4G9: An antibody that potentially imitates a cancer-related carbohydrate

Kristoffer Gudesen Solbakke

Master's Thesis in Chemistry

60 credits

Department of Chemistry
Faculty of Mathematics and Natural Science

UNIVERSITY OF OSLO

May 2019

2019

4G9: An antibody that potentially imitates a cancer-related carbohydrate

Kristoffer Gudesen Solbakke

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

Trykk: Reprosentralen, Universitetet i Oslo

Acknowledgments

After many years of education at the University of Oslo, it seems like the end is near. With a strained relationship to school did I by luck stumble upon chemistry. After being accepted to the Bachelor's program in Molecular biology and biological chemistry didn't it take long for my attraction to biochemistry and especially protein chemistry to evolve. The immense complexity of proteins and all-consuming functional abilities did fascinate me for the day I discovered them. My pursuit to understand proteins is essentially why I elected to apply for a Master's position in the group of my supervisor Ute Krengel[1].

I have to start by thanking my main supervisor, Ute Krengel for being a clear voice in the chaos that is my mind. For directing my Master's project through a path that at the moment of revelation was not clear to me. Also, a huge thanks to my co-supervisor Hedda Johannesen[2], who through guidance and assistance have been there by my side throughout my entire Master's, always ready to help, answer questions, and discussing me through my every problem.

Thanks to my co-supervisor, Geir Åge Løset[3] who takes more time than it seems like he has available to answer questions and discuss the status of my project whenever we meet. To Lene Støkken Høydalh[4] and Rahel Frick[5], who gifted us the two vectors (pLNOH2 & pLNOκ), and who opened your lab to me and assisted me to troublesome cloning times.

Thanks to co-supervisor Kaare Bjerregaard-Andersen[6] Gabriele Cordara[6], Joel Heim[6], Henrik Sørensen[2], and Helen Thorbjørnsrud[6], for answering my questions whenever asked. And thank you to all past and present members of Ute Krengel's group for making it a great research group. A special thanks to Helene Mykland Hoås[5] for helping me in the introduction phase of my Master's project, and making the transition between our two theses a smooth one.

[1] Professor, Department of Chemistry, University of Oslo, Norway
[2] M.Sc., Department of Chemistry, University of Oslo, Norway
[3] Scientist, Center for Immune Regulation and Department of Biosciences, University of Oslo, Norway
[4] Doctor, Institute of Clinical Medicine, University of Oslo, Norway
[5] M.Sc., Department of Bioscience, University of Oslo, Norway
[6] Doctor, Department of Chemistry, University of Oslo, Norway

And last but not least, thank you to my mother, father, and brother for being my family and always being there for me. Finally, thanks to my friends who have helped me throughout the years, you know who you are.

UiO, May 2019

Kristoffer Gudesen Solbakke

Summary

The faith of receiving a diagnosis as abysmal as cancer is one no one deserves. The diagnosis has, for a long time, been linked to one of the world's leading causes of death. Being able to alter this perception by developing medicines that not only can increase the patient's lifespan with a short time but making the patient able to outlive their cancer without compromised life quality, this is what we are striving for.

Knowing the complexity of the immune defense, and being able to manipulate it to detect and destroy cancer cells selectively is the ideal though. Now, this thought seems to be in reach. Already in use is trastuzumab[1], cetuximab[2], and multiple others[3], which recognizes cancer based on cell surface protein receptors. The new wave of immunotherapy with mAbs and receptor inhibitors (e.g., Afatinib) targeted cancer therapy has improved the overall cancer survival rate (progression-free survival) significantly, especially for lung cancer patients[4].

A multitude of cell surface receptor drugs have been developed, but few of these are targeting gangliosides. Even with the well-documented ganglioside *N*-glycolyl GM3 as a pronounced anti-tumor marker[5]. Some drugs are on the marked, such as Racotumomab (1E10)[6], but considering the regularity for cancers to express *N*-glycolyl GM3, one would expect more. Our lab has the last decade been working on the anti-*N*-glycolyl GM3 antibody 14F7, and the 14F7 anti-idiotypic antibody 4G9. We want to know if 4G9 binds to 14F7 in the same way as *N*-glycolyl GM3 does. Hence, mapping the possibility of 4G9 as an anti-*N*-glycolyl GM3 vaccine to induce host 14F7 production in an immune response.

The goal of this thesis was to make a stable expression and purification protocol for scFv 14F7 and scFv 4G9. Followed by solving their co-crystal structure and evaluating the similarity between 4G9's and *N*-glycolyl GM3's interaction with 14F7. Hence, concluding the potential of 4G9 as an anti-tumor vaccine.

We found that scFv 14F7 C4* and C4# expresses best with the autoinduction set up. And that the highest protein extraction yield is with two-step osmotic stress, first with 25 % sucrose extraction, followed by 5.0 mM $MgCl_2$ + 0.15 $^{mg}/_{mL}$ lysozyme extraction. 4G9, on the other hand, did not express with either of the tried protocols. Also, when the switch from a bacterial to a humanized (HeLa) expression system has proven tougher than anticipated, because of multiple hiccups with the cloning procedure.

Abbreviations

AC	Affinity chromatography
AP	Alkaline phosphate
ASC	Antibody-secreting cell
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
$CXCL_{13}$	Chemokine ligand 13
$CXCR_5$	Chemokine receptor 5
Da	Dalton
DC	Dendritic cell
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, crystallizable
FcRn	Neonatal Fc receptor
Fv	Fragments of the variable domains
GC	Germinal center
His-tag	Polyhistidine-tag (6x)
HRP	Horseradish peroxidase
ICOS	Inducible T cell costimulatory
ICOS-L	ligand B7RP-1
Ig	Immunoglobulin
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Lac	Lactose
LB	Lysogeny broth
LB-AG	LB containing 100 mg/l ampicillin and 0.1 M glucose
L_C	Elongated Cuba linker
L_R	Rikshospital linker

mAb	Monoclonal antibody
mAU	Milli absorption unit
MES	2-(*N*-monopholino)ethanesulfonate
MHC	Major histocompatibility complex
MS	Mass spectrometry
mqH_2O	Milli-Q filtered and ion-exchanged water
N-terminus	Amino-terminus
NEB	New England Biolabs
NeuAc	*N*-acetylneuraminic acid
N-acetyl	*N*-acetylneuraminic acid
NeuGc	*N*-glycolyl neuraminic acid
N-glycolyl	*N*-glycolyl neuraminic acid
ON	Over night
PCR	Polymerase chain reaction
PBS	Phosphate buffered saline
PBSB-BT	PBS containing 2.0 % BSA and 0.05 % Tween-20
PBS-T	PBS containing 0.1% Tween-20
PBS-TM	PBS containing 0.1% Tween-20 and 5% skimmed milk
PEG	polyethylene glycol
RE	Restriction enzyme
RNase	Ribonuclease
RS	Recognition site
scFv	Ringle-chain Fv

SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SLO	Secondary lymphoid organ
SPR	Surface plasmon resonance
TAE	Tris acetate EDTA
TCR	T cell receptor
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 content

1. Introduction

1.1 Cancer and cancer treatment

Cancer is a devastating disease that seems to strike a person at random because of the complexity in its causes. Globally, 1 of 6 death is due to cancer, which makes it the second largest cause for death[7]. Because of the abysmal outcome of the disease, a lot of resources are put in place to try and tackle the disease and to make reason out of how we can improve the outcome for the people inflicted by harmful tumor growth.

1.1.1 Cancer

Tumor growth is the result of some cellular function related to control of cell division and proliferation gone wrong. A tumor can be a simple abnormal growth that can be removed (benign), or become cancerous and spread to a different place in the body (malignant) and invade a distant tissue (metastasis). Tumor evolves from a cell's lost ability to halter cell division at the appropriate time. A multitude of proteins has upon their inactivation the ability to cause tumor growth by the loss of downstream regulation. Even though multiple regulatory proteins can inflict tumor growth, most are a result of loss in the regulation of the retinoblastoma protein (pRb)[8]. pRb controls the transition from growth and neutral cellular function to the replication of new DNA for cell division (Restriction point). Inactivation of pRb results in premature restriction point transition and cell division. Repeating the premature cell divisions can alter what the cells produce in regards to their molecular composition[8]. The alteration in the surface molecules expression is significant from an immunotherapeutic standpoint.

1.1.2 Traditional cancer treatment

Chemotherapy is giving the patient a chemical (either intrathecal, intra-arterial, intraperitoneal intravenous, injection, oral, or tropical) that are carried in the blood to the cancer tumor and halters its cell division (cytotoxicity)[9]. All though chemotherapy is effective in killing cancer cells, it does also inflict healthy cells with the same cytotoxic effect. Chemotherapy results in much discomfort for the patient when receiving this treatment.

Radiation therapy (a.k.a. radiotherapy) is a more targeted therapeutic approach than chemotherapy, but do damage the healthy cells local to the tumor as much as the tumor itself. The purpose of radiotherapy is to damage the cancer cells DNA, which results in necrosis or apoptosis of the tumor cells. Radiation therapy is quite swift, but the outcome of the treatment may take some time to set in, and cells with damaged DNA may die up to months after ended treatment[10]. Radiotherapy is mainly divided into two types, internal and external beam radiotherapy. In external beam radiotherapy is a bream source for photons (θ), protons (p^+), and electrons (e^-) concentrated to the cancerous area[11]. Internal radiation therapy works through giving a patient orally, intravenously, or by precisely placing a solid source (brachytherapy[12]) of radiation emitting particles (p^+, e^-, and gamma (γ)) inside the tumor.

Surgery is used in an attempt to remove the cancerous tumor to stop it from spreading. The incision can be either to remove a part of or the entire tumor. A reason for only removing a part of the tumor may be that the tumor grows to close to a vital organ that will be in danger of damage (debulking)[13].

1.1.3 Targeted Cancer Therapy

Targeted therapy[14], is the types of treatment that are directed at the location, type, and molecular composition of a given cancerous tumor. Targeted treatment involves hormone therapy, immunotherapy, and small-molecular drug treatment.

Small-molecule drug therapy is giving the patient, in the form of a tablet, molecules that are small enough to enter the cancer cells without the help of transmembrane transport (passive transport).

Hormone therapy is providing a drug that inhibits the production of a specific hormone. The hormone is often linked with activation of cell division, and the act of inhibiting its production halters tumor growth[15].

Immunotherapy is the stimulation of the patient's immune system to better fight the cancer. Immunotherapy can be divided into two main approaches. One, giving the patient a monoclonal antibody (mAb) that stimulates the patient immune defense, improving its fight against the cancerous growth (active immunotherapy). The second, providing a mAb that does not stimulate the immune defense but instead directly attacks the cancer cells (passive immunotherapy)[16].

The different types of cancer treatment are often combined with regards to each patient's specific case. It is, for example, not uncommon to combine chemotherapy, radiotherapy, and tumor-related targeted treatment, to kill the primary tumor and the potential metastasis. Followed by the removal of the tumor with surgery.

1.2 Tumor antigens

The composition of molecules that covers the surface of each cell is well known to the body's immune system. So, when an unknown (non-host) organism enters the body is it recognized by the body's immune defense as "not one of its own," and defensive measures are initiated. It is the surface molecules of the non-host organism that the immune system recognizes. These surface molecules are collectively known as antigens[17]. Even though most antigens are of non-host origin, some are also produced in the host, often by tumor cells that malfunction in their mechanism in the production of healthy host surface molecules. These tumor cells should be recognized by the host immune system and killed. However, on the off chance that they are overlooked, the tumor cells are able to multiply to the point of metastasis, a cancerous tumor.

1.2.1 Gangliosides as tumor antigens

On the outer surface of all cells, there are a variety of different molecules. The most common molecules have their basis in carbohydrates. Either, carbohydrates bound to a membrane protein (glycoprotein), or carbohydrates directly bound to the phospholipids layer of the cell membrane (glycolipid). The synthesis of membrane proteins and lipids happens in the endoplasmic reticulum (ER), which also is the site for adding carbohydrates to membrane proteins, making membrane-bound glycoprotein. The membrane proteins and glycoproteins are fused into a membrane-enclosed droplet (vesicle). The vesicle is transported to the Golgi apparatus for additional modification or directly to the cell membrane for extracellular display. A vesicle that fuses with the Golgi apparatus can have the glycoproteins undergo modification. The Golgi apparatus is the organelle where the fusion of carbohydrates with lipids to form glycolipids takes place. The completely modified lipid molecules are released from the Golgi apparatus as vesicles. The inside of the vesicle resembles the external of the cell's membrane, and the outside of the vesicle resembles the cytoplasmic facing layer of the cell. So when fusing, the interior of the vesicle is becoming the exterior of the cell and the modified glycoproteins and glycolipids are presented as surface molecules of the cell[18].

Ganglioside

Gangliosides are a type of sialic-acid-containing glycosphingolipid (GSL). The gangliosides are anchored to the lipid part of the cell membrane by a ceramide structure. The ceramide of interest to us comprises of an 18 carbon amino alcohol, a sphinganine, and an 18 carbon (C18) fatty acid chain[19]. The sphinganine and fatty acid C18 are linked by an amide bond catalyzed by the enzyme ceramide synthetase 1 (CerS1), making dihydroceramide[20]. Finally, is dihydroceramide dehydrogenized by enzyme dihydroceramide desaturase 1 (DEGS1) to C18:1(4) amide alcohol, also known as sphingosine. The final product is a ceramide[21]. The ceramide is synthesized in the ER compartment and transported the Golgi apparatus where the carbohydrates ($_D$-glucose, $_D$-galactose, and *N*-acetylneuraminic acid, respectively) are added[5]. The full extent of ceramide transport is not yet known but is considered a non-vesicular transport aided by the ATP-dependent transport factor CERT[22]. Their respective enzyme glycosyltransferases add the carbohydrates. First is D-Glucose added to ceramide by ceramide glucosyltransferase (Cer-Glc-T), followed by the activity of galactosyltransferase (Gal-T_2; Lac-Cer synthetase) to generate lactosylceramide (Lac-Cer). Finally, the addition of a sialic acid to the galactose of the Lac-Cer by sialyltransferase I (ST-I) to finalize the molecule GM3[23]. *N*-glycolylneuraminic acid (*N*-glycolyl) and *N*-acetylneuraminic acid (*N*-acetyl) are two of the most abundant sialic acids among all vertebrates. However, humans, due to the deletion of a 92 bp sequence, lack the gene necessary for generating *N*-glycolyl. The 92 bp sequence is the gene for enzyme cytosine monophosphate *N*-acetylneuraminic acid (CMP-NeuAc) hydroxylase (*cmah*), which oxidizes the terminal methyl of *N*-acetyl to methanol and an *N*-glycolyl (Figure 1.2-1)[24]. *N*-glycolyl is still present on the surface of fetal and cancerous cells, coining it an "onco-fetal" antigen. Patients with non-small-cell lung carcinoma (NSCLC) with high expression level of *N*-glycolyl exhibits low survival rate[25]. Delay in cell growth, regulation of proliferation and differentiation is broth on by N-acetyl GM3 binding to the extracellular domain of epidermal growth factor receptor (EGFR) tyrosine kinase inhibiting its dimerization, even when ligand is bound[26,27]. This inhibiting activity is lost when *N*-acetyl GM3 is reduced to *N*-glycolyl GM3[28]. Silencing of *cmah* gene in *N*-glycolyl GM3-expressing L1210 mice lymphocytic B cells shifts the expression to *N*-acetyl GM3, and the tumor starts shrinking[28].

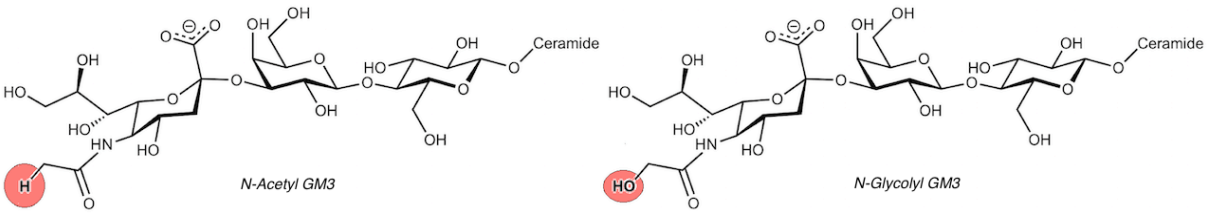


Figure 1.2- 1. Illustration of the trisaccharide elements of *N*-acetyl GM3 and *N*-glycolyl GM3, their only difference circled in red.

1.3 Anti-tumor antibodies

Since the ganglioside *N*-glycolyl GM3 is absent in normal healthy human tissue, its presence makes it a good indicator that something abnormal is evolving. Hence, can *N*-glycolyl be used as an indicator as well as an active target for stimulated immune responses, an antigen. Furthermore, since *N*-glycolyl GM3 is highly expressed on tumor cell surfaces, can it be considered a tumor-antigen.

1.3.1 Antibodies and Medicine

Upon the evolution of a molecule, as *N*-glycolyl GM3 which for a human host is considered a non-host antigen (or epitope), the immune defense generates a response. The response consists of making proteins that recognize the non-host antigen, binds strongly to it and signals for the epitope owner's destruction. Antibodies are one such type of host-defending molecule, but they are only made once the host has become infected (adaptive immunity).

The immune defense is generally divided into two main brackets, the adaptive and innate immune system. When we are born, we are given by our mother and father a basal composition of defensive molecules against the most common pathogens encountered by generations before us. This is the innate immune system, the immune defense that we are born with. But the moment we leave our mother's womb we are going to encounter a whole world of new pathogens that neither your mother nor your father can protect you from. Here is where the adaptive immune defense comes into play.

The adaptive immune defense is composed of a multitude of different molecules and complex mechanism, so for simplicity, the explanatory road from pathogen encounter to antibody activation and activity will here go through the mechanism of the dendritic cells (DCs). The DCs are a member of the innate immune system and are present in all peripheral tissue. This makes them the first line of defense once we get a wound, exposing our internal systemic system to the outside world's pathogens. The DCs can destroy the pathogen upon interaction by consuming the pathogen (phagocytosis). DCs perform degradation by relying on nucleophilic radicals ($O_2^{\bullet}$ or $N_2^{\bullet}$), antimicrobial peptides, and proteolytic enzymes, the same mechanisms as macrophages and neutrophils (other members of the innate immune system). The result of the

phagocytic digestion is short peptide fragments. The DCs then presents short peptide fragments from the pathogen surface, on their major histocompatibility complex (MHC). Once the DCs have an antigen, they start migrating towards the lymph node (Figure 1.3-1). The DC upregulates the expression of their chemokine receptor CCR7, which is guided by the chemoattractants CCL21 and CCL19 toward the lymph node through the lymphatic system and to the spleen through the blood stream, collectively known as secondary lymphoid organs (SLOs). The DCs enters the SLO and presents the antigen to helper T-cells, killer T-cells, and B-cells (T-cell independent activation)[17]. The DC's MHC class II (with antigen) interact with the T-cell receptor (TCR) of the immature CD4 T-cell together with the co-stimulator CD278 (ICOS). The ICOS (inducible T-cell co-stimulator) binds to its ligand B7RP-1 (ICOS-L) on DC[29], stimulating the release of IL-21 from DC. IL-21 binding determines the maturation of the T cell to become a follicular helper T-cell (T_{FH} cell). The mature follicular helper T-cell downregulates CCR7 whiles upregulating the chemokine receptor 5 ($CXCR_5$), and the T_{FH} cell is drawn towards the B-cell area of the SLO by chemokine ligand 13 ($CXCL_{13}$), expressed by the stromal network in the boundary region[30]. $CXCL_{13}$ draws $CXCR_5$ expressing T_{FH} and naïve B cells together in the boundary region (or follicular region). The T_{FH} cell TCR bind to the MHC class II on the naïve B cell, this in correlation with IL-21 released by the T_{FH} cell maturates the naïve B cell into an antibody-secreting cell (ASC)[31]. The B cells mostly relying on the maturation by T_{FH} cell (T cell dependent) antigens are classified as follicular B cells. After activation into an ASC, depending on the stimuli, the follicular B cell can enter one of two ASC pathways[32]. In one, known as the extrafollicular response, the ASCs are traveling from the follicular region of SLO to the site of infection as plasmablasts, short lived ASCs with low antibody diversity. The extrafollicular response is responsible for the majority of the early antibody production[32]. In the second, the activated B cells reenter the B cell area and start proliferating to form a germinal center (GC). The GC generates highly selective (high affinity) long lived plasma cells (the second class of ASC). It is in the GC that memory B cells are generated, this by suppression of the *BACH2* gene. The transcription factor Bach2 is regulating the transition from long lived memory B cell to long lived plasma cell[33]. Mature ASC leaves the SLO through the systemic fluids circulating the host in search for their favored antigen.

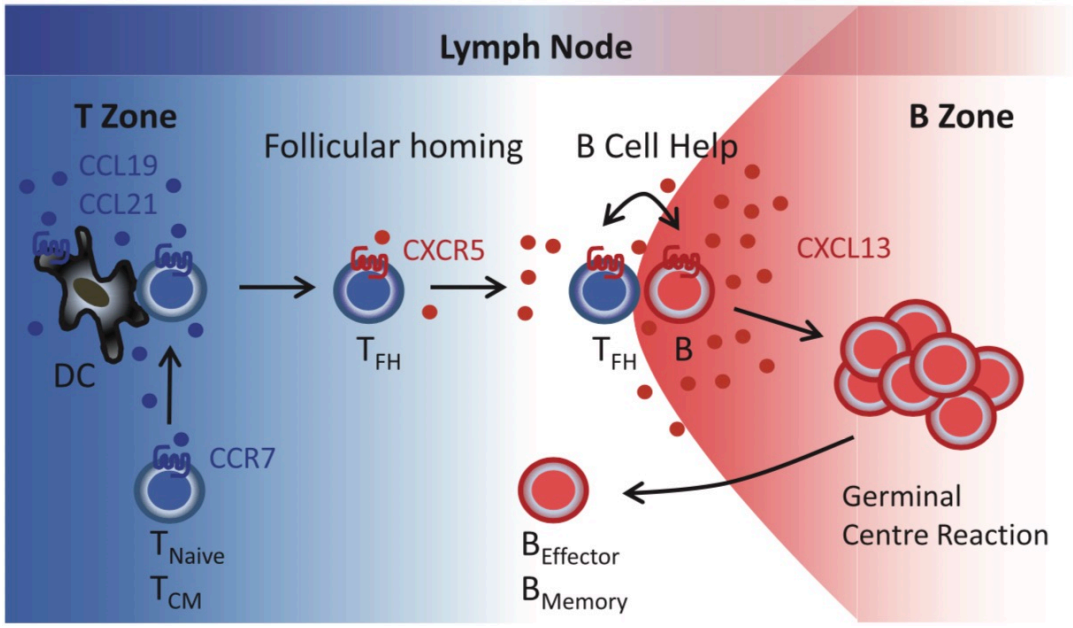


Figure 1.3- 1. The process of T cell dependent B cell activation. Illustration is reproduced form Bernhard Moser (2015)[34]

1.3.2 Antibody structure and classification

Antibodies are produced by mature B cells when activated by antigen interaction. Antibodies can be imagined as simple Y-shaped proteins that uses its arms to bind to its antigen. However, antibodies are not one type of protein structure, even though the foundation of their structure is the Y-shape. There are, in total, five classes of immunoglobulin (Ig): IgA, IgD, IgE, IgG, and IgM (Figure 1.3-2).

The simples Y-shape is the one of IgG and IgD, which are monomers. They consist of eight heavy chain (HC) fragments and four light chain (LC) fragments. Each of the different fragments has a β-sandwich structural conformation. The arms of the Y-shape are the antigen binding fragments (Fabs) and consists of two HC and two LC fragments in each arm. The LC and HC are connected by a disulfide bond between the constant heavy (CH) and constant light (CL) chains of their Fab. The LC can take one of two isoforms, as kappa (κ) or lambda (λ) light chains. The "top" of the arms is mainly where the antigen binding happens, with the help of six highly diversifiable loops (three on each HC and LC) named complimentary determining region (CDR) loops 1, 2 and 3. The two "top" fragments of the Fab with the CDR loops are conveniently named the variable heavy (V_H) and variable light (V_L) fragment or chain. On the "bottom" of the two Fabs, the HCs extends in additional β-sandwich structural fragments. These

highly conserved regions of the antibody structure were the one first crystallizable fragment after papain treatment and exercised no antigen binding, hence are named fragment crystallizable (F_C)[35]. The F_C region of IgG and IgD consists of four β-sandwich fragments (two tailing each of the two Fab arms) and is where the phagocytes bind to the Ig when the Ig is bound to its antigen.

In the Fab region, most Ig's are built the same, so the diversification between the five Ig classes is in the F_C region and their host localization. IgE is similar to IgG and IgD a monomer but has six F_C β-sandwich fragments. Similar to IgG and IgD, has IgA four F_C β-sandwich fragments that IgA can dimerize. The two Y-shapes of IgA are linked between their terminal F_C regions by disulfide bonds, orienting the Fab arms of each IgA monomer 180° to each other. IgM has six F_C fragments joining together five IgM monomers their circular pentameric structure, also linked together by disulfide bonds.[17] Numeration of the different constant heavy (CH) fragments starts at the top, CH1 is a part of the Fab chain whiles CH2 and CH3 makes up the F_C chain. IgM and IgD do also have a third CH F_C fragment, CH4. All classes of Ig's have multiple glycosylation hot spots in their F_C chain. One of these, Asparagine (Asn)-180 (IgG1), Asn-176 (IgG2), Asn-227/322 (IgG3), Asn-177 (IgG4), Asn-225 (IgD), Asn-275 (IgE), and Asn-279 (IgM) are collectively known as CH2-84.4 (and CH3-84.4 for IgE and IgM)[36]. As mentioned, phagocytes bind to the F_C chain of Ig's but do so with their F_C gamma receptor ($F_C\gamma R$). The $F_C\gamma R$ binds to *N*-glycosylated CH2-84.4 but only to the IgG(1-4) class. This binding is significant when considering an antibody from a screening assay for medical purposes[36]. What type of glycan bound to CH2-84.4 dictates what type of $F_C\gamma R$ that can bind the IgG, hence determine the destiny of the antibody bound cell[36].

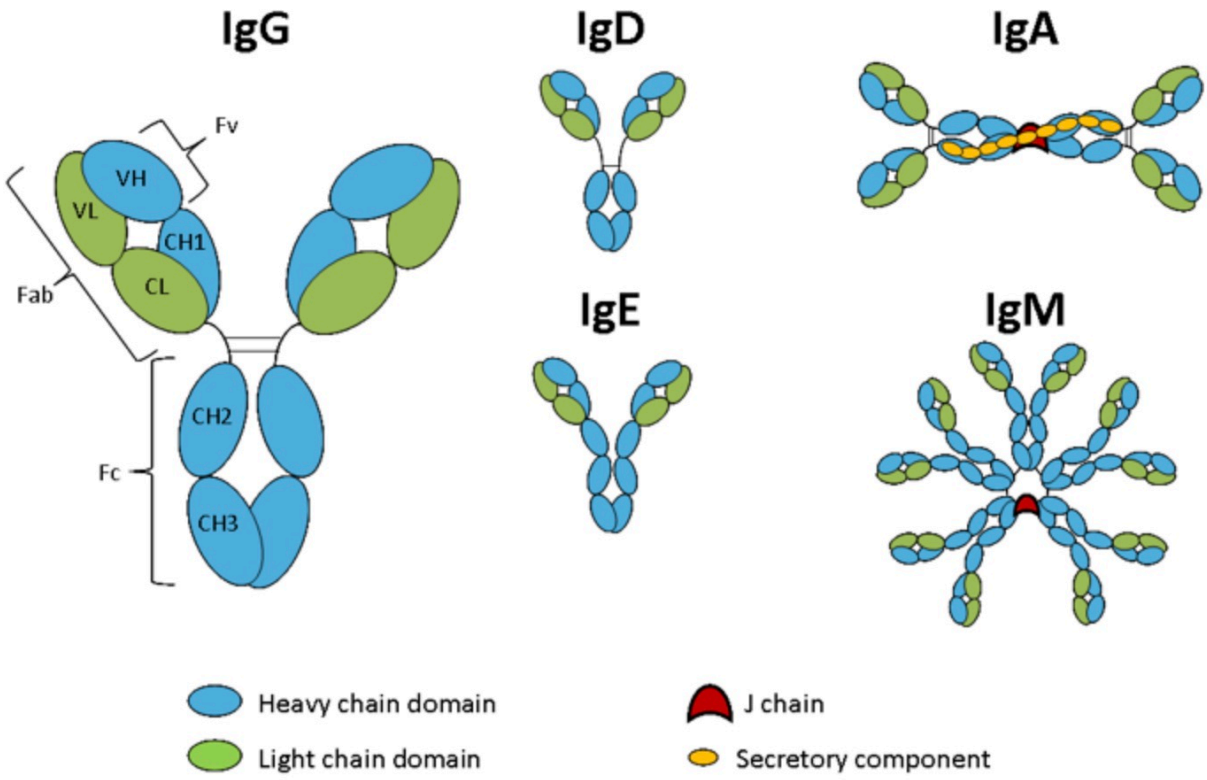


Figure 1.3- 2. Antibody subclasses. Schematic presentation of the different immunoglobulin subclasses. CH, constant heavy chain; CL, constant light chain; VH, variable light chain; VH variable heavy chain; F_C, fragment crystallizable; Fab, fragment antigen binding; Fv, variable fragment. This illustration is reproduced from Absolute antibody[37].

1.3.3 The anti-tumor antibody 14F7

Stimulation and manipulation of a process as the adaptive immune system are of central importance in medicine. So with the discovery that *N*-glycolyl GM3 was a tumor-related antigen[5], the research in targeted immunotherapy towards this antigen begun. Multiple antibodies have been generated in response to *N*-glycolyl GM3 as antigen. Screening with *N*-glycolyl GM3 as antigen did mostly generate monoclonal antibodies (mAbs) of the class IgM[38]. However, the research of Adriana Carr et al. (2000) discovered an IgG1 mAb, namely 14F7[38]. 14F7 was found through histochemical binding experiments to be selective towards melanoma and breast tumors expressing *N*-glycolyl GM3[38].

Even though 14F7 is not alone able to destroy an entire tumor that is already proclaimed in a patient, the presence of 14F7 helps to prevent the growth of new tumors with *N*-glycolyl GM3 as surface molecules. This means that 14F7 might not be a suitable primary treatment option

against a tumor but instead as a potential vaccine or as a supplement in adjuvant therapy[39]. When exploring the effect of 14F7 upon interaction with murine leucocytic leukemia cells (L1210), a cytotoxic effect was discovered. Binding of 14F7 mAb and $F(ab)_2$ induced and oncosis-like cell death mechanism towards *N*-glycolyl GM3 L1210 cells, whiles 14F7 Fab did not[40]. $F(ab)_2$ are two Fabs connected through their original hinge region (an IgG mAb with CH2 and CH3 cleaved off).

Before using a synthetically produced antibody as anti-tumor treatment, the precise mechanism of interaction between antibody and antigen needs to be known. Otherwise, the interactions that are empirically found may be different from what is actually happening. Observation of the interaction between an antibody and its antigen can be done by X-ray crystallography and has been done for 14F7 Fab and *N*-glycolyl GM3[41]. The solved structure of 14F7 Fab indicated that the interaction with *N*-glycolyl GM3 is with the variable CDR loops. It was discovered that 14F7 has an unusually long CDR H3 loop (16 aa), which is responsible for most of the antigen binding. These findings were supported by point mutations done on the LC residues that are in a position to interact with the antigen. The mutation result indicated that interaction with *N*-glycolyl GM3 is with HC residues, and mainly the CDR H3 loop[41]. However, removing the LC completely resulted in the loss of binding, meaning that the LC is essential for the integrity of the HC.

The crystal structure of 14F7 Fab by Krengle et al. (2004) gave a lot of information, but the refined density map did not cover for the residues in the CDR H3 loop responsible for binding. Making crystal structure out of protein is an unfavorable physiological exercise. So, reproducing the crystallization of 14F7 Fab has not since been successful. This has led the 14F7 history through many years of optimization experiments. It is generally easier to make crystal structures of smaller and less complex structures. Since 14F7 seemingly only needs the variable HC and LC to perform binding, the Fab constant HC and LC were removed to create a single chain fragment variable (scFv) construct of 14F7[42]. It was not only made one 14F7 scFv construct but four, scFv 14F7 C1*, C2*, C3*, and C4* (Figure 1.3-3). The linker region (disulfide bond) that holds the HC and LC of Fabs together is located between the constant chains of HC and LC. The generation of scFv constructs removes these disulfide bonds and artificial linkers where made. The two linkers for scFv 14F7 were synthetically made, one by Carr et al. in Cuba (Centre of Molecular Immunology, Havana, Cuba, CIM) and one at Rikshospitalet in Norway (Centre for Immune Regulation and Department of Biosciences, University of Oslo, Norway, CIR). The Cuba linker (L_C) are linking the variable heavy (V_H)

and variable light (V_L) chain of scFv 14F7 C3* and C4* constructs. Whiles the Rikshopitalet liker (L_R) is used in the scFv 14F7 C1* and C2* constructs. ScFv 14F7 C4* is the construct with the original Fab V_H and V_L. The scFv 14F7 C2* did also have the original V_L chain. The C1* and C3* constructs sheared an expression and purification optimized alternative V_L chain ($V_{L, A}$)[43]. By using the scFv 14F7 C1* construct the residual composition, and *N*-glycolyl GM3 interaction by the CDR H3 loop was documented[42]. Moreover, it still resonates that the long CDR H3 loop is responsible for recognizing the antigen, but that indirect interactions from V_L CDR loop contribute in the overall affinity[42].

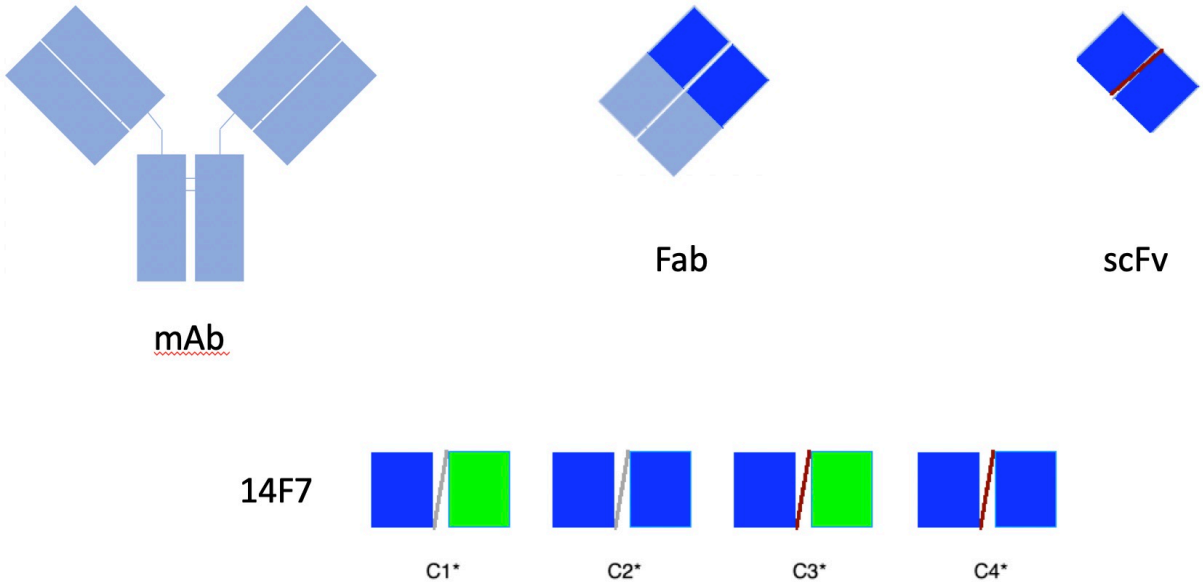


Figure 1.3- 3. Illustration of 14F7 from mAb to scFv. On the bottom is the four scFv constructs that were created44. (Bottom line) Rikshospitalet linker (LR, gray), Cuba linker (LC, red), alternative light chain (VL, A, green).

1.3.4 The anti-idiotypic antibody 4G9

When the mAb 14F7 was discovered as a potential anti-tumor therapeutic antibody, screening for potential anti-idiotypic antibodies revealed the mAb 4G9. A pair of anti-idiotypic antibodies are antigens to each other, i.e.; their respective CDRs interact in binding. MAb 4G9 (Ab2) showed the ability to inhibit mAb 14F7 (Ab1) binding to *N*-glycolyl GM3. Ab2 where also able to generate a mAb Ab3 (anit-anti-idiotypic antibody) when used to immunize mice (Figure 1.3-4). The Ab3 are able to bind specifically to *N*-glycolyl GM3, therefore are thought to be similar to Ab1[45]. This indicates that the mAb 4G9 idiotype (epitope-binding region) has the same chemical properties as *N*-glycolyl GM3, and might be a potential vaccine against *N*-glycolyl GM3 tumor cells.

Before applying 4G9 as an anti-tumor vaccine, the specific interactions between 4G9 and 14F7 needs to be discovered. X-ray crystallography to solve the co-crystal of 14F7:4G9 would be preferable. This was the starting point for the work in this thesis.

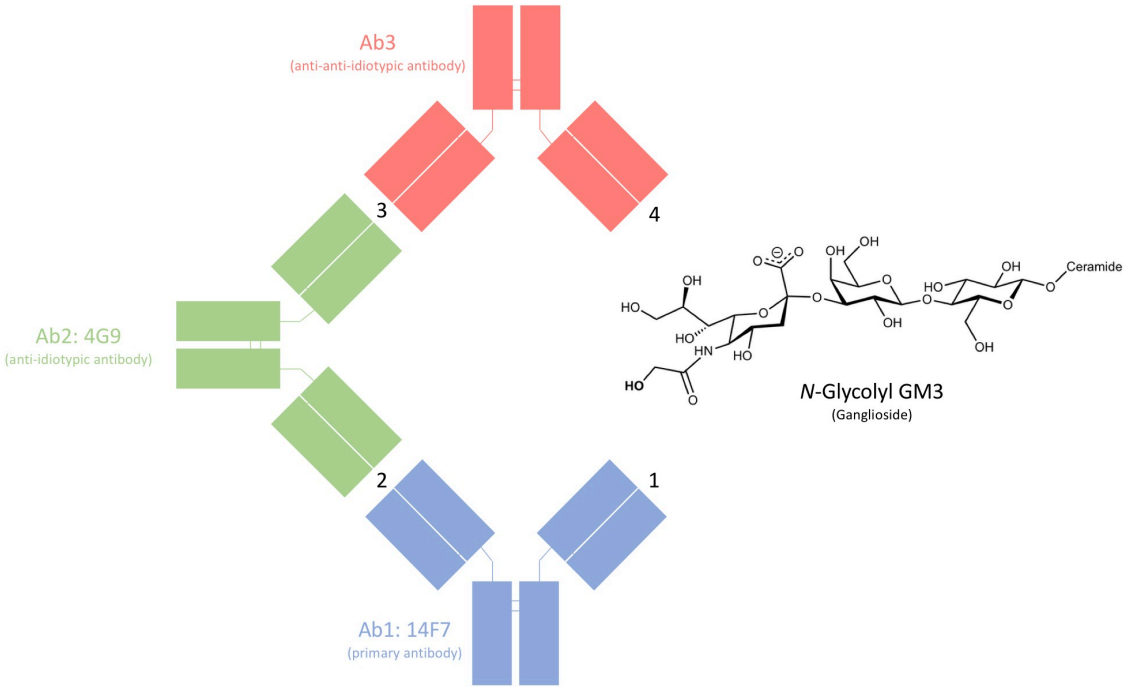


Figure 1.3- 4. Schematic representation of the relationship between the ganglioside N-glycolyl GM3 and the antibodies produced in response to each other. Mice were immunized with N-glycolyl GM3, 14F7 (Ab1) was created (1), succeeded by injection into a different batch of mice resulting in the production of 4G9 (Ab2) (2). Injection of 4G9 (Ab2) into N-glycolyl lacking chicken, Ab3 is the resulting chicken immune response (3). The chicken-made antibody (Ab3) was selective towards N-glycolyl GM3 (4), in a similar way as 14F7 (1)4.

1.4 Method related theory

One obstacle when working with eukaryotic protein expression system such as the murine Balb/c and L1210 used to discover the mAb 14F7 and mAb 4G9 is the long generation time of the host. If you like to do some simple genetic alterations to see the influences of the final protein product you would have to wait longer than the result is worth. This is one of the main reasons why when exploring the manipulation of the residual composition of your protein, one usually switches from a complex eukaryotic to a more simplistic bacterial host.

1.4.1 Bacterial expression of recombinant scFv antibodies

The kingdom of bacteria is of the domain prokaryote and are generally described as microorganism. Their size is on the micrometer scale and is divided into different classes based on their intercellular mechanism of energy consumption (cellular respiration). Bacteria are also divided based on their extracellular buildup, the biochemical composition of their cell wall and cell membrane. The cell membrane composition is of phospholipid bilayer similar to the one of the eukaryote cell membrane. The cell wall is composed of a peptidoglycan layer, a network of the glycans *N*-acetylglucosamine and *N*-acetylmuramic acid connected by β-(1-4)-glycosidic bond. The *N*-acetylmuramic acid has 4-5 amino acids extended form their 3'-carboxyl group. For the bacteria, *Escherichia coli* (*E. coli*) are these residues; $_{L}$-alanine, $_{D}$-glutamate, $_{L}$-lysine, and $_{D}$-alanine. The residues link the glycan chains together in a comprehensive network with cross-linking between the amino acid. The cell wall/membrane of bacteria undergo two different buildups. One, where the interior (cytoplasm) of the bacteria are closed off by on cell membrane layer and coated by a thick peptidoglycan layer (weight ratio 9:1). This thick peptidoglycan layer can visibly be colored purple by "gram staining," and is hence named gram-positive. The bacteria with this cell wall/membrane composition are considered gram-positive bacteria. On the contrary, bacteria that are not colored by gram staining have a slightly different cell wall/membrane composition, the gram-negative bacteria. The gram-negative has two membranes or phospholipid bilayers separated by a "thinner" peptidoglycan layer (weight ratio 1:9). The spacing between the two membranes is named the periplasmic region, so the peptidoglycan layer of gram-negative bacteria are a part of the periplasmic region[46].

The bacterial genome consists of the nucleoid where most of the vital information for bacterial survival and division is stored. However, bacteria do also carry smaller bits of genetic information, enclosed in circular vectors or plasmids. Since bacteria do not have any organelle organization, do their entire molecular components float around in the same cytoplasmic fluid. So, all the molecules necessary for DNA transcription and translation into proteins are excisable for whatever genomic molecule they encounter; they only need an origin of replication[47].

E. coli is a gram-negative bacterium proven to accept extracellular vectors and express their information. In this way can an artificial vector with antibiotic resistance and the ability to express proteins of choice be introduced into a fast-growing and dividing bacterial host. *E. coli* possesses the ability to translate and fold proteins directly in the cytoplasmic space, but also with the help of a tag sequence perform the folding in their periplasmic region. The amount of protein that is translated into the periplasmic region is lower than the amount transcribed in the cytoplasm. An N-terminal pelB sequence ensures the periplasmic expression of the protein. N-terminal pelB stabilizes the formation of disulfide bond making in the periplasm, and helps in correct folding and increases solvability of the proteins[48].

When using a bacterial host to express your protein, you want the bacteria to be at their fittest state and large in number. To ensure maximum fitness, one takes an overgrown culture (optical density at 600 nm (OD_{600}) > 1 absorption unit (AU)) of bacteria and dilutes it OD_{600} = 0.0250 AU, and lets the diluted bacteria grow until the OD_{600} = 0.6-0.8 AU. At which point, the growth has reached the exponential growth phase. To control the expression of your favorite protein to start at OD_{600} = 0.6-0.8 AU, position a regulatory promoter upstream of your favorite protein gene sequence. Here, we used the *lac* promoter, which is activated when lactose is present, which can be suppressed by the presence of glucose, which for the bacteria is an easier consumable source of energy.

1.4.2 Extraction of periplasmic scFv antibodies

Once you have your protein folded in the periplasm of the bacteria, you need a way of getting them out. Mixing the bacteria with a concentrated sucrose solution is one way. The higher concentration of carbohydrate (sucrose) on the outside of the bacteria membrane leads the bacteria to open their outer membrane porin channels (ScrY[49]) for transport against the

concentration gradient ($[sucrose]_{extracellular}$ >>> $[sucrose]_{intracellular}$). The influx of sucrose leads the periplasmic protein to be leaking out of the bacteria (Osmotic shock)[50,51].

Lysozyme digests the β-(1-4)-glycosidic bond of the peptidoglycan layer in the *E. coli* periplasmic region. Causing the periplasmic protein retained by the peptidoglycans to be free for movement in the extracellular direction. Lysozyme in combination with a low concentration (0.5-20 mM) of $MgCl_2$ force periplasmic protein release through osmotic shock[50,51].

1.4.3 scFv antibody purification

The osmotic shock extracted proteins are purified based on a specific chemical property (His-tag or κ-LC affinity chromatography, AC) and size (Size Exclusion Chromatography, SEC). The His-tag is an N- or C-terminal six-histidine sequence which can selectively bind to the stationary cation (Ni^{2+}, Co^{2+}, or Zn^{2+}) of the HisTrap™ (Amersham Biosciences) chromatography column[52]. Alternatively, can antibodies be purified based on the LC type, kappa (κ) or lambda (λ) LC. Our favorite protein, 14F7 has a κ-LC. Capto™ L is a resin with an affinity towards κ-LC[53]. After AC is the protein fraction purified based on size. SEC consists of resin with pores of different size. The smaller molecules that enter the pores are retained compared to the larger molecules that bypass the resin[54].

1.4.4 Crystallization

To be able to determine the structure of a protein with X-ray crystallography, one needs to make the proteins crystals. Protein crystals are a grid-like packing of proteins into repeating unit cells. The proteins interaction and orientation inside a unit cell can be repeated throughout the crystal, making a three-dimensional basis for the entire crystal structure.

Forming a crystal structure is not a favorable physiological operation for proteins. So, to make protein crystallize, the environment of the proteins needs to be even less favorable (buffer solution), forcing the proteins together and eventually forming crystals. In the the technique of hanging drop and sitting drop, are the protein diluted in a low volume drop together with a buffer solution. The droplet is placed in an isolated system together with a larger volume

(reservoir) of the buffer solution (Figure 1.4-1A). Since the concentration of the buffer solution in the droplet is lower than in the buffer reservoir, liquid water diffuses from the droplet to the reservoir (vapor diffusion) to reach a buffer concentration equilibrium (Figure 1.4-1B). In the prosses of removing water from the droplet is the concertation of protein and buffer components increased. Reducing the free space, making the general environment less favorable and forcing the proteins together into small crystal nuclei (nucleation). Over time, as the free protein are forced together, the protein nuclei grow and form larger crystals (Figure 1.4-1C) that can be used for X-ray diffraction (> 0.1 mm).[55]

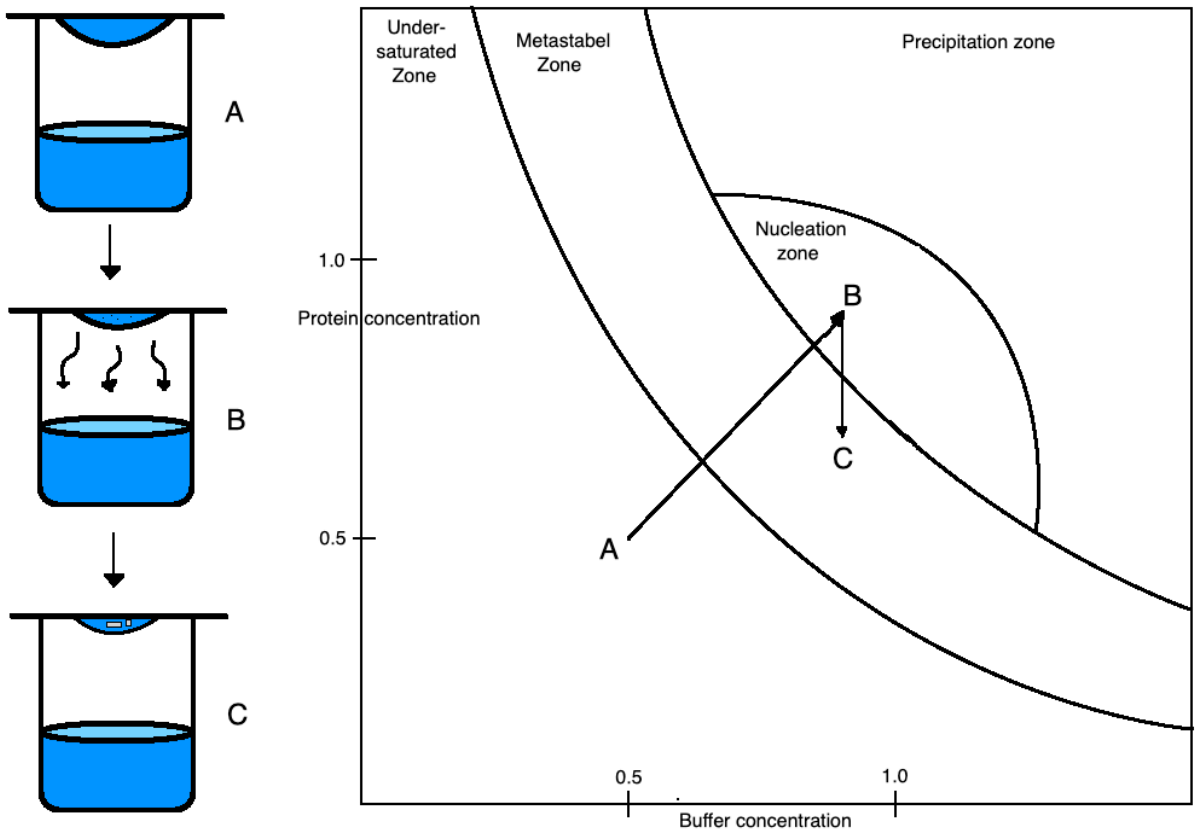


Figure 1.4- 1. Vapor diffusion of hanging drop and phase diagram. (A) Start, (B) nucleation, (C) crystal growth. (A) A 1:1 ratio of protein and buffer solutions makes the starting droplet. With vapor diffusion of water from the droplet to the more concentrated buffer reservoir increases the relative protein and buffer concentration (B). In the nucleation zone does the protein adhere together and create nuclei, which reduces the concentration of "free" protein in solution, accelerating crystal growth into the metastable zone C).

1.4.5 Indirect ELISA

Enzyme-linked immunosorbent assay (ELISA) is a technique to detect binding between small quantities of antibody to its antigen. A small volume well is coated with the antigen. The primary antibody is then applied to the antigen coated well for binding. The unbound antibody is washed out of the well. In the instance of indirect ELISA, a secondary antibody is added to the well, which binds noncompetitively to the primary antibody. The secondary antibody contains the ability to react with a chromogen, that emits a wavelength for quantitative measurement of antigen:primary antibody:secondary antibody concentration[56].

1.4.6 ThermoFluor Assay

ThermoFluor Assay, when used to measuring protein melting point, is based on the ability of tryptophan to absorb light at wavelength 330 nm and emit light at 350 nm. Tryptophan has a hydrophobic functional group, which favors the interior of the folded protein, away from the more hydrophilic environment. The sudden environmental appearance of tryptophan when raising the temperature of the protein solution indicates that the protein no longer can sustain its natural fold. The protein becomes unstructured with increased temperature hence revealing their tryptophan to the environment, but more importantly, the removal of the hindering or quenching effect by surrounding amino acids on tryptophan. Removing the quenching effect of neighbor residues can be observed as a shift in 350/330 ration to a higher 350 nm emission[57].

2. Aim of thesis

The structure of scFv 14F7 C1* has been solved and the conformation of the residues of the long CDR H3 loop have been determined[42]. Next, knowing how 14F7 and the CDR H3 loop interacts with its anti-idiotypic antibody 4G9 — establishing if 4G9 is a potential anti-tumor vaccine for *N*-glycolyl GM3 expressing tumor cells.

With the use of an *E. coli* expression host produce scFv constructs of 14F7 C4* and 4G9 by periplasmic expression. When a sufficient expression and purification protocol for both constructs are established, and sufficient amounts of protein are collected, screen for crystallization conditions favoring co-crystallization of the two anti-idiotypes.

3. Materials and methods

All the chemicals, recipes, equipment, and instruments are listed in 8. Appendix, together with their provider.

3.1 Cloning

3.1.1 Restriction cleavage

For expression of our proteins scFv 14F7 C4* and scFv 4G9 the starting vector pFKPEN were used. The restrictions sites *Nco*I (Thermo Scientific) and *Hind*III (Thermo Scientific) are flanking the V_H, and the restriction sited *Mlu*I (Thermo Scientific), and *Not*I (Thermo Scientific) are flanking the V_L domain regions. The RE sites are the same for both scFv 14F7 C4* and scFv 4G9. An artificial linker is located between the C-terminal end of the V_H chain region and the downstream N-terminal end of the V_L chain. The restriction sites *Hind*III and *Mlu*I flank the linker.

The plasmid 4G9 Fab was made by fusing the Eurofins® synthesized humanized 4G9 V_H sequence (pEX-A128-4G9_VH) with the humanized Fab HC vector (pLNOH2) and the humanized 4G9 V_L (pEX-A128-4G9_VL) with Fab LC vector (pLNOκ). Both the humanized Fab vectors (pLNOH2 and pLNOκ) were provided by Lene Støkken Høydahl (Centre for Immune Regulation and Department of Immunology, University of Oslo and Oslo University Hospital, NO-0372 Oslo, Norway). The same two restriction sites flank the 4G9 V_H and V_L. *Mva1269*I (Thermo Scientific)/ *Bsm*I (NEBiolabs) restriction site at 5' end, and *Pfl23*II (ThermoScientific)/ *Bsi*WI (NEBiolabs) restriction site at 3'end. The pLNOH2 and pLNOκ vector have the same two restriction sites flanking a dummy V_H and V_L sequence, respectively.

The restriction enzyme cutting of the plasmids was done with reagents provided by NEBiolabs and their suggested protocol (Appendix C). For 4G9 Fab construction, the restriction sites *Bsi*WI and *Bsm*I were manipulated to exchange V_H and V_L chain sequences with the two dummy sequences in pLNOH2 and pLNOk, respectively. Digestion mixture of 20 µL total

volume consists of 15 µL nuclease-free miliQ water (mqH_2O), 2.0 µL CutSmart® Buffer (NEBiolabs), 1.0 µL plasmid DNA (or 1.0 µg), and 1.0 µL *BsiW*I-HF restriction enzyme. The mixture was incubated 1 hour at 37°C. Then 1.0 µL *Bsm*I restriction enzyme was added, with incubation at 65°C for 1 hour. The digestion mixture of 20 µL is loaded on a 1.0 % agarose gel for electrophoresis at 90 V for 50 minutes, with 1xTAE buffer pH 8.3 (Appendix B). The lower band of pEX-A128-4G9_VH (380 bp) and pEX-A128-4G9_VL (347 bp), in between the 500 bp and 250 bp reference bands of the GeneRuler 1 kb DNA Ladder (ThermoScientific). The DNA fragments were extracted from the agarose gels with the Gel extraction kit (Qiagen).

The RE digestion with *Pfl23*II and *Mva1269*I followed the same reaction set up as NEB protocol (Appendix C). The only deviation between the two protocols was the temperature for optimal RE activity. Both *Pfl23*II and *Mva1269*I prefers 37°C incubation temperature. So, both RE was mixed into the reaction mixture at start.

3.1.2 Ligation

Mix, 15.0 µL mqH_2O (nuclease free), 2.0 µL T4 DNA Ligation buffer (NEBiolabs), 60 ng plasmid DNA fragment, with the mass of insert DNA equal four times the molar ratio of plasmid fragment, and finally 1.0 µL T4 DNA Ligase (NEBiolabs) (Appendix D). 60 ng (1.5 µL) pLNOκ were mixed with 8.4 ng (0.5 µL) 4G9_VL, and 60 ng (0.9 µL) pLNOH2 with 9.6 ng (0.6 µL) 4G9_VH based on the concentration of the isolated fragments (Table S-22). Incubate at 22°C (room temperature) for 1 hour. Transform 1-5 µL ligation mixture to 10-50 µL $CaCl_2$ competent XL1-BLue E. coli cells.

3.1.3 $CaCl_2$ competent XL1-Blue E. coli stock

1.0 mL 2xYT (16 $^g/_L$ Peptone, 10 $^g/_L$ Yeast extract, 5 $^g/_L$ NaCl, pH 6.8) media were inoculated with 2.0 µL Supercompetent XL1-Blue E. coli for 12-16 hours, 37°C 300 rpm. 1.0 mL inoculation mix are diluted with 99.0 mL 2xYT for growth to OD_{600} of 0.4 (3-4 hours). Centrifuged 4000 rpm at 4°C for 10 minutes. Discarded supernatant and dissolve the pellet in 20.0 mL ice-cold 0.1 M $CaCl_2$ solution, and incubated on ice for 30 minutes. Centrifuged 4000 rpm at 4°C for 10 minutes, discarded the supernatant. Resuspended pellet in 5 mL 0.1 M $CaCl_2$ 15 % Glycerol solution. Stored stocks at -80°C.

3.1.3 Transformation

1 µL plasmid DNA (500 ng) was gently mixed with 30 µL $CaCl_2$ competent XL1-Blue *E. coli* cells. The mixture was incubated on ice for 30 minutes followed by 45 seconds heat shock in 42°C water bath, directly followed by replacement on ice for two minutes. After the two minutes, addition of 900 µL 2xYT to the 30 µL transformation mix. The diluted transformation mix was incubated for 1-2 hours at 37°C 300 rpm. 50 µL of the liquid culture was plated with L-shaped bacteria spreader. Plates were left for overnight culture growth (12-16 hour) on 30 mL 2xYT-AG 1.0 % agar plates. Singular cultures were selected for 50 mL of 2xYT-AG growth (12-16 hours). The liquid culture was centrifuged at 4000 rpm for 10 minutes, followed by plasmid isolation with DNA midiprep kit (Sigma) to a final concentration of 200-1000 $^{ng}/_{\mu L}$. Alternatively, making of 1.0 mL 25 % glycerol stock by dilution of the 50 mL bacteria pellet, stored at -80°C.

3.2 Protein expression

3.2.1 Prokaryotic expression

The DNA sequences of scFv 14F7 C1*, C2*, C3*, C4*, and scFv 4G9 are expressed from the vector pFKEN (Appendix C) provided by Geir Åge Løset (Centre for Immune Regulation and Department of Immunology, University of Oslo and Oslo University Hospital, NO-0372 Oslo, Norway). The plasmids pFKPEN-scFv14F7C1*, pFKPEN-scFv14F7C2*, pFKPEN-scFv14F7C3*, pFKPEN-scFv14F7C4*, and pFKPEN-scFv4G9 are transformed into the *Escherichia coli* (*E. coli*) bacteria strain XL1-Blue Supercompetent cells (Alignet®). The plasmid (pFKEN) carry the gene for Ampicillin resistance for positive selection, and the gene *lac* operon for activation of scFv protein production in the absence of glucose and glycerol.

3.2.1 ScFv protein production

One single culture of *E. coli* is selected and inoculated with 50 mL 2xYT-AG (2xYT, 0.1 mg/mL Ampicillin, 0.1 M Glucose) for 14-18 hours (Preculture).

Manual induction

The preculture was diluted to OD_{600} = 0.025 AU with 500 mL 2xYT-AG (2.0 L 2xYT-AG divided by four 2 L baffled flasks), and incubated for 4-6 hours at 37°C 125 rpm until OD_{600} = 0.6-0.8 AU. The culture was pelletized by centrifugation at 4000 rpm. The pellet was resuspended in 500 mL 2xYT-A (2.0 L 2xYT-A divided by four 2 L baffled flasks) and left for overnight incubation at 30°C 125 rpm, for protein expression.

Autoinduction

The preculture was diluted to OD_{600} = 0.025 AU with 400 mL autoinduction media (Appendix B) and left for 24-hour incubation at 30°C 125 rpm (16 hours bacteria proliferation, and 8 hours protein production).

3.2.2 Periplasmic isolation

The induced bacteria culture was centrifuged 4°C 4000 rpm for 40 minutes. The bacteria pellet was dissolved in 25% sucrose solution (pH 7.4) in ration 4-5 $^{mL\ sucrose\ (aq)}/_{g\ bacteria\ pellet}$, and incubated on ice for 30 minutes, followed by centrifugation at 4°C 18000 rpm for 40 minutes. The sucrose supernatant is kept for protein purification. The pellet is dissolved in a $MgCl_2$ solution (5.0 mM $MgCl_2$, 1 tablet/100 mL c*0*mpete® inhibitor, pH 7.4) in ration 4-5 $^{mL\ MgCl_2\ (aq)}/_{g\ bacteria\ pellet}$, and incubated on ice for 30 minutes. Lysozyme was added to a final concentration of 0.15 $^{mg}/_{mL}$ and incubation on ice for an additional 30 minutes. The $MgCl_2$/lysozyme bacteria solution was centrifuged at 4°C 18000 rpm for 40 minutes, and the supernatant was kept for protein purification. Bacteria pellet was discarded in appropriate hazardous waste.

3.3 Protein purification

Every protein purification was done first with affinity chromatography (AC) and followed by size exclusion chromatography (SEC).

3.3.1 Affinity Chromatography

For the purification of the scFv 14F7 constructs and scFv 4G9 Äkta Purifier (GE Healthcare) at 10°C was mainly used. The scFv 4G9 was also purified at 4°C with Äkta Start (GE Healthcare).

Capto™ L column

Used a 1.0 mL Protein L (GE Healthcare) column for AC, which is selective for the κ-LC of all the scFv 14F7 constructs and the scFv 4G9. For both the 10°C Äkta Purifier and the 4°C Äkta Start was the protein samples loaded at a flow of 2.0 $^{mL}/_{min}$. The column washed for 15 mL of Protein L binding buffer (0.10 M Tris-HCl, 0.15 M NaCl, pH 7.4) with the flow 2 $^{mL}/_{min}$. Elution of scFvs from the Protein L column with Protein L elution buffer (0.10 M Glycine, pH 2.5) at a flow of 2 $^{mL}/_{min}$. The column was equilibrated with 15 mL Protein L binding buffer between different protein samples.

HisTrap™ column

For the purification of scFv 14F7 C1#, C2#, C3#, C4#, and scFv 4G9, which all has a C-terminal hexahistidine sequence tag (His-tag), 5.0 mL HisTrap™ column (Amersham Biosciences) was used. The protein samples were loaded onto the column with flow 5 $^{mL}/_{min}$. After completed loading, was the column washed with 15 mL HisTrap binding buffer (20 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, pH 7.4) at flow 5.0 $^{mL}/_{min}$. The bound protein was eluted by HisTrap elution buffer ((20 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole, pH 7.4) at flow 5.0 $^{mL}/_{min}$. The column equilibrated with 15 mL HisTrap binding buffer flow 5.0 $^{mL}/_{min}$ between each protein sample.

HisTrap column strping and recharging

The HisTrap column was stripped of Ni^{2+} with 25 mL HisTrap Stripping buffer (20 mM Tris-HCl, 500 mM NaCl, 50 mM EDTA, pH 7.4). Washed with 25 mL HisTrap binding buffer, and reloaded with 2.5 mL 0.1 M $NiSO_4$ (aq). Finally washed with 15 mL mqH_2O and 15 mL HisTrap binding buffer, respectively[52].

3.3.2 Size exclusion chromatography

The protein containing eluate from AC was concentrated to a final volume of 250-500 µL. The concentrated fraction was loaded onto a 1.0 mL loading loop connected to the 10°C Äkta Purifier. The sample was loaded directly onto a 25.0 mL Superdex™ 75 SEC column. Loaded fractions was mobilized by SEC buffer (20mM Tris-HCl, 100 mM NaCl, pH 7.4) at flow 1.0 $^{mL}/_{min}$.

3.4 Concentration Measurement

NanoPhotometer™ Pearl was used to measure both DNA and protein concentration. Applied 1.5 µL DNA or protein sample to the submicroliter Sample Compression Technology™ with 50 mm cap for measurement.

3.4.1 DNA

The NanoPhotometer™ Pearl measured the DNA concentration based on the 280/260 nm absorption ratio. The instrument was "blanked" by mqH_2O or Elution buffer (0.1 M Tris-HCl, pH 8.0) according to the DNA sample solution.

3.4.2 Protein

Measurements with NanoPhotometer™ Pearl of the different scFv samples required the manual setting of the extinction coefficient (ε) in relation to the molecular weight (Mw) of the protein sample to be measured ($^{\varepsilon}/_{Mw}$). The settings used are given in Table 3.4-1.

Table 3.4- 1. Protein parameters. Parameters used for concentration measurements and pH calibrations for buffer solutions.

Protein/scFv	ε *	Mw (Da) *	$^{Mw}/\varepsilon$	pI *
14F7 C1*	53080	28292.56	0.533	6.73
14F7 C2*	54110	28301.58	0.523	7.63
14F7 C3*	54570	29877.24	0.548	8.23
14F7 C4*	54110	28211.50	0.521	8.79
4G9	54110	28029.17	0.518	6.22
14F7 C1#	54570	29667.01	0.544	6.70
14F7 C2#	55600	29676.02	0.534	7.12
14F7 C3#	54579	29576.93	0.542	8.24
14F7 C4#	55600	29585.94	0.532	8.58

* - Calculated with ExPaSy ProtParam tool for sequences in Appendix G.

3.5 Electrophoresis

3.5.1 Agarose gel

1.0 % agarose gels were made manually by adding 5.0 g agarose (Lonza) to 45 mL 1xTAE buffer (40 mM Tris, 20 mM Acetic acid, 1.0 mM EDTA, pH 8.3), and heated to boiling. After cooling to the becoming viscous (≈ 50°C), was 5.0 µL Ethidium bromide added and poured into a cassette with an eight-well comb for solidification in room temperature. After solidifying, the comb was removed the gel placed in electrophoresis instrument filled with 1xTAE buffer to cover the entire gel. Loaded 20.0 µL of the samples and 5.0 µL of 1 kb GeneRuler DNA ladder (ThermoScientific). Applied 90 V for 50 minutes.

3.5.2 SDS-PAGE

Used 10, 15, and 17 well NuPAGE Bis-Tris 4-12% gel (Life technology) with MES SDS Running Buffer, 20x (Invitrogen). Sample purification with electrophoresis settings 200 V, 270 mA, 100 W, for 25 minutes. SeeBlue® Plus2 standard (Invitrogen), was used as the molecular weight marker. After electrophoresis, washed the NuPAGE Bis-Tris 4-12% gels in hot water for 2x 5min. Stained the protein in gel with Coomassie quick stain solution (Appendix B), 5x 15 seconds in a microwave oven. The gel is ready stained after 2 hours. Distrained with heated water.

3.6 Indirect ELISA

A Polysorp Nunc-Immuno 96-well plate (Sigma) was coated with 100 µL 10 $^{\mu g}/_{mL}$ *N*-glycolyl GM3 and *N*-acetyl GM3 diluted in methanol to each well. Let the methanol evaporate (overnight). Washed each well three times with 100 µL PBS-T (Phosphate-buffered saline (Life technology) with 0.1 % Tween -20 (Sigma)) solution. Incubate for one hour at 22°C (room temperature) with 200 µL PBS-2 % BSA (1.0 M PBS with 2 % bovine serum albumin), covered with parafilm. During incubation, diluted the scFv 14F7 C1* monomer, dimer, and mAb (positive control) with PBS-BT (PBS with 2 % BSA and 0.05 % Tween-20) to the final concentration of 200 nM, and kept samples on ice. After the one-hour incubation, were the plate washed three times with 100 µL PBS-T. Followed by the addition of 100µL sample (scFv/mAb) and PBS-BT as blank to each well, and incubated at 22°C for one hour. Poured off the samples, and washed the wells three times with 100 µL PBS-T, followed by one-hour incubation with 100 µL 0.5 $^{\mu g}/_{mL}$ pl-HRP (peanut lectin-horseradish peroxidase diluted in PBS). Pour off excess pl-HRP and wash each well three times with 100 µL PBS-T. Add 100 µL TMB (3,3',5,5'-tetramethylbenzidine (Calbiochem)), incubated for 30 minutes. Discarded the TMB, and added 1 M HCl to each well for equilibration for 15 minutes.

3.7 Crystallization

Used Mosquito® Crystal (TTP Labtech) to set up crystal screens in Triple Sitting Drop 96-well iQ plate (TTP Labtech) with scFv 14F7 C4* and the five commercial screens, Structure screen 1 & 2 HT-96 (Molecular Dimensions); JCSG-*plus*™ HT-96 (Molecular Dimensions); Morpheus® HT-96 (Molecular Dimensions); PACT *premier*™ HT-96/FX-96 (Molecular Dimensions); and Wizard Cryo 1 & 2 (Rigaku). The concentration of C4* was measured to 1.577 $^{mg}/_{mL}$ with NanoDrop One (Thermo Scientific). Each sitting drop had the final volume of 500 nL. Drop one on each plate had a 1:1 protein-buffer volumetric ratio, drop two a 1:3 protein-buffer volumetric ratio, and drop three was 500 nL buffer (control). The plates were stored at 22°C.

3.8 ThermoFluor Assay

We used a JASCO-8500 fluorimeter (Jasco) to measure the denaturation temperature of scFv 14F7 C1* alone and compared it to the C1* in solution with the *N*-glycolyl GM3 trisaccharide (tricer, Figure 3.8-1). 2 µM C1* was measured with and without 20 µL tricer in a 200 µL cuvette. The sample was heated from 25°C to 90°C $1^{°C}/_{min}$. Tryptophan fluorescence measurements of the ration between 350 nm and 330 nm (F350/F330) intensity.

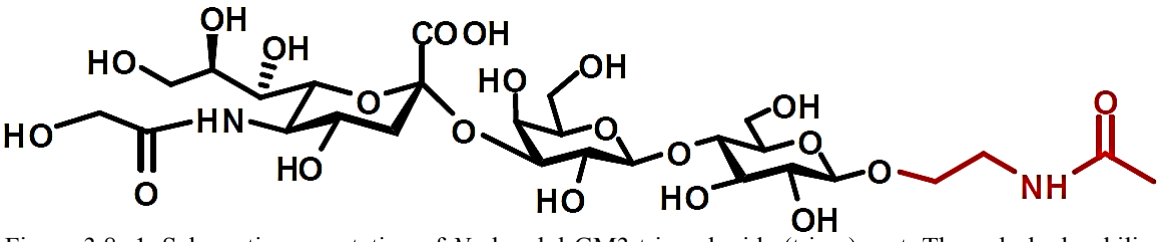


Figure 3.8- 1. Schematic presentation of *N*-glycolyl GM3 trisaccharide (tricer) part. The only hydrophilic inker between the sphingosine and the C18:0 fatty acid of the ceramide remains (red). The illustration is reproduced from Jonhannesen (2014)[44].

4 Results and Discussion

4.1 Status at project start

When starting my project, all the constructs for scFv 14F7 were already made. Hedda Johannesen had during her Master's Thesis made scFv 14F7 constructs with N-terminal His-tag (C1-C4) and without His-tag (C1*-C4*)[44]. The scFv14F7 with C-terminal His-tag (C1#-C4#) did Helene Mykland Hoås construct in her Master's Thesis[58]. Hoås did thermoflourecence of 14F7 C1* thermos stability, and indirect ELISA together with *N*-acetyl GM3 and *N*-glycolyl GM3 for binding conformation of scFv14F7 C1*. Hoås ended her thesis imploring the work to continue in the direction of interaction studies between 4G9 and 14F7 and ultimately co-crystalize the two ideotypes in binding.

The crystal structure of scFv 14F7 C1* was recently solved[42] and the data collection was compared to the crystal structure of Fab 14F7[41]. The comparison of the electron density maps of Fab and scFv 14F7 revealed the key residues of the long CDR H3 loop[42].

4.2 Expression and purification of scFv 14F7 constructs

Our favorite proteins scFv 14F7 C4* and scFv 4G9 tends to dimerize when at high expression levels instead of fold as intended. Hence, keeping a low expression level to sustain the correct folding of the protein is the point of focus. To allow the produced protein time to fold, we selected the bacteria strain *E.* coli XL1-Blue because it has a lower rate of growth. In addition to the low level of expression, the proteins that are expressed was aided in folding by the chaperon FkpA. FkpA assists the scFvs in the correct folding once in the periplasmic space of the bacterial cell[59].

4.2.1 Purification of scFv 14F7 C1*

Purification of scFv 14F7 C1* was at thesis start the construct that gave the highest protein yield with correctly folded proteins. ScFv 14F7 C1* has a molecular weight of 28292.56 Da, so the elution volume with a Superdex 75 (GE Healthcare) SEC column will be at approximately 11 mL with the flow rate of 1 $^{mL}/_{min}$ (Appendix F). The scFv 14F7 C1* monomeric peak begins elution at approximately 11 mL and ends at approximately 14 mL (Figure 4.2-1). When using the protocol of Helene Mykland's Master Thesis[58], 2.0 L of bacteria culture (OD_{600} = 2-7 AU) the protein yield after Protein L AC and Superdex 75 SEC amounted to 20-25 µL of 2-3 $^{mg}/_{mL}$ protein (Table 4.2-1). Not a very high amount but enough for crystallization screens, ELISA, tryptophan fluorescence, or other types of physical studies.

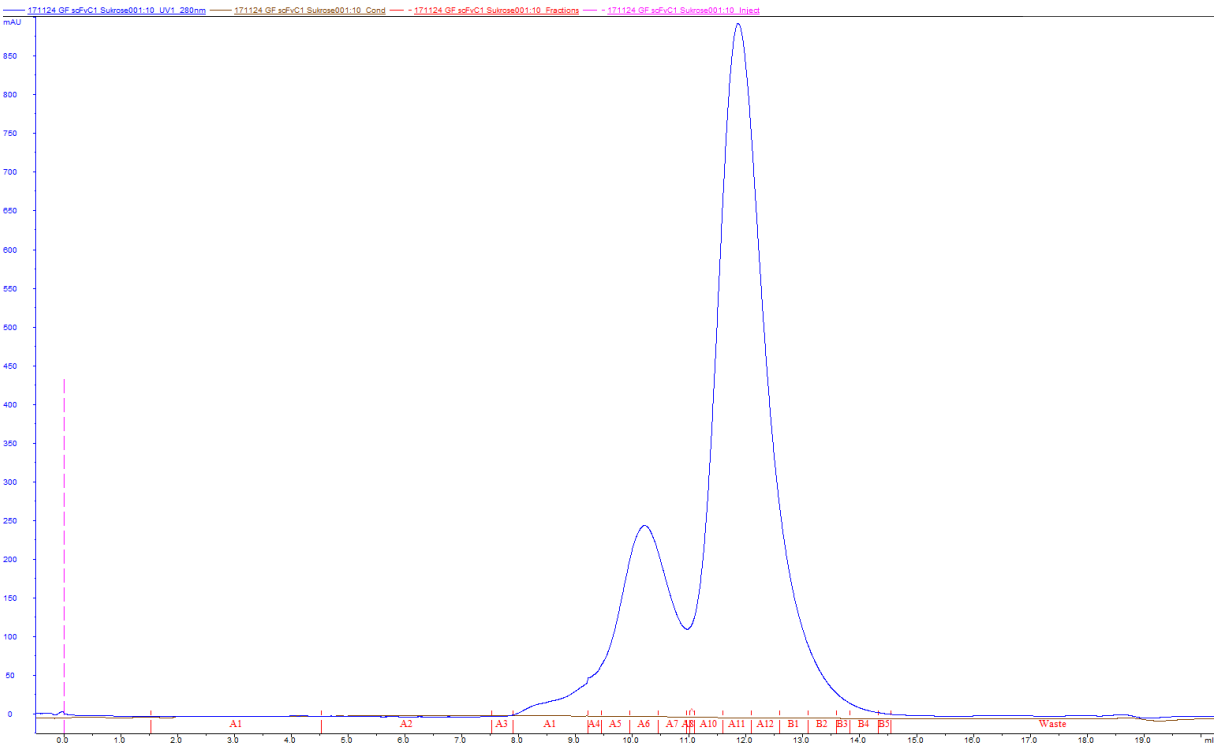


Figure 4.2- 1. Superdex 75 SEC of scFv 14F7 C1*. Firs peak (8-11 mL, A3-A9) is a mixture of dimerized and aggregated scFv 14F7 C1*. The second peak (11-14 mL, A10-B3), is of monomeric scFv 14F7 C1*.

Table 4.2- 1. Comparison of scFv 14F7 C1* production. OD_{600} is measured of bacteria culture at the end of induction. Volume and final protein concentration after SEC. [a] – sucrose fraction, [b] – $MgCl_2$ + lysozyme fraction.

Batch number	OD_{600}	Volume	Concentration
1	2.436 AU	25 µL	2.7 $^{mg}/_{mL}$
2 [a]	6.732 AU	20 µL	2.4 $^{mg}/_{mL}$
2 [b]	6.732 AU	34 µL	6.4 $^{mg}/_{mL}$

Protocol comparisons

After establishing that we are able to reproduce the new protocols workflow, did we try comparing this new protein extraction protocol of Helene's thesis[58] (one step sucrose followed by one step $MgCl_2$ and lysozyme treatment), with the original one of Hedda's thesis[44] (one step with sucrose and lysozyme) (Table 4.2-2). OD_{600} after induction was only measured for one of the samples. The volumes and concentrations measured in Table 4.2-2 were after Protein L AC. When comparing the three dimer/aggregation peaks of Figure 4.2-2, is it the original protocol that indicates the most significant occurrence of the unwanted fraction. Since the SEC purification of the three fractions indicated a high abundance of dimer contamination (Figure 4.2-2) was the fractions A11-B3 of Suc + Lys (Figure 4.2-2A); A10-B3 of sucrose (Figure 4.2-2B); and A10-B3 of $MgCl_2$ fraction (Figure 4.2-2C) combined and concentrated to 37 µL of 6.4 $^{mg}/_{mL}$. This points to both a substantial loss in protein during the last purification step with SEC and that a high amount of the measured concentration after AC was from the dimerized fraction. Unfortunately, is the SEC purification step a necessary one, seeing that the dimerized fraction of scFv will obstruct crystallization and other physiochemical experiments. This combined fraction of scFv 14F7 C1* where used in ELISA an experiment (4.8 ELISA with scFv 14F7 C1* and *N*-glycolyl GM3). Also, the remaining C1* SEC (Table 4.1-1, 2 [b]) product was purified with a second round of SEC (Figure 4.7-1), also with Superdex 75 (GE Healthcare) to a final concentration of 2.4 $^{mg}/_{mL}$ (20 µL), and used for tryptophan fluorescence measurement for thermostability (4.7 Tryptophan fluorescence with scFv 14F7 C1* and *N*-glycolyl GM3)

Table 4.2- 2. Comparison of protocols. Head-to-head comparison of the original and the new purification protocols. Suc, 25 % sucrose solution; Lys, 0.15 $^{mg}/_{mL}$ lysozyme; $MgCl_2$, 5 mM $MgCl_2$ solution.

Protocol	Fractions	OD_{600}	Volume	Concentration
Hedda's original	Suc + Lys	6.732 AU	115 µL	12.8 $^{mg}/_{mL}$
Helene's new	Suc	6.732 AU	130 µL	9.3 $^{mg}/_{mL}$
	$MgCl_2$ + Lys	6.732 AU	130 µL	12.9 $^{mg}/_{mL}$

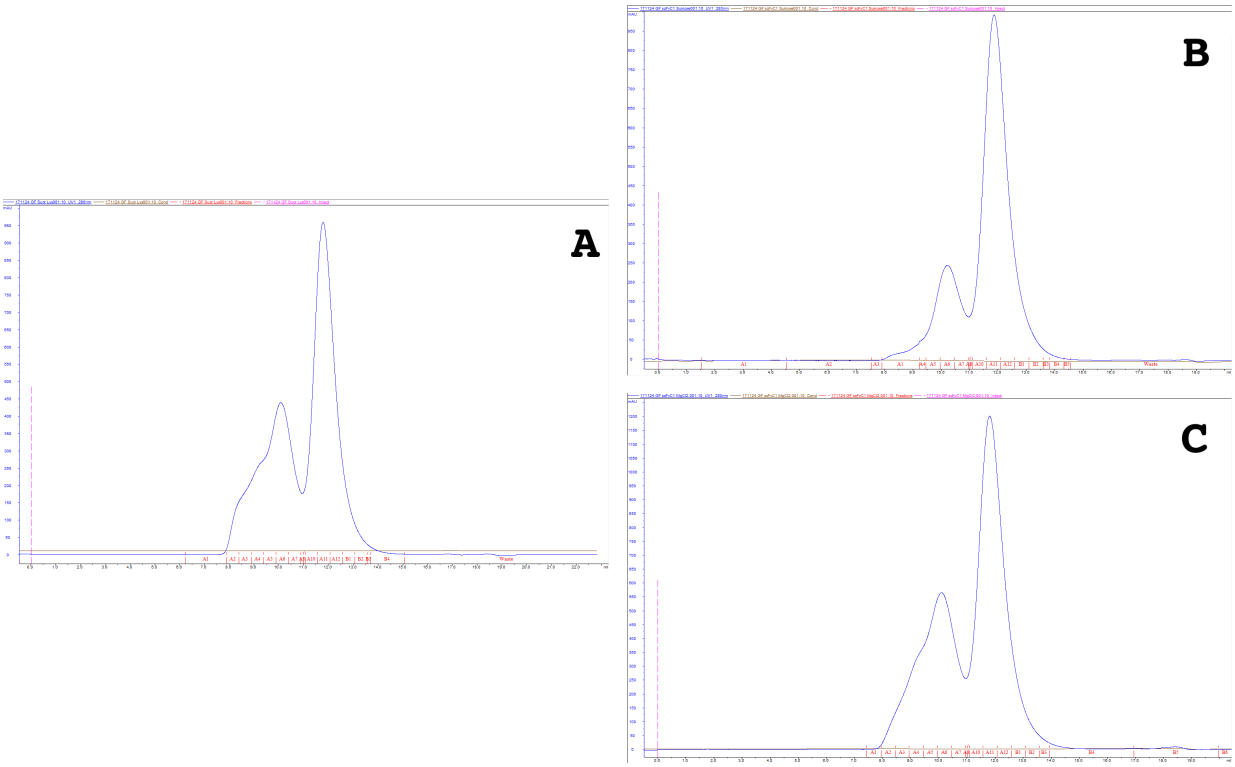


Figure 4.2- 2. SEC of scFv 14F7 C1*, comparing Hedda's original protocol (A) with Helene's new protocol (B and C). (A) is Suc + Lys, (B) Suc, and (C) $MgCl_2$ + Lys. The first peak, 8-11 mL, is the unwanted dimer/aggregations fraction. The second peak, 11-14 mL, is of the desired monomer fraction. Suc, 25 % sucrose solution; Lys, 0.15 $^{mg}/_{mL}$ lysozyme; $MgCl_2$, 5 mM $MgCl_2$ solution

Expression of all scFv 14F7 constructs

Since it had been established that the new protocol created during Helene's Master thesis provides a higher yield of operative scFv 14F7 C1* compared to the original protocol. Would it be interesting to investigate if this new protocol also the same effect on the other scFv 14F7 constructs and in particular the C4* construct [V_H-L_C-V_L]. ScFv 14F7 C4* is the original construct form Cuba that as a mAb gave the anti-idiotypic mAb 4G9 in an immunized murine response[45]. So, exploring the interaction between 14F7 C4* and 4G9 is desirable. Besides, it would be interesting to know if the scFv 14F7 C2* [V_H-L_R-V_L], which also have the original LC[43], express well with the new protocol. Then we have a valid alternative to C4* if C4* prove not to express well or able to crystallize.

We did not have all the equipment to purify all the four constructs with the new protocol in parallel. So, to get a general feel for the protein expression levels and folding capability, we chose to go for parallel induction followed by protein extraction with bacteria lysis by sonication. The four scFv 14F7 constructs were expressed in parallel in 500 mL 2xYT-A media each. All the four 500 mL 2xYT-A produced a bacteria pellet of circa 8 g. The bacteria pellets were mixed with 4-5 $^{\text{mL Lysis buffer}}/_{\text{g bacteria}}$ (Table S-12) and sonicated six rounds of 10 seconds with 35% intensity of 20 kHz. The resulting purification with SEC (Figure 4.2-3) indicates that neither scFv 14F7 C2* or C4* have optimal expression levels with our induction protocol, but some are produced. Comparing the amount (mg) scFv 14F7 C1* and C4*, C4* bacteria produced approximately one-sixth of the monomeric C4* at the same conditions as C1* (Table 4.2-3). This result, together with the almost same low expression level of C2* meant that the expression and purification protocol for the C4* construct need to be optimized.

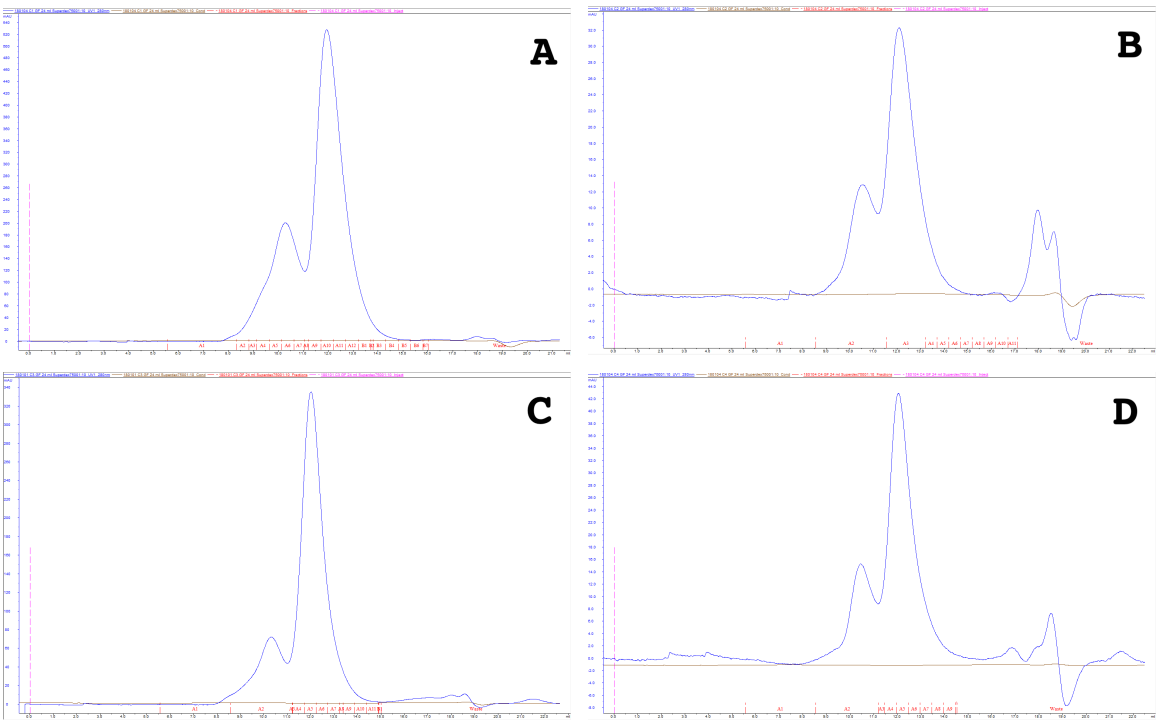


Figure 4.2- 3. SEC of scFv 14F7 C1*-C4* (A-D). The concentrated monomeric peaks are given in Table 4.2-3

Table 4.2- 3. Purification of scFv 14F7 C1*-C4*. OD_{600} is measured of bacteria culture at the end of induction. The volumes and concentrations are from monomeric Superdex 75 SEC purified fractions (Figure 4.2-3).

Construct	OD_{600}	Volume	Concentration
C1*	7.78 AU	850 µL	0.042 $^{mg}/_{mL}$
C2*	7.10 AU	1100 µL	0.011 $^{mg}/_{mL}$
C3*	7.25 AU	500 µL	0.042 $^{mg}/_{mL}$
C4*	6.54 AU	400 µL	0.015 $^{mg}/_{mL}$

4.2.2 Purification of scFv 14F7 C4*

To get a more accurate impression of the actual state of scFv 14F7 C4* expression, the new protocol for C1* expression was applied. From 2 L 2xYT-A with OD_{600} = 0.765 AU at induction start, a collected 32.77 g (OD_{600} = 6.92 AU) bacteria pellet was acquired after 18-hour induction at 30°C, 125 rpm. The two extraction fractions (Suc and $MgCl_2$ + Lys) was purified separately with Protein L (GE Healthcare) AC, followed by separate purification on Superdex 75 SEC column (Figure 4.2-4).

The optical density of the induced fraction of C4* measured the bacteria density similar to the OD_{600} of the C1* expression, and pellet size was consequently of same size (Table 4.2-4). When examining the peaks, the area under the peak of C4* is substantially smaller than that of C1* (Table 4.2-4). Moreover, when comparing the C4* SEC fractions, is the dimer/aggregate fraction more significant then preferable, especially for the MgCl2 fraction (Figure 4.2-4E).

We collecting the fractions from SEC equal the monomeric peak (11-14 mL) and measuring the concentration. It is safe to assume that some of the measured protein is contamination from the dimer fraction (Appendix F). Hence additional SEC purification was performed when purer monomeric samples were required (Figure 4.7-1). The expression levels are not very high but sufficient, so the project then turned towards the scFv 14F7 C4*'s anti-idiotype scFv 4G9.

Table 4.2- 4. Comparison of scFv 14F7 C1* and C4* bacterial growth and protein yield. Suc, 25 % sucrose solution; Lys, 0.15 $^{mg}/_{mL}$ lysozyme; $MgCl_2$, 5 mM $MgCl_2$ solution.

	scFv 14F7 C1*		scFv 14F7 C4*	
OD_{600} (post-induction)	6.732 AU		6.92 AU	
Bacteria pellet	32 g		32.77 g	
	Suc	$MgCl_2$ + Lys	Suc	$MgCl_2$ + Lys
Protein L AC peak area	1515.4 mAU•mL	2398.2 mAU•mL	1609.5 mAU•mL	1468.1 mAU•mL
Monomeric SEC peak area	944.2 mAU•mL	1340.0 mAU•mL	1031.4 mAU•mL	793.1 mAU•mL

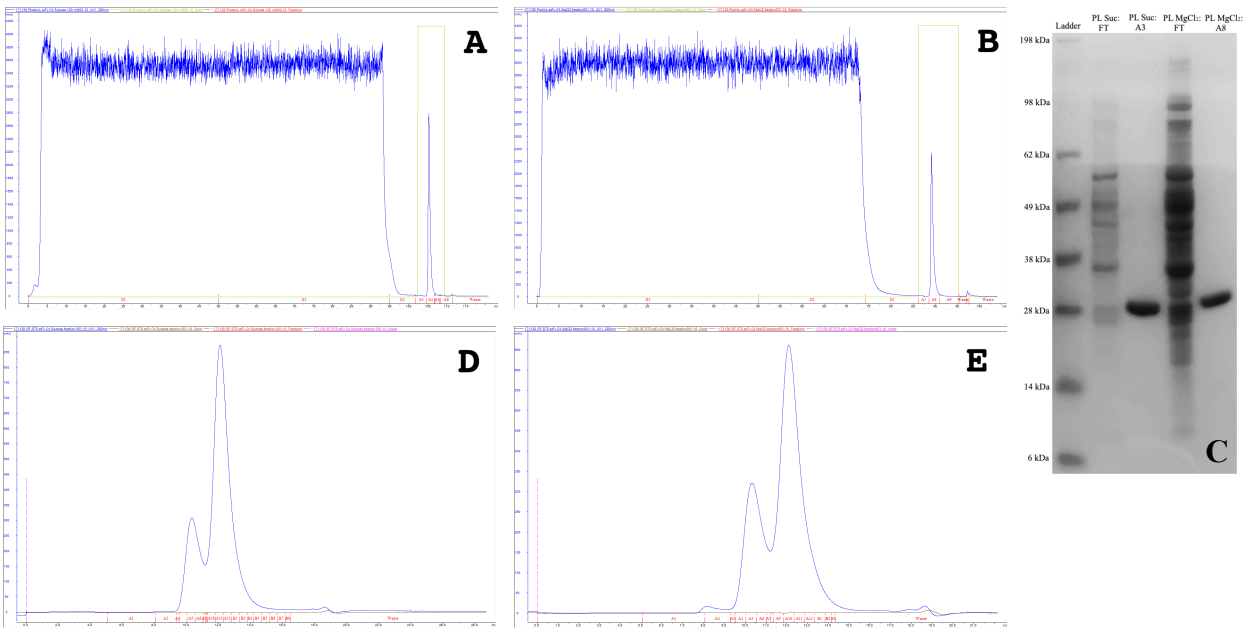


Figure 4.2- 4. Purification of scFv 14F7 C4*. (A) Protein L (GE Healthcare) AC of sucrose fraction, and $MgCl_2$ fraction (B). (C) SDS-PAGE gel of Flow through (FT) and elution fractions (A3 & A8). The eluate appears to a mass of 28 kDa compared to standard molecular weight ladder (SeaBlue®). The following SEC (GE Healthcare) purification of AC fraction A3 for sucrose (D) and A8 for $MgCl_2$ (E).

4.3 Expression and purification of scFv 4G9

ScFv 4G9 is 28029.17 Da in size and which is similar to scFv 14F7 C4* (Table 3.4-1) and a pI of 6.22, so the same buffers and solutions at pH 7.4 used with C4* can be used for 4G9.

4.3.1 Purification of scFv 4G9 with Protein L & HisTrap™

From the very first run-through, the expression procedure of scFv 4G9 seemed promising. The sample was first purified Protein L (GE Healthcare) based on 4G9's κ-LC and gave a peak amounting to 28 kDa in size (Figure 4.3-1). Since scFv 4G9 has a C-terminal His-tag we next purification run tried to purifying the extracted protein with HisTrap™ (Amersham Biosciences) column. The HisTrap™ column did not prove as selective for our protein. The chromatogram had three observable peaks (Figure 4.3-2A), which composed of different proteins (Figure 4.3-2B), all apparent with an affinity towards the Ni^{2+} material of HisTrap™ column. When dissecting the SDS-PAGE gel and sending the most distinct fragments to mass spectroscopy (MS) analysis did we only get hits on *E. coli* proteins when aligning with the *E. coli* proteome, and no hit for alignment with our scFv 4G9. The bands at 28 kDa (Figure 4.3-2B) and 14 kDa (Figure 4.3-2B) was sent for sequencing. A "hit" in the sense of protein alignment of MS measured trypsin treated sequences, means that trypsin treatment of the scFv 4G9 sequence will create fragments of specific mass. Trypsin digest polypeptides at the C-terminal peptide bond of arginine and lysine, if the peptide bond does not bind to a proline.

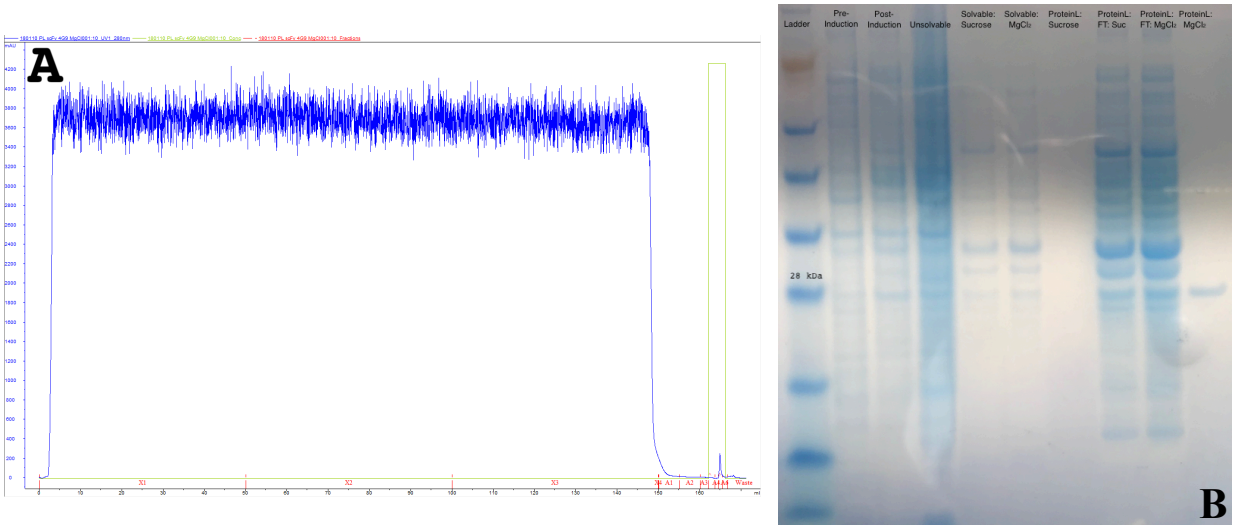


Figure 4.3- 1. AC with Protein L (GE Healthcare) of scFv 4G9. (A) The peak in fraction A5 was assumed to be scFv 4G9 based on the Mw being 28 kDa from the bond in resulting SDS-PAGE(B), "Protein L: $MgCl_2$".

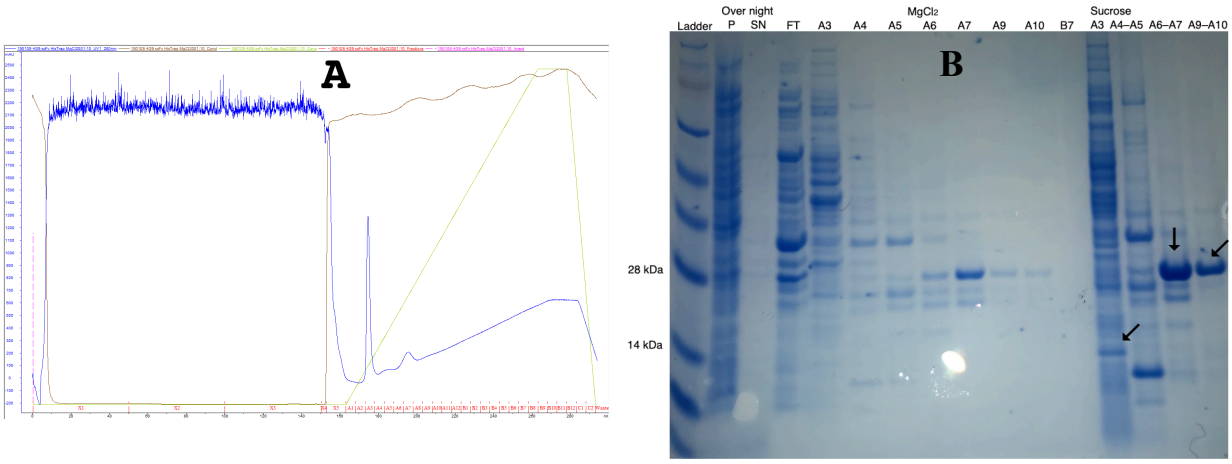


Figure 4.3- 2. AC with HisTrap™ (Amersham Biosciences) of scFv 4G9. (A) Chromatogram of gradient elution of 4G9 $MgCl_2$ fraction by HisTrap elution buffer (3.3.1 – HisTrap™ column). (B) Resulting SDS-PAGE gel, bands sent for sequencing marked by arrows. According to SeaBlue® Plus2 Molecular weight standard marker (Invitrogen); A3, 14 kDa band; A6-A7, 28 kDa band; A9-A10, 28 kDa band.

4.3.2 Screening for optimal for scFv 4G9 Protein production

Since repeated tries in purifying scFv 4G9 ended unsuccessfully, the search for protocol optimization began. We started screening with different growth media because this was the most straightforward step in the protocol to alter. Based on, that the bacteria were producing no apparent 4G9, we thought that maybe the media we used favors bacteria division over protein production, even with the removal of glucose. We did also think that maybe the bacteria itself was the problem, or maybe that the 4G9 plasmid (pFKPEN-4G9scFv) had become mutated, but these were considered to be more time consuming and costly measures to determine. Hence, we started looking at the growth media influence.

4.3.3 Media screening, incubation times, and influence of inducer

We began simply by comparing our usual 2xYT medium with terrific broth (TB) medium (Appendix B), and inducing the two media with isopropyl-β-D-thiogalactopyranoside (IPTG) at concentrations 0.05 mM and 0.1 mM (Figure 4.3-3). The IPTG was added to 2xYT-A and TB-A induction cultures. Form the measured OD_{600} of pre-induction cultures, indications that TB medium is favoring even quicker bacterial growth than 2xYT were observed (Appendix F).

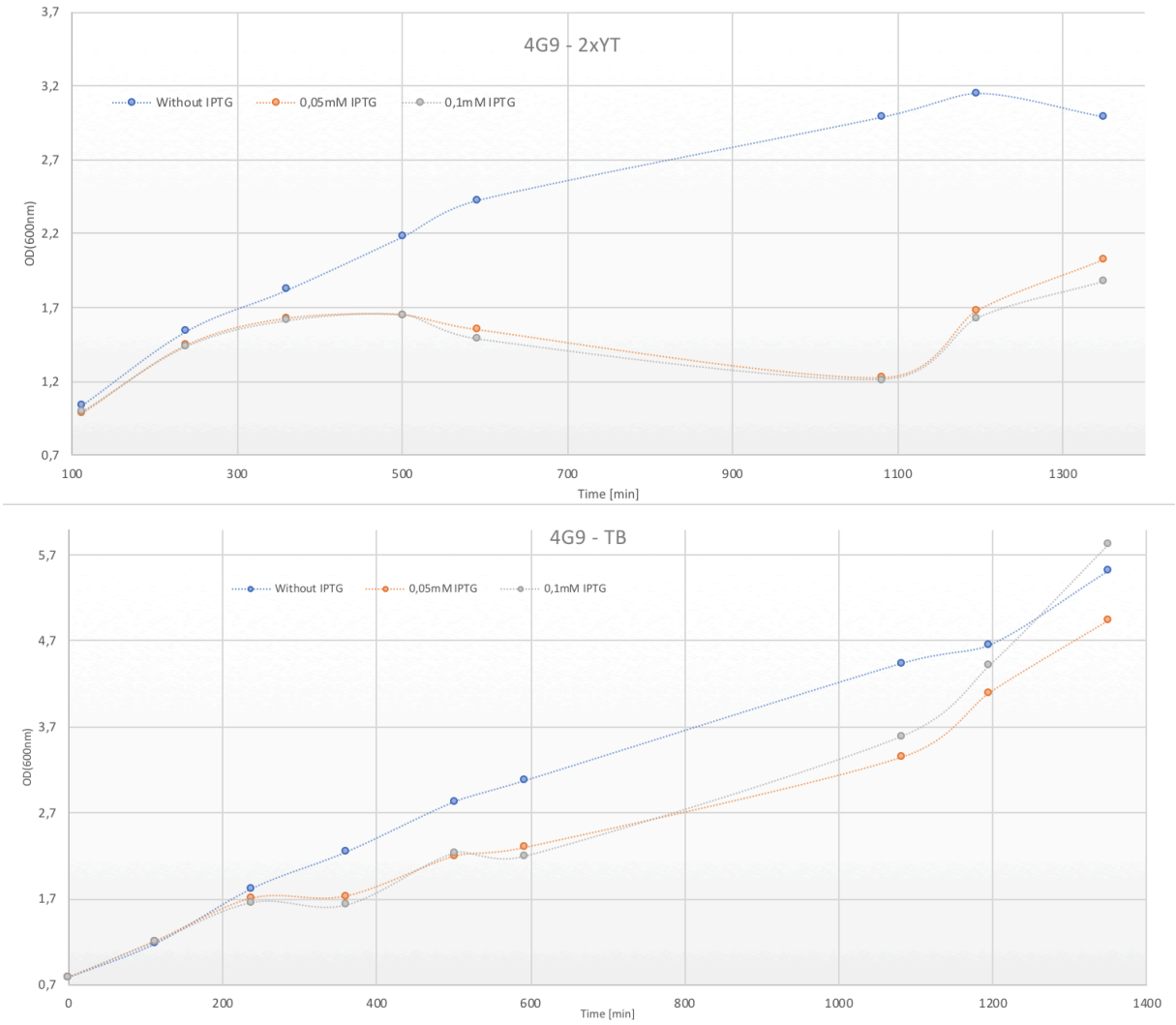


Figure 4.3- 3. Growth curves for scFv 4G9 in 2xYT (A) compared to TB medium (B), with and without the addition of IPTG induction. Raw data for growth curves in Appendix F, Table S-26.

Gel samples were taken for every OD_{600} measurement, but the final SDS-PAGE gel did not show any band for protein presence (gel picture not included). So, a large-scale protein production with TB medium was tested. The TB run did not include IPTG as inducer because of the negative influence reported by Gunnarsen et. al.[60], combined with the increase of inclusion bodies and no increase in soluble scFv found by over own lab[42]. It has also been found by our lab (Hedda Johannesen, unpublished data) that the expression of scFv 14F7 C1* has an optimal expression level after 10 hours (600 min) of induction. Moreover, after 10 hours induction a clear trend in unwanted digestion of the scFv into separate HC and LC (bond appearance at 14 kDa).

The usage of TB medium did not improve protein production. Besides, OD_{600} (post-induction) increased to 16.7 AU and 72.61 g bacteria compared to OD_{600} (post-induction) of 7.11 AU and 32.78 g bacteria. The bacteria mass increase is thought to be a result of continues division by the bacteria disregarding the influence of glucose removal. This could be caused by the bacteria sudden ability of ampicillin resistance and disposal of pFKPEN-4G9scFv. To be sure, did we isolate pFKPEN-4G9scFv and sent it for DNA sequencing. The sequence proved identical with the sequence we assumed we had for scFv 4G9 (Appendix G).

4.3.4 Alternative BACTERIA STRAINS

Since the plasmid is present and the 2xYT media is preferred from the protocol, we tried some alternative strains of bacteria. We selected three *E. coli* strains in addition to the original XL1-Blue (Stratagens) to transfect, BL21 (NEBiolabs), T7 (NEBiolabs), and Rosetta (Merck). The transformations were plated on 25 mL LB-AG 1.5 % agar plates. The transformations into BL21 did not produce any colonies, so initially was the three other strains check for improvement in growth and protein expression. The transformation into fresh *E. coli* proved to reduce the amount of bacteria down to 40-45 g after 18 hours induction. Also, the periplasmic protein expression returned with the new transformations. However, none of the different strains gave any difference in protein quality Judging from the SDS-PAGE gels, the same protein pattern as in Figure 4.3-2 were still being produced and extracted, and still no scFv 4G9. No apparent difference in protein between the different strains of bacteria. MS sequencing of the 14 kDa, 28 kDa, and 60 kDa SDS-PAGE protein bands proved to be the same periplasmic *E. coli* proteins as previously detected.

4.4 Autoinduction media

Bacteria stocks are kept at -80°C for general storage. However, over time, with multiple exposures to room temperature at culture picking, have the stocks indicated reduced protein production capability. The amount of protein produced goes down (from 2-3 $^{mg}/_{mL}$ down to 0.2-0 $^{mg}/_{mL}$), and the mass amount of bacteria after induction goes up (from 30-35 g to 70-80 g). So, it seems to be a close correlation between the two measurables and the number of times the bacteria are exposed to the room temperature upon culture picking. It has also been an undesirable that incubation over 10 hours have indicated an increase in the digested scFv product. So, timing the induction start to begin late in the evening has become a necessary inconvenience. Hence, a search for an alternative protocol that bypasses the induction inconvenience and increases the yield of monomeric scFv has been continuous.

Luckily, such a protocol was provided to us by the help of Anders Tveita (Rikshospitalet, Institute of Clinical Research, UiO, Norway). A protocol and media recipe that are mixed to a final bacteria concentration of OD_{600} of 0.025 AU, and left for incubation for 24 hours. The bacteria will in this media grow and divide for the duration of glucose and glycerol present, which takes approximately 16 hours. The next 8 hours are dedicated to protein production.

Switching to the autoinduction protocol reduced the yielded to once again 28-43 g bacteria (OD_{600} = 9.5-11 AU). Also, the extracted protein amount was generally higher than for the reduced media protocol of Helene's Thesis (Table 4.4-1), especially for scFv 14F7 C4*.

The switch to autoinduction media increase in scFv 14F7 C4* monomeric fraction but also the other unwanted fractions (Figure 4.4-1). Even though the amount of dimer and digested C4* are high, are they relatively easy to remove with SEC. Moreover, the final monomeric C4* concentration is still higher than with autoinduction than other protocols tried (Table 4.4-1).

Table 4.4- 1. Comparson of autoinduction protocol and reduced media protocol influence on protein concentration between scFv 14F7 C1*, C4*, and 4G9. Reduced media concentrations for scFv 14F7 C1* and C4* from Table 4.1-1 and Table 4.2-3, respectively.

Media type	scFv 14F7 C1*	scFv 14F7 C4*	scFv 4G9
Reduced Media	37 µL 6.4 $^{mg}/_{mL}$	200 µL 0.5 $^{mg}/_{mL}$	0 $^{mg}/_{mL}$
Autoinduction media	400 µL 2.5 $^{mg}/_{mL}$	350 µL 1.6 $^{mg}/_{mL}$	0 $^{mg}/_{mL}$

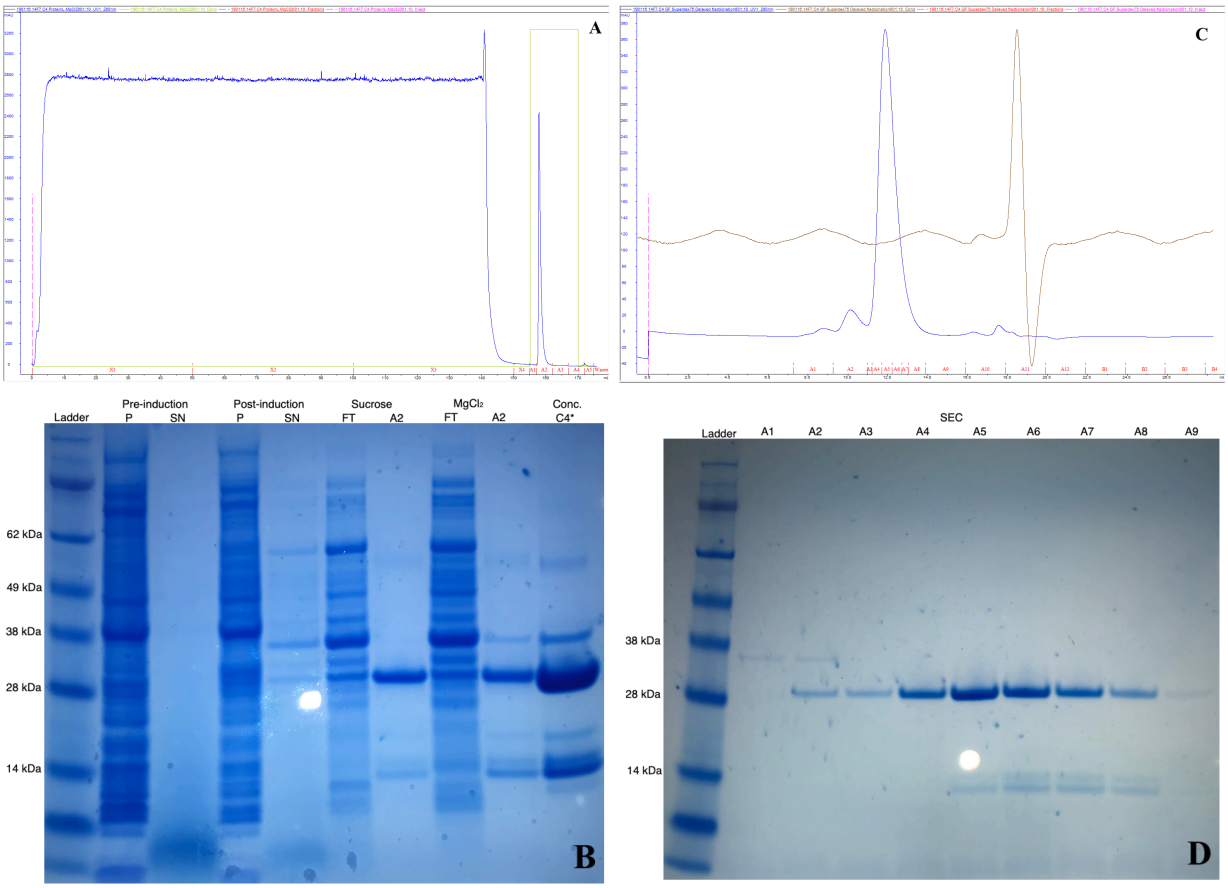


Figure 4.4- 1. Superdex 75 (GE Healthcare) SEC of auto-induced scFv 14F7 C4* (C) after Protein L (GE Healthcare) purification (A), with respective SDS-PAGE of fractions (D and B). (B) Conc. C4* is scFv 14F7 C4* Sucrose and $MgCl_2$ fractions A2 collectively concentrated, that are purified with SEC (C and D).

4.5 Creating Fab 4G9

Compared to other tested protocols did the autoinduction protocol improved the workflow, protein yield, and overall quality of scFv 14F7 C1* and C4*. However, the autoinduction protocol did not improve scFv 4G9 outcome. Since 4G9 is central in this project did we focus on try and fix the protein production. We selected to make a potentially more stable Fab construct of 4G9. Since we have a Fab vector pFABEFN-HaLb380 (Appendix E) for *E. coli* production system, we started here.

Our collaborators at Rikshospitalet, Lene Støkken Høydahl, have generated a huminized Fab for protein production in HeLa system that has proven the ability to produce Fab of troublesome scFv.

4.5.1 Restriction enzyme digestion

The pFABEFN-HaLb380 VH are flanked by RE *Nco*I (Thermo Scientific) and *Hind*III (Thermo Scientific), and VL is flanked by *Mlu*I (Thermo Scientific) and *Not*I (Thermo Scientific). The same four RE flanking the VH and VL of all our scFv constructs. The RE proved unsatisfactory in digestions activity, resulting in weak agarose gel bands for complete digest (Figure 4.5-1). After dissecting the agarose gel fragments and purified the fractions with Gel Extraction Kit (Qiagen), the measured concentrations were in the range of 5-15 $^{ng}/_{\mu L}$. When setting up DNA ligation reactions with these samples, none of the ligations resulted in colony growth after transformation. The transformations were done with $CaCl_2$ competent XL1-Blue *E. coli* cells, with pUC18 as positive control.

The troubleshooting started by looking at the restriction enzymes and its protocol. Alter the DNA concentration between 100 ng to 2 µg (5-100 $^{ng}/_{\mu L}$), use fresh restriction enzymes, and increasing incubation time for ER digestion. None of these had any positive effect on the RE digestion outcome. Calf-intestinal alkaline phosphorylase (CIP) was at an occasion added to the ER mixtures for the vector digestion, to inhibit potential relegation of vector during ligation. It did not give any colony growth as a result. The ligation mixtures were left overnight at 16°C to see if increased incubation time for the T4 DNA ligase (NEBiolabs) would alter the outcome, no effect was observed.

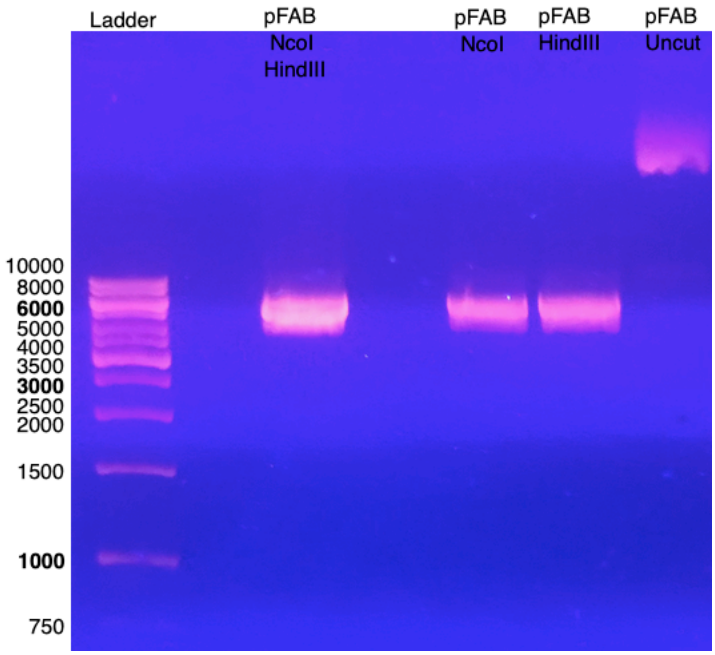


Figure 4.5- 1. RE digestion of pFABEFN-HaLb380 (pFAB). 1.0 % agarose gel 90 V for 50 min. Used 1 kb ladder (Thermo scientific).

HeLa Fab expression system

The inability to get positive results with the bacterial system led us to move forward in the direction of a eukaryotic HeLa expression system instead. A collaborating group at Rikshospitallet have a protocol for recombinant Fab production. We started by humanizing the V_H and V_L of our 4G9 sequence, flanking V_H and V_L with RE sites according to the insert position of the two humanized vectors (pLNOH2 and pLNOκ, respectively[61]). The HeLa expression system based one vector for each Fab (HC and LC). The same two RE sites flanked both V_H and V_L, *Bsm*I (NEBiolabs)/*Mva1269*I (Thermo Scientific) and *Bsi*WI-HF (NEBiolabs)/*Pfl23*II (Thermo Scientific) (Appendix C). So, once again, the same bottleneck had to be overcome, restriction enzyme digestion.

The two humanized 4G9 vectors (pEX-A128-4G9_VH and pEX-A128-4G9_VL, Appendix E) were RE digested in parallel with pLNOκ and pLNOH2 using NEB protocol and *Bsm*I/*Bsi*WI-HF (Figure 4.5-2). Used the Gel Extraction Kit (Qiagen) to extract the 380 bp fragment of pEX-A128-4G9_VH and 347 bp fragment of pEX-A128-4G9_VL and ligated them with 10262 bp fragment of pLNOH2 and 8691 bp of pLNOκ, respectively. Only a few colonies appeared on the pLNOH2:4G9_VH and on pLNOκ:4G9_VL plates, which one colony from each plate was picked and sent for sequencing. The sequencing of these plasmids proved it to be religation of

the two pLNO vectors. The indication of compromised digestion was dedicated to reduced activity of RE *Bsm*I (Figure 4.5-2), even though the presence of a smaller fragment suggested that sufficient cutting was performed (not present in Figure 4.5-2). When trying to redo the RE digestion with the *Bsm*I/*Bsi*WI-HF, had it appeared that the RE activity was lost, with the smaller fragment band (300-400 bp) visually gone. With the ordering of the new restriction enzymes *Pfl23*II (Thermo Scientific) and *Mva1269*I (Thermo Scientific), the smaller band reappeared (Figure 4.5-3). The following T4 DNA ligation (NEBiolabs) reactions and transfection of electrocompetent XL1-Blue *E. coli* (1 µL ligation mix, 29 µL cells, 1.3 V for 4.7 seconds) gave a relatively high number of colonies (Table 4.5-1). Unluckily, the number of vector:insert colonies was at a similar count as to the number of religated vector colonies. So, selecting one of the over 100 colonies present would most likely prove to be a colony with the religated vector. We instead chose to treat the pLNOκ and pLNOH2 vectors with CIP at the RE digestion step. The outcome of the transformation was one single culture on pLNOκ:4G9_VL plate (>200 colonies on positive control transformation with pUC18 plasmid). The plasmid has been sent for sequencing.

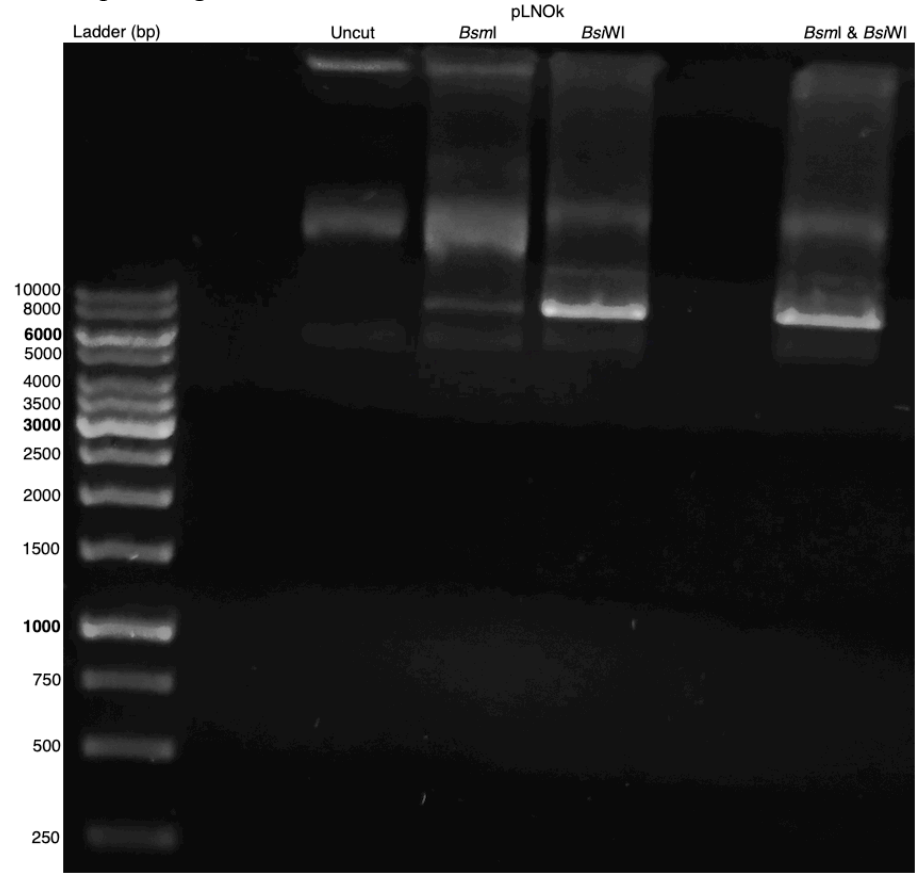


Figure 4.5- 2. 1.0 % agarose electrophoresis gel purification for 90 minutes at 90 V of pLNOκ digested with the RE enzymes *Bsm*I (NEBiolabs) and *Bsi*WI-HF (NEBiolabs) for 60 minutes + 60 minutes.

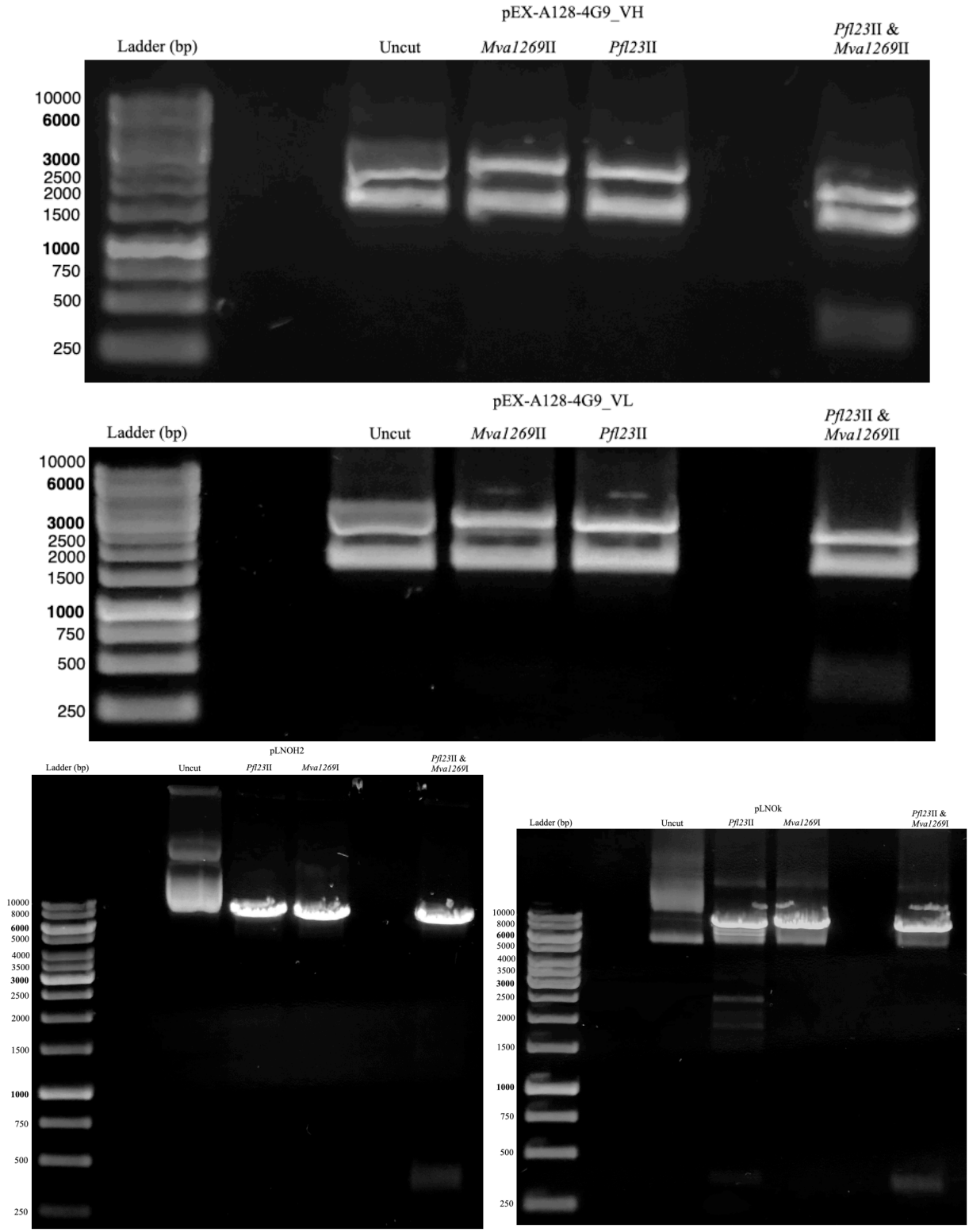


Figure 4.5- 3. 1.0 % agarose gel electrophoresis purification of pEX-A128-4G9_VH (top), pEX-A128-4G9_VL (middle) 50 minutes at 90 V, and pLNOH2 (Bottom left), and pLNOκ (bottom right) 90 minutes at 90 V. 4G9_VH, 4G9_VL, and pLNOH2 was digested for 30 minutes, whiles pLNOk digested ON with the RE enzymes *Pfl23*II (Thermo Scientific) and *Mva1269*I (Thermo Scientific).

Table 4.5- 1. Number of colonies after overnight T4 DNA ligation (NEBiolabs) and transformation into electrocompetent XL1-Blue *E. coli*.

Vector:Insert	Number of colonies
pLNOH2:4G9_VH	187
pLNOH2 alone	160
pLNOκ:4G9_VL	115
pLNOκ alone	107

4.6 Crystallization of scFv 14F7 C4*

Out of all the five screens (Appendix A) only one condition, B9-1 (1:1 v/v of 250 nL 1.6 $^{mg}/_{mL}$ C4* and 250 nL 40% (v/v) 1,2-Propanediol, 100 mM Sodium acetate/ Acetic acid pH 4.5 , 50 mM Calcium acetate) (Figure 4.6-1), in the Wizard Cryo 1 & 2 (Rigaku) gave promising phase separation. The phase separations are formed as spherulite (Figure 4.6-1).

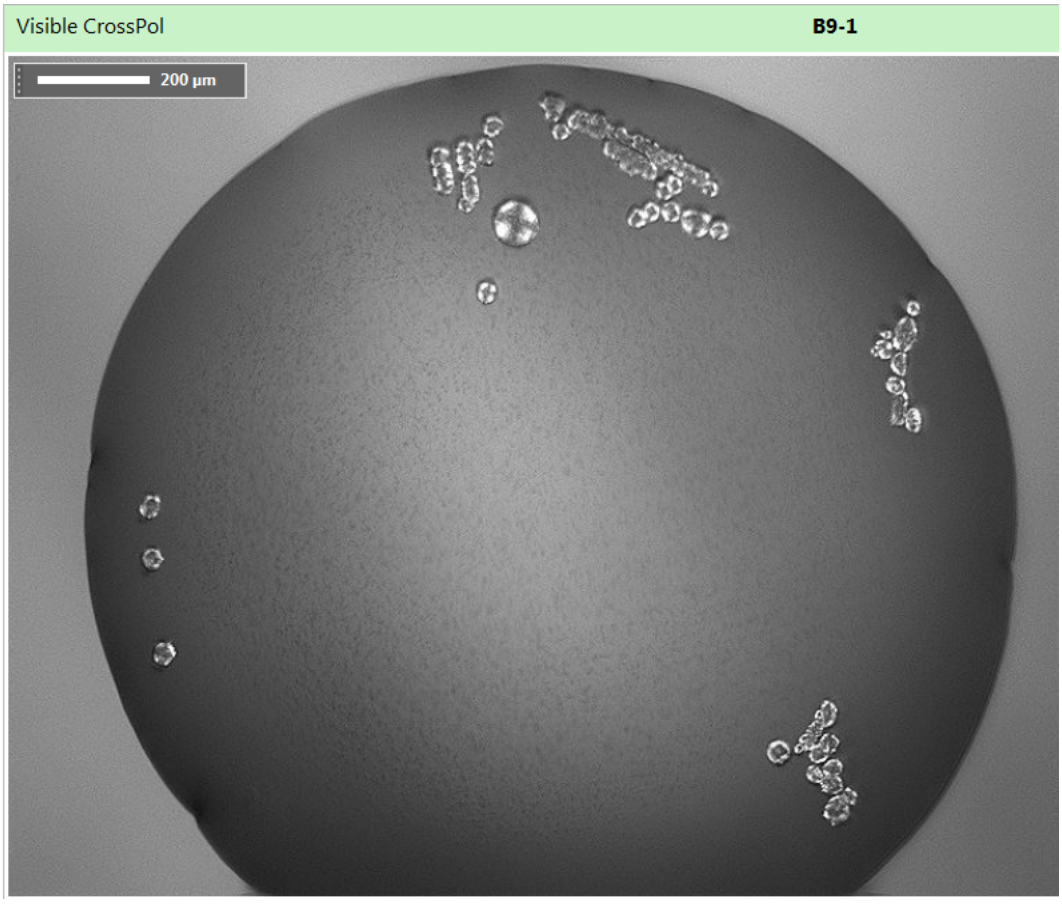


Figure 4.6- 1. Image of scFv14F7 C4* spherulites under Cross-polarized light. Image taken with RockMaker® (Formulatrix).

4.7 Tryptophan fluorescence with scFv 14F7 C1* and *N*-Glycolyl GM3 “tircer”

It has previously been published that the scFv 14F7 C1* iw stable to the temperature of 72°C, and that the protein is stable both with and without the ligand trisaccharide part of *N*-glycolyl GM3 (tricer) in binding[58].

Out reenactment of the thermostability with and without the ligand tricer bound was done with JASCO-8500 fluorimeter (Jasco), and gave similar finding as previously published. With a melting temperature at 73°C, no noticeable influence by the presence of ligand.

The scFv 14F7 C1* sample used had to be purified twice by Superdex 75 SEC (Figure 4.7-1) before use, because of the high occurrence of aggregate/dimerized contamination (elution volume 8-11 mL). The concentration of the monomeric peak after the first SEC purification was 6.4 $^{mg}/_{mL}$ (34 µL) (Figure 4.7-1 A), and 2.4 $^{mg}/_{mL}$ (20 µL) after the second round of SEC purification.

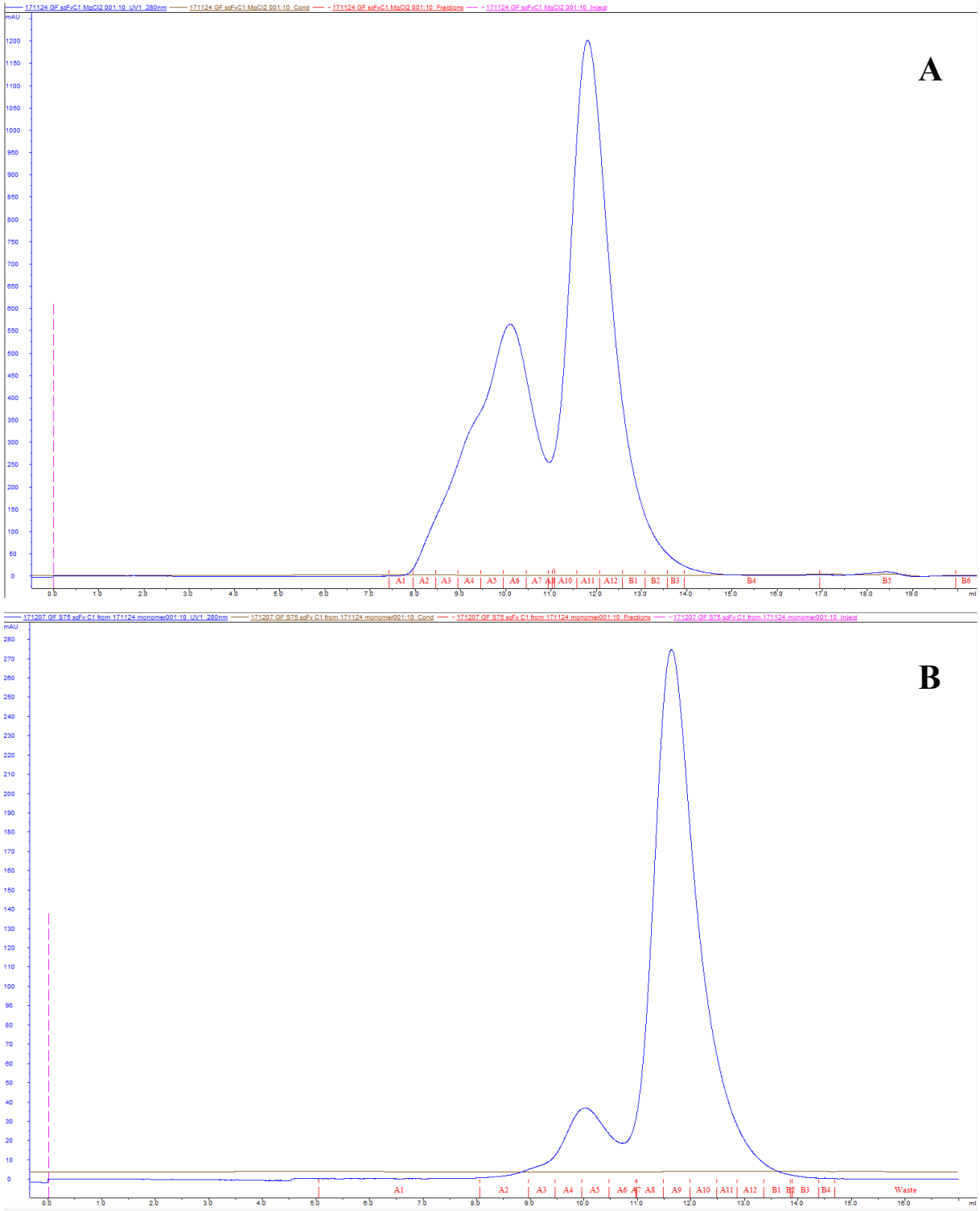


Figure 4.7- 1. Effect of two times SEC with Superdex 75 of scFv 14F7 C1*. (A) The first round of SEC purification, where fractions A10-B3 are SEC purified an additional time (B). Fractions A8-A12 was used for thermostability measurements.

4.8 ELISA with scFv 14F7 C1* and C4# with *N*-Glycolyl GM3

We wanted to see if the 4G9 (scFv and/or Fab) can bind to with scFv 14F7 C4#, and also see if C4# favors binding with N-glycolyl GM3 over N-acetyl GM3. We did a trial run with scFv 14F7 C1* monomer, dimer, and Mab (control), against wells coated with the two gangliosides (Figure 4.8-1). The results were inconclusive when it came to C1* monomer and dimer selectivity towards the gangliosides. This result was thought to be because of miscalculations in concentration C1* monomer and dimer used, or that the coating with the two gangliosides was not done properly. When measuring the concentrations, the lesser dimeric SEC peak (Figure 4.2-2) gave a concentration between 20 and 30 mg/mL whiles the more dominant monomeric peak only 2.4 mg/mL. Since the C1* selectivity is well documented[44,58] did we not redo this experiment, with the motivation that we would continue ELISA experiment once we had 4G9 and scFv 14F7 C4#.

ScFv 14F7 C4# (Mw = 29585.94 Da) has been gathered to some extent, but with a high amount of digested product (Figure 4.8-2).

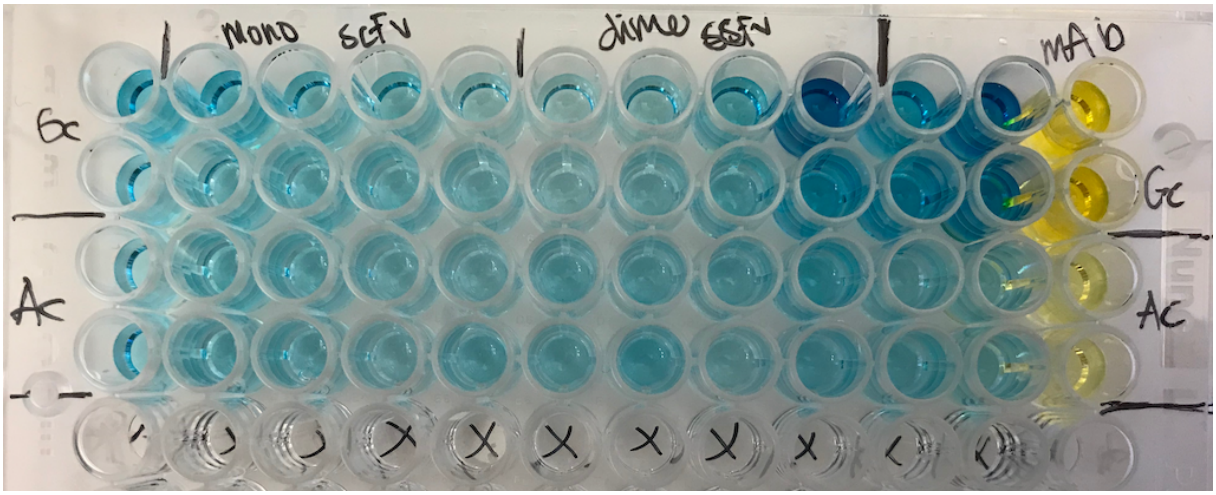


Figure 4.8- 1. Results of ELISA with *N*-glycolyl GM3 (Gc) and *N*-acetyl GM3 (Ac, negative control) cowering the wells and scFv 14F7 C1* monomer (mono scFv), dimer (dimer scFv), and mAb 14F7 (positive control) as primary antibodies. The anti-protein L antibody pl-HRP was used as the secondary antibody.

Table 4.8- 1. The reaction set-up of ELISA with concentrations of primary antibody (nM). Same as viewed in Figure 4.8-1

	Blank	14F7 scFv C1* (nM) Monomer				14F7 scFv C1* (nM) Dimer				14F7 mAb (nM) Positive control		
Gc	0.0	7.4	22.2	66.7	200.0	7.4	22.2	66.7	200.0	22.2	66.7	200.0
Gc	0.0	7.4	22.2	66.7	200.0	7.4	22.2	66.7	200.0	22.2	66.7	200.0
Ac	0.0	7.4	22.2	66.7	200.0	7.4	22.2	66.7	200.0	22.2	66.7	200.0
Ac	0.0	7.4	22.2	66.7	200.0	7.4	22.2	66.7	200.0	22.2	66.7	200.0

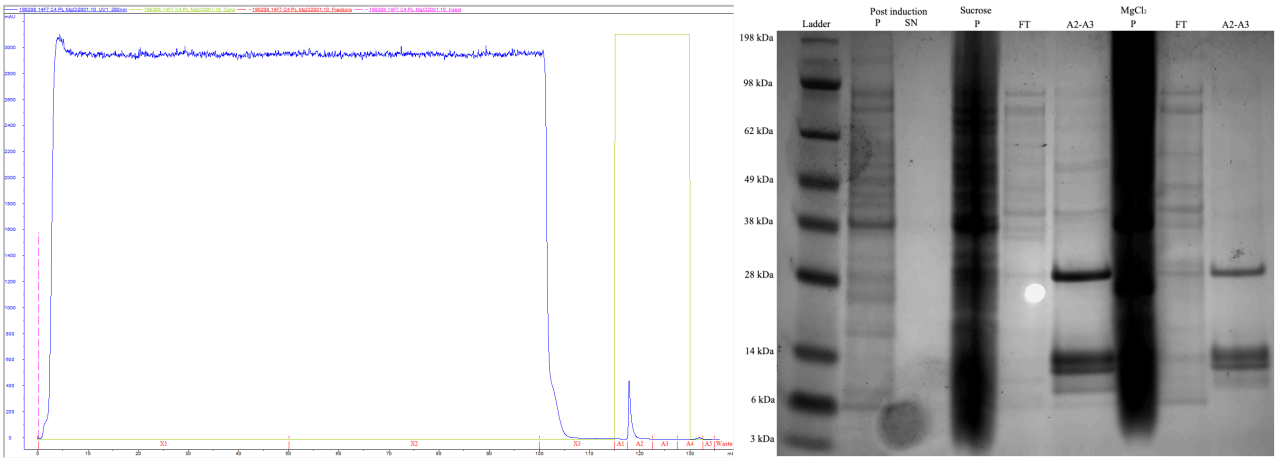


Figure 4.8- 2. Protein L AC of scFv 14F7 C4# (left) and the resulting SDS-PAGE of the different purification fractions during the protein procedure and purification (right). C4# has Mw of 29585.94 Da, which correlates with the single band at around 28 kDa SeaBlue® marker (Ladder).

5 Summary and Conclusion

Considering the protocol for protein extraction with osmotic stress by sucrose, $MgCl_2$, and lysozyme in stepwise treatment of the bacteria is recommended over the one-step addition of sucrose and lysozyme previously used. This conclusion is drawn for proteins produced both with induction by removal of glucose and when using the autoinduction protocol. Periplasmic protein production in *E. coli* (XL1-Blue) with autoinduction is to be the preferred technique, considering both the simplicity of the technique and the protein yield. When purifying the protein fractions with affinity chromatography, Protein L based on the κ-LC is to be recommended over HisTrap for C-terminal His-tag. Protein L purification has proven much more selective when purifying from the periplasm because no apparent periplasmic *E. coli* protein has indicated Protein L affinity. HisTrap purification has, on the other hand, proven high in contamination from periplasmic *E. coli* proteins. Maybe even so high in contamination that it has hindered primary binding of our His-tagged proteins. Additional purification with Superdex 75 SEC has proven necessary because of the relatively high occurrence of aggregated and dimerized scFv protein fractions. One round of SEC purification is generally sufficient enough to get a homogenous monomer fraction for succeeding studies (ELISA, Crystallization), but two rounds of SEC purification is the preferred.

After spending months on trying to clone the scFv 4G9 DNA V_H and V_L sequences into both a bacterial (pFABEFN) and a humane (pLNO) vector, do I have to conclude that recombination of plasmid with restriction enzyme digestion, agarose gel purification, DNA isolation from agarose gel, DNA ligation, and transformation into chemocompetent *E. coli* is not reliable cloning technique. After performing multiple RE digestion and T4 DNA ligation with high activity enzymes from two providers both with and without CIP treatment of vector fragments, did not bring my pursuit for a Fab 4G9 constructs any closer. If I were to retry cloning of Fab 4G9 would I go for a more modern recombination technique for the generation of the Fab vectors. Preferably, Gibbson assembly of the humanized Fab vectors[62].

6 Future prospects

First and for most, the completion of the two humanized recombinant Fab constructs of 4G9, pLNO-4G9_VH and pLNO-4G9_VL. I would implore the future of this project to move in the direction of one of the new generation techniques in the recombination of vector DNA. Gibson assembly[63], *In vivo* assembly (IVA)[64], or Sequence- and ligation-independent cloning (SILC)[65] should be tested. So, maybe starting with Gibson assembly since this technique bypasses RE digestion and DNA ligation to some degree.

Both the reduced media induction protocol (removal of glucose) and the autoinduction protocol have been established as reliable methods for periplasmic expression of scFv 14F7 C4* in an XL1-Blue *E. coli* system. The only optimization of scFv 14F7 C4# are necessary if the humanized Fab 4G9 expression system is accomplished, and C4# is needed for ELISA[56]. Alternatively, can surface plasmon resonance (SPR)[66] be used with Fab 4G9 and scFv 14F7 C4*.

When binding between 4G9 and 14F7 is established co-crystallizing Fab 4G9 and scFv 14F7, maybe start by optimizing the Wizard Cryo 1 & 2 B9 conditions for scFv 14F7 C4* to screen for crystals. Moreover, see the effect of soaking the eventual C4* crystals with Fab 4G9, and resulting X-ray diffraction studies. Begin a new screening procedure for co-crystallization, because most likely is an entirely different buffer condition necessary for Fab 4G9 and scFv 14F7 C4* to co-crystallize.

7. References

1. Bange, J., Zwick, E. & Ullrich, A. Molecular targets for breast cancer therapy and prevention. *Nature Medicine* **7**, 548-552 (2001).

2. Chung, K.Y. et al. Cetuximab Shows Activity in Colorectal Cancer Patients With Tumors That Do Not Express the Epidermal Growth Factor Receptor by Immunohistochemistry. *Journal of Clinical Oncology* **23**, 1803-1810 (2005).

3. Boonstra, M.C. et al. Selecting Targets for Tumor Imaging: An Overview of Cancer-Associated Membrane Proteins. *Biomarkers in cancer* **8**, 119-133 (2016).

4. Silva, A.P.S., Coelho, P.V., Anazetti, M. & Simioni, P.U. Targeted therapies for the treatment of non-small-cell lung cancer: Monoclonal antibodies and biological inhibitors. *Human vaccines & immunotherapeutics* **13**, 843-853 (2016).

5. Krengel, U. & Bousquet, P.A. Molecular Recognition of Gangliosides and Their Potential for Cancer Immunotherapies. *Frontiers in Immunology* **5**, 325 (2014).

6. Talavera, A. et al. Crystal structure of an anti-ganglioside antibody, and modelling of the functional mimicry of its NeuGc-GM3 antigen by an anti-idiotypic antibody. *Molecular Immunology* **46**, 3466-3475 (2009).

7. WHO. Cancer Fact sheet. Vol. 2019 (World Health Organization, www.who.int 2018).

8. Weinberg, R.A. *The biology of Cancer, 2nd edition*, 876 (Garland Science, Taylor & Francis Group, LLC, 2014).

9. NIH. Chemotherapy to Treat Cancer. Vol. 2019 (April 29, 2015, https://www.cancer.gov/about-cancer/treatment/types/chemotherapy, 2015).

10. NIH. Radiation Therapy to Treat Cancer. Vol. 2019 (NIH, https://www.cancer.gov/about-cancer/treatment/types/radiation-therapy, 2019).

11. NIH. External Beam Radiation Therapy for Cancer. Vol. 2019 (https://www.cancer.gov/about-cancer/treatment/types/radiation-therapy/external-beam, 2018).

12. NIH. Brachytherapy to Treat Cancer. Vol. 2019 (National Cancer Institute, https://www.cancer.gov/about-cancer/treatment/types/radiation-therapy/brachytherapy, 2019).

13. NIH. Surgery to Treat Cancer. (https://www.cancer.gov/about-cancer/treatment/types/surgery, 2015).

14. NIH. Targeted Cancer Therapies. Vol. 2019 www.cancer.gov (https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies/targeted-therapies-fact-sheet, 2019).

15. NIH. Hormone Therapy to Treat Cancer. (National Cancer Institute, https://www.cancer.gov/about-cancer/treatment/types/hormone-therapy, 2015).

16. Westburg. Immunetherapy: the fight agains cancer. Vol. 2019 (www.westenburg.eu, https://www.westburg.eu/immunotherapy-for-cancer 2019).

17. Parham, P. *The Immune System, 4th edition*, 532 (Garland Science, Taylor & Francis Group, LLC, 2015).

18. Campbell, N.A. *Biology A Global Approach, 10th edition*, 1351 (Pearson Education Limited, 2015).

19. Rabionet, M., Gorgas, K. & Sandhoff, R. Ceramide synthesis in the epidermis. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1841**, 422-434 (2014).

20. Venkataraman, K. et al. Upstream of Growth and Differentiation Factor 1 (uog1), a Mammalian Homolog of the Yeast Longevity Assurance Gene 1 (LAG1), RegulatesN-Stearoyl-sphinganine (C18-(Dihydro)ceramide) Synthesis in a Fumonisin B1-independent Manner in Mammalian Cells. *Journal of Biological Chemistry* **277**, 35642-35649 (2002).

21. Ternes, P., Franke, S., Zähringer, U., Sperling, P. & Heinz, E. Identification and Characterization of a Sphingolipid Δ4-Desaturase Family. *Journal of Biological Chemistry* **277**, 25512-25518 (2002).

22. Hanada, K. et al. Molecular machinery for non-vesicular trafficking of ceramide. *Nature* **426**, 803-809 (2003).

23. Maccioni, H.J.F. Glycosylation of glycolipids in the Golgi complex. *Journal of Neurochemistry* **103**, 81-90 (2007).

24. Irie, A., Koyama, S., Kozutsumi, Y., Kawasaki, T. & Suzuki, A. The Molecular Basis for the Absence ofN-Glycolylneuraminic Acid in Humans. *Journal of Biological Chemistry* **273**, 15866-15871 (1998).

25. Hayashi, N. et al. Detection of N-glycolyated gangliosides in non-small-cell lung cancer using GMR8 monoclonal antibody. *Cancer Science* **104**, 43-47 (2013).

26. Bremer, E.G., Schlessinger, J. & Hakomori, S. Ganglioside-mediated modulation of cell growth. Specific effects of GM3 on tyrosine phosphorylation of the epidermal growth factor receptor. *Journal of Biological Chemistry* **261**, 2434-2440 (1986).

27. Wang, X.-Q., Sun, P. & Paller, A.S. Ganglioside GM3 Blocks the Activation of Epidermal Growth Factor Receptor Induced by Integrin at Specific Tyrosine Sites. *Journal of Biological Chemistry* **278**, 48770-48778 (2003).

28. Casadesús, A.V. et al. A shift from N-glycolyl- to N-acetyl-sialic acid in the GM3 ganglioside impairs tumor development in mouse lymphocytic leukemia cells. *Glycoconjugate Journal* **30**, 687-699 (2013).

29. Akiba, H. et al. The Role of ICOS in the CXCR5⁺ Follicular B Helper T Cell Maintenance In Vivo. *The Journal of Immunology* **175**, 2340-2348 (2005).

30. Ueno, H. Chapter 7 - T Follicular Helper Cells as a Therapeutic Target for Autoimmune Diseases. in *Translational Immunology* (ed. Tan, S.-L.) 185-204 (Academic Press, Boston, 2016).

31. Bryant, V.L. et al. Cytokine-Mediated Regulation of Human B Cell Differentiation into Ig-Secreting Cells: Predominant Role of IL-21 Produced by CXCR5⁺ T Follicular Helper Cells. *The Journal of Immunology* **179**, 8180-8190 (2007).

32. Nutt, S.L., Hodgkin, P.D., Tarlinton, D.M. & Corcoran, L.M. The generation of antibody-secreting plasma cells. *Nature Reviews Immunology* **15**, 160 (2015).

33. Kometani, K. et al. Repression of the Transcription Factor Bach2 Contributes to Predisposition of IgG1 Memory B Cells toward Plasma Cell Differentiation. *Immunity* **39**, 136-147 (2013).

34. Moser, B. CXCR5, the Defining Marker for Follicular B Helper T (TFH) Cells. *Frontiers in Immunology* **6**(2015).

35. Porter, R.R. The hydrolysis of rabbit y-globulin and antibodies with crystalline papain. *The Biochemical journal* **73**, 119-126 (1959).

36. Maverakis, E. et al. Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review. *Journal of Autoimmunity* **57**, 1-13 (2015).

37. antibody, A. Antibody Isotypes & Subtypes. Vol. 2019 (https://absoluteantibody.com/antibody-resources/antibody-overview/antibody-isotypes-subtypes/, 2015).

38. Carr, A. et al. A Mouse IgG1 Monoclonal Antibody Specific for N-Glycolyl GM3 Ganglioside Recognized Breast and Melanoma Tumors. *Hybridoma* **19**, 241-247 (2000).

39. Carr, A., Mesa, C., del Carmen Arango, M., Vázquez, A.M. & Fernández, L.E. In Vivo and In Vitro Anti-Tumor Effect of 14F7 Monoclonal Antibody. *Hybridoma and Hybridomics* **21**, 463-468 (2002).

40. Roque-Navarro, L. et al. Anti-ganglioside antibody-induced tumor cell death by loss of membrane integrity. *Molecular Cancer Therapeutics* **7**, 2033-2041 (2008).

41. Krengel, U. et al. Structure and Molecular Interactions of a Unique Antitumor Antibody Specific for N-Glycolyl GM3. *Journal of Biological Chemistry* **279**, 5597-5603 (2004).

42. Bjerregaard-Andersen, K. et al. Crystal structure of an L chain optimised 14F7 anti-ganglioside Fv suggests a unique tumour-specificity through an unusual H-chain CDR3 architecture. *Scientific Reports* **8**, 10836 (2018).

43. Rojas, G. et al. 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 (2004).

44. Johannesen, H. The Assassin - Recombinant single-chain Fv production of the anti-tumour antibody 14F7. Vol. 2019 Master's Thesis (UiO, 2014).

45. Rodríguez, M., Llanes, L., Pérez, A., Pérez, R. & Vázquez, A.M. Generation and Characterization of an Anti-Idiotype Monoclonal Antibody Related to GM3(NeuGc) Ganglioside. *Hybridoma and Hybridomics* **22**, 307-314 (2003).

46. Michael Madigan, J.M., Kelly Bender, Daniel Buckley, David Stahl. *Brock Biology of Microorganisms, 14th edition*.

47. Thanbichler, M., Wang, S.C. & Shapiro, L. The bacterial nucleoid: A highly organized and dynamic structure. *Journal of Cellular Biochemistry* **96**, 506-521 (2005).

48. Sockolosky, J.T. & Szoka, F.C. Periplasmic production via the pET expression system of soluble, bioactive human growth hormone. *Protein expression and purification* **87**, 129-135 (2013).

49. Sun, L., Bertelshofer, F., Greiner, G. & Böckmann, R.A. Characteristics of Sucrose Transport through the Sucrose-Specific Porin ScrY Studied by Molecular Dynamics Simulations. *Frontiers in bioengineering and biotechnology* **4**, 9-9 (2016).

50. Nossal, N.G. & Heppel, L.A. The Release of Enzymes by Osmotic Shock from Escherichia coli in Exponential Phase. *Journal of Biological Chemistry* **241**, 3055-3062 (1966).

51. Heppel, L.A. Selective Release of Enzymes from Bacteria. *Science* **156**, 1451-1455 (1967).

52. Biosciences, A. HisTrap HP, 1 ml and 5 ml. Vol. 2019 (http://wolfson.huji.ac.il/purification/PDF/Tag_Protein_Purification/Ni-NTA/AMERSHAM_HisTrapHP_Instruct.pdf, 2004).

53. Healthcare, G. Capto™ L. Vol. 2019 (http://wolfson.huji.ac.il/purification/PDF/affinity/GE_CaptoL.PDF, 2012).

54. Healthcare, G. Size exclution chromatiography columns and media. (http://www.gelifesciences.co.kr/wp-content/uploads/2016/07/2016_Selection_guide_Size_exclusion_chromatography_columns_and_media18112419.pdf, 2015).

55. McPherson, A. & Gavira, J.A. Introduction to protein crystallization. *Acta crystallographica. Section F, Structural biology communications* **70**, 2-20 (2013).

56. Gan, S.D. & Patel, K.R. Enzyme Immunoassay and Enzyme-Linked Immunosorbent Assay. *Journal of Investigative Dermatology* **133**, 1-3 (2013).

57. Ghisaidoobe, A.B.T. & Chung, S.J. Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Förster resonance energy transfer techniques. *International journal of molecular sciences* **15**, 22518-22538 (2014).

58. Hoås, H.M. Recombinant production, purification and characterisation of the anti-tumour antibody 14F7 single chain Fv. Vol. 2019 Master's Thesis (UiO, https://www.duo.uio.no/handle/10852/59663, 2017).

59. Bothmann, H. & Pluckthun, A. The periplasmic Escherichia coli peptidylprolyl cis,trans-isomerase FkpA. I. Increased functional expression of antibody fragments with and without cis-prolines. *J Biol Chem* **275**, 17100-5 (2000).

60. Gunnarsen, K.S. et al. Periplasmic expression of soluble single chain T cell receptors is rescued by the chaperone FkpA. *BMC Biotechnology* **10**, 8-8 (2010).

61. Michaelsen, T. et al. *Construction and Functional Activities of Chimeric Mouse-Human Immunoglobulin G and Immunoglobulin M Antibodies against the Neisseria meningitidis PorA P1.7 and P1.16 Epitopes*, 5714-23 (2003).

62. Gibson, D.G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343 (2009).

63. SGIDNA. Gibson Assembly, Cloning Guid, 2. edition. Vol. 2019 (https://www.biocat.com/bc/files/Gibson_Guide_V2_101417_web_version_8.5_x_11_FINAL.pdf, 2017).

64. García-Nafría, J., Watson, J.F. & Greger, I.H. IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. *Scientific Reports* **6**, 27459 (2016).

65. Jeong, J.-Y. et al. One-Step Sequence- and Ligation-Independent Cloning as a Rapid and Versatile Cloning Method for Functional Genomics Studies. *Applied and Environmental Microbiology* **78**, 5440-5443 (2012).

66. Tang, Y., Zeng, X. & Liang, J. Surface Plasmon Resonance: An Introduction to a Surface Spectroscopy Technique. *Journal of chemical education* **87**, 742-746 (2010).

67. Healthcare, G. Superdex 75 Increase columns. Vol. 2019 (GE Healthcare, https://cdn.gelifesciences.com/dmm3bwsv3/AssetStream.aspx?mediaformatid=10061&destinationid=10016&assetid=18066, 2016).

8. Appendix

Appendix A: Materials

Table S- 1. Reagents

Chemicals	Vendor
α–lactose	Sigma
β-mercaptoethanol	Sigma
1 kbp DNA ladder	New England Biolabs
100 bp DNA ladder	New England Biolabs
3,3',5,5'-tetramethylbenzidine (TMB), insoluble	Calbiochem
3,3',5,5'- tetramethylbenzidine (TMB), soluble	Calbiochem
Agarose	Lonza
Acetic Acid	Merck
Alkaline Phosphatase, Calf Intestinal, (CIP)	New England Biolabs
Anti-poly Histidine-Alkaline phosphatase antibody	Sigma
Ampicillin	AppliChem
Ammonium chloride (NH_4Cl)	Sigma
Ammonium sulphate ($(NH_4)_2SO_4$)	Sigma
Benzonase	Novagen®
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
Calf-intestinal alkaline phosphatase	New England Biolabs
Coomassie Brilliant Blue G250	Amersham Biosciences
CutSmart® Buffer	New England Biolabs
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 BL21	New England Biolabs
E. coli Rosetta™	Merck
E. coli T7	New England Biolabs
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 FD restriction enzyme	Thermo Scientific

Hydrogen-Chloride (HCl)	Merck
Imidazole	Sigma
Isopropyl-β-D-Thiogalactopyranoside (IPTG)	Sigma
Ligase buffer, 10x	New England Biolabs
Lysozyme	Sigma
Magnesium chloride ($MgCl_2$)	Fluka
Magnesium sulphate heptahydrate ($MgSO_4 \cdot 7H_2O$)	Fluka
Methanol	Prolabo
MES SDS Running buffer, 20x	Invitrogen
Milli-Q H_2O (mqH_2O)	Millipore
*Mlu*I FD restriction enzyme	Thermo Scientific
Mono Sodium phosphate (NaH_2PO_4)	Fluka
*Mva1269*I restriction enzyme	Thermo Scientific
NeuGc GM3, from collaboration partners	CIM
*Nco*I FD restriction enzyme	Thermo Scientific
NuPAGE loading buffer (4x)	Invitrogen
Nickel(II) sulphate ($NiSO_4$)	Sigma
*Not*I FD restriction enzyme	Thermo Scientific
NuPAGE LDS Sample buffer	Invitrogen
NuPAGE® LDS Sample Buffer (4X)	Invitrogen
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_2PO_4)	Fluka
Primers, Sequencing	Eurofins Genomics
2-Propanol	Sigma
*Pfl23*II restriction enzyme	Thermo Scientific
Protein L-HRP	Genscript
RNase A	Sigma
SeeBlue® Plus2 standard, molecular weight marker	Invitrogen
Sodium carbonate (Na_2CO_3)	Sigma
Sodium chloride (NaCl)	Prolabo
Sodium citrate ($Na_3C_6H_5O_7$)	Sigma
Sodium hydroxide (NaOH)	Kebo Lab
Sodium sulfate (Na_2SO_4)	VWR
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 S- 2. Chromatography columns

Materials	Manufacturer
Chelating Sepharose	GE healthcare
Capto™ Protein L resin	GE Healthcare,
HisTrap™ Ni^{2+} resin	Amersham Biosciences
Superdex 75 10/300 GL	GE Healthcare
Superdex 200 R10/300 Column	GE Healthcare

Table S- 3. Lab equipment

Equipment	Manufacturer
Amicon Ultra free tubes	Millipore
Amicon ultra filter, Ultracell-10 K, 2 ml	Millipore
Amicon ultra filter, Ultracell-10 K, 15 ml	Millipore
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
NuPAGE Bis-Tris 4-12% gel	Life technology
Nunc-Immuno 96 MicroWell MaxiSorp and PolySorb solid plates	

	Sigma
Optifit Refill Tips	Sartorius
Petman pipettes	Gilson
Polysorp Nunc-Immuno 96-well plate	Sigma
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
Triple Sitting Drop 96-well iQ plate	TTP Labtech

Table S- 4. Hardware equipment

Machines	**Vendor**
ÄKTApurifier-900	GE Healthcare
ÄKTA Start + Frac 30	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
Eppendorf Thermomixer comfort	Eppendorf
Gel logic 200 image system	Kodak
JASCO-8500 fluorimeter	Jasco inc
Mini Horizontal Submarine Unit	GE Healthcare
Miniorbital shaker	Stuart

Mosquito® Crystal	TTP Labtech
Multitron II	Infors
MQ-H_2O, Direct Q	Millipore
NanoDrop One	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
VWR Model AS12 Bench-Top Sterilizer	Tuttnauer
VWR Vapour-line autoclave	VWR
Ultraspec III KKB	Pharmacia

Table S- 5. Software

Software	Vendor
Clustal2W	EMBL-EBI
Microsoft Excel	Microsoft
Microsoft Word	Microsoft
Mosquito® Crystal user interface	TTPlabtech
Paintbrush	Hockey SDK
ProtParam, ExPASy	Swiss Institute of Bioinformatics
PyMOL	Schrödinger

RockMaker®	Formulatrix
Snapgene® Viewer 4.3.7	GSL Biotech LLC
Unicorn™ 5.11	GE Healthcare
Unicorn™ start 1.0	GE Healthcare
Wasp Run	Douglas Instruments

Table S- 6. Kits

Kit	Vendor
Nucleo Spin Plasmid kit	Macherey-Nagel GmbH & Co. KG
QIAprep Spin Miniprep Kit	QIAGEN
QIAprep Spin Midiprep Kit	QIAGEN
QIAquick Gel Extraction Kit	QIAGEN

Table S- 7. Crystallization kits

Crystallization kit	Vendor
Structure screen 1 & 2 HT-96	Molecular Dimensions
JCSG-*plus*™ HT-96	Molecular Dimensions
Morpheus® HT-96	Molecular Dimensions
PACT *premier*™ HT-96/FX-96	Molecular Dimensions
Wizard Cryo 1 & 2	Rigaku

Appendix B: Buffers and solutions

All the reagents are listed in Table S- 1. Reagents.

Bacteria growth media

Table S- 8. Medium recipes

LB medium		**2x YT**	
10 g	Peptone	16 g	Peptone
5 g	Yeast extract	10 g	Yeast extract
10 g	NaCl	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*	

TB medium	
20 g	Peptone
24 g	Yeast extract
4 mL	Glycerol
pH 7.2	
MQ-H_2O to final volume of 900 mL	
Autoclaved	
100 mL	Phosphate buffer (17 mM KH_2PO_4, 72 mM K_2HPO_4)
Store at 4°C	

Table S- 9. Protein producing media mixtures

2xYT-AG – Growth media		**2xYT-A – Induction media**	
1861.2 mL	2xYT	1998 mL	2xYT
181.8 mL	1.1 M Glucose		
2.0 mL	100 $^{mg}/_{mL}$ Ampicillin	2.0 mL	100 $^{mg}/_{mL}$ Ampicillin
pH 7.8		pH 7.8	

Autoinduction media

The protocol is provided by Anders Tveita (Rikshospitalet, Institute of Clinical Research, UiO, Norway)

Table S- 10. Autoinduction media, the reagents.

10xZY		
1 %	Tryptone	100 $^{g}/_{L}$
0.5 %	Yeast extract	50 $^{g}/_{L}$
Mix and autoclave		

25xM		
1.25 M	Na_2HPO_4-anhydrid	177.5 $^{g}/_{L}$
1.25 M	KH_2PO_4	170 $^{g}/_{L}$
2.5 M	NH_4Cl	134 $^{g}/_{L}$
0.25 M	Na_2SO_4	35.5 $^{g}/_{L}$
pH 6.75		
Mix and autoclave		

20xP

1.0 M	Na_2HPO_4-anhydrid	142 $g/_L$
1.0 M	KH_2PO_4	136 $g/_L$
0.5 M	$(NH_4)_2SO_4$	66 $g/_L$

pH 6.75

Mix and autoclave

50x5052

Prepare and autoclave each component separately in water, autoclave, and then combine!

25 %	Glycerol	250 $g/_L$
2.5 %	Glucose	25 $g/_L$
10 %	α-lactose	100 $g/_L$

Mix and autoclave

500x$MgSO_4$

1.0 M	$MgSO_4 \cdot 7H_2O$	246.5 $g/_L$

Mix and autoclave

Table S- 11. Autoinduction medium, recipe.

1.0 L Autoinduction medium	
100 mL	10xZY
40 mL	25xM
50 mL	20xP
20 mL	50x5052
2 mL	500x$MgSO_4$
778 mL	mqH_2O

Protein extraction solutions

Table S- 12. Buffers for Protein extraction

Periplasmic extraction solutions		**Cell lysis buffer – Sonication**	
Sucrose solution		0.15 $^{mg}/_{mL}$	Lysozyme
25%	Sucrose	50 mM	Tris-HCl
50 mM	Tris-HCl	150 mM	NaCl
pH 7.4		1 mM	EDTA
			Benzonase
$MgCl_2$ solution		1 tab	Per 100 ml c*O*mplete inhibitor
5 mM	$MgCl_2$	pH 7.4	
1 tab	Per 100 ml c*O*mplete inhibitor		
pH 7.4			

Protein purification buffers

Table S- 13. His-Trap buffers

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 7.4		pH 7.4	
The buffer is filtered and degassed before use		*The buffer is filtered and degassed before use*	

Table S- 14. Capto™ Protein L buffers

Protein L binding buffer		**Protein L elution buffer**	
0.1 M	Tris-HCl	0.1 M	Glycine
0.15 M	NaCl		
pH 7.4		pH 2.5	
The buffer is filtered and degassed before use		*The buffer is filtered and degassed before use*	

Table S- 15. Tris Crystallization buffer for SEC

SEC running buffer	
20 mM	Tri-HCl
100 mM	NaCl
pH 7.4	
The buffer is filtered and degassed before use	

Buffer for Agarose gel electrophoresis

Table S- 16. Agarose running and gel casting buffer

1x TAE-Running buffer	
40 mM	Tris base
20 mM	Acetic acid
1 mM	EDTA
pH 8.3	

Buffer for SDS-PAGE electrophoresis

Table S- 17. Coomassie quick stain.

Coomassie quick stain – 1.0 L	
80 mg	Coomassie (G-250)
5 mL	12.1 M HCl
995 mL	mqH_2O
pH 2.0	

Appendix C: Restriction enzyme digestion

Table S-18. Reaction set-up for RE digestion of pLNOH2 and pLNOκ vectors with New England Biolabs reagents. Uncut and singularly cut vector fragments are used as controls for RE activity. Double digested vectors are in mixture column 4 and 5 (BmiWI-HF & BmsI).

	pLNOκ				pLNOH2			
	Uncut	BmsI	BmiWI-HF	BmiWI-HF & BmsI	BmiWI-HF & BmsI	BmiWI-HF	BmsI	Uncut
mqH_2O	17 µL	16 µL	16 µL	15 µL	15 µL	16 µL	16 µL	17 µL
CutSmart® Buffer	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL
DNA	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL
BsiWI-HF	–	–	1 µL	1 µL	1 µL	1 µL	–	–
BsmI	–	1 µL	–	1 µL	1 µL	–	1 µL	–
Total volum	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL

Table S- 19. Reaction set-up for RE digestion of 4G9 V_H and V_L fragments with New England Biolabs reagents. Uncut and singularly cut vector fragments are used as controls for RE activity. Double digested vectors are in mixture column 4 and 5 (BmiWI-HF & BmsI).

	pEX-A128-4G9_VL				pEX-A128-4G9_VH			
	Uncut	BmsI	BmiWI-HF	BmiWI-HF & BmsI	BmiWI-HF & BmsI	BmiWI-HF	BmsI	Uncut
mqH_2O	17 µL	16 µL	16 µL	15 µL	15 µL	16 µL	16 µL	17 µL
CutSmart® Buffer	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL
DNA	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL
BsiWI-HF	–	–	1 µL	1 µL	1 µL	1 µL	–	–
BsmI	–	1 µL	–	1 µL	1 µL	–	1 µL	–
Total volum	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL

Table S-20. Reaction set-up for RE digestion of 4G9 V_H and V_L fragments with Thermo Scientific reagents. Uncut and singularly cut vector fragments are used as controls for RE activity. Double digested vectors are in mixture column 4 and 5 (Pfl23II & Mva1269I).

	pEX-A128-4G9_VL				pEX-A128-4G9_VH			
	Uncut	Mva1269I	Pfl23II	Pfl23II & Mva1269I	Pfl23II & Mva1269I	Pfl23II	Mva1269I	Uncut
mqH_2O	17 µL	16 µL	16 µL	15 µL	15 µL	16 µL	16 µL	17 µL
FD Green Buffer	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL
DNA	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL
Pfl23II	–	–	1 µL	1 µL	1 µL	1 µL	–	–
Mva1269I	–	1 µL	–	1 µL	1 µL	–	1 µL	–
Total volum	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL

Table S-21. Reaction set-up for RE digestion of pLNOH2 and pLNOκ vectors with Thermo Scientific reagents. Uncut and singularly cut vector fragments are used as controls for RE activity. Double digested vectors are in mixture column 4 and 5 (Pfl23II & Mva1269I).

	pLNOκ				pLNOH2			
	Uncut	Mva1269I	Pfl23II	Pfl23II & Mva1269I	Pfl23II & Mva1269I	Pfl23II	Mva1269I	Uncut
mqH_2O	17 µL	16 µL	16 µL	15 µL	15 µL	16 µL	16 µL	17 µL
FD Green Buffer	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL
DNA	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL
Pfl23II	–	–	1 µL	1 µL	1 µL	1 µL	–	–
Mva1269I	–	1 µL	–	1 µL	1 µL	–	1 µL	–
Total volum	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL	20 µL

*Pfl23*II	↓ 5′- C GTAC G - 3′ 3′- G CATG C - 5′ ↑
*Mva1269*I	↓ 5′- GAATGCN - 3′ 3′- CTTAC GN - 5′ ↑

*Bsi*WI	↓ 5′- C GTAC G - 3′ 3′- G CATG C - 5′ ↑
*Bsm*I	↓ 5′- GAATGCN - 3′ 3′- CTTAC GN - 5′ ↑

Table S- 22. Isolated fragments. The wanted fragments form agarose gel electrophoresis purification of RE samples. Concentration measured with NanoDrop™ One (Thermo Scientific).

Vector	Fragment size	Concentration
pLNOκ	8691 bp	67.8 $^{ng}/_{\mu L}$
pLNOH2	10626 bp	40.9 $^{ng}/_{\mu L}$
4G9_V_H	380 bp	25.9 $^{ng}/_{\mu L}$
4G9_V_L	347 bp	14.7 $^{ng}/_{\mu L}$

Appendix D: DNA ligation

Equation S- 1. Calculation of volumetric ration of insert : vector after RE treatment to the final molar ration of 4:1 insert : vector.

$$\frac{m(Vector) \bullet n(Insert) \bullet 4}{n(Vector) \bullet 1} = m(Insert)$$

Equation S- 2. The general relation between concentration and volume at two points in time (1 and 2). The same relation can be translated to Vector : Insert relations.

$$c_1 \bullet V_1 = c_2 \bullet V_2$$

$$c(Vector) \bullet V(Vector) = c(Insert) \bullet V(Insert)$$

Since the calculation form bp to mol goes through the same constant molar mass per base pair (Mm(bp) ≈ 650 $^{g}/_{mol \bullet bp}$), one simple uses the size of the insert an vector directly. The vector mass is set to be 60 ng, vector and insert data are from Table S-22.

pLNOκ : 4G9_V_L

$$\frac{60\ ng \bullet 370\ bp \bullet 4}{10626\ bp} = 8.36\ ng \rightarrow 8.36\ {}^{ng}/_{\mu L}$$

$$Volume(pLNO\kappa) = \frac{60.0\ ng}{40.9\ {}^{ng}/_{\mu L}} = 1.47\ \mu L$$

$$Volume(4G9\_V_L) = \frac{8.36\ {}^{ng}/_{\mu L} \bullet 1.47\ \mu L}{25.9\ {}^{ng}/_{\mu L}} = 0.47\ \mu L$$

pLNOH2 : 4G9_V_H

$$\frac{60\ ng \bullet 347\ bp \bullet 4}{8691\ bp} = 9.58\ ng \rightarrow 9.58\ {}^{ng}/_{\mu L}$$

$$Volume(pLNOH2) = \frac{60.0\ ng}{67.8\ {}^{ng}/_{\mu L}} = 0.88\ \mu L$$

$$Volume(4G9\_V_H) = \frac{9.58\ {}^{ng}/_{\mu L} \bullet 0.88\ \mu L}{14.7\ {}^{ng}/_{\mu L}} = 0.58\ \mu L$$

Table S- 23. DNA ligation setup. The same volumes and ratios where used with both NEB reagents and ThermoScinetific reagents. To calculate the molar ratio between vector and insert to 1:4, Equation 1 and 2 where used, and measurements in Table S-22.

	pLNOκ : 4G9_V_L	pLNOH2 : 4G9_V_H
mqH_2O	15.0 µL	15.5 µL
10x T4 DNA Ligase buffer	2.0 µL	2.0 µL
Vector (pLNO)	1.5 µL	0.9 µL
Insert (4G9)	0.5 µL	0.6 µL
T4 DNA Ligase	1.0 µL	1.0 µL

Final total volume of 20 µL was incubated at 22°C for 1 hour.

Appendix E: Plasmid for 4G9 Fab

Vector map of scFv 14F7 and 4G9

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.

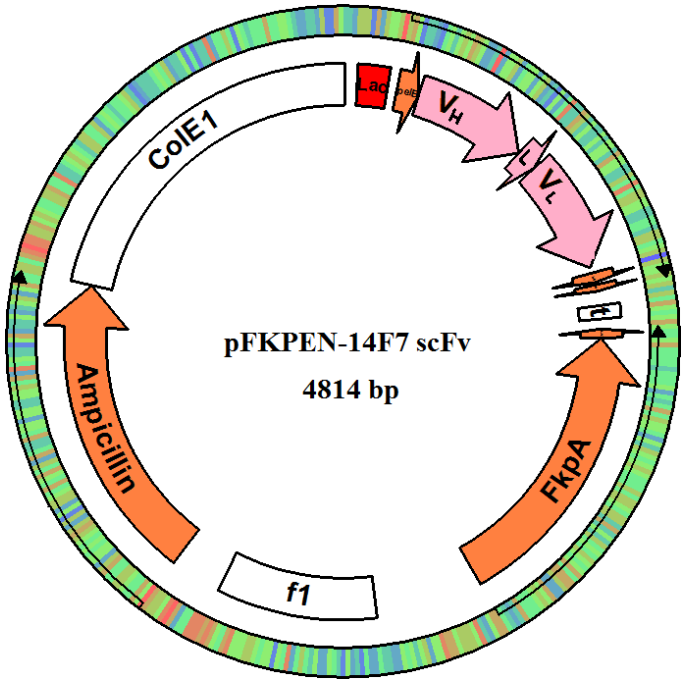


Figure S- 1. Vector map of pFKPEN plasmid used to express all scFv 14F7 constructs and scFv 4G9. V_H are flanked by RE sites *Noc*I (Thermo Scientific) and *Hind*III (Thermo Scientific) and V_L by *Mlu*I (Thermo Scientific) and *Not*I (Thermo Scientific).

pFABEFN-HaLb380:

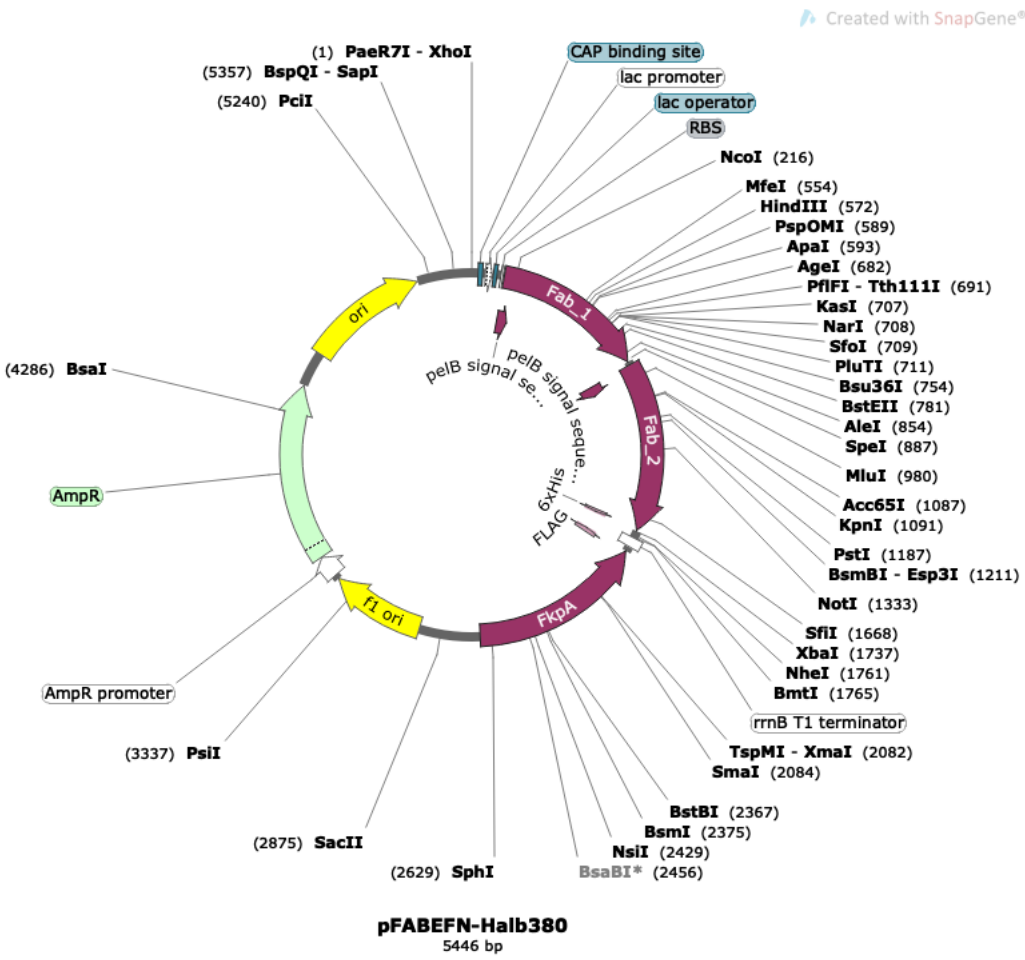


Figure S- 2. pFABEFN-HeLa380 vector map generated by SnapGene® for bacterial Fab production. 4G9_VH was to be inserted in Fab-1 between RE sites *Noc*I (Thermo Scientific) and *Hind*III (Thermo Scientific) and 4G9_VL to be inserted in Fab_2 between *Mlu*I (Thermo Scientific) and *Not*I (Thermo Scientific).

pEX-A128-4G9_VH and pEX-A128-4G9_VL:

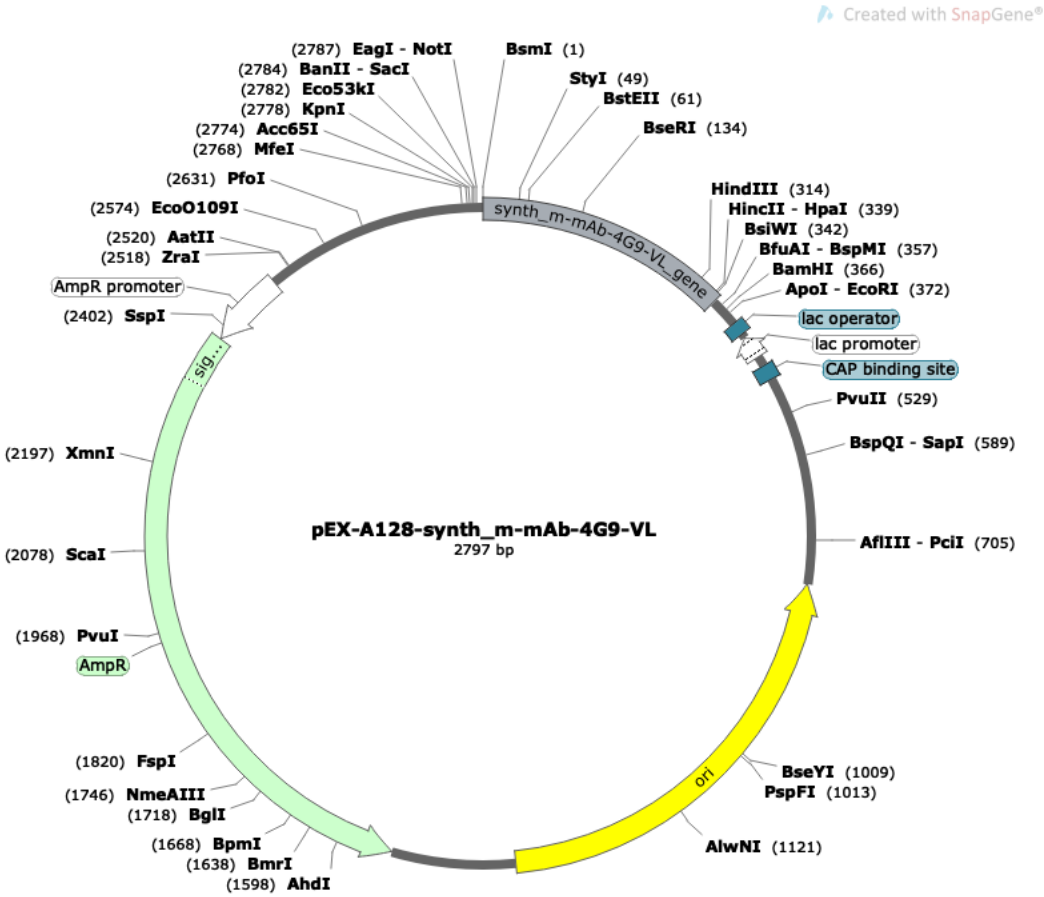


Figure S- 3. Vector map generated by SnapGene® of the humanized 4G9_VL fragment flanked by *Bsm*I (NEBiolabs)/*Mva1269*I (Thermo Scientific) and *Bsi*WI-HF (NEBiolabs)/ *Pfl23*II (Thermo Scientific) RE sites. pEX-A128 is a commercial 2450 bp vector form Eurofins who constructed our humanized gene sequence.

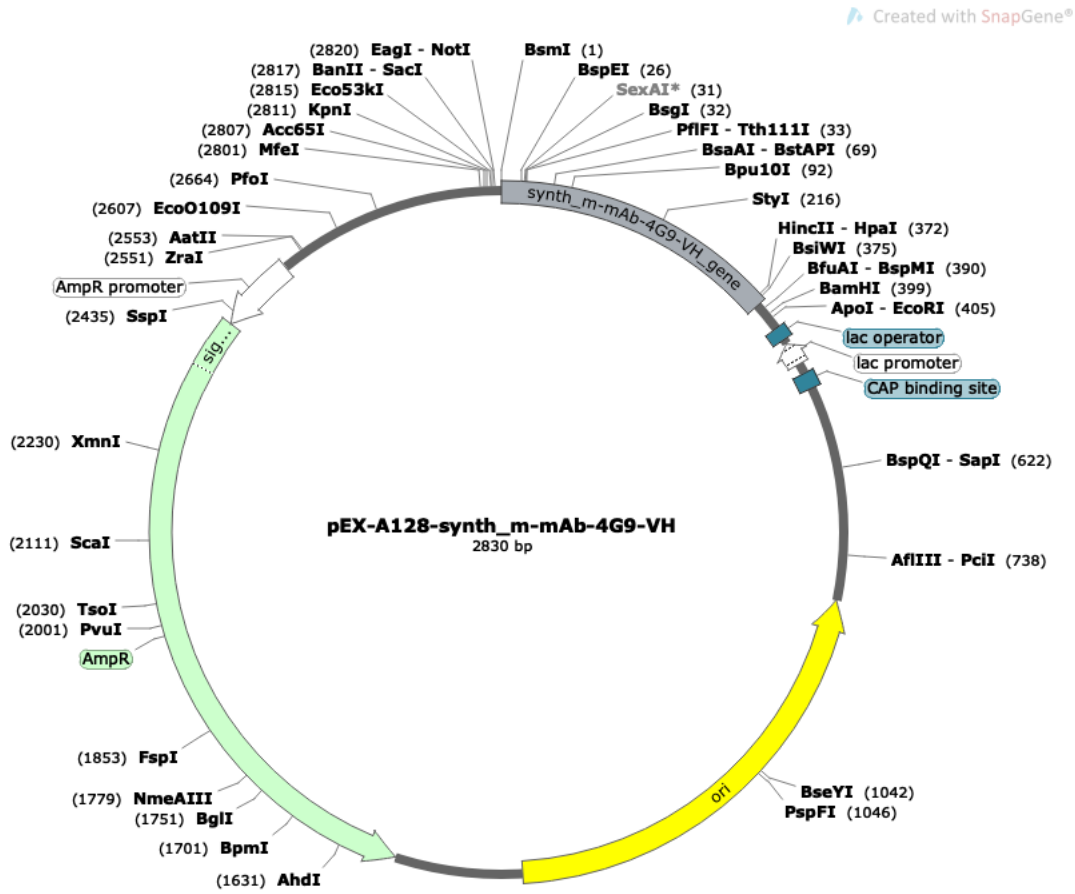


Figure S- 4. Vector map generated by SnapGene® of the humanized 4G9_VH fragment flanked by *Bsm*I (NEBiolabs)/*Mva1269*I (Thermo Scientific) and *Bsi*WI-HF (NEBiolabs)/ *Pfl23*II (Thermo Scientific) RE sites. pEX-A128 is a commercial 2450 bp vector form Eurofins who constructed our humanized gene sequence.

pLNOH2 and pLNOκ:

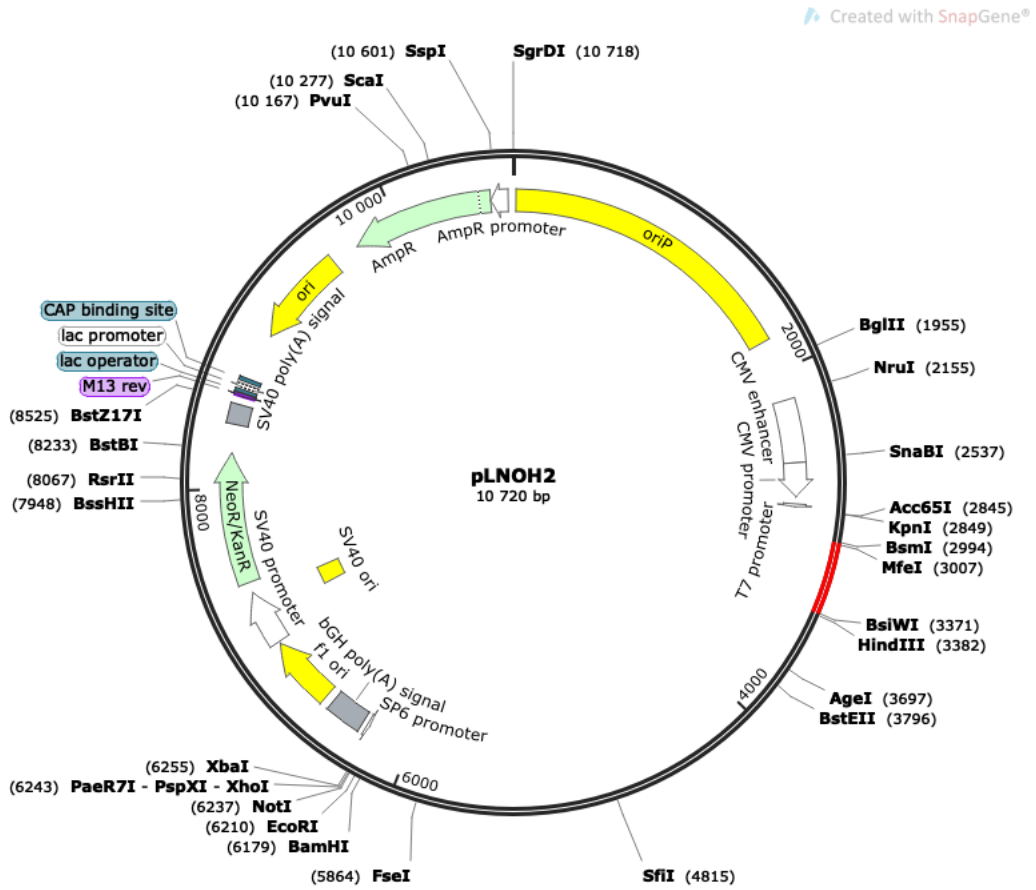


Figure S- 5. Vector map of pLNOH2 generated by SnapGene®. Our 4G9_VH gene is inserted between *Bsm*I (NEBiolabs)/*Mva1269*I (Thermo Scientific) and *Bsi*WI-HF (NEBiolabs)/ *Pfl23*II (Thermo Scientific) RE sites (red).

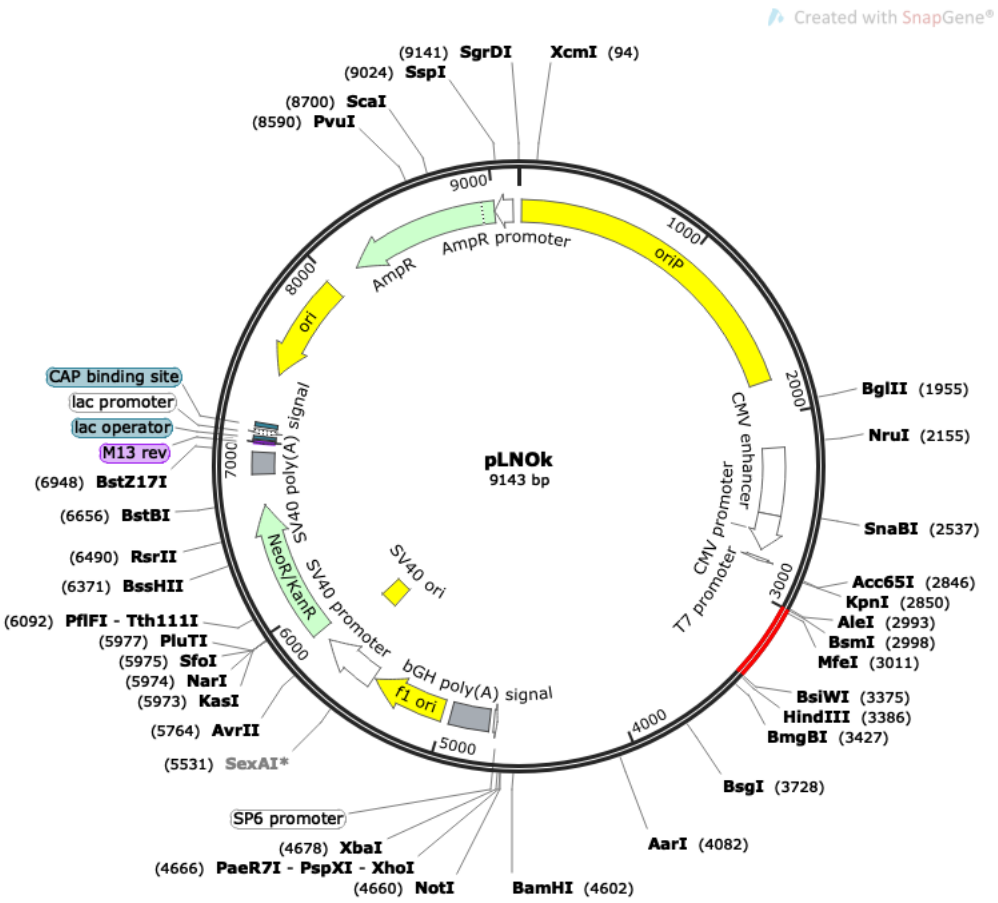


Figure S- 6. Vector map of pLNOκ generated by SnapGene®. Our 4G9_VL gene is inserted between *Bsm*I (NEBiolabs)/*Mva1269*I (Thermo Scientific) and *Bsi*WI-HF (NEBiolabs)/ *Pfl23*II (Thermo Scientific) RE sites (red).

Appendix F: Standards and calculations

Standard curve for Superdex 75 10/300 column

Table S- 24. Superdex 75 10/300 (GE Healthcare) standard curve. Calculated values data sheet provided by GE Healthcare for Superdex 75 (Figure 4B)[67]. Peak top is calculated from 0.8 $^{mg}/_{mL}$ flow rate to 1.0 $^{mg}/_{mL}$ flow rate that we used. Microsoft Excel was used for calculations.

Protein Standard	Mw (Da)	Peak top (mL)	log (Mw)
Conalbumin	75000	9,30	4,88
Ovalbumin	44000	10,01	4,64
Carbonic anhydrase	29000	11,31	4,46
Cytochrome C	12300	13,32	4,09
Aprotinin	6500	15,72	3,81
Vitamin B_{12}	1300	18,38	3,11

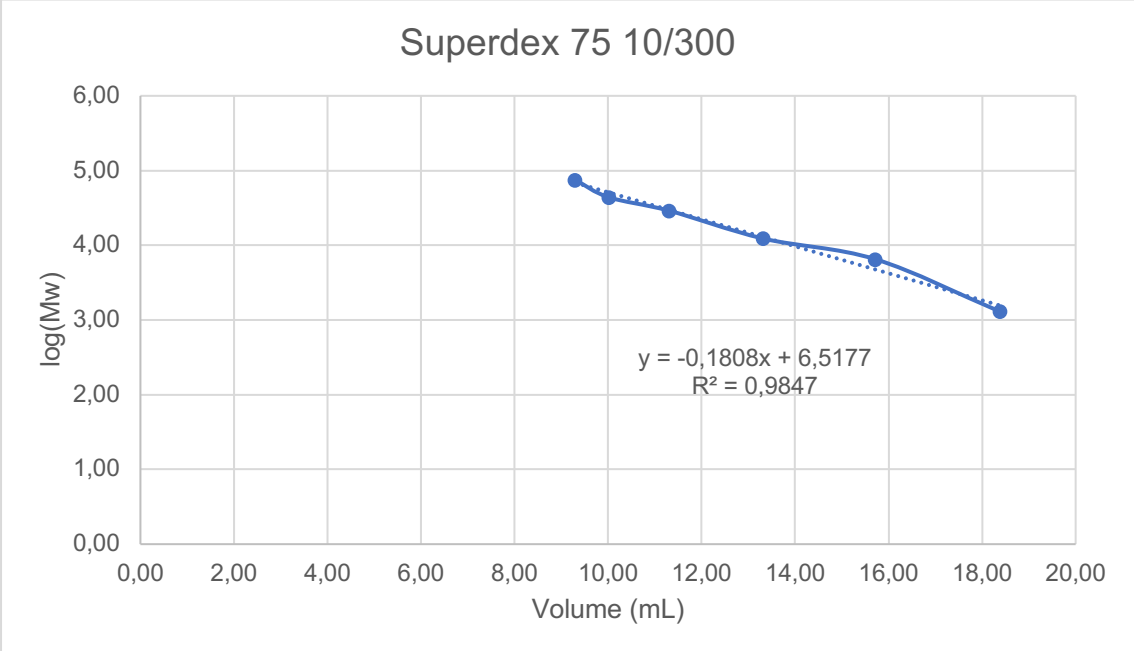


Figure S- 7. Standard curve for Suoerdex 75 10/300 (GE Healtcare), from values in Table S-24. Graf was calculated and constructed with Microsoft Excel.

Equation S- 3. The standard curve of Superdex 75 10/300

$$y = -0.1808x + 6{,}5177$$

Where "y" is the logarithm of the molecular weight of your protein in Dalton (Da), and "x" is the expected volume in mL for the proteins peak top.

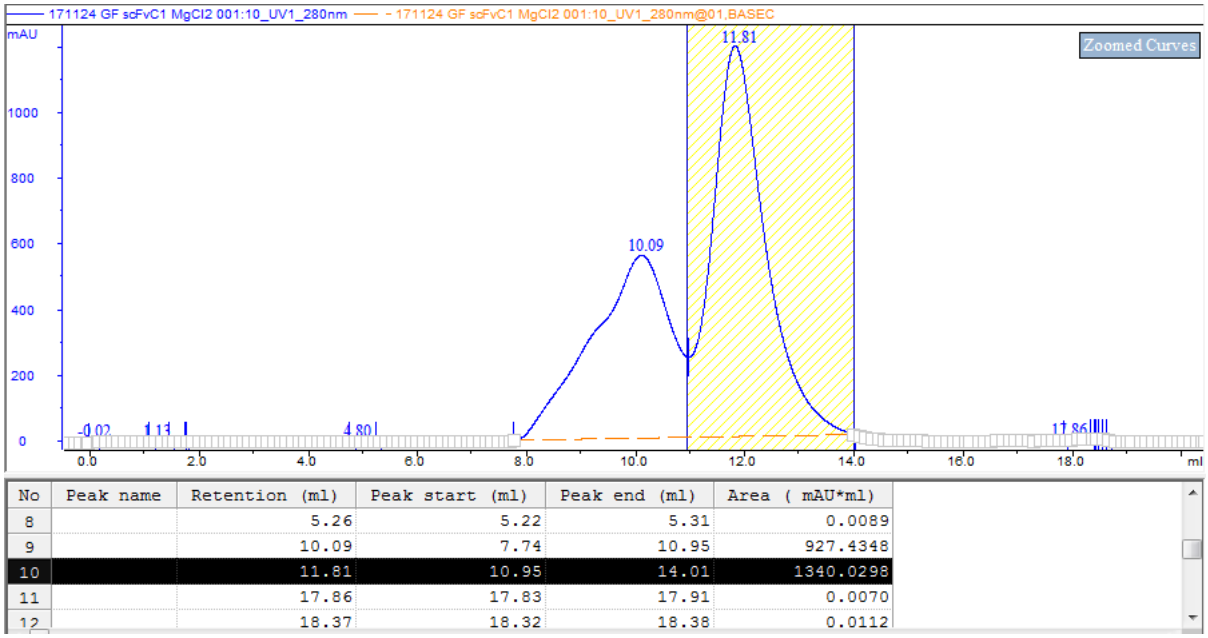
ScFv 14F7 C1*-C4* and scFv 4G9 has a molecular weight of 28-30 kDa (Table 3.4-1), the approximately same as carbonic anhydrase (Table S-24).

With the flow rate of 1.0 $^{mL}/_{min}$ will scFv 14F7 C4* and scFv 4G9 have their respective peak top at:

scFv 14F7 C4*: $x = \frac{\log(28211.50\ Da) - 6.5177}{-0.1808} = 11.43\ mL$

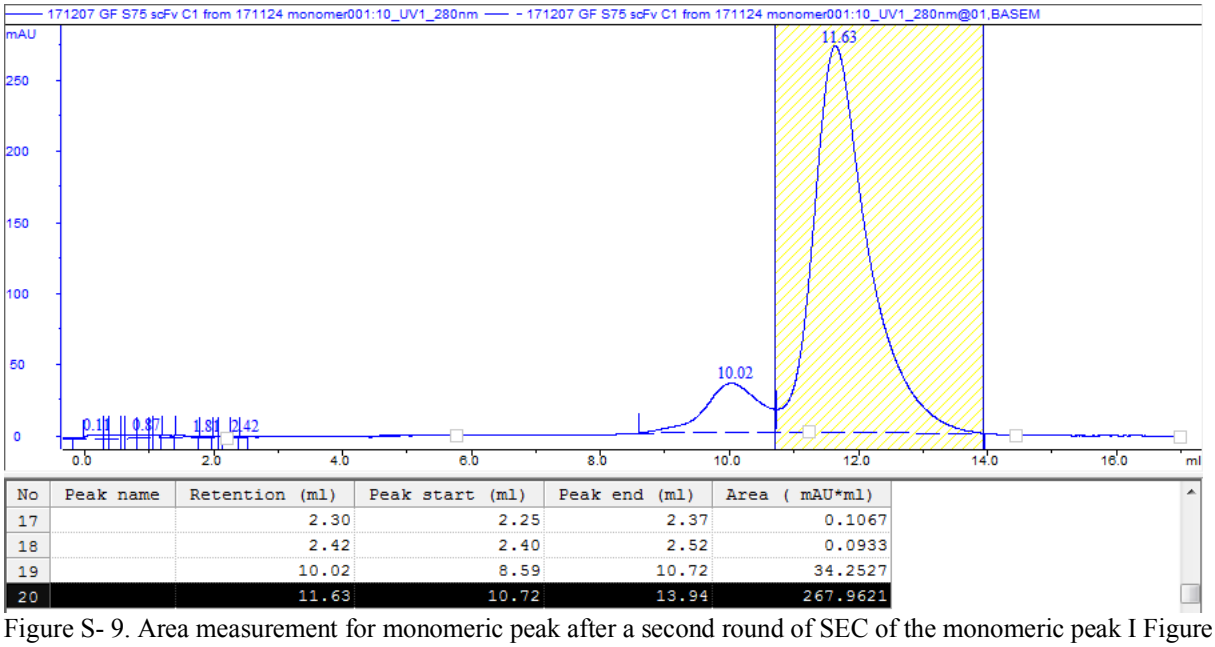
scFv 4G9: $x = \frac{\log(28029.17\ Da) - 6.5177}{-0.1808} = 11.45\ mL$

Fraction collection and area vs measured concentration



No	Peak name	Retention (ml)	Peak start (ml)	Peak end (ml)	Area (mAU*ml)
8		5.26	5.22	5.31	0.0089
9		10.09	7.74	10.95	927.4348
10		11.81	10.95	14.01	1340.0298
11		17.86	17.83	17.91	0.0070
12		18.37	18.32	18.38	0.0112

Figure S- 8. Area measurement for monomeric peak from SEC of 14F7 C1*. The area corresponds to 37 µL of 6.4 $^{mg}/_{mL}$.



No	Peak name	Retention (ml)	Peak start (ml)	Peak end (ml)	Area (mAU*ml)
17		2.30	2.25	2.37	0.1067
18		2.42	2.40	2.52	0.0933
19		10.02	8.59	10.72	34.2527
20		11.63	10.72	13.94	267.9621

Figure S- 9. Area measurement for monomeric peak after a second round of SEC of the monomeric peak I Figure S-8. This peak corresponds to the concentration of 2.4$^{mg}/_{mL}$ (20 µL)

From Figure S-8; The elution volume of 14.01 mL – 10.95 mL = 3.06 mL, and peak area of 1340.03 mAU•mL should give 37 µL of 6.4 $^{mg}/_{mL}$:

$$\frac{6.4\left[{}^{mg}/_{mL}\right] \bullet 37 \bullet 10^{-3}\ [mL]}{\frac{1340.03\ [mAU \bullet mL]}{3.06\ mL}} = 5.407 \bullet 10^{-4}\left[{}^{mg}/_{mAU}\right]$$

If on translates this estimation form Figure S-8 to Figure S-9 with peak area of 267.96 mAU•mL and elution volume of (13.94 mL – 10.72 mL) 3.22 mL, on would get 20 µL of concentration:

$$\frac{5.41 \bullet 10^{-4}\left[{}^{mg}/_{mAU}\right] \bullet 267.3621\ [mAU \bullet mL]}{3.22\ [mL] \bullet 20 \bullet 10^{-3}\ [mL]} = 2.24\ \left[{}^{mg}/_{mL}\right]$$

Which is acceptable considering the uncertainty in programs parameter free peak area calculation and the uncertainty in NanoPhotometer instrument measurments.

Growth curve with different media

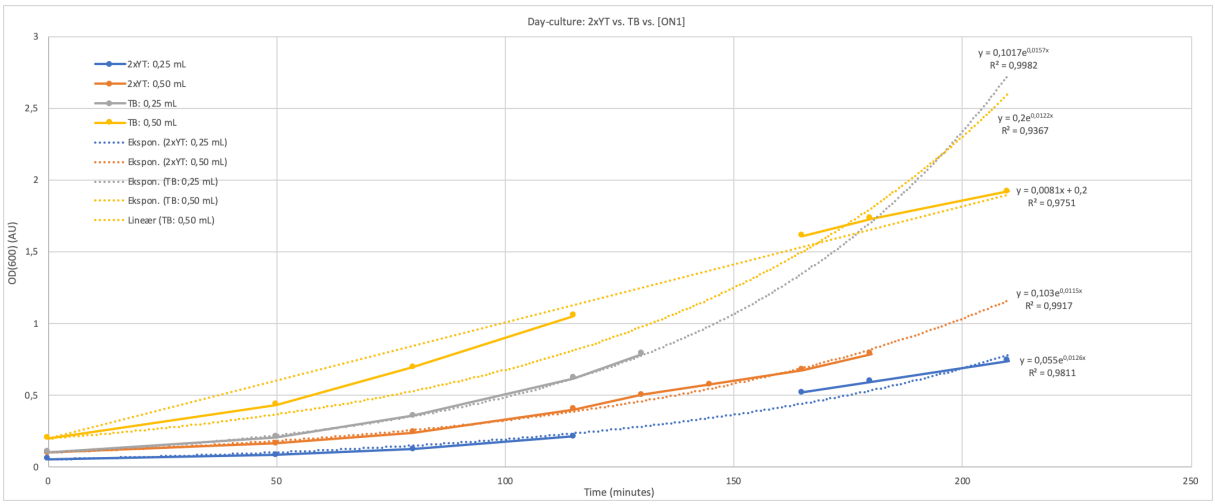


Figure S- 10. Comparison between pre-induction cultures growth for XL1-Blue *E. coli* in 2xYT-AG media and TB-AG media. Induction starts at OD_{600} = 0.7-0.8 AU by removing glucose.

Table S- 25. Measured optical densities at 600 nm for 2xYT-AG and TB-AG at different times. OD_{600} are plotted against time in Figure S-4.

	2xYT		TB	
Time (min)	2xYT: 0,25 mL Bacteria	2xYT: 0,50 mL Bacteria	TB: 0,25 mL Bacteria	TB: 0,50 mL Bacteria
0	0,055	0,103	0,105	0,200
50	0,085	0,165	0,210	0,436
80	0,126	0,240	0,360	0,695
115	0,215	0,402	0,620	1,057
130		0,504	0,790	
145		0,576		
165	0,520	0,675		1,610
180	0,595	0,790		1,730
210	0,741			1,922

Table S- 26. Raw data for induction curves of Figure 4.3-3. There were used three replicates of each reaction set-up (A, B, and C).

		2xYT					TB				
Time:	pre-induction	0,773					pre-induction	0,790			
Time (min)		A	B	C	Average	Standard dviation	A	B	C	Average	Standard diviation
112	Without IPTG	1,045	1,017	1,071	1,044	0,027	1,175	1,182	1,183	1,180	0,004
112	0.05 mM IPTG	0,985	0,990	1,005	0,993	0,010	1,220	1,206	1,207	1,211	0,008
112	0.10 mM IPTG	1,002	1,001	0,983	0,995	0,011	1,212	1,199	1,198	1,203	0,008
238	Without IPTG	1,500	1,465	1,665	1,543	0,107	1,775	1,830	1,865	1,823	0,045
238	0.05 mM IPTG	1,380	1,500	1,465	1,448	0,062	1,665	1,600	1,865	1,710	0,138
238	0.10 mM IPTG	1,430	1,480	1,400	1,437	0,040	1,635	1,655	1,690	1,660	0,028
360	Without IPTG	1,805	1,775	1,885	1,822	0,057	2,305	2,215	2,220	2,247	0,051
360	0.05 mM IPTG	1,605	1,630	1,650	1,628	0,023	1,755	1,720	1,730	1,735	0,018
360	0.10 mM IPTG	1,630	1,625	1,585	1,613	0,025	1,570	1,695	1,625	1,630	0,063
502	Without IPTG	2,075	2,240	2,245	2,187	0,097	2,875	2,825	2,810	2,837	0,034
502	0.05 mM IPTG	1,650	1,700	1,600	1,650	0,050	2,120	2,200	2,285	2,202	0,083
502	0.10 mM IPTG	1,655	1,655	1,650	1,653	0,003	2,180	2,250	2,290	2,240	0,056
592	Without IPTG	2,380	2,375	2,535	2,430	0,091	3,125	3,025	3,085	3,078	0,050
592	0.05 mM IPTG	1,575	1,625	1,450	1,550	0,090	2,335	2,300	2,260	2,298	0,038
592	0.10 mM IPTG	1,415	1,500	1,545	1,487	0,066	2,200	2,195	2,200	2,198	0,003
1082	Without IPTG	2,970	3,010	3,000	2,993	0,021	4,410	4,390	4,530	4,443	0,076
1082	0.05 mM IPTG	0,990	1,310	1,390	1,230	0,212	3,540	3,390	3,120	3,350	0,213
1082	0.10 mM IPTG	1,200	1,170	1,270	1,213	0,051	3,680	3,360	3,750	3,597	0,208
1196	Without IPTG	3,140	3,065	3,250	3,152	0,093	4,540	4,680	4,755	4,658	0,109

1196	0.05 mM IPTG	1,250	1,845	1,945	1,680	0,376	4,385	3,965	3,960	4,103	0,244
1196	0.10 mM IPTG	1,655	1,575	1,665	1,632	0,049	4,380	4,550	4,350	4,427	0,108
1350	Without IPTG	3,070	3,000	2,905	2,992	0,083	5,500	5,450	5,620	5,523	0,087
1350	0.05 mM IPTG	1,650	2,125	2,295	2,023	0,334	4,580	5,270	4,970	4,940	0,346
1350	0.10 mM IPTG	1,895	1,885	1,865	1,882	0,015	5,400	5,360	6,730	5,830	0,780

Appendix G: DNA and amino acid sequences

All nucleotides and amino acids partaking in the polyhistidine tag are colored red

Underlined nucleotides are involved in one of four restriction enzyme sites, upstream NcoI or downstream *Not*I, *Hind*III separating V_L and linker or *Mlu*I separating the linker and V_H.

Restriction sites (underlined):

*Noc*I: 5' – C|CATG G – 3'
3' – G GTAC|G – 5'

*Not*I: 5' – GC|GGCC CG – 3'
3' – CG CCGG|GC – 5'

*Hind*III: 5' – A|AGCT T – 3'
3' – T TCGA|A – 5'

*Mlu*I: 5' – A|CGCG T – 3'
3' – T GCGC|A – 5'

All the nucleic acid sequences below are given in 5'→3' direction.

Amino acids and nucleotides in grey are recognized as involved in tobacco etch virus (TEV) protease cleavage site.

TEV: ENLYFQ|G

The amino acid sequences below are given in N-terminal → C-terminal direction.

The mark " * " is describing a translation stop codon. All the sequences below carry's the same stop codon "TAA".

scFv 14F7 C1*

DNA sequence (780 bp)

GCC ATG GCC CAG GTG CAG CTG CAG CAG AGC GGC GCG GAA CTG GCG AAA CCG GGC GCG AGC ATG AAA ATG AGC TGC CGC GCG AGC GGC TAT AGC TTT ACC AGC TAT TGG ATT CAT TGG CTG AAA CAG CGC CCG GAT CAG GGC CTG GAA TGG ATT GGC TAT ATT GAT CCG GCG ACC GCG TAT ACC GAA AGC AAC CAG AAA TTT AAA GAT AAA GCG ATT CTG ACC GCG GAT CGC AGC AGC AAC ACC GCG TTT ATG TAT CTG AAC AGC CTG ACC AGC GAA GAT AGC GCG GTG TAT TAT TGC GCG CGC GAA AGC CCG CGC CTG CGC CGC GGC ATT TAT TAT TAT GCG ATG GAT TAT TGG GGC CAG GGC ACC ACC GTG ACC GTG AGC AGC AAG CTT TCA GGG AGT GCA TCC GCC CCA AAA CTT GAA GAA GGT GAA TTT TCA GAA GCA CGC GTA GAC ATC CAG ATG ACC CAG ACC CCG TCT TCT CTG TCT GCT TCT CTG GGT GAC CGT GTT ACC ATC TCT TGC CGT GCT TCT CAG GAC ATC TCT AAC TAC CTG AAC TGG TAC CAG CAG AAA CCG GAC GGT ACC GTT AAA CTG CTG ATC TAC TAC ACC TCT CGT CTG CAC TCT GGT GTT CCG TCT CGT TTC TCT GGT TCT GGT TCT GGT ACC GAC TAC TCT CTG ACC ATC TCT AAC CTG GAA CAG GAA GAC ATC GCT ACC TAC TTC TGC CAG CAG GGT AAC ACC CTG CCG CCG ACC TTC GGT GCT GGT ACC AAA CTG GAA CTG AAA TAA GCG GCC GCT

Amino acid sequence (260 aa)

AMAQVQLQQS GAELAKPGAS MKMSCRASGY SFTSYWIHWL KQRPDQGLEW IGYIDPATAY TESNQKFKDK AILTADRSSN TAFMYLNSLT SEDSAVYYCA RESPRLRRGI YYYAMDYWGQ GTTVTVSSKL SGSASAPKLE EGEFSEARVD IQMTQTPSSL SASLGDRVTI SCRASQDISN YLNWYQQKPD GTVKLLIYYT SRLHSGVPSR FSGSGSGTDY SLTISNLEQE DIATYFCQQG NTLPPTFGAG TKLELK*AAA

scFv 14F7 C2*

DNA sequence (780 bp)

GCC ATG GCC CAG GTG CAG CTG CAG CAG AGC GGC GCG GAA CTG GCG AAA CCG GGC GCG AGC ATG AAA ATG AGC TGC CGC GCG AGC GGC TAT AGC TTT ACC AGC TAT TGG ATT CAT TGG CTG AAA CAG CGC CCG GAT CAG GGC CTG GAA TGG ATT GGC TAT ATT GAT CCG GCG ACC GCG TAT ACC GAA AGC AAC CAG AAA TTT AAA GAT AAA GCG ATT CTG ACC GCG GAT CGC AGC AGC AAC ACC GCG TTT ATG TAT CTG AAC AGC CTG ACC AGC GAA GAT AGC GCG GTG TAT TAT TGC GCG CGC GAA AGC CCG CGC CTG CGC CGC GGC ATT TAT TAT TAT GCG ATG GAT TAT TGG GGC CAG GGC ACC ACC GTG ACC GTG AGC AGC AAG CTT TCA GGG AGT GCA TCC GCC CCA AAA CTT GAA GAA GGT GAA TTT TCA GAA GCA CGC GTA GAC CTG GTT CTG ACC CAG TCT CCG GCT ACC CTG TCT GTT ACC CCA GGT GAC TCT GTT TCT TTC TCT TGC CGT GCT TCT CAG TCT ATC TCT AAC AAC CTG CAC TGG TAC CAG CAG CGT ACC CAC GAA TCT CCG CGT CTG CTG ATC AAA TAC GCT TCT CAG TCT ATC TCT GGT ATC CCG TCT CGT TTC TCT GGT TCT GGT TCT GGT ACC GAC TTC ACC CTG TCT ATC TCT TCT GTT GAA ACC GAA GAC TTC GGT ATG TAC TTC TGC CAG CAG TCT AAC CGT TGG CCG CTG ACC TTC GGT GCT GGT ACC AAA CTG GAA CTG AAA TAA GCG GCC GCT

Amino acid sequence (260 aa)

AMAQVQLQQS GAELAKPGAS MKMSCRASGY SFTSYWIHWL KQRPDQGLEW IGYIDPATAY TESNQKFKDK AILTADRSSN TAFMYLNSLT SEDSAVYYCA RESPRLRRGI YYYAMDYWGQ GTTVTVSSKL SGSASAPKLE EGEFSEARVD LVLTQSPATL SVTPGDSVSF SCRASQSISN NLHWYQQRTH ESPRLLIKYA SQSISGIPSR FSGSGSGTDF TLSISSVETE DFGMYFCQQS NRWPLTFGAG TKLELK*AAA

scFv 14F7 C3*

DNA sequence (780 bp)

GCC ATG GCC CAG GTG CAG CTG CAG CAG AGC GGC GCG GAA CTG GCG AAA CCG GGC GCG AGC ATG AAA ATG AGC TGC CGC GCG AGC GGC TAT AGC TTT ACC AGC TAT TGG ATT CAT TGG CTG AAA CAG CGC CCG GAT CAG GGC CTG GAA TGG ATT GGC TAT ATT GAT CCG GCG ACC GCG TAT ACC GAA AGC AAC CAG AAA TTT AAA GAT AAA GCG ATT CTG ACC GCG GAT CGC AGC AGC AAC ACC GCG TTT ATG TAT CTG AAC AGC CTG ACC AGC GAA GAT AGC GCG GTG TAT TAT TGC GCG CGC GAA AGC CCG CGC CTG CGC CGC GGC ATT TAT TAT TAT GCG ATG GAT TAT TGG GGC CAG GGC ACC ACC GTG ACC GTG AGC AGC AAG CTT GCG CCG CAG GCG AAA AGC AGC GGC AGC GGC AGC GAA AGC AAA GTG GAT GCA CGC GTA GAC ATC CAG ATG ACC CAG ACC CCG TCT TCT CTG TCT GCT TCT CTG GGT GAC CGT GTT ACC ATC TCT TGC CGT GCT TCT CAG GAC ATC TCT AAC TAC CTG AAC TGG TAC CAG CAG AAA CCG GAC GGT ACC GTT AAA CTG CTG ATC TAC TAC ACC TCT CGT CTG CAC TCT GGT GTT CCG TCT CGT TTC TCT GGT TCT GGT TCT GGT ACC GAC TAC TCT CTG ACC ATC TCT AAC CTG GAA CAG GAA GAC ATC GCT ACC TAC TTC TGC CAG CAG GGT AAC ACC CTG CCG CCG ACC TTC GGT GCT GGT ACC AAA CTG GAA CTG AAA TAA GCG GCC GCT

Amino acid sequence (260 aa)

AMAQVQLQQS GAELAKPGAS MKMSCRASGY SFTSYWIHWL KQRPDQGLEW IGYIDPATAY TESNQKFKDK AILTADRSSN TAFMYLNSLT SEDSAVYYCA RESPRLRRGI YYYAMDYWGQ GTTVTVSSKL APQAKSSGSG SESKVDARVD IQMTQTPSSL SASLGDRVTI SCRASQDISN YLNWYQQKPD GTVKLLIYYT SRLHSGVPSR FSGSGSGTDY SLTISNLEQE DIATYFCQQG NTLPPTFGAG TKLELK*AAA

scFv 14F7 C4*

DNA sequence of (780bp)

GCC ATG GCC CAG GTG CAG CTG CAG CAG AGC GGC GCG GAA CTG GCG AAA CCG GGC GCG AGC ATG AAA ATG AGC TGC CGC GCG AGC GGC TAT AGC TTT ACC AGC TAT TGG ATT CAT TGG CTG AAA CAG CGC CCG GAT CAG GGC CTG GAA TGG ATT GGC TAT ATT GAT CCG GCG ACC GCG TAT ACC GAA AGC AAC CAG AAA TTT AAA GAT AAA GCG ATT CTG ACC GCG GAT CGC AGC AGC AAC ACC GCG TTT ATG TAT CTG AAC AGC CTG ACC AGC GAA GAT AGC GCG GTG TAT TAT TGC GCG CGC GAA AGC CCG CGC CTG CGC CGC GGC ATT TAT TAT TAT GCG ATG GAT TAT TGG GGC CAG GGC ACC ACC GTG ACC GTG AGC AGC AAG CTT GCG CCG CAG GCG AAA AGC AGC GGC AGC GGC AGC GAA AGC AAA GTG GAT GCA CGC GTA GAC CTG GTT CTG ACC CAG TCT CCG GCT ACC CTG TCT GTT ACC CCA GGT GAC TCT GTT TCT TTC TCT TGC CGT GCT TCT CAG TCT ATC TCT AAC AAC CTG CAC TGG TAC CAG CAG CGT ACC CAC GAA TCT CCG CGT CTG CTG ATC AAA TAC GCT TCT CAG TCT ATC TCT GGT ATC CCG TCT CGT TTC TCT GGT TCT GGT TCT GGT ACC GAC TTC ACC CTG TCT ATC TCT TCT GTT GAA ACC GAA GAC TTC GGT ATG TAC TTC TGC CAG CAG TCT AAC CGT TGG CCG CTG ACC TTC GGT GCT GGT ACC AAA CTG GAA CTG AAA TAA GCG GCC GCT

Amino acid sequence (261 aa)

AMAQVQLQQS GAELAKPGAS MKMSCRASGY SFTSYWIHWL KQRPDQGLEW IGYIDPATAY TESNQKFKDK AILTADRSSN TAFMYLNSLT SEDSAVYYCA RESPRLRRGI YYYAMDYWGQ GTTVTVSSKL APQAKSSGSG SESKVDARVD LVLTQSPATL SVTPGDSVSF SCRASQSISN NLHWYQQRTH ESPRLLIKYA SQSISGIPSR FSGSGSGTDF TLSISSVETE DFGMYFCQQS NRWPLTFGAG TKLELK*AAA

scFv 4G9

DNA sequence (783 bp)

GCC ATG GCC CAG GTG CAG CTG AAA GAA AGC GGC CCG GGC CTG GTG GCG CCG AGC CAG AGC CTG AGC ATT ACC TGC ACC GTG AGC GGC TTT AGC CTG AGC AAC TAT GGC GTG CAT TGG GTG CGC CAG CCG CCG GGC AAA GGC CTG GAA TGG CTG GGC GAA ATT TGG GCG GGC GGC AGC ACC AAC TAT AAC AGC GCG CTG ATG AGC CGC CTG AGC ATT AGC AAA GAT AAC AGC AAA AGC CAG GTG TTT CTG AAA ATG AAC AGC CTG CAG ACC GAT GAT ACC GCG ATG TAT TAT TGC GCG CGC GCG TAT GAT TAT GAT GGC GCG TGG TTT GCG TAT TGG GGC CAG GGC ACC CTG GTG ACC GTG AGC GCG AAG CTT TCA GGG AGT GCA TCC GCC CCA AAA CTT GAA GAA GGT GAA TTT TCA GAA GCA CGC GTA GAT ATT CAG ATG ACC CAG ACC ACC AGC AGC CTG AGC GCG AGC CTG GGC GAT CGC GTG ACC ATT AGC TGC CGC GCG AGC CAG GAT ATT AGC AAC TAT CTG AAC TGG TAT CAG CAG AAA CCG GAT GGC ACC GTG AAA CTG CTG ATT TAT TAT ACC AGC CGC CTG CAT AGC GGC GTG CCG AGC CGC TTT AGC GGC AGC GGC AGC GGC ACC GAT TAT AGC CTG ACC ATT AGC AAC CTG GAA CAG GAA GAT ATT GCG ACC TAT TTT TGC CAG CAG GGC GAT ACC CTG CCG TAT ACC TTT GGC GGC GGC ACC AAA CTG GAA CTG AAA GCG CAC CAC CAC CAC CAC CAC TAA GCG GCC GCT

Amino acid sequence (261 aa)

AMAQVQLKES GPGLVAPSQS LSITCTVSGF SLSNYGVHWV RQPPGKGLEW LGEIWAGGST NYNSALMSRL SISKDNSKSQ VFLKMNSLQT DDTAMYYCAR AYDYDGAWFA YWGQGTLVTV SAKLSGSASA PKLEEGEFSE ARVDIQMTQT TSSLSASLGD RVTISCRASQ DISNYLNWYQ QKPDGTVKLL IYYTSRLHSG VPSRFSGSGS GTDYSLTISN LEQEDIATYF CQQGDTLPYT FGGGTKLELK AHHHHHH*AA A

scFv 14F7 C1#

DNA sequence (813 bp)

ACT TCC ATG GCG CAG GTG CAG CTG CAG CAG AGC GGC AAC GAA CTG GCG AAA CCG GGC GCG AGC ATG AAA ATG AGC TGC CGC GCG AGC GGC TAT AGC TTT ACC AGC TAT TGG ATT CAT TGG CTG AAA CAG CGC CCG GAT CAG GGC CTG GAA TGG ATT GGC TAT ATT GAT CCG GCG ACC GCG TAT ACC GAA AGC AAC CAG AAA TTT AAA GAT AAA GCG ATT CTG ACC GCG GAT CGC AGC AGC AAC ACC GCG TTT ATG TAT CTG AAC AGC CTG ACC AGC GAA GAT AGC GCG GTG TAT TAT TGC GCG CGC GAA AGC CCG CGC CTG CGC CGC GGC ATT TAT TAT TAT GCG ATG GAT TAT TGG GGC CAG GGC ACC AGC GTG ATT GTG AGC AGC AAG CTT AGC GGC AGC GCG AGC GCG CCG AAA CTG GAA GAA GGC GAA TTT AGC GAA GCA CGC GTG GAT ATT CAG ATG ACC CAG ACC CCG AGC AGC CTG AGC GCG AGC CTG GGC GAT CGC GTG ACC ATT AGC TGC CGC GCG AGC CAG GAT ATT AGC AAC TAT CTG AAC TGG TAT CAG CAG AAA CCG GAT GGC ACC GTG AAA CTG CTG ATT TAT TAT ACC AGC CGC CTG CAT AGC GGC GTG CCG AGC CGC TTT AGC GGC AGC GGC AGC GGC ACC GAT TAT AGC CTG ACC ATT AGC AAC CTG GAA CAG GAA GAT ATT GCG ACC TAT TTT TGC CAG CAG GGC AAC ACC CTG CCG CCG ACC TTT GGC GCG GGC ACC AAA CTG GAA CTG CTG TAT TTT CAG AGC CAC CAT CAC CAT CAC CAT TAA GCG GCC GCT

Amino acid sequence (271 aa)

TSMAQVQLQQ SGNELAKPGA SMKMSCRASG YSFTSYWIHW LKQRPDQGLE WIGYIDPATA YTESNQKFKD KAILTADRSS NTAFMYLNSL TSEDSAVYYC ARESPRLRRG IYYYAMDYWG QGTSVIVSSK LSGSASAPKL EEGEFSEARV DIQMTQTPSS LSASLGDRVT ISCRASQDIS NYLNWYQQKP DGTVKLLIYY TSRLHSGVPS RFSGSGSGTD YSLTISNLEQ EDIATYFCQQ GNTLPPTFGA GTKLELLYFQ SHHHHHH*AA A

scFv 14F7 C2#

DNA sequence (813 bp)

ACT TCC ATG GCG CAG GTG CAG CTG CAG CAG AGC GGC AAC GAA CTG GCG AAA CCG GGC GCG AGC ATG AAA ATG AGC TGC CGC GCG AGC GGC TAT AGC TTT ACC AGC TAT TGG ATT CAT TGG CTG AAA CAG CGC CCG GAT CAG GGC CTG GAA TGG ATT GGC TAT ATT GAT CCG GCG ACC GCG TAT ACC GAA AGC AAC CAG AAA TTT AAA GAT AAA GCG ATT CTG ACC GCG GAT CGC AGC AGC AAC ACC GCG TTT ATG TAT CTG AAC AGC CTG ACC AGC GAA GAT AGC GCG GTG TAT TAT TGC GCG CGC GAA AGC CCG CGC CTG CGC CGC GGC ATT TAT TAT TAT GCG ATG GAT TAT TGG GGC CAG GGC ACC AGC GTG ATT GTG AGC AGC AAG CTT AGC GGC AGC GCG AGC GCG CCG AAA CTG GAA GAA GGC GAA TTT AGC GAA GCA CGC GTG GAT CTG GTG CTG ACC CAG AGC CCG GCG ACC CTG AGC GTG ACC CCG GGC GAT AGC GTG AGC TTT AGC TGC CGC GCG AGC CAG AGC ATT AGC AAC AAC CTG CAT TGG TAT CAG CAG CGC ACC CAT GAA AGC CCG CGC CTG CTG ATT AAA TAT GCG AGC CAG AGC ATT AGC GGC ATT CCG AGC CGC TTT AGC GGC AGC GGC AGC GGC ACC GAT TTT ACC CTG AGC ATT AGC AGC GTG GAA ACC GAA GAT TTT GGC ATG TAT TTT TGC CAG CAG AGC AAC CGC TGG CCG CTG ACC TTT GGC GCG GGC ACC AAA CTG GAA CTG CTG TAT TTT CAG AGC CAC CAT CAC CAT CAC CAT TAA GCG GCC GCT

Amino acid sequence (271 aa)

TSMAQVQLQQ SGNELAKPGA SMKMSCRASG YSFTSYWIHW LKQRPDQGLE WIGYIDPATA YTESNQKFKD KAILTADRSS NTAFMYLNSL TSEDSAVYYC ARESPRLRRG IYYYAMDYWG QGTSVIVSSK LSGSASAPKL EEGEFSEARV DLVLTQSPAT LSVTPGDSVS FSCRASQSIS NNLHWYQQRT HESPRLLIKY ASQSISGIPS RFSGSGSGTD FTLSISSVET EDFGMYFCQQ SNRWPLTFGA GTKLELLYFQ SHHHHHH*AA A

scFv 14F7 C3#

DNA sequence (813 bp)

ACT TCC ATG GCG CAG GTG CAG CTG CAG CAG AGC GGC AAC GAA CTG GCG AAA CCG GGC GCG AGC ATG AAA ATG AGC TGC CGC GCG AGC GGC TAT AGC TTT ACC AGC TAT TGG ATT CAT TGG CTG AAA CAG CGC CCG GAT CAG GGC CTG GAA TGG ATT GGC TAT ATT GAT CCG GCG ACC GCG TAT ACC GAA AGC AAC CAG AAA TTT AAA GAT AAA GCG ATT CTG ACC GCG GAT CGC AGC AGC AAC ACC GCG TTT ATG TAT CTG AAC AGC CTG ACC AGC GAA GAT AGC GCG GTG TAT TAT TGC GCG CGC GAA AGC CCG CGC CTG CGC CGC GGC ATT TAT TAT TAT GCG ATG GAT TAT TGG GGC CAG GGC ACC AGC GTG ATT GTG AGC AGC AAG CTT GCG CCG CAG GCG AAA AGC AGC GGC AGC GGC AGC GAA AGC AAA GTG GAT GCA CGC GTG GAT ATT CAG ATG ACC CAG ACC CCG AGC AGC CTG AGC GCG AGC CTG GGC GAT CGC GTG ACC ATT AGC TGC CGC GCG AGC CAG GAT ATT AGC AAC TAT CTG AAC TGG TAT CAG CAG AAA CCG GAT GGC ACC GTG AAA CTG CTG ATT TAT TAT ACC AGC CGC CTG CAT AGC GGC GTG CCG AGC CGC TTT AGC GGC AGC GGC AGC GGC ACC GAT TAT AGC CTG ACC ATT AGC AAC CTG GAA CAG GAA GAT ATT GCG ACC TAT TTT TGC CAG CAG GGC AAC ACC CTG CCG CCG ACC TTT GGC GCG GGC ACC AAA CTG GAA CTG CTG TAT TTT CAG AGC CAC CAT CAC CAT CAC CAT TAA GCG GCC GCT

Amino acid sequence (271 aa)

TSMAQVQLQQ SGNELAKPGA SMKMSCRASG YSFTSYWIHW LKQRPDQGLE WIGYIDPATA YTESNQKFKD KAILTADRSS NTAFMYLNSL TSEDSAVYYC ARESPRLRRG IYYYAMDYWG QGTSVIVSSK LAPQAKSSGS GSESKVDARV DIQMTQTPSS LSASLGDRVT ISCRASQDIS NYLNWYQQKP DGTVKLLIYY TSRLHSGVPS RFSGSGSGTD YSLTISNLEQ EDIATYFCQQ GNTLPPTFGA GTKLELLYFQ SHHHHHH*AA A

scFv 14F7 C4#

DNA sequence (813 bp)

ACT TCC ATG GCG CAG GTG CAG CTG CAG CAG AGC GGC AAC GAA CTG GCG AAA CCG GGC GCG AGC ATG AAA ATG AGC TGC CGC GCG AGC GGC TAT AGC TTT ACC AGC TAT TGG ATT CAT TGG CTG AAA CAG CGC CCG GAT CAG GGC CTG GAA TGG ATT GGC TAT ATT GAT CCG GCG ACC GCG TAT ACC GAA AGC AAC CAG AAA TTT AAA GAT AAA GCG ATT CTG ACC GCG GAT CGC AGC AGC AAC ACC GCG TTT ATG TAT CTG AAC AGC CTG ACC AGC GAA GAT AGC GCG GTG TAT TAT TGC GCG CGC GAA AGC CCG CGC CTG CGC CGC GGC ATT TAT TAT TAT GCG ATG GAT TAT TGG GGC CAG GGC ACC AGC GTG ATT GTG AGC AGC AAG CTT GCG CCG CAG GCG AAA AGC AGC GGC AGC GGC AGC GAA AGC AAA GTG GAT GCA CGC GTG GAT CTG GTG CTG ACC CAG AGC CCG GCG ACC CTG AGC GTG ACC CCG GGC GAT AGC GTG AGC TTT AGC TGC CGC GCG AGC CAG AGC ATT AGC AAC AAC CTG CAT TGG TAT CAG CAG CGC ACC CAT GAA AGC CCG CGC CTG CTG ATT AAA TAT GCG AGC CAG AGC ATT AGC GGC ATT CCG AGC CGC TTT AGC GGC AGC GGC AGC GGC ACC GAT TTT ACC CTG AGC ATT AGC AGC GTG GAA ACC GAA GAT TTT GGC ATG TAT TTT TGC CAG CAG AGC AAC CGC TGG CCG CTG ACC TTT GGC GCG GGC ACC AAA CTG GAA CTG CTG TAT TTT CAG AGC CAC CAT CAC CAT CAC CAT TAA GCG GCC GCT

Amino acid sequence (271 aa)

TSMAQVQLQQ SGNELAKPGA SMKMSCRASG YSFTSYWIHW LKQRPDQGLE WIGYIDPATA YTESNQKFKD KAILTADRSS NTAFMYLNSL TSEDSAVYYC ARESPRLRRG IYYYAMDYWG QGTSVIVSSK LAPQAKSSGS GSESKVDARV DLVLTQSPAT LSVTPGDSVS FSCRASQSIS NNLHWYQQRT HESPRLLIKY ASQSISGIPS RFSGSGSGTD FTLSISSVET EDFGMYFCQQ SNRWPLTFGA GTKLELLYFQ SHHHHHH*AA A

scFv 14F7 C1

DNA sequence (819 bp)

GCC ATG GCC CAC CAC CAC CAC CAC CAC GAA AAC CTG TAC TTC CAG GGT
CAG GTG CAG CTG CAG CAG AGC GGC GCG GAA CTG GCG AAA CCG GGC GCG
AGC ATG AAA ATG AGC TGC CGC GCG AGC GGC TAT AGC TTT ACC AGC TAT
TGG ATT CAT TGG CTG AAA CAG CGC CCG GAT CAG GGC CTG GAA TGG ATT
GGC TAT ATT GAT CCG GCG ACC GCG TAT ACC GAA AGC AAC CAG AAA TTT
AAA GAT AAA GCG ATT CTG ACC GCG GAT CGC AGC AGC AAC ACC GCG TTT
ATG TAT CTG AAC AGC CTG ACC AGC GAA GAT AGC GCG GTG TAT TAT TGC
GCG CGC GAA AGC CCG CGC CTG CGC CGC GGC ATT TAT TAT TAT GCG ATG
GAT TAT TGG GGC CAG GGC ACC ACC GTG ACC GTG AGC AGC AAG CTT TCA
GGG AGT GCA TCC GCC CCA AAA CTT GAA GAA GGT GAA TTT TCA GAA GCA
CGC GTA GAC ATC CAG ATG ACC CAG ACC CCG TCT TCT CTG TCT GCT TCT CTG
GGT GAC CGT GTT ACC ATC TCT TGC CGT GCT TCT CAG GAC ATC TCT AAC TAC
CTG AAC TGG TAC CAG CAG AAA CCG GAC GGT ACC GTT AAA CTG CTG ATC
TAC TAC ACC TCT CGT CTG CAC TCT GGT GTT CCG TCT CGT TTC TCT GGT TCT
GGT TCT GGT ACC GAC TAC TCT CTG ACC ATC TCT AAC CTG GAA CAG GAA
GAC ATC GCT ACC TAC TTC TGC CAG CAG GGT AAC ACC CTG CCG CCG ACC
TTC GGT GCT GGT ACC AAA CTG GAA CTG AAA TAA GCG GCC GCT

Amino acid sequence (273 aa)

AMAHHHHHHENLYFQGQVQLQQSGAELAKPGASMKMSCRASGYSFTSYWIHWLK
QRPDQGLEWIGYIDPATAYTESNQKFKDKAILTADRSSNTAFMYLNSLTSEDSAVYY
CARESPRLRRGIYYYAMDYWGQGTTVTVSSKLSGSASAPKLEEGEFSEARVDIQMTQ
TPSSLSASLGDRVTISCRASQDISNYLNWYQQKPDGTVKLLIYYTSRLHSGVPSRFSGS
GSGTDYSLTISNLEQEDIATYFCQQGNTLPPTFGAGTKLELK*AAA