 1973). Compared to the Fv, an isolated Fab has a very slow unfolding rate, caused by mutual stabilisation between the variable and constant domains (Röthlisberger et al., 2005). Because of the low stability of the Fv, construct engineering techniques have evolved to improve Fv stability. One such has been the introduction of a soluble and flexible 15-25 amino acid linker, connecting the two variable domains. The resulting fragment was named single-chain Fv fragment (scFv) (Bird, 1988). The scFv construct has revolutionised the crystallisation field because of its small size, and low complexity (Wörn and Plückthun, 2001). The scFv should optimally have the same binding affinity as the whole antibody. However, most often this is not the case, due to the increased entropy-dependent penalty that transpire when the stabilisation provided by the constant domains are removed. Even with a linker connecting the two variable domains, the scFv often remains unstable, and tends denature and aggregate at higher rates compared with Fab and mAb. Therefore, assessment of linker length and sequence, as well as expression and isolation techniques, are important for optimisation of antibody production (Wörn and Plückthun, 2001).

In humans, antibodies of subclass Immunoglobulin (Ig) G are one of the most abundant proteins in the blood serum (Nelson, 2008). IgG antibodies are widely used in diagnostic and therapeutic methods because of their binding specificity, and their advantage of interacting with the complement system, resulting in a strong therapeutic effect. IgG has an especially long half-life in serum due to interaction with the neonatal Fc receptor (FcRn). The intracellular FcRn has in later years been shown to be almost ubiquitously expressed, and is the primary component mediating the exceptionally long serum half-life of the two most abundant serum proteins, namely albumin and IgG. FcRn binds internalised IgG at low pH in the acidified endosomal compartment and re-routs it back to the cell surface for release at physiological pH, thus rescuing the cargo from lysosome degradation (Cohen-Solal et al., 2004, Goebl et al., 2008).

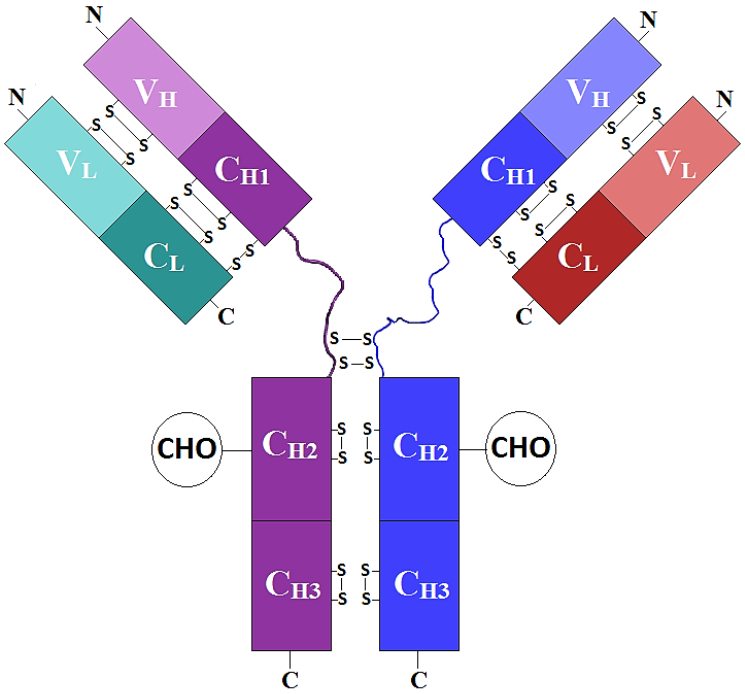

Figure 2: An IgG model. Heavy chains are in blue and violet. Light chains in red and teal green. The constant domains (C) are in a darker colour tone than the variable domains (V). Sulphide bridges are annotated "S-S". The carbohydrate moiety connected to C_{H2} is annotated as "CHO".

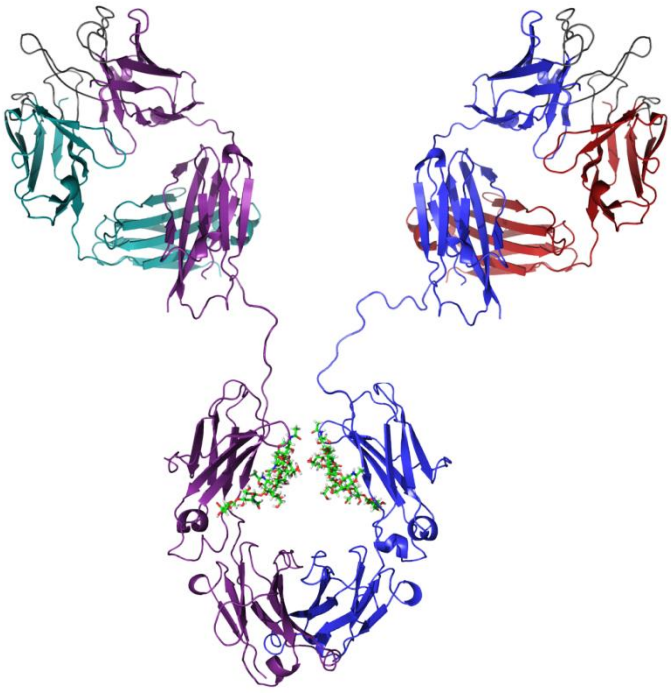

Figure 3: 14F7. The modified structure is adapted from Krengel et al. (Krengel et al., 2004), with heavy chains in blue/violet, light chains in red/teal green and CDRs in grey. The carbohydrate moiety is coloured light green. The Fab and Fc structures were visualised using PyMOL (Schrödinger) (PDB ID 1RIH), and fused MS Paint. The β-sheets are shown as arrows

1.3.3 The monoclonal antibody 14F7

Monoclonal antibodies emerged as diagnostic and therapeutic agents in 1985 when the first murine monoclonal IgG_2 Orthoclone OKT3 antibody was approved for human treatment by the United States Food and Drug Administration (Smith, 1996). Orthoclone OKT3 prevents the human body from rejecting solid organs during clinical transplantation. Since it was murine, it was only efficient for 3 weeks before it was inactivated due to the HAMA response. With time, antibodies recognising cancer antigens were discovered, and, in the year 2000, an IgG_1 antibody, recognising the NeuGc GM3 ganglioside, was reported for the first time. The antibody was named 14F7 and was the first antibody of the IgG subclass recognising a ganglioside. (K_D = 25 nM) (Rojas et al., 2004). 14F7 was generated by the Centre of Molecular Immunology in Havana, Cuba, by immunising Balb/c mice with a vaccine containing the NeuGc GM3 ganglioside hydrophobically conjugated to a human very-low-density lipoprotein in the presence of Freund's adjuvant. The Ig was obtained by a mature humoral response resulting in a hypermutated murine monoclonal IgG_1 antibody (Carr A., 2000). A chimeric version of 14F7 was made by cloning the variable domains of the murine 14F7 onto human IgG_1 and κ constant domains (Roque-Navarro et al., 2008). Although the constant part of the chimeric antibody was human, a chimeric antibody can still evoke a HAMA response. Therefore, the non-human epitopes in the variable domains of the chimeric 14F7 were changed, producing a humanised 14F7.

The selective binding of 14F7 is strictly restricted to NeuGc GM3 ganglioside with a binding affinity of K_D = 25 nM, and does not interact with other NeuGc variants or the closely related NeuAc GM3 ganglioside (Carr A., 2000). The specificity of 14F7 mAb is dependent on both the sialic acid linkage and the external position of the *N*-glycolyl neuraminic acid group (Carr A., 2000). 14F7 mAb has been shown to inhibit growth of solid tumours and is an effective killer of cancer cells both *in vitro* and *in vivo* (Carr A., 2000, Carr A., 2002). Initially the antibody was targeted against breast cancer and melanoma, but today a wide range of cancers overexpressing the NeuGc GM3 ganglioside are known (Carr A., 2000). 14F7 mAb was demonstrated to be inefficient in killing cancers by more standard antibody pathways, as for example antibody-dependent cytotoxicity and complement-dependent cytotoxicity. (Carr A., 2002). Instead, 14F7 can induce a non-apoptotic cell death resulting in loss of cytoskeleton integrity, cell swelling, and large membrane lesions (Roque-Navarro et al., 2008). The lesions are larger than the pores formed by complement, perforin or bacterial toxins (Geny and

Popoff, 2006). This mechanism resembles a complement independent oncosis that might be a consequence of ion pump failure (Majno and Joris, 1995). This oncosis-like cell death mechanism mediated by a certain group of antibodies is believed to be related to positive charges in the CDR3 region of the heavy chain (Bhat et al., 1997). The arginine motif described in the next section (Arg98-X-Arg100A-Arg100) matches the postulated description (Rodríguez et al., 2007). It is especially fascinating that isolated Fab fragments lose their cytotoxic function while $F(ab)_2$ are still functionally active, suggesting that the binding sites might cooperate (Roque-Navarro et al., 2008). Today, 14F7 can for example be used in passive immunotherapy combined with chemotherapy or radiotherapy of solid tumours (Carr A., 2002). In 2006, 14F7 was included in a clinical phase (1 trial), where breast cancer patients were treated successfully with one dose (up to 3 mg) of murine 14F7 mAb, and today it has proceeded into phase 2 (Oliva et al., 2006).

Rojas et at. failed to produce the 14F7 scFv in a bacterial expression system both as a soluble molecule and displayed on phage (Rojas et al., 2004). They therefore experimented with variable light chain shuffling, producing five functional 14F7 scFv variants adapted to a prokaryote expression system by use of phage display. The clones contained the original heavy chain, but different light chains, all with retained specificity and affinity towards the NeuGc GM3 ganglioside (Rojas et al., 2004).

14F7 Structure

14F7 is a highly mutated antibody produced by an mature humoral response (Vázquez et al., 1995). The majority of amino acid replacements are located in the V_H region, suggesting that somatic mutation has occurred separately for V_H and V_L (David and Zouali, 1995). The V_L gene is a member of the J558 family. The $V_{L\kappa}$ belongs to the $V_\kappa 23$ family and is encoded by gene 23-43 (Rodríguez et al., 2007).

When reporting an antibody structure, the Kabat numbering scheme applies for the CDRs. Deletions, insertions and mutations may introduce an unpredictable number of amino acids in these areas, making it difficult to compare two antibody sequences. The solution is to give the amino acids within the CDRs an alphabetic abbreviation. The CDR H3 in 14F7 can be used as an example. The amino acid sequence is N'-Arg98-Leu99-Arg100-Arg100A-Gly100B-Ile100C-Tyr100D-Tyr100E-Tyr100F-Ala100G-Met100H-Asp101-Tyr102-C'. By always

ending the CDR H3 on number 102, the rest of the molecule is still in frame with other antibodies.

Krengel et al. solved the Fab structure in 2004 by X-ray crystallography to 2.5 Å, see Figure 4 (Krengel et al., 2004). The crystal structure was similar to other reported Fab structures with the exception of the CDR 3, of the heavy chain (CDR H3). The binding site was divided into two separated zones by an exceptionally long CDR H3 loop, consisting of 16 amino acid residues. The unusually long CDR H3 segment arose from the fusion of two D minigenes originating from the DSP2 family, giving the segment extra length (Rodríguez et al., 2007). These types of long loops are generally thought to be very flexible, but in the crystal, it was quite rigid. The rigidity is probably partly due to crystallographic stabilisation within the crystal, but a cluster of three tyrosine residues located at the base of the CDR H3 loop (Tyr100D, Tyr100E and Tyr100F) may provide extra stabilisation by interacting with other aromatic residues, giving the CDR H3 loop more rigidity than what is normally observed (Krengel et al., 2004).

By functionally mapping the paratope of 14F7, using tolerated/not-tolerated mutation analysis the results indicate that Trp33 and Tyr100D plays a critical role in recognising the NeuGc GM3 ganglioside (Rojas et al., 2012). Even though Trp33 and Tyr100D are completely buried within the structure, they could only be exchanged with other aromatic amino acid residues to preserve NeuGc GM3 ganglioside binding during the tolerated/not-tolerated mutation analysis. Thus confirming aromatic stacking of amino acid side chains as an important contributor for the stabilisation of the paratope structure (Rojas et al., 2012). The strongest interaction within the CDR H3 loop of the Fab crystal involved three arginine side chains (Arg98, Arg100, and Arg100A) (Krengel et al., 2004). In the mutation analysis, Arg100 and Arg100A were relatively conserved, while Arg98 was highly conserved. Two of the arginines (Arg98, Arg100A) are located in the top part of the CDR H3 loop, creating a small hydrophilic groove, together with amino acids from CDR H1, H2 and one amino acid from the light chain (Trp94) (Rojas et al., 2012).

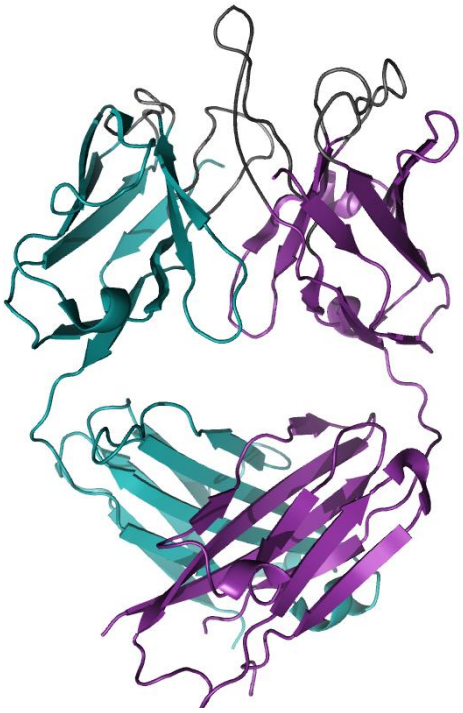

Figure 4: 14F7 Fab structure solved by Krengel et al. (Krengel et al., 2004). The structure is visualised using PyMOL (Schrödinger), PDB code 1RIH. V_L is coloured teal green, V_H is in violet, and the CDRs are grey. The prominent CDR in the middle of the binding site is CDR H3.

14F7 binding model

To simulate the ganglioside binding, Krengel et al. proposed a theoretical model of the 14F7-NeuGc GM3 ganglioside complex (Krengel et al., 2004). Since it was possible to exchange the V_L without losing affinity for the NeuGc GM3 ganglioside, it was assumed that the ganglioside primarily interacted with V_H (Rojas et al., 2004). That is why the terminus disaccharide moiety of NeuGc GM3 (NeuGcα3Galβ) was docked exclusively at the V_H part of the structural model. After docking the disaccharide onto the $V_{H,}$ the third carbohydrate residue (4Glcβ1) was added. There were no major differences between the reported structure and the docking model. It was later discovered by functionally mapping the paratope of 14F7 that the model is at least partly wrong. Residues (V_H: Asp52, Trp33 and Tyr50) identified by the docking model as important for recognising the hydroxyl group of the *N*-glycolyl moiety of the sialic acid was in reality not essential for the affinity (Rojas et al., 2012, Krengel et al., 2004). Since it is this hydroxyl group that separate NeuAc from NeuGc the amino acids recognising this specific hydroxyl group of NeuGc GM3, must be vital for the affinity, therefore it is clear that the ganglioside is docked incorrectly.

1.4 Method-related theory

1.4.1 Recombinant antibody expression

Immunotherapy it is common to produce functional antibodies in non-human organisms, as for example mice. To be able to treat humans with such antibodies, the murine antibodies must be humanised to escape allergic reactions, hypersensitivity and to escape the quenching of a therapeutic effect by human anti-mouse antibodies (HAMA). In order to avoid the cross-species reactivity problem two significant innovations emerged early in the antibody therapy field. The ability to make chimeric antibodies by domain grafting the variable domains onto human heavy chains reduced the xenogeneic sequence proportion to only about 25% of the complete Ab drug. CDR grafting further reduced the xenogeneic sequence proportion to less than 10% of the complete Ab drug. Thus, a significant reduction in anti-drug responses was achieved, making therapeutic Abs a viable intervention option for sustained treatment regimens (Kashmiri et al., 2005).

Antibodies can be isolated from living organisms such as mice, but these can also be produced *in vitro*. For producing antibodies *in vitro*, the antibody-producing B cells must be immortal. A normal cell will only go through a limited number of cell divisions before the cell dies. To make an immortal murine B lymphocyte, the antibody-producing B cell must be fused with a cancerous cell, e.g. from the myeloma cell line producing a murine hybridoma (Kohler and Milstein, 1975). However, there are limitations connected to human antibody production. Generating fully human hybridomas are problematic.

To overcome this challenge, B cells for hybridoma production are isolated from transgenic mice with gene repertoire consisting of human variable domains and murine constant domains producing chimeric antibodies (Brüggemann et al., 1989). If successful in producing a chimeric antibody, a HAMA response might still take place. However, it can be taken a step further, creating a human antibody consisting of murine CDRs. This humanised antibody will have a better chance in avoiding depletion due to anti-drug responses.

Since the Fc region of natural-occurring human antibodies are glycosylated, the therapeutic full-size antibodies are usually expressed in mammalian cells where glycosylation is a part of the eukaryotic posttranslational modification system (Schaefer and Plückthun, 2012). Fab or scFv, on the other hand, are generally not glycosylated and can therefore readily be expressed

in prokaryotic cells, for example in *E. coli* (Schirrmann, 2008). Their small size improves tissue penetration for therapeutic application such as in solid cancer therapy, and provides an excellent opportunity to study the structure of the antibody paratope (Batra et al., 2002, Yokota et al., 1992).

1.4.2 Prokaryote expression of recombinant scFvs

The increased demand for scFvs with regard to quality, stability and amount resulted in the development of a variety of recombinant production methods, vector systems and expression strategies (Schirrmann, 2008).The first functional antibody fragment was produced in *E. coli* in 1988 (Skerra, 1988). Today, there are two main strategies to obtain scFv constructs from *E. coli*. One includes producing the fragments as cytosol inclusion bodies followed by *in vitro* refolding. The disadvantage of this method is that the disulphide bridges can form incorrectly without an *in vivo* folding system (Buchner J., 1992). The other alternative is to utilise the secretion machinery of prokaryotic cells. The reducing environment inside the cytosol of a bacterium greatly sabotages scFv folding, as the conserved disulphide bonds can only form under oxidative conditions (Gilbert, 1990). The disulphide bridges are so important for the stability that only intrinsically very stable scFv constructs are able to fold in the cytosol (Glockshuber, 1992). With the addition of a signal sequence, construction can be directed to the oxidising environment of the periplasm, where the scFv can be isolated in a correctly folded state (Pugsley, 1993, Glockshuber, 1990). The main obstacle for high yield production in a periplasmic expression system is the formation of insoluble aggregates over correctly folded protein, as a consequence of high protein concentration in the periplasmic space (Kipriyanov et al., 1997). But by expressing scFv constructs utilising the secretory pathway of *E. coli*, protein can leak out into the medium, resulting in lower host toxicity, and provide the opportunity for protein isolation directly from the medium (Takkinen et al., 1991).

In *E. coli*, the energetic barrier for RNA polymerase to associate with the promoter is much lower compared to eukaryotic cells. Consequently, most promoters in *E. coli* are regulated by a repressor. By coupling scFv expression to a repressor-regulated promoter, the scFv gene is not transcribed without repressor dissociation. This prevents the *E. coli* secretory pathway from being overloaded with a foreign molecule. Another technique to improve scFv folding is the introduction of periplasmic chaperones, which help to decrease the amount of aggregation inside the cells.

Periplasmic expression

It has been discovered that the amount of scFv aggregates depends on the primary sequence of the variable domain, being directly responsible for successful folding (Knappik and Plückthun, 1995). When using a periplasmic expression system, Kipriyanov et al. found large amounts of insoluble scFv aggregate in the spheroplasts pellet (cells that have had their cell wall removed) (Kipriyanov et al., 1997). The amount of aggregates could be decreased if the incubation was done in test tubes, increasing the periplasm yield. Incubation in Erlenmeyer flasks also increased the yield, but caused a higher release of scFv into the medium, which could be decreased by expression at lower temperatures. In order to utilise the secretion machinery of prokaryotic cells, a signal sequence must be added to the construct. A common method is to add a pelB signal sequence at the amino-terminus (N-terminus) end of the constructs. This will make the scFv products translocate across the inner membrane of *E. coli*, through the secB-dependent pathway, into the oxidising environment of the periplasm, where disulphide bridges can readily form (Sapriel et al., 2002). Another advantage with the secB pathway is that binding of preproteins to SecB enables them to exist in a translocation competent state that will neither aggregate nor fold (Hardy et al., 1993).

The periplasmic expression system used to express the 14F7 scFv construct in this thesis was developed in *E. coli* XL1-Blue cells, and the pFKPEN vector was developed especially for this system (Gunnarsen et al., 2010, Løset et al., 2007). The vector contains the recombinant protein controlled by the *lac* promoter and constitutively expresses the recombinant FkpA, peptidyl-prolyl isomerases, from its native promoter xPO (Gunnarsen et al., 2010, Løset et al., 2007).

Chaperone assisted scFv expression

Chaperones are known to improve expression yield by facilitating folding, reducing protein degradation and preventing aggregation (Ramm and Plückthun, 2000, Ying et al., 2004). The periplasmic peptidyl-propyl cis-trans isomerase FkpA is found naturally in *E. coli* and has a beneficial effect on scFv expression (Ramm and Plückthun, 2000, Ramm and Plückthun, 2001, Bothmann and Plückthun, 2000). By co-expressing FkpA together with the scFv protein, the degree of scFv proteins trapped in the cytosol decreases, instead protein accumulates in the periplasm and a positive growth rate is observed (Gunnarsen et al., 2010).

scFv controlled by the lac promoter

14F7 scFv expression is regulated by the *lac* promoter. When glucose is present, the *lac* repressor will be bound to the *lac* operator, a 24-nucleotide sequence located just downstream of the *lac* promoter (Gilbert, 1973). The repressor prevents the RNA polymerase to transcribe the *lac* promoter-regulated genes, giving *E. coli* a chance to grow without experiencing any potential toxic effect. However, since the *lac* repressor is leaky, a low level of the scFv will probably be expressed at all times. The *lac* repressor disassociates in the presence of lactose. Lactose is transformed into its isomer allolactose in the cell. Allolactose is the inducer of the *lac* promoter: it associates with the *lac* repressor, which in turn dissociates from the DNA, making the promoter available for transcription. Artificial inducers have also been synthesised, with IPTG (Isopropyl β-D-1-thiogalactopyranoside) being the most common for the *lac* promoter. *E. coli* prefers glucose over lactose as an energy source, which is why the *lac* promoter will be downregulated in the presence of a mixture between glucose and lactose. In order to have effective transcription of the genes controlled by the *lac* promoter must not only the *lac* repressor be removed by inducer-association but a co-activator is needed. In the presence of glucose is the co-activator, the cAMP receptor protein (CAP) associated with its own repressor inactivating. When glucose is removed it can associate with the DNA upstream of the promoter. CAP bends the DNA in a conformation that aids RNA polymerase to associate with the Pribnow-Schaller box (TATAAT) upstream of the promoter (Gaston et al., 1990). Kipriyanov et al. (1997) reported that a reduction in the IPTG concentration and the addition of sucrose to the growth medium just before induction of the *lac* promoter, significantly increased the yield of secreted soluble scFv (Kipriyanov et al., 1997). They hypothesised that this was due to an increase in the osmotic pressure, resulting in an enlarging of the periplasmic space based on the findings of Kiefhaber (Kiefhaber, 1991). The positive effect of reducing the IPTG concentration was confirmed by Gunnarsen et al. who removed IPTG completely (Gunnarsen et al., 2010). The addition of sucrose to increase the protein concentration was also thoroughly investigated indicating that sucrose addition did not affect the yield (Gunnarsen et al., 2010).

2 Aims of the thesis

To date it has not been possible to structurally characterise the ligand complex of 14F7 with its tumour antigen the NeuGc GM3 ganglioside, mainly because of difficulties reproducing the 14F7 Fab crystal. In this thesis, the aim was to generate recombinant single-chain Fv of 14F7 and express those for binding analysis and crystallographic studies. We generated scFv constructs, consisting of the two variable antibody domains connected with a linker peptide. Restriction sites flanked the construct and the linker region, making it possible to switch between the two different linker regions and the two variable light chain (V_L) domains, to construct four scFv versions of 14F7. Proteins were expressed in *E. coli* using a periplasmic expression system. Once the proteins were successfully expressed and purified, we attempted co-crystallisation with a hydrophilic synthetically derivative of the NeuGc GM3 ganglioside, named Tricer. We conducted complementary investigations using ELISA, SPR and ThermoFluor and screened for crystals. The experimental part of the thesis can be subdivided into three sections as follows:

1. Cloning, expressing and purifying the scFv constructs

Design expression constructs of recombinant scFv 14F7, clone the constructs, express and purify sufficient amounts of scFv for further analysis.

2. Binding analysis of scFv and the NeuGc GM3 ganglioside

Evaluate the affinity between the different scFvs and its tumour antigen using ELISA, SPR, and ThermoFluor.

3. Crystallisation

Screen for crystallisation conditions of scFv - Tricer complexes, as a first step towards obtaining diffraction data and determine the 3D structure by X-ray crystallography.

3 Materials & procedures

All materials and chemicals are listed in the Appendix, Section A: Materials, page S1.

All solution recipes can be found in the Appendix, Section B: Solutions, buffers and gels, page S9.

All digestion, ligation, PCR primers and corresponding programs are listed in the Appendix, Section C: Restriction- and ligation reagents, page S13, and Section D: PCR mixtures and PCR programs, page S23.

All nucleic acid and amino acid sequences are listed in the Appendix, Section E: Nucleic acid sequences, page S24 and Section F: Amino acid sequences, page S29.

3.1 Cloning

3.1.1 Preparing the scFv constructs

The sequences (V_H-L_R-$V_{L.A}$ and L_C-V_L) were codon-optimised for *E. coli* expression and ordered as synthetic genes (Life technology). A sketch of the constructs is given in Section 4.1.1, Figure 9. The scFv constructs were generated by combining the original variable heavy chain (V_H) with two different murine variable light chains, both from the κ family. The two light chains were either the original 14F7 (V_L) or an alternative variable light chain adapted for prokaryote expression systems ($V_{L.A}$) (Rojas et al., 2004). The V_H is flanked by the restriction sites *Nco*I and *Hind*III (5'-*Nco*I-V_H-*Hind*III-3'), while V_L/$V_{L.A}$ is flanked by *Mlu*I and *Not*I (5'-*Mlu*I-V_L/$V_{L.A}$-*Not*I-3'). To make the scFv constructs, V_H-V_L and V_H-$V_{L.A}$ were combined with two different synthetic linkers. One linker, called the original Cuba linker ($L_{C.o}$), originates from Gertrudis Rojas, Centre of Molecular Immunology, Havana, Cuba (Rojas et al., 2004). It was elongated by introducing restriction sites flanking the sequence (5'-*Hind*III-L_C-*Mlu*I-3'), the new version of the linker was named L_C. The sequence for the second linker, called Rikshospital-linker (L_R), was provided by Geir Åge Løset (Centre for Immune Regulation and Department of Biosciences, University of Oslo, Norway) and already contained the selected restriction sites. The four resulting constructs were named Construct 1, 2, 3 and 4 (C1, C2, C3 and C4). The constructs also contain a 6x Polyhistidine-tag (His-tag), a TEVp recognition site at the N-terminus and a stop codon at

their C-terminus. The soluble expression vector pFKPEN, containing an ampicillin resistance gene and the FkpA gene regulated by its native promoter xPO, was a kind gift from Geir Åge Løset (Gunnarsen et al., 2010). The vector also contains a *lac* promoter, a Shine-Dalgarno sequence and the pelB sequence encoding stretch of the pSEX81 (GenBank accession no. Y14584), upstream of the multiple cloning site containing the *Nco*I and *Not*I restriction sites.

Subcloning *Construct 1* DIGESTION

The vector pFKPEN from Geir Åge Løset (Centre for Immune Regulation and Department of Biosciences, University of Oslo, Norway) and the V_H-L_R-$V_{L.A}$ DNA sequence (Life technology) corresponding to C1 were digested with *Nco*I/*Not*I restriction enzymes (NEB) for 1.5 hour at 37 °C. The DNA was run on a 1% agarose (Lonza) gel at 90 V for 45 minutes together with the two standards ΦX174 (NEB) and λ*Hind*III (Fermentas) and controls where the two DNA sequences were mixed with only one restriction enzyme. The resulting vector ^{NotI}5'-pFKPEN-3'^{NcoI} and insert ^{NcoI}5'-V_H-L_R-$V_{L.A}$-3'^{NotI} were extracted from the gel using the QIAquick gel extraction kit (QIAGEN). In the last DNA purifying step, 30 µl of the elution buffer was used instead of 50 µl to get a higher end concentration. This procedure was initiated every time the gel extraction protocol was used. Concentration was measured at 260 nm using a spectrophotometer (Nanodrop 2000c, Thermo scientific).

Subcloning *Construct 1* LIGATION

Standard protocol: For calculating the correct amount of insert vs. vector for the ligation, the following equations were used with a molar ratio of 4:1.

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

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

It is recommended to have around 10-100 ng vector. 60 ng of the vector was mixed with the calculated amount of insert and incubated over night with T4 polynucleotide DNA ligase (NEB) and T4 ligase buffer at room temperature. Isolated vector without insert was used as a negative control.

Subcloning *Construct 1* TRANSFORMATION

Standard protocol: Half of the ligation mix was mixed with 300 µl $CaCl_2$ competent XL1 Blue *E. coli* cells (recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 *lac* [F' proAB lacIqZDM15 Tn10 (Tetr)] (Stratagene) and left on ice for 45 minutes. After 3 minutes heat shock at 42 °C (Dri-Bloc, Techne), were the cells again put on ice, and 1 ml LB medium was added to each eppendorf-tube. The samples were placed at 37 °C for 30 minutes to give the vector a chance to express antibiotic resistance. The samples were thereafter centrifuged for 4 minutes at 4000 × *g* (Biofuge Fresco Heraeus, Thermo Scientific). ~1.2 ml of the supernatant was removed and the remaining ~100 µl was resuspended and plated on ampicillin containing agar plates overnight at 37 °C.

Two colonies from the transformation were incubated overnight in 6 ml 1x LB containing 100 mg/l ampicillin (AppliChem) and 0.1 M glucose (Sigma) (LB-AG) at 225 rpm (Multitron II, Infors). Vector DNA was isolated using the QIAprep spin Miniprep kit (QIAGEN); the resulting DNA concentration was measured at 260 nm (Nanodrop 2000c, Thermo scientific). The sample was split into three. One part was run on an agarose gel for size control, another was used for sequencing and the remaining part was used to make the other scFv 14F7 constructs. The rest of the bacteria' cultures was used to make glycerol stocks (20% glycerol, Prolabo) and stored at -80 °C.

Sequencing

Standard protocol: The constructs were validated by sequencing: 200 ng DNA mixed separately with either the forward primer pQE-FP (5'-CGGATAACAATTTCACACAG-3') (Metabion) or the revers primer pHogb (5'-CTAGATTAGTGATGGTGATG-3') (Metabion). The primer sequences correspond to the vectors' 5' and 3' sequences flanking the antibody fragment. The primers were used at a concentration of 0.5 µM, and the samples were sent for sequencing at the ABI lab, Department of Biosciences, UiO, Oslo.

Cloning *Construct* 2

The pFKPEN containing C1 and the ordered DNA fragment L_C-V_L (Life Technology) were digested for 1.5 hours at 37 °C with *Not*I/*Mlu*I restriction enzymes (NEB). CIP (NEB) was also added to the C1-containing vector to prevent re-ligation. As a control, the pFKPEN containing C1 was digested separately with *Not*I or *Mlu*I (NEB). The DNA was run on a 1.2%

agarose gel at 90 V for 45 minutes together with the two standards ΦX174 (NEB) and λ*Hind*III (Fermentas). The (^{MluI}5'-V_L-3'^{NotI}) insert fragment and the (^{NotI}5'-pFKPEN-V_H-L_R-3'^{MluI}) vector fragment were extracted from the gel using the QIAquick gel extraction kit (QIAGEN). Concentrations were measured using a spectrophotometer (Nanodrop 2000c, Thermo scientific) at 260 nm. Ligation and transformation were done following the standard protocols.

Two colonies from the transformation plate were incubated overnight with 6 ml LB-AG at 37 °C at 225 rpm (Multitron II, Infors). Vector DNA was isolated using the QIAprep spin Miniprep kit (QIAGEN). The isolated vector was used for sequencing and a sample fraction was run on a 1% agarose gel for size control. The remaining bacterial culture was used to make glycerol (Prolabo) stocks and stored at -80 °C.

Cloning *Construct 4*

The pFKPEN containing C1 and the ordered DNA fragment L_C-V_L (Life Technology) were digested with *Hind*III/*Not*I restriction enzymes (NEB) at 37 °C (Dri-Bloc, Techne) of 1.5 hours. CIP (NEB) was also added to the C1-containing vector to prevent re-ligation. As a control, pFKPEN containing scFv C1 was digested separately with *Hind*III or *Not*I (NEB). The DNA samples were run on a 1.2% agarose gel at 90 V for 45 minutes. The (^{HindIII}5'- L_C-V_L-3'^{NotI}) insert and the (^{NotI}5'-pFKPEN-V_H-3'^{HindIII}) vector fragments were isolated from the gel using the QIAquick gel extraction kit (QIAGEN). Concentrations were measured using a spectrophotometer (Nanodrop 2000c, Thermo scientific) at 260 nm and ligation and transformation were done following the standard protocols.

Two colonies from the transformation plate were incubated overnight with 6 ml LB-AG at 37 °C at 225 rpm (Multitron II, Infors). Vector DNA was isolated using the QIAprep spin Miniprep kit (QIAGEN). The isolated vector was used for sequencing and one sample fraction was run on a 1% agarose gel for size control. The C4-containing vector was also used to make construct C3. The remaining bacterial culture was used to make glycerol (Prolabo) stocks and stored at -80 °C.

Cloning *Construct 3*

The pFKPEN vectors containing C1 and C4 were digested with *Not*I/*Mlu*I restriction enzymes (NEB) for 1.5 hour at 37 °C with controls. Digested C4-containing vector was run on a 1% agarose gel while digested C1 containing vector was run on a 1.2% agarose gel. The resulting vector fragment from C4 (^{NotI}5'-PFKPEN-V_H-L_C-3'^{MluI}) and insert fragment from C1 (^{MluI}5'-$V_{L.A}$-3' ^{NotI}) were isolated from the gel using QIAquick gel extraction kit (QIAGEN). The concentrations were measured at 260 nm using the spectrophotometer (Nanodrop 2000c, Thermo scientific). Insert and vector were mixed in a 4:1 ratio using equations (1) and (2). Ligation mix and controls were made following standard protocol and transformed into XL1 Blue competent cells using standard technique.

Two colonies from the transformation plate were incubated overnight with 6 ml LB-AG at 37 °C at 225 rpm (Multitron II, Infors). Vector DNA was isolated using the QIAprep spin Miniprep kit (QIAGEN). The isolated vector was used for sequencing and a sample fraction was run on a 1% agarose gel for size control. The remaining bacterial culture was used to make glycerol (Prolabo) stocks and stored at -80°C.

3.1.2 Oligonucleotide-directed mutagenesis

Optimised protocol: PCR mediated oligonucleotide-directed mutagenesis was performed for removal of the His-tag and the TEVp recognition site. The forward primer was designed including the Pribnow box, and the 5'-end *Nco*I tag and 20 nucleotides of the V_H region, excluding the 33 nucleotides of the His-tag and the TEVp cleavage site. 20 nucleotides of the V_H region were added to the forward primer (5'-ATATCCATGGCCCAGG TGCAGCTGCAGCAG-3') (Eurofins Genomics). The reverse primer contained the 3'-end *Not*I and 20 eversed nucleotides of the V_L region (5'-TATAGCGGCCGCTTATTTCAGTTC CAGTTTGG-3') (Eurofins Genomics).

The vectors were subjected to PCR using Phusion DNA Polymerase (Thermo Scientific), one of the most accurate thermostable polymerases available today. The resulting DNA fragments were run on a 1% agarose gel at 90 V for 50 minutes and purified using a PCR Purification Kit (250) (QIAGEN). 1000 ng DNA per scFv construct and 2000 ng of the original plasmid pFKPEN were digested with *Nco*I/*Not*I restriction enzymes (NEB). The original vector pFKPEN was also mixed separately with *Nco*I or *Not*I restriction enzymes (NEB) as controls.

All samples were incubated at 37 °C for 1.5 hours together with a mix of 1 kbp and 100 bp ladders (NEB). The DNA was separated according to size on a 1% agarose gel at 90 V for 50 minutes before the relevant bands were isolated and purified using the QIAquick gel extraction kit (QIAGEN). As for the initial cloning protocol, 30 µl of the elution buffer was used during the last step of the extraction instead of 50 µl to get a higher concentration of purified DNA. For calculating the correct amount of insert vs. vector for the ligation, equations (1) and (2) were used with a molar ratio of 4:1.

Ligation and transformation were performed according to standard protocol. Glycerol stocks (20% glycerol, Prolabo) were made from two colonies per construct. Transformed plasmids were purified using the miniprep Nucleo Spin Plasmid kit (Macherey-Nagel GmbH & Co. KG) and sent to sequencing. Sequencing was performed using the standard protocol. The new constructs were named 14F7 scFv: C1*, C2*, C3* and C4*.

3.2 Expression protocols

3.2.1 Growth conditions

Transformed XL1-Blue *E. coli* cells (Stratagene) were grown over night in 100 ml 2x YT medium with 100 mg/l ampicillin (AppliChem) and 20% glucose (Sigma) (2x YT-AG) at 125 rpm, 37 °C (Multitron II, Infors). The next day the main cultures were inoculated to OD_{600} of 0.025 in 1 L 2x YT-AG in 2 litre Erlenmeyer flasks and incubated at 125 rpm, 37 °C (Multitron II, Infors) until the OD_{600} reached 0.6-0.8. The bacteria were pelleted by centrifugation (4000 × *g*, 40 min, 4 °C) (Centrifuge 5810 R, Eppendorf) and resuspended in an equal volume of fresh 2x YT medium containing 100 mg/ml ampicillin (AppliChem) (YT-A). The cultures were left overnight at 125 rpm, 30 °C (Multitron II, Infors).

3.2.2 Isolation of soluble periplasmic proteins

Standard protocol

If the OD_{600} was close to or over 1 for the overnight cultures, the cells were harvested by centrifugation (4000 × *g*, 40 min, 4 °C) (Avanti centrifuge J-26 XP, Beckman Coulter). The pellets were resuspended in ice-cold periplasmic extraction buffer with 80 µg/ml lysozyme

(Sigma) and 80 µg/ml RNase A (Sigma), and incubated for 1 hour at 4 °C with shaking (Multitron II, Infors). The samples were centrifuged (4000 × *g*, 40 min, 4 °C (Centrifuge 5810 R, Eppendorf), giving the soluble periplasmic extracts as the supernatants, and spheroplasts with insoluble periplasmic material as the pellets. The supernatants were filtered through 0.2 µm filter (Rapid Flow, Nalgene) into clean glass bottles and stored at 4 °C. If the storage periods exceeded one week, the soluble periplasmic extracts were frozen to -80 °C.

Cracking the cells using high-pressure homogeniser

If the OD_{600} was close to or over 1 for the overnight cultures, the cells were harvested by centrifugation (4000 × *g*, 40 min, 4 °C) (Avanti centrifuge J-26 XP, Beckman Coulter). The pellets were resuspended in cell cracking buffer, 5-10 ml buffer per 1 g wet pellet. The cells went through two rounds of homogenisation using the high pressure homogeniser (EmulsiFlex-C3, Avestin). The cell lysate was centrifuged at 48000 × *g* for 20 minutes at 4 °C (Centrifuge 5810 R, Eppendorf). The supernatants were filtered through 0.2 µm filter (Rapid Flow, Nalgene) into clean glass bottles and stored at 4 °C.

3.3 Protein purification

3.3.1 Affinity chromatography

Immobilised metal affinity chromatography (IMAC)

The soluble periplasmic extract was adjusted to a buffer composition similar to the binding buffer used in IMAC (Immobilised metal affinity chromatography) (see buffer compositions in the Appendix, Section B: Solutions, buffers and gels, page S1). IMAC was performed at 4 °C using a 5 ml column of chelating Sepharose (GE Healthcare) charged with Ni^{+2}, on the ÄKTApurifier-900 (Biosciences, GE Healthcare). The column was equilibrated with a binding buffer with pH 7.8. The sample was loaded onto the column and washed with 4 column volumes of binding buffer. The bound proteins were eluted with a gradient of 0-100% of elution buffer with pH 7.8, over 6 column volumes. Fractions of 1 ml were collected throughout the procedure and absorbance was recorded at 254 and 280 nm and analysed using the Unicorn 5.11, GE Healthcare software. SDS-PAGE was performed according to standard procedure, see Section 3.4.2, page S27. IMAC fractions with proteins of the correct size

(~30 kDa) were pooled, concentrated and dialysed against 50 mM Tris-HCl (Chalbiocem) at pH 7.5 (Amicon ultra filter, Ultracell-10 K, 15 ml and 2 ml, Millipore). The concentration was measured (A_{280nm}, NanoPhotometer, IMPLEN) and the sample stored at -80 °C. The column was washed with binding buffer and water, and stored in 20% ethanol (Arcus) at 4 °C.

Protein L affinity chromatography

Soluble periplasmic extract was adjusted to a buffer composition similar to the binding buffer used in Protein L affinity chromatography (buffer compositions in the Appendix, Section B: Solutions, buffers and gels, page S9). Affinity chromatography was performed using gravity columns containing 1 ml of Protein L capto resin (GE healthcare). A separate column was used for each construct. The columns were equilibrated with binding buffer. The samples were loaded onto the columns and washed with 20 ml of binding buffer pH 7.8. The proteins were eluted with elution buffer of pH 3 and neutralised to pH 7.8 with 50 mM Tris-HCl (Chalbiocem). SDS-PAGE was performed according to standard protocol (see Section 3.4.2, page 27), and fractions with proteins of the correct size (~28 kDa) were pooled and concentrated (Amicon ultra filter, Ultracell-10 K, 15 ml and 2 ml, Millipore) before purified further by size-exclusion chromatography (SEC). The columns were neutralised with 20 ml 1 M Tris-HCl (Chalbiocem) with pH 9 and stored in 20% ethanol (Arcus) at 4 °C.

Size-exclusion chromatography (SEC)

Analytical gel filtration was performed after affinity chromatography using a Superdex 200 R10/300 (GE healthcare) column installed on the ÄKTApurifier-900 (Biosciences, GE Healthcare) with 50 mM Tris-HCl (Chalbiocem) with pH 7.8 as the running buffer. Sample volume and flow rate were 500 µl and 0.5 ml/min respectively. Fractions of 0.5 ml were collected throughout the procedure and absorbance was recorded at 254 nm and 280 nm (Unicorn 5.11, GE Healthcare). SDS-PAGE was performed according to standard procedure (see Section 3.4.2. page 27), and SEC fraction with proteins of the correct size (~28 kDa) were pooled, concentrated and dialysed against 50 mM Tris-HCl (Chalbiocem) at pH 7.5 (Amicon ultra filter, Ultracell-10 K, 15 ml and 2 ml, Millipore). The concentration was measured (NanoPhotometer, IMPLEN) and the sample stored at -80 °C.

3.4 Concentration measurements

3.4.1 DNA

The DNA concentrations were measured at 260 nm using a spectrophotometer (Nanodrop 2000c, Thermo scientific).

3.4.2 Protein

The protein concentrations were measured at 280 nm using the NanoPhotometer (IMPLEN).

Throughout the thesis the protein concentrations were measured incorrectly. When measuring the concentration the absorbance coefficient for the specific protein must be inserted in the NanoPhotometer program. The NanoPhotometers manual does not specify if it is the direct or the inverted absorbance coefficient that should be used. It was therefore assumed that the direct coefficient could be inserted. But, just before finishing the thesis it was discovered that the machine requires the inverted coefficient to calculate the correct protein concentration. For proteins with a coefficient close to 1, this would not give noteworthy errors. However, since the scFvs has an absorbance coefficient (Abs_{280nm}) of ~1.87, (see scFv info in Table 4) the change in reported protein concentration is significant, as the inverted number is 0.535. By measuring scFv concentrations with the inverted coefficient it was discovered that earlier measurements overestimated the concentrations by a factor of ~3.2.

3.5 Electrophoresis

3.5.1 Agarose gel

The agarose concentration in the gel is directly correlated to the pore size, controlling DNA migration. DNA migrates towards the anode due to its negative charge. Linear double stranded DNA migrates with a velocity that is inversely proportional to the logarithm of the number of base pairs. In practice, this means that small DNA molecules migrate faster through the gel than larger molecules. Circular molecules on the other hand do not follow the same rule. Circular DNA molecules can be packed densely due to supercoiling, making them move faster through the gel than their linear counterpart. To visualise the DNA, a DNA-binding compound is added to the gel that fluoresces when exposed to ultraviolet light. As the

DNA moves through the gel, it will bind the fluorescent compound. The strength of the accumulated fluorescent signal reflects the DNA concentration.

Standard procedure: Agarose (Lonza) was mixed during heating with 1xTAE buffer to a final agarose concentration of 1% or 1.2%, depending on the DNA fragment size. The agarose solution was stored at 37 °C. When casting a gel, 50 ml agarose solution was mixed with 5 µl SYBR safe DNA stain (Invitrogen), to visualise the DNA bands through fluorescence. The gel was set after 20 minutes (Mini Horizontal Submarine Unit), 1xTAE buffer was added until the gel was completely covered. Then the DNA samples were loaded into the wells using ΦX174 (NEB), λ*Hind*III (Fermentas), as molecular weight standards or a mixture of 1kbp (NEB), and 100 bp (NEB) molecular weight ladders, to indicate fragment size (see Figure 5 for molecular weight information) The gels were run at 90 V for 45 minutes (PowerPack HC power supply).

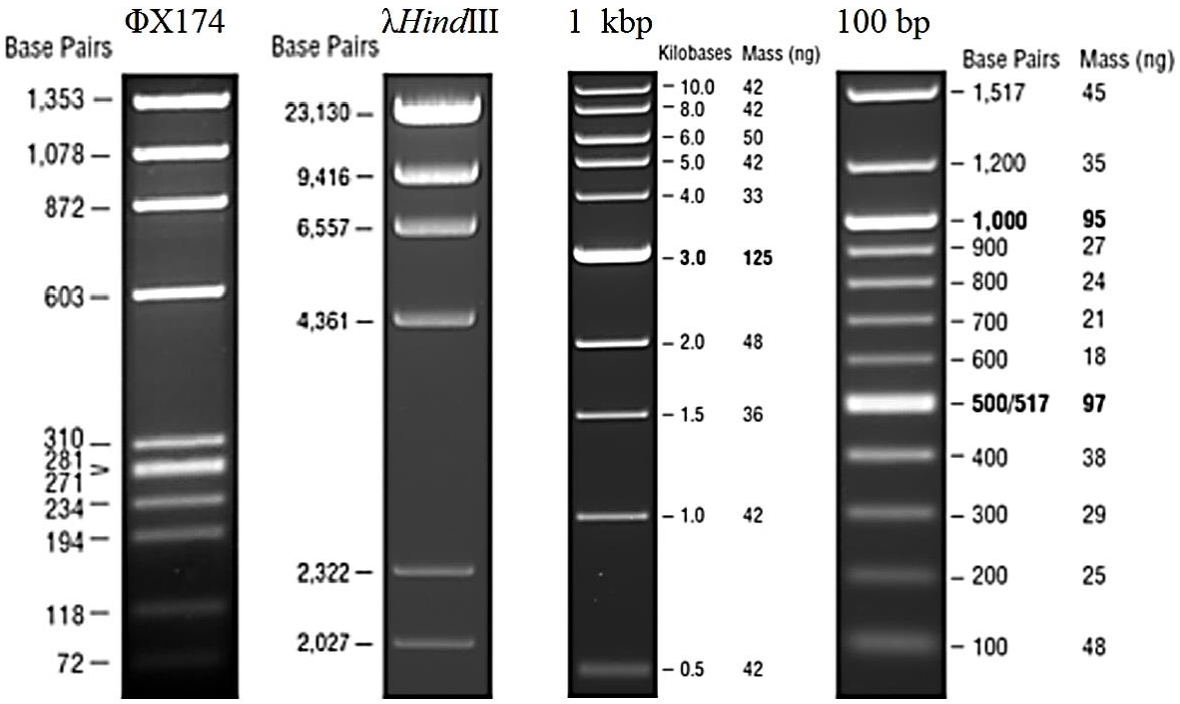

Figure 5: From left: ΦX174 standard on a 1.7% agarose gel, λ*Hind*III standard on a 1.0% agarose gel,1 kbp ladder on a 0.8% agarose gel, and 100 bp ladder on a 1.2% agarose gel.

3.5.2 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate proteins according to their respective sizes. Sodium dodecyl sulphate (SDS) is a denaturing detergent: it binds to the protein with its hydrophobic part in a 1.4:1 ratio and unfolds it giving it a rodlike shape. Its hydrophilic section is negative, masking the proteins' intrinsic charge, giving all proteins the same mass to charge ratio (Voet and Voet, 2011). During electrophoresis, the negative charge will make all proteins move towards the anode, the migration rate will be determined by the gel's pore size and the molecular size of the proteins.

Standard procedure: 7.5 µl of each sample were mixed with 4 µl 4x NuPAGE LDS Sample buffer (Invitrogen,) and 3.5 µl MQ-H_2O. The samples were heated for 10 minutes at 70 °C, spun down (using Capsulefuge, TOMY PMC-060, Tomyteck) and loaded on a NuPAGE Bis-Tris 4-12% gel (Life technology). SeeBlue® Plus2 prestained standard (Invitrogen) was used to get an indication of protein size (see Figure 6 for molecular weight information). The gel was coupled to an Electrophoresis powersupply-EPS601 (GE healthcare) and run for 35 minutes at 200 V in 1x MES SDS Running buffer (Invitrogen).

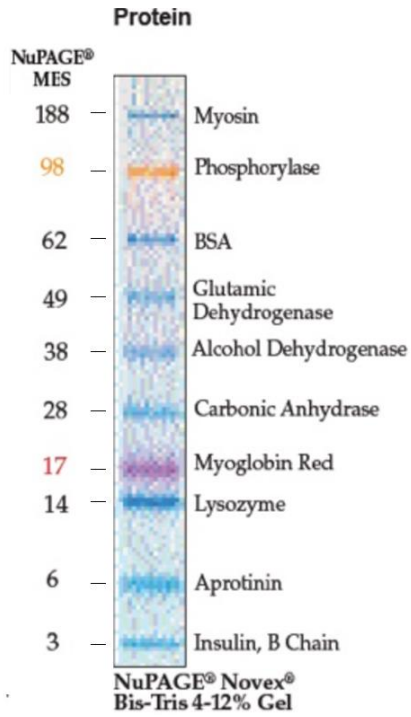

Figure 6: SeeBlue® Plus2 prestained standard (Young-Pearse, 2009)

3.5.3 Coomassie staining

The polyacrylamide gel was stained for one hour in Coomassie staining solution after SDS-PAGE. The gel was thereafter de-stained with MQ-H_2O, until a clear contrast between background colour and protein bands was detected. The gel was stored in MQ-H_2O until a picture was taken.

3.5.4 Silver staining

After SDS-PAGE, the polyacrylamide gel was incubated with shaking for 10 or more minutes in MQ-H_2O water. Thereafter the gel was incubated with shaking in 50 ml fixation mix for 20 minutes, followed by incubation with 50 ml Farmer's Reducer for 2 minutes. The gel was washed multiple times in MQ-H_2O until the yellow colour disappeared. After washing, the gel was silver stained with 0.1% 50 ml silver nitrate solution ($AgNO_3$) (Merck) for 15 minutes with shaking. The gel was then subjected to two rounds of washing: first 30 seconds with 50 ml MQ-H_2O, followed by 30 seconds with 50 ml 2.5% sodium carbonate (Na_2CO_3) (Sigma). The silver staining was developed by placing the gel in 50 ml developing solution until the protein bands were visible. The development was stopped by adding 100 ml of 10% v/v acetic acid (Merck). The gel was incubated with shaking for 10 minutes, followed by two rounds of washing with 10 ml MQ-H_2O. The silver stained gel was stored in MQ-H_2O until a picture was taken.

3.6 Analysis by Mass Spectrometry

Depending on the protein's mass, *m*, and its positive or negative charge, *z*, every protein possesses a *mass/charge*, (*m/z*) ratio. A powerful magnetic field as used in mass spectroscopy (MS) will be able to deflect ionized samples based on this ratio.

IMAC purified C1 from 13.05.2013, was separated according to size on a NuPAGE Bis-Tris 4-12% gel (Life technology). The band corresponding to the scFv size (~30 kDa) was excised and further analysed by mass spectrometry using the Orbitrap-XL (Thermo Fisher).

Protein C1*, C3* and C4* -fractions from SEC were run on a NuPAGE Bis-Tris 4-12 % gel (Life technology). Bands with the sizes of scFvs (~28 kDa) and the bands corresponding to ~14 kDa were excised for all three scFvs and sent to MS analysis, where it was compared

with the different 14F7 scFv, FkpA and lysozyme sequences. The results were returned and analysed using the Scaffold 4 software (Proteomic Software Inc.)

MS analyses were performed by Bernd Thiede's group, at The Biotechnology Centre of Oslo.

3.7 ELISA

Enzyme-linked immunosorbent assays (ELISA) were used to determine the affinity between scFv and its antigen.

3.7.1 Direct ELISA

Proteins diluted in 1x PBS to the desired protein concentrations were incubated overnight in Nunc-Immuno 96 MicroWell MaxiSorp solid plates (Sigma) wells (see Appendix for solution recipes, Section B: Solutions, buffers and gels, page S9). The plates were washed the next day with 1x PBS containing 0.1% Tween 20 (Sigma) (PBS-T) and blocked with 1x PSB-T containing 5% skimmed milk (AppliChem) (PBS-TM) for one hour at room temperature. The binding was developed as described below. Every experiment was duplicated to ensure reproducible results. 14F7 mAb (CIM) was included as a positive control and BSA (bovine serum albumin) (NEB) was used as a negative control.

3.7.2 Indirect ELISA

The Nunc-Immuno 96 MicroWell PolySorp solid plate (Sigma) wells were incubated with 100 µl NeuGc GM3 ganglioside (10 µg/ml) in methanol (Prolabo) overnight, allowing the methanol to evaporate. Plates were washed with 300 µl of PBS-T and blocked with 300 µl 1x PSB-TM for one hour at room temperature. The scFv constructs were diluted in 1x PBS to the desired concentration, and 100 µl solutions were added to the wells and incubated on the plate for 1.5 hours at room temperature. The plate was washed with 300 µl PBS-T and developed following the standard protocol, described below. Every experiment was done in duplicates. 14F7 mAb (CIM) was included as a positive control for NeuGc GM3 ganglioside recognition, and BSA (NEB) was used as a negative control.

3.7.3 Detecting protein binding in ELISA

Anti-His-tag antibody

For scFv containing the His-tag, a murine monoclonal anti-polyhistidine-alkaline phosphatase antibody (Sigma) was utilised. It was diluted in 1x PSB to a 1: 2000 compared to the scFv concentration before adding 100 µl of solution to each well. The plate was incubated for 1 hour at room temperature, and washed with PBS-T before the addition of *p*-nitrophenylphosphate substrate (pNPP) (Sigma). The signal was observed by visual inspection before and after stopping the reaction with 1 M NaOH (Kebo Lab). Every experiment was done in a parallel to ensure a reproducible result. 14F7 mAb (CIMM) was included as a positive control for NeuGc GM3 recognition, and BSA (NEB) was used as a negative control. Since no micro plate reader was available at our laboratory, the plates were frozen at -20 °C for later signal detection. Unfortunately, a side reaction occurs when the plates are frozen and the solutions turn brown, making signal detection unreliable.

Protein L-HRP

To develop scFvs without His-tag, Protein L coupled to horseradish peroxidase (Genscript) (HRP) was utilised (Protein L-HRP). It was diluted with 1x PBS to a 1:1000 ratio compared to the scFv concentration, 100 µl was added to the each well and the plate was incubated for 1 hour at room temperature. The plate was again washed in PBS-T before the addition of the substrate TMB (Chalbiocem) solution. The chromatographic development was observed by visual inspection. The reaction was stopped by adding 100 µl 1 M HCl. Every experiment was duplicated on the same plate to ensure a reproducible result. 14F7 mAb was included as a positive control for NeuGc GM3 recognition, and BSA (NEB) was used as a negative control.

3.8 Western blotting

3.8.1 Blotting

Protein separated on a gel (NuPAGE Bis-Tris 4-12% gel, Life technology) by SDS-PAGE was placed in transfer buffer for 10 minutes. A piece of Immobilon-P membrane (Millipore) corresponding to the size of the gel was dipped in 100% methanol (Prolabo) for 15 seconds, then immediately soaked 5 minutes in MQ-H_2O, followed by 5 minutes in Transfer buffer. An

extra thick paper blot (Bio-Rad Laboratories) was soaked in Transfer buffer and placed on the Trans-Blot® SD Semi-Dry Transfer Cell (Bio-Rad Laboratories) apparatus. Then, the wet Immobilon-P membrane (Millipore) was placed over the wet paper blot, before a new wet paper blot was placed on top. Trapped air was removed and the blotting was run at 25 V for 30 minutes.

3.8.2 Immunodetection

The membrane was incubated in PBS-TM for 1 hour with agitation, followed by 20 seconds in new PBS-TM solution. Then the membrane was incubated for 1 hour in PBS-TM containing 1:2000 Protein L-HRP (Genscript) with agitation. After the incubation time, the membrane was washed twice for 5 minutes in PBS-T and once for 5 minutes in 1x PBS.

Two different substrates were used to detect scFv:

1. Insoluble TMB: The membrane was placed in 15 ml insoluble TMB solution (Chalbiocem) and incubated for 10 minutes with agitation. The TMB solution was removed and the membrane rinsed in MQ-H_2O, before it was dried at room temperature and photographed.
2. SuperSignal West Femto Chemiluminescent Substrate was used combined with Protein L-HRP to detect scFvs (Thermo Scientific). To detect a signal Amersham Hyperfilm ECL (GE healthcare) was used with an X-Ray film developer (Fuji) localised in a dark room.

3.9 Surface plasmon resonance

In Surface plasmon resonance (SPR), a CM5 chip (GE Healthcare) was used to detect scFv binding. The chip contains 4 channels, but only 2 were used per experiment: one reference channel (nr 1) and a binding channel (nr 2). The ganglioside was immobilised in channel nr 2. During SPR, the scFv was first run through the reference channel, measuring the background binding to the dextran matrix of the chip, followed by the binding channel. Next, the background signal was subtracted from the signal, resulting in a graph giving information about the molecular interaction kinetics.

3.9.1 Immobilising NeuGc GM3 on the CM5 chip

~0.15 mg Neu GcGM3 was dissolved in 100 µl methanol, then diluted in ethanol (Arcus) in a 1:9 ratio followed by a second round of dilution in a 1:3 ratio using HBS-N buffer (GE Healthcare). The system containing a CM5 (GE Healthcare) chip was washed with HBS-N (GE Healthcare) buffer containing 5 µM NaCl (Prolabo) and 0.1% P20 (GE Healthcare) until the signal stabilised. The ganglioside solution was run over channel nr 2 at a low speed (5 µl per minute for 30 minutes). A reference value of 50 or above indicates a coated chip, if the value is below 50 must the chip be re-coated.

3.9.2 scFv affinity measured by SPR

Humanised 14F7 mAb (CIM) at the same molar concentration as the mid concentration chosen for the scFv samples was run over the chip before and after an experiment, the affinity graphs were compared as a control to make sure the signal did not degenerate. All proteins were diluted in HPS-EP buffer (GE Healthcare) to the desired concentrations. The mid concentration of scFv was always duplicated to check for a reproducible result. The proteins were run separately in a single-cycle with the concentration range of 25, 50, 151, 151, 455, 910 nM (believed at the time to be 81, 162, 486, 486, 1458, 2916 nM, because of a misinterpretation in how to measure the protein concentration, see Section 3.4.2, page 27). After completing the experiment, Rune Johansen Forstrøm (Engineer at Oslo university Hospital, Oslo) performed curve fitting and computer calculations, for kinetic evaluation.

3.10 ThermoFluor

After mixing protein and dye in the wells of a 96-well plate, LightCycle 480 (Roche) with LightCycle 480 sealing foil (Roche) was used to test the thermal stability. The temperature was gradually raised from 20 °C to 90 °C, with steps of 0.3 °C every 3 seconds. The proteins eventually unfold due to high temperature, and the dye binds nonspecifically to exposed hydrophobic regions, emitting a signal detected by LightCycle 480 using the LightCycle software (Roche). The SYPRO Orange (Sigma) dye was used in these experiments, but different dyes can be used to optimise the signal. The resulting melting curve will give the melting temperature for the protein, which indicates its stability.

An initial ThermoFluor experiment was performed where C1, C3 and C4 were screened for the optimal scFv/dye concentration ratio. Since C2 was not obtained in high enough amounts, it was not included in the experiment. The result indicated that the highest protein concentration at 30 µg/ml (believed at the time to be 100 µg/ml, see explanation in Section 3.4.2, page 27) combined with the middle dye concentration (1:1000 SYPRO, Sigma) were the best for further analysis.

Ligand binding might stabilise the protein structure, giving the proteins increased stability and correspondingly a higher melting temperature. Two different antigens were available: NeuGc GM3 and a hydrophilic synthetic derivative of NeuGc GM3 called Tricer (see Figure 7). In order to investigate scFv stability with and without antigen, the scFv concentration was held constant at 30 µg/ml, the dye constant at a 1:1000 ratio, and an increasing concentration of either the NeuGc GM3 ganglioside or the Tricer (0 µg/ml, 10 µg/ml, 20 µg/ml, 30 µg/ml) was added. The ganglioside had first been solubilised in methanol (Prolabo), before the methanol solution was added to the scFv samples. Whit an increased NeuGc GM3 concentration followed a correspondent increase in the methanol concentration. The scFvs were diluted to a final concentration of 30 µg/ml, with 1:1000 dye, in 50 mM Tris-HCl (Chalbiocem) containing 750 mM NaCl. The antigen/scFv mix was incubated for 2 hours before subjecting it to ThermoFluor. The result was analysed by Gabriele Cordara, University of Oslo, Oslo, with the data software GraphPad Prism v5 (GraphPad software Ink).

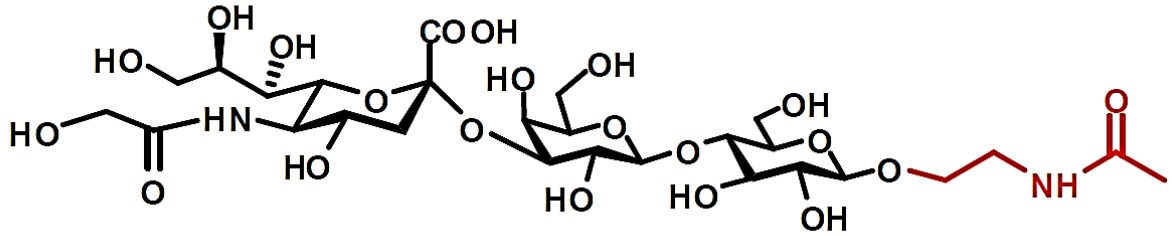

Figure 7: Molecular composition of the soluble NeuGc GM3 variation named Tricer**.** Tricer only contains the first ethylene group of the ceramide lipid, and was made more hydrophilic by the addition of an acetamide group (the ethylene group and the acetamide group are shown in red)**.**

3.11 Crystallisation

C1* scFv in 50 mM TrisHCl was mixed with a soluble derivative of the NeuGc GM3 ganglioside called Tricer (see Figure 7) in approximately ten times higher molar concentration.

The mixture was incubated for 1 hour before setting up the crystallisation screen using Crystallisation Plate, 96 MRC (SWISSCI) plates with vapour diffusion. The screening was done with the Oryx4 (Douglas Instruments) robot using the Wasp Run computer program (Douglas Instruments), with drops of a final volume of 0.4 µl, consisting of 50% or 25% protein solution. The different buffer screens used for crystallisation screening are listed in the Appendix, Section A: Materials, Table S7. The plates were stored in the dark at 4 °C and checked after 2 weeks, 1 month, 2 month and 4 month, respectively.

Optimised screening was done with larger drops of 2 µl with either 50% or 25% protein. The screening conditions were optimised using Morpheus MDI-46 kit (Molecular dimensions) containing all the different reagents. The protein mix was placed on siliconised cover slides (Hampton Research) in a hanging drop experiment using TPP® tissue culture plates, 24 wells (Sigma). The plates were stored in the dark at 4 °C and checked after 2 weeks, 1 month, 2 month and 4 month, respectively.

4 Results and Discussion

4.1 Status at project start

The 14F7 Fab structure was solved by Krengel et al. at a resolution of 2.5 Å (see Figure 4) (Krengel et al., 2004). Since it was not possible to reproduce the crystal, the mystery of how 14F7 binds to the NeuGc GM3 ganglioside is still not solved experimentally. Therefore, we wanted to investigate the possibility of crystallising a less complex 14F7 version, the single-chain Fv (scFv). This thesis is not the first attempt to produce functional 14F7 scFvs. Rojas et al. at the Center of Molecular Immunology, Havana, Cuba were not able to express 14F7 scFv consisting of the original V_H and V_L combined with their in-house constructed linker (original Cuba linker, $L_{C.o}$) (Rojas et al., 2004). However, by combining V_H with different variable light chains, they succeeded in producing five scFvs with conserved antigen affinity, solving the expression problem. The clone with the highest NeuGc GM3 affinity was named 3Fm (Rojas et al., 2004).

4.1.1 Synthesising the genes

We received the sequence for both the original and the 3Fm scFv, and Geir Åge Løset (Centre for Immune Regulation and Department of Biosciences, University of Oslo, Norway) discovered a truncation of 21 amino acids at the N-terminus of the V_H sequences in the original 14F7 scFv. It was confirmed by Gertrudis Rojas that the obtained sequence was indeed used experimentally for producing the original scFv, and that the truncation probably occurred during PCR (G. R. personal communication). This is likely the reason why Rojas et al. did not manage to produce the original scFv, as the truncation might have caused stability issues. When Geir Åge Løset synthesised the V_H gene, the truncation was corrected by reconstructing the missing 21 amino acids based on comparison with its V gene germ line segment.

When designing recombinant proteins, tags are practical for purification and detection. However, one of the aims of this thesis was to produce crystals, and flexible tags, which might disturb crystal formation, should be avoided. To be able to remove such a tag, in this case a Polyhistidine-tag (His-tag, 6x), a TEVp recognition site was included in the DNA construct. TEVp recognises the amino acid sequence ENLYFQ\S where '\' denotes the

position of the cleaved peptide bond. If the His-tag with the TEVp recognition were to be placed C-terminally, the six amino acids from the TEVp recognition site would not be removed, again resulting in a flexible C-terminus. To avoid this issue, the His-tag combined with the TEVp recognition site was placed N-terminally for V_H in the scFv. In retrospect, this was not well thought through. By looking at the 14F7 structure in PyMOL (Schrödinger) (PDB code: 1RIH (Krengel et al., 2004), see Figure 8), the N-terminus of V_H is placed next to the CDRs. Therefore, an N-terminal His-tag may affect the scFvs affinity towards the NeuGc GM3 ganglioside, as discussed in Section 4.2.4 page 46. A synthetic scFv linker normally has a length between 15-25 amino acids (Bird, 1988). We hypothesised that the linker used in the original scFvs (original Cuba linker, named: $L_{C.o}$) might have been too short, making it difficult for the scFv constructs to fold. Therefore, we added four extra amino acids to the linker introducing *Hind*III/*Mlu*I flanking restriction sites, resulting in a 21 amino acid long linker (^{HindIII}5'-L_C-3'^{MluI}) named L_C. In an alternative construct, we used a different linker sequence, in order to explore the effect that the amino acid composition may have on the scFvs ability to form soluble proteins. The alternative linker was an integrated part of the vector system obtained from Geir Åge Løset (Centre for Immune Regulation and Department of Biosciences, University of Oslo, Norway) and named L_R. After codon optimisation for *E. coli* expression, the synthetic genes were ordered from Life technology and the resulting DNA sequences for the four scFvs can be found in the Appendix, Section F: Amino acid sequences, page S29.

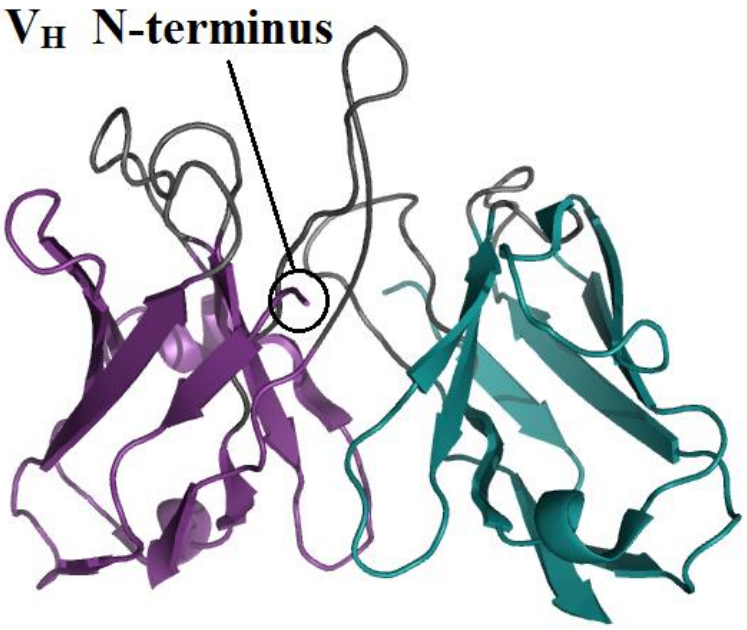

Figure 8: 14F7 Fv structure based on the Fab structure solved by Krengel et al. by X-ray crystallography, PDB ID: 1RIH (Krengel et al., 2004). The V_H chain is coloured violet, while the V_L chain is coloured teal green, the CDRs in grey, the N-terminal position of V_H is highlighted and the β-sheets are shown as arrows.

4.2 scFv constructs containing a His-tag

4.2.1 Cloning, subcloning and transformation

By generating four scFv constructs, we hoped to increase the chance for success both with respect to the yield and antigen affinity. The four resulting constructs (see Figure 9) were successfully cloned into the pFKPEN vector (see Figure 10) and transformed into *E. coli* XL1 Blue cells (Stratagene). In Figure 11 the finished C1 and C4 scFv constructs was digested with *Nco*I and *Mlu*I restriction enzymes for subcloning the C3 scFv. Figure 12 shows two clones of the generated C3 scFv that has been isolated from *E. coli* cultures.

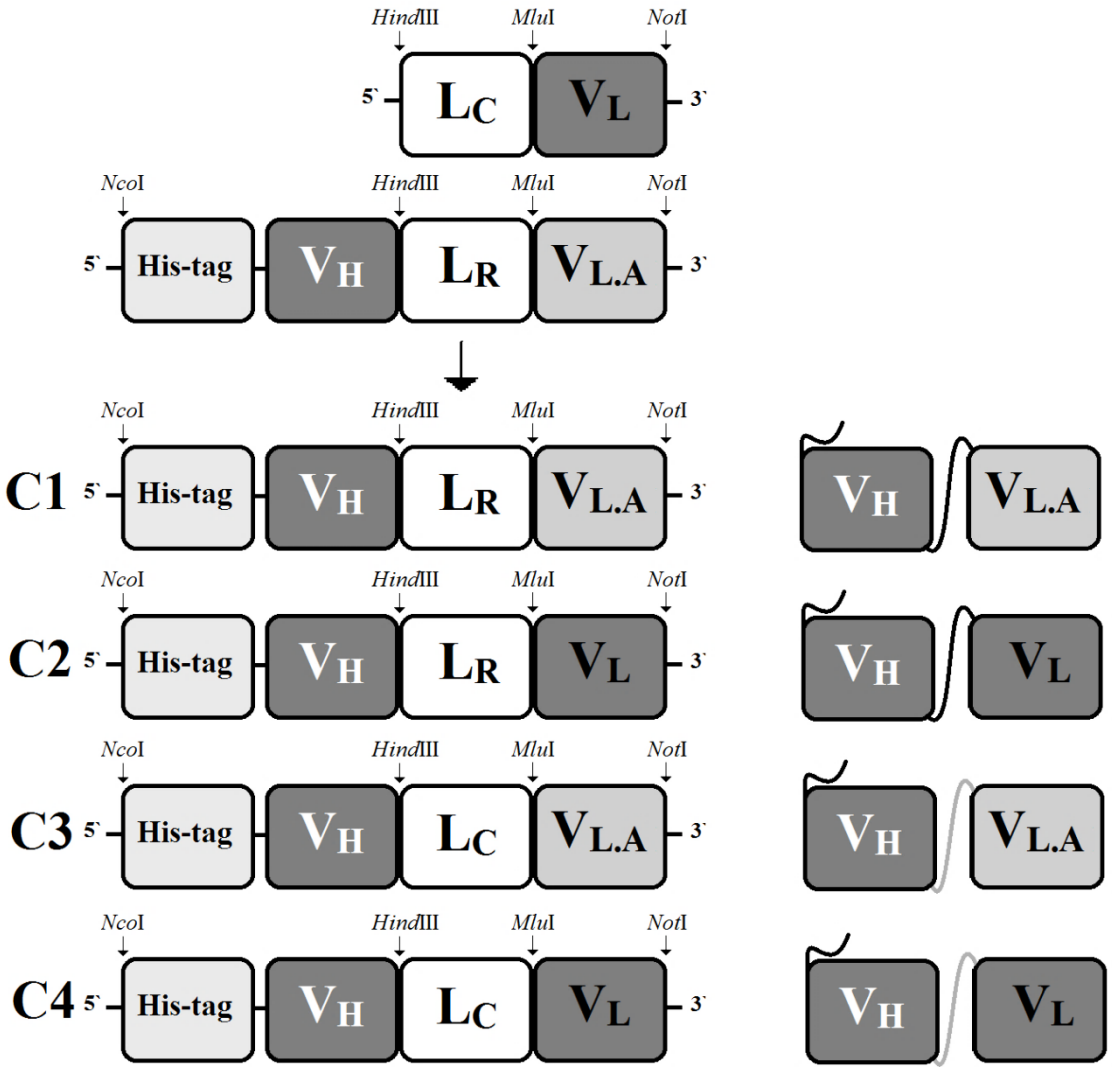

Figure 9: Overview, Cloning and subcloning of the 14F7 scFv constructs. To the left: the constructs' names (C1, C2, C3 and C4) followed by a block sketch of the DNA composition. To the right: a domain sketch of the scFvs. V_H represents the variable heavy chain, V_L represents the 14F7 variable light chain, $V_{L.A}$ represents an alternative variable light chain, L_C represents a modified version of the linker obtained from our collaborators at CIM in Cuba, and L_R represents the linker obtained from Geir Åge at CIR, Rikshospital. The sketch-designed is inspired by Gunnarsen et al. (Gunnarsen et al., 2010).

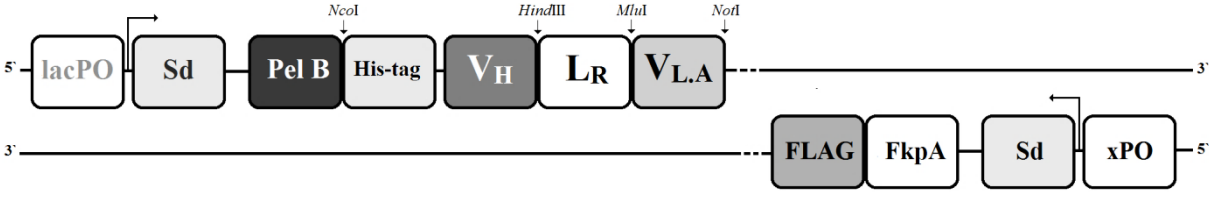

Figure 10: Overview, C1 subcloned into the pFKPEN vector. From the left: the *lac* promoter (lacPO), Shine-Dalgarno sequence (Sd), followed by the scFv construct 1 including the pelB sequence and the His-tag. On the opposite strand: The chaperon FkpA with FLAG-tag controlled by its native promoter xPO. The sketch-designed is inspired by Gunnarsen et al. (Gunnarsen et al., 2010).

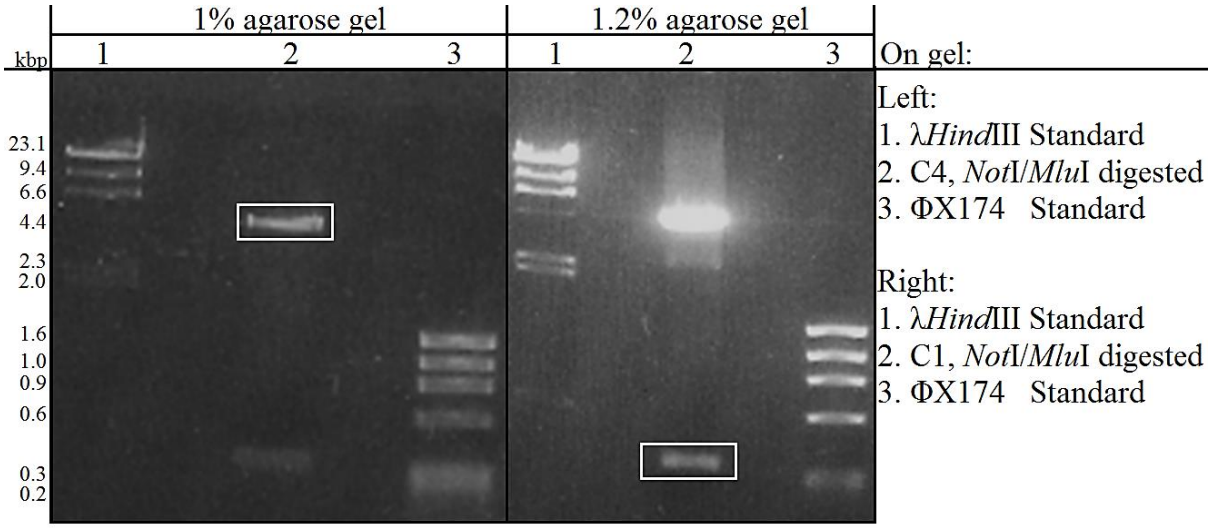

Figure 11: Subcloning of C3. C3 was generated from fragments of C1 and C4 already subcloned into pFKPEN. C1 and C4 in pFKPEN were both digested with *Not*I and *Mlu*I. V_L (size 341 bp) of C1 was isolated from the 1.2% agarose gel (right), while C4 in pFKPEN (size: 4470 bp) without its $V_{L.A}$ was isolated from a 1% agarose gel (left) and used for ligation to generate C3.

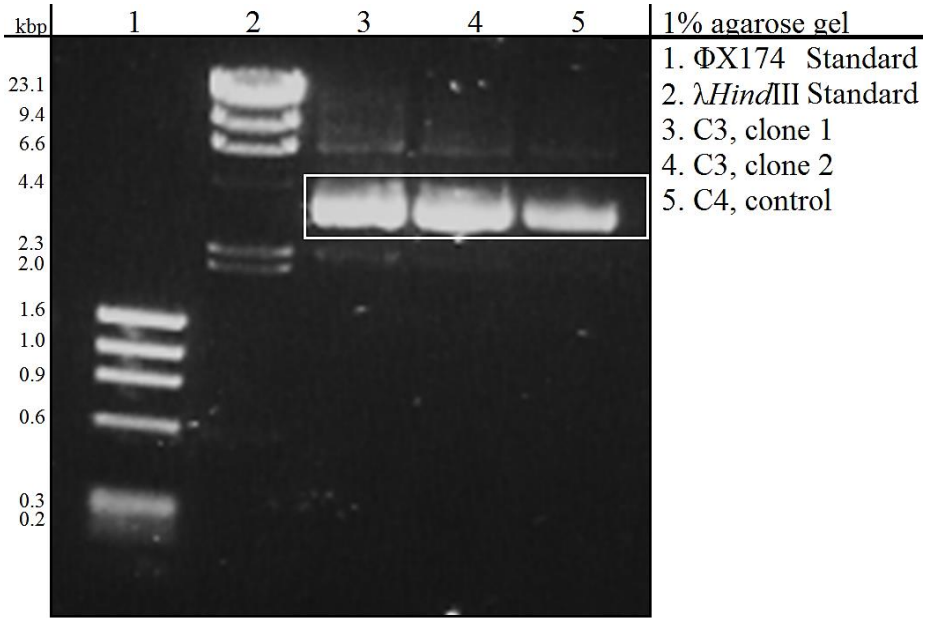

Figure 12: Subcloning ofC3. After generating C3 within pFKPEN (4815 bp) and transforming the plasmid in *E. coli*, plasmids from two different *E. coli* colonies were isolated on a 1% agarose gel. Isolated C4, pFKPEN (4815 bp) was used as a size control. The isolated plasmids were sent to sequencing and their identities were confirmed.

4.2.2 Protein expression and isolation I

The scFv expression is controlled by the *lac* promoter in *E. coli*. In order to help the cells grow, the scFv expression was downregulated by the addition of 20% glucose until the OD_{600} reached ~0.7. Then, the glucose was removed, but no inducer was added. The repressor is leaky and will allow for sporadic RNA transcription even without an inducer present, resulting in a low expression profile that prevents the formation of insoluble aggregates and increases the overall yield (Kipriyanov et al., 1997, Gunnarsen et al., 2010). The next step was to incubate the cells at 30 °C overnight without glucose, to give the bacteria time to express the proteins.

After expression and purification sufficient amounts of C1, C3 and C4 were obtained. Unfortunately, did not C2 express in high enough amounts for effective isolation. The sequencing confirmed that the C2 gene was successfully introduced into the pFKPEN vector, and transformed into *E. coli*. Therefore, the expression issue probably originates in the expression protocol In the beginning, we thought that C2 did not express well because the original pH of the growth medium (2x YT medium, pH 7.1) was exactly the same as the pI for the C2 scFv (information of pI for the different scFvs are listed in Table 1). The growth medium's pH was consequently changed to 7.8 to avoid the destabilising effect, but the expression did not improve. Another possibility was that the cell growth peaks earlier for C2 in the 30 °C overnight culture, than for the other scFvs. If the C2 bacteria started to die during the night, the C2 scFv might have been degraded. Although, it would probably be possible to optimise C2 expression no further effort was made, since it was possible to obtain sufficient amounts of C1, C3, and C4. After expression, isolation, and purification, the average yield for the C1 scFv was 5 mg per litre of *E. coli* culture, while the C3 scFv had a yield of 3 mg, and the C4 scFv a yield of 0.3 mg per litre of *E. coli* culture.

Rojas et al. did not manage to express scFv containing the original variable light chain in a prokaryotic expression system (Rojas et al., 2004). However, from the initial results in this thesis, it is confirmed that swapping the variable light chain improves scFv expression significantly. In addition to expression scFvs containing $V_{L.A}$ we also managed to express C4, a scFv containing the original 14F7 V_L. C4 differs from the original 14F7 by the elongated linker (4 amino acids), and the corrected truncation of the 21 amino acids at the N-terminus of V_H. This truncation is probably the main reason for why Rojas et al. did not manage to express the original 14F7 scFv (Rojas et al., 2004).

Table 1: scFv protein information calculated based on the amino acid composition using ProtParam, ExPASy (Swiss Institute of Bioinformatics) (Artimo et al., 2012).

	C1	**C2**	**C3**	**C4**
amino acids	268	268	268	268
weight (kDa)	30.0	30.0	29.9	29.9
pI	6.7	7.1	8.2	8.6
ext. coefficient M^{-1} cm^{-1}	54570	55600	54570	55600
absorption coefficient	1.82	1.86	1.83	1.85
instability index (Below 40 indicate stable proteins)	36.9	49.3	37.5	50.0

4.2.3 Purification using immobilised metal affinity chromatography

When purifying a protein with chromatography there are several different methods available. It is recommended to begin with the most specific technique, as for example affinity chromatography. Then, continue with more general separation techniques, as for example ion-exchange chromatography (IEC), followed by size-exclusion chromatography (SEC).For every purification step a fraction of protein is lost. Therefore, it is advisable to only include as many purification steps as is strictly necessary, depending on the degree of purity that is required for future experiments.

The scFv constructs contained a His-tag, giving the opportunity to selectively isolate the scFvs by affinity chromatography loaded with nickel ions, this specific affinity chromatography technique is named: immobilised metal affinity chromatography (IMAC). The absorbance was measured at 280 nm to detect the proteins migrating through the column. Proteins where eluted by outcompeting the proteins nickel affinity with free imidazole. The elution buffer had an imidazole concentration of 0.5 M and was raised in a gradient from 0% to 100%. The chromatogram displayed two peaks from the area where the elution buffer was

added (see Figure 13 to Figure 15). For C2 consisting of V_H combined with L_R and V_L there were no clear peaks, the elution profile exhibited two shoulders that are barely distinct from the background absorbance (A_{280}) caused by the increasing imidazole concentration. The first peak eluted when the gradient of elution buffer had reached ~40% (~200 mM imidazole) and the second peak eluted when the buffer over the column contained ~80% elution buffer (~400 mM imidazole). The fractions were separated on a polyacrylamide gel and it was concluded that the second peak corresponded to scFv with the correct size, see Figure 16. Since the flow-through still exhibited a band in the area corresponding to the size of scFv, the samples were loaded on the column a second time. After repeating the purification step, no scFvs were present in the elution fractions. The band in the flow through with a size similar to the scFvs is therefore probably not scFv, but rather the periplasmic chaperon FkpA. With a size of ~28 kDa, it is just a little smaller than the scFvs (~30 kDa). Since only one band was visible in the elution fractions (Figure 16), the samples were presumed to be pure. Therefore no further purification steps were done. The scFvs were dialysed against 50 mM Tris buffer and concentrated.

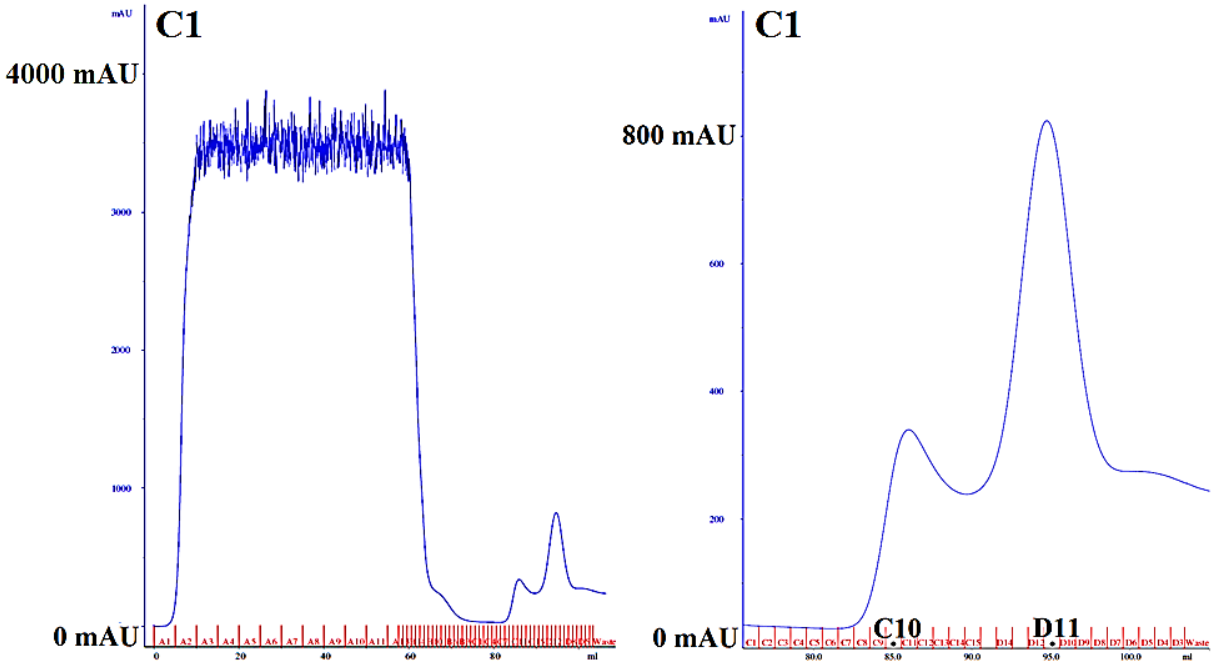

Figure 13: IMAC of C1. Left: full chromatogram Right: zoomed elution peaks. The marked fractions, C10 and D11, correspond to peak 1 and 2 in the chromatogram, and are shown on gel in Figure 16. *Blue graph: absorption at 280 nm. Red graph: absorption at 254 nm.*

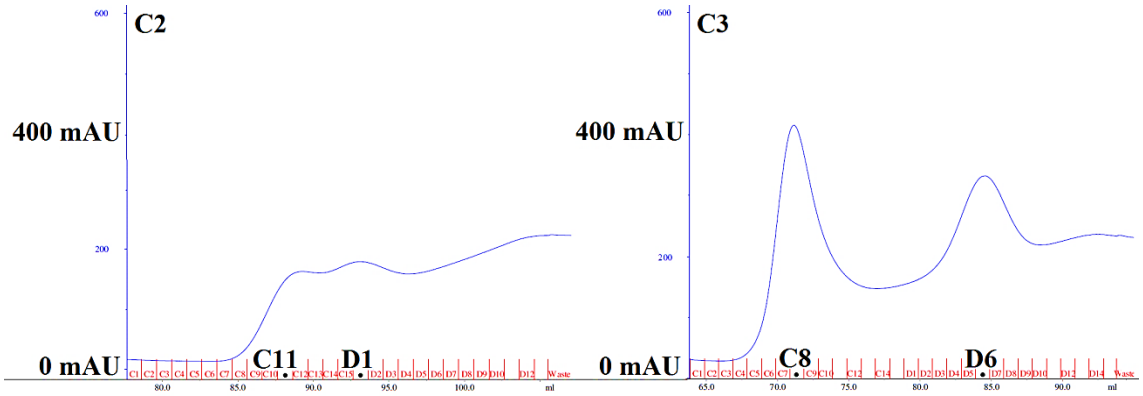

Figure 14: IMAC of C2 (left) and C3 (right), zoomed in on elution fractions. The marked fractions, C11 and D1 for C2 and C8, D6 for C3, correspond to peak 1 and 2 in the chromatogram, and are shown on gel in Figure 16. *Blue graph: absorption at 280 nm.*

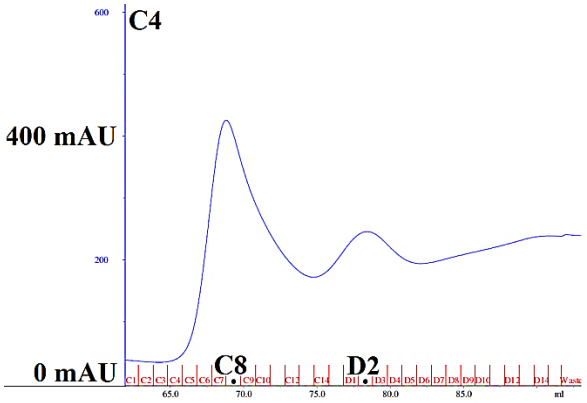

Figure 15: IMAC of C4 zoomed in on the elution fractions. The marked fractions, C8 and D2 correspond to peak 1 and 2 in the chromatogram, and are shown on gel in Figure 16. *Blue graph: absorption at 280 nm.*

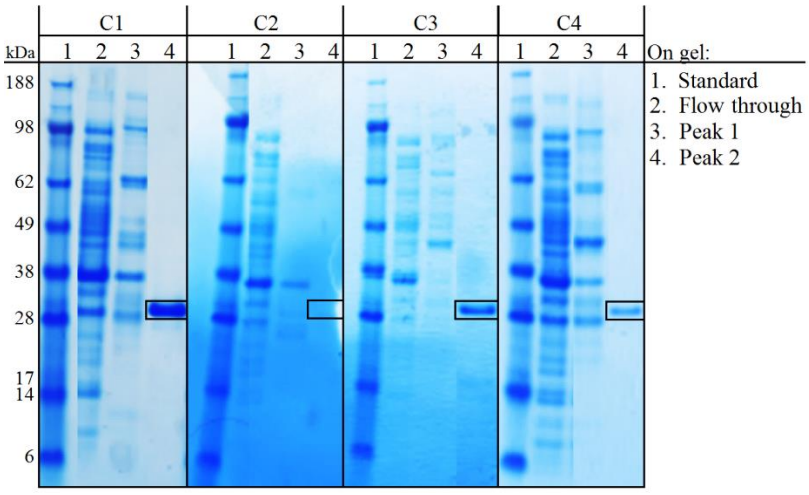

Figure 16: IMAC fractions on polyacrylamide gels. It is difficult to see a band corresponding to C2 scFv, but there is a faint band presence seen by visual inspection. *Collected fraction for peak number 1 and 2 are collected at highest absorbance (A_{280}). The scFv are located in peak number 2, and are highlighted with a square in the figure.*

4.2.4 Estimating the initial scFv constructs by ELISA

MicroWell PolySorp solid plate (Sigma) wells were incubated with 100 µl NeuGc GM3 ganglioside (10 µg/ml) in methanol (Prolabo), and allowed to evaporate overnight. Unspecific binding was blocked using 1x PBS containing 0.1% Tween 20 (Sigma) and 5% skimmed milk (AppliChem) (PBS-TM) and washed with 1x PBS with 0.1% Tween 20 (Sigma), between the different steps. After several trials of indirect ELISA with C1, C3 and C4, using a protein concentration range of 0.3 µg/ml – 15.6 µg/ml, no signal was seen (we thought at the time that the calculated concentration range was between 1 µg/ml – 5 µg/ml, see Section 3.4.2, page 27). We therefore started questioning the functionality of the scFv constructs. As discussed in Section 4.1.1, page 37, the His-tag was positioned N-terminally, placing it right next to the CDRs. The placement of the His-tag might have affected the scFvs activity. The following explanations for the failed experiments were considered and probed:

1. Human mistakes made in the protocol.
2. The NeuGc GM3 does not stick to the Polysorp surface of the wells and is washed away, leaving no ganglioside for scFv binding.
3. The scFv concentrations used in the experiments were too low to give a readable signal.
4. The secondary anti-His-tag antibody does not recognise the His-tag
5. The scFv constructs do not bind to the NeuGc GM3 ganglioside. One reason might be that the His-tag disrupts antigen binding.

1. Human mistakes are always a factor and can give variable results. Since the indirect ELISA was repeated at several occasions, all with a positive signal for the positive control and a negative signal for the negative control, this explanation was excluded.

2. If the ganglioside does not bind to the plates, the blocking would make it difficult for other proteins to bind, as long as the proteins do not interact with the blocking agent itself. Since the negative control gave no signal, no cross reactivity between the scFvs and the blocking agent was observed. If the ganglioside does not bind to the plate, the positive control should not be able to bind either. A positive signal for mAb 14F7 at a concentration of 10 µg/ml was seen, confirming the presence of the ganglioside.

3. A new indirect with a higher concentration range of 1 µ/ml – 100 µg/ml was conducted in a triplicate, but again no signal was observed. It is worth mentioning that the measured concentrations were initially overestimated by a factor of 3.2 (see explanation in Section

3.4.2, page 27). When adjusting the concentrations from the ELISA experiment, it was not 1 µ/ml – 100 µg/ml, but 0.31 µ/ml – 31 µg/ml that were used. At the time it was assumed that the original concentrations were correct, and therefore this explanation was considered unlikely.

4. Since the His-tag was placed at the N-terminus of the scFv constructs, it was possible that the His-tag was not accessible to the secondary antibody when the scFv was associated with the ganglioside. In order to investigate this explanation, a direct ELISAs was conducted with scFvs coated to the wells of the Nunc-Immuno 96 MicroWell MaxiSorp solid plates (Sigma), to confirm that the anti-His-tag antibody was capable of binding to the scFvs His-tag. The direct ELISA gave a concentration-dependent signal confirming the presence of the His-tag, and the functionality of the anti-His-tag antibody.

5. Since the His-tag was localised close to, or within, the binding site, its presence might disturb NeuGc GM3 binding. In order to find out if the activity of the scFvs improves when the His-tag was removed, an alternative detection method for ELISA was required. Protein L, a protein with affinity for the variable light chain was suggested by Geir Åge Løset (Centre for Immune Regulation and Department of Biosciences, University of Oslo, Norway). A direct ELISA was performed where C1, C3 and C4 at a concentration of 15 µg/ml was detected with Protein L linked to horseradish peroxidase (HRP) (Protein L-HRP). A signal was observed for all three constructs, confirming that Protein L was able to bind to the scFv constructs. It was difficult to assess how the His-tag influenced the scFvs binding affinity. We therefore decided to remove the tag permanently from the vectors, and rather purify and detect the recombinant scFv constructs utilising Protein L. Of course, a way of detecting the His-tag's influence would be to remove the His-tag using TEVp, and test the activity with Protein L-HRP in an indirect ELISA. But even after the removal of the His-tag, there would be two amino acids of the TEVp recognition site left, possibly disturbing NeuGc affinity. In order to test this, two different experiments were initiated in parallel: an activity test of C1 with a TEVp-cleaved His-tag, and the cloning of His-tag free scFvs. An in-house-produced TEVp was used for cleaving off C1's His-tag before an ELISA affinity test. Since the successful cleavage of the His-tag was difficult to determine, commercially available TEVp was ordered. No clear conclusion was reached as to the activity of the in-house-produced TEVp, and the ordered commercial TEVp did not arrive before the cloning and transformation of His-tag free scFvs were finished. Since Protein L gave the opportunity to

purify and detect scFvs without being dependent on a tag, we did not test His-tag scFvs when the commercial TEVp finally arrived. Instead we decided that future effort should be focussed on producing scFvs without a tag.

There are several other possible reasons why the scFv constructs did not bind to NeuGc GM3. Maybe the Fv needs the stabilisation effect provided by the constant domains of the Fab for antigen binding. Another possibility is that the linker might restrict the scFvs, giving it the wrong conformation. One thing is certain: the scFv must have the correct 3D structure to be able to bind the ganglioside. Since it was possible to isolate three out of four scFvs from the periplasm, the isolated scFvs certainly form soluble proteins.

4.3 Removing the His-tag

To find out whether the His-tag influences the scFv activity, the His-tag and the TEVp-recognition site were removed using oligonucleotide-directed mutagenesis. In the first attempt to remove the His-tag, the pFKPEN vectors containing the scFv' His-tag constructs were digested with *Nco*I/*Not*I restriction enzymes and isolated using the QIAquick gel extraction kit (QIAGEN). The genes were amplified by PCR using a primer (Eurofins Genomics) that did not include the His-tag and TEVp recognition site. After isolating the PCR fragments, ligation and transformation, the negative control had 26 colonies, while the construct plates had around. Four colonies from each construct were sent to sequencing, none of the chosen colonies had the His-tag removed. In the second attempt the genes were amplified with PCR directly on the plasmid, using the same primers. The annealing temperature was raised from 55 °C to 60 °C to avoid unspecific annealing, and as a control, every scFv gene was mixed separately with only forward or reverse primer and subjected to the same PCR conditions (Figure 17 shows the resulting PCR amplified DNA on a gel). The DNA concentrations from the isolated PCR fragments were low, therefore only 1000 ng DNA was used for re-digestion with *Nco*I/*Not*I restriction enzymes instead of the recommended 2000 ng. When digesting the original pFKPEN vector with *Nco*I/*Not*I restriction enzymes, CIP (NEB) was added to prevent vector re-ligation. After ligation and transformation into *E. coli* XL1-Blue cells, there were no colonies on the negative control, while the constructs had around 60 colonies. The sequencing confirmed complete removal of the His-tag sequence encoding portion and the TEVp recognition site, and preservation of the start codon and its relative position towards the *lac* promoter and Shine-Dalgarno sequence. The new constructs were denoted 14F7 scFv:

C1*, C2*, C3* and C4*. Table 2 gives general information about the different scFvs and a sketch of the new constructs are shown in Figure 18.

Table 2: scFv protein information calculated based on the amino acid composition using ProtParam, ExPASy (Swiss Institute of Bioinformatics) (Artimo et al., 2012).

	C1*	**C2***	**C3***	**C4***
amino acids	255	255	255	255
weight (kDa)	28.3	28.3	28.2	28.2
pI	6.7	7.63	8.5	8.8
ext. coefficient M^{-1} cm^{-1}	53080	54110	53080	54110
absorption coefficient	1.88	1.90	1.88	1.92
instability index (Below 40 indicate stable proteins)	38.6	51.67	39.3	52.4

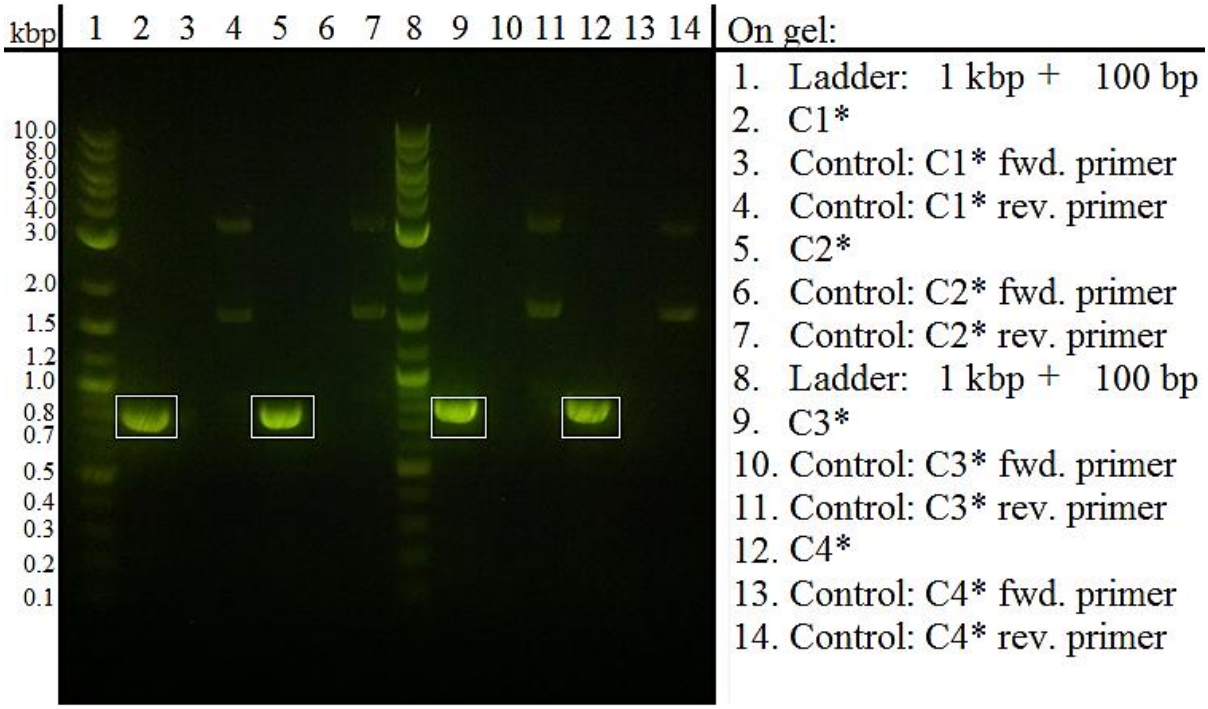

Figure 17: scFvs' genes without a His-tag on a 1% agarose gel. The His-tag was removed during PCR by oligonucleotide-directed mutagenesis. The size of the new scFvs named C1*, C2*, C3* and C4* were reduced from 819 to 780 bp. In the wells with the reverse primer control (4, 7, 11 and 14) there are impurities. This might indicate that the reverse primer recognise a second site in the vector, or an impurity due to human mistakes.

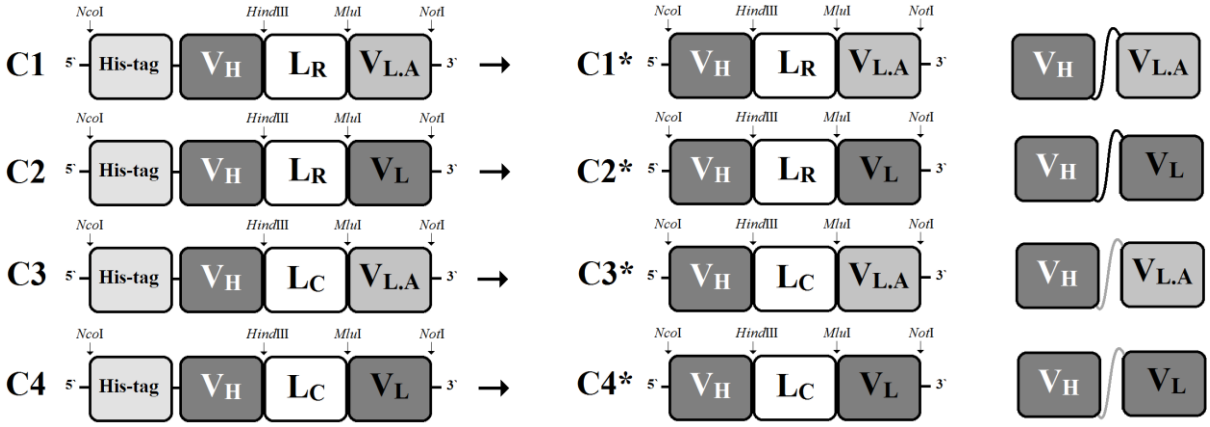

Figure 18: Overview. Removal of the His-tag from the scFv constructs. From the left: names and a block sketch of the scFvs with a His-tag. Mid-section: new names and a block sketch of the tag-free scFvs. To the right: a block sketch of the scFv domains connected by a linker. V_H: variable heavy chain. V_L: original 14F7 variable light chain. $V_{L.A}$: alternative variable light chain. L_C: elongated version of the linker obtained from our collaborators at CIM, Cuba. L_R: linker obtained from Geir Åge at CIR, Rikshospital. The sketch-designed is inspired by Gunnarsen et al. (Gunnarsen et al., 2010).

4.4 scFv constructs without a His-tag

4.4.1 Protein expression II

During expression, the OD_{600} of the bacterial cultures was measured in a time-dependent fashion. An example of OD_{600} measurements during expression is provided in Table 3 Bacterial culture samples were loaded on a polyacrylamide gel after normalising the amount according to the OD_{600} of the cultures by using the formula: $OD^{-1} \times 50 \times$ nr of samples, and separated by SDS-PAGE, see Figure 19 and Figure 20. The factor 50 was chosen since 50 µl gave the best band-resolution for a bacterial culture with an OD of 1. The scFv size has been reduced from ~30 kDa to approximately 28 kDa when removing the His-tag. Since the chaperone FkpA has a very similar size to the scFvs (~28 kDa) a band was expected in all cell fractions. The chaperone is controlled by its native promoter and is supposed to be expressed continuously, and in a range of 20 to 30 times higher than the scFvs (Geir Åge Løset, G.R personal communication). A band at ~28 kDa was only seen in the medium of the overnight culture before removing the glucose (fraction 9 in Figure 19 and Figure 20) for C1*, C3* and C4*, and in the final fractions for both whole cells and medium for all four constructs (fraction 8 and 15 in Figure 19 and Figure 20). Since the band only occurs in overnight cultures it was probably the scFvs and not FkpA. The chaperone should have been expressed in high amounts throughout the cells' growth cycle. If there had been more time, the question could have been answered using Western blotting. The scFvs can be detected with Protein L, followed by a re-blotting of the membrane to detect the FkpA utilising its FLAG-tag (see Figure 10 for a sketch of the FkpA chaperon). The result would make it clear which protein is present in which fractions.

scFvs can be isolated from both cells and medium. Isolating protein from medium was time-consuming as the medium had to be concentrated ten times to reduce the volume before affinity chromatography. Consequently, isolation from medium was only done for C4*, which had a much lower yield than to C1* and C3*. C2* was not optimised as sufficient material was available for the other scFvs. After expression, isolation, and purification, the average yield for the C1* scFv was ~5.5 mg per litre of *E. coli* culture, the C3 scFv had a yield of ~3.0 mg, and the C4 scFv had a yield of ~0.1 mg per litre of *E. coli* culture. Compared to scFvs with a His-tag the yield was almost identical for C1* and C3*, both containing $V_{L.A}$, but the yield was reduced significantly for C4*, with a decrease in the yield from 0.3 to 0.1 mg per

litre culture. A decrease could also be observed for C2*, the polyacrylamide gel had to be silver stained to be able to see a band, while a faint band could be observed for purified C2 after IMAC with Coomassie staining The difference in expression, purification and activity is discussed further in Section 4.9, page 73.

Table 3: An example of OD_{600} measurements taken during protein expression

	C1*	**C2***	**C3***	**C4***
ON +glucose, 37 °C	1.3	1.6	1.6	1.5
OD_{600} 0.6 -0.8,+glucose, 37 °C	0.7	0.7	0.8	0.8
t = 0 h, 30 °C	0.7	0.8	0.9	1.3
t = 1 h, 30 °C	0.9	0.8	1.1	0.8
t = 2 h, 30 °C	1.0	0.9	1.1	1.5
t = 4 h, 30 °C	2.4	1.6	1.6	2.8
ON 30 °C	4.5	3.3	3.4	5.1

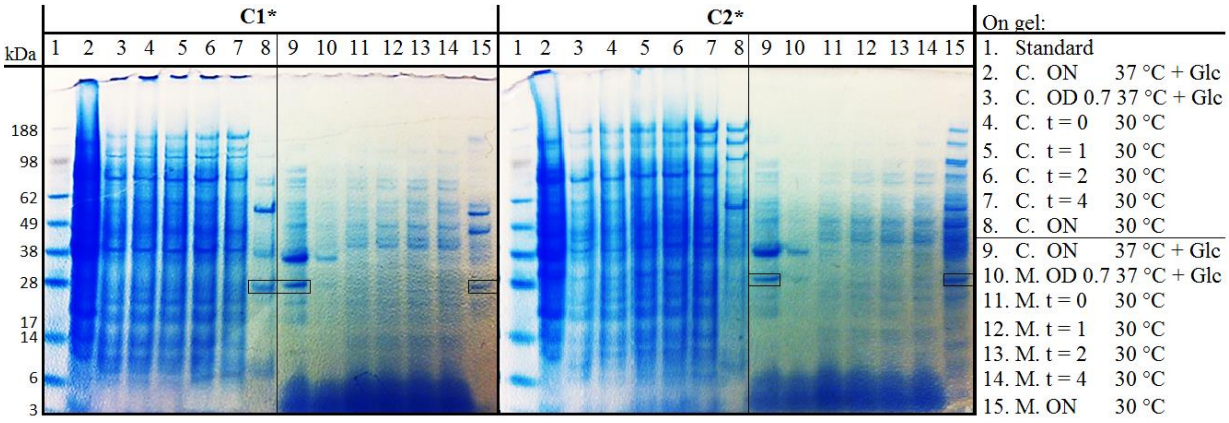

Figure 19: Expression test for C1* and C2*, the proteins are separated by SDS-PAGE. The whole cell fractions (C.) were normalised according to the OD. The medium was not normalised: 10 µl of medium (M) was loaded per sample for gel fractions 9 to 15. Samples were taken at different time points (t) during cell growth. A band corresponding to the size of scFv can be observed in gel fractions 8, 9 and 15 for C1* and in fraction 8 and 15 for C2*.

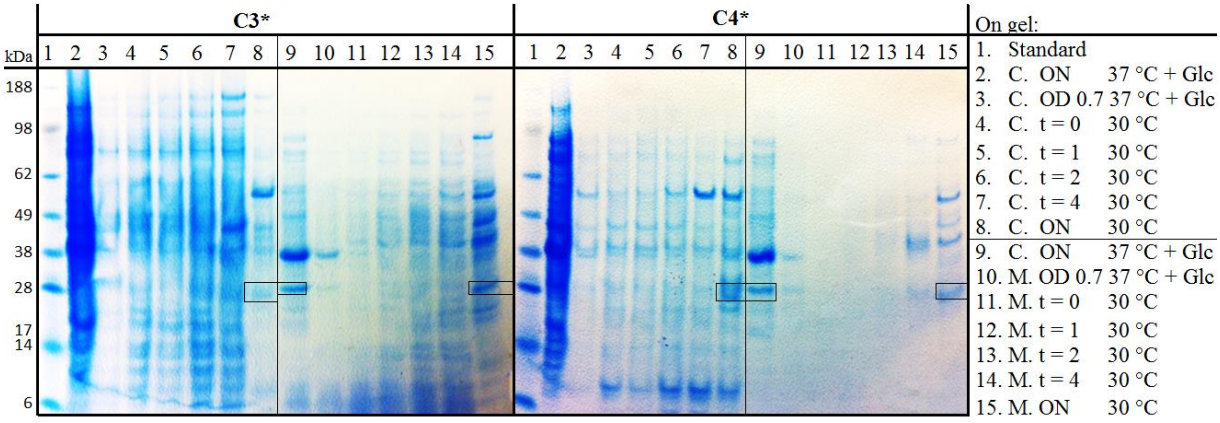

Figure 20: Expression test for C3* and C4*, the proteins are separated by SDS-PAGE. The whole cell fractions (C.) were normalised according to the OD. The medium was not normalised: 10 µl of medium (M) was loaded per sample for fractions 9 to 15. Samples were taken at different time points (t) during cell growth. A band corresponding to the size of scFv can be observed in gel fractions 8, 9 and 15 for C3* and C4*.

4.4.2 Purification

Protein L affinity chromatography

Protein L is a dimer with four binding sites for the variable light chain of antibodies. Protein L binds human κ-light chain 1, 3 and 4 as well as many murine κ-light chains (Kastern et al., 1992, Graille et al., 2001). Both V_L and $V_{L.A}$ are members of the murine κ subgroup, and should bind Protein L. Since the constructs were similar in size a separate column was used for each construct. This proved to be a good decision as the constructs containing the $V_{L.A}$ had a stretched-out elution profile (see Protein L elution fractions on gel in Figure 21 and Figure 22). This was probably due to the fact that Protein L has several binding sites for the variable light chain, generating avidity, making it difficult to elute the scFvs containing $V_{L.A}$. In order to investigate if there were more available scFvs in the flow through, the flow through was added to the column a second time. Since there were almost no protein observed in the elution fractions it was concluded that one round over the column was sufficient.

The elution pattern indicates a difference in Protein L and V_L/$V_{L.A}$ binding, as scFvs with V_L eluted quite quickly (see Figure 21 and Figure 22). When comparing the amino acid sequences of the two variable light chains with other light chains known to interact with Protein L, we found that V_L was lacking two amino acids in the main Protein L binding region, while $V_{L.A}$ was lacking one amino acid (see Figure 23 for detailed information). This might be the reason why scFvs consisting of $V_{L.A}$ sticks to the column. The difference may

also be related to the amount of protein loaded onto the column. C1* and C3* consisting of the $V_{L.M}$ have a better expression pattern compared to C4* and C2*, which both contain the $V_{L.}$ The difference in affinity for Protein L might have implications for a potential quantitative comparison for scFvs affinity to the ganglioside in ELISA, since Protein L was used to detect binding.

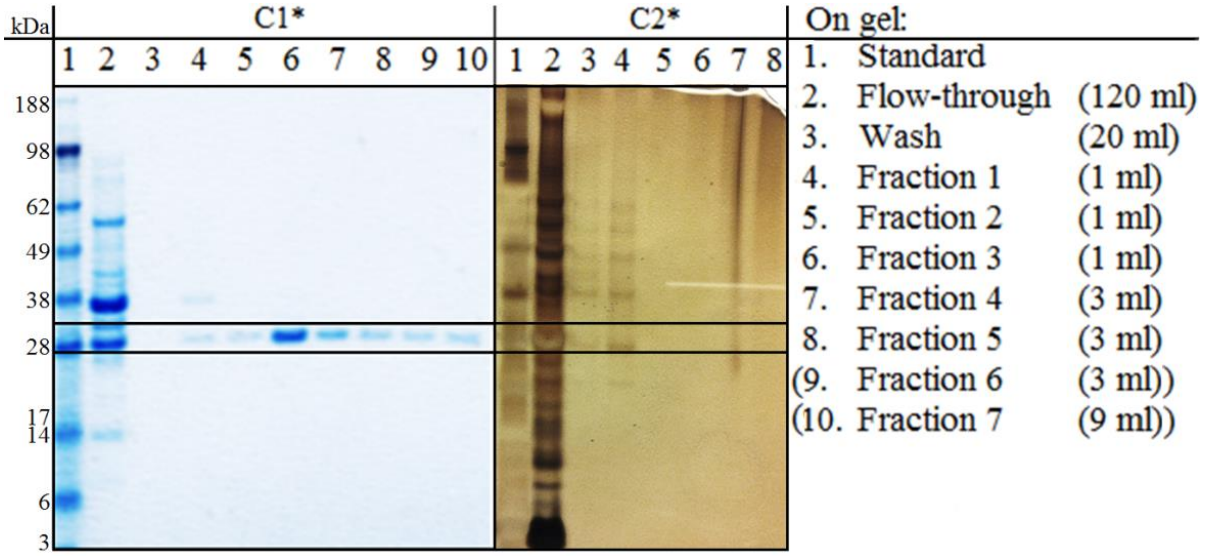

Figure 21: C1* (left) and C2* (right) isolated fractions by Protein L affinity chromatography column. *C1*(to the left) is coloured Coomassie staining, C2* (to the right) is coloured by silver staining.*

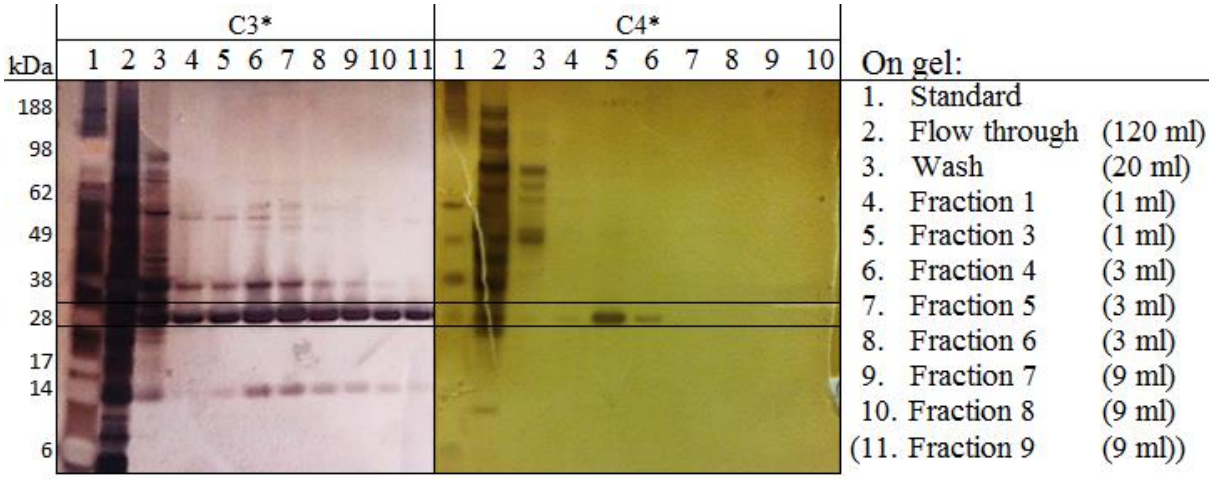

Figure 22: C3* and C3* isolated fractions by Protein L affinity chromatography. *C3*(to the left) and C4* (to the right) are both silver stained.*

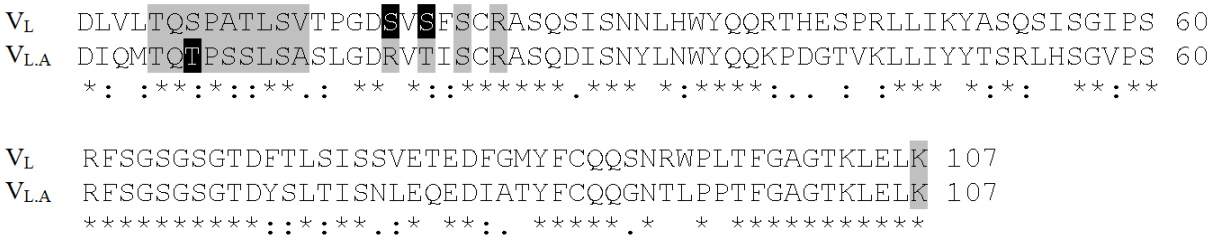

Figure 23: Alignment of amino acid sequence for $V_{L.A}$ and V_L C1* and C3* consist of $V_{L.A}$ while C2 and C4 consist of $V_{L.}$ *Coloured in grey are amino acids with the correct R-group for Protein L interaction. Coloured in black are mismatch amino acids in positions that are important for Protein L interaction (Graille et al., 2001).*

Size-exclusion chromatography (SEC)

When performing SPR with C1* and C4*, the resulting graph suggested a high degree of heterogeneity. To combat this, a second purification step was included in the purification protocol for the constructs without His-tag. The resulting chromatograms for the constructs showed two overlapping peaks (see Figure 24 and Figure 25). When analysing the fractions by SDS-PAGE, it was discovered that one of the peaks corresponded to a protein with a size of ~14 kDa (see Figure 26). The presence of this unknown band is discussed further in the next section. In order to separate the scFvs from the unknown protein, the proteins could be denatured by pre-incubating the sample with for example urea or guanidinium before SEC. Since the unknown protein elutes at approximately the same time as the scFv, it is assumed that it dimerise. By first denaturing the samples, the proteins will elute as monomers during SEC, separating the ~14 kDa from the ~28 kDa proteins. After SEC, the scFv must be refolded *in vitro*, but this might lead to a heterogenic sample, which is especially problematic for crystallisation. Also, by denaturing and refolding the scFvs, a portion of scFv will most likely be lost due to refolding problems, making this method sub-optimal for obtaining a high scFv concentration.

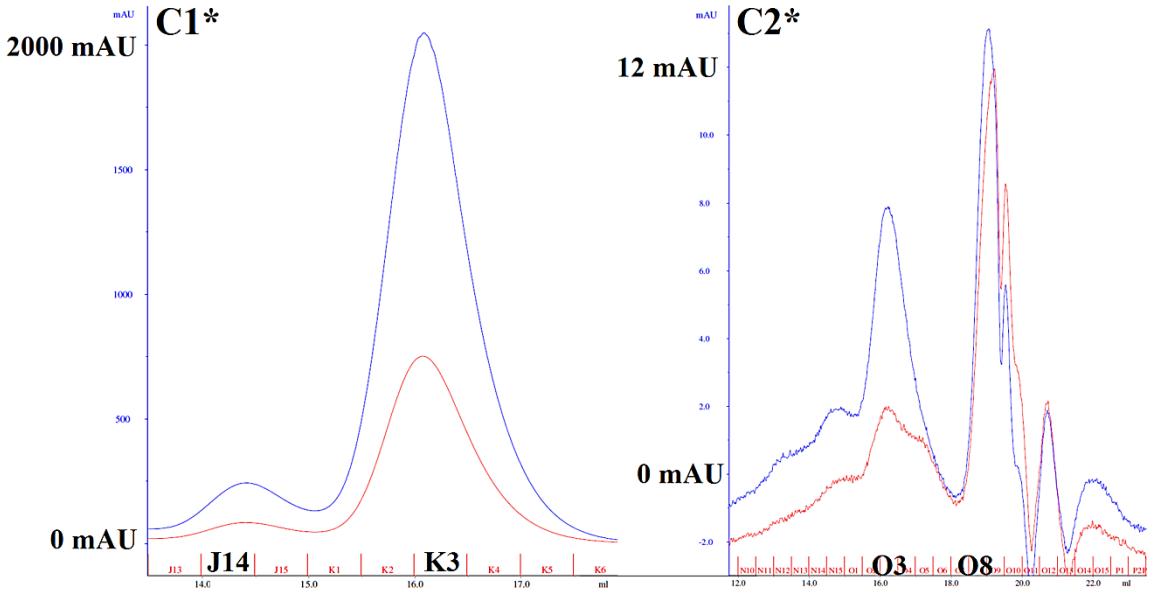

Figure 24: SEC chromatogram of C1* (left) and C2* (right). Marked fractions (J14, K3 for C1* and O3, O8 for C2) can be observed on gel in Figure 26. *Blue-line: absorbance measured at 280 nm. Red-line: absorbance measured at 254 nm.*

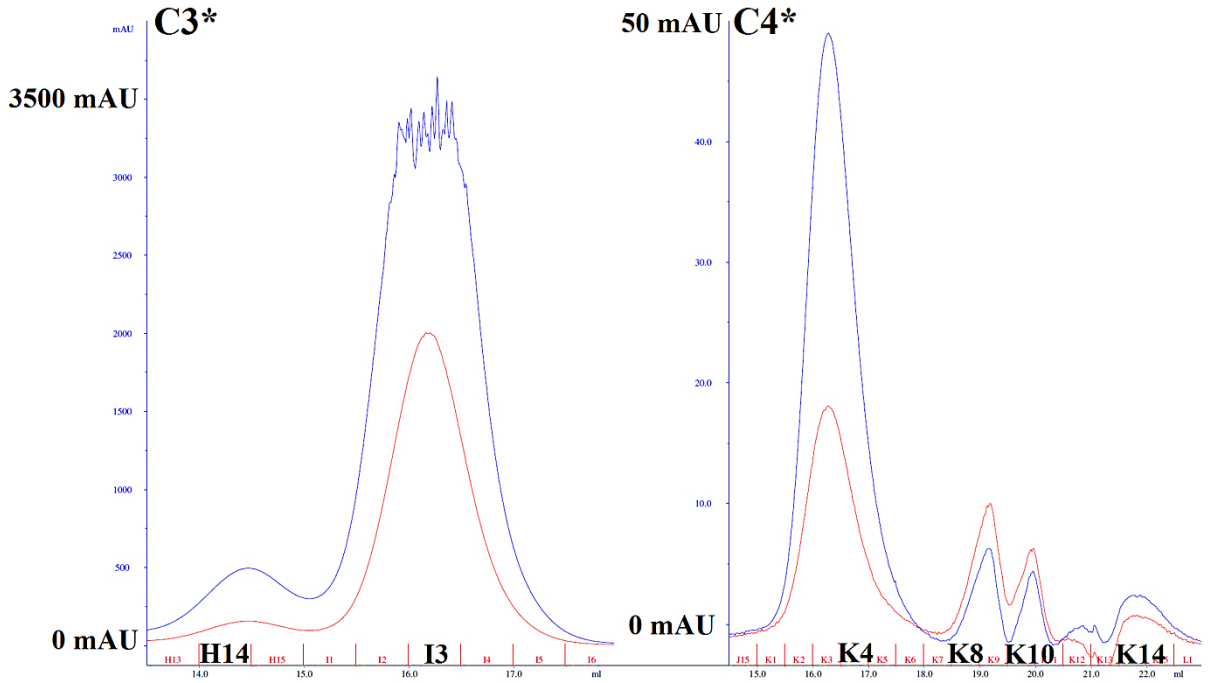

Figure 25: SEC chromatogram of C3* (left) and C4* (right). Marked fractions can be observed on gel in Figure 26 *Blue-line: absorbance measured at 280 nm. Red-line: absorbance measured at 254 nm.*

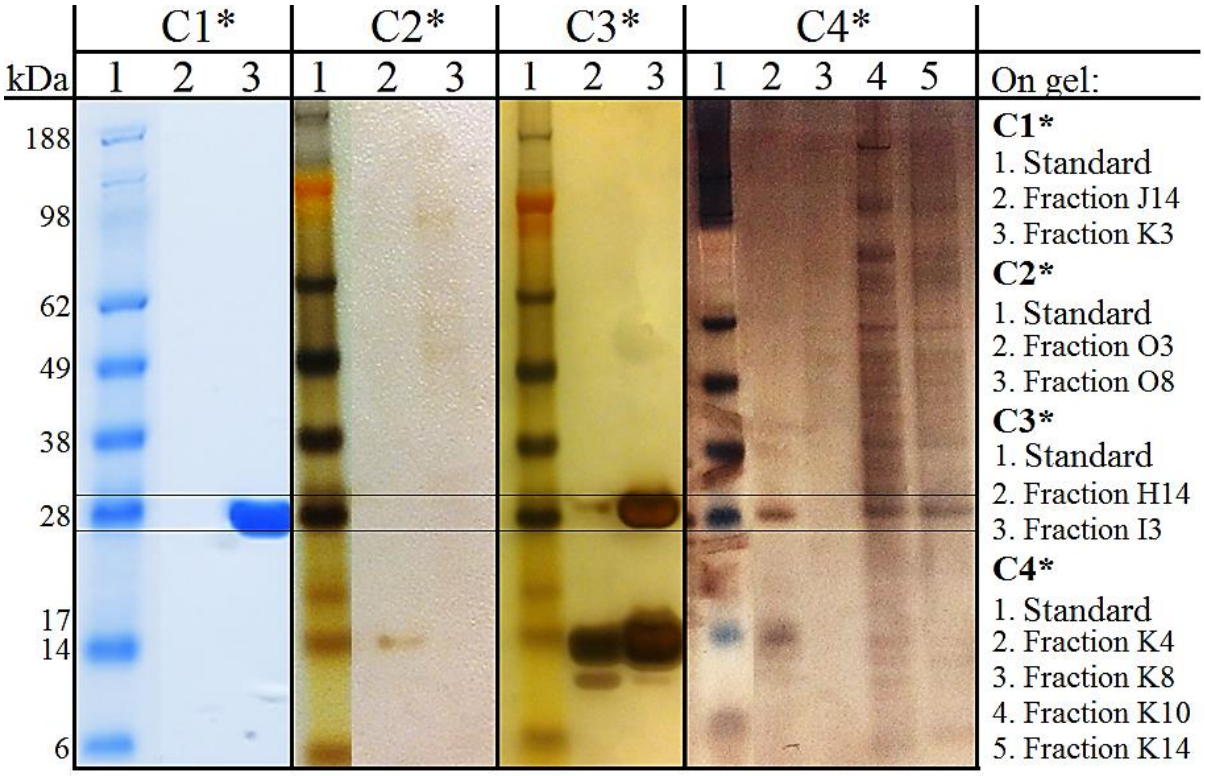

Figure 26: SEC fractions from (from left) C1*, C2*, C3*, C4* separated on NuPAGE Bis-Tris 4-12% gel (Life technology) by SDS-PAGE. *C1* is Coomassie stained, while C2*, C3* and C4* are silver stained*.

4.5 Additional band at ~14 kDa

The unidentified protein with the approximate size of ~14kDa probably has the ability to dimerise non-covalently, forming a ~28kDa complex in solution. This would explain why it elutes during SEC in a peak overlapping with the scFvs. Since the unknown protein was co-purified with scFv during Protein L chromatography it must have some degree of Protein L affinity. Two explanations for the origination of the protein were considered:

1. The unknown protein is the scFv construct cleaved in two.
2. The unknown protein might be lysozyme.

1. The unidentified protein is approximately half the size of the scFv, making the suggestion that it is a cleaved version of the scFvs an attractive explanation. The most likely explanation would be that the scFvs are cleaved by a protease. If the cleaving is site specific, the recognition sequence is probably located in the consensus region flanking the linkers, since the cleaved scFvs consist of different linkers and variable light chains. When loading high concentrations of purified scFv on a NuPAGE Bis-Tris 4-12% gel (Life technology) several

bands between 10 and 28 kDa were observed, see Figure 27, possibly showing scFv split in several pieces by unspecific cleavage. Several proteases in *E. coli* cleave unspecifically by recognising exposed unstructured or hydrophobic regions. Perhaps the proteases degP or OmpT are responsible for the fragmented scFvs. DegP is the most likely candidate, located in the periplasm of *E. coli*, while OmpT is localised in the cytosol (Jones et al., 2002, Baneyx and Georgiou, 1990, White et al., 1995). An autolysis peptide sequence could also be responsible for the cleaving, or the fragmentation might occur due to a chemical added during protein isolation and purification. This is probably not the case, since the isolation and purification protocols are found in similar versions in several articles where single-chain proteins are produced (Schaefer and Plückthun, 2012, Gunnarsen et al., 2010, Kipriyanov et al., 1997). The bands for C1* highlighted in Figure 27, and the corresponding bands from C3* and C4* (not shown) were sent to mass spectrometry (MS) for identification. The protein was compared with the sequence of the four scFv constructs, the FkpA chaperone sequence and lysozyme. While waiting for the MS results, the second possibility that the unknown protein represents lysozyme was investigated further.

2. Lysozyme is not known for interacting with Protein L, but since it was used in high concentrations (1 mg/ml) in the periplasmic expression protocol, it might be able to associate non-specifically with Protein L or the column material. Lysozyme also has the correct size, 14 kDa, and has the ability to dimerise. Ion exchange chromatography (IEC) was considered as a method to separate the scFv with the unidentified protein based on its pI. However, since lysozyme (Sigma) has a pI of 9.3 and the C1*, C3*, C4* have a pI of 6.7, 8.5 and 8.8, respectively, the difference in pI is only large enough to perform IEC on C1*, but would not work for C3 and C4*. Since it was considered important to remove the unidentified proteins from all the samples, other separation techniques were conducted. To test if lysozyme binds Protein L at all, a concentrated lysozyme solution could have been added to the Protein L column, but this was not done since a method to separate the scFvs from the unidentified protein was still desired.

To find out if the proteins elute at different pH, Protein L affinity chromatography was performed on C3* (with the protein composition seen in Figure 26), where the pH was lowered half a unit at the time. However, both proteins eluted at pH 3.

C3* (with the protein composition seen in Figure 26) was mixed with an *E. coli* sample. Cell degradation was investigated after 1, 2, 4, 6 hours and compared with a negative control. If

the unknown protein was lysozyme, there should be an increase in cell death since lysozyme degrades bacterial cell walls. No increased degradation was observed.

Since more protein was needed for further experiments, an alternative isolation without lysozyme was tested. The proteins were isolated by cracking the whole cells using a high pressure homogeniser (EmulsiFlex-C3, Avestin) (see the lab protocol in Section 3.2.2, page 2424). After Protein L affinity chromatography and SEC, followed by SDS-PAGE, the gel still showed the presence of the unidentified protein as shown in Figure 27. We could therefore conclude that the unidentified protein was not lysozyme.

In the meantime, the MS results had arrived. By mapping the peptide sequence for the scFvs, it was revealed that the unidentified band at ~14 kDa consisted of scFv fragments (see Figure 27 and Figure 28). It is clear that for the cleaved scFv, both the heavy and light chain are connected through Protein L affinity chromatography and SEC, as the MS results show peptide coverage from both variable chains. It is possible that the restriction sites *Hind*III/*Mlu*I are part of the recognition site for proteolytic cleavage, as they together with the flanks of the variable domains create a consensus sequence in all four scFvs. By replacing the restriction sites, the fractionation might be prevented, but the replacements must be considered thoroughly as most changes will affect the properties of the linker. One suggestion would be a linker constructed by Gunnarsen et al. where a leucine is mutated into a proline (L2P, linker numbering), destroying the *Hind*III recognition site in the gene sequence, and has a positive entropic effect on the linker's flexibility (Geir Åge Løset, G.R personal communication) (Gunnarsen et al., 2010). But if the cleavage is performed by an unspecific protease such as degP, it might be necessary to transform the constructs into a degP-deficient *E. coli* strain. In order to find the best solution it would be beneficial to known why and where the scFvs are cut. One way of investigating where the scFvs are cleaved is by cleaving the scFvs with papain and send the fractions to MS. Papain cleaves C-terminally after Arg and Lys, all peptides formed by papain cleavage would therefore end on either Arg or Lys. If a peptide that does not end on Arg or Lys was detected, this would have to be cleaved by something else, and would indicate the cleavage site.

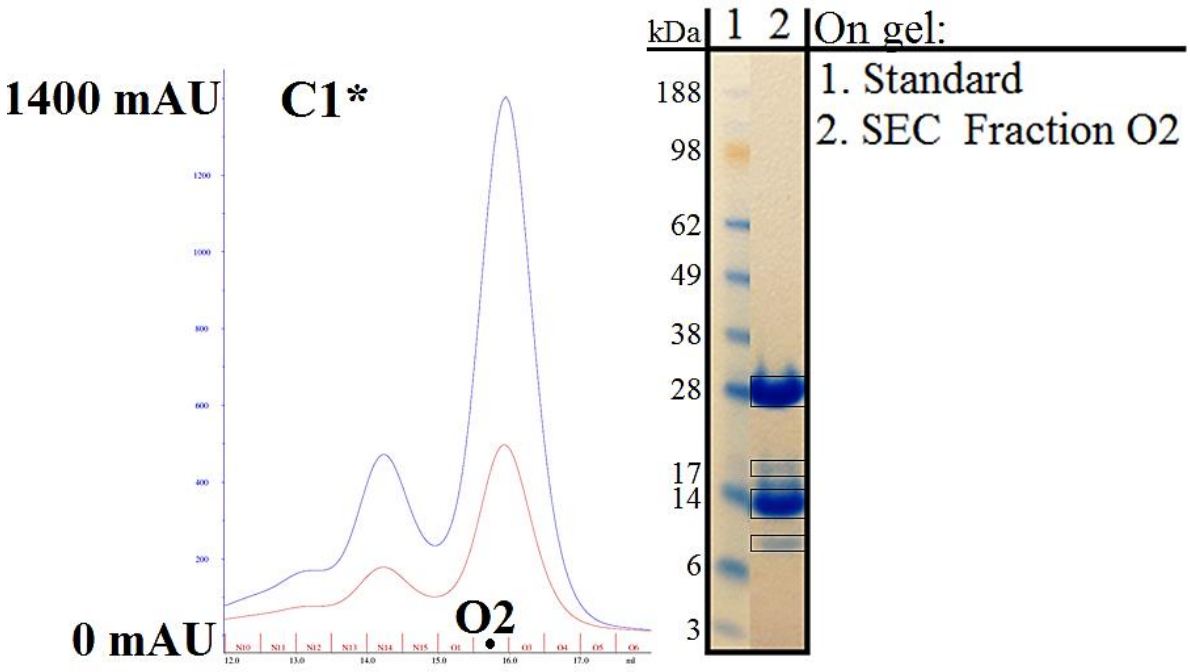

Figure 27: C1* SDS chromatograph and proteins isolated from cracked cells and separated by SDS-PAGE. Fraction at ~28, ~20, ~14 and ~10 kDa were sent to MS analysis.

C1* SEC fraction O2 Bond size: ~28 kDa

MAQVQLQQSGAELAKPGASMKMSCRASGYSFTSYWIHWLK
QRPDQGLEWIGYIDPATAYTESNQKFKDKAILTADRSSNTAF
MYLNSLTSEDSAVYYCARESPRLRRGIYYYAMDYWGQGTT
VTVSSKLSGSASAPKLEEGEFSEARVDIQMTQTPSSLSASLG
DRVTISCRASQDISNYLNWYQQKPDGTVKLLIYYTSRLHSGV
PSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPPTFGAG
TKLELK

C1* SEC fraction O2 Bond size: ~20 kDa

MAQVQLQQSGAELAKPGASMKMSCRASGYSFTSYWIHWLK
QRPDQGLEWIGYIDPATAYTESNQKFKDKAILTADRSSNTAF
MYLNSLTSEDSAVYYCARESPRLRRGIYYYAMDYWGQGTT
VTVSSKLSGSASAPKLEEGEFSEARVDIQMTQTPSSLSASLG
DRVTISCRASQDISNYLNWYQQKPDGTVKLLIYYTSRLHSGV
PSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPPTFGAG
TKLELK

C1* SEC fraction O2 Bond size: ~14 kDa

MAQVQLQQSGAELAKPGASMKMSCRASGYSFTSYWIHWLK
QRPDQGLEWIGYIDPATAYTESNQKFKDKAILTADRSSNTAF
MYLNSLTSEDSAVYYCARESPRLRRGIYYYAMDYWGQGTT
VTVSSKLSGSASAPKLEEGEFSEARVDIQMTQTPSSLSASLG
DRVTISCRASQDISNYLNWYQQKPDGTVKLLIYYTSRLHSGV
PSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPPTFGAG
TKLELK

C1* SEC fraction O2 Bond size: ~10 kDa

MAQVQLQQSGAELAKPGASMKMSCRASGYSFTSYWIHWLK
QRPDQGLEWIGYIDPATAYTESNQKFKDKAILTADRSSNTAF
MYLNSLTSEDSAVYYCARESPRLRRGIYYYAMDYWGQGTT
VTVSSKLSGSASAPKLEEGEFSEARVDIQMTQTPSSLSASLG
DRVTISCRASQDISNYLNWYQQKPDGTVKLLIYYTSRLHSGV
PSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPPTFGAG
TKLELK

Figure 28: MS results, peptide coverage, displayed in yellow are mapped using Scaffold 4 software (Proteomic Software Inc.).

4.6 Western blotting

scFv run on a NuPAGE Bis-Tris 4-12% gel (Life technology) was stained by either Coomassie or silver staining. Coomassie staining makes it possible to estimate the quantity by comparing the stained scFv band with bands of protein with known concentrations, but it is not very sensitive. Silver staining is very sensitive, but does not say much about the amount of protein. Western blotting was used because it is both sensitive and quantitative, although the method usually needs some optimisation to get a good result.

SEC fractions from C1*, C3* and C4* were subjected to SDS-PAGE. The fractions were blotted over to an Immobilon-P membrane. In the initiation trial the scFv were detected with Protein L-HRP using insoluble Tetramethylbenzidine (TMB), thereafter the insoluble TMB was exchanged for SuperSignal West Femto Chemiluminescent Substrate to detect scFvs (Thermo Scientific), the signal was visualised using Amersham Hyperfilm™ ECL in a dark room. When SuperSignal West Femto Chemiluminescent Substrate (Thermo Scientific) was used to detect scFv, a peculiar phenomenon occurred for C1* and C3* (see Figure 29 and Figure 30). Once the substrate was added, a clear yellow glow in full daylight was observed. The membranes were rushed into the darkroom for signal detection, but all the films showed halos around the areas that initially had projected such strong signal. Afterwards, the membranes of C1* and C3* were permanently stained in the areas where the halos were located, making re-blotting impossible (see permanently stained regions at the right hand side of Figure 30). The only normal Western blot was detected from C4* (see Figure 31). The halo phenomenon is probably due to too high protein concentrations. Possibly, the scFv concentration was so high that all the substrate was used up within the halo area, leaving no substrate left for Protein L-HRP. The outer part of the scFv band has a lower scFv concentration and can produce a continuous signal, creating the halo. Another explanation could be that the high scFv concentration packed the scFv so tightly that it was unavailable for Protein L-HRP binding. Consequently, Protein L-HRP would only be able to bind to the scFvs in the outline of the scFv band, where the protein concentration was lower, resulting in a signal distribution shaped like a halo. If there were more time, the western blot would have been repeated, this time with lower protein concentrations for constructs containing the $V_{L.A}$ (C1* and C3*) using the same detection system, and/or with a less sensitive detection system as for example SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific). Why the bands inside the halo were permanently stained is unclear. The permanent stains might be

a product of an unknown side reaction. Another possibility is that the substrate was cleaved so quickly that heat developed, permanently staining the membrane.

Western blotting was also performed on an expression test, the same as discussed in Section 4.4.1, and seen in Figure 19 and Figure 20. The expression test contained OD_{600} normalised whole cells, and medium detected at different times during the expression. The goal was to detect scFv, re-blot and detect the chaperone, by the FkpA's anti-FLAG-tag. This would be very informative since C1*, C3*, C4* and the FkpA are of similar sizes, and the protein bands cannot be separated on a standard NuPAGE Bis-Tris 4-12% gel by SDS-PAGE. FkpA should have been expressed throughout the E. coli cell cycle, while the scFvs should have been expressed after glucose was removed. In the expression test, a band corresponding to scFv or FkpA was present in the whole cell fraction when glucose was removed, but not in the other whole cell fractions, indicating that FkpA is not expressed. In the initial Western blotting of the expression test, only C1* gave a readable signal. It was therefore decided to repeat the experiment, but load a larger number of whole cells on the polyacrylamide gel to increase the protein concentration. Regrettably there was not enough time left at this point to repeat the experiment.

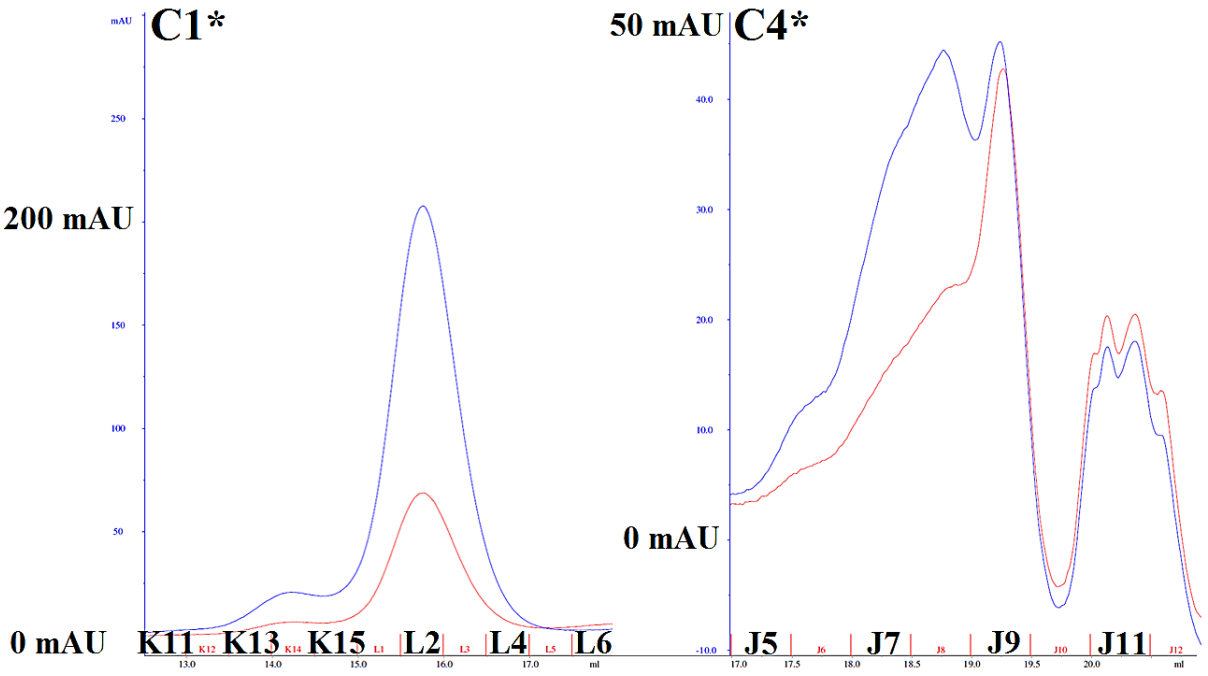

Figure 29: SEC chromatogram of C1* (left) and C4* (right) used for Western blotting. Emphasised fractions for C1* are separated on polyacrylamide gel in Figure 30, and separated in Figure 31 for C4*. *Blue graph: absorption at 280 nm. Red graph: absorption at 250 nm.* The absorbance is low because only a small volume of bacteria was produced

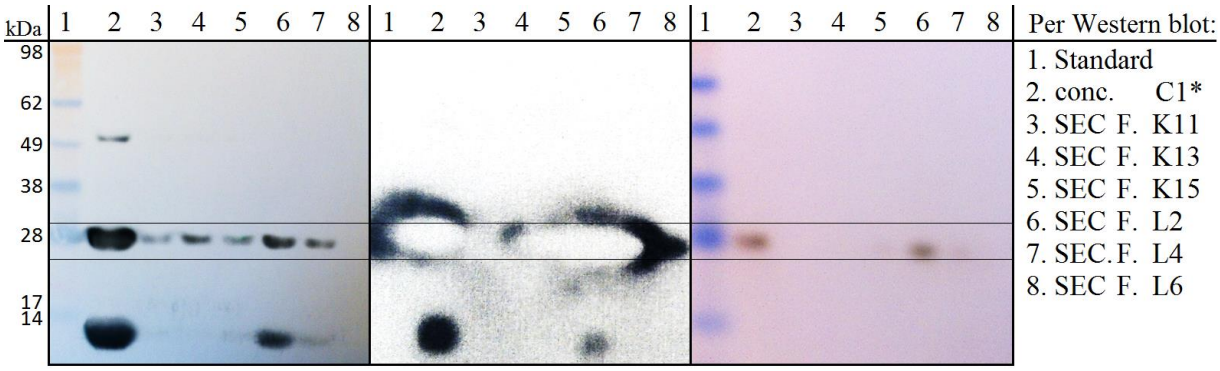

Figure 30: SEC fractions of C1* on Western blot. *Membrane to the left:* Signal developed from insoluble TMB. *Membrane in the middle*: Signal developed with SuperSignal West Femto Chemiluminescent (Thermo Scientific). *Membrane to the right:* Membrane after Chemiluminescent signal developing soaked in PBS for over one week. Concentrated C1*(conc. C1*) was obtained after Protein L affinity chromatography. The right picture is scaled to correspond to the molecular weight markers of the Western blot detected with insoluble TMB.

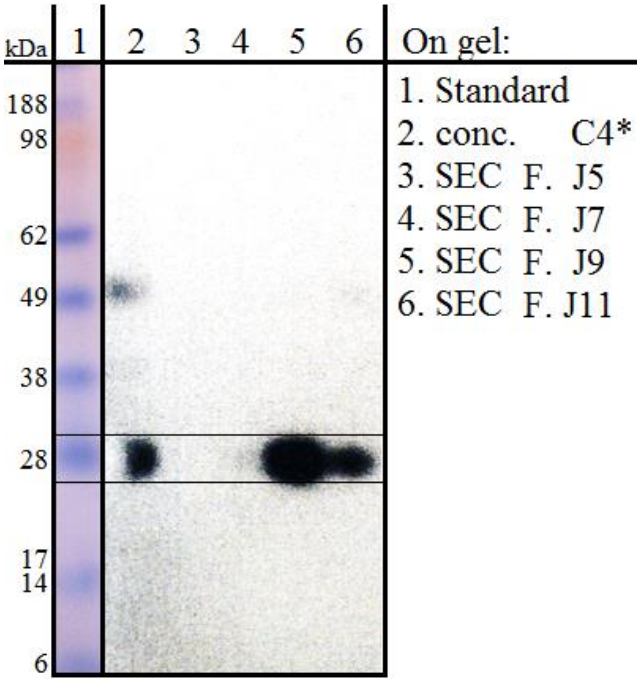

Figure 31: SEC fraction of C4* on Western blot. The Standard originates from the original membrane and is proportionated using the permanently stained band from C3* originating from the same membrane as a guide. Concentrated C4* originates was obtained after Protein L affinity chromatography

4.7 scFv's affinity for NeuGc GM3

4.7.1 Testing the scFv-ganglioside affinity with ELISA

Using indirect ELISA, ganglioside binding was tested using C1*, C3* and C4*. For signal-detection Protein L-HRP was used. A signal was seen for the wells coated with NeuGc GM3, but not for the wells coated with the close relative NeuAc GM3, see Figure 32. This indicates that the removal of the His-tag solved the binding problems. When repeating the ELISAs a decrease in binding affinity was observed over time, first for C4* and then for C1*. Thus, a direct ELISA was performed, comparing the activity of an old and a new batch of C1* and C4* with the same concentration range. A weaker signal was perceived for the old batches compared to the new, demonstrating that the scFvs lose activity over time when stored at 4 °C. An affinity test was performed to see if the scFv was stable after being frozen, and if glycerol should be added. The result showed that there was no loss in affinity after freezing at -80 °C, see Figure 33. The Figure also shows concentration-dependent binding for C1* towards the NeuGc GM3 ganglioside. The decrease in signal when 20% glycerol was added might be related to the viscosity of glycerol, making the sample concentration heterogenic. From that day all proteins were frozen at -80 °C in small aliquots and only fresh batches were used to ensure the highest degree of functionality for the scFvs.

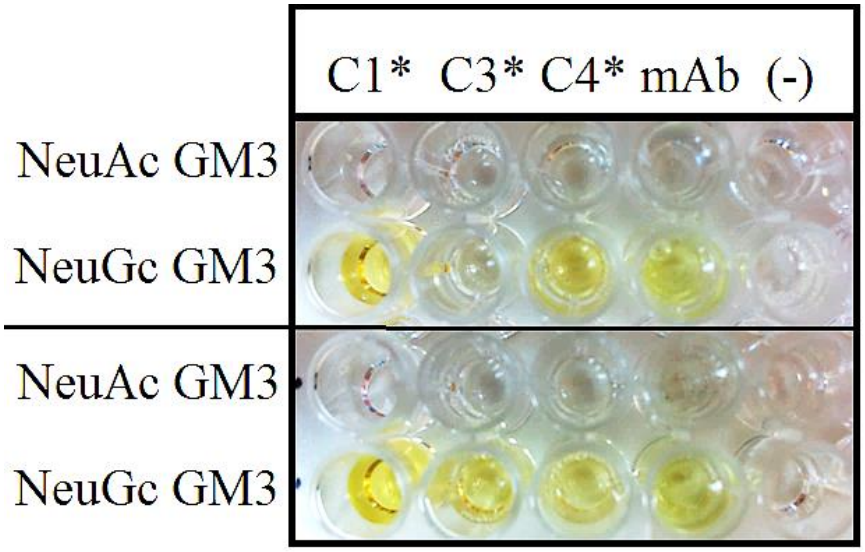

Figure 32: ELISA assay for plates coated with NeuGc GM3 or NeuAc GM3. Binding of the scFvs: C1* C3* and C4* in a concentration of 16 µg/ml are detected with Protein L-HRP in a double repeat. Since the ELISA assays were performed in a lab without ELISA reader, pictures were taken of the plates. The plate shows that the scFvs binds specifically to NeuGc GM3, and has no cross-reactivity towards NeuAc GM3. C3* only displays a weak signal in the upper repetition due to human errors.

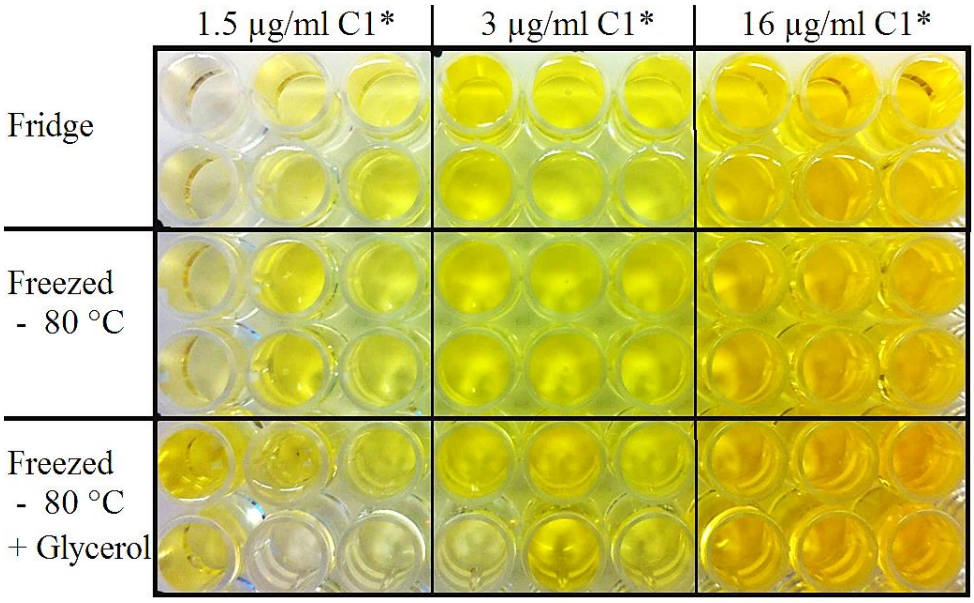

Figure 33: Three aliquots of C1* with at three different concentrations were used to see if the protein was harmed by being frozen at -80 °C with and without 20% glycerol. The aliquots were fresh and stored for 24 hours in either fridge or freezer before being used in the ELISA assay. The affinity is measured at three different concentrations, each consisting of three aliquots, displaying concentration-dependent binding to NeuGc GM3. Since the ELISA assays were performed in a lab without ELISA reader, pictures were taken of the plates. The positive and negative controls are outside the scope of this picture but reported a positive and negative signal, respectively. The result was a concentration-dependent signal, and showed that C1*could be frozen at -80 °C without the addition of glycerol.

4.7.2 Surface plasmon resonance (SPR)

SPR was performed for C1* and C4* in a single-cycle experiment, on a CM5 chip (GE Healthcare) coated with the NeuGc GM3 ganglioside. When ganglioside was immobilised on the CM5 chip, the reference value was 75, if the value had been below 50 the chip would have had to be recoated. The 14F7 mAb was used as a control and displayed the same signal before and after the experiment, verifying that no measurable amount of ganglioside was lost during the experiment. The signal was interpreted by Rune Johansen Forstrøm (Engineer at Oslo University Hospital, Oslo). SPR was only performed for C1* and C4*, as these were the two constructs with the highest diversity in amino acid sequence and available in appropriate concentrations. For C4*, the graph was judged invalid, probably due to protein aggregation. It was possible to obtain a binding constant for C1*, but it required extensive, optimistic fitting. C1* (Figure 34) display heterogeneity and artefacts, it was possible to obtain a binding constant for C1* (~28 µM), but it does not properly fit the data, and is not trustworthy. But, the graph does display binding towards the NeuGc GM3 ganglioside, confirming the ELISA results. The samples used in SPR had been stored at 4 °C for several weeks. Thus, the low

quality data might originate from bad samples (see Section 4.7.1, page 65). To obtain more reliable data, the SPR experiment should be repeated with fresh samples.

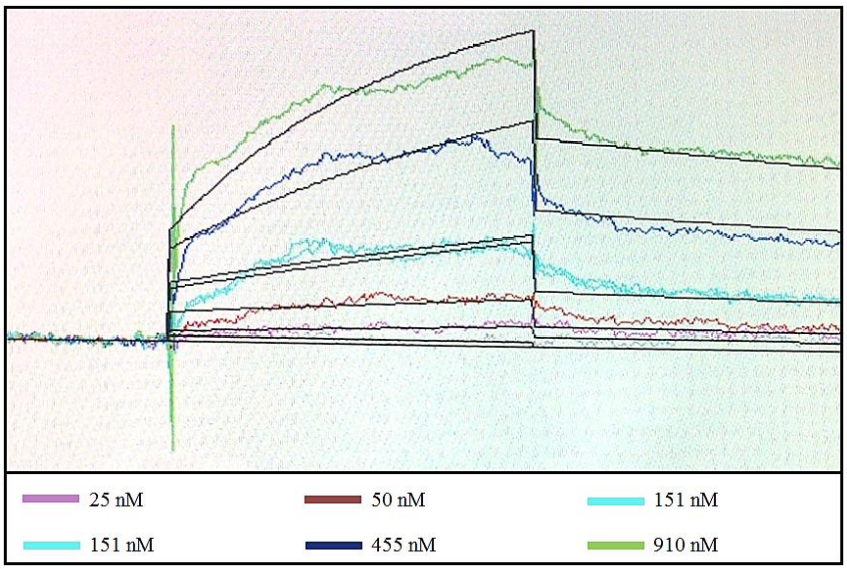

Figure 34: SPR result for C1* binding to NeuGc GM3 in a single-cycle experiment.

4.7.3 ThermoFluor

Since binding of a ligand often stabilises protein structure, we wanted to use ThermoFluor to see if we could detect a shift in the melting temperature when the antigen was added to the scFvs. C1*, C3* and C4* were used at a concentration of 30 µg/ml (believed at the time to be 0.1 mg/ml, see explanation in Section 3.4.2, page 27). The proteins were mixed with increasing concentrations of the soluble NeuGc GM3 derivative called Tricer, or the NeuGc GM3 ganglioside.

In Figure 35, the three scFvs with and without Tricer are plotted in a normalised graph. C1* and C4* are the two scFvs with the highest diversity in the amino acid sequence. C1* consists of L_R and $V_{L.A}$, while C4* consists of L_C combined with V_L. C1* was significantly more stable than C4 *, with a difference in melting temperature of around 20 °C. We could not observe any differences in the C1* and C4* graphs after adding Tricer. It is important to remember that it was no guarantee that Tricer would bind to 14F7 at all, since the consequences of removing a most of the ceramide moiety have never been assessed, due to low available amounts of Tricer. The fact that there were no shifts in the melting curves might indicate that Tricer did not bind to the scFvs, or that binding did not lead to increased stability.

For C3*, consisting of L_C combined with $V_{L.A}$, we saw a small shift in the curve when Tricer was present, indicating higher stability, but the overall melting temperature stayed the same. The cleaved scFv fragments, with an approximate size of 14 kDa, probably dissociate when the temperature is increased. This will make them denature faster, giving them a different melting curve with a lower average melting temperature compared to a sample consisting of pure full-size scFv. A heterogeneous sample used for ThermoFluor, consisting of both cleaved and full-size scFv, will display a melting curve that is in reality an average of the two melting curves. The consequence is that the percentage of fragmented scFv in a sample will affect the resulting melting curve, and the calculated melting temperature. Especially C3* had a high percentage of fragmented scFv, as seen in Figure 26, and the calculated melting temperature can therefore not be trusted. The heterogeneity in the sample might also explain the shoulder present in the C3* curve without Tricer, in Figure 35. Perhaps it is in the shoulder area that the cleaved scFvs are completely denatured, resulting in a short levelling period before the curve continues to rise, due to denatured full-size scFv. In order to obtain trustworthy ThermoFluor data the fragmented scFv has to be removed.

When the scFvs were mixed with the NeuGc GM3 ganglioside, methanol was added to make the ganglioside soluble. The hydrophobic environment decreased the overall melting temperatures for all scFvs. There is a visible difference between the C3* curves with and without NeuGc GM3 in Figure 36. The shoulder shown in Figure 36 for C3* levels out in the presence of NeuGc GM3, resulting in a plateau, before the curve continues to rise. If the perceived shoulder is due to cleaved scFvs, NeuGc GM3 binding has a positive effect on the stability of the cleaved scFvs. Both curves overlap in the main peak, indicating that NeuGc GM3 does not influence full-size scFv in the same manner. If the shoulder originates from the cleaved scFv, the correct melting temperature would be higher. An expected melting curve is modelled into Figure 36 marked by a blue dotted line. In order to obtain valid scFv melting curves, ThermoFluor has to be repeated with samples containing only full-size scFvs. C4* had a small shift in the curve when the ganglioside was present (data not shown) while the curve for C1* together with ganglioside had an invalid fit, probably due to the methanol concentration.

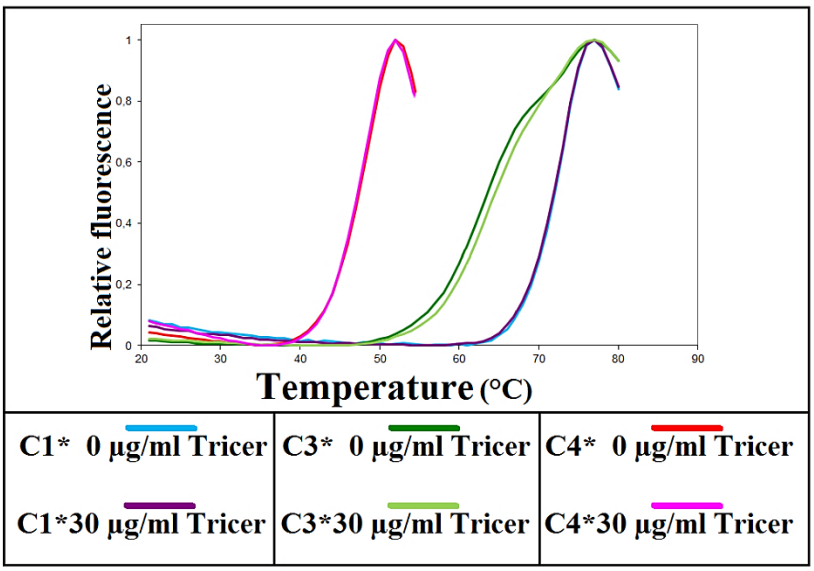

Figure 35: ThermoFluor. C1*, C3*,and C4* have melting temperatures of 71.5 °C, 63.7 °C and 51 °C, respectively.

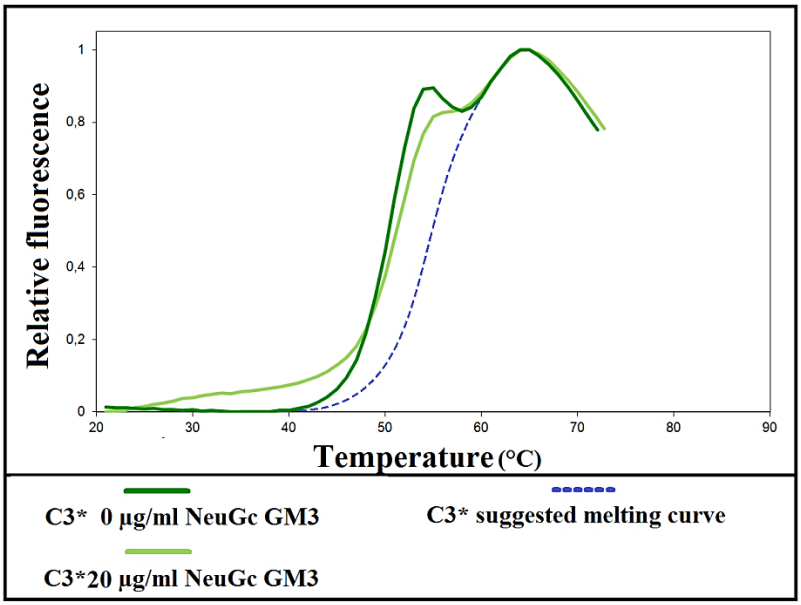

Figure 36: ThermoFluor result of C3* with and without the NeuGc GM3 ganglioside. The shoulder seen at approximately 53 °C levels out in the presence of ganglioside and the melting temperature shifts from 50 °C to 51 °C. The shoulder might originate from cleaved scFv, while the main peak might correspond to full-size scFv. A blue dotted curve is modelled into the graph representing the expected melting curve for purified full-size C3* scFv with a melting temperature of approximately 54 °C.

4.8 Protein Crystallisation

The two most common techniques for determining protein structure are nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography. The methods can be used to find the structure of a free protein or for a protein associated with a binding partner (Blow, 2002). For a protein binding a small ligand, X-ray crystallography might have an advantage compared to NMR for low-affinity binding. Since NMR obtains the structure from proteins in solution, the proteins can move around freely, making it difficult to detect the small ligand's precise binding conformation. In X-ray crystallography it is possible to obtain a higher precision, as the structure is detected from a crystal where the proteins are arranged in a rigid lattice structure.

To obtain a 3D model of a protein using X-ray crystallography it is not enough to just expose it to X-rays, the scattering effect for a single molecule is too weak. A crystal is necessary, because the lattice structure amplifies the scattering signal resulting in a molecule-dependent scattering pattern known as the diffraction pattern (Sheehan, 2009, Blow, 2002). The rate-limiting step in X-ray crystallography is very often protein crystallisation. In order to obtain a crystal, proteins are mixed with different additives and salt concentrations under a variety of pH ranges and temperatures. It is especially challenging to crystallise a full-size antibody, as it is a highly complex and flexible molecule. By using scFv the complexity is reduced increasing the chance for crystal formation.

Purified C1* at a concentration of 3.2 mg/ml in 50 mM Tris-HCl (Chalbiocem) buffer was mixed with the soluble derivative of the NeuGc GM3 ganglioside named Tricer and incubated for 1 hour, see Figure 7 for Tricer's molecular composition. The optimal molar antigen concentration should be 10 to 50 times higher compared to the molar concentration of scFv. Since Tricer was only available in very small amounts, we used approximately 10 times more Tricer compared to scFv before setting up the crystallisation screens.

The crystal screening was a continuing process throughout the master thesis. For initial screening the Oryx4 (Douglas Instruments) was used. One drop had a volume of 0.4 µl, containing 50% reservoir solution and 50% protein. Several commercial kits were used for screening as indicated in the Appendix, Table S7.

After trying out several screening conditions, two hits were detected in the Morpheus screen, C8 and G8 (Molecular dimensions). The conditions are listed in the Appendix Section G: Crystallisation screens, page S31. The compositions of Morpheus screen buffers C8 and G8 were used for setting up an optimisation screen containing several large drops with slight deviations from the original conditions and with differences in protein, screening buffer composition. See the Appendix, Section G: Crystallisation screens Optimisation, crystallisation hits Table S29 and Table S30, page S33 for an example. From the optimised screen, crystals formed readily when the pH was varied between 7 and 8. Two promising crystals also formed in the well for G8, where the only additive was 0.5 M sodium oxamate. The larger of the two crystals was sent to the European Synchrotron and Radiation Facility (ESRF), Grenoble, France, for diffraction experiments, but the crystal did not diffract.

In several wells in the optimisation screen many small crystals formed (see Figure 37 and Figure 38 and the Appendix Section G: Crystallisation screens, Optimisation, crystallisation hits, page S33). A large crystal was desired, as it has better probability to diffract. A more extensive lattice might result in diffraction data with higher intensity spots, giving better resolution. To achieve this, new optimisation screens were prepared using 75% buffer and 25% protein. As in the initial optimisation screen several small crystals formed, but this time some larger crystals were observed. These were sent to the ESRF, Grenoble, France, for diffraction experiments, without success. The crystals did either not diffract, or displayed large spots, indicating salt crystals. An example is seen in Figure 39.

If there were more time available, seeding would be a method that might help form larger scFv-Tricer crystals. The small crystals can act as seeds for further crystal growth, resulting in larger crystals. If more time and protein were accessible, the scFvs should be crystallised alone, without the addition of Tricer, since the antigen is available only in small amounts. If diffracting crystals are obtained, crystals formed under the same or similar conditions could be soaked with Tricer. If successful, this procedure would make it possible to obtain the structure of scFv with and without Tricer, giving the ability to investigate potential conformation differences between the crystals, and minimising the use of Tricer.

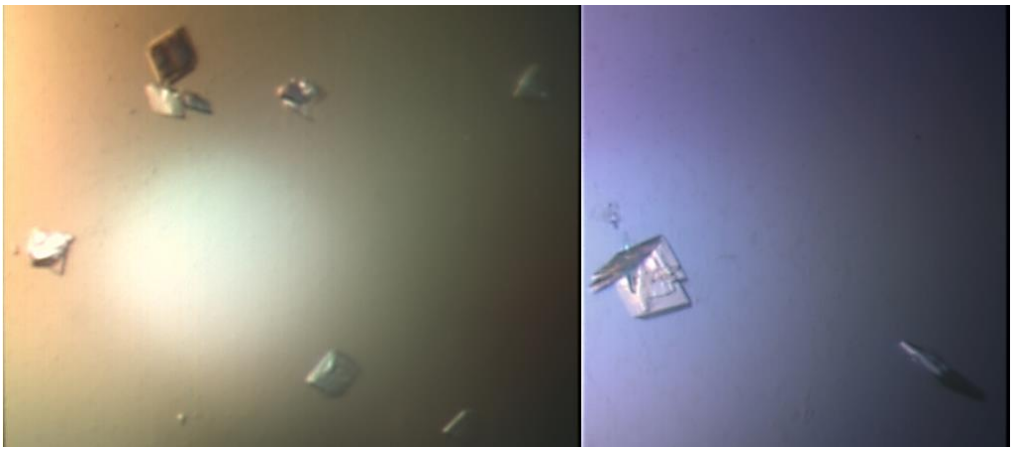

Figure 37: Crystals formed in C8 Morpheus screen. Left: Crystals formed during initial screening conditions, Right: Crystal produced during optimised screening conditions. *The size proportions for the two pictures are not correlated.*

Figure 38: Crystals formed in G8 Morpheus screen. Left: Crystals formed during initial screening conditions, Right: Crystal produced during optimised screening conditions. *The size proportions for the two pictures are not correlated.*

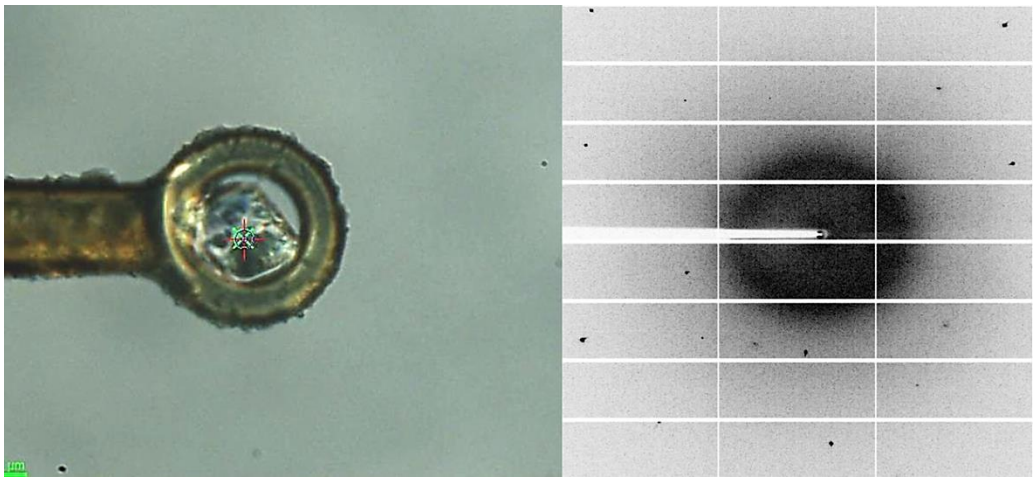

Figure 39: Left: A crystal in a crystal loop. Right: Diffraction of the crystal, the well-separated, large spots indicates salt crystals.

4.9 Comparison of the scFv constructs

The stability index seen in Table 4 indicates that the presence of a hydrophilic His-tag helped to solubilise the scFvs, making them more stable. The pI is in a range between 6.7 for C1* and 8.8 for C4* (see Table 4). The linker is mainly responsible for the difference in pI, since L_R has a total charge of minus one, while L_C has a total charge of plus two (see aligned linker sequence in Figure 40). When removing the His-tag from the scFvs a decrease in the yield was perceived for C4*, consisting of the original 14F7's V_L. The yield decreased from 0.3 mg to 0.1 mg per litre bacterial culture, when compared to C4 with C4*, respectively (C4 has a His-tag and was purified using IMAC) while C2* was not obtainable in sufficient amounts for effective isolation. The yield for C1* and C3*, both consisting of $V_{L.A}$, stayed approximately at the same level, with a yield of 5 and 3 mg per litre bacterial culture, respectively. It can therefore be concluded that that the scFvs containing $V_{L.A}$ (C1*/C3*) compared to scFvs containing V_L (C2*/C4*), express better in a prokaryotic system. The decrease in yield for scFvs containing the V_L might have originated in reduced solubility when the His-tag was removed, or because the V_L has low affinity to Protein L and was therefore not efficiently purified. Since no scFv was detected the second time over the Protein L column a possibility would be that Protein L only associated with C2* and C4*, explaining the decrease in yield.

All ELISA experiments were detected by visual inspection, since no ELISA reader was available at the laboratory and the plates could not be frozen. Therefore, the difference in affinity towards the NeuGc GM3 ganglioside was not estimated quantitatively, but binding was detected for all four scFvs confirming conserved affinity towards the NeuGc GM3 ganglioside. In ThermoFluor, C1* containing $V_{L.A}$ had ~20 °C higher melting temperature compared to C4* containing V_L. It can therefore be concluded that scFvs containing $V_{L.A}$ are more stable.

Table 4: The scFv constructs with and without His-tag. The information was obtained using ProtParam, ExPASy (Swiss Institute of Bioinformatics) (Artimo et al., 2012).

	C1	C1*	C2	C2*	C3	C3*	C4	C4*
amino acids	268	255	268	255	268	255	268	255
weight (kDa)	30.0	28.3	30.0	28.3	29.9	28.2	29.96	28.2
pI	6.7	6.7	7.1	7.63	8.2	8.5	8.6	8.8
ext. coefficient	54570	53080	55600	54110	54570	53080	55600	54110
absorption coefficient	1.82	1.88	1.86	1.90	1.83	1.88	1.85	1.92
instability index (Below 40 indicate stable proteins)	36.9	38.6	49.3	51.67	37.5	39.3	50.0	52.4

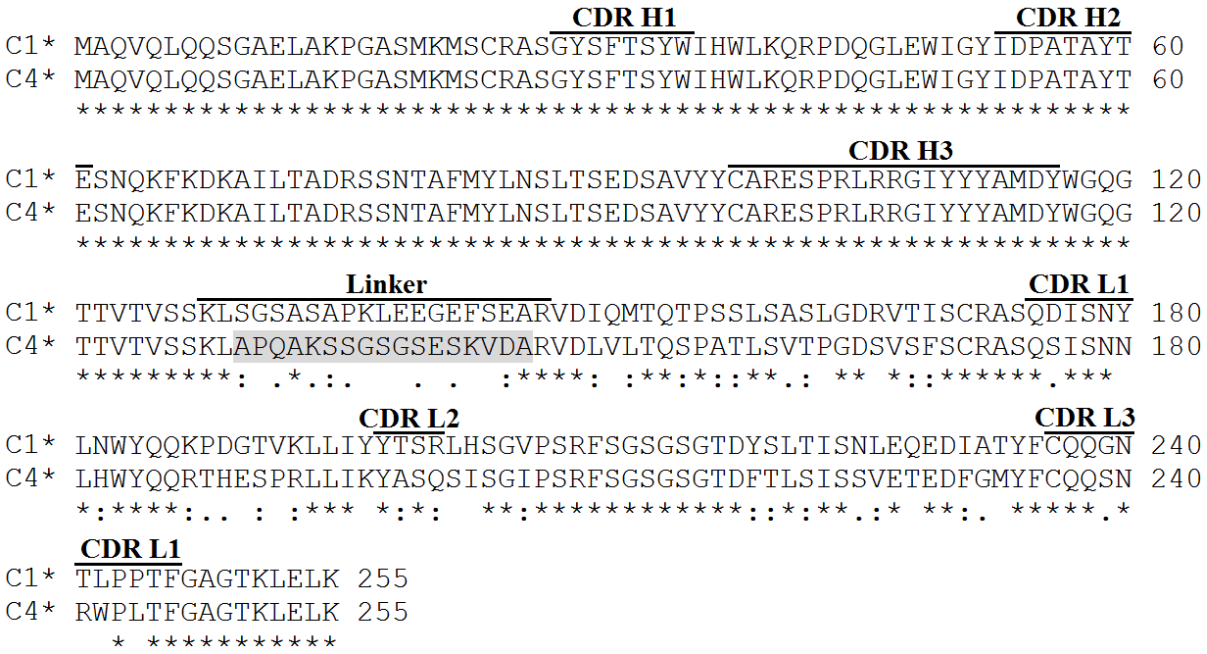

Figure 40: ClustalW2, EMBL-EBI alignment of the amino acid sequence of C1* and C4*. The CDRs and the linkers are marked with a black line above the amino acid sequence, the CDRs are localised using IMGT/V-QUEST. The original linker from Cuba ($L_{C.O}$) is highlighted in light grey (Larkin et al., 2007, Brochet et al., 2008, Giudicelli et al., 2011).

5 Methodological considerations and future perspectives

5.1 Optimising the scFv constructs

In the beginning, it was not known if it was possible to produce one, several, or all the scFv constructs. When three out of four constructs were expressed in sufficient amounts to continue with further experiments, no additional effort was made to optimise the C2/C2* expression. The sequence analysis shows that the C2/C2* constructs are correct, but it might be advisable to subclone the construct once more in case there is something wrong within the vector.

To prevent scFv cleavage the cause should be found. If it was cleaved by an unspecific protease, it might be possible to prevent the cleaving by isolating the scFvs earlier without incubating the cells overnight. To investigate this, scFvs can be isolated from the cells at different time points and, and after normalising the samples and SDS-PAGE the ratio of full-size scFvs vs cleaved scFvs can be compared with Western blotting. Also, by looking at the expression tests (see Figure 19 and Figure 20) it is unclear if the periplasmic chaperone FkpA is expressed at all since it is not visible on the polyacrylamide gel. If this is the case, the introduction of FkpA in *E. coli* might help solubilising the scFv, increasing the yield and minimising scFv cleavage. One method would be to subclone the construct once more into pFKPEN, or to change to the pFKPEI vector, where the FkpA chaperone is transcribed as the second open reading frame in a discistronic constellation with the scFv, where both are controlled by the lacPO (Gunnarsen et al., 2010).

In the addition to FkpA expression, the yield might be increased by reintroducing a, both in respect to solubility, purification and to be able to detect the binding affinity quantitatively. Since an N-terminus tag probably disrupt antigen binding the tag should preferably be placed at the scFvs C-terminus. A C-terminal tag cannot be removed by TEVp without leaving a flexible peptide tail at the C-terminus, possibly disturbing crystal formation. If possible, an alternative would be to introduce a recognition site for a different protease, which cleaves such that a minimum of amino acids are left at the C-terminus. Tagged proteins can still crystallise, although flexible elements are generally not beneficial for crystal formation.

Therefore, the choice of solution depends on the ultimate target: to characterise the different scFvs or to optimise the scFvs for crystallisation studies.

5.2 How the presence of Tricer effects the experiments

When ThermoFluor and crystallisation screening was executed, Tricer was used as a replacement for the NeuGc GM3 ganglioside. Tricer contains part of the ceramide moiety and was made soluble by the addition of an acetamide group. It is not known how the partial removal of the ceramide portion affects antigen binding. The binding affinity could be assessed using ELISA or SPR, but since only small amounts of Tricer were available, it was not possible to conduct any quantitative experiments. Since we do not know if Tricer is a worthy replacement of the NeuGc GM3 ganglioside, ThermoFluor cannot be used as a method to assess if the NeuGc GM3 ganglioside stabilise the scFvs. It would also be beneficial to produce scFv crystals without the addition of Tricer, and rather soak in Tricer afterwards, when diffracting crystals are obtained.

5.3 Ganglioside affinity estimated by ELISA

Since there was no ELISA reader available at the laboratory, and the plates could not be read after having been frozen, all-important ELISAs should have been repeated in a laboratory containing an ELISA reader giving quantitative data. But since there is a probability that Protein L has a higher affinity towards scFvs containing $V_{L.A}$ (C1*/C3*), compared to scFvs consisting of V_L (C2*/C4*), another detection method would be advisable so that a more trustworthy quantitative comparison of the scFvs affinity towards the NeuGc GM3 ganglioside could be executed. In order to detect the scFvs a tag would be beneficial. Since the original N-terminally His-tag probably disturbed antigen binding, an alternative would be to place the His-tag at the C-terminus. This would create the opportunity to purify the scFvs using IMAC and to detect the NeuGc GM3 affinity with an anti-His-tag antibody.

5.4 General thoughts for future experimental procedures

Several of the experiments should be repeated to give more weight to the results, and to rule out uncertainties: ThermoFluor and SPR should both be repeated with fresh, homogenous samples, as heterogeneity can affect the results. SPR should preferable be repeated in a multiple-cycle mode instead of single-cycle mode. In the single-cycle mode, all protein samples were injected continuously, giving possible overlap, while in a multiple-cycle mode the channels would be cleaned with NaOH between sample injections. An experiment that might be recommended for further stability characterisation is circular dichroism (CD). CD gives information about the secondary structure. By comparing the CD spectra from the scFvs with known scFv spectra of clean scFv samples, it is possible to detect if there is a higher or lower degree of unstructured elements presence. In order to get a reliable result, the sample must consist of only full-size scFv, without any cleaved fragments, since the cleaved fragments are probably less structured.

To increase the chance to obtain a diffracting crystal, it would be beneficial to undertake crystallisation with scFv variants containing different variable light chains. Future work could be narrowed down to C1 or C3 combined with C4, all with a C-terminal tag. C1 containing L_R and $V_{L.A}$ give the best yield, and has the highest amino acid diversity compared to C4 with L_C and V_L. But C3 with L_C and $V_{L.A}$ has the same linker as C4, making it easier to verify the effect of the light chain, without the presence of a second parameter, the linker. If the expression problem with C2 was solved, and the scFv gave the same or a better yield compared to C4, then C1 and C2 would be the optimal constructs to use for further experiments because of C1's high expression yield.

6 Summary and conclusion

Cancer is a disease that is treated by a range of different methods, where the most common ones also weaken the healthy cells within the body. This makes cancer medicine a research field with a high potential for improvement. 14F7 mAb has the nickname The Assassin, because it specifically kills tumours without harming healthy human cells. In order to fully exploit the favourable qualities of 14F7, it is important to understand how it binds to its antigen, the tumour specific NeuGc GM3 ganglioside. This knowledge can be used to generate the next generation of ganglioside-specific antibodies and provide general insight into how antibodies interact with gangliosides. Different methods have been tested to understand the interaction between the antibody and its antigen (Abshiru, 2010, Rojas et al., 2012, Krengel et al., 2004). In 2004, the structure of Fab was solved by X-ray crystallography, but the crystals are difficult to reproduce (Krengel et al., 2004). To obtain an experimentally valid model of the complex, the first step would be to reproduce crystals, then soak in NeuGc GM3 or co-crystallise the complex. We decided to initiate an alternative strategy by producing 14F7 scFvs.

In this thesis, we designed four scFv versions of 14F7, which were expressed and isolated from the periplasm of *E. coli*. Since scFvs are less complex than Fabs, we hoped that we would improve the chances of obtaining crystals. We managed to grow crystals, but since they did not diffract, further optimisation is needed. scFv consists of a variable heavy chain, and a variable light chain, connected with a linker. In order to improve the chances of producing functional scFvs, the variable heavy chain of 14F7was combined with one of two different linkers, and either the original or an alternative variable light chain. All constructs proved to bind to the NeuGc GM3 ganglioside, as shown by SPR and ELISA. The scFvs consisting of the alternative variable light chain had a significant higher yield compared to the scFvs with the original 14F7 variable light chain. The scFvs with the alternative light chain also had an increased stability as shown with ThermoFluor.

This thesis sets the foundation for further scFv characterisation, and provides the tools to produce 14F7 scFv crystals. We believe, and hope, that diffracting crystals are now within reach, and that this thesis will contribute to solve the mystery of how 14F7 binds and interacts with its tumour-specific antigen.

References

Abshiru, N. 2010. *Oxidative footprint and mass spectrometry based structural characterization of mAb 14F7-NeuGc Gm3 ganglioside complex.* Master, University Of Oslo.

Artimo, P., et al. 2012. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research,* 40**,** W597-W603.

Baneyx, F. & Georgiou, G. 1990. In vivo degradation of secreted fusion proteins by the Escherichia coli outer membrane protease OmpT. *Journal of Bacteriology,* 172**,** 491-494.

Bardor, M., Nguyen, D. H., Diaz, S. & Varki, A. 2005. Mechanism of Uptake and Incorporation of the Non-human Sialic Acid N-Glycolylneuraminic Acid into Human Cells. *Journal of Biological Chemistry,* 280**,** 4228-4237.

Batra, S. K., Jain, M., Wittel, U. A., Chauhan, S. C. & Colcher, D. 2002. Pharmacokinetics and biodistribution of genetically engineered antibodies. *Current Opinion in Biotechnology,* 13**,** 603-608.

Bhat, N. M., et al. 1997. Rapid cytotoxicity of human B lymphocytes induced by VH4-34 (VH4.21) gene-encoded monoclonal antibodies, II. *Clinical & Experimental Immunology,* 108**,** 151-159.

Bird, R. E. H., Karl D; Jacobson, James W; Johnson, Syd; Kaufman, Bennett M; Et Al. 1988. Single-Chain Antigen-Binding Proteins. *Science,* 242**,** 423-426.

Blanco, R., et al. 2013. Immunoreactivity of the 14F7 Mab Raised against N-Glycolyl GM3 Ganglioside in Primary Lymphoid Tumors and Lymph Node Metastasis. *Pathology research international,* 2013**,** 920972.

Blanco, R., et al. 2012. Immunoreactivity of the 14F7 Mab (Raised against N-Glycolyl GM3 Ganglioside) as a Positive Prognostic Factor in Non-Small-Cell Lung Cancer. *Pathology research international,* 2012**,** 235418.

Blanco, R., et al. 2011a. Immunoreactivity of the 14F7 Mab Raised against N-Glycolyl GM3 Ganglioside in Epithelial Malignant Tumors from Digestive System. *ISRN gastroenterology,* 2011**,** 645641.

Blanco, R., et al. 2011b. Immunohistochemical Reactivity of the 14F7 Monoclonal Antibody Raised against N-Glycolyl GM3 Ganglioside in Some Benign and Malignant Skin Neoplasms. *ISRN dermatology,* 2011**,** 848909.

Blow, D. 2002. *Outline of Crystallography for Biologists,* Oxford, OXFORD University Press.

Bothmann, H. & Plückthun, A. 2000. The Periplasmic Escherichia coli Peptidylprolyl cis,trans-Isomerase FkpA: I. Increased Functional Expresion of Antibody Fragments with and without cis-Prolines. *Journal of Biological Chemistry,* 275**,** 17100-17105.

Brochet, X., Lefranc, M.-P. & Giudicelli, V. 2008. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Research,* 36**,** W503-W508.

Brüggemann, M., et al. 1989. A repertoire of monoclonal antibodies with human heavy chains from transgenic mice. *Proceedings of the National Academy of Sciences,* 86**,** 6709-6713.

Buchner J., P. I., Brinkmann U. 1992. A method for increasing the yield of properly folded recombinant fusion proteins: single-chain immunotoxins from renaturation of bacterial inclusion bodies. *Analytical Biochemistry,* 205**,** 263-270.

Carr A., M. A., Mazorra Z., Vázquez A. M., Alfonso M., Mesa C., Rengifo E., Pérez R., Fernández L. E. 2000. A Mouse IgG1 Monoclonal Antibody Specific for N-Glycolyl GM3 Ganglioside Recognized Breast and Melanoma Tumors *Hybridoma,* 19**,** 241-247.

Carr A., M. C., Arango M. C., Vázquez A. M., Fernández L. El 2002. *In Vivo* and *In Vitro* Anti-Tumor Effect of 14F7 Monoclonal Antibody. *Hybridoma and Hybridomics,* 21**,** 6.

Chou, H.-H., et al. 2002. Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proceedings of the National Academy of Sciences,* 99**,** 11736-11741.

Chou, H.-H., et al. 1998. A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. *Proceedings of the National Academy of Sciences,* 95**,** 11751-11756.

Cohen-Solal, J. F. G., Cassard, L., Fridman, W.-H. & Sautès-Fridman, C. 2004. Fc γ receptors. *Immunology Letters,* 92**,** 199-205.

Croce, C. M. 2008. Oncogenes and Cancer. *New England Journal of Medicine,* 358**,** 502-511.

David, D. & Zouali, M. 1995. Variable region light chain genes encoding human antibodies to HIV-1. *Molecular Immunology,* 32**,** 77-88.

Davies, D. R. & Chacko, S. 1993. Antibody structure. *Accounts of Chemical Research,* 26**,** 421-427.

Dudley, M. E. & Rosenberg, S. A. 2003. Adoptive-cell-transfer therapy for the treatment of patients with cancer. *Nature Reviews Cancer,* 3**,** 666-675.

Ehrenstein, M. R. & Notley, C. A. 2010. The importance of natural IgM: scavenger, protector and regulator. *Nat Rev Immunol,* 10**,** 778-786.

Fenton, R. G. & Longo, D. L. 1995. Genetic Instability and Tumor Cell Variation: Implications for Immunotherapy. *Journal of the National Cancer Institute,* 87**,** 241-243.

Gagneux, P. & Varki, A. 1999. Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology,* 9**,** 747-755.

Gaston, K., Bell, A., Kolb, A., Buc, H. & Busby, S. 1990. Stringent spacing requirements for transcription activation by CRP. *Cell,* 62**,** 733-743.

Geny, B. & Popoff, M. R. 2006. Bacterial protein toxins and lipids: pore formation or toxin entry into cells. *Biology of the Cell,* 98**,** 667-678.

Gilbert, H. F. 1990. Molecular and cellular aspects of thiol-disulfide exchange. *Advances in enzymology and related areas of molecular biology,* 63**,** 96-172.

Gilbert, W., Maxam A. 1973. The Nucleotide Sequence of the lac Operator. *Proc Natl Acad Sci USA,* 70**,** 3581-3584.

Giudicelli, V., Brochet, X. & Lefranc, M.-P. 2011. IMGT/V-QUEST: IMGT Standardized Analysis of the Immunoglobulin (IG) and T Cell Receptor (TR) Nucleotide Sequences. *Cold Spring Harbor Protocols,* 2011**,** pdb.prot5633.

Glockshuber, R., Malia, M. , Pfitzinger, I., Plückthun A. 1990. A comparison of strategies to stabilize immunoglobulin Fv-fragments. *Biochemistry,* 29**,** 1362-1369.

Glockshuber, R., Schmidt T., Plückthun A. 1992. The disulfide bonds in antibody variable domains: effects on stability, folding in vitro, and functional expression in Escherichia coli. *Biochemistry,* 31**,** 1270-1279.

Goebl, N. A., et al. 2008. Neonatal Fc Receptor Mediates Internalization of Fc in Transfected Human Endothelial Cells. *Molecular Biology of the Cell,* 19**,** 5490-5505.

Graille, M., et al. 2001. Complex between Peptostreptococcus magnus Protein L and a Human Antibody Reveals Structural Convergence in the Interaction Modes of Fab Binding Proteins. *Structure,* 9**,** 679-687.

Gunnarsen, K., et al. 2010. Periplasmic expression of soluble single chain T cell receptors is rescued by the chaperone FkpA. *BMC Biotechnology,* 10**,** 8.
Hakomori, S.-I. 2002. The glycosynapse. *Proceedings of the National Academy of Sciences,* 99**,** 225-232.
Hardy, S. J. S., et al. 1993. Recognition of Ligands by SecB, a Molecular Chaperone Involved in Bacterial Protein Export. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences,* 339**,** 343-354.
Irie, A., Koyama, S., Kozutsumi, Y., Kawasaki, T. & Suzuki, A. 1998. The Molecular Basis for the Absence ofN-Glycolylneuraminic Acid in Humans. *Journal of Biological Chemistry,* 273**,** 15866-15871.
Jefferis, R., Lund, J., Mizutani, H., Nakagawa, H., Kawazoe, Y., Arata, Y., Takahashi, N. 1990. A comparative study of the N-linked oligosaccharide structures of human IgG subclass proteins. *Biochemical and Biophysical Research Communications,* 268**,** 529-537.
Jones, C. H., et al. 2002. Escherichia coli DegP Protease Cleaves between Paired Hydrophobic Residues in a Natural Substrate: the PapA Pilin. *Journal of Bacteriology,* 184**,** 5762-5771.
Kashmiri, S. V. S., De Pascalis, R., Gonzales, N. R. & Schlom, J. 2005. SDR grafting—a new approach to antibody humanization. *Methods,* 36**,** 25-34.
Kastern, W., Sjöbring, U. & Björck, L. 1992. Structure of peptostreptococcal protein L and identification of a repeated immunoglobulin light chain-binding domain. *Journal of Biological Chemistry,* 267**,** 12820-5.
Kiefhaber, T., Rudolph, R., Kohler, H. H., Buchner, J. 1991. Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation. *Biotechnology (N Y),* 9**,** 825-829.
Kipriyanov, S. M., Moldenhauer, G. & Little, M. 1997. High level production of soluble single chain antibodies in small-scale Escherichia coli cultures. *Journal of Immunological Methods,* 200**,** 69-77.
Knappik, A. & Plückthun, A. 1995. Engineered turns of a recombinant antibody improve its in vivo folding. *Protein Engineering,* 8**,** 81-89.
Kohler, G. & Milstein, C. 1975. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature,* 256**,** 495-497.
Krengel, U., et al. 2004. Structure and Molecular Interactions of a Unique Antitumor Antibody Specific for N-Glycolyl GM3. *Journal of Biological Chemistry,* 279**,** 5597-5603.
Larkin, M. A., et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-2948.
Løset, G. Å., Lunde, E., Bogen, B., Brekke, O. H. & Sandlie, I. 2007. Functional phage display of two murine α/β T-cell receptors is strongly dependent on fusion format, mode and periplasmic folding assistance. *Protein Engineering Design and Selection,* 20**,** 461-472.
Majno, G. & Joris, I. 1995. Apoptosis, oncosis, and necrosis. An overview of cell death. *The American journal of pathology,* 146**,** 3-15.
Malykh, Y. N., Schauer, R. & Shaw, L. 2001. N-Glycolylneuraminic acid in human tumours. *Biochimie,* 83**,** 623-634.
Marquina, G., et al. 1996. Gangliosides Expressed in Human Breast Cancer. *Cancer Research,* 56**,** 5165-5171.
Muchmore, E. A., Diaz, S. & Varki, A. 1998. A structural difference between the cell surfaces of humans and the great apes. *American Journal of Physical Anthropology,* 107**,** 187-198.

Nelson, D. L., Cox, M. M. (ed.) 2008. *Principles of biochemistry* W. H. Freeman and Companey.
Oliva, J., et al. 2006. Clinical evidences of GM3 (NeuGc) ganglioside expression in human breast cancer using the 14F7 monoclonal antibody labelled with 99mTc. *Breast Cancer Research and Treatment,* 96**,** 115-121.
Parish, C. R. 2003. Cancer immunotherapy: The past, the present and the future[ast]. *Immunology and Cell Biology,* 81**,** 106-113.
Poljak, R. J., Amzel, L. M. , Avey, H. P., Chen, B. L. , Phizackerley, R. P., Saul F. 1973. Three-Dimensional Structure of the Fab′ Fragment of a Human Immunoglobulin at 2.8-Å Resolution. *Proceedings of the National Academy of Sciences of the United States of America,* 70**,** 3305-3310.
Pugsley, A. P. 1993. The complete general secretory pathway in gram-negative bacteria. *Microbiological Reviews,* 57**,** 50-108.
Ramm, K. & Plückthun, A. 2000. The Periplasmic Escherichia coli Peptidylprolyl cis,trans-Isomerase FkpA: II. Isomerase-Independent Chaperon Activity In Vitro. *Journal of Biological Chemistry,* 275**,** 17106-17113.
Ramm, K. & Plückthun, A. 2001. High enzymatic activity and chaperone function are mechanistically related features of the dimeric E. coli peptidyl-prolyl-isomerase FkpA. *Journal of Molecular Biology,* 310**,** 485-498.
Rodríguez, M., et al. 2007. Insights into the immunogenetic basis of two ganglioside-associated idiotypic networks. *Immunobiology,* 212**,** 57-70.
Rojas, G., Pupo, A., Gómez, S., Krengel, U. & Moreno, E. 2012. Engineering the Binding Site of an Antibody against N-Glycolyl GM3: From Functional Mapping to Novel Anti-ganglioside Specificities. *ACS Chemical Biology,* 8**,** 376-386.
Rojas, G., et al. 2004. Light-chain shuffling results in successful phage display selection of functional prokaryotic-expressed antibody fragments to N-glycolyl GM3 ganglioside. *Journal of Immunological Methods,* 293**,** 71-83.
Roque-Navarro, L., et al. 2008. Anti-ganglioside antibody-induced tumor cell death by loss of membrane integrity. *Molecular Cancer Therapeutics,* 7**,** 2033-2041.
Röthlisberger, D., Honegger, A. & Plückthun, A. 2005. Domain Interactions in the Fab Fragment: A Comparative Evaluation of the Single-chain Fv and Fab Format Engineered with Variable Domains of Different Stability. *Journal of Molecular Biology,* 347**,** 773-789.
Sapriel, G., Wandersman, C. & Delepelaire, P. 2002. 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,* 277**,** 6726-6732.
Schaefer, J. V. & Plückthun, A. 2012. Transfer of engineered biophysical properties between different antibody formats and expression systems. *Protein Engineering Design and Selection,* 25**,** 485-506.
Schirrmann, T., Al-Halabi, .L, Dübe,L S., Hust, M. 2008. Production systems for recombinant antibodies. *Frontiers in Bioscience,* 1**,** 4876-94.
Scursoni, A. M., et al. 2011. Detection of N-glycolyl GM3 ganglioside in neuroectodermal tumors by immunohistochemistry: an attractive vaccine target for aggressive pediatric cancer. *Clinical & developmental immunology,* 2011**,** 245181.
Sheehan, D. 2009. *Physical Biochemistry,* Chichester, West Sussex, John Wilson & Sons Ltd.
Skerra, A., Plückthun, A. 1988. Assembly of a functional immunoglobulin Fv fragment in Escherichia coli. *Science,* 240**,** 1038-1041.
Smith, S. L. 1996. Ten years of Orthoclone OKT3 (muromonab-CD3): a review. *Journal of Transplant Coordination,* 6**,** 109-121.

Stoler, D. L., et al. 1999. The onset and extent of genomic instability in sporadic colorectal tumor progression. *Proceedings of the National Academy of Sciences,* 96**,** 15121-15126.

Stults, C. L. M., Sweeley, C. C. & Macher, B. A. 1989. [14] Glycosphingolipids: structure, biological source, and properties. *In:* Victor, G. (ed.) *Methods in Enzymology.* Academic Press.

Takkinen, K., et al. 1991. An active single-chain antibody containing a cellulase linker domain is secreted by Escherichia coli. *Protein Engineering,* 4**,** 837-841.

Tangvoranuntakul, P., et al. 2003. Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proceedings of the National Academy of Sciences,* 100**,** 12045-12050.

Varki, A. 2001. N-glycolylneuraminic acid deficiency in humans. *Biochimie,* 83**,** 615-622.

Vázquez, A. M., et al. 1995. Generation of a murine monoclonal antibody specific for N-glycolylneuraminic acid-containing gangliosides that also recognizes sulfated glycolipids. *Hybridoma,* 14**,** 551-556.

Voet, D. & Voet, J. G. 2011. *Biochemstry*, John Wiley & Sons, INC.

White, C. B., Chen, Q., Kenyon, G. L. & Babbitt, P. C. 1995. A Novel Activity of OmpT.: PROTEOLYSIS UNDER EXTREME DENATURING CONDITIONS. *Journal of Biological Chemistry,* 270**,** 12990-12994.

Wörn, A. & Plückthun, A. 2001. Stability engineering of antibody single-chain Fv fragments. *Journal of Molecular Biology,* 305**,** 989-1010.

Ying, B.-W., Taguchi, H., Ueda, H. & Ueda, T. 2004. Chaperone-assisted folding of a single-chain antibody in a reconstituted translation system. *Biochemical and Biophysical Research Communications,* 320**,** 1359-1364.

Yokota, T., Milenic, D. E., Whitlow, M. & Schlom, J. 1992. Rapid Tumor Penetration of a Single-Chain Fv and Comparison with Other Immunoglobulin Forms. *Cancer Research*, 52**,** 3402-3408.

Young-Pearse, T. 2009. *Protein Markers* [Online]. Available: http://youngpearselab.bwh.harvard.edu/Young-Pearse_Lab/Protocols_-_Protein/Entries/2009/8/23_Protein_Markers.html.

7 Appendix

Section A: Materials

Table S1: Reagents

Chemicals	Vendor
β-mercaptoethanol	Sigma
λ*Hind*III, molecular weight marker	Fermentas
Φ*X*174, molecular weight marker	New England Biolabs
1 kbp DNA ladder	New England Biolabs
100 bp DNA ladder	New England Biolabs
14F7 mAb form collaboration partners	CIM
3',5,5'-tetramethylbenzidine (TMB), insoluble	Chalbiocem
3',5,5'- tetramethylbenzidine (TMB), soluble	Chalbiocem
agarose	Lonza
Acetic Acid	Merck
Alkaline Phosphatase, Calf Intestinal, (CIP)	New England Biolabs
Anti-poly Histidine-Alkaline phosphatase antibody	Sigma
Ampicillin	AppliChem
Bio-Rad Protein Assay	Bio-Rad Laboratories
Bovine serum albumin (BSA, 100x)	New England Biolabs
Buffer #3, 10x	New England Biolabs
Buffer #3.1, 10x	New England Biolabs
c*O*mplete inhibitor	Roche
Coomassie Brilliant Blue G250	Amersham Biosciences
Dibasic sodium phosphate (Na_2HPO_4)	G-Biosciences
Dithiothreitol (DTT)	Bio-Rad Laboratories

DNA constructs	Invitrogen
DNA Loading Dye Solution (6x)	Lonza
DNase	AppliChem
dNTP mix	Invitrogen
E. coli XL1-Blue	Stratagene
Ethanol (Absolut Prima)	Arcus
Ethylenediaminetetraacetic acid (EDTA)	Fluka
Formaldehyde (CH_2O)	Prolabo
D-(+)-Glucose	Sigma
Glycerol	Prolabo
Glycine	Sigma
HBS-EP buffer	GE Healthcare
HBS-N buffer	GE Healthcare
*Hind*III restriction enzyme	New England Biolabs
Hydrogen-Chloride (HCl)	Merck
Imidazole	Sigma
Isopropyl-β-D-Thiogalactopyranoside (IPTG)	Sigma
Ligase buffer, 10x	New England Biolabs
Lysozyme	Sigma
Methanol	Prolabo
MES SDS Running buffer, 20x	Invitrogen
Milli-Q H_2O (MQ- H_2O)	Millipore
*Mlu*I restriction enzyme	New England Biolabs
Mono Sodium phosphate (NaH_2PO_4)	Fluka
NeuGc GM3, from collaboration partners	CIM
*Nco*I restriction enzyme	New England Biolabs

NuPAGE loading buffer (4x)	Invitrogen
Nickel(II) sulphate ($NiSO_4$)	Sigma
*Not*I restriction enzyme	New England Biolabs
NuPAGE LDS Sample buffer	Invitrogen
NuPAGE® LDS Sample Buffer (4X)	Invitrogen
P20 detergent	GE Healthcare
PEG solutions	Fluka
Peptone from casein	Merck
Phosphate buffered saline (PBS)	Life technology
Phusion DNA polymerase, 2U/µl	Thermo Scientific
Phusion reaction buffer, 5x	Thermo Scientific
Potassium Ferricyanide ($K_3Fe(CN)_6$)	Sigma
Potassiumphosphate, mono (KH_4PO_4)	Fluka
Primers, PCR	Eurofins Genomics
Primers, Sequencing	Metabion
p-nitrophenylphosphate substrate (pNPP)	Sigma
2-Propanol	Sigma
Protein L-HRP	Genscript
RNase A	Sigma
SeeBlue® Plus2 standard, molecular weight marker	Invitrogen
Silver nitrate ($AgNO_3$)	Merck
Skimmed milk (dried)	AppliChem
Sodium carbonate (Na_2CO_3)	Sigma
Sodium chloride (NaCl)	Prolabo
Sodium citrate ($Na_3C_6H_5O_7$)	Sigma
Sodium hydroxide (NaOH)	Kebo Lab

Sodium thiosulfate ($Na_2S_2O_3$)	Sigma
D-(+)-Sucrose	Fluka
SYBR Safe DNA gel stain (100x)	Invitrogen
SYPRO Orange	Sigma
Synthesised 14F7 scFv genes	Life technology
T4 DNA ligase	New England Biolabs
T4 DNA ligase buffer (10x)	New England Biolabs
Tris Base	Chalbiocem
Tris-hydrogencloride (Tris-HCl)	Chalbiocem
Tween 20	Sigma
Urea	Merck
Yeast Extract, granulated	Merck

Table S2: Chromatographic Column Materials

Materials	**Manufacturer**
Chelating Sepharose	GE healthcare
Protein L capto resin	GE Healthcare,
Superdex 200 R10/300 Column	GE Healthcare

Table S3: Equipment

Equipment	Manufacturer
Amersham Hyperfilm ECL	GE healthcare
Amicon Ultra free tubes	Millipore
Amicon ultra filter, Ultracell-10 K, 2 ml	Millipore
Amicon ultra filter, Ultracell-10 K, 15 ml	Millipore
CM5 chip	GE Healthcare
Crystal Clear Sealing Tape	Hampton
Cuvettes	Sarstedt
Crystallisation Plate, 96 MRC	SWISSCI
Eppendorf tubes, 1.5 ml and 2 ml	Eppendorf
Extra thick paper blot paper	Bio-Rad Laboratories
Immobilon-P membrane	Millipore
LightCycle, 480 Sealing foil	Roche
LightCycle, 480 Plate 384	Roche
NuPAGE Bis-Tris 4-12% gel	Life technology
Nunc-Immuno 96 MicroWell MaxiSorp and PolySorb solid plates	Sigma
Optifit Refill Tips	Sartorius
PCR Cuvette	Eppendorf
Petman pipettes	Gilson
Rapid Flow, 20 µm, 150 ml, filter	Nalgene
Siliconized cover slides	Hampton Research
SuperSignal West Femto Chemiluminescent Substrate	Thermo Scientific
TPP® tissue culture plates, 24 well	Sigma

Table S4: Software

Software	Vendor
Ape	By Wayne Davis
Clustal2W	EMBL-EBI
GraphPad Prism v5	GraphPad software Ink
LightCycle software	Roche
Paint	Microsoft
ProtParam, ExPASy	Swiss Institute of Bioinformatics
PyMOL	Schrödinger
Scaffold 4	Proteomic Software Inc.
Unicorn 5.11	GE Healthcare
V-QUEST	IMGT
Wasp Run	Douglas Instruments

Table S5: Equipment

Machines	Vendor
ÄKTApurifier-900	GE Healthcare
Avanti Centrifuge J-26 XP	Beckman Coulter
Biofuge Fresco, Heraeus	Thermo Scientific
Capsulefuge, TOMY PMC-060	Tomyteck
Centrifuge 5810 R	Eppendorf
Electrophoresis power supply-EPS 601	GE healthcare
EmulsiFlex-C3	Avestin
Eppendorf Thermomixer comfort	Eppendorf
Gel logic 200 image system	Kodak
LightCycle 480	Roche
Mini Horizontal Submarine Unit	GE Healthcare
Miniorbital shaker	Stuart
Multitron II	Infors
MQ-H_2O, Direct Q	Millipore
Nanodrop 2000c	Thermo Scientific
NanoPhotometer	IMPLEN
Novex Mini-Cell	Invitrogen
Orbitrap-XL	Thermo Fisher
Oryx4 robot	Douglas Instruments
Peltier Thermal Cycler (PTC) -200	MJ Research
PowerPack HC power supply	Bio-Rad Laboratories
Trans-Blot SD Semi-Dry Transfer Cell	Bio-Rad Laboratories
Tuttnauer 3870-ML, autoclave	Tuttnauer
Ultraspec III KKB	Pharmacia

Table S6: Kits

Kit	Vendor
Nucleo Spin Plasmid kit	Macherey-Nagel GmbH & Co. KG
PCR Purification Kit (250)	QIAGEN
QIAprep Spin Miniprep Kit	QIAGEN
QIAquick gel extraction kit	QIAGEN

Table S7: Crystallisation Kits

Crystallisation kit	Vendor
JCSG	Molecular Dimensions
JCSG++	Molecular Dimensions
Morpheus	Molecular Dimensions
Morpheus (added Ca^{2+})	Molecular Dimensions
Morpheus MDI-46	Molecular Dimensions
PGA	Molecular Dimensions
PACT	Molecular Dimensions

Section B: Solutions, buffers and gels

All reagents are listed in Table S1: Reagents

Media

LB medium		**2x YT**	
10 g	Peptone	16 g	Peptone
5 g	Yeast extract	10 g	Yeast extract
10 g	HCl	5 g	NaCl
pH 7.1		pH 7.8	
MQ-H_2O to final volume of 1 L		MQ-H_2O to final volume of 1 L	
Autoclaved, and stored at 4 °C		*Autoclaved, and stored at 4 °C*	

Protein extraction

Periplasmic extraction solution		**Cell cracking buffer**	
20%	Sucrose	50 mM	Tris-HCl
50 mM	Tris-HCl	150 mM	NaCl
1 mM	EDTA	1 mM	EDTA
pH 7.8		2 µg/ml	DNase
		1 tab	Per 100 ml cOmplete inhibitor
		pH 7.8	

Cell cracking wash buffer		**Protein storage buffer**	
50 mM	Tris-HCl	50 mM	Tris-HCl
150 mM	NaCl		
pH 7.8			

Purification buffers

Pre-purification protein adjustment-IMAC	
80 ml	Soluble periplasm
4 ml	1 M Tric-HCl
1.8 ml	1 M Imidazole
24 ml	IMAC binding buffer
pH 7.8	

IMAC binding buffer		**IMAC elution buffer**	
20 mM	Imidazole	0.5 M	Imidazole
50 mM	Tris-HCl	50 mM	Tris-HCl
0.5 M	NaCl	0.5 M	NaCl
pH 7.8		pH 7.8	
The buffer is filtered and degassed before use		*The buffer is filtered and degassed before use*	

Pre-purification protein adjustment -Protein L, C1 and C3		**Protein L binding buffer**	
80 ml	Soluble periplasm	20 mM	Sodium phosphate
14 ml	1 M Tric-HCl	150 mM	NaCl for *C1* and C3**
12 ml	200 mM Sodium Phosphate	300 mM	NaCl *for C2* and C4**
4.5 ml	4 M NaCl for *C1* and C4**	pH 7.8	
9 ml	4 M NaCl for *C2* and C4**		
pH 7.8		*The buffer is filtered before use*	

Protein L elution buffer		**SEC running buffer**	
0.1 M	Sodium Citrate	50 mM	Tris-HCl
pH 3 *The buffer is filtered before use*		pH 7.8	*The buffer is filtered and degassed before use*

Running buffer for agarose gel

1x TAE-Running buffer (agarose gel)	
4.84 g	Tris base
1.14 g	Acetic acid
0.74 g	EDTA * $2H_2O$
MQ-H_2O to final volume of 1 L	

Staining and destaining solutions

Coomassie Blue staining solution	
0.05%	Coomassie Blue
10%	Acetic Acid
25%	2-Propanol

Fixation mix-Silver staining	
40%	Ethanol
10%	Acetic acid

Farmer's Reducer-Silver staining	
1%	30 mM $K_3Fe(CN)_6$
0.75%	30 mM $Na_2S_2O_3$x 5 H_2O

Silver nitrate solution-Silver staining	
0.1%	$AgNO_3$

Washing solution-Silver staining	
2.5%	Na_2CO_3

Developing solution-Silver staining	
2.5%	Na_2CO_3
0.1%	Formaldehyde (CH_2O)

Stop development solution-Silver staining	
10%	Acetic acid

Buffers used in ELISA and Western blotting

1x PBS	
0.8%	NaCl
0.02%	KCl
0.144%	Na_2HPO_4
0.024%	KH_4PO_4
pH 7.4	

1x PBS-T	
500 ml	1x PBS
0.5 ml	Tween-20

1x PSB-TM – (5% milk)	
50 ml	1x PBS-T
2.5 g	Dried skimmed milk

ECF buffer-Western	
15 ml	5 M NaCl
50 ml	1 M Tris pH 7.5
MQ-H_2O to a final volume of 0.5 L	

Transfer buffer-Western	
15 g	Glycine
25 ml	1 M Tris pH 8.0
200 ml	Methanol
1 ml	20% SDS
MQ-H_2O to a final volume of 1 L	

Section C: Restriction- and ligation reagents

All reagents are listed in Table S1: Reagents

Double-digest reactions: Cloning and subcloning the scFv constructs into pFKPEN

Table S8: Digestion setup for cloning the C1 construct

Double digest reaction of pFKPEN: mother clone (vector)		**Double digest reaction for V_H-L_R-$V_{L.A}$ (insert)**	
2.5 µl	800 ng/µl pFKPEN, mother clone	10 µl	200 ng/µl V_H-L_R-$V_{L.A}$
3 µl	Buffer #3	3 µl	Buffer #3
1 µl	*Nco*I	1 µl	*Nco*I
1 µl	*Not*I	1 µl	*Not*I
3 µl	100x BSA	3 µl	100x BSA
19.5 µl	MQ-H_2O	19.5 µl	MQ-H_2O
30 µl		30 µl	

Table S9: Digestion setup controls *Nco*I and *Not*I for cloning the C1 construct

Control digestion of pFKPEN, mother clone (vector)		**Control digestion of pFKPEN, mother clone (vector)**	
2.5 µl	800 ng/µl pFKPEN, mother clone	2.5 µl	800 ng/µl pFKPEN, mother clone
3 µl	Buffer #3	3 µl	Buffer #3
1 µl	*Nco*I	1 µl	*Not*I
3 µl	100x BSA	3 µl	100x BSA
20.5 µl	MQ-H_2O	20.5 µl	MQ-H_2O
30 µl µl		30 µl	

Table S10: Digestion setup for subcloning the C2 construct

Double digest reaction of pFKPEN: containing scFv C1 (vector)		**Double digest reaction for L_C-V_L (insert)**	
10 µl	290 ng/µl pFKPEN, C1	20 µl	100 ng/µl L_C-V_L
3 µl	Buffer #3	3 µl	Buffer #3
1 µl	*Mlu*I	1 µl	*Mlu*I
1 µl	*Not*I	1 µl	*Not*I
3 µl	100x BSA	3 µl	100x BSA
12 µl	MQ-H_2O	2 µl	MQ-H_2O
30 µl		30 µl	

Table S11: Digestion setup for subcloning the C4 construct

Double digest reaction of pFKPEN: containing scFv C1 (vector)		**Double digest reaction for L_C-V_L (insert)**	
10 µl	290 ng/µl pFKPEN, C1	20 µl	100 ng/µl L_C-V_L
3 µl	Buffer #3	3 µl	Buffer #3
1.5 µl	*Hind*III	1.5 µl	*Hind*III
1 µl	*Not*I	1 µl	*Not*I
3 v	100x BSA	3 µl	100x BSA
10.5 µl	MQ-H_2O	1.5 µl	MQ-H_2O
1 µl	CIP		
30 µl		30 µl	

Table S12: Digestion setup controls HindIII and NotI for subcloning the C2 and C4 construct s

Control digestion of pFKPEN, C1		**Control digestion of pFKPEN, C1**	
5 µl	290 ng/µl pFKPEN, C1	5 µl	290 ng/µl pFKPEN, C1
3 µl	Buffer #3	3 µl	Buffer #3
1.5 µl	*Hind*III	1 µl	*Not*I
3 µl	100x BSA	3 µl	100x BSA
17.5 µl	MQ-H_2O	18 µl	MQ-H_2O
30 µl		30 µl	

Table S13: Digestion setup control *Mlu*I for subcloning the C2 and C4 constructs

Control digestion of pFKPEN, C1	
5 µl	290 ng/µl pFKPEN, C1
3 µl	Buffer #3
1 µl	*Mlu*I
3 µl	100x BSA
18 µl	MQ-H_2O
30 µl	

Table S14: Digestion setup for subcloning the C3 construct

Double digest reaction of pFKPEN: containing scFv C4 (vector)		**Double digest reaction of pFKPEN: containing scFv C1 (insert)**	
10 µl	230 ng/µl pFKPEN, C4	7 µl	370 ng/µl pFKPEN, C1
3 µl	Buffer #3	3 µl	Buffer #3
1 µl	*Mlu*I	1 µl	*Mlu*I
1 µl	*Not*I	1 µl	*Not*I
3 µl	100x BSA	3 µl	100x BSA
12 µl	MQ-H_2O	15 µl	MQ-H_2O
30		30 µl	

Table S15: Digestion setup controls *Mlu*I and *Not*I for subcloning the C3 construct

Control digestion of pFKPEN, C4		**Control digestion of pFKPEN, C4**	
10 µl	230 ng/µl pFKPEN, C4	10 µl	230 ng/µl pFKPEN, C4
3 µl	Buffer #3	3 µl	Buffer #3
1 µl	*Mlu*I	1 µl	*Not*I
3 µl	100x BSA	3 µl	100x BSA
13 µl	MQ-H_2O	13 µl	MQ-H_2O
30 µl		30 µl	

Double-digest reactions: Removing the His-tag

Table S16: Digestion setup of pFKPEN for subcloning of C1* C2*, C3* and C4*

Double digest reaction of pFKPEN C3 (vector) for subcloning of C1*, C2*, C3*and C4*	
3.6 µl	560 ng/µl pFKPEN, C3
3 µl	Buffer #3
1.5 µl	*Nco*I
1.5 µl	*Not*I
3 µl	100x BSA
16.4 µl	MQ-H_2O
1 µl	CIP
30 µl	

Table S17: Digestion setup for subcloning of C1 and C2* into the pFKPEN vector

Double digest reaction of PCR product C1* (insert)		**Double digest reaction of PCR product C2* (insert)**	
13 µl	152 ng/µl PCR DNA: C1*	18 µl	111 ng/µl PCR DNA: C1*
3 µl	Buffer #3	3 µl	Buffer #3
1.5 µl	*Nco*I	1.5 µl	*Nco*I
1.5 µl	*Not*I	1.5 µl	*Not*I
3 µl	100x BSA	3 µl	100x BSA
8 µl	MQ-H_2O	3 µl	MQ-H_2O
30 µl		30 µl	

Table S18: Digestion setup for subcloning of C3* and C4* into the pFKPEN vector

Double digest reaction of PCR product C3* (insert)		**Double digest reaction of PCR product C4* (insert)**	
12 µl	166 ng/µl PCR DNA: C1*	11 µl	180 ng/µl PCR DNA: C1*
3 µl	Buffer #3	3 µl	Buffer #3
1.5 µl	*Nco*I	1.5 µl	*Nco*I
1.5 µl	*Not*I	1.5 µl	*Not*I
3 µl	100x BSA	3 µl	100x BSA
9 µl	MQ-H_2O	10 µl	MQ-H_2O
30		30 µl	

Table S19: Digestion setup controls *Nco*I and *Not*I with pFKPEN for subcloning the C1*, C2 *, C3*, C4* constructs

Control digestion of pFKPEN, C3		**Control digestion of pFKPEN, C3**	
3.6 µl	560 ng/µl pFKPEN, C3	3.6 µl	560 ng/µl pFKPEN, C3
3 µl	Buffer #3	3 µl	Buffer #3
1.5 µl	*Nco*I	1.5 µl	*Not*I
3 µl	100x BSA	3 µl	100x BSA
17.9 µl	MQ-H_2O	17.9 µl	MQ-H_2O
1 µl	CIP	1 µl	CIP
30		30 µl	

All restriction reactions were incubated at 37 ℃ for 1.5 hours.

Ligation reactions: Cloning and subcloning the scFv constructs into pFKPEN

Table S20: Ligation setup for cloning the scFv C1 construct into pFKPEN

4:1 ligation reaction of pFKPEN, mother clone (vector) and V_H-L_R-$V_{L.A}$ (insert)		**Negative control: Ligation reaction for pFKPEN, mother clone (vector)**	
13 µl	3.7 ng/µl V_H-L_R-$V_{L.A}$, 815 bp	0 µl	Insert
3 µl	21 ng/µl pFKPEN, 39995 bp	3 µl	21 ng/µl pFKPEN, 3995 bp
2 µl	10x ligation buffer	2 µl	10x ligation buffer
1 µl	T4 DNA ligase	1 µl	T4 DNA ligase
1 µl	MQ-H_2O	14µl	MQ-H_2O
20 µl		20 µl	

Table S21: Ligation setup for subcloning of C2 into the pFKPEN vector

4:1 ligation reaction of pFKPEN, C1 (vector) and V_L (insert)		**Negative control: Ligation reaction for pFKPEN, C1 (vector)**	
5 µl	3.9 ng/µl $V_{L.}$, 340 bp	0 µl	Insert
5 µl	13 ng/µl pFKPEN, C1, 4474 bp	5 µl	13 ng/µl pFKPEN, C1, 4474 bp
2 µl	10x ligation buffer	2 µl	10x ligation buffer
1 µl	T4 DNA ligase	1 µl	T4 DNA ligase
7 µl	MQ-H_2O	7 µl	MQ-H_2O
20 µl		20 µl	

Table S22: Ligation setup for subcloning of C4 into the pFKPEN vector

4:1 ligation reaction of pFKPEN, C1 (vector) and L_C-V_L (insert)		**Negative control: Ligation reaction for pFKPEN, C1 (vector)**	
5.7 µl	4.2 ng/µl $V_{L.}$, 396bp	0 µl	Insert
3 µl	22.3 ng/µl pFKPEN, C1, 4418bp	3 µl	22.3 ng/µl pFKPEN, C1, 4418bp
2 µl	10x ligation buffer	2 µl	10x ligation buffer
1 µl	T4 DNA ligase	1 µl	T4 DNA ligase
8.3 µl	MQ-H_2O	14 µl	MQ-H_2O
20 µl		20 µl	

Table S23: Ligation setup for subcloning of C3 into the pFKPEN vector

4:1 ligation reaction of pFKPEN, C4 (vector) and V_L (insert)		**Negative control: Ligation reaction for pFKPEN, C4 (vector)**	
3.4 µl	3.5 ng/µl $V_{L.}$, 340 bp	0 µl	Insert
13 µl	3 ng/µl pFKPEN, C1, 4474 bp	13 µl	3 ng/µl pFKPEN, C1, 4474 bp
2 µl	10x ligation buffer	2 µl	10x ligation buffer
1 µl	T4 DNA ligase	1 µl	T4 DNA ligase
0.6 µl	MQ-H_2O	4 µl	MQ-H_2O
20 µl		20 µl	

Ligation reactions: Subcloning the scFv PCR fragments into pFKPEN

Table S24: Ligation setup for subcloning of PCR product C1* and C2* into the pFKPEN vector

4:1 ligation reaction of pFKPEN, C3 (vector) and C1* (insert)		**4:1 ligation reaction of pFKPEN, C3 (vector) and C2* (insert)**	
1 µl	38 ng/µl, C1*, 780 bp	1 µl	40 ng/µl, C2*, 780 bp
4 µl	12.5 ng/µl pFKPEN, C1, 3995 bp	4 µl	12.5 ng/µl pFKPEN, C1, 3995 bp
2 µl	10x ligation buffer	2 µl	10x ligation buffer
1 µl	T4 DNA ligase	1 µl	T4 DNA ligase
12 µl	MQ-H_2O	12 µl	MQ-H_2O
20 µl		20 µl	

Table S25: Ligation setup for subcloning of PCR product C3* and C4* into the pFKPEN vector

4:1 ligation reaction of pFKPEN, C3 (vector) and C3* (insert)		**4:1 ligation reaction of pFKPEN, C3 (vector) and C4* (insert)**	
1 µl	40 ng/µl, C2*, 780 bp	2 µl	20 ng/µl, C2*, 780 bp
4 µl	12.5 ng/µl pFKPEN, C1, 3995 bp	4 µl	12.5 ng/µl pFKPEN, C1, 3995 bp
2 µl	10x ligation buffer	2 µl	10x ligation buffer
1 µl	T4 DNA ligase	1 µl	T4 DNA ligase
12 µl	MQ-H_2O	11 µl	MQ-H_2O
20 µl		20 µl	

Table S26: Ligation setup control for the subcloning of the PCR DNA fragments into vector pFKPEN

Negative control: ligation reaction for pFKPEN, C3 (vector)	
0 µl	Insert
4 µl	12.5 ng/µl pFKPEN, C1, 3995 bp
2 µl	10x ligation buffer
1 µl	T4 DNA ligase
12 µl	MQ-H_2O
20 µl	

All ligation samples are incubated overnight at room temperature

Section D: PCR mixtures and PCR programs

All reagents are listed in Table S1: Reagents

Table S27: PCR mixture for amplification of circular pFKPEN containing the scFv constructs

36.1 µl	MQ-H_2O
10 µl	5x Phusion HF buffer
2 µl	5 mM dNTP
0.2 µl	100 µM forward primer
0.2 µl	100 µM reverse primer
1 µl	40 ng/ µl pFKPEN containing the scFv construct
0.5 µl	2U/ µl Phusion DNA polymerase
50 µl	

Table S28: PCR program for circular pFKPEN containing the scFv constructs

1x	Initial melting	98 °C	1 minute
28x	Denaturation	98 °C	10 seconds
	Annealing	60 °C	15 seconds
	Elongation	72 °C	20 seconds
1x	Final elongation	72 °C	5 minutes
	Maintain	4 °C	

Section E: Nucleic acid sequences

The restriction sites are underlined.

DNA sequence of C1 (819 bp)

1 GCCATGGCCCACCACCACCACCACCACGAAAACCTGTACTTCCAGGGTCA
51 GGTGCAGCTGCAGCAGAGCGGCGCGGAACTGGCGAAACCGGGCGCGAGCA
101 TGAAAATGAGCTGCCGCGCGAGCGGCTATAGCTTTACCAGCTATTGGATT
151 CATTGGCTGAAACAGCGCCCGGATCAGGGCCTGGAATGGATTGGCTATAT
201 TGATCCGGCGACCGCGTATACCGAAAGCAACCAGAAATTTAAAGATAAAG
251 CGATTCTGACCGCGGATCGCAGCAGCAACACCGCGTTTATGTATCTGAAC
301 AGCCTGACCAGCGAAGATAGCGCGGTGTATTATTGCGCGCGCGAAAGCCC
351 GCGCCTGCGCCGCGGCATTTATTATTATGCGATGGATTATTGGGGCCAGG
401 GCACCACCGTGACCGTGAGCAGCAAGCTTTCAGGGAGTGCATCCGCCCCA
451 AAACTTGAAGAAGGTGAATTTTCAGAAGCACGCGTAGACATCCAGATGAC
501 CCAGACCCCGTCTTCTCTGTCTGCTTCTCTGGGTGACCGTGTTACCATCT
551 CTTGCCGTGCTTCTCAGGACATCTCTAACTACCTGAACTGGTACCAGCAG
601 AAACCGGACGGTACCGTTAAACTGCTGATCTACTACACCTCTCGTCTGCA
651 CTCTGGTGTTCCGTCTCGTTTCTCTGGTTCTGGTTCTGGTACCGACTACT
701 CTCTGACCATCTCTAACCTGGAACAGGAAGACATCGCTACCTACTTCTGC
751 CAGCAGGGTAACACCCTGCCGCCGACCTTCGGTGCTGGTACCAAACTGGA
801 ACTGAAATAAGCGGCCGCT

DNA sequence of C1* (780bp)

1 GCCATGGCCCAGGTGCAGCTGCAGCAGAGCGGCGCGGAACTGGCGAAACC
51 GGGCGCGAGCATGAAAATGAGCTGCCGCGCGAGCGGCTATAGCTTTACCA
101 GCTATTGGATTCATTGGCTGAAACAGCGCCCGGATCAGGGCCTGGAATGG
151 ATTGGCTATATTGATCCGGCGACCGCGTATACCGAAAGCAACCAGAAATT
201 TAAAGATAAAGCGATTCTGACCGCGGATCGCAGCAGCAACACCGCGTTTA
251 TGTATCTGAACAGCCTGACCAGCGAAGATAGCGCGGTGTATTATTGCGCG
301 CGCGAAAGCCCGCGCCTGCGCCGCGGCATTTATTATTATGCGATGGATTA
351 TTGGGGCCAGGGCACCACCGTGACCGTGAGCAGCAAGCTTTCAGGGAGTG
401 CATCCGCCCCAAAACTTGAAGAAGGTGAATTTTCAGAAGCACGCGTAGAC
451 ATCCAGATGACCCAGACCCCGTCTTCTCTGTCTGCTTCTCTGGGTGACCG
501 TGTTACCATCTCTTGCCGTGCTTCTCAGGACATCTCTAACTACCTGAACT
551 GGTACCAGCAGAAACCGGACGGTACCGTTAAACTGCTGATCTACTACACC
601 TCTCGTCTGCACTCTGGTGTTCCGTCTCGTTTCTCTGGTTCTGGTTCTGG
651 TACCGACTACTCTCTGACCATCTCTAACCTGGAACAGGAAGACATCGCTA
701 CCTACTTCTGCCAGCAGGGTAACACCCTGCCGCCGACCTTCGGTGCTGGT
751 ACCAAACTGGAACTGAAATAAGCGGCCGCT

DNA sequence of C2 (819 bp)

1 GCCATGGCCCACCACCACCACCACCACGAAAACCTGTACTTCCAGGGTCA
51 GGTGCAGCTGCAGCAGAGCGGCGCGGAACTGGCGAAACCGGGCGCGAGCA
101 TGAAAATGAGCTGCCGCGCGAGCGGCTATAGCTTTACCAGCTATTGGATT
151 CATTGGCTGAAACAGCGCCCGGATCAGGGCCTGGAATGGATTGGCTATAT
201 TGATCCGGCGACCGCGTATACCGAAAGCAACCAGAAATTTAAAGATAAAG
251 CGATTCTGACCGCGGATCGCAGCAGCAACACCGCGTTTATGTATCTGAAC
301 AGCCTGACCAGCGAAGATAGCGCGGTGTATTATTGCGCGCGCGAAAGCCC
351 GCGCCTGCGCCGCGGCATTTATTATTATGCGATGGATTATTGGGGCCAGG
401 GCACCACCGTGACCGTGAGCAGCAAGCTTTCAGGGAGTGCATCCGCCCCA
451 AAACTTGAAGAAGGTGAATTTTCAGAAGCACGCGTAGACCTGGTTCTGAC
501 CCAGTCTCCGGCTACCCTGTCTGTTACCCCAGGTGACTCTGTTTCTTTCT
551 CTTGCCGTGCTTCTCAGTCTATCTCTAACAACCTGCACTGGTACCAGCAG
601 CGTACCCACGAATCTCCGCGTCTGCTGATCAAATACGCTTCTCAGTCTAT
651 CTCTGGTATCCCGTCTCGTTTCTCTGGTTCTGGTTCTGGTACCGACTTCA
701 CCCTGTCTATCTCTTCTGTTGAAACCGAAGACTTCGGTATGTACTTCTGC
751 CAGCAGTCTAACCGTTGGCCGCTGACCTTCGGTGCTGGTACCAAACTGGA
851 ACTGAAATAAGCGGCCGCT

DNA sequence of C2* (780bp)

1 GCCATGGCCCAGGTGCAGCTGCAGCAGAGCGGCGCGGAACTGGCGAAACC
51 GGGCGCGAGCATGAAAATGAGCTGCCGCGCGAGCGGCTATAGCTTTACCA
101 GCTATTGGATTCATTGGCTGAAACAGCGCCCGGATCAGGGCCTGGAATGG
151 ATTGGCTATATTGATCCGGCGACCGCGTATACCGAAAGCAACCAGAAATT
201 TAAAGATAAAGCGATTCTGACCGCGGATCGCAGCAGCAACACCGCGTTTA
251 TGTATCTGAACAGCCTGACCAGCGAAGATAGCGCGGTGTATTATTGCGCG
301 CGCGAAAGCCCGCGCCTGCGCCGCGGCATTTATTATTATGCGATGGATTA
351 TTGGGGCCAGGGCACCACCGTGACCGTGAGCAGCAAGCTTTCAGGGAGTG
401 CATCCGCCCCAAAACTTGAAGAAGGTGAATTTTCAGAAGCACGCGTAGAC
451 CTGGTTCTGACCCAGTCTCCGGCTACCCTGTCTGTTACCCCAGGTGACTC
501 TGTTTCTTTCTCTTGCCGTGCTTCTCAGTCTATCTCTAACAACCTGCACT
551 GGTACCAGCAGCGTACCCACGAATCTCCGCGTCTGCTGATCAAATACGCT
601 TCTCAGTCTATCTCTGGTATCCCGTCTCGTTTCTCTGGTTCTGGTTCTGG
651 TACCGACTTCACCCTGTCTATCTCTTCTGTTGAAACCGAAGACTTCGGTA
701 TGTACTTCTGCCAGCAGTCTAACCGTTGGCCGCTGACCTTCGGTGCTGGT
751 ACCAAACTGGAACTGAAATAAGCGGCCGCT

DNA sequence of C3 (819 bp)

1 GCCATGGCCCACCACCACCACCACCACGAAAACCTGTACTTCCAGGGTCA
51 GGTGCAGCTGCAGCAGAGCGGCGCGGAACTGGCGAAACCGGGCGCGAGCA
101 TGAAAATGAGCTGCCGCGCGAGCGGCTATAGCTTTACCAGCTATTGGATT
151 CATTGGCTGAAACAGCGCCCGGATCAGGGCCTGGAATGGATTGGCTATAT
201 TGATCCGGCGACCGCGTATACCGAAAGCAACCAGAAATTTAAAGATAAAG
251 CGATTCTGACCGCGGATCGCAGCAGCAACACCGCGTTTATGTATCTGAAC
301 AGCCTGACCAGCGAAGATAGCGCGGTGTATTATTGCGCGCGCGAAAGCCC
351 GCGCCTGCGCCGCGGCATTTATTATTATGCGATGGATTATTGGGGCCAGG
401 GCACCACCGTGACCGTGAGCAGCAAGCTTGCGCCGCAGGCGAAAAGCAGC
451 GGCAGCGGCAGCGAAAGCAAAGTGGATGCACGCGTAGACATCCAGATGAC
501 CCAGACCCCGTCTTCTCTGTCTGCTTCTCTGGGTGACCGTGTTACCATCT
551 CTTGCCGTGCTTCTCAGGACATCTCTAACTACCTGAACTGGTACCAGCAG
601 AAACCGGACGGTACCGTTAAACTGCTGATCTACTACACCTCTCGTCTGCA
651 CTCTGGTGTTCCGTCTCGTTTCTCTGGTTCTGGTTCTGGTACCGACTACT
701 CTCTGACCATCTCTAACCTGGAACAGGAAGACATCGCTACCTACTTCTGC
751 CAGCAGGGTAACACCCTGCCGCCGACCTTCGGTGCTGGTACCAAACTGGA
801 ACTGAAATAAGCGGCCGCT

DNA sequence of C3* (780bp)

1 GCCATGGCCCAGGTGCAGCTGCAGCAGAGCGGCGCGGAACTGGCGAAACC
51 GGGCGCGAGCATGAAAATGAGCTGCCGCGCGAGCGGCTATAGCTTTACCA
101 GCTATTGGATTCATTGGCTGAAACAGCGCCCGGATCAGGGCCTGGAATGG
151 ATTGGCTATATTGATCCGGCGACCGCGTATACCGAAAGCAACCAGAAATT
201 TAAAGATAAAGCGATTCTGACCGCGGATCGCAGCAGCAACACCGCGTTTA
251 TGTATCTGAACAGCCTGACCAGCGAAGATAGCGCGGTGTATTATTGCGCG
301 CGCGAAAGCCCGCGCCTGCGCCGCGGCATTTATTATTATGCGATGGATTA
351 TTGGGGCCAGGGCACCACCGTGACCGTGAGCAGCAAGCTTGCGCCGCAGG
401 CGAAAAGCAGCGGCAGCGGCAGCGAAAGCAAAGTGGATGCACGCGTAGAC
451 ATCCAGATGACCCAGACCCCGTCTTCTCTGTCTGCTTCTCTGGGTGACCG
501 TGTTACCATCTCTTGCCGTGCTTCTCAGGACATCTCTAACTACCTGAACT
551 GGTACCAGCAGAAACCGGACGGTACCGTTAAACTGCTGATCTACTACACC
601 TCTCGTCTGCACTCTGGTGTTCCGTCTCGTTTCTCTGGTTCTGGTTCTGG
651 TACCGACTACTCTCTGACCATCTCTAACCTGGAACAGGAAGACATCGCTA
701 CCTACTTCTGCCAGCAGGGTAACACCCTGCCGCCGACCTTCGGTGCTGGT
751 ACCAAACTGGAACTGAAATAAGCGGCCGCT

DNA sequence of C4 (819 bp)

1 GCCATGGCCCACCACCACCACCACCACGAAAACCTGTACTTCCAGGGTCA
51 GGTGCAGCTGCAGCAGAGCGGCGCGGAACTGGCGAAACCGGGCGCGAGCA
101 TGAAAATGAGCTGCCGCGCGAGCGGCTATAGCTTTACCAGCTATTGGATT
151 CATTGGCTGAAACAGCGCCCGGATCAGGGCCTGGAATGGATTGGCTATAT
201 TGATCCGGCGACCGCGTATACCGAAAGCAACCAGAAATTTAAAGATAAAG
251 CGATTCTGACCGCGGATCGCAGCAGCAACACCGCGTTTATGTATCTGAAC
301 AGCCTGACCAGCGAAGATAGCGCGGTGTATTATTGCGCGCGCGAAAGCCC
351 GCGCCTGCGCCGCGGCATTTATTATTATGCGATGGATTATTGGGGCCAGG
401 GCACCACCGTGACCGTGAGCAGCAAGCTTGCGCCGCAGGCGAAAAGCAGC
451 GGCAGCGGCAGCGAAAGCAAAGTGGATGCACGCGTAGACCTGGTTCTGAC
501 CCAGTCTCCGGCTACCCTGTCTGTTACCCCAGGTGACTCTGTTTCTTTCT
551 CTTGCCGTGCTTCTCAGTCTATCTCTAACAACCTGCACTGGTACCAGCAG
601 CGTACCCACGAATCTCCGCGTCTGCTGATCAAATACGCTTCTCAGTCTAT
651 CTCTGGTATCCCGTCTCGTTTCTCTGGTTCTGGTTCTGGTACCGACTTCA
701 CCCTGTCTATCTCTTCTGTTGAAACCGAAGACTTCGGTATGTACTTCTGC
751 CAGCAGTCTAACCGTTGGCCGCTGACCTTCGGTGCTGGTACCAAACTGGA
851 ACTGAAATAAGCGGCCGCT

DNA sequence of C4* (780bp)

1 GCCATGGCCCAGGTGCAGCTGCAGCAGAGCGGCGCGGAACTGGCGAAACC
51 GGGCGCGAGCATGAAAATGAGCTGCCGCGCGAGCGGCTATAGCTTTACCA
101 GCTATTGGATTCATTGGCTGAAACAGCGCCCGGATCAGGGCCTGGAATGG
151 ATTGGCTATATTGATCCGGCGACCGCGTATACCGAAAGCAACCAGAAATT
201 TAAAGATAAAGCGATTCTGACCGCGGATCGCAGCAGCAACACCGCGTTTA
251 TGTATCTGAACAGCCTGACCAGCGAAGATAGCGCGGTGTATTATTGCGCG
301 CGCGAAAGCCCGCGCCTGCGCCGCGGCATTTATTATTATGCGATGGATTA
351 TTGGGGCCAGGGCACCACCGTGACCGTGAGCAGCAAGCTTGCGCCGCAGG
401 CGAAAAGCAGCGGCAGCGGCAGCGAAAGCAAAGTGGATGCACGCGTAGAC
451 CTGGTTCTGACCCAGTCTCCGGCTACCCTGTCTGTTACCCCAGGTGACTC
501 TGTTTCTTTCTCTTGCCGTGCTTCTCAGTCTATCTCTAACAACCTGCACT
551 GGTACCAGCAGCGTACCCACGAATCTCCGCGTCTGCTGATCAAATACGCT
601 TCTCAGTCTATCTCTGGTATCCCGTCTCGTTTCTCTGGTTCTGGTTCTGG
651 TACCGACTTCACCCTGTCTATCTCTTCTGTTGAAACCGAAGACTTCGGTA
701 TGTACTTCTGCCAGCAGTCTAACCGTTGGCCGCTGACCTTCGGTGCTGGT
751 ACCAAACTGGAACTGAAATAAGCGGCCGCT

PCR-primers for Oligonucleotide-directed mutagenesis

Forward (*Nco*I restriction site underlined):

5’-ATATCCATGGCCCAGGTGCAGCTGCAGCAG-3’

Reverse (*Not*I restriction site underlined):

(5’-TATAGCGGCCGCTTATTTCAGTTCCAGTTTGG-3’)

Section F: Amino acid sequences

ClustalW2, EMBL-EBI alignment of the four scFv nucleotide sequence (Larkin et al., 2007).

The V_H is not marked and starts form nucleotide 7 to 423.
The V_L is not marked and starts from nucleotide 486 to 810
The $V_{L.A}$ is marked in light grey: ATCG
The L_C linker is marked in medium grey: ATCG
The L_R linker is marked in dark grey: ATCG
The restriction sites are marked in black: ATCG

```
C1          GCCATGGCCCACCACCACCACCACCACGAAAACCTGTACTTCCAGGGTCAGGTGCAGCTG 60
C2          GCCATGGCCCACCACCACCACCACCACGAAAACCTGTACTTCCAGGGTCAGGTGCAGCTG 60
C3          GCCATGGCCCACCACCACCACCACCACGAAAACCTGTACTTCCAGGGTCAGGTGCAGCTG 60
C4          GCCATGGCCCACCACCACCACCACCACGAAAACCTGTACTTCCAGGGTCAGGTGCAGCTG 60
            ************************************************************

C1          CAGCAGAGCGGCGCGGAACTGGCGAAACCGGGCGCGAGCATGAAAATGAGCTGCCGCGCG 120
C2          CAGCAGAGCGGCGCGGAACTGGCGAAACCGGGCGCGAGCATGAAAATGAGCTGCCGCGCG 120
C3          CAGCAGAGCGGCGCGGAACTGGCGAAACCGGGCGCGAGCATGAAAATGAGCTGCCGCGCG 120
C4          CAGCAGAGCGGCGCGGAACTGGCGAAACCGGGCGCGAGCATGAAAATGAGCTGCCGCGCG 120
            ************************************************************

C1          AGCGGCTATAGCTTTACCAGCTATTGGATTCATTGGCTGAAACAGCGCCCGGATCAGGGC 180
C2          AGCGGCTATAGCTTTACCAGCTATTGGATTCATTGGCTGAAACAGCGCCCGGATCAGGGC 180
C3          AGCGGCTATAGCTTTACCAGCTATTGGATTCATTGGCTGAAACAGCGCCCGGATCAGGGC 180
C4          AGCGGCTATAGCTTTACCAGCTATTGGATTCATTGGCTGAAACAGCGCCCGGATCAGGGC 180
            ************************************************************

C1          CTGGAATGGATTGGCTATATTGATCCGGCGACCGCGTATACCGAAAGCAACCAGAAATTT 240
C2          CTGGAATGGATTGGCTATATTGATCCGGCGACCGCGTATACCGAAAGCAACCAGAAATTT 240
C3          CTGGAATGGATTGGCTATATTGATCCGGCGACCGCGTATACCGAAAGCAACCAGAAATTT 240
C4          CTGGAATGGATTGGCTATATTGATCCGGCGACCGCGTATACCGAAAGCAACCAGAAATTT 240
            ************************************************************

C1          AAAGATAAAGCGATTCTGACCGCGGATCGCAGCAGCAACACCGCGTTTATGTATCTGAAC 300
C2          AAAGATAAAGCGATTCTGACCGCGGATCGCAGCAGCAACACCGCGTTTATGTATCTGAAC 300
C3          AAAGATAAAGCGATTCTGACCGCGGATCGCAGCAGCAACACCGCGTTTATGTATCTGAAC 300
C4          AAAGATAAAGCGATTCTGACCGCGGATCGCAGCAGCAACACCGCGTTTATGTATCTGAAC 300
            ************************************************************

C1          AGCCTGACCAGCGAAGATAGCGCGGTGTATTATTGCGCGCGCGAAAGCCCGCGCCTGCGC 360
C2          AGCCTGACCAGCGAAGATAGCGCGGTGTATTATTGCGCGCGCGAAAGCCCGCGCCTGCGC 360
C3          AGCCTGACCAGCGAAGATAGCGCGGTGTATTATTGCGCGCGCGAAAGCCCGCGCCTGCGC 360
C4          AGCCTGACCAGCGAAGATAGCGCGGTGTATTATTGCGCGCGCGAAAGCCCGCGCCTGCGC 360
            ************************************************************

C1          CGCGGCATTTATTATTATGCGATGGATTATTGGGGCCAGGGCACCACCGTGACCGTGAGC 420
C2          CGCGGCATTTATTATTATGCGATGGATTATTGGGGCCAGGGCACCACCGTGACCGTGAGC 420
C3          CGCGGCATTTATTATTATGCGATGGATTATTGGGGCCAGGGCACCACCGTGACCGTGAGC 420
C4          CGCGGCATTTATTATTATGCGATGGATTATTGGGGCCAGGGCACCACCGTGACCGTGAGC 420
            ************************************************************

C1          AGCAAGCTTTCAGGGAGTGC-ATCCGCCCCAAAACTTGAAGAAGGTGAATTTTCAGAAGC 479
C2          AGCAAGCTTTCAGGGAGTGC-ATCCGCCCCAAAACTTGAAGAAGGTGAATTTTCAGAAGC 479
C3          AGCAAGCTTGCGCCGCAGGCGAAAAGCAGCGGCAGCGGCAGCGAAAGCAAAGTG-GATGC 479
C4          AGCAAGCTTGCGCCGCAGGCGAAAAGCAGCGGCAGCGGCAGCGAAAGCAAAGTG-GATGC 479
```

```
         ********* *.  *.. ** *:..**. *...*   *.**.....:*.*:: *  **:**

C1         ACGCGTAGACATCCAGATGACCCAGACCCCGTCTTCTCTGTCTGCTTCTCTGGGTGACCG 539
C2         ACGCGTAGACCTGGTTCTGACCCAGTCTCCGGCTACCCTGTCTGTTACCCCAGGTGACTC 539
C3         ACGCGTAGACATCCAGATGACCCAGACCCCGTCTTCTCTGTCTGCTTCTCTGGGTGACCG 539
C4         ACGCGTAGACCTGGTTCTGACCCAGTCTCCGGCTACCCTGTCTGTTACCCCAGGTGACTC 539
         **********.*  : .*********:* *** **:* ******* *:* * .******

C1         TGTTACCATCTCTTGCCGTGCTTCTCAGGACATCTCTAACTACCTGAACTGGTACCAGCA 599
C2         TGTTTCTTTCTCTTGCCGTGCTTCTCAGTCTATCTCTAACAACCTGCACTGGTACCAGCA 599
C3         TGTTACCATCTCTTGCCGTGCTTCTCAGGACATCTCTAACTACCTGAACTGGTACCAGCA 599
C4         TGTTTCTTTCTCTTGCCGTGCTTCTCAGTCTATCTCTAACAACCTGCACTGGTACCAGCA 599
         ****:* :********************* . *********:*****.**************

C1         GAAACCGGACGGTACCGTTAAACTGCTGATCTACTACACCTCTCG-TCTGCACTCTGGTG 658
C2         GCGTACCCACGAATCTCCGCGTCTGCTGATCAAATACGCTTCTCAGTCT-ATCTCTGGTA 658
C3         GAAACCGGACGGTACCGTTAAACTGCTGATCTACTACACCTCTCG-TCTGCACTCTGGTG 658
C4         GCGTACCCACGAATCTCCGCGTCTGCTGATCAAATACGCTTCTCAGTCT-ATCTCTGGTA 658
         *.:.*  ***.::*    ..:**********:*.*.***.* ****. *** .:********.

C1         TTCCGTCTCGTTTCTCTGGTTCTGGTTCTGGTACCGACTACTCTCTGACCATCTCTAACC 718
C2         TCCCGTCTCGTTTCTCTGGTTCTGGTTCTGGTACCGACTTCACCCTGTCTATCTCTTCTG 718
C3         TTCCGTCTCGTTTCTCTGGTTCTGGTTCTGGTACCGACTACTCTCTGACCATCTCTAACC 718
C4         TCCCGTCTCGTTTCTCTGGTTCTGGTTCTGGTACCGACTTCACCCTGTCTATCTCTTCTG 718
         * ***************************************.*:* ***:* ******:.

C1         TGGAACAGGAAGACATCGCTACCTACTTCTGCCAGCAGGGTAACACCCTGCCGCCGACCT 778
C2         TTGAAACCGAAGACTTCGGTATGTACTTCTGCCAGCAGTCTAACCGTTGGCCGCTGACCT 778
C3         TGGAACAGGAAGACATCGCTACCTACTTCTGCCAGCAGGGTAACACCCTGCCGCCGACCT 778
C4         TTGAAACCGAAGACTTCGGTATGTACTTCTGCCAGCAGTCTAACCGTTGGCCGCTGACCT 778
         * ***.. ******:*** **  *************** ****.   ***** *****

C1         TCGGTGCTGGTACCAAACTGGAACTGAAATAAGCGGCCGCT 819
C2         TCGGTGCTGGTACCAAACTGGAACTGAAATAAGCGGCCGCT 819
C3         TCGGTGCTGGTACCAAACTGGAACTGAAATAAGCGGCCGCT 819
C4         TCGGTGCTGGTACCAAACTGGAACTGAAATAAGCGGCCGCT 819
         *****************************************
```

Section G: Crystallisation screens

Original conditions for crystallisation hits

Morpheus screen condition C8 0.5 M Buffer, 12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% w/v PEG MPD, 0.03 M of each NPS, pH 7.8

C1* (1.6 mg/ml) with ~10x Tricer stored at 4 °C

Morpheus screen condition G8: 0.5 M Buffer, 12.5% w/v PEG 1000, 12.5% w/v PEG 3350, 12.5% w/v PEG MPD, 0.02 M of each carboxyl acid, pH 7.8 C1* (1.6 mg/ml) with ~10x Tricer stored at 4 °C

Table S29: Optimised Morpheus screen C8. The parameter that is varied from the original condition is marked in bold. C1 contains the original conditions

	1	**2**	**3**
A	0.5 M Buffer, **pH 6.5** 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD 0.03 M of each NPS:	0.5 M Buffer, pH 7.5 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD -	0.5 M Buffer, pH 7.5 **38.75% PEG 1000 (50%)** - - -0.03 M of each NPS:
B	0.5 M Buffer, **pH 7.0** 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD 0.03 M of each NPS:	0.5 M Buffer, pH 7.5 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD **1 M Sodium Nitrate**	0.1 ml Buffer, pH 7.5 - **30% PEG 3350** - 0.03 M of each NPS:
C	0.5 M Buffer, pH 7.5 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD 0.03 M of each NPS:	0.5 M Buffer, pH 7.5 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD **1 M Ammonium sulphate**	0.1 ml Buffer, pH 7.5 - - **35% PEG MME** 0.03 M of each NPS:
D	0.5 M Buffer, **pH 8.0** 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD 0.03 M of each NPS:	0.5 M Buffer, pH 7.5 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD **0.5 M Sodium phosphate**	0.1 ml Buffer, pH 7.5 - - **30% ml MPD** 0.03 M of each NPS:

Table S30 Optimised Morpheus screen G8. The parameter that is varied from the original condition is marked in bold. C1 contains the original conditions

	4	**5**	**6**
A	0.5 M Buffer, **pH 6.5** 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD 0.02 M of each carboxy l acid	0.5 M Buffer, pH 7.5 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD - -	0.1 ml Buffer, pH 7.5 **38.75% PEG 1000** - - 0.02 M of each carboxyl acid:
B	0.5 M Buffer, **pH 7.0** 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD 0.02 M of each carboxy l acid	0.5 M Buffer, pH 7.5 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD **1 M ammonium acetate**	0.1 ml Buffer, pH 7.5 - **30% PEG 3350** - 0.02 M of each carboxyl acid
C	0.5 M Buffer, pH 7.5 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD 0.02 M of each carboxy l acid:	0.5 M Buffer, pH 7.5 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD **1 M sodium potassium l -tartrate**	0.1 ml Buffer, pH 7.5 - - **35% PEG MME**- 0.02 M of each carboxyl acid
D	0.5 M Buffer, **pH 8.0** 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD 0.02 M of each carboxy l acid:	0.5 M Buffer, pH 7.5 12.5% w/v PEG 1000 12.5% w/v PEG 3350 12.5% w/v MPD **0.5 M sodium oxamate**	0.1 ml Buffer, pH 7.5 - - **30% MPD** 0.02 M of each carboxyl acid

Optimisation, crystallisation hits

The number system refers to crystallisation conditions in Table S29 and Table S30.

50% protein:

Crystals:

1D (several very small crystals)
2A (1 crystal and precipitation)
2C (several very small crystals)
2D (several small crystals and precipitation)
4D (several small crystals), 5B (two, larger crystals),
5D (several small crystals)

Precipitation:

1A, 1B, 2B, 3B, 3C, 4B, 4C, 5C, 6B,

25% protein:

Crystals:

1A (several small crystals, and one larger)
1D (1 larger crystal and presentation)
2D (1 small crystal and precipitation)
4A (three small crystals)
5D (precipitation and 1 larger crystal)

Precipitation:

1B, 1C, 2A, 2B, 2C, 2C,3A, 3B, 3C, 3D 4B, 4C, 4D, 5A, 5B, 5C

Section H: Vector

The original pFKPEN model was created by Geir Åge Løset (Centre for Immune Regulation and Department of Biosciences, University of Oslo, Norway), in this version the scFv have been sketched into the multiple cloning site.