The development of antibodies against tumour-associated carbohydrate antigens

Inaugural Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of the Freie Universität Berlin
by

Sana Khan Khilji

2023

This work was performed between July 2018 and February 2023 under the supervision of Dr Oren Moscovitz in the Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Potsdam, Germany.

1st reviewer: Dr Oren Moscovitz (Max Planck Institute of Colloids and Interfaces, Potsdam)

2nd reviewer: Prof. Dr Helge Ewers (Freie Universität Berlin)

Date of oral defence: 07.07.2023

Acknowledgements

I am a different person to the one who entered this PhD journey. I have fallen, cried, and doubted myself, and then picked myself up, wiped away tears, and carried on, achieving the goals I had set for myself. Reflecting on this time, I am reminded of something my 10th class English teacher once said to me, “Your best quality is that you persevere”. Indeed, I persevered. It would be imprudent of me though if I did not acknowledge the ‘village’ of wonderful supportive people without whom this journey would have been impossible to start, let alone finish.

First and foremost, I would like to thank my supervisor, Dr Oren Moscovitz for taking a chance on me and for his guidance throughout my time in his group. I’ve learnt a lot from you. I also really appreciate your support and understanding through the last year, which continued to motivate me to write and complete the dissertation.

I would also like to thank Prof. Dr Dr Peter Seeberger and Oren for the opportunity to work in Max Planck Institute of Colloids and Interfaces and to get to know all the amazing colleagues there. Furthermore, I would like to take the opportunity to thank Prof. Dr Helge Ewers for agreeing to be my second reviewer.

Also, a thank you to my collaborators: Dr Alvaro Mallagaray for the STD-NMR experiments, Janine Arndt and Dr med. Carsten Kamphues for the IHC experiments, and Michael Krummhaar and Dr Christian Roth for the crystallography experiments. Moreover, I would like to thank Sutapa Chakrabarti (PhD) and Dr Jens Dernedde for being part of my informal thesis advisory committee. Sutapa, thank you also for your support and guidance over the years. It means a lot!

The ‘Dahlem crew’ is a bunch of wonderful people who were as much a part of my journey the last few years as I was of theirs. Dr Christian Roth, Michael Krummhaar, Ruben Ananian, Dr Paulina Kapłonek, Dr Jonnel Jaurigue, Dr Bruna Seco, Dr Ling Yao, Patricia Priegue, Fabienne Weber, Emelie Reuber, Magdalena Zaslona, Dr Felix Goerdeler, Anika Freitag, Jost Lühle, Charlotte Op’t Hoog, Kristin Frensemeier, Sophia Borowski, and my students, David Warschkau, Zeinab Fandi, and Sophia Schrinner, I miss our mealtimes, coffee breaks, pumpkin days, and secret Santa parties. Thank you for making coming to work enjoyable. We lived through the pandemic and survived together. Magda, thank you for my bridal shower, welcome back from the wedding, and farewell as well as for the wonderful times together. And thank you Kristin,

Anika, David, Zeinab, and Sophia for your contributions to this work. I would also like to thank Paul Meißner for the technical support and Katrin Sellrie for her work in establishing the hybridomas and printing the glycan array slides.

Christian, thank you for your steadfast support and your ear whenever I needed to rant. I could not have managed parts of my project without your guidance with the crystallisation experiments as well as many other things. And thank you for proofreading my thesis.

Felix, I could not have asked for a better ‘comrade in arms’. You made working side-by-side fun, and I loved our talks about anything and everything. I miss those the most! Thank you for your constant encouragement and thank you for proofreading my thesis as well as translating the abstract to German.

Anika, remember my sweatshirt that said, “I love Fridays”? One could call that a prediction! No matter what you were going through on your own, you were always there for me. You listen, and that means a lot! Thank you for your steadfast support and motivation.

Mama and Baba, I wouldn’t be standing where I am today without you two. You are my rocks and I feel I can conquer the world with you two behind me. Thank you for always being there for me, for unconditionally supporting me, and encouraging me to reach for the stars. You are my heroes.

Uzair, thank you for bearing with me, day after day! You have been unconditionally supportive and loving, and I could not have achieved this without you. Even though living apart was not easy, I will always look back at the last few years with fondness because it was when we found each other and got married. I look forward to more adventures with you in our future.

I would also like to give a shout out to all my family and friends, including my parents-in-law, Amma and Papa, siblings, Nauman and Kiran, and siblings-in-law, Asad, Awais, and Manahil, who always encourage me to move forward and never give up. I am blessed to have you all in my life.

Finally, I dedicate this thesis to Mama, Baba, and Uzair.

Declaration

Hierdurch versichere ich, dass ich meine Dissertation selbstständig verfasst und keine anderen als die von mir angegeben Quellen und Hilfsmittel verwendet habe. Ich bestätige, dass die elektronische Kopie den gedruckten Kopien gleicht.

This work has not previously been submitted for any degree or professional qualification. This thesis is the result of my own work, except where otherwise stated. All sources have been acknowledged through explicit references in the text and a Bibliography is appended.

I confirm that the electronic copy is identical to the bound copies of the thesis.

Berlin, 27 February 2023

…….

Sana Khan Khilji

Table of Contents

List of abbreviations

α1,4GalT	α1,4-galactosyltransferase
β1,3GalT-V	β1,3- galactosyltransferase V
Ab	antibody
AAb	autoantibody
ADC	Ab-drug conjugate
ADCC	Ab-dependent cellular cytotoxicity
ADCP	Ab-dependent cellular phagocytosis
AF	amplification factors
APC	allophycocyanin
APS	ammonium persulphate
BCG	Bacille Calmette-Guerin
BRCA1	breast cancer 1 protein
C	constant domain
CA	carbohydrate antigens
CAR	chimeric antigen receptor
Casp	caspase
CAV1	caveolin-1
CDC	complement-dependent cytotoxicity
cDNA	Complementary DNA
CDR	complementarity determining region
CEA	carcinoembryonic antigen
Cer	ceramide
Ch	chimeric
CK5/6	Cytokeratin 5 and 6
CRC	colorectal carcinoma
CRM_{197}	cross-reactive material 197
CT	computed tomography
CTC	circulating tumour cell
ctDNA	Circulating tumour DNA
CV	column volumes
DAPI	4’,6-Diamidino-2-phenylindole
DMEM	Dulbecco’s modified Eagle’s medium
DNA	Deoxyribonucleic acid
DSF	differential scanning fluorimetry
DSS-d_6	3-(trimethylsilyl)-1-propanesulfonic acid-d_6
DTT	Dithiothreitol
E-	Endothelial-
E. coli	*Escherichia coli*
EGFR	epidermal growth factor receptor
EMT	epithelial to mesenchymal transition
EpCAM	epithelial cell adhesion molecule

ER	oestrogen receptor
ERCP	endoscopic retrograde cholangiopancreatography
EUS	endoscopic ultrasound
EV	Extracellular vesicle
Fab	antigen binding fragment
FADD	Fas death domain
FAK	focal adhesion kinase
F_c	fragment crystallisable
F_cR	fragment crystallisable receptor
FCS	Foetal calf serum
FDA	U.S. food and drug administration
FITC	fluorescein isothiocyanate
FPLC	Fast Protein Liquid Chromatography
FR	framework regions
FSC-A	forward scatter amplitude
FSC-H	FSC-Height
FSC-W	FSC-Width
Fuc	Fucose
ΔG	change in free Gibbs energy
Gal	Galactose
GalNAc	N-acetylgalactosamine
GFP	green fluorescence protein
Glc	Glucose
GlcA	Glucoronic acid
GlcNAc	N-acetylglucosamine
GSH	Gluthatione
GSL	Glycosphingolipids
ΔH	change in enthalpy
H	Heavy (chain)
hcAb	heavy-chain only Abs
HER2	human epidermal growth receptor two
hESC	human embryonic stem cell
Hip-DD	Hsp70-interacting protein dimerization domain
H. pylori	*Helicobacter pylori*
HSP60	heat shock chaperone protein
IgG	immunoglobulin G
IHC	Immunohistochemistry
IPTG	isopropyl-β-D-thiogalactoside
ITC	Isothermal Calorimetry
K_D	apparent affinity
KLH	keyhole limpet hemocyanin
L	Light (chain)
Lac-Cer	lactosylceramide
LB	lysogeny broth

Le A	Lewis A
Le B	Lewis B
Le X	Lewis X
Le Y	Lewis Y
(lnc)RNA	long non-coding RNA
2-ME	2-mercaptoethanol
Man	Mannose
MAPK	mitogen-activated protein kinase
MFI	mean fluorescence intensity
MHC	major histocompatibility complex
(m)RNA	messenger RNA
miRNA	MicroRNA
MRI	magnetic resonance imaging
MST	Microscale Thermophoresis
MWCO	molecular weight cut-off
N	binding sites
Nb	Nanobody
Neu5Ac	N-acetylneuraminic acid
Neu5Gc	N-glycolylneuraminic acid
NK	Natural killer cells
NOE	Nuclear Overhauser Effect
NMR	Nuclear magnetic resonance
OD_{600}	Optical density at 600 nm
P-	Platelet-
PBMC	peripheral blood mononuclear cell
PBS	phosphate buffered saline
PBST	PBS + 0.05% Tween-20
PC	phosphatidylcholine
PCR	Polymerase chain reaction
PDAC	pancreatic ductal adenocarcinoma
PDB	Protein Data Bank
PE	phosphatidylethanolamine
PET	Positron emission tomography
PR	progesterone receptor
PTM	post-translational modification
RIA	radioimmunoassay
RIP	receptor-interacting protein kinase
RNA	Ribonucleic acid
RU	response units
ΔS	change in entropy
scFv	single-chain variable fragment
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SEM	standard error of mean

Siglecs	sialic acid-binding immunoglobulin-like lectins
sLe A	sialyl Lewis A
sLe X	sialyl Lewis X
SLPG	sialosyl lactosaminyl paragloboside
SPECT	single-photon emission computed tomography
SPR	Surface Plasmon Resonance
Srt	sortase-tag
SSC-A	sideward scatter amplitude
SSEA	Site specific embryonic antigen
ST3GAL-II	α2,3-sialyltransferase 2
STD-NMR	Saturation transfer difference nuclear magnetic resonance
sTF	sialyl Thomsen–Friedenreich antigen
sTn	sialyl Thomson nouveau
sTRA	sialylated tumour-related antigen
T	absolute temperature
TAA	Tumour-associated antigens
TACA	tumour-associated carbohydrate antigens
TB	terrific broth
TEMED	tetramethylethylenediamine
TF	Thomsen–Friedenreich antigen
TME	tumour microenvironment
T_M	melting temperature
Tn	Thomsen-nouvelle antigen
TRAX	translin-associated factor X
trNOE	transferred Nuclear Overhauser Effect
T-VEC	Herpes Simplex Virus type 1- based talimogene laherparepvec
V	variable region
VEGF	vascular endothelial growth factor
WHO	world health organisation
Xyl	Xylose
κ	kappa
λ	lambda

List of figures

List of tables

“Cancer is one of the world’s leading causes of death, and its burden is growing. In 2021, the world crossed a sobering new threshold – an estimated 20 million people were diagnosed with cancer, and 10 million died. These numbers will continue to rise in the decades ahead. And yet all cancers can be treated, and many can be prevented or cured.”

–‘World Cancer Day: closing the care gap’,

World Health Organization (WHO), 3 February 2022.

Abstract

Cancer is the second leading cause of death for humans around the world. Aberrant cell surface glycosylation is a hallmark for cancer and can be used for targeted cancer therapy. Two such aberrant tumour-associated carbohydrate antigens (TACAs) are Globo-H and sialyl Lewis A (sLe A). Globo-H is expressed in various epithelial cancers, such as breast cancer, the highest incidence cancer in women. Its expression leads to cancer progression through angiogenesis, immunosuppression, and tumour survival. On the other hand, sLe A is so far the only FDA approved biomarker for pancreatic cancer that is often only detected at a later stage and, therefore, has a high mortality rate. Sialyl Le A has also been implicated in metastasis, tumour-associated inflammation, and immune evasion. Therefore, both TACAs are important targets, but the number of tools available to study them are limited.

Synthetic, pure, and well-defined glycans were used to obtain antibodies (Abs) and nanobodies (Nbs) against sLe A and Globo-H, respectively. I demonstrated through various methods that Abs and Nbs generated through this technique are highly specific for their targets both *in vitro* with synthetic glycans as well as *in vivo* with native glycans.

For the first part of my thesis, I studied sLe A-targeting Abs, GB11 and HA8. I compared their amino acid sequences to that of a commercial Ab, 1116-NS-19-9, clinically used to target sLe A and diagnose pancreatic cancer. I compared the binding of the three Abs to sLe A and obtained a crystal structure for GB11 to unravel the molecular origin of its specificity and higher affinity, compared to the commercial counterpart, 1116-NS19-9.

For the second part of my thesis, I showed that a Nb, GH46, is the best binder for Globo-H in a previously generated Nb library. I deduced its binding epitope using different methods and generated a multivalent construct of GH46 by successfully cloning and expressing GH46-trimer. The trimer has better binding to native Globo-H on cells than GH46-monomer. Furthermore, I used a site-directed mutagenesis strategy to improve upon solubility, thermostability, and affinity of GH46. I demonstrated that the framework region 3 of the Nb plays an important role in Nb stability. Through this, I generated a GH46 mutant which retains specificity for Globo-

H but gives 100-fold higher yield, 7-fold better binding affinity *in vitro*, and 2-fold better binding *in vivo*.

In summary, I show with this thesis how the immunisation with synthetic glycans can generate highly specific Abs and Nbs, which could be used in biomedical and biotechnological applications in the future particularly in the framework of cancer therapy. As proof of concept, two new Abs targeting sLe A were introduced and characterised that could potentially be used for early diagnosis of pancreatic cancer. Additionally, the first in class Nb binding to a TACA was characterised and functionalised to a multimer. The Nb structure was studied in depth and may add to the current knowledge available on structural stability of Nbs.

Zusammenfassung

Krebs ist weltweit die zweithäufigste Todesursache für Menschen. Veränderte Zelloberflächenglykosylierung ist ein Kennzeichen für Krebs und kann zur zielgerichteten Krebstherapie eingesetzt werden. Zwei solcher veränderter tumorassoziierter Kohlenhydratantigene (*tumour-associated carbohydrate antigens*, TACAs) sind Globo-H und Sialyl-Lewis A (sLe A). Globo-H ist in verschiedenen epithelialen Krebsarten exprimiert – zum Beispiel in Brustkrebs, der Krebsart mit der höchsten Inzidenz bei Frauen. Die Expression von Globo-H trägt maßgeblich zur Angiogenese, Immunsuppression und zum Überleben des Tumors bei, was zur Progredienz der Krankheit führt.
SLe A ist der bisher einzige von der FDA zugelassene Biomarker für Bauchspeicheldrüsenkrebs, eine Krebsart, die oft erst in einem fortgeschrittenen Stadium entdeckt wird und dadurch eine hohe Mortalitätsrate aufweist. Die Expression von sLe A steht auch im Zusammenhang mit der Bildung von Metastasen, tumorassoziierten Entzündungsreaktionen und Immunevasion. Deshalb sind beide TACAs wichtige Ziele für die Krebstherapie, aber die Zahl der Werkzeuge zu ihrer Untersuchung ist begrenzt.

Synthetische, reine und definierte Glykane wurden eingesetzt, um gegen sLe A gerichtete Antikörper (Antibodies, Abs) und gegen Globo-H gerichtete Nanokörper (Nanobodies, Nbs) zu erhalten. Ich konnte durch verschiedene Experimente zeigen, dass die mit dieser Methode generierten Abs und Nbs eine hohe Spezifität für ihre Zielmoleküle aufweisen, sowohl *in vitro* mit synthetischen Glykanen als auch *in vivo* mit nativen Glykanen.

Im ersten Teil meiner Doktorarbeit untersuchte ich die anti-sLe A Abs GB11 und HA8. Ich verglich ihre Aminosäuresequenzen mit der Sequenz des kommerziellen Abs 1116-NS-19-9, welcher ebenfalls gegen sLeA gerichtet ist und zur klinischen Diagnose von Bauchspeicheldrüsenkrebs verwendet wird. Ich verglich die Bindung der drei Antikörper an sLe A und erhielt eine Kristallstruktur von GB11, um den molekularen Ursprung aufzudecken für die Spezifität und höhere Affinität von GB11 im Vergleich zum kommerziellen Gegenstück 1116-NS-19-9.

Im zweiten Teil meiner Doktorarbeit zeigte ich, dass von allen Nbs aus einer zuvor generierten Sequenzbibliothek Nb GH46 die beste Bindung zu Globo-H zeigt. Ich ermittelte das Bindungsepitop von GH46 unter Verwendung verschiedener Methoden und generierte ein multivalentes Konstrukt von GH46 durch erfolgreiche Klonierung und Expression eines GH46-Trimers. Das Trimer verfügte über eine bessere Bindung zu nativem Globo-H auf Zellen im Vergleich zum GH46-Monomer. Des Weiteren setzte ich eine ortsgerichtete Strategie zur Mutagenese ein, um die Löslichkeit, Thermostabilität und Affinität von GH46 zu verbessern. Ich demonstrierte dadurch, dass die *Framework*-Region 3 eine wichtige Rolle für die Stabilität des Nb spielt. Hierbei generierte ich eine GH46-Mutante, die die Spezifität für Globo-H beibehält, aber mit einer 100-mal höheren Ausbeute gewonnen werden kann und sich durch eine siebenfach verbesserte Bindungsaffinität *in vitro* sowie zweifach verbesserte Bindung *in vivo* auszeichnet.

Insgesamt zeige ich mit meiner Doktorarbeit, wie die Immunisierung mit synthetischen Glykanen hochspezifischen Abs und Nbs generiert, welche für zukünftige biomedizinische und biotechnologische Anwendungen im Rahmen der Krebstherapie eingesetzt werden können. Zum Nachweis der Machbarkeit wurden zwei neue sLe A-bindende Abs vorgestellt und charakterisiert, die potenzielle Verwendung in der rechtzeitigen Diagnose von Bauchspeicheldrüsenkrebs finden könnten. Außerdem wurde der erste TACA-bindende Nb charakterisiert und als Multimer funktionalisiert. Die Struktur des Nb wurde detailliert untersucht und ergänzt das bestehende verfügbare Wissen über die Stabilität von Nbs.

1. Introduction

1.1. Cancer and pitfalls of cancer therapy

Cancer is an aberration of the genome, causing unrestricted proliferation of cells, which can invade neighbouring tissue as well as metastasize to other parts of the body. While it can be hereditary, it is more often caused by epigenetic factors[1]. These include age, smoking of tobacco products, infectious agents such as the gram-negative bacterium *Helicobacter pylori* (*H. pylori*), Human Pappilomavirus (HPV), and Hepatitis B virus, alcohol consumption, sunlight and ultraviolet radiation, diet, and obesity, among others[1]. Cancer counts for one of the major causes for global mortality and, according to WHO, was responsible for almost 10 million deaths in 2021[2]. It has even been projected to surpass infectious as well as other non-communicable chronic diseases as the leading cause of death in every country during the twenty-first century[3,4].

Genetic alterations lead to a loss of normal regulatory cellular processes and result in the expression of neoantigens, differentiation antigens, or cancer testis antigens[5,6]. In an ideal scenario, the immune system initiates a series of events that take place iteratively to neutralise the threat that is cancer. The immune system usually acts upon captured antigens that will be displayed on major histocompatibility complex (MHC) I and MHCII molecules and lead to activation of effector T cells. Effector T cells would then infiltrate the tumour microenvironment (TME), interact with T cell receptors and the antigen peptide bound to MHCI molecules, consequently killing their target. During this process the dying threat would release additional antigens to reinitiate the cycle. Furthermore, this type of active immunity is expected to activate B-cells in a T-cell-mediated mechanism. B-cell activation would then lead to cell proliferation to produce targeted and highly specific antibodies (Abs; see **section 1.3** for detail). Alternatively, recombinant Abs could be used to augment the active immune system, which should behave in the same way as the endogenous Abs (**Figure 1**)[7,8]. This is known as passive immunisation, in comparison to the active immunisation described above.

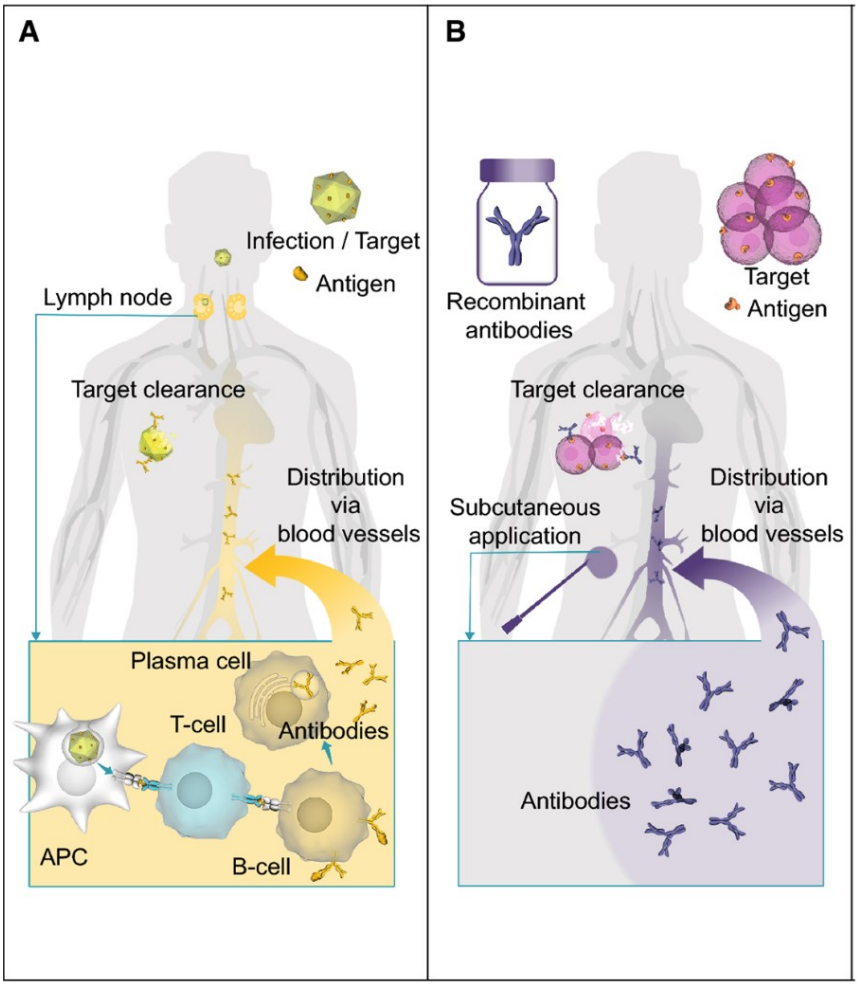


Figure 1: A schematic illustration of active and passive immunisation. A. *Active immunisation is caused by an antigen presented by antigen-presenting cells (APC) in case of infection or after a vaccination. T cells are activated which leads to activation of B cells and production of specific Abs.* ***B.*** *In comparison, passive immunisation is the administration of recombinantly produced specific Abs through subcutaneous application. The recombinant Abs enter blood circulation through diffusion and behave like endogenous Abs. Modified figure*[8]*.*

In principle the same reaction should take place in the case of aberrant cells such as tumour cells. Neoantigens and damage-associated molecular patterns released by cancer cells into the TME would be captured by immune cells and processed, and the targeted cancer cells should release additional tumour-associated antigens upon death allowing the cycle to start over[6,9].

However, in cancer patients, many factors contribute to an inefficient immune response. Each step requires the cooperation of multiple factors, both stimulatory to enhance immunity as well as inhibitory to inhibit immunity, which normally are necessary to keep the process in check and balance and help to prevent the

development of autoimmunity. Thus, tumour-associated antigens may either not be detected or may be treated as self rather than foreign by immune cells, thereby not eliciting an effective T cell response[6]. Even when effective T cell responses occur, they rarely provide protective immunity as they may not be able to properly home in on to the tumour, may be prevented from penetrating the TME, or could be neutralised by factors therein[6].

Glycans form a thick layer as part of the cell surface and are the first contact for immune cells screening host cells to identify any aberrance or presence of pathogens. In this aspect, sialic acids are of particular interest as they are vertebrate-specific (self-associated molecular patterns) and allow distinction from pathogens that do not express them[10]. Lectins specific for sialic acids called sialic acid-binding immunoglobulin-like lectins (Siglecs), have been found on almost all immune cells and are believed to mostly have immune inhibitory effects[11,12]. However, pathogens with evolved strategies, such as molecular mimicry to express host sialic acids, have also been identified[13]. Similarly, tumour cells have been shown to overexpress sialic acids to mask surface antigens and evade the immune system[10,11].

Therapeutic cancer vaccines have mostly failed to yield desired results despite vaccinations existing since the eighteenth century and there being several extensive studies about developing cancer vaccines over the years[14–19]. To date, there are only three FDA approved therapeutic cancer vaccines (DC-based sipuleucel-T against metastatic prostate cancer[20], *Mycobacterium bovis*-based Bacille Calmette-Guerin (BCG) against early-stage bladder cancer[21], and Herpes Simplex Virus type 1- based talimogene laherparepvec (T-VEC) against metastatic melanoma[22]). For a successful vaccination leading to immune system stimulation, factors such as large amounts of high-quality antigen for proper and sustainable immune activation, T cell infiltration of the TME, B cell activation and production of the correct Ab type[9,23], along with maintenance of all these factors, is required. Several methods have been studied so far including vaccination with antigen and adjuvant, autologous dendritic cells loaded with specific tumour antigens, viral or bacterial particles causing a local immune response, or antigens from dead or dying tumours (in situ vaccines)[9,21,22]. The failure at large to achieve vaccines can primarily be attributed to low immunogenicity of tumour-associated antigens[9] caused by incorrect display of the antigen, lack of

detection, or recognition as self rather than foreign, as mentioned before. While this may be overcome by using an adjuvant in the vaccine, there have only been six adjuvants included in licensed vaccines for humans since the first adjuvant was introduced in the 1920s, due to safety and tolerability concerns. Moreover, despite their widespread use, the molecular mechanisms by which the currently used adjuvants work in humans have not been well-established[24]. These issues could be successfully circumvented by administrating recombinant Ab-based target immunotherapies (see **Figure 1 and section 1.3**).

Traditionally, methods such as chemotherapy, radiation therapy, and surgery are also used for cancer management, however, the success of such methods depends on various factors, including type of tumour and stage of the disease. As most tumours are diagnosed at an advanced stage, often the treatment fails. This accounts for ~90% of all cancer related deaths[25]. Along with a wide range of other acute and chronic regimen-related toxicities, chemotherapy and radiotherapy can also trigger matrix remodelling by inducing elevation of profibrotic growth factor levels[26,27]. Such changes in the biochemistry and biomechanics of the matrix can increase metastatic progression, treatment resistance, and recurrence of cancer[26].

1.1.1. Pancreatic cancer

Pancreatic cancer is a generic term for malignancies originating from the exocrine (>95%) or endocrine (<5%) tissue of the pancreas[28]. The most common and aggressive form of pancreatic cancer is the pancreatic ductal adenocarcinoma (PDAC), which comprises of over 90% of all cases, therefore pancreatic cancer is often interchangeably used for PDAC[29,30]. Surgical resection so far is the only potentially curable curative treatment[31]. Due to the lack of early-stage symptoms, pancreatic cancer is often only detected at an advanced, metastatic, and incurable stage so that only 20% of patients profit from or can be cured by tumour resection[32].

Pancreatic cancer is ranked as the 14th most common cancer as well as the 7th most common cause of cancer-related deaths in the world with number of deaths projected to double by 2060[28,31]. In Germany, after mesothelioma (the cancer of the lining covering the outer surface of some of the body's internal organs), pancreatic cancer has the lowest cancer survival rate of all cancers[33]. It is one of the most lethal

malignancies with a very low five-year survival rate of about 9% and a poor average prognosis of around five months after diagnosis[28,34]. As the incidence rates of pancreatic cancer are almost four times higher in high-income countries like the US and Western Europe than in low-income countries, different lifestyles may play a role. Indeed, one-third of all cancers could be prevented with healthy lifestyle habits, and adherence to healthy behaviours was associated with 16-23% lower risk of pancreatic cancer[31]. Family history, tobacco use, chemical exposure (especially to benzene, petrochemicals, pesticides, and certain dyes), old age, diabetes, chronic pancreatitis, *H. pylori* infection, and obesity have all been implicated as serious risk factors for pancreatic cancer[28,31,35]. The incidence, prevalence, and mortality of pancreatic cancer, follow a similar age-related trend of gradually increasing after the age of 30 years and reaching their highest burden after the age of 80 years for all genders[28].

There are multiple fundamental reasons for the high mortality in pancreatic cancer. Firstly, the retroperitoneal position of pancreas deep in the abdomen protects the tumour from imaging detection. Also, the aggressive nature of the tumour leads to early and high rate of metastasis to liver, peritoneum, lung, and, less frequently, bone[31,32,36]. Therefore, almost 50% of patients have metastatic cancer at presentation, while many patients develop metastases within four years of surgery, suggesting micrometastasis in apparently localised tumours. Pancreatic cancer also shows resistance to many anti-tumour therapies. Furthermore, pancreatic cancer patients are quickly drained of strength and find it hard to deal with aggressive treatments [31].

Diagnostic methods and their limitations

The best way to counter pancreatic cancer is timely diagnosis[35]. Several physical methods can be adopted, with imaging diagnosis being the most common one[31]. The golden standard is computed tomography (CT). A protocol pancreatic CT can be performed if pancreatic cancer is suspected, or routine CT was inconclusive[37,38]. Contrast-enhanced magnetic resonance imaging (MRI) may be useful in determining the cancer stage of patients at their initial presentation and may be the best detection technique for small liver metastasis[39]. Endoscopic Ultrasound (EUS) is the most sensitive technique for early neoplasia detection, though may lack specificity in presence of inflammation and sensitivity for distant metastasis[40]. However, CT, MRI, and EUS may not detect some early-stage tumours or may only find a localised

stenosis of the main pancreatic duct especially for carcinoma in situ. Therefore, it may be critical to perform an endoscopic retrograde cholangiopancreatography (ERCP)[31], which demonstrates 92% sensitivity and 96% specificity in diagnosis of pancreatic cancer[41]. On the other hand, ERCP is a highly invasive procedure that could easily lead to acute pancreatitis, bleeding, and cholangitis[42]. Positron emission tomography (PET) has generally limited use in pancreatic cancer diagnosis due to high number of false positive and negative results[43], while staging laparoscopy is suggested to be only used for patients with the highest likelihood of metastases that could not be detected by other means[44,45]. A biopsy of the tumour mass either percutaneously or via EUS can also be performed. However, in most cases pancreatic cancer cannot be confirmed histologically, hence, biopsy is mostly suggested for patients who may be unfit for surgery or if other diagnoses need to be excluded[31,46,47]. Taken together, many physical methods for diagnosis of pancreatic are in practise, however, they are either not sensitive enough or are too invasive in nature. Therefore, there is a dire need for better diagnostic methods.

Various biomarkers in urine, saliva, pancreatic juice, and cyst fluid, as well as blood are being investigated with the aim of improving early diagnosis of pancreatic cancer that is not detectable by imaging and to differentiate between cancer and a benign tumour[31,48]. Protein as well as MicroRNA (miRNA) biomarkers have been studied in urine, which may be able to help diagnose and distinguish between the different stages noninvasively. Nevertheless, more studies are required to validate clinical use of these biomarkers[49,50]. In the saliva in case of cancer, several genes and messenger (m)RNA, miRNA, and long non-coding (lnc)RNAs have so far been identified that may be up- or downregulated compared to healthy controls or benign tumours. These include the miR-1246, which is related to the serum levels of the biomarker CA 19-9 (see below)[51–54]. Along with that, miR-1246 can also be measured in serum and urine with a combined sensitivity of about 85%, making it a useful biomarker for pancreatic cancer[31,54]. Pancreatic juice and pancreatic cyst fluid can similarly be studied after obtaining them at the time of ERCP or EUS, respectively, and could help in early pancreatic cancer diagnosis. Next generation sequencing could be carried out to identify genetic mutations even at low levels[31,55–57]. Other markers in the pancreatic juice and cyst fluid could also be indicative of cancer, including mucins and interleukins, which are found at low levels in normal tissue, yet are upregulated in malignancy[31,58,59].

Lastly, serum also shows variations in various factors that can help identify pancreatic cancer, including miRNAs and lncRNAs[60–64]. Cell-free double-stranded DNA in plasma and serum and its methylation state can also be studied to detect early-stage cancer and its tissue of origin[65,66].

Several carbohydrate antigens (CA) have been extensively analysed in pancreatic cancer, including CA 50 (glycoprotein or glycolipid with epitope similar to CA 19-9 or its afucosylated form[67,68]), CA 72.4 (glycoprotein TAG 72 with sialyl Thomson nouveau (sTn; NeuAcα(2-6)GalNAcα, where NeuAc is N-acetylneuraminic acid and GalNAc is N-acetylgalactosamine) epitope[69]), CA 242 (mucin with epitope similar to CA 50 and CA 19-9 but otherwise unresolved to date[70,71]) , and CA 125 (mucin (MUC)16, a large membrane glycoprotein with complex and heterogeneous glycosylation, including complex N-glycans and sTn[72]) [31]. CA 19-9 (also called sialyl Lewis A (see **section 1.2.1, Figure 3,** and **Figure 4**), with the epitope Neu5Acα2,3Galβ1,3(Fucα1,4)GlcNAc[73], where NeuAc is N-acetylneuraminic acid, Gal is Galactose, Fuc is Fucose, and GlcNAc is N-acetylglucosamine), especially, has been identified as a good indicator of malignancy[31]. DUPAN-2, the precursor of CA 19-9, and Span-1, another glycoprotein biomarker could be useful for testing in cases where CA 19-9 is not expressed, however they do not show high specificity and sensitivity for early detection and need to be studied further[74,75]. Two other glycan biomarkers for PDAC are sialylated tumour-related antigen (sTRA; Neu5Acα2,3Galβ1,3GlcNAcβ1,3Galβ1,4GlcNAc), which is expressed in up to 50% of patients with low CA 19-9 expression, and sialyl Lewis X (ca; Neu5Acα2,3Galβ1,4(Fucα1,3)GlcNAc (see **Figure 3** and **section 1.2.1**)) that is also highly expressed in 30-50% of low CA 19-9 expressing cases[76–78]. Sialyl Le X, however, has a high false-positive rate of 10% in benign pancreatic diseases[77], while sTRA has very low false-positive rate and has also been suggested as an identifier for chemotherapy-resistant PDAC[77,79].

One of the most reliable tests for diagnosing pancreatic cancer in symptomatic patients is the measurement of elevated serum tumour marker CA 19-9. The sensitivity and specificity of CA 19-9 in detection of pancreatic cancer are 79% and 82%, respectively[31]. While other tests may also claim high specificity and sensitivity, it is still recommended to test levels of CA 19-9 in combination to improve accuracy of

diagnosis[31,77,80]. Moreover, CA 19-9 is the only FDA approved marker for pancreatic cancer[81].

1.1.2. Breast cancer

While the tumour-associated carbohydrate antigens (the Globo series, see **section 1.2.2**) studied in this thesis are highly expressed on the surface of many cancers, such as prostate, breast, lung, or pancreatic cancer[82–84], for this study I focused on a breast cancer model only.

Breast cancer is one of the oldest described cancers in human history having already been mentioned in the pyramid age in ancient Egypt (3500-2500 BC), while its various stages were described by Hippocrates in 450 BC[85]. Despite this and the high mortality rate, breast cancer is a taboo subject in many cultures resulting in patients often being reluctant to openly discuss their disease and its symptoms[85]. As a visual and important part of women's physiology, breasts have symbolic association to femininity, fertility, and motherhood. As such, breast cancer not only causes distress over the prognosis of the disease, but also feelings of social and family isolation as well as loss of identity and self-doubt in patients[86]. However, social media, moral and social reforms as well as various breast cancer organisations and initiatives like the Global Breast Cancer Initiative by WHO are helping provide awareness and support even in remote areas of the world. Initiatives such as the common use of the pink symbol for breast cancer around the world also give a sense of solidarity[85,87].

Accounting for 1 in 8 cancer diagnoses today, breast cancer is the most commonly diagnosed cancer in the world and was estimated to have 2.3 million new cases in all genders combined in 2020[87]. Male breast cancer accounts for 1% of all cancers in males and less than 1% of total breast cancer cases in the world[88]. On the other hand, a quarter of all cancer cases in females are represented by breast cancer and it was the cause of every 1 in 6 cancer-related deaths with an estimation of 685,000 women passing away in 2020[87].

Genetic or hereditary causes are responsible for 5-10% of all breast cancer cases[89]. Mutations in the BRCA genes (BRCA1 and 2) attribute to up to 40% of these cases[90]. In comparison, eight out of nine females diagnosed with breast cancer do not have an affected first degree relative[89]. While the age standardised incidence rate in females globally is around 48 in 100,000, the number varies from under 30 in sub-Saharan

Africa to over 70 in 100,000 in western Europe and North America[91]. Societal prevalence of reproductive, hormonal, and behavioural risk factors such as menarche at early age, menopause at later age, first birth at older age, fewer number of children, less breastfeeding, menopausal hormone-replacement therapy, oral contraceptives, as well as increasing life expectancy, alcohol consumption, obesity, and physical inactivity is reflected in the higher number of new cancer cases in high-income countries[87]. Nevertheless, over half of all breast cancer cases are diagnosed in less developed regions with larger populations and cause a significant burden of disease[91]. The low- to middle-income countries also have a disproportionately higher amount (almost two-third[92]) of all breast cancer-related deaths which could be decreased through earlier diagnosis and improved treatment strategies. If the current burden continues as such, numbers as high as 3 million new cases and 1 million deaths a year are expected by 2040 due to population growth and ageing alone[87].
Age is the most important risk factor for breast cancer with the highest age-associated global incidence rates being observed in postmenopausal (above 50 years of age) females[93,94]. In less developed countries, however, more than half of breast cancer cases are diagnosed in females under the age of 50. This is mainly due to the shorter life expectancy and an average younger population in the region[94]. Accordingly, breast cancer incidence is expected to increase in less developed countries also, as they develop more economically and lead to an increase in average life expectancy[92].

Breast cancer can be classified into six intrinsic subtypes based on their immunohistochemical profile: luminal A, luminal B, HER2 enriched, normal-like, basal-like, and claudin-low; the latter two are also known as triple-negative breast cancer subtypes. The differences are based on the presence or absence of three main markers which are also routinely tested in clinic to establish a treatment plan[95,96]: the hormone receptors, progesterone receptor (PR) and oestrogen receptor (ER), and human epidermal growth receptor two (HER2). Along with this, the basal markers (Cytokeratin 5 and 6 (CK5/6), epidermal growth factor receptor (EGFR)), and the Ki-67 (nuclear protein associated with cell proliferation) index are also considered[96]. Both luminal A and B are positive for PR and ER, while luminal B is also positive for HER2[97]. HER2 enriched subtype is only positive for HER2 and not for PR and ER[98]. On the other hand, the normal-like tumour tissue has a low percentage of tumour cells and lacks expression of proliferation genes. The triple-negative breast cancer subtypes are

ER, PR, and HER2 negative[96]. Therefore, triple-negative breast cancer does not respond to targeted therapy methods, such as hormone therapy, and can only be treated with cytotoxic drugs instead[99].

Diagnostic methods and their limitations

Breast imaging can be carried out in various ways. Mammography is the golden standard. The sensitivity of mammography decreases with increase in breast density and, therefore, American Cancer Society recommends mammography for females over 45 years of age, with younger women undergoing mammography at personal preference[100–102]. The disadvantages of the method are false-positive results, radiation exposure, pain, anxiety, and other psychologically adverse effects[101]. Between the ages of 40 and 50 years, a 61% chance of a false-positive result exists, which decreases with older age[100,101]. Healthy women at 55 years of age onwards are recommended to continue mammography screening annually or biannually[100].

Ultrasound is a viable option for risk women with dense breast tissue[100,101]. It is also recommended for initial imaging of palpable masses in lactating and pregnant women, for evaluating problems in breast implants, as a guidance for breast biopsy and for planning radiation therapy, as well as for assessing any abnormalities detected in mammography or MRI[100]. However, here too there is a high-rate of false-positives. Furthermore, results are dependent on the expertise of the operator[103].

MRI can be used for detecting familial breast cancer early on regardless of patient's age, breast density, or risk status[104]. It is, however, not suitable for patients who have metal material in their bodies, are hypersensitive to contrast, or are claustrophobic as it runs for a long time. Also, it does not help distinguish between all breast cancer stages[105]. PET, CT, and single-photon emission computed tomography (SPECT) are not recommended for detecting breast cancer, however, can be used as auxiliary diagnostic methods in some special cases, including screening for metastatic breast cancer, or presence of bone and lymphatic metastases[105].

For histopathological diagnosis, needle biopsy is the main method to obtain tumour tissue[101,105]. The quality of the histology sections greatly depends on characteristics such as tissue handling, ischemic time, cautery, use of frozen sections, fixation, decalcification, and processing, which may give rise to potential artifacts[101]. Some of the disadvantages of needle biopsy could be tumours getting transferred, neoplastic seeding through high-grade non-coaxial biopsies, or multiple insertions[105–107].

Immunohistochemistry (IHC) is commonly used to analyse paraffin sections for expression of ER, PR, and HER2 to guide hormone therapy and prognosis. IHC allows for distinction between benign and malignant tumours as well as ductal and lobular tumours. It can also evaluate interstitial infiltration[108–111]. However, histologic tumour markers may show significant intratumoural variation, even within the same biopsy sample[101]. Additionally, IHC requires fluorescence labelling which is time-consuming and difficult to prepare[105].

During the formation and growth of tumours, cancer cells release multiple components through apoptosis, necrosis, or active release into the blood, which can be studied with liquid biopsy[112,113]. One of these are the circulating tumour cells (CTCs)[114]. An increase in CTC number in blood is related to low survival rate and their persistence in peripheral blood is indicative of poor response to treatment[95]. CTCs can be specifically isolated from other blood cells with positive selection for surface molecules specific for CTCs. An example is the epithelial cell adhesion molecule (EpCAM), whose expression level is also used as an indicator for the disease prognosis[114]. CTCs have also been shown to be a marker for the transition towards metastasis. CTC-based analyses have an advantage over solid tumour biopsies as blood samples are easier to obtain and can be taken regularly over the period of treatment[95,115]. However, the frequency of CTCs in blood is very low, which, along with the heterogeneity of antigens present on the CTC surface, makes them difficult to detect and limits their diagnostic value for early-stage breast cancer, for example[116].

Fragments of tumour genomic DNA, known as the circulating tumour DNA (ctDNA)[105], are released actively by tumour cells, or through apoptosis and necrosis of the cells[114]. Circulating tDNA can help monitor the progression of cancer as it has a high turnover and reflects the genomics of the tumour mutations as well as epigenetic alterations[117,118]. It can be taken up by other cells in the body and, undergoing horizontal gene transfer, can result in alterations in the recipient cell that may lead to metastasis[119]. Not only is ctDNA more specific for malignancy and a more sensitive marker for monitoring tumour burden than CTCs[120,121], but it can also be used to detect cancers relatively early as well as identify the organ of origin of cancer[122]. However, ctDNA comprises only of <0.1% of total cell-free DNA, which is also released from normal cells, especially in early-stage cancer patients[123]. This makes detection of ctDNA challenging[120]. Studies have also shown that combinations of certain circulating miRNAs can be used to distinguish between breast cancer and healthy cells as well

as cancer and benign lesions with high specificity and sensitivity, however the panels of cell free miRNA studied in various studies are inconsistent and, so far, there is no miRNA panel that can be used in a clinical setting[124].

Extracellular vesicles (EVs) are secreted from almost all cells including cancer cells. These contain lipids, proteins, DNA, various RNAs as well as other biomolecular structures that depict the origin and the state of tumour[125]. For example, the overexpression of focal adhesion kinase, EGFR proteins, as well as various miRNAs is implicated in breast cancer. These were found to be significantly enriched in plasma isolated EVs of breast cancer patients[124]. Therefore, EVs have great potential to be analysed for their content and lead to detection of early-stage breast cancer as well as help differentiate between cancer and benign, non-cancerous diseases[125,126]. However, identification and isolation of EVs still poses an issue as cell contamination and platelet remnants are common[127]. Even though it has been shown that EV profiles from same source could vary depending on isolation methods, a standardised method of isolation and purification of EVs is currently lacking[124,128].

As in the case of pancreatic cancer, body biofluids such as urine can also be tested for biomarkers of breast cancer[124]. Among others, the analysis includes that of phospholipids such as phosphatidylcholine (PC) and phosphatidylethanolamine (PE) or their precursors, which have been observed to have increased metabolism in breast cancer tissues[129–131]. Also, urinary miRNA profile has been shown to vary between healthy and primary breast cancer patients and could help identify cancer with a sensitivity and specificity of more than 90%[132]. Exosomes can also be isolated from urine and analysed for cancer biomarkers. This includes exosomal miRNA, which has also been shown to distinguish between cancer and healthy patients[133,134]. While urine-derived biomarkers including proteins, metabolites, miRNAs, and other cellular components could potentially be used for diagnosis of breast cancer, they are still being discovered and more studies are needed to establish their sensitivity and specificity[135].

Tumour-associated antigens (TAA) include aberrant macromolecular structures, changes in protein expression levels, and alterations in posttranslational modifications like glycosylation, phosphorylation, acetylation, methylation, etc. that can cause an immunological response as the immune system does not recognise it as ‘self’ anymore[136]. These trigger the production of autoantibodies (AAbs), which can be used for diagnosis as well as monitoring of disease[136]. While the sheer amount of total AAbs

in the human body remains unknown, the number is estimated to be more than 1 billion[137]. One of the main advantages of AAbs targeting TAAs is that they can be used for early detection of the tumourigenesis process, years before clinical symptoms develop[138–140]. Additionally, AAbs are highly stable and can be analysed easily. They can be detected on archived samples and have well-developed secondary reagents for easy identification[136,141,142]. Lastly, they circulate in the blood longer than TAAs that suffer from low concentrations as well as quick degradation and, consequently, less circulation time[141,142]. While the AAbs could be specific for their antigens, they may still be universally found across various cancers, which limits their use for early detection of any one cancer[136,143]. Combination of multiple AAbs in screening of antigens might increase their sensitivity, while maintaining their specificity[136]. Some of the AAbs related to breast cancer that have been studied so far target TAAs including p53 (nuclear phosphoprotein)[144,145], MUC1 (highly glycosylated transmembrane protein)[146], HER2 (tyrosine kinase receptor protein (also known as human epidermal growth receptor two, as mentioned previously))[145], HSP60 (heat shock chaperone protein)[147], BRCA1 (breast cancer 1 protein normally helps repair DNA)[148], Globo-H (also known as stage specific embryonic antigen 3b is a tumour-associated carbohydrate antigen with the structure Fucα1-2Galβ1-3GalNAcβ1-3Galα1-4Galβ1-4Glcβ1; mentioned in detail in **section 1.2.2**)[149], and Galectin 3 (β-galactoside-binding lectin)[150].

The FDA approved biomarkers for breast cancer detection include ER, PR, and HER2 as tissue markers detected by IHC and, the TAAs, CA 15-3 and CA 27-29 as circulating carcinoma proteins for serum assays[124,151]. The latter two are soluble forms of the glycoprotein MUC1[151,152]. Additional biomarkers include CA 125 (also described in **section 1.1.1**) and carcinoembryonic antigen (CEA), which is a glycoprotein as well[124,151]. Tumour-associated carbohydrate antigens also show aberrant expression of glycans in breast cancer. These include Tn, sTn, Thomsen–Friedenreich antigen (TF; Galβ1,3-GalNAcα) and sLe X, as well as glycosphingolipids such as the globo family[153]. Like MUC16 [72], MUC1 is also highly enriched with Tn and sTn structures[154]. The globo family was studied as part of this thesis and is described in detail later (**section 1.2.2**).

1.2. Tumour-associated carbohydrate antigens

Glycosylation is a result of non-template driven glycan biosynthesis, mediated by coordinated functions of glycosyltransferases and glycosidases and takes place in the endoplasmic reticulum/Golgi compartment of essentially all cells where glycans are then attached to proteins or lipids. The human glycome is built up from nine monosaccharides, which can be differently linked leading to the large number of glycan structures in nature. Furthermore, glycosylation can differ among cells and tissues, and it is affected by the cell microenvironment causing structural complexity and heterogeneity in glycans[155–157]. As the most abundant form of protein post-translational modification (PTM), glycosylation has a substantial impact on every cellular process, including intra- and intercellular recognition, signalling, and interaction processes. Glycans play important roles in protein folding and conformation, trafficking, and degradation. They are also essential in adhesion, cell-matrix interaction, protection from proteases, immune recognition, and membrane organisation as part of the glycocalyx, among others. Therefore, it is no surprise that aberrant alterations in glycosylation are a hallmark for many diseases, including cancer[25,158,159]. These alterations may include synthesis of truncated glycans like the Tn antigen, neo-synthesis leading to abnormal glycosylation patterns like the sLe X, or deletion of a normal glycan like the core 3-derived O-glycans in colorectal tumours[160–162]. Such tumour-associated carbohydrate antigens (TACAs), overexpressed on cancer cells, are molecular markers that can be used to distinguish them from healthy cells. They are not only representative of the changes in neoplastic cell behaviour, but also play critical roles in the development and metastatic progression of cancer[25,158,159]. TACAs can be divided into five categories: branched N-glycans, truncated O-glycans, the Lewis family, embryonic glycolipids, and glycosaminoglycans (**Figure 2**).

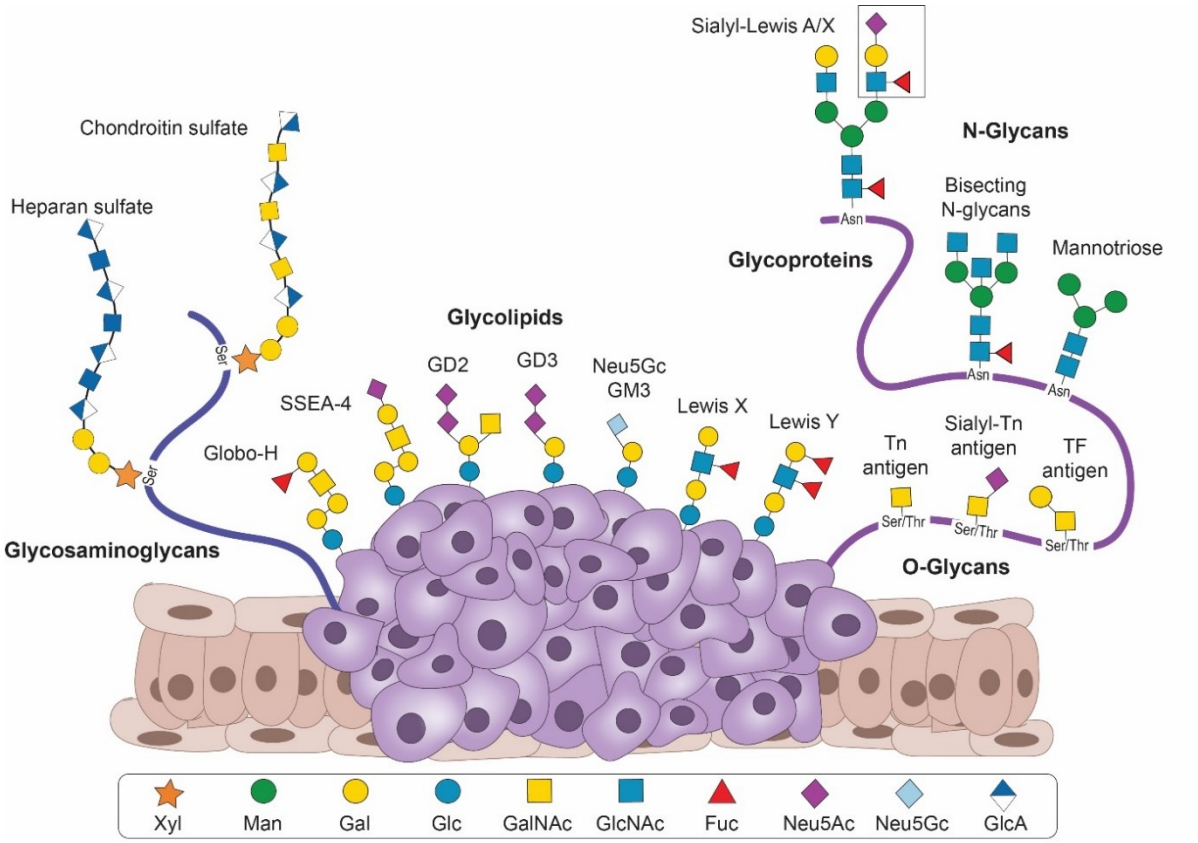


Figure 2: Five main categories of TACAs. *These are glycosaminoglycans such as heparan sulphate and chondroitin sulphate, glycolipids including glycosphingolipids such as the globo- and the ganglio-family, the Lewis antigens found on both glycosphingolipids and glycoproteins, cancer-associated N-glycans such as mannotriose and bisecting N-glycans, and O- glycans carried by proteins (glycoproteins) such as Tn, sTn, and TF antigen. SSEA-4: Site specific embryonic antigen-4. Monosaccharide abbreviations are described in the list of abbreviations. Figure credits: Dr Oren Moscovitz.*

N-glycosylation is characteristically an addition of a precursor glycan structure onto Asn in the consensus sequence Asn-X-Ser/Thr[163]. N-glycan TACAs are highly branched structures and play significant roles[164]. For example, an increased expression of β1-6 branched N-oligosaccharides is linked to cancer progression and aggressiveness in breast, colon, oesophagus, gliomas, and endometrium cancers[165,166].

In O-linked glycosylation, glycans are added onto the side chain of Ser or Thr[163]. Tumour-associated O-glycans can be divided into O-glycans with altered expression profile, oncofoetal glycans (which are rarely found in healthy adult tissue, but are expressed in embryonic tissue), and neoantigens (which are neither found in healthy adult tissue nor in embryonic tissue and are, therefore, novel structures)[167]. Tn (also called CD175), sTn (CD175s), TF (CD176 as well as T antigen), and sTF (Neu5Acα2,6- and Neu5Acα2,3-Galβ1,3-GalNAcα) are all truncated O-glycans

implicated in cancer[168]. Some other extended O-glycans found on tumour cells include the ABO(H) antigens, the Lewis antigens, and poly-N-acetyllactosamine[167]. Lewis antigens are described in detail in **section 1.2.1**. The expression of heavily O-glycosylated mucins in cancer is associated with a more malignant phenotype. An example of such a mucin is MUC16, which, when overexpressed in ovarian cancer and pancreatic cancer, helps in cancer cell survival by protecting cells from being recognised by natural killer cells as well as facilitating them to spread and metastasize to the peritoneal cavity[169].

Glycolipids comprise of a carbohydrate group linked to a lipid moiety via a glycosyl linkage. Glycosphingolipids have a basic structure made up of a ceramide having monosaccharides and oligosaccharide chains attached to it[170]. Examples of these are gangliosides and globo family. Gangliosides are present on healthy tissue and are often overexpressed in cancers like lung cancer, melanoma, and neuroblastomas[171]. On the other hand, the globo family TACAs are normally not expressed on healthy tissue but on embryonic tissue instead[172] (**section 1.2.2** and **Figure 6**).

Glycosaminoglycans (GAGs) are linearly formed with disaccharide repeats of GlcNAc or GalNAc, in combination with a uronic acid or galactose[173,174]. They are often sulfated and linked to anchor proteins called proteoglycans[175]. The GAGs include heparan sulfate, chondroitin sulfate, and dermatan sulfate[176]. Low levels of heparan sulfate, for example, have been correlated to high metastatic activity of tumours such as colon, mesothelioma, lung, hepatocarcinoma, and breast cancer[177].

The understanding of TACA functions on cancer cells has led to novel immunotherapeutic strategies in clinical settings. While a majority of these have not been given FDA approval yet, there are specific monoclonal Abs against Tn/sTn-MUC1 and MUC16 glycopeptides as well as GD2 (Dinutuximab) that have met FDA approval and are being used in cancer treatment[163,178–181]. Recently developed TACA-based anti-cancer immunotherapy includes, among others, engineering of chimeric antigen receptor (CAR) T cells[171,182]. For example, CAR T cells specific for GD2 were shown to reach solid tumour sites and produce cytokines reprogramming the tumour microenvironment and the anti-sTF-MUC1 peptide CAR T cells reportedly induced apoptosis and necrosis of tumour cells which was related to production of proinflammatory cytokines and delaying of tumour growth[183–186]. TACAs present a promising target and many available therapeutic opportunities can be explored for targeting of TACAs as a new field within cancer immunotherapy[171].

1.2.1. CA 19-9 (Sialyl Lewis A; sLe A)

Discovery, biosynthesis, and clinical importance of CA 19-9

In 1979, Herlyn *et al.* immunised mice with cells from five colorectal carcinoma (CRC) cell lines and obtained 104 hybridomas. In a radioimmunoassay (RIA), only two (1083-17-1A and 1116-56-2) of the hybridomas showed reactivity against CRC cells[187]. An IgG2 was obtained from 1083-17-1A, which was approved in Europe in 1995 (but not by FDA) and was used in clinical trials to cure CRC, till it was eventually removed due to poor prognosis of recurrence[188]. The second hybridoma (1116-56)-derived IgG1 mAb, 1116-NS-19-9, was tested on SW 1116 CRC cells to detect an antigen present only in stomach and pancreatic cancers[189,190]. The high specificity of the Ab to the antigen was attributed to the "nature of the antigen", while it was also suggested that "its presence in tumors of endodermal origin (…pancreatic and ...gastric carcinomas) may prove to be advantageous, given the usual difficulty in diagnosing these types of cancer during the early growth stages"[189]. The antigen's nature was established by Magnani *et al.* in the same year to be a monosialoganglioside not detected in any human tissue other than colorectal cancer and human meconium, which is usually only present in embryonic tissue[191]. 1116-NS-19-9 is still used for clinical diagnosis to date.

Today known as CA 19-9, CD43, or sialyl Lewis A (sLe A; in this thesis, CA 19-9 and sLe A have been used interchangeably), the 1116-NS-19-9-targeted antigen is found in embryonic tissue and, only at low levels, in healthy tissue, but overexpressed in certain epithelial cancers[192]. It has been shown to actively drive the tissue to pancreatitis, which can cause pancreatic cancer. As pancreatitis can be reversed by targeting sLe A with mAbs, sLe A provides an important therapy target for pancreatitis and pancreatic cancer[193]. It is usually attached to different proteins, like mucins[194,195], Kininogen, and circulating apolipoproteins[196]. CA 19-9, as mentioned previously (**section 1.1.1**), is a tetrasaccharide composed of Neu5Acα2,3Galβ1,3(Fucα1,4)GlcNAc. It is synthesised by glycosyltransferases on both N- and O-linked glycans[73]. Sialyl Le A is part of the human histo-blood system, known as the Lewis antigen system, which consists of the type I and type II Lewis antigens[197]. H1, H2, Lewis A (Le A), Lewis B (Le B), Lewis X (Le X), and Lewis Y (Le Y) are made of the same three monosaccharides, namely GlcNAc, Gal, and Fuc, and differ from one another only in their corresponding glycosidic bonds (type I: Galβ1-

3GlcNAc; type II: Galβ1-4GlcNAc). Further addition of sialic acids to Le A and Le X yields sLe A and sLe X, respectively (**Figure 3**)[197].

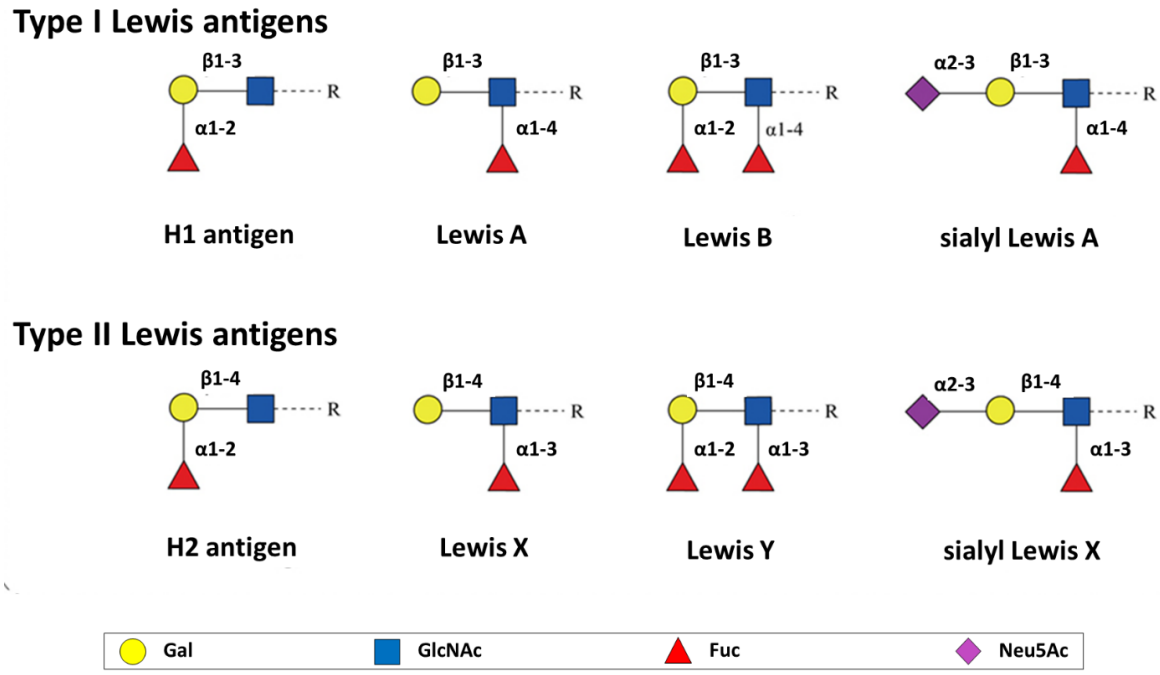


Figure 3: The Lewis antigens. *Type I consists of H1 antigen, Le A, Le B, and sLe A, while type II consists of H2, Le X, Le Y, and sLe X. Monosaccharide abbreviations are described in the list of abbreviations. Adapted figure* [197].

The Lewis antigens are not only expressed on the surface of red blood cells, but also in the epithelium lining of the gastrointestinal, respiratory, urinary, and reproductive tracts[198,199]. While Lewis antigens play an important role in many inter- and intracellular biological processes in embryogenesis and later development, including cell adhesion and cell communication, their overexpression in mature tissue is implicated in cancer. Le B is a receptor for the gram-negative bacterium, *H. pylori*, which is a leading cause of gastric cancer and can provoke inflammation leading to chronic gastritis and duodenal ulcers, and eventually adenocarcinoma[198]. Once overexpressed in inflammation, sLe A and sLe X are also receptors of *H. pylori*[198]. Furthermore, Lewis antigens help in cancer development and progression by being involved in epithelial to mesenchymal transition (EMT), cancer cell and endothelial and/or immune cell interaction, multidrug resistance induction, and cancer stemness[197]. The overexpression of these antigens is attributed to the overexpression of corresponding fucosyltransferases (α1,2, 1,3, and/or 1,4 fucosyltransferase[200]), as fucose addition to

the structure is the final step in the synthesis of these TACA structures (**Figure 4**)[73,197,201].

For the expression of the final sLe A structure the FUT3 enzyme with α1,3/4 fucosyltransferase activity adds the fucose to 3'-OH of the GlcNAC[197] (**Figure 4**). Approximately 6% Caucasians and 22% non-Caucasians with genotype Le A^-B^- lack the 1,4-fucosyltrasferase gene and, consequently, do not synthesize sLe A[73]. In normal epithelial cells, disialyl Le A is synthesized from sLe A and interacts with immunosuppressing Siglecs on monocytes and macrophages to maintain immunological order in the mucosal membrane of the digestive system[202]. In the early malignancy stages, the gene for the 2,6 sialyltransferase is silenced, so that the synthesis is incomplete and sLe A begins to accumulate (**Figure 4**). Due to cancer progression, the malignant cells are eventually faced with lack of oxygen and evolve by inducing integral and irreparable transcription of several glycogenes to adapt to their hypoxic environment. These are directly or indirectly involved in further accelerating the synthesis of sLe A. Hypoxia-resistant cancer tissue makes carbohydrate metabolism adjustments to produce ATP from anaerobic glycolysis (the Warburg effect) as well as produce vascular endothelial growth factor (VEGF) that supports tumour angiogenesis. Not only does this help the cells survive the hypoxic conditions, but also leads to selective survival and expansion of hypoxia-resistant cancer cells which are highly invasive and metastatic[202].

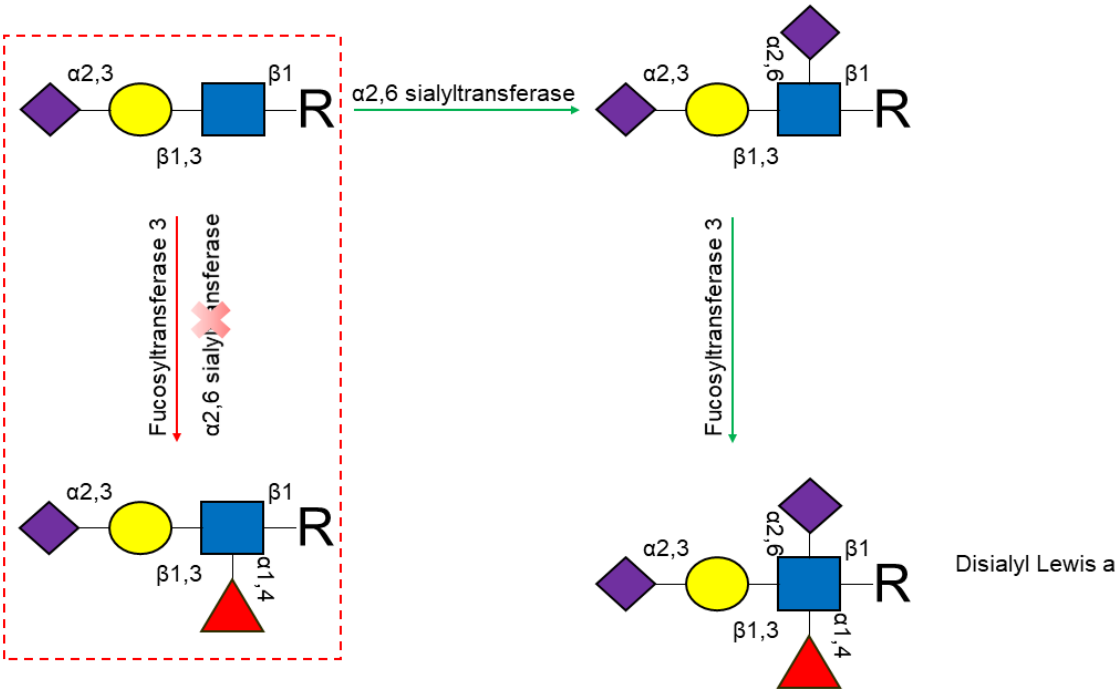


Figure 4: The synthesis of sLe A. *Sialyl Lewis A is synthesised in early cancer stage in place of Disialyl Le A, when the gene expressing α2,6-sialyltransferase is silenced. Figure based on Borenstein et al. 2021* [203].

The metastatic cascade involves steps, such as migration of metastatic cell from primary tumour site, intravasation, circulation in blood, interaction with endothelium, extravasation, and formation of TME at the metastatic site. For this, the cancer cell requires specific adhesive interactions with other cells as well as with components of the extracellular matrix[192]. Sialyl Le A facilitates vascularisation of tumours by interacting with endothelial (E)- or platelet (P)-selectins present on endothelial cells and causing a 'tumbling' effect of the cancer cells to the metastatic site[202] (**Figure 5**). This is similar to the mechanism leukocytes and neutrophils normally use in an immune reaction with help of sLe X to get to a site of inflammation[204].

Tumour-associated stromal cells are vital for tumour invasion and metastasis and can arise from different cellular origins: fibroblasts, pericytes, bone marrow mesenchymal stromal cells, adipocytes, macrophages as well as, endothelial cells that have undergone endothelial mesenchymal transition, and cancer cells that have undergone EMT[205]. The stromal cells secrete various pro-inflammatory cytokines such as IL6, IL-1β, TNF-α, and CXCL1, which attract immune cells and stimulate further tumour growth and metastasis. Tumour-associated inflammation is also caused by the hypoxic conditions of the TME, which allow for granulocyte and monocyte/macrophage

infiltration and activation in the TME. Furthermore, sLe A also recruits circulating lymphocytes to peripheral lymph nodes and inflamed tissue[206] (**Figure 5**). Additionally, while in normal tissue, disialyl Le A serves as a ligand for the immunosuppressive siglec-7 and -9 on macrophages/monocytes and $CD8^+$ T cells and maintains the immunological homeostasis, the shift of expression towards sLe A in tumour instead abrogates the interaction causing chronic inflammatory stimuli that may allow progression of cancer[207]. Inflammation causes glycan alteration as cytokines like IL6 and IL-1β also regulate expression of glycosyltransferases and lead to more production of sLe A among other glycans[202,206]. Additionally, VEGF produced by hypoxic cancer cells as well as TNF-α, IL-1α, and IL-1β in the TME have been shown to induce E-selectin expression on endothelial cells, therefore further helping in invasion and metastasis[202].

The tumour-infiltrating immune cells recruited by inflammation develop immunosuppressive features and push towards tumour growth and survival[208]. Glycans play a vital role in immune suppression also. Fucosylated Lewis antigens also interact with macrophages and dendritic cells to induce production of anti-inflammatory cytokines. Hyper-fucosylation and sialylation have been implicated in helping cancer cells escape the immune system and cause anti-inflammatory microenvironments, which includes influencing Natural Killer cell-dependent tumour surveillance (**Figure 5**) [206].

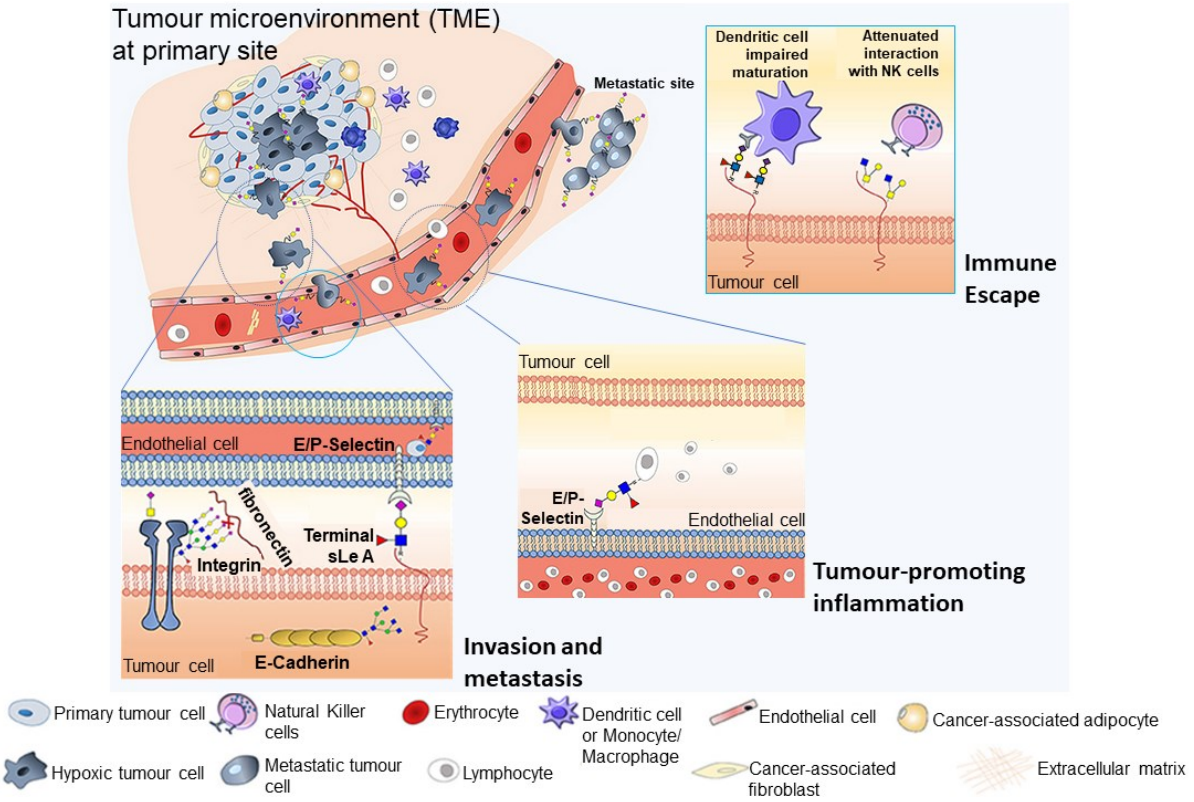


Figure 5: Roles of sLe A in cancer. *Sialyl Le A induces the expression of E-selectins on endothelial cells to help the cancer cells in invasion and metastasis. It also helps promote tumour-associated inflammation so that tumour-infiltrating immune cells would develop immunosuppressive features. Fucosylated Lewis antigens also interact with dendritic cells to produce anti-inflammatory cytokines and with Natural killer (NK) cells to evade the immune system. Modified figure*[206]*.*

In summary, sLe A is highly overrepresented on cancer cells and plays a significant role in metastasis, tumour-associated inflammation, and immune evasion. It is also the most reliable indicator and, so far, the only FDA approved marker for pancreatic cancer[31,81,209]. High levels of sLe A are also associated with other cancer and non-cancer diseases, such as digestive tract, liver, breast, lung, and ovarian cancer, as well as pancreatitis, diabetes mellitus, and obstructive jaundice[192,202,210].

Despite its significant importance as well as boasted high sensitivity (79%) and specificity (82%)[31], CA 19-9 assays have been shown to give false positive and negative readings, posing a great challenge for timely detection of pancreatic cancer[81,211]. The serological assay available since almost three decades and used for detection and to monitor clinical response to therapy, involves mAb capturing of CA 19-9 antigen. However, anti-carbohydrate mAbs generally have low affinity and specificity compared to Abs against proteins, because carbohydrates are less immunogenic.

Indeed, there have also been studies showing broader specificities of Abs targeting CA 19-9 and, therefore, may lead to false detection of the antigen[81].

Abs with better affinity are required to improve detection of pancreatic cancer. The mAbs, murine 1116-NS-19-9 [189] and human 5B1[212], are currently the most clinically used Abs for diagnosis and therapy of pancreatic cancer, respectively[203]. However, these Abs too have low binding affinity to sLe A and tools are being developed to improve on the affinity of the Abs[203]. Borenstein-Katz *et al.* 2021 studied both Abs in detail and improved on the binding affinity of 1116-NS-19-9 with an Ab-design method called AbLIFT approximately ten-fold from 14.7 μM to 1.7 μM[203].

1.2.2. Globo-family

Glycosphingolipids (GSLs) consist of oligosaccharides glycosidically bonded to the terminal primary hydroxyl group of ceramide lipids[213]. Based on the LIPIDS MAPS Structure Database[214], there are over 400 unique glycan structures ranging from 1 to 20 sugar residues that are attached to ceramide in vertebrates[215]. Globo-, ganglio-, isoglobo-, lacto-, and neolacto-series are the five families of GSLs that share a common lactosylceramide precursor (**Figure 6**). GSLs play a crucial role in embryonic development and differentiation, and are also a part of many cellular processes including migration, morphogenesis, cell-to-cell and cell-to-matrix interactions[213]. The key enzyme responsible for the globo-family biosynthesis, α1,4-galactosyltransferase (α1,4GalT), is encoded by the *A4GALT* gene. It has been shown through depletion of this gene that GSLs play major roles in the EMT of cancer cells and that α1,4GalT and globo-family are required for cell-cell adhesion via E-cadherin[216]. Their metabolism is highly dynamic with rearrangements in GSL composition occurring during normal embryonic development and tissue lineage differentiation. These events, however, also take place in cancer, therefore, GSLs play key roles in malignant transformation[217]. For example, while globo- and lacto-series GSLs are predominantly expressed in undifferentiated human embryonic stem cells (hESCs), the ganglio-series shows higher expression in differentiated cells. Therefore, during differentiation, the glycosyltransferases responsible for the globo- and lacto-series are downregulated, while the ones responsible for the synthesis of the ganglio-series are upregulated[218]. In line with that, the globo-series GSLs have been shown to be (over)expressed in various cancers unlike the healthy tissue[219]. The globo-series (**Figure 6**) includes

globotriaosylceramide (GB3), globotetraosylceramide (GB4), globopentaosyl-ceramide (GB5, also SSEA3 as mentioned before), Globo-H (SSEA3b as mentioned before), and SSEA-4. In this thesis, I worked with the first four and focused specifically on Globo-H.

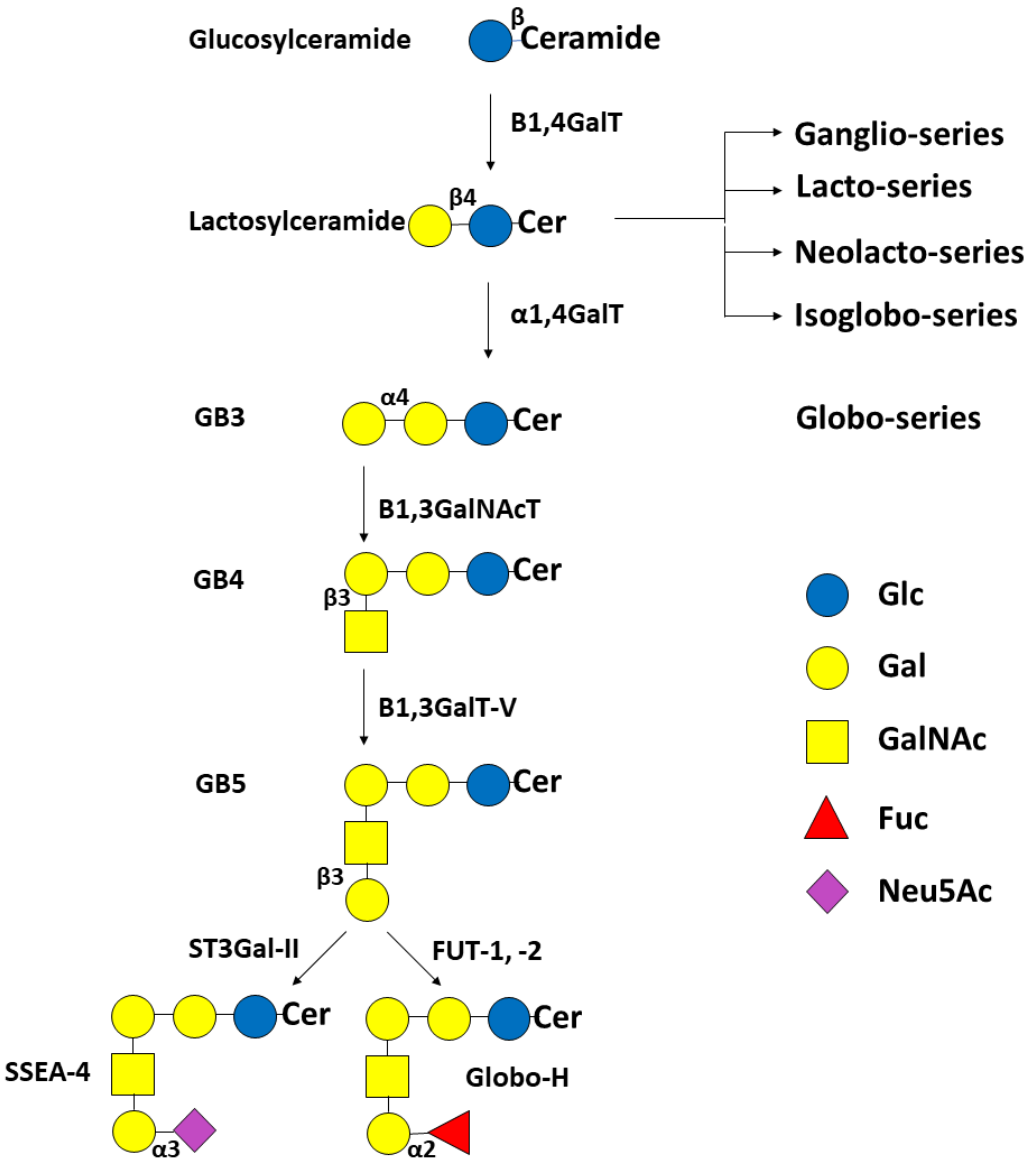


Figure 6: Schematic diagram of biosynthesis of globo-family glycosphingolipids. *Five different families of glycosphingolipids are formed from lactosylceramide (Lac-Cer; also known as GB2). The globo-series is built sequentially through the work of identified glycosyltransferases. The linkages of the bonds are given. Figure based on Jacob et al. 2018*[216] *and Suila et al. 2011*[220].

GB3

GB3 (or CD77), initially identified on human erythrocytes as the Pk antigen of the P blood group system[221], was later shown to be overexpressed in many cancers including Burkitt lymphoma[222,223], and solid tumours such as ovarian, breast, colon,

pancreatic and gastric cancer[224,225]. GB3 is also a surface receptor for shiga-like toxins produced by pathogenic *Escherichia coli (E. coli)* strains. The receptor leads to endocytosis-mediated internalisation. In addition to this, the bacterial surface lectin LecA also binds GB3 and engulfs the pathogen *Pseudomonas aeruginosa*. The LecA-GB3 interaction induces a "lipid zipper" formation so that actin polymerisation is not required for bacterial uptake[226]. Consequently, GB3 is currently being investigated as a possible target candidate for toxin-based therapeutics against cancer[224,225].

GB4

Another member of the P blood group system, GB4 was initially characterised as the P antigen with high expression on the erythrocyte membrane[227,228]. It is, however, scantily expressed on other differentiated cell types. Furthermore, GB4 plays a role during embryogenesis and cancer and is predominantly expressed in both as well[229]. GB4 is involved in cell adhesion[230] and promotes EGFR-induced mitogen-activated protein kinase (MAPK) signalling by interacting and forming a complex with EGFR[229]. This may lead to cancer cell proliferation and cancer progression[229].

GB5

Frequently used as a surface marker for hESCs[231], GB5 (or SSEA-3 as mentioned earlier), has also been detected to be overexpressed in breast[82] and colorectal tumours[232]. While it is found in various normal tissues such as kidney, rectum, testis, and thymus, its expression is mainly confined to the cytoplasm or the apical surface of epithelial cells and, therefore, inaccessible to the immune system. This makes GB5 a good target for cancer immunotherapy[82]. On one hand, upregulation of the enzyme responsible for galactosylation of GB4 to GB5, β1,3- galactosyltransferase V (β1,3GalT-V), was seen in metastatic hepatocellular carcinoma cells and liver cancer-initiating cells[233]. On the other hand, knockdown of β1,3GalT-V not only inhibited proliferation, but also induced caspase-3-mediated apoptosis in breast cancer cells showing that GB5, Globo-H, and SSEA-4 are important for cancer cell survival (**Figure 7**)[233]. This is because GB5/Globo-H form a complex with focal adhesion kinase (FAK) and GB5/SSEA-4 form a complex with caveolin-1 (CAV1), which further interact with the kinase AKT and receptor-interacting protein kinase (RIP), respectively. β1,3GalT-V knockdown causes dissociation of RIP from the complex and leads to interaction with the Fas death domain (FADD) instead, activating the Fas-dependent pathway and causing FAK degradation and apoptosis[233].

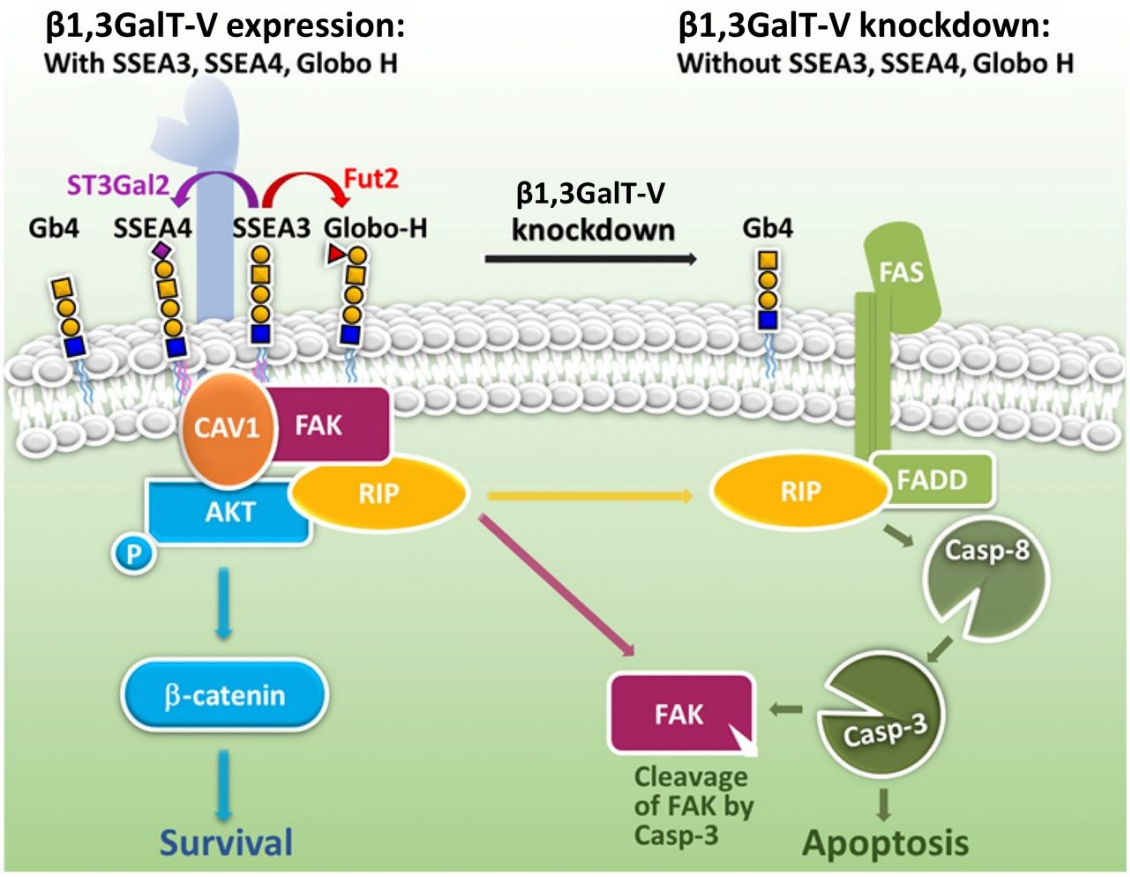


Figure 7: Importance of GB5, Globo-H and SSEA-4 for cancer cell survival. *GB5, Globo-H and SSEA-4 play key roles in the survival of cancer cells. This is highlighted by the knockdown of β1,3GalT-V, which causes dissociation of the GB5/Globo-H from their complex with FAK, which in return is interacting with RIP and CAV1, both of which in return are interacting with the AKT kinase. This causes RIP to dissociate and interact with FADD instead, initiating an apoptosis cascade through caspase (Casp)-8 and Casp-3 which leads to the death of the cancer cell. FAK: focal adhesion kinase, CAV1: caveolin-1, RIP: receptor-interacting protein kinase, FADD: Fas-death domain. Adapted figure* [233].

Globo-H

Globo-H is a hexasaccharide comprising of Fucα1-2Galβ1-3GalNAcβ1-3Galα1-4Galβ1-4Glcβ1-Cer as mentioned earlier (**section 1.1.2**). Much like most embryonic antigens, Globo-H is vital for normal embryonic development and differentiation[213] with expression decreasing thereafter. Globo-H is then only found on the apical epithelial cells at the luminal border with low expression, and, therefore, similar to GB5 cannot be accessed by immune cells[234,235]. Globo-H was first identified in 1983 in human teratocarcinoma and in the human breast cancer cell line MCF-7[236]. Its overexpression has since been reported in many epithelial cancers, including breast, colon, ovarian, gastric, pancreatic, lung, and prostate cancers[234,237]. For this, the Globo-H-specific mAbs, MBr1 (IgM)[238] and VK9 (IgG3)[239], were used. VK9 has high affinity (~1 nM as

calculated with glycan array) as well as specificity for Globo-H and requires the four sugars at the non-reducing terminal of Globo-H for binding. The terminal fucose is essential as removing it disrupts the binding[149,239]. Globo-H expression has also been studied in breast cancer stem cells. It was shown that Globo-H^{high} cells were more tumourigenic with higher blood vessel densities than Globo-H^{low} cells leading to significantly increased tumour growth[82,235,240].

Tumour cells shed Globo-H-ceramide (Cer) containing microvesicles into their microenvironment which are either taken up by endothelial cells[240] or immune cells[241]. Uptake by endothelial cells results in interaction with translin-associated factor X (TRAX) that leads to Ca^{2+} influx, tube formation, and expression of growth factors like VEGF-A, and, consequently, migration and promotion of angiogenesis[235,240]. The importance of the terminal fucose is highlighted here by the fact that only Globo-H-Cer, and not GB5-Cer, can bind to TRAX and induce angiogenesis [235,240]. Therefore, removing the fucose could be of therapeutic interest making the two fucosyltransferases, FUT1 and 2, responsible for the addition of the α1,2-fucose to Globo-H in breast cancer cells, potential targets for cancer therapy[82].

On the other hand, uptake of Globo-H-Cer microvesicles by immune cells promotes immunosuppression by downregulating the Notch1 signalling pathway, which is an important pathway for T cell activation[241]. Additionally, as described earlier, Globo-H also promotes tumour cell survival along with GB5 as part of a complex with FAK (**Figure 7**)[233].

Globo-H overexpression on cancer cells as well as its critical role in cancer progression make it a good target for therapy. Clinical trials based on Globo-H targeting Abs as well as Globo-H based vaccines are underway. These include a phase III trial with a Globo-H vaccine. Though only half the candidates with metastatic breast cancer developed an Ab response, they showed significantly prolonged survival[242,243]. Phase I/II trials with an mAb (OBI-888) targeting Globo-H, and phase I/II trials with an Ab-drug conjugate (OBI-999) composed of a Globo-H-specific mAb and an anti-mitotic drug[244,245].

Globo-H and SSEA-4, another well studied TACA member of the globo family, share a common core glycan structure (Galβ(1–3)GalNAcβ(1–3)Galα(1–4)Galβ(1–4)Glc) of GB5-ceramide. While FUT1 and 2 are responsible for adding the α-Fu to the Gal-5 to synthesize Globo-H, SSEA-4 is synthesised with the addition of a sialic acid by α2,3-sialyltransferase 2 (ST3GAL-II). In 2020, C. Soliman *et al.* described a chimeric (ch)

Ab (ch28/11), which binds specifically to SSEA-4. With the help of X-ray structures, they identified the interacting amino acids in the antigen binding fragment (Fab) of the Ab including the ones interacting with Gal3 and GalNAc4 in the common Globo-H and SSEA-4 structures[246].

1.3. Antibodies and Nanobodies

1.3.1. Antibodies

An immunoglobulin G (IgG)-type Ab is a 150 kDa, Y-shaped, soluble protein made up of two heavy (H) and two light (L) chains linked through disulphide bonds. The constant (C) domains C_H2 and C_H3 form the so-called fragment crystallisable (F_C) region, while the C_H1, the C_L, and the variable domains of heavy and light chain, V_H and V_L, form the Fab. Divided into two parts, the IgG interacts with antigen molecules directly with its variable region (V_H and V_L) and mediates various effector functions via its constant region (C_H and C_L). While the human H chain can be divided into five classes (IgM, IgG, IgA, IgD, and IgE), the L chain comprises only of two classes, (kappa (κ) and lambda (λ))[247]. The various effector functions, including blocking of receptor-induced proliferative signalling and activation of anti-tumour immune responses, such as Ab-dependent cellular cytotoxicity (ADCC), complement-dependent cytotoxicity (CDC), and Ab-dependent cellular phagocytosis (ADCP), could lead to pathogen cell death. They are caused by the F_C region, which can engage multiple F_C receptors (F_CRs) expressed at the surface of various immune cell populations, like NK cells, neutrophils, monocytes, dendritic cells, and eosinophils. The resulting activated mechanisms then aim to eliminate the cell the Ab is bound to. The type of mechanism that is activated depends also on which IgG subclass the Ab belongs to. For example, IgG1 and IgG3 subclasses are able to trigger ADCC and CDC activation, while IgG2 and IgG4 subclasses are not[248].

The establishment of the hybridoma technology almost 50 years ago led to laboratory production of Abs for multiple applications and is one of the most crucial developments in the field of biotechnology[249,250]. When a host organism is immunised with purified target antigen, the Ab-secreting B-cells are challenged and an immune response is triggered, thereby B-cells are activated and start producing Abs. The spleen of the animal is isolated and the activated B-cells are fused with myeloma cells for

immortalisation and, subsequently, generation of a hybrid, clonal, and Ab-producing cell line, or hybridoma[248]. Hybridoma technology is preferable over other methods for the generation of monoclonal Abs (mAbs) due to a comparatively convenient, cost-effective, and limitless production of mAbs[250]. Because of their high specificity and superior pharmacokinetics, Abs have attracted great attention from the pharmaceutical industry, especially since the first anti-tumour mAb drug, rituximab, was approved in 1997[251].

Monoclonal Abs offer various advantageous characteristics such as high specificity and affinity for secreted as well as cell-surface targets, along with their low toxicity and long half-lives[25]. Different (functional) formats of Abs can be used to target circulating antigens, block signalling pathways, cause internalisation and degradation of cell-surface receptors, carry toxic small-molecules to cancer cells (Ab-drug conjugates (ADCs)), recruit immune cells to cancer cells, and more. Immunotherapeutic signalling pathway blockers, for example PD1/PDL1 blockers, can also have lasting positive responses in various cancers, making Abs highly lucrative against cancer[252]. Indeed, cancer dominates the number of clinical trials being run with Abs and, in 2020, more than twice as many cancer trials entered the clinic compared to non-cancer trials[252]. Hence, it is not surprising to see that while there are only three FDA-approved vaccines[20–22], at least 41 mAbs[252] and several mAb-based drugs against cancer have been approved by FDA to date. Ab engineering to control half-life, affinity, biological function, and safety of the Ab is also a vast and an ever-growing area of research[252], which helps develop better Ab-based therapies against cancer.

Most of these Abs are targeting proteins[253]. However, protein antigens have their own pitfalls: to evade being targeted, cancers can downregulate protein expression or mutate the binding epitope so that it is not displayed on the surface anymore and cells become resistant to the Ab/drug[254]. Also, protein TAAs often express in only a fraction of cancer cells and in a low percentage of patients. For example, the protein HER2 is only expressed in approximately 15% of breast cancer cases[255]. In comparison to this, TACAs could prove to be better biomarkers as their expression is not based on the expression of a single enzyme that could be altered by cells, thereby affecting TACA expression in return. Instead, multiple enzymes are involved making cells less likely to develop resistance against therapy targeting TACAs[256]. Furthermore, TACAs generally

have a much higher incidence for certain cancers. For example, sTn is found on more than 90% of colorectal cancers specimens[257], while GB5 and Globo-H are expressed in 77.5% and 61% of breast cancer specimens, respectively[82]. Additionally, TACAs often have a higher expression compared to proteins on the cancer cell surface. For example, it has been estimated that HER2 has about 10^6 copies per cell, while TF is expressed with about 10^7 copies per cell instead[258].

Antibodies against TACAs

With all their advantages over proteins, there are still very few mAbs targeting TACAs in comparison[253]. Developing Abs against glycans is difficult due to several reasons[259]: the enormous structural complexity of the glycans makes extraction of glycans from natural sources strenuous and time-consuming. Glycan antigens are not pure and are usually taken as part of a glycopeptide/glycoprotein[260], glycolipid[261], or simply as surface glycans as part of cell surface[187]. The availability of a sufficient amount of TACAs also poses a big problem[259]. All these problems can be solved by using pure, synthetic, homogeneous glycans produced with automated glycan assembly instead[262]. They can be conjugated to a carrier protein/mixed with an adjuvant and used for immunisation to produce anti-glycan Abs specifically targeting the aberrant glycan only[263].

Abs that do target TACAs specifically and have high relevance in clinic, often have binding affinities in the micromolar range. This is the result of glycans having inherently low immunogenicity which poses a major bottleneck in the development of mAbs. Immunisations often lead to a T-cell independent immune response of plasma cells that produce low-affinity IgM Abs[264,265]. Conjugation of glycans to immunogenic carrier proteins[264], such as keyhole limpet hemocyanin (KLH) or cross-reactive material 197 (CRM_{197})[266], together with the use of adjuvants[267] in immunisation leads to the presentation of processed glycopeptides by MHCII on antigen presenting cells, eliciting a T-cell-dependent immune response. Here, affinity maturation and memory B cell formation will follow, hence, leading to generation of high-affinity IgG Abs[264].

1.3.2. Nanobodies

Almost three decades ago, Raymond Hamers and colleagues at the Vrije Universiteit Brussel discovered that camelids, such as camels, llamas, and alpacas, possess

conventional Abs as well as heavy-chain only Abs (hcAbs) that lack the light chain[268]. The camelid serum contains three IgG subtypes, namely the conventional Ab IgG1 which makes up for 50%, and the hcAbs IgG2 and IgG3, which make up for 30% and 20% of all IgGs, respectively[268,269]. The structure is caused by a genetic deletion of the heavy chain constant domain C_H1, which is responsible for linking the heavy and light chains. The antigen-binding region is, therefore, only made up of a single domain (V_{HH}), which can be expressed recombinantly as a Nanobody (Nb). This means that the paratope consists of only three complementarity determining regions (CDRs1-3) instead of the conventional six with V_H and V_L combined. These are interspaced with four highly conserved framework regions (FR1-4). Furthermore, Nb-CDR3 is on average three amino acids longer than in conventional Abs (15 vs. 12, respectively) allowing Nbs to recognise normally less accessible and hidden epitopes (**Figure 8**)[270]. In comparison, the antigen-binding domain of a conventional IgG can be recombinantly expressed as a single-chain variable fragment (scFv), which consists of both V_H and V_L linked via an oligopeptide and has a size of approximately 27 kDa[270,271].

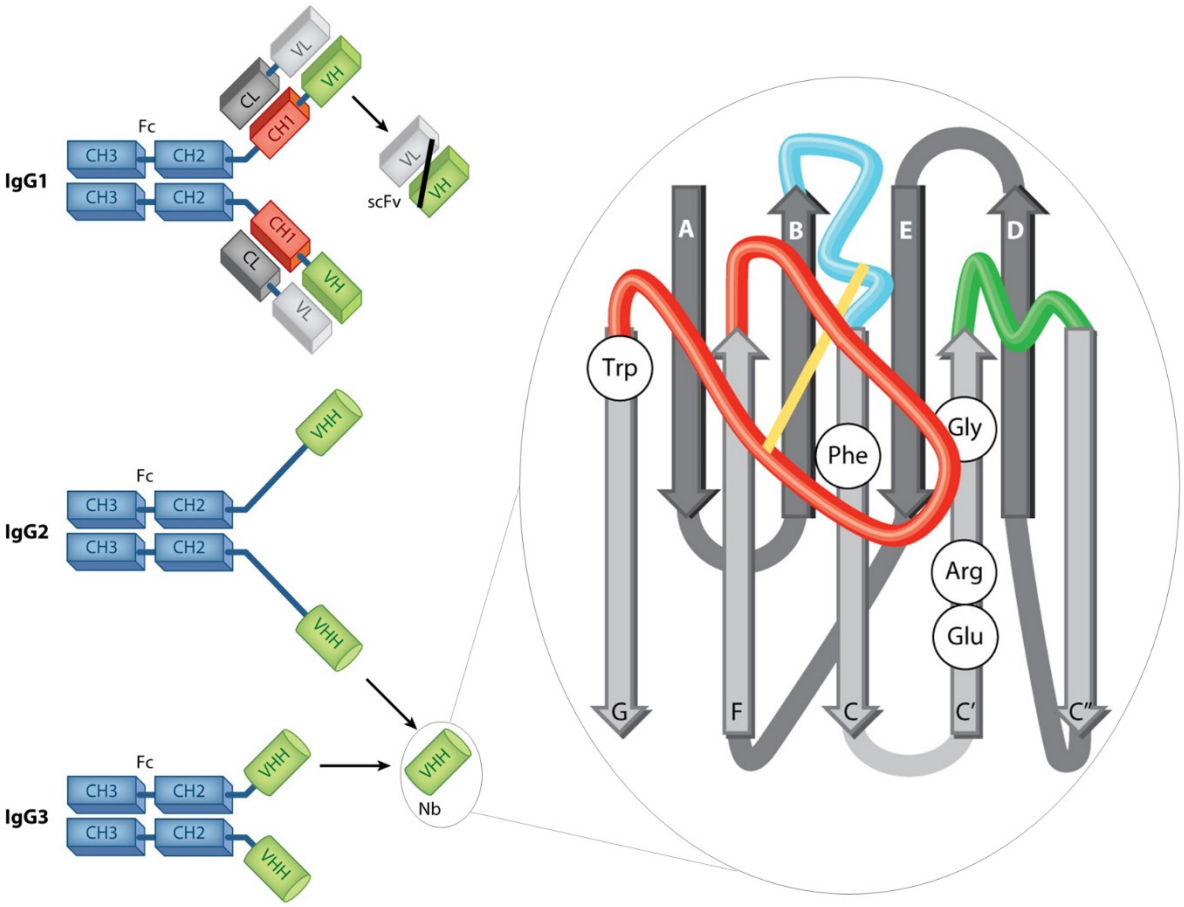


Figure 8: Nanobody origin and structure. *Camelid serum contains three types of IgG. Of these, the conventional IgG1 makes up for 50% of all IgGs, while the hcAbs IgG2 and 3 make up the rest together. Recombinantly expressed, their antigen-binding domains are called scFv or Nb (V_{HH}), respectively. Nb comprises of three variable CDR domains interspaced with four highly conserved FR regions (right panel with zoom in). It contains disulfide bonds (yellow) that stabilize the structure. Some hydrophobic amino acids in the conventional V_H are replaced by polar/hydrophilic amino acids in the V_{HH} as the latter has direct contact with the surrounding solvent. The V_{HH} also boasts a longer CDR3 domain, which helps access cryptic epitopes. CDR1: blue, 2: green, 3: red. Adapted figure*[270].

With a molecular weight of ~15 kDa, Nbs are one-tenth the size of conventional IgGs[272]. This gives them the advantage of accessing dense tumour environments and homogeneously distributing in the tissue, which are important characteristics for both therapeutics and diagnostics[273]. Additionally, the renal clearance for Nbs is rapid, which is advantageous for diagnostics and imaging even if not for therapeutics[273,274]. The serum half-life of Nbs can, however, be increased by the addition of an anti-albumin moiety to the Nb construct that targets endogenous albumin, for example[275]. Furthermore, with such a small size and a single domain only, a bacterial system can be used to easily produce Nbs recombinantly. Therefore, one can obtain high protein

yield at low cost[270]. Quite often, Nbs are also described as highly stable and soluble proteins[276,277]. However, this is not always the case and there have been studies undertaken to improve the stability and solubility of Nbs through directed and random mutagenesis in CDR as well as FR regions[278,279].

As the V_H domain is partly covered by the V_L domain in Abs, it is not as exposed to the surrounding solvent as the V_{HH}. Therefore, V_H and V_{HH} differ in their so-called hallmark amino acids, which in the case of latter are more polar and hydrophilic in nature. These include Val42, Gly49, Leu50, and Trp52 in V_H-FR2, whereas V_{HH} often contains Phe42, Glu49, Arg50, and Gly52 instead[280]. While these mutations provide higher stability to Nbs, stable Nbs also exist with only some or even none of these amino acids[280]. A conserved disulfide bond between FR1 and FR3 (**Figure 8**) is also responsible for Nb stability, while another disulfide bond between CDR1 and CDR3 may come forth in some cases, allowing for higher rigidity in the longer CDR3[270,281].

While Abs are established through hybridomas, Nbs can be generated from sequence libraries that have been generated by high throughput sequencing from the RNA pool of peripheral blood mononuclear cells (PBMCs) isolated from the animal's blood. The process is more economical and much less time consuming than obtaining conventional Abs[270]. Moreover, in the latter case only a sustainable amount of blood from the animal is used, instead of a spleen. Thus, the process is less harmful and more ethical than the hybridoma technology because no animal needs to be sacrificed. Additionally, naïve libraries from non-immunised animals have proven in many cases to be a rich source of various high affinity binders without the need of immunisaton. Furthermore, fully synthetic libraries have been developed that might circumvent the use of any animals in the future[282,283].

The single-domain characteristic of Nbs allows for relatively easy modifications for functionalisation. One of these is "mix and matching" Nbs to obtain a single construct which binds multiple targets (multispecific), several identical targets (multivalent), or different epitopes in the same target[284]. Multimerisation can increase avidity effects and, therefore, apparent affinity by maintaining the k_{on} rate of an interaction but reducing the k_{off} rate[273,285]. Additionally, multivalent constructs of Nbs can improve efficacy by increasing the neutralisation potency in comparison to the Nb monomer[286] and anti-tumour activity in comparison to a conventional Ab[287].

Similar to anti-glycan Ab generation, Nbs targeting glycans face bottlenecks like low glycan immunogenicity and heterogeneous samples (not pure glycan, instead part of a glycopeptide/glycoprotein[260], glycolipid[261], or surface glycans as part of cell surface[187]) for immunisation. Nanobodies generated against glycans so far have been derived from glycoprotein or whole cell immunisation and do not recognise pure glycan epitopes, instead binding either to glycoprotein or heterogenous, uncharacterised polysaccharides[288–290]. Nanobodies targeting Globo-H were previously generated in our lab following immunisation of alpacas with synthetic Globo-H in six rounds (**Figure 9**)[291]. The immunisation was carried out with thiol-linked Globo-H coupled to CRM_{197}. A Nb library was created from mRNA extracted from isolated PBMCs and next generation sequencing was carried out to obtain the library, which was translated *in silico* into corresponding protein sequences. In parallel, the alpaca serum was used to isolate Globo-H targeting hcAbs with synthetic amine-linked Globo-H immobilised on beads. The different linkers in immunisation and hcAb purification helped to avoid isolation of Abs that would target the linker instead of the glycan. Purified hcAbs were analysed with LC-MS/MS and compared with the established Nb library to obtain 36 potential candidates for Globo-H targeting Nbs[291].

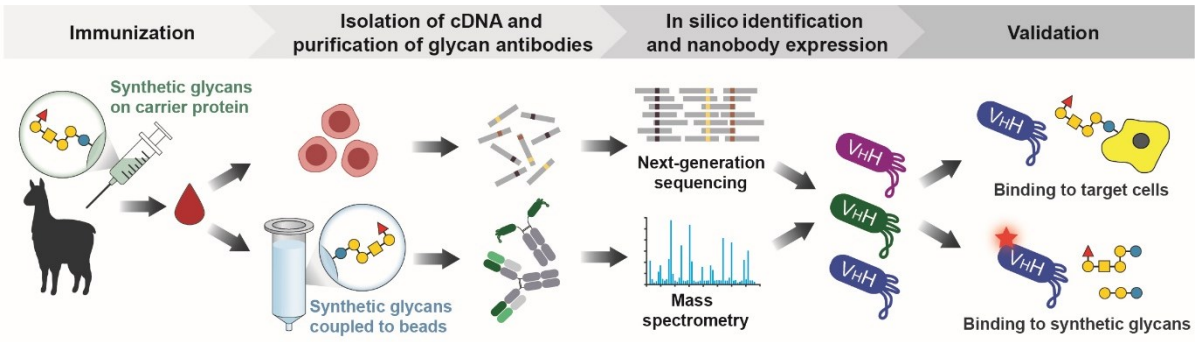


Figure 9: Schematic overview of the workflow for generating Globo-H-targeting Nbs. *After alpacas had been immunised with Globo H, their blood was extracted and used to isolate hcAbs against immobilised synthetic Globo-H as well as PBMCs for RNA extraction. V_{HH} from hcAbs were analysed with LC-MS/MS. Meanwhile, a Nb library was generated from mRNA. These two were compared and potential Globo-H binders were identified. The plasmids expressing the selected Nbs were ordered. As part of my thesis, I tested the Nbs for Globo-H-binding. Adapted figure*[291].

1.4. Aims of the Study

TACAs, such as sLe A and Globo-H, play key roles in cancer progression and survival, and, therefore, make excellent biomarkers for cancer diagnosis and therapy. However, Abs targeting glycans are not frequently generated because TACA-specific Ab development faces multiple challenges.

Cancers like pancreatic cancer have high mortality rates, due to the often-late diagnosis. While Abs targeting sLe A exist, Abs with higher specificity and binding affinity that might hopefully help in earlier detection are needed. With that in mind, the goals for the first part of my thesis (**Figure 10A**) were:

I. **to characterise murine Abs generated against synthetic CA 19-9/sLe A, which would potentially be more specific and stronger binders than the FDA-approved and clinically used 1116-NS-19-9 and**
II. **to chart the binding epitopes and paratopes of the Abs.**

Due to their size and single-domain characteristic, Nbs have the advantages over Abs that they can enter dense tumour microenvironments, distribute evenly in the tissue, and access hidden targets. This makes Nbs highly lucrative for cancer therapy, but methods to develop glycan targeting Nbs are so far lacking. Previously, synthetic Globo-H had been used to immunise alpacas and obtain Nbs[291] which had not yet been characterised (**Figure 9**). The goals for the second part of my thesis (**Figure 10B**) were:

I. **to test Nbs for Globo-H binding and determine the best binder for synthetic and native Globo-H,**
II. **to chart the binding epitope of the best binder, GH46, among the Globo-series,**
III. **to determine the binding affinity of GH46 for Globo-H, and**
IV. **to functionalise GH46. For this, I aimed to generate a trivalent construct that would show higher binding affinity than the monomer.**

While working with GH46, I realised that the protein is not stable. The Nb expresses well, but is highly prone to aggregation and, therefore, has a low end yield. Also, it

cannot be concentrated above ~1 mg/mL as, again, the protein tends to aggregate. Furthermore, GH46 has a binding affinity for Globo-H in the high micromolar range. Therefore, a my final goal developed into:

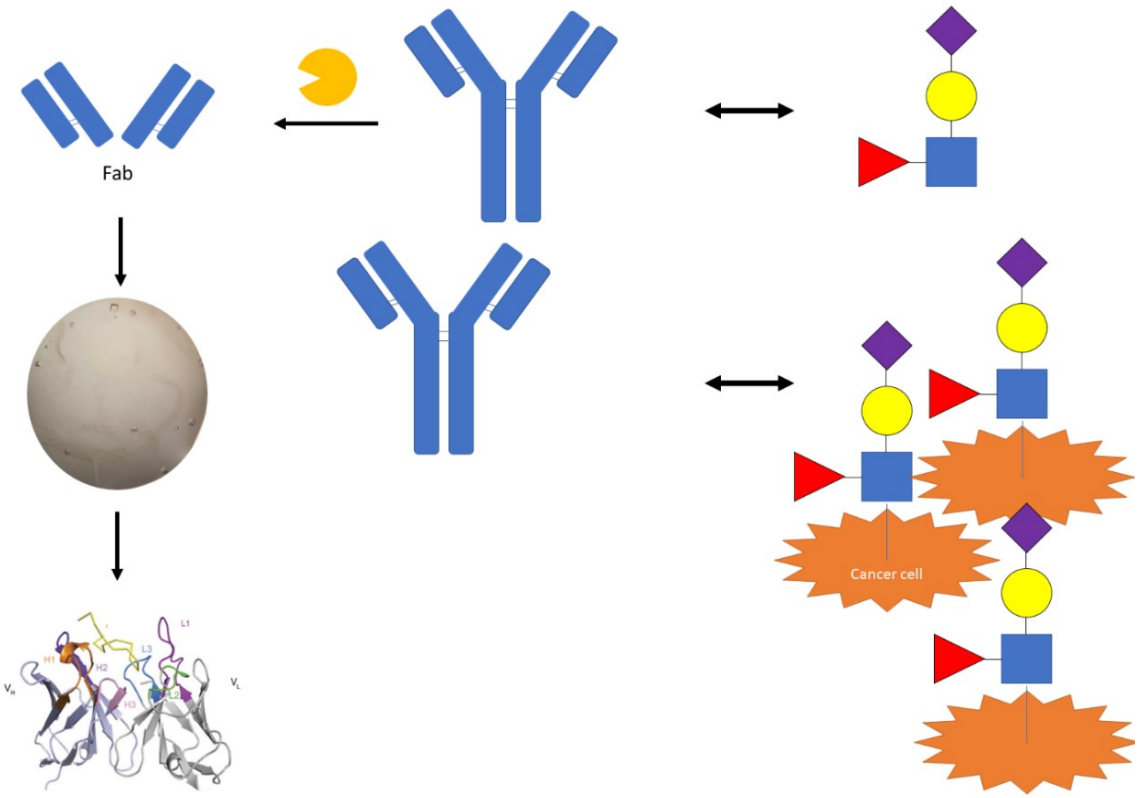
V. **studying the amino acid sequence of GH46 and introducing mutations to improve solubility, thermostability, and binding affinity for Globo H.**

With this, I wanted to address the following questions:

1. What mutations could be carried out to improve the Nb solubility and thermostability?
2. Would the FR mutations affect binding of Nb to Globo-H?
3. Would the Nb retain its specificity for Globo-H?
4. Could directed mutations in the CDRs improve the binding affinity of GH46 to Globo-H?

Because Abs and Nbs against TACAs are rare and their generation is challenging, ultimately, this thesis aimed to characterise potentially good binders for sLe A and Globo-H to iterate the benefits of using pure, synthetic glycans for immunisation in cancer therapy. On one hand, I present two Abs targeting sLe A with higher binding to the glycan than 1116-NS-19-9, which is currently used against sLe A for pancreatic cancer diagnosis. On the other hand, I present a novel glycan-only binding Nb that specifically recognises Globo-H. With a straightforward mutation strategy, I also provide evidence for the importance of FR3 in the Nb stability paving the way for better structural understanding of Nbs.

A



B

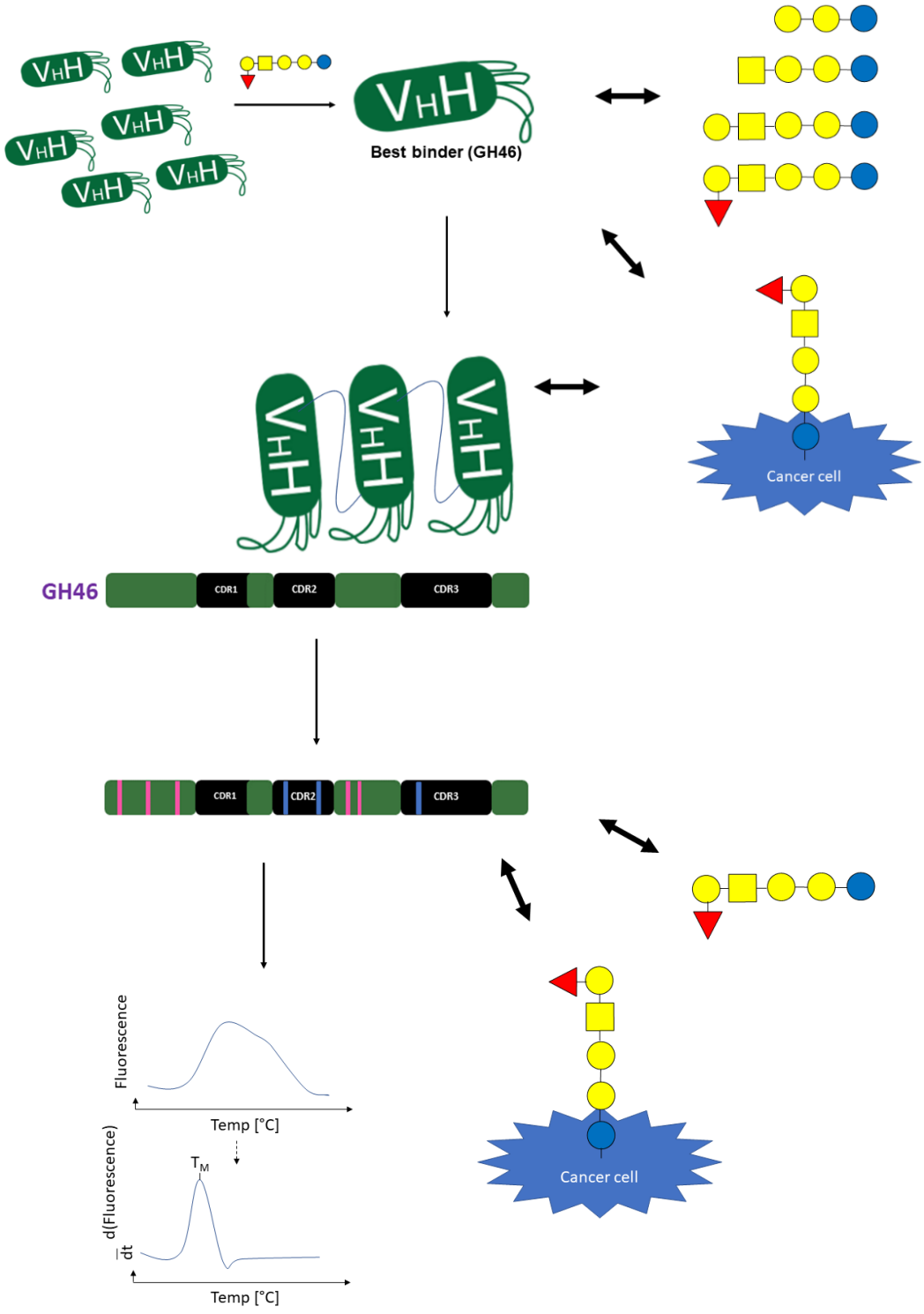


Figure 10: Aims of the study. A. *For the first part of the thesis, I aimed to determine the binding affinity and specificity of mAbs to sLe A in various methods. The binding to native glycan also needed to be analysed. Secondly, I aimed to map the binding paratope and epitope of the mAbs.* ***B.*** *In the second part of the thesis, my goal was to determine the best binder of Globo-H from a pool of Nbs and to test its binding to Globo-H to calculate binding affinity and chart binding epitope. Additionally, the Nb was to*

be functionalised to a trivalent construct that would bind better to Globo-H than the monomer. Furthermore, directed single mutations in the FRs and CDRs would be carried out to generate a final construct with better thermostability, solubility, yield, and binding affinity that would, however, still retain binding specificity for Globo-H.

2. Materials and Methods

The materials used in the experiments are mentioned in the methods.

2.1. CA 19-9 in pancreatic cancer

2.1.1. Expansion of hybridoma cultures

Frozen stocks of hybridoma cells at a concentration of $2x10^6$/mL in RPMI + 10% FCS + 10% DMSO were taken from Katrin Sellrie. For a new culture, cells were thawed at 37°C for 30 s and immediately pipetted into ISF-1 + 10% FCS to wash away DMSO. They were centrifuged at 300 g/RT for 5 min after which the supernatant was removed. The cell pellet was diluted in 10 mL ISF-1 + 10% FCS and grown at 37°C/5% CO_2. ISF-1 media with decreasing amounts of FCS (8%, 6%, 4%, 2%, and 0%) were prepared and used to decrease FCS in the hybridoma cultures over the next few days. For GB11, the decrease was smooth, and FCS was completely removed in 5 days, however, HA8 was more sensitive to the decrease and needed to be weaned off slower. It took 10 days to remove FCS completely from HA8 culture. The cultures were then passaged every other day according to their cell count. Small flasks were maintained at ~10^6/mL, while the cultures were expanded to 300 cm^2 flasks. These were allowed to grow, and cells were observed for 10-14 days. Once most of the cells were dead, the cultures were centrifuged at 1,200 g at RT for 20 min. The supernatant containing Abs were decanted and filtered with 0.2 µm filters. These were stored in the cold room with 0.02% NaN_3, till Ab purification could be carried out, to avoid contamination.

2.1.2. RNA isolation and sequence determination of GB11 and HA8

The RNA isolation from 10^7 hybridoma cells of clone GB11 and HA8 was carried out by Anika Freitag using RNeasy MiniKit (Qiagen, Hilden, Germany). Complementary DNA (cDNA) was synthesized via PCR using the RNA as template by Jost Lühle. Chain specific primers for mouse IgG heavy chain, kappa light chain, and lambda light chain were used along with template-switch oligo (AAGCAGTGGTATCAACGCAGAGTACATrGrGrG, where r stands for RNA base) and a 3' adaptor sequence was added for the subsequent sequencing. For reverse

transcription, the SuperScript IV Reverse Transcriptase system (Invitrogen) was used. The PCR products were run on a 2% agarose gel and purified with NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel, Germany). The fragments were blunt end ligated with the pCRZeroT plasmid (a gift from Ken Motohashi), which was previously digested with SmaI enzyme (NEB). The ligated plasmids were transformed into chemically competent *E. coli*, DH5α strain. Colonies with successful transformation were picked and sent for sequencing (LGC genomics) with M13 standard primers to obtain the sequences of variable domains.

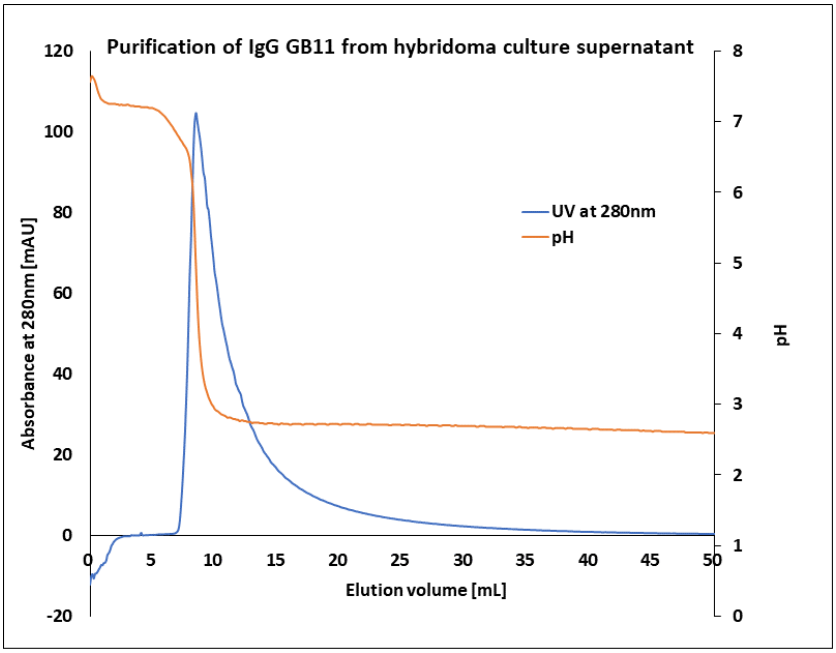
2.1.3. Monoclonal Antibody Purification

5 mL Pierce Protein A/G Chromatography cartridge (Thermo Fischer; Rockland, IL, USA) was equilibrated with 10 column volumes (CV) of binding buffer (25 mM sodium phosphate pH 7.4, 150 mM NaCl) using Äktapurifier UPC10 System (GE Healthcare; Uppsala, Sweden). 2 L of hybridoma culture supernatant were run over the column in a loop at a flowrate of 0.5-1.0 mL/min overnight so that most of the mAb would bind. The column was washed again with 10 CV of binding buffer and then 5 CV of 80% binding buffer + 20% protein A/G elution buffer (100 mM Glycine-HCl, pH 2.7) to remove any non-specifically bound proteins. The mAb was eluted with 15-20 CV of protein A/G elution buffer at a flowrate of 5 mL/min. Eluate was collected in 5 mL fractions containing previously determined volume (155 µL) of neutralization buffer (1 M Tris-HCl, pH 9) already. The final pH of the fractions was 7.4. Supernatant was run over the column once more to ensure all Ab had been isolated. As example GB11 purification is shown in ***Figure 11***. The column was washed with neutralisation buffer, water, binding buffer, and then water again to ensure all protein from the cartridge was removed. It was stored in 20% ethanol.

Fractions containing eluted mAb were collected and concentration using Amicon® Ultra-4 filter with a molecular weight cut-off (MWCO) of 30 kDa (Merck Millipore; Tullagreen, Ireland). For glycan array, isothermal calorimetry, surface plasmon resonance immunohistochemistry, and cell binding assays, the purified Abs were used directly, however, for crystallisation the Abs were first digested with Ficin, the digested product was purified, and then crystallisation plates were set up with the Fab region only. Depending on the following experiment, the Ab was either buffer exchanged to 1x PBS, pH 7.4, or Ficin digestion buffer (see below) with HiLoad 16/600 Superdex

200 pg (Cytiva). Eluate was collected in fractions of 2 mL. To ensure Ab purity, sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE; see **section 2.2.3** for details) was run to determine which fractions should be pooled together. Pooled fractions were concentrated using Amicon® Ultra-4 filter (MWCO 30 kDa) and protein concentration was determined with NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, MA, USA) using the settings for mouse IgG mass extinction coefficient (E1% = 14). The mAb was stored in the cold room.

A



B

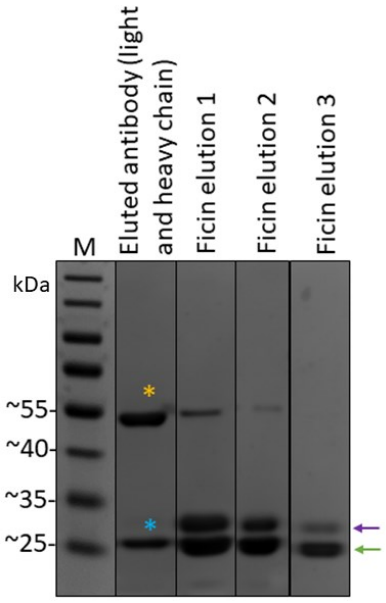
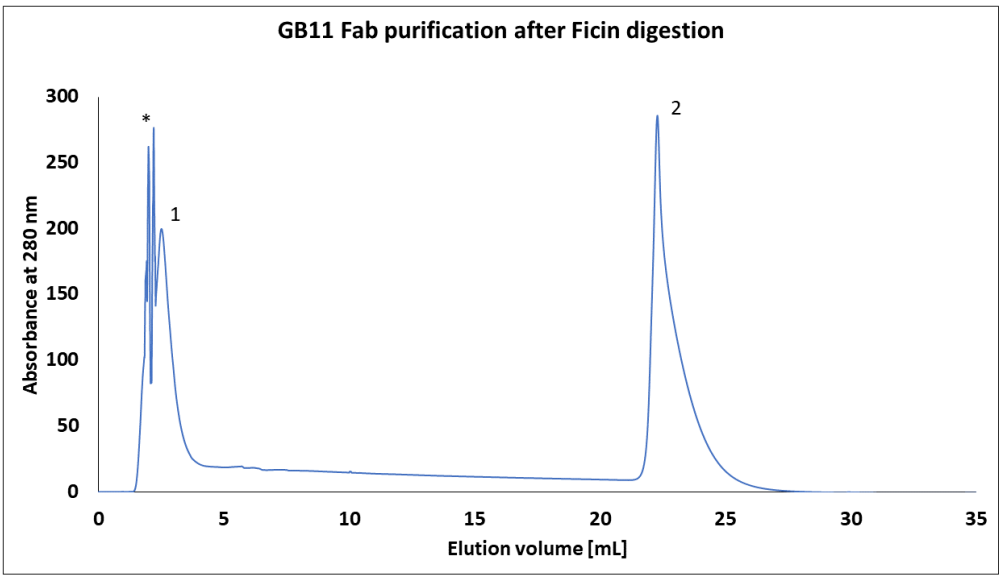


Figure 11: IgG purification. A. *GB11 purification on the A/G column is shown as example. Ab started eluting as pH dropped to 3. It was collected directly in the neutralisation buffer to ensure Ab was around pH 7 immediately after elution.* ***B.*** *Fractions of eluted Ab were collected and concentrated before being analysed with SDS-PAGE. The gel is shown. Ficin digestion was subsequently carried out three times to ensure most of the Ab was digested. M: Prestained protein ladder as marker; yellow asterik (*): Ab light chain; blue *: Ab heavy chain; purple arrow: Fc region of digested Ab; green arrow: Fab region of digested Ab.*

2.1.4. Ficin digestion

For the subsequent Ficin digestion, purified and concentrated mAbs were injected onto a HiLoad 16&600 Superdex 200 pg (Cytiva) column for buffer exchange to Ficin digestion buffer (100 mM Na-Citrate pH 6, 6.36 mM EDTA). Fractions containing mAb were concentrated. To this 36.2 mM end concentration cysteine was added and incubated O/N with equilibrated Ficin immobilised on agarose beads at 37°C. The Fab product from the digestion was purified from the mixture with the protein A/G column (**Figure 12**). The flow through was collected separately, while the bound F_c region was eluted with protein A/G elution buffer. To ensure most of the Ab was digested, three rounds of digestion were carried out. While pure Fab was obtained, some of the Fab protein eluted with the Fc region (peak 2, **Figure 12B**). This could be denatured protein that stuck to the column but was eluted with the acidic elution buffer along with the Fc region. The Fab fractions were buffer exchanged to the crystallization buffer (10 mM Tris-HCl pH 7, 150 mM NaCl) either on the HiLoad 16&600 Superdex 200 pg (Cytiva) column or overnight with dialysis at 4°C.

A



B

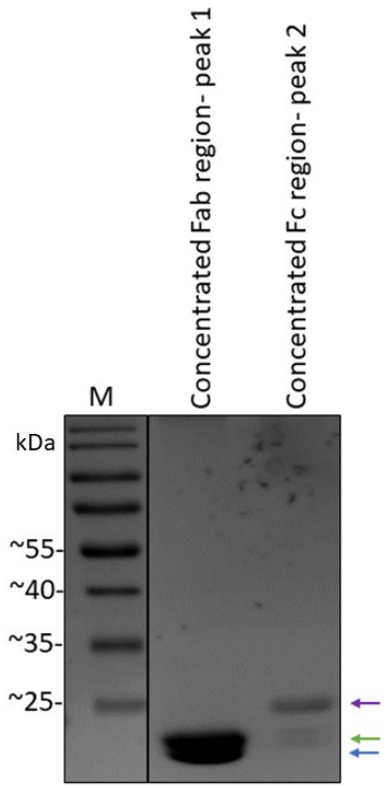


Figure 12: ***IgG-Fab purification after digestion. A.*** *GB11 was digested thrice with Ficin and the Fab (peak 1) and Fc (peak 2) regions separated using the protein A/G chromatography cartridge. The peaks were collected and concentrated.* *: air-bubble. ***B.*** *Concentrated peak 1 shows pure Fab region of GB11* in SDS-PAGE. *Peak 2 shows concentrated Fc region with some contamination with Fab protein. M: prestained protein ladder as marker; purple arrow: Fc region; green arrow: Fab heavy chain; blue arrow: Fab light chain*

Following the purification, Abs were used for binding specificity and affinity analyses, while Fabs were used to study the protein structure and interactions with sLe A.

2.1.5. Glycan array

Glass slides with synthetic glycans (0.1 mM) printed on them were prepared by Katrin Sellrie as previously described[292]. The printed pattern is shown in **Figure 13**. Anika Freitag generated the raw data, which was analysed by me.

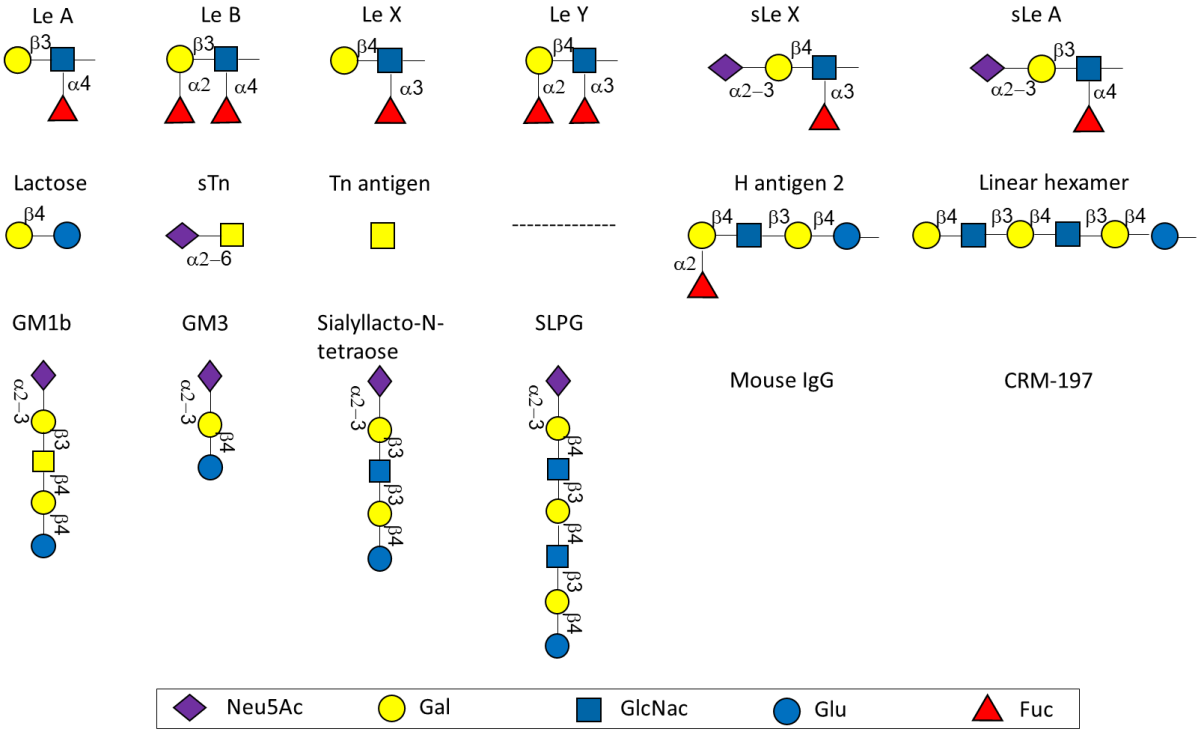


Figure 13: Glycan array printing pattern. *0.1 mM of glycan was printed on the glycan array glass slides per well in the following pattern: (Upper right to left) Le A, Le B, Le X, Le Y, sLe X, sLe A; (middle right to left) Lactose, sTn, Tn, empty, H antigen 2, a linear hexamer; (lower right to left) GM1b, GM3, sialyllacto-N-tetraose, sialosyl lactosaminyl paragloboside (SLPG), an unrelated mouse IgG to confirm secondary binding, CRM_{197} to ensure GB11 and HA8 target glycan only and CRM_{197} is not part of the binding epitope. Each line of glycans was repeated twice.*

All steps were carried out at 37°C unless stated otherwise. For each step, 50 µL per well of solution were used. The slide was first blocked with PBS + 3% BSA per well for 1 h. The wells were washed once with PBS and then incubated with 5 µg/mL of GB11, HA8, or 1116-NS-19-9 (OriGene Technologies, Inc., #CF190083) in PBS + 1% BSA at 400 rpm for 1 h. Following this, the slide was washed twice with PBS for 15 min. Next, the secondary goat α-mouse-IgG (H+L) Ab Alexa Fluor™ 635 (Invitrogen by Thermo Fisher Scientific) was used in a dilution of 1:500 in PBS + 1% BSA and incubated for

1 h. The wells were washed twice again with PBS for 5 min each. Finally, the 96-well grid was removed and the whole slide dipped several times into ddH_2O in a 50 mL falcon. The slide was dried by centrifugation at 300 g for 5 min at RT and was scanned using Glycan Array Scanner Axon GenePix® 4300A (Molecular Devices, LLC, San Jose, CA, USA). The glycan array was analysed for binding using GenePix Pro7 (Molecular Devices, LLC). At least three independent replicates were carried out.

2.1.6. Isothermal Calorimetry (ITC)

The mAbs and 3'-sLe A (BIOSYNTH Carbosynth, OS00745) were diluted in PBS buffer, pH 7.4. For each experiment 300 µL of Ab were loaded into the sample cell. Initially, 20 µM of Ab were tested and it was established that a molar ratio of Ab to sLe A of 1:20 to 1:24 gave the best results. The concentration of the Ab was taken down to as low as 12.5 µM for subsequent replicates. The sLe A was titrated into the Ab-containing cell with the syringe via 18 injections. Isothermal Calorimetry (ITC) measurements were carried out on an iTC200 instrument (MicroCal) in the AG Wahl lab (FU Berlin). The first injection was 4 µL, followed by 2 µL each. Experiments were conducted at RT. For both mAbs, at least three repetitions were performed.

To analyse the data, MicroCal PEAQ-ITC Analysis Software was used which gives out biophysical data to analyse the binding. The change in enthalpy (ΔH) and change in entropy (ΔS) can be used to calculate the change in free Gibbs energy (ΔG) using the following equation:

$$\Delta G = \Delta H - T\Delta S, \qquad \text{Eq. 1}$$

where T is absolute temperature (298 K for the ITC experiments). A negative ΔG value shows that the reaction taking place is spontaneous in nature.

2.1.7. Surface Plasmon Resonance (SPR)

For the Surface Plasmon Resonance (SPR), the Biacore T100 (GE Healthcare) was operated using the Biacore T200 control software. Mouse Antibody Capture Kit (GE Healthcare) was used according to manufacturer's protocol to first immobilise ~ 8,000 RU of α-mouse IgG capture Ab on two flow cells on a CM5 chip. One of these was used for Ab immobilisation and the other as a 'blank'-immobilised flow cell for reference to compensate for unspecific binding of glycan to sensor chip surface and

α-mouse IgG capture Ab. Immobilisation of Abs and glycan binding assays were performed in PBS at 25°C. Approximately 400 RU of mAbs were captured at either a concentration of 50 µg/mL for IgG GB11 and HA8 or 1:100 dilution factor for 1116-NS-19-9 (Thermo Fisher #MA5-12421). The surface contact time was 180 s with a flow rate of 30 µL/min and a stabilisation period of 60 s to allow any uncaptured Ab to be washed away. Following this, sLeA was run over the immobilised Abs in 2 cycles with increasing concentrations ranging from 0.375 µM to 100 µM so that 10 different concentration points could be measured. The highest concentration of the first cycle was taken as lowest concentration of the second and both cycles were plotted on the same graph using single cycle kinetics. Parameters used were contact time of 60 s and dissociation time of 180 s at a flow rate of 30 µL/min. Flow cells were regenerated with 10 mM glycine-HCl pH 1.7 with a contact time of 120 s at a flow rate of 20 µL/min, and a stabilisation time of 300 s was added to ensure the baseline returned to its original value and all Ab molecules were removed. As negative control, glycan was passed over the anti-mouse Ab only. For all Abs, at least 5 repetitions were carried out. Analyses were carried out using the Biacore T200 evaluation software (GE Healthcare). As the on- and off-rates were outside of the measurable ranges of the instrument, the affinity (K_D) values were determined using a steady-state affinity model.

2.1.8. Cell culture

B16 and B16-FUT3$^+$ cells were a gift by Prof. Dr. J. V. Ravetch[293]. B16-FUT3$^+$ have been modified to express FUT3 gene and, therefore, the enzyme FUT3, which allows for surface expression of sLe A. For the cell culture, the Dulbecco's Modified Eagle Medium (DMEM; PAN™ Biotech) was separately prepared for both cell lines. This was used supplemented with 10% FCS (PAN™ Biotech), 2 mM/mL L-glutamine (PAN™ Biotech), and 10 U/mL penicillin/10 µg/mL streptomycin (PAN™ Biotech). As selection antibiotic for B16-FUT3$^+$, 500 µg/mL geneticin (Gibco™) was also added to its growth medium. B16 and B16-FUT3$^+$ were passaged twice a week at 1:20 and 1:25, respectively. Cultures were maintained at 37°C and 5% CO_2.

2.1.9. Flow cytometry

The raw data for the cell binding assay was generated by Anika Freitag using a previously described protocol[291] and analysed by me. Briefly, cells were trypsinised

(using Trypsin/EDTA; PAN™ Biotech) and counted. 1 x 10^6 cells per sample were taken for the binding assay. These were washed with PBS and transferred to 1.5 mL Eppendorf tubes. The samples were centrifuged for 5 min at 300 g/RT and the PBS removed. The cell pellets were taken up again in PBS + 5% BSA containing 5 µg/mL of GB11, HA8, or 1116-NS-19-9 (OriGene Technologies, Inc., #CF190083) and incubated for 1 h/RT/400 rpm. As negative control cells were incubated in PBS + 5% BSA only. The cells were centrifuged at 300 g/RT and washed twice with PBS. They were incubated for 1 h/RT with secondary goat α-mouse-IgG (H+L) Ab Alexa Fluor™ 635 (Invitrogen by Thermo Fisher Scientific) in a dilution of 1:500 in PBS + 5% BSA. Cells were washed thrice with PBS and finally resuspended in 200 µL of PBS for analysis.

Stained cells were measured on a FACSCanto™ II device (BD Biosciences, San Jose, CA, USA). The main cell population was selected for intact cells via forward and sideward scatter amplitude (FSC-A and SSC-A) and events were gated for single cells via FSC-Width (FSC-W) and FSC-Height (FSC-H) using the software FlowJo (v10.8.1, FLowJo LLC, Ashland, OR, USA). For qualitative data analysis, histograms of the allophycocyanin (APC; for the wavelength 635 nm) channel were generated (**Figure 14**). Cell binding was quantified by exporting mean fluorescence intensity (MFI) of the APC-A signal for at least three independent experiments and comparing it to examine differences.

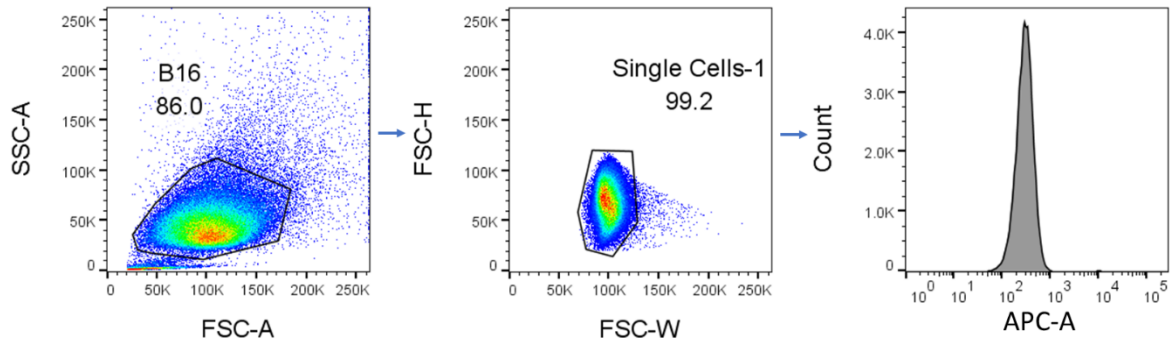


Figure 14: Gating strategy for flow cytometry. *As an example, B16 cells are shown. First, the SSC-A and FSC-A range is established ensuring all populations are clearly shown in the dot plot. Then, for the first gating, small contaminants and cell debris are gated out by selecting the main cell population in SSC-A vs. FSC-A dot plots (left). Gating for single cells only is carried out in FSC-H vs. FSC-W dot plots (centre) and the resulting population is displayed as a histogram of the APC-A (right) or used for MFI analyses.*

2.1.10. Crystallisation

To establish crystallisation conditions, 4x 96-well plates were initially setup with various crystallisation conditions for both GB11-Fab (c = 12.4 mg/mL) and HA8-Fab (c = 9.8 mg/mL). 0.2 µL of sitting drops were added to the wells using Oryx4 pipetting robot (Douglas Instruments). Plates were stored at RT and viewed every other day till crystals could be seen. HA8 aggregated in the preliminary crystallization steps and did not give crystals. For GB11-Fab, the conditions giving best crystals were set up in 48 well plates (condition 1: 2.06 M DL-Malic acid pH 7.0, 2 mM citric acid pH 3.5, 0.5 % PEG3350; condition 2: 1.6 M sodium malonate pH 7, 30 mM Hepes pH 7.0, 8.3 % PEG3350). 2 µL of GB11-Fab were mixed in a 1:1 ratio in the two conditions in the hanging drop method. Multiple crystals in each condition typically appeared within five days. For the apo structure, crystals were fished and directly frozen in liquid nitrogen. For the holo structure, the crystals were incubated for seven days in mother liquor supplemented with 5 mM sLeA, and then frozen in liquid nitrogen.

Diffraction data was collected and processed by Michael Krummhaar, who also solved the structure and created the figures as part of a collaboration with Dr. rer. Nat. Christian Roth, Max-Planck Institute of Colloids and Interfaces, Potsdam, Germany. The method along with other methods from various collaborators is described in the supplementary information (**section 5.1.1**).

2.2. Globo-H in breast cancer

2.2.1. Cloning

One of the functionalisations attempted for GH46 was fusion with the lytic peptide Tachyplesin I (from *Tachypleus tridentatus*). Bacteria transformed with the cloned plasmid, however, did not express the construct. To test whether multimerising the Nb would produce protein, a pET-28b(+) plasmid containing the sequence of 6xHis-tag-GH46-trimer-Tachyplesin I was commercially ordered from Synbio Technologies LLC, USA. However, here too, no protein expression was detected. While the mini-project was discontinued, the plasmid was used as template to produce the GH46-trimer alone instead. To do so, a stop codon sequence (TAATAA) was introduced at the C-terminal of the Nb before the peptide sequence.

To amplify the Nb46-trimer sequence by PCR, T7 forward primer and a designed reverse primer with the stop codon primer were used with the Q5 High-Fidelity DNA Polymerase (New England Biolabs, USA). The reaction mixture was mixed on ice and included 1x Q5 Reaction buffer, 1x Q5 High GC Enhancer, 200 µM dNTPs (New England Biolabs, USA), 0.5 µM of each primer, 8 ng plasmid DNA, and 0.01 units of Q5 polymerase. The PCR reaction conditions are listed in table 1. The 1448 bp PCR product was analysed on a 1% agarose gel, extracted using the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel, Germany), and inserted back into the original pET-28b(+) vector through restriction and ligation. For this, the insert and vector were digested with *Xba*I and *Xho*I (New England Biolabs, USA) at 37°C for 25 min. The restricted insert and vector products (1385 bp and 5192 bp, respectively) were run on a 1% agarose gel and extracted. For the ligation, 1:3 ratio of insert to vector was used. For the ligation, T4 DNA ligase (New England Biolabs, USA) was used for 10 min at RT and then inactivated at 65°C for 10 min. The mixture was transformed into DH5α cells as described above and purified DNA was sequenced (Eurofins Genomics) to confirm that the Tachyplesin I sequence was removed.

For the molecular cloning of GH46 mutants, either the pET-28b(+) plasmid containing GH46-6xHis-tag or the commercially ordered pET-28b(+) plasmid containing GH46 mutant 9 (M9; Synbio Technologies LLC, USA) was used as a template. M9 is described in detail in **section 3.2.7**. Briefly, 8 mutations were introduced into the original GH46 sequence, of which 3 were in FR1, 2 in CDR2, 2 in FR3, and 1 in CDR3. Eight further mutants (M1-8) were generated in total via cloning. The FR 1 mutants are M1 with Gln5Val, M2 with His13Gln, M3 with Val23Ala, and M4 with a combination of Gln5Val, His13Gln, and Val23Ala. M5 and 6 include FR1 mutations along with FR3 mutations Met64Val only, or Met64Val and Gln62Asp, respectively. M7 and 8 include all FR mutations along with CDR3 only (Asp108 deletion), and CDR2 (Asp57Ser and Tyr59Asn substitution into sequence), respectively, while M9 contains all eight FR and CDR mutations.

The PCR reaction mix was set up similar to GH46-trimer. Overhang extension PCR was applied[294]. Briefly, two PCR products were amplified separately, one with T7 forward and a designed reverse primer, and the other with a designed forward and a T7 reverse primer. These were then used as templates for the final overlapping PCR

amplification step with T7 forward and reverse primers. All products were analysed on a 2% agarose gel and the gel extracted inserts were digested along with pET-28b(+) vector backbone with *Xho*I and *Xba*I overnight at 37°C. In addition, to prevent re-ligation, the linearised plasmid was digested for 1 h at 37 °C with Antarctic phosphatase (NEB). For ligation 1:7 ratio of insert to vector were used. Ligation mixture was then transformed into DH5α. Cloning was confirmed by sequencing (LGC Genomics, Germany). The primers and the cloning strategy were designed by me, and the experiments were carried out either by me or by David Warschkau, Zeinab Fandi, or Sophia Schrinner as part of their master's projects supervised by me.

Table 1: List of primers used for the overhang extension PCRs for cloning of GH46 mutants M1-8.

Protein of interest	Forward primer	Reverse primer
M1	ATGTTCAACTGGTTGAATC TGGCGGC	GCCGCCAGATTCAACCAGT TGAACAT
M2	TCTGGTTCAACCGGGCGGT	ACCGCCCGGTTGAACCAGA
M3	TCTGTCTTGCGCAGCGAGC GGCTTT	AAAGCCGCTCGCTGCGCAA GACAGA
M4	CTACACCCAAAGCATGAAA GGCCGTTTCACCA	GGCCTTTCATGCTTTGGGTG TAGTAGGTATCATCG
M5	ACACCCAAAGCGTCAAAGG CCGTTTC	ACGCTTTGGGTGTAGTAGGT ATCATCG
M6	ACACCCAAAGCGTCAAAGG CCGTTTC	ACGCTTTGGGTGTAGTAGG TATCATCG
M7	AACTGGAACGGCGATGATA CCTACTACA	CTATCGGTGTAGTAGGTATC ATCGCCGTTC
M8	AATTGGGTTTCTGGCGATTA TGGTCTGG	AATAATCCAGACCATAATCG CCAGAAACC

2.2.2. Transformation of DH5alpha and Shuffle cells

The pET-28b(+) plasmids containing Nb sequences fused to a C-terminal 6xHis-tag were transformed into 50 µL of the chemically competent *E. coli,* DH5 α (NEB). The

6xHis-tag lies on the N-terminal of the commercially ordered pET-28b(+) of the GH46-trimer. Cells were kept on ice for 30 min before they were heat shocked at 42°C for 45 s and incubated for 2 min on ice again. 950 µL of lysogeny broth (LB) medium were added to the cells, which were then incubated for 1 h at 37°C/800 rpm. Cells were spun down at 3000 g for 1 min before 850 µL of supernatant were removed. Pelleted cells were resuspended in the remaining 150 µL solution and plated onto LB agar plates containing 50 µg/mL kanamycin. Plates were incubated overnight at 37°C. Single colonies were inoculated in 5 mL LB medium with 50 µg/mL kanamycin overnight at 37°C again. Culture was centrifuged at 3000 g for 1 min and pellet was either directly used for plasmid extraction or stored at -20°C. Plasmid DNA was extracted with a NucleoSpin Plasmid kit (Macherey-Nagel) and measured for concentration by A_{230} with NanoDrop® ND-1000.

The extracted plasmid was then transformed into chemically competent *E. coli*, SHuffle cells (NEB) following the same procedure, except incubations were carried out at 30°C instead of 37°C. After overnight incubation of plate, colonies were scraped together and inoculated into 10 mL of LB medium with 50 µg/mL kanamycin, which was grown at 30°C/280 rpm overnight.

2.2.3. Expression and Purification of nanobodies

Culture expression

The overnight pre-culture was diluted into 1 L of Terrific Broth (TB) medium supplemented with 50 µg/mL of kanamycin and grown at 30°C/280 rpm until the OD_{600} of 1.5-1.7 was reached. After this, cultures were cooled down to 16°C and induced with 0.4 mM isopropyl-β-D-thiogalactoside (IPTG) overnight. Cells were harvested at 6,000 rpm for 15 min at 4°C and pellets were either stored at -20°C or purified directly. For the protein yield comparison of GH46 mutants, 100 mL of cultures was grown in TB medium and induced at an OD_{600} of 1.5 with 0.4 mM IPTG.

Lysis of Shuffle cells

Pellets were resuspended in 3 mL sample buffer (40 mM sodium phosphate pH 8, 150 mM NaCl, 10 mM imidazole) per 1 g of wet pellet supplemented with protease inhibitor mix HP (1:100 dilution; SERVA, Germany) and Dnase I (1:10,000 dilution; New England Biolabs, USA). Cells were then lysed by passing through the French press

(Emulsiflex C3, Avestin Europe GmbH, Germany) twice at 10,000-15,000 psi. Lysed cells were centrifuged at 42,000 g for 45 min at 4°C. Lysate was filtered using 0.22 µm filter and purified. Samples of the pellet and lysate were taken for SDS-PAGE analysis.

Purification of nanobodies

A two-step purification strategy was implied which included affinity purification followed by size exclusion chromatography (SEC). For the affinity purification, 1 mL of Ni^{2+}/NTA beads per 1 L of bacterial culture volume was loaded into a gravity chromatography column and equilibrated with 3 CV of sample buffer. The filtered lysate was added to the column and incubated with the beads for 1 h at 4°C on a rotating wheel. Flow-through was collected and the beads with immobilized protein was washed with 20 CV of sample buffer, 20 CV of wash buffer 1 (40 mM sodium phosphate pH 8, 1 M NaCl, 10 mM imidazole), and 20 CV of wash buffer 2 (40 mM sodium phosphate pH 8, 150 mM NaCl, 30 mM imidazole). All samples were collected and analysed later with SDS-PAGE. The beads were incubated with 10 CV of elution buffer (40 mM sodium phosphate pH 8, 150 mM NaCl, 300 mM imidazole) for 10 min at RT. Another 10 CV of elution buffer were added and collected without incubation time to wash out any remaining protein in the column. The eluate fractions were collected and concentrated with an equilibrated AmiconR Ultra MWCO 3 kDa at 3200 g/4°C. Concentration was determined by measuring UV absorbance at 280 nm using NanoDrop ND-1000 Spectrophotometer. To prevent aggregation of GH46, the protein was only concentrated till 0.9 mg/mL. For the mutants, however, protein was concentrated up to 5 mg/mL. If the purified eluate from the affinity chromatography could not be loaded onto the SEC column on the same day, it was dialysed against 5 L of PBS pH 7.4 at 4°C using SnakeSkin™ Dialysis Tubing (3.5 kDa cut-off; ThermoFisher Scientific, USA) overnight to remove imidazole.

For the SEC of Nbs, Fast Protein Liquid Chromatography (FPLC) was carried out using either HiLoad Superdex 16/600 75 pg (S75) or Superdex 200 Increase 10/300 GL (S200) column depending on the volume of the eluate that was to be injected. For the analytical size exclusion analysis of mutants in comparison to GH46, S200 column was used. The column was washed with 1.5 CV of degassed MilliQ water and equilibrated with 1.5 CV filtered and degassed PBS pH 7.4. For the S75 column, sample was injected into a 5 mL sample loop and was run with a flowrate of 0.75 mL/min. For S200,

a 0.5 mL loop was used instead, and sample was run with a flowrate of 0.5 mL/min. The UV absorbance was measured at 280 nm and sample was collected in 1.5 mL of fractions. The fractions with an absorbance peak were analysed with SDS-PAGE and the Nb-containing fractions were pooled and concentrated using AmiconR Ultra MWCO 3 kDa at 3200 g/4°C. For S75, the peak lay between 75 and 85 mL, for the S200, it lay between 16 and 18 mL. Nb concentration was determined by measuring UV absorbance at 280 nm on a NanoDrop ND-1000 Spectrophotometer. For GH46, the protein was again only concentrated to a maximum of 0.9 mg/mL. Nb was aliquoted into 50 µL samples and snap-frozen in liquid nitrogen to store at -80°C. As an example, the SDS-PAGE of the final purification step as well as the SEC chromatogram of a Nb are shown in **Figure 15**.

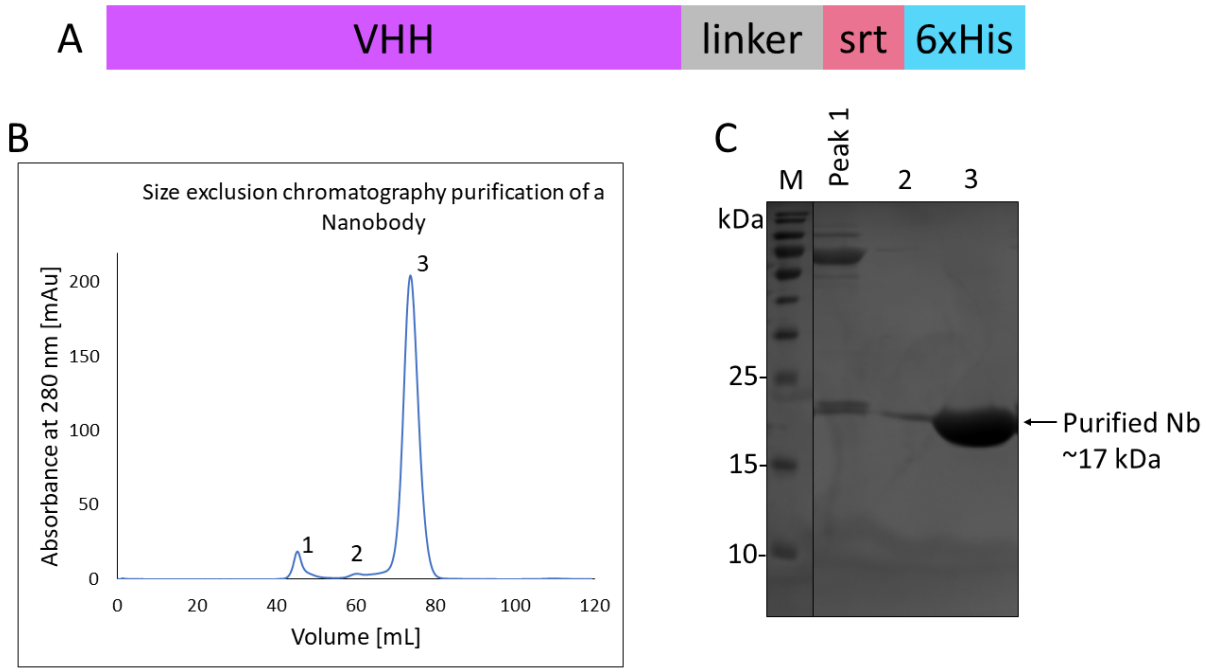


Figure 15: Example of purification of nanobodies. A. *A schematic representation of a Nb sequence. The Nb (VHH) sequence was followed by a linker, a sortase-tag (srt), and a 6xHis-tag.* ***B.*** *SEC chromatogram of a Nb purification. 3 peaks were seen in the chromatogram that were separately analysed on SDS-PAGE.* ***C.*** *SDS-PAGE from SEC of Nb. Peaks represent the marked peaks of the SEC chromatogram. The Nb has a theoretical molecular weight of ~17 kDa (marked in figure). The Nb eluted in peak 3 as seen in the SDS-PAGE. M: Marker.*

Based on the calibration standard run of HiLoadR 16/600 S75 prep grade column, a Nb with ~15 kDa molecular weight is expected to elute at ~85 mL, therefore the Nb elutes at the expected elution volume (between 70 and 80 mL). All purified Nbs were tested for binding to Globo-H expressing cells.

For the protein yield comparison of GH46 mutants, proteins eluted from Ni^{2+}/NTA beads were concentrated to 3 mL and run over an S200 column.

Purification of GH46-trimer from inclusion bodies

The purification was carried out by David Warschkau as part of his Masters' thesis project under my supervision. SHuffle cells were transformed, cultured, induced, and harvested as described above. For the inclusion body isolation, purification, and refolding, the protocol from Li Xu *et al.* (2017) was followed[295]. In brief, cell pellet was resuspended in 40 mL of 20 mM Tris-HCl, pH 8. 20 mL of mixture at a time were sonicated for 10 min with 50% amplitude and 30% power to lyse cells. This was carried out twice with 10 min on ice in between. Cell lysate was then centrifuged at 10,000 g/4°C for 5 min to remove cell debris. The supernatant was removed. Pellet was washed thrice with 40 mL of 20 mM Tris-HCl pH 8, 0.5% (v/v) Triton X-100, 1 M urea and thrice with 40 mL of just 20 mM Tris-HCl pH 8. In between centrifugation steps were carried out at 10,000 g/4°C for 5 min. The pellet was then solubilised overnight at 4°C/200 rpm in 40 mL of 20mM Tris-HCl pH 8, 6 M urea, 2 mM 2-mercaptoethanol (2-ME). Any remaining insolubilities were removed by centrifugation at 12,000 g/4°C for 15 min. Supernatant was filtered through 0.45 µm filter and purified in Äktapurifier UPC 10 System using a HisTrap™ HP 5 mL Ni^{2+}/NTA column. The column was equilibrated with binding buffer (20 mM Tris-HCl pH 8, 6 M urea, 10 mM imidazole) and supernatant was loaded onto the column. Elution was carried out in a 5-tep gradient with increasing elution buffer concentration (20 mM Tris-HCl pH 8, 6 M urea, 300 mM imidazole). Peak fractions were collected and analysed with SDS-PAGE. **Figure 16** shows the chromatogram and the corresponding gel.

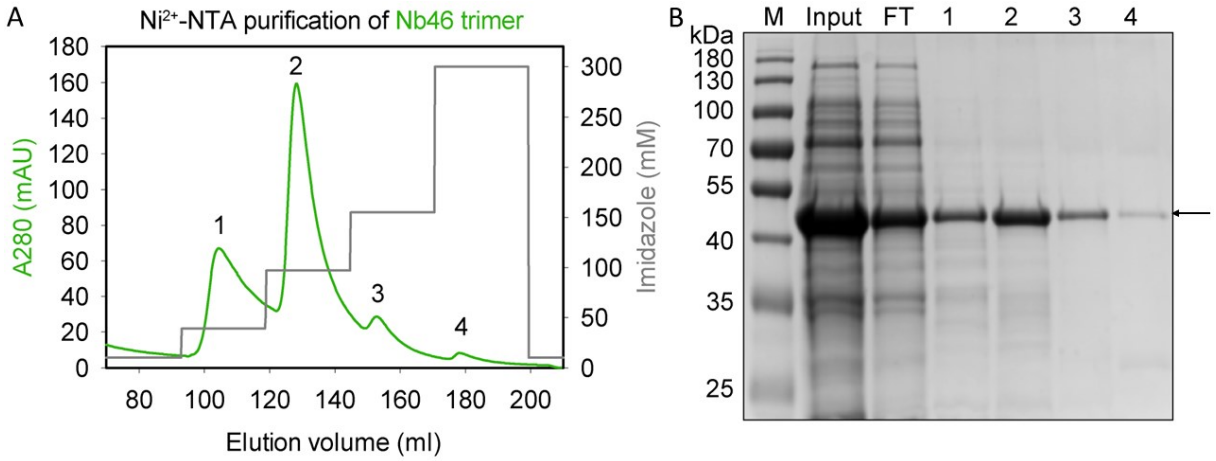


Figure 16: GH46-trimer purification from inclusion bodies. A. *Chromatogram from Ni^{2+}/NTA purification. Imidazole concentration was increased in steps and all eluting peaks were collected separately.* ***B.*** *Peaks 1-4 from purification were analysed SDS-PAGE. Expected molecular weight of GH46-trimer is ~46 kDa, which was obtained as shown with an arrow. While contaminations were seen, especially in peaks 1 and 2, GH46-trimer was obtained mostly purified. Peaks were combined.*

Concentration of purified GH46-trimer was determined using NanoDrop ND-1000 Spectrophotometer. Imidazole was removed by ultrafiltration using Amico filter units and 20 mM Tris-HCl pH 8, 6 M urea. Following this, GH46-trimer was concentrated to 0.3 mg/mL. 2-ME was added to protein in a final concentration of 2 mM so that mismatched disulphide bonds could be broken. The denatured Nb was finally refolded by rapid dilution method of 1 mL protein each in 9 mL cold refolding buffer (20 mM Tris-HCl pH 8, 0.2 M L-Arginine, 5 mM Gluthatione (GSH), 1 mM Glutathione disulphide) and vortexed for mixing before incubation at 4°C overnight. Mixture was centrifuged at 12,000 g/4°C for 10 min to remove any remaining precipitates. The refolded protein was concentrated using AmiconR Ultra MWCO 3 kDa and buffer exchanged via dialysis to PBS pH 7.4.

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Samples for SDS-PAGE were denatured using 5x sample buffer (0.25 M Tris-HCl pH 6.8, 50% glycerol (v/v), 10% SDS (w/v), 5% β-Mercaptoethanol (v/v), 0.05% Bromophenol blue (w/v)) and heating at 95°C for 5 min. Samples were centrifuged down for 1 min at 14,000 rpm using a table-top centrifuge so that precipitates would settle down. A separating gel (375 mM Tris-HCl pH 8.8, 10 or 15% of 37.5:1 acrylamide:bisacrylamide, 0.1% SDS, 0.1% ammonium persulphate (APS), 0.04% tetramethylethylenediamine (TEMED)) and a stacking gel (125 mM Tris-HCl pH 6.8,

5% of 37.5:1 acrylamide:bisacrylamide, 0.1% SDS, 0.1% APS, 0.1% TEMED) were poured into an assembled glass chamber with 1 mm spacing, loaded with denatured samples and run in the MiniProtean system (Bio-Rad, USA). The gel and running chamber were filled with running buffer (25 mM Tris-HCl pH 8.3, 192 mM glycine, 1% SDS) and 120 V, 25 mA were applied to separate proteins. As a size marker, 4 µL of PageRuler Plus Prestained Protein Ladder (ThermoFisher Scientific, USA) were used. Once the dye front reached the bottom of the gel, the gel was removed and stained with Coomassie Brilliant Blue G250 (70 µM Coomassie in 40 mM HCl) by boiling for 5 s and then agitating gel in Coomassie for 20 min. Gel was then boiled and agitated in water to remove excess dye. The gel was imaged using Gel Doc EZ imaging system (BioRad, USA).

2.2.4. Microscale Thermophoresis

Microscale Thermophoresis (MST) raw data was either generated by me or by David Warschkau, as part of his Masters' thesis project under my supervision. The data was analysed by me. 6x-His-tagged Nbs were diluted to 200 nM in PBS + 0.05% Tween-20 (PBST). For labelling, Monolith RED-tris-NTA 2nd Generation dye (Nanotemper Tech, Germany) was also diluted to 100 nM in PBST. Nb and dye were mixed in 1:1 volume ratio and incubated in the dark for 30 min at RT. For the MST, dilution series of glycans were required. Synthetic Globo-series glycans (Elicityl, France) were dissolved in MilliQ water for 10 mM stock solutions. For each of the glycans, a 16-step serial dilution was established with 5 µL concentrations ranging from 10 mM to 305 nM. These were mixed 1:1 with 100 nM Nb giving end concentrations for Nb of 50 nM and for glycans ranging from 5 mM to 153 nM and loaded into Monolith NT. 115 capillaries. Experiments were carried out using the instrument Monolith NT. 115 with medium MST and 100% LED power. Data was analysed on the MO. Affinity Analysis software.

2.2.5. Cell culture

Frozen stocks of cells were thawed rapidly at 37°C for 30 s before resuspending immediately in 10 mL of fresh media. Cells were then centrifuged at 300 g for 5 min to remove DMSO and resuspended again in fresh media to be seeded in 25 cm^2 flasks. Cells were then expanded so that they could be used in experiments. MCF-7 and HEK293 were cultured in DMEM supplemented with 10% FCS, 1%

penicillin/streptomycin, and 2 mM glutamine. For MCF-10A, DMEM and F12 medium were taken 1:1 supplemented with an end concentration of 5% horse serum, 20 ng/mL epidermal growth factor (EGF), 0.5 mg/mL hydrocortisone, 100 ng/mL cholera toxin, 10 µg/mL insulin, and 1% penicillin/streptomycin. Cell cultures were maintained at 37°C and 5% CO_2 and passaged every 3-4 days after they reached 85-90% confluency.

2.2.6. Flow cytometry

For the flow cytometry analysis of Nb binding to cells, the protocol described in **section 2.1.9**.[291] was carried out. Briefly, 0.5 x 10^6 cells per sample were incubated with 50 µL of either 0.4 mg/mL of purified His-tagged Nb, 5 µg/mL of VK9 in PBS + 1% BSA, or buffer only. As secondary Abs, α-6xHis Atto647N Ab (1:500; Rockland USA) and goat α-mouse-IgG (H+L) Ab Alexa Fluor™ 635 (1:400; Invitrogen by Thermo Fisher Scientific USA) were used. For the mutants, the binding assays were carried out at 37°C. Raw data for Nb binding was generated either by me or by masters' students (Kristin Frensemeier, David Warschkau, Zeinab Fandi, and Sophia Schrinner, with the last three working under my supervision). The data was analysed by me.

For the apparent K_D determination of GH46 and GH46 trimer in comparison to VK9, varying concentrations of the proteins were incubated with the MCF7 cells. Relative binding of cells was calculated based on the relative MFI and plotted against logarithmic protein concentration. Apparent K_D values were determined by sigmoidal fitting in Origin (v2021.b, OriginLab).

The co-staining of GH46 and VK9 was carried out with the same concentrations used in the single staining protocol. For detection, α-6xHis Atto647N Ab and α-mouse IgG Ab Alexa Fluor™ 488 (1:400; Invitrogen by Thermo Fisher Scientific USA) were used together. Data was analysed in 2D dot plots.

2.2.7. Confocal laser scanning microscopy

Cell binding of Nbs was also shown by confocal microscopy. Data was generated either by Kristin Frensemeier, as part of her masters' thesis, or David Warschkau, as part of his masters' thesis under my supervision. 12 mm coverslips were used. These were first cleaned with 70% ethanol and air-dried before being coated with 0.01% poly-L-lysine solution for 1 h/RT. Coverslips were washed thrice with PBS before being placed

face up in 24-well plates. 20,000 cells were seeded 48 h before experiments to ensure ~70% cell confluency.

Medium from the wells was removed, and cells were rinsed twice with PBS. Cells were fixed in 4% paraformaldehyde for 5 min at RT before being washed thrice again with PBS for 5 min each. Cells were then blocked for 1 h/RT in PBS + 1% BSA. 50 µL drops of 0.4 mg/mL of Nb, 5 µg/mL of VK9 or PBS + 1% BSA only were placed on parafilm in a humidified chamber and covered with coverslip with cells facing solution and incubated for 1 h/RT. Coverslips were washed thrice with PBS + 1% BSA before being similarly incubated with secondary Abs (anti-His Atto 647 (1:5,000; Rockland USA) and anti-mouse IgG Alexa™ Fluor 635 (1:400; Invitrogen by Thermo Fisher Scientific USA)). Coverslips were washed thrice with PBS + 1% BSA and placed on 5 µL drops of Roti®Mount FluorCare DAPI mounting solution (Carl Roth, Germany) on microscopy slides. Coverslips were sealed onto slides with nail polish and dried overnight. The Axio Imager.M2 confocal LSM 800 microscope (Carl Zeiss, Germany) equipped with EC Plan-Neofluar 40x/1.30 Oil DIC M27 objective and ZEN software (Carl Zeiss) was used at settings 22% laser power and pinhole = 80 µm for all images. For GH46 and the control Nb, the same brightness/contrast have been used. To display VK9 images, different brightness/contrast settings were taken to compensate for high intensity of the staining. Also, a different pseudo-colour for VK9 images was selected to highlight the distinct display settings as well as the fact that VK9 and GH46/control Nb were detected using different secondary Abs.

For the comparison of GH46 and GH46-trimer binding to MCF-7, brightness/contrast were adjusted identically.

2.2.8. Differential Scanning Fluorimetry (DSF)

To assess the stability of GH46 and its mutants, the melting temperatures (T_M) of the proteins were determined with differential scanning fluorimetry (DSF). The raw data was generated by Sophia Schrinner as part of her masters' thesis under my supervision. The data was analysed by me. 0.5 mg/ml of protein in PBS were mixed with the hydrophobic SYPRO™ Orange protein stain (Thermo Fisher Scientific, USA) in sealed UV transparent 96 well plates and monitored for increase in fluorescence upon denaturation in the QuantStudio™ 5 Real-Time-PCR-System (Thermo Fisher Scientific, USA). With the Protein Thermal Shift™ 1.3 software (Thermo Fisher

Scientific, USA), the melting curve plots with fluorescence over temperature were used to determine first derivative functions and their corresponding T_M values.

3. Results and Discussion

3.1. CA 19-9 in pancreatic cancer

3.1.1. Sequence analysis of GB11 and HA8 in comparison to 1116-NS-19-9

GB11 and HA8 were generated against sLe A and compared to 1116-NS-19-9, currently a commonly used Ab targeting sLe A. By RNA sequencing, the light chains of GB11 and HA8 were identified previously to be of κ-class. I compared the heavy and light chain sequences to those of 1116-NS-19-9 (pdb: 6XTG) and charted out the CDR regions (**Figure 17**).

Figure 17: The sequence alignments of heavy (top) and light chains (bottom) of GB11 and HA8 with those of 1116-NS-19-9. *Only the amino acid differences in the sequences are given and are marked in red. CDRs are marked in magenta. The arrow marks one out of seven core mutations carried*

out by Borenstein-Katz et al. in the 1116-NS-19-9 sequence to generate an Ab with higher affinity for sLe A[203]*. The sequence of 1116-NS-19-9 was taken from pdb file 6XTG and alignment was generated with protein BLAST*[296]*.*

Interestingly, there is a 97% (GB11) and 96% (HA8) identity similarity in the light chains including the same CDR regions. Similarly, 91% (GB11) and 89% (HA8) similarities in the heavy chain sequences is seen. Such high similarities in the sequences of the Abs were surprising as 1116-NS-19-9 was generated over four decades ago with immunisation of mice with colorectal cancer cells[187], while GB11 and HA8 were generated from immunisation of mice with pure, synthetic, and CRM_{197}-conjugated sLe A. Because all three Abs show high specificity for sLe A (see glycan array and cell binding results, **Figure 18** and **Figure 21**, respectively), it can be hypothesised that the germline Abs produced by the murine immune system against sLe A are already highly selective in nature and allow for affinity-matured Abs with similar sequences. Indeed, germline Abs against carbohydrate antigens can be highly selective and produce affinity-matured Ab sequences which are similar even when different initial immunisation approaches have been employed[297]. Having highly selective germline Abs could be nature's way of avoiding autoimmunity as carbohydrate antigens often differ from other self-associated glycans only in stereochemistry, at times of just a single hydroxyl group. Therefore, generating polyreactive Abs could have catastrophic outcomes[297].

The only differences in the binding regions are found in CDR2 in heavy chains (CDRH2) of both GB11 and HA8. Instead of the charged Lys and the Asn, GB11 has the hydrophobic Tyr and Ile at positions 54 and 56, respectively (also **Figure 23**). HA8 also displays a more hydrophobic CDRH2 with Trp and the bigger amino acid Val instead of Lys and Gly at positions 54 and 55, respectively. The differences in binding affinities of the Abs, as shown below (**sections 3.1.3** and **3.1.4**), may be attributed to CDR2.

Furthermore, Borenstein-Katz *et al.* studied and mutated seven amino acids in the core region of 1116-NS-19-9 and generated an Ab with higher affinity for sLe A[203]; one of those mutations is found at position 100 (Threonine to Valine) in the heavy chains of both GB11 and HA8 (in the publication, the mutation is indicated as Thr*94*Val[203]), which lies close to CDRH3 and could also be the reason for higher binding affinities of our

Abs to sLe A. Another mutation carried out by them is Thr99Ala (indicated as Thr*93*Ala in the publication[203]), which in case of HA8 has evolutionarily been changed to Asn and remains the same in GB11. In the light chain, they mutated Tyr87Trp, which naturally occurs as Phe in GB11. However, Tyr or Phe is conserved for murine Abs at this position[203], therefore, Phe might not be contributing to higher binding affinity here.

In order to test whether any of the differences in the sequences translated to better Ab binding to sLe A, I employed various methods to analyse and compare the Abs 1116-NS-19-9, GB11, and HA8 in binding, specificity, and affinity for sLe A.

3.1.2. Glycan array

Figure 18 shows the analysis of the glycan array carried out with GB11 and HA8 and compared with 1116-NS-19-9 to show binding to immobilised synthetic sLe A. Relative MFI was calculated and plotted. GB11 and HA8 bind similarly to immobilised synthetic sLe A, however the binding is more than 20 times better than that of 1116-NS-19-9, showing GB11 and HA8 to be far better binders of synthetic sLe A. All three Abs showed no cross-reaction binding to any other structure on the glycan array slide (see glycan array scheme in **Figure 13**) as no other signal was detected. This includes structures similar to sLe A such as sLe X (Neu5Acα2,3Galβ1,4(Fucα1,3)GlcNAc; **Figure 3**), which only differs from sLe A (Neu5Acα2,3Galβ1,3(Fucα1,4)GlcNAc) in its glycosidic bonds, proving the Abs' high specificity for the antigen. As seen with the sequence comparison, the differences between the Ab sequences are minimal. However, the more hydrophobic residues in CDRH2 might play a role in better binding of GB11 and HA8. The differences could also be because of how the Abs were generated: GB11 and HA8 were generated against synthetic sLe A, while 1116-NS-19-9 was generated against native glycan displayed on cells[187]. Alternatively, the few amino acid differences in the core regions of the three Abs could also be affecting the binding of the Abs and in case of GB11 and HA8, making them better binders (**section 3.1.1**). Similarly, Borenstein-Katz *et al.* made only seven mutations in the core region of 1116-NS-19-9 to improve its affinity approximately ten-fold[203].

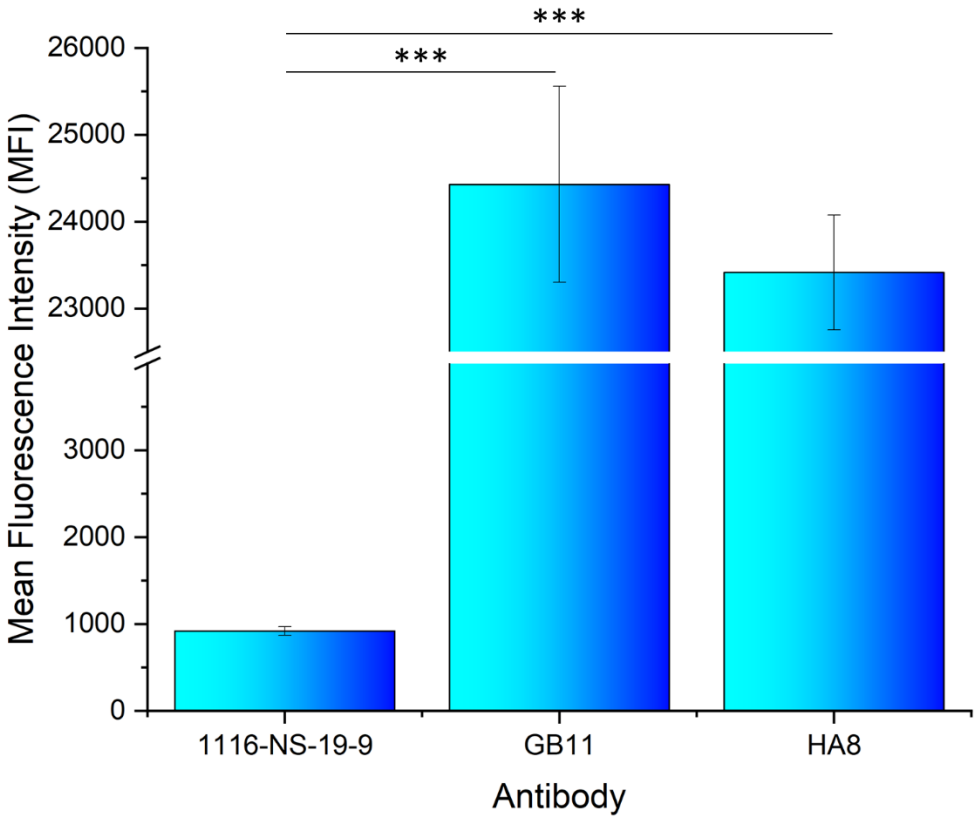


Figure 18: Glycan array quantification of sLe A-binding Abs. *MFI was calculated from three independent experiments and compared by two-sample t-test with (***) $p < 0.001$. Error bars represent standard error of mean (SEM).*

Next, I carried out assays to determine the K_D values of the Ab interactions with sLe A.

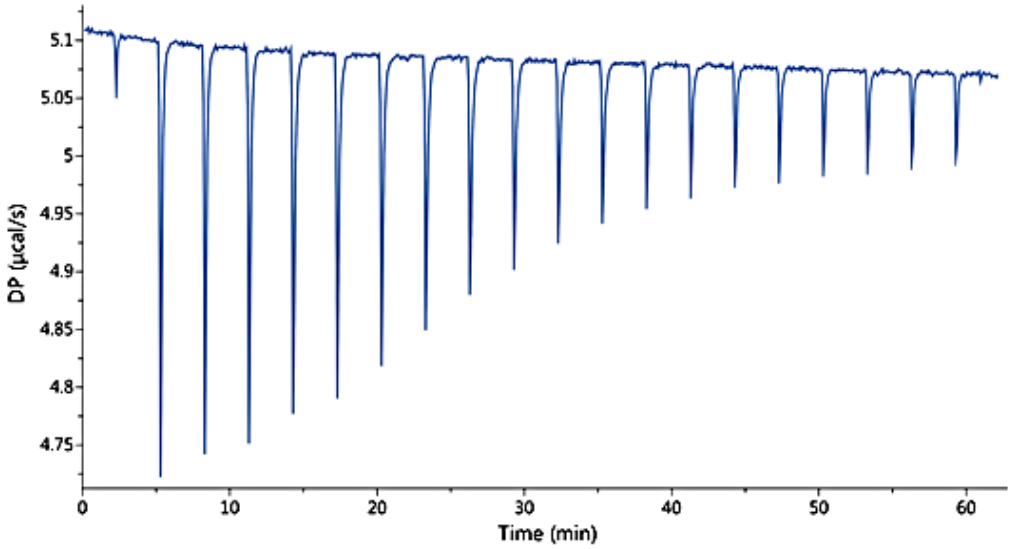
3.1.3. Isothermal Titration Calorimetry (ITC)

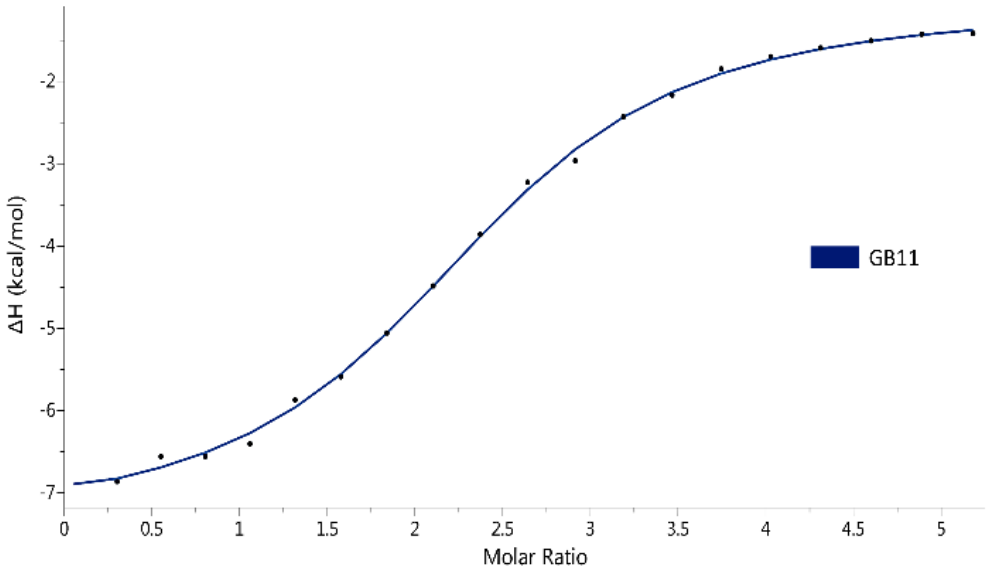
GB11 and HA8 were analysed with ITC to determine their binding affinities to synthetic sLe A in solution. Because 1116-NS-19-9 was commercially bought and expensive, a positive control for ITC was not carried out. Instead, the measured binding K_D were compared to the literature value of 1116-NS-19-9[203]. As negative control, the glycan was titrated into 1x PBS buffer only. **Figure 19** shows the compiled results.

The binding sites (N) are derived from the calculated molar ratio between ligand molecules and Ab and, therefore, the expected N sites for an Ab is 2, which would mean that both the variable Fab regions on the Ab are active and binding to ligand molecules similarly. For both GB11 and HA8, the value lies around 2 and, therefore, both binding sites are active and binding to ligand (**Figure 19D**).

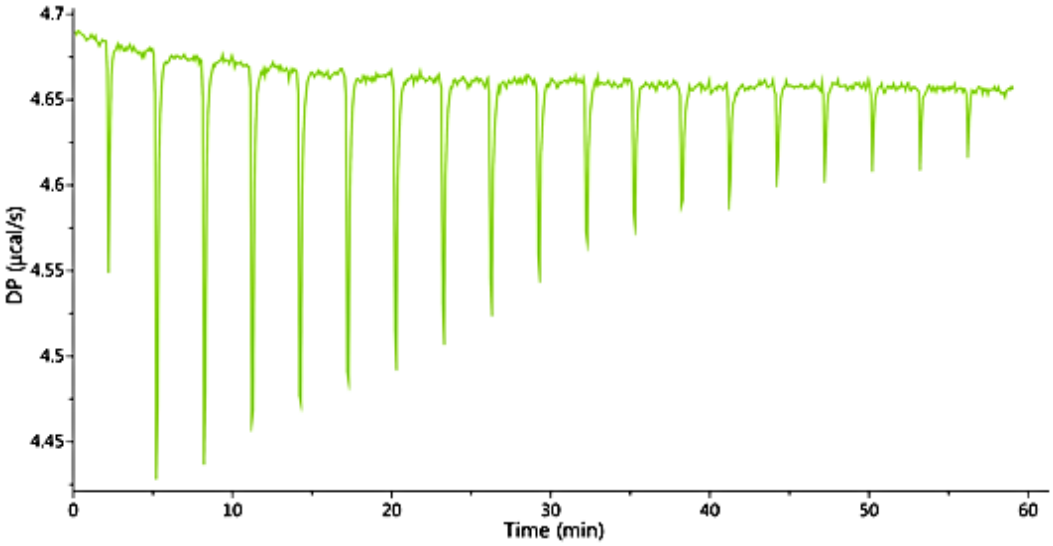
A negative ΔG value shows that the reaction taking place is spontaneous in nature. As both GB11 and HA8 have ΔG, ΔH, and TΔS values in error range of each other, the Abs react with similar spontaneity as well as enthalpy (generating similar amounts of heat), and, therefore, show similar binding to ligand. The average K_D also lies in range of each other. The Abs are low micromolar binders of sLe A, which is generally acceptable for carbohydrate binding Abs[203]. The murine mAb, 1116-NS-19-9, targeting sLe A and currently the most clinically used Ab for diagnosis of pancreatic cancer has a measured K_D of 1116-NS-19-9 is 14.7 μM[203]. In comparison to this, GB11 and HA8 show ~6 and ~4 times better binding affinities, respectively. However, the K_D for 1116-NS-19-9 was measured using SPR, therefore a comparison of the Abs in SPR is also required and follows below. There is no enthalpy change with ligand titrated into buffer (**Figure 19C**), showing that ΔH measured for the Abs is because of the Ab interaction with the ligand only.

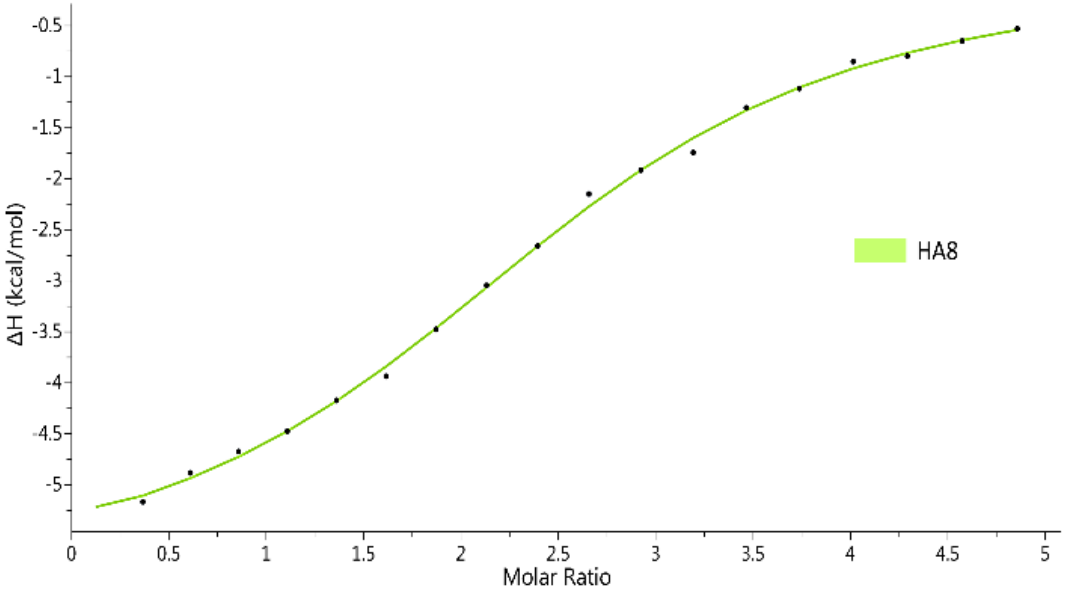
A


DP (µcal/s)
5.1
5.05
5
4.95
4.9
4.85
4.8
4.75
0
10
20
30
40
50
60
Time (min)

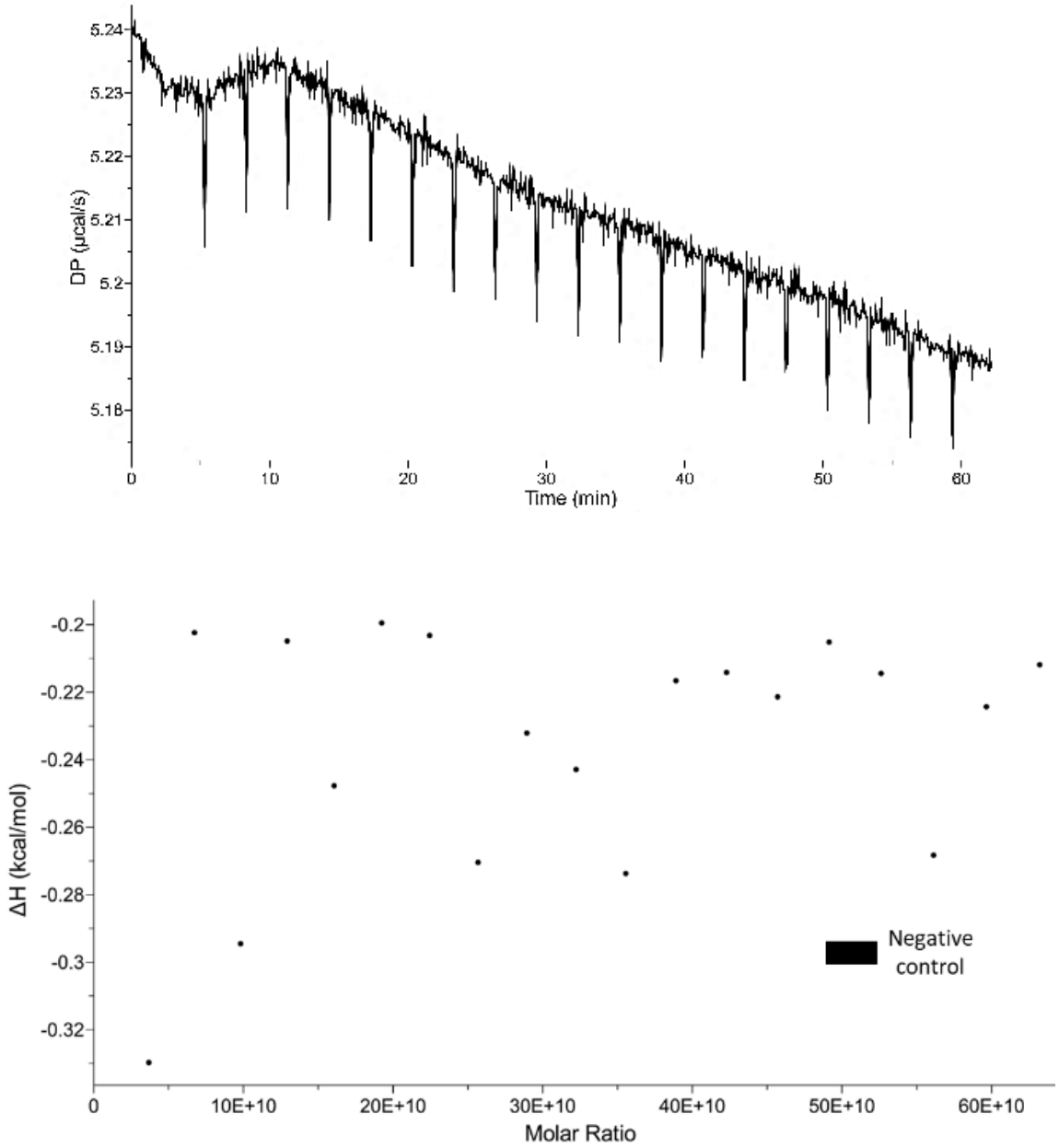

ΔH (kcal/mol)
-2
-3
-4
-5
-6
-7
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Molar Ratio
GB11

B


DP (µcal/s)
4.7
4.65
4.6
4.55
4.5
4.45
0
10
20
30
40
50
60
Time (min)


ΔH (kcal/mol)
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
-5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Molar Ratio
HA8

C



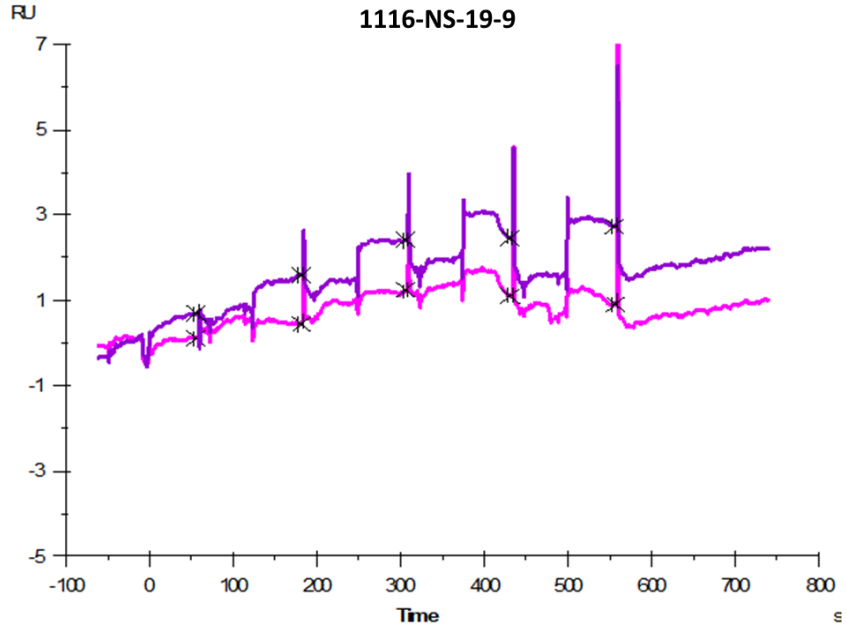
D	Binding sites (N)	Average K_D [µM]	Average ΔH [kcal/mol]	Average ΔG [kcal/mol]	Average TΔS [kcal/mol]
GB11	1.9 ± 0.4	2.42 ± 0.81	-6.36 ± 0.52	-7.68 ± 0.18	-1.32 ± 0.70
HA8	2.2 ± 0.3	3.56 ± 0.78	-5.57 ± 0.52	-7.45 ± 0.12	-1.88 ± 0.60

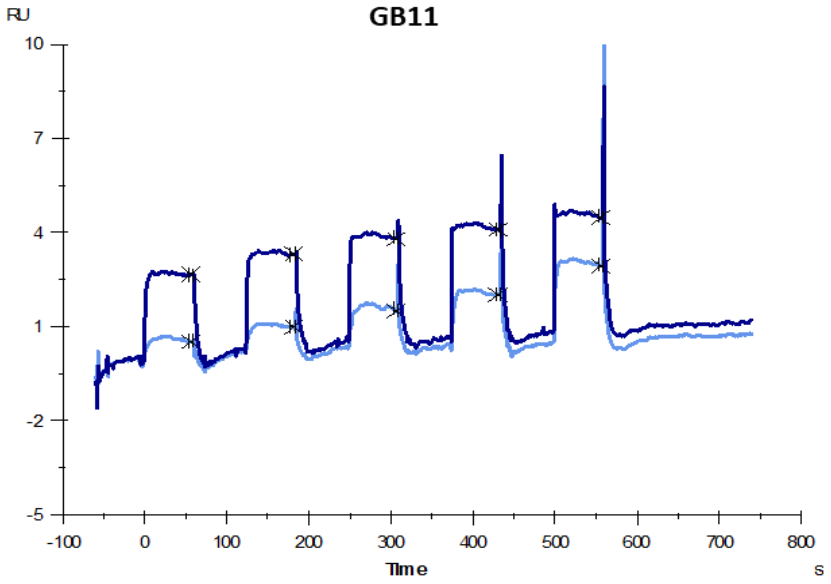
Figure 19: Isothermal Titration Calorimetry Analysis*. The Ab was in the cell, while sLe A was injected. Examples of graphs generated for each Ab are shown. As negative control, glycan was injected into PBS buffer only.* ***A.*** *GB11 is blue,* ***B.*** *HA8 is green, and* ***C.*** *negative control is black.* ***D.*** *The values for N are given. Average values were calculated from triplicates for K_D, ΔH, ΔG, and TΔS. SEM is given.*

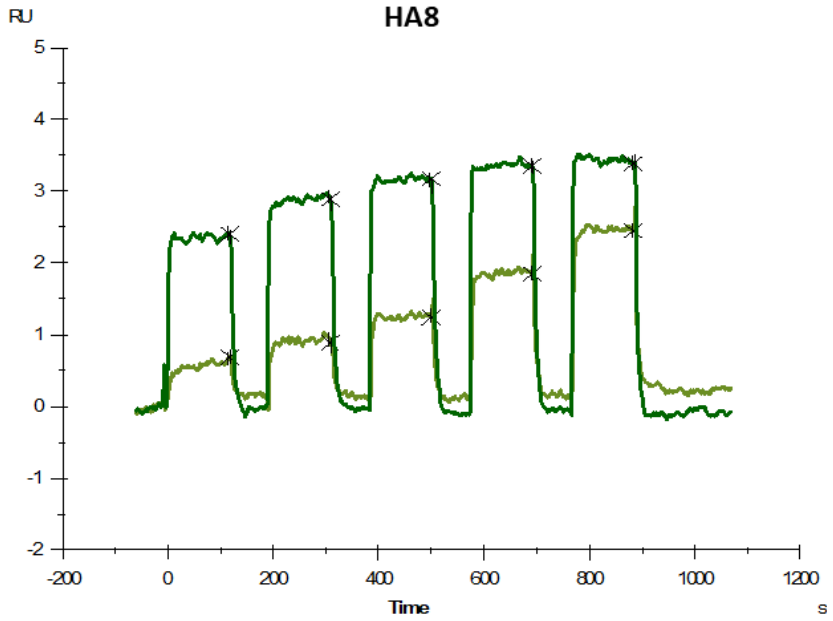
3.1.4. Surface Plasmon Resonance (SPR)

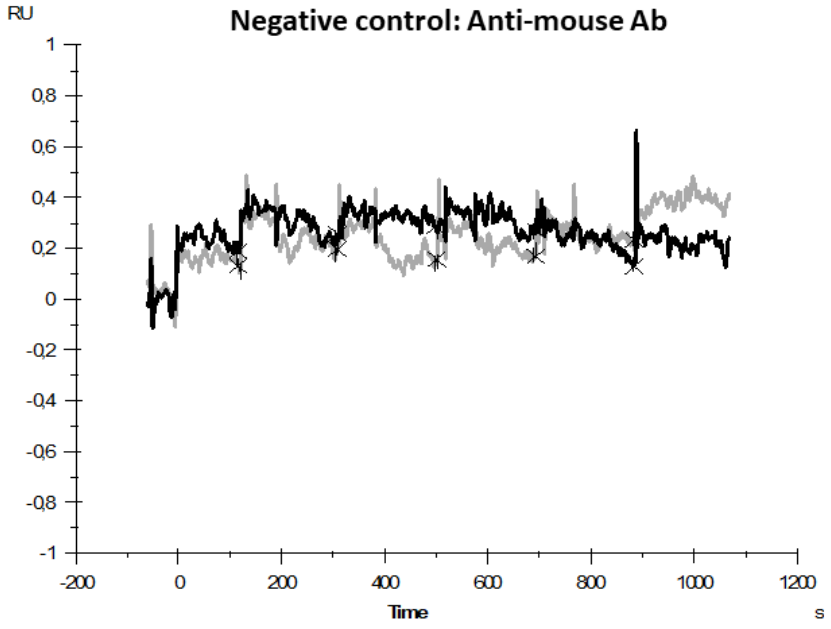
SPR affinity measurements were also performed to determine the K_D of GB11 and HA8 in comparison to 1116-NS-19-9. The K_D values are 3.2 ± 0.87 µM (GB11), 2.9 ± 0.47 µM (HA8), and 7.1 ± 3.4 µM (1116-NS-19-9; **Figure 20**), respectively. The SPR values for GB11 and HA8 corroborate the results from ITC. While the average measured K_D of 1116-NS-19-9 is almost half of that found in literature (14.7 µM[203]), a high SEM of 3.4 µM is also seen (**Figure 20B** and **C**). As the Ab was not pure but instead in medium and, the immobilisation of 1116-NS-19-9 ranged from 150-500 RU, it could have contributed to the higher SEM. Nevertheless, from both SPR results and the literature value as well as a comparison of ITC results, I derived that GB11 and HA8 were better binder of synthetic sLe A, in solution as well as when immobilised. GB11 and HA8 boast similar binding affinities to sLe A, which is in correlation with the glycan array results. Glycan array, however, also shows more than 20 times more binding of GB11 and HA8 than 1116-NS-19-9 to sLe A, therefore, it can be argued that the conformation of the immobilised glycan plays a vital role in Ab and glycan binding. In nature, glycan and not Ab is found immobilised on cells. Hence, glycan array may provide the closest scenario *in vitro* with synthetic glycan. Following these results, I wanted to test how the Abs would interact with cells expressing sLe A.

A

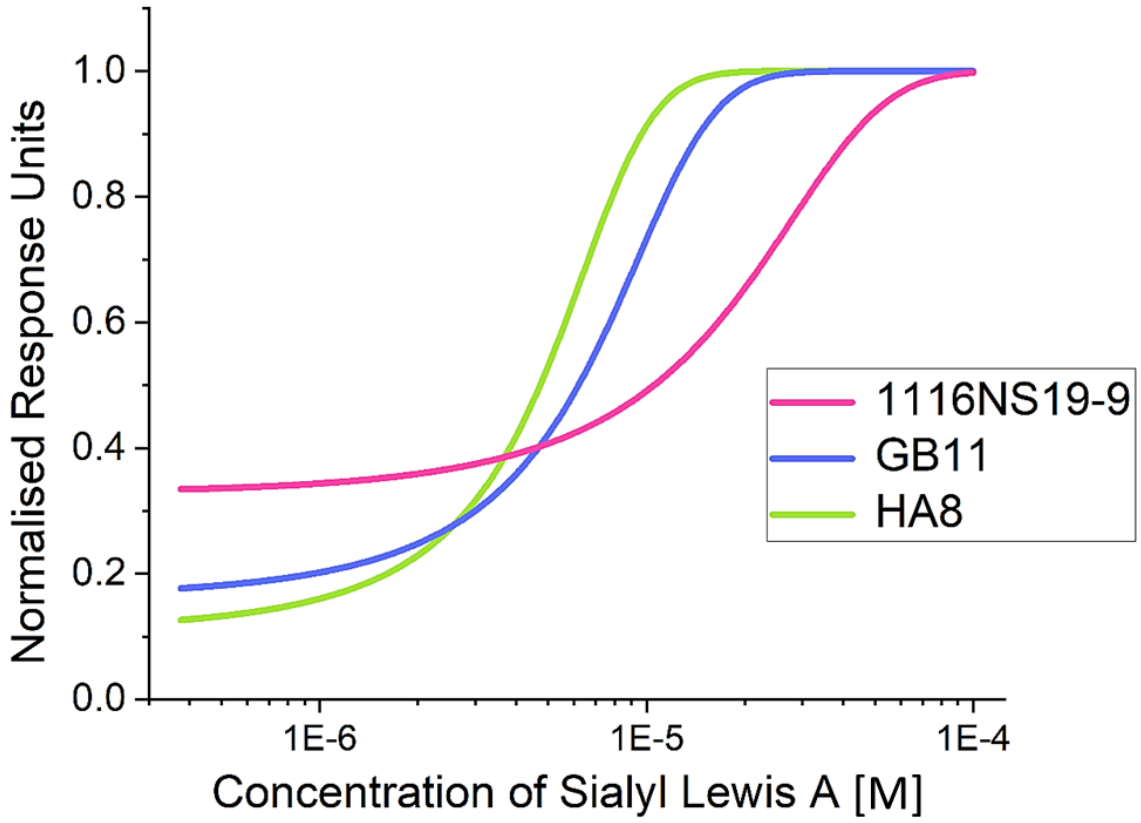

1116-NS-19-9
RU
7
5
3
1
-1
-3
-5
-100
0
100
200
300
400
500
600
700
800
Time
s


GB11
RU
10
7
4
1
-2
-5
-100
0
100
200
300
400
500
600
700
800
Time
s


HA8
RU
5
4
3
2
1
0
-1
-2
-200
0
200
400
600
800
1000
1200
Time
s


Negative control: Anti-mouse Ab
RU
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
-200
0
200
400
600
800
1000
1200
Time
s

B



C	**Average K_D [µM]**
1116-NS-19-9	7.1 ± 3.4
GB11	3.2 ± 0.87
HA8	2.9 ± 0.47

Figure 20: Surface Plasmon Resonance Analysis of Abs vs. sLe A*. Abs were immobilised using a commercial mouse Ab immobilisation kit and synthetic sLe A was used as analyte.* ***A.*** *Raw data of the response units (RU) over time with increasing sLe A concentrations from one of the repetitons as an example. The anti-mouse Ab from the immobilisation kit was used as negative control. Signal was measured at the marked points which came after the washing steps.* ***B.*** *The corresponding results with normalised response units against increasing concentration of sLe A over a logarithmic scale. The plot was generated in OriginPro 2021b using the ‚dose response' parameters. GB11: blue, HA8: green, 1116-NS-19-9: magenta, anti-mouse Ab: Black.* ***C.*** *Average K_D in µM. SEM is given.*

3.1.5. Flow cytometry analysis with B16 and B16 FUT3$^+$ cells

Next, I wanted to test how the Abs would interact with native glycan in a biological scenario as immobilised on cells. B16-FUT3$^+$ cells express sLe A on their cell surface, while B16 cells can be used as negative control[293]. Results are shown in **Figure 21**. In line with the previous results, all three Abs are highly specific for sLe A and they show no non-specific binding to B16 cells. Furthermore, GB11 shows significantly higher MFI than 1116-NS-19-9. Contrary to expectation, however, HA8 binding to native sLe A is drastically lower. Instead, 1116-NS-19-9 is a significantly better binder in comparison. As previously seen with the glycan array also, the conformation of the native immobilised glycan could play an important role here. Also, living cells provide a more dynamic environment so that other molecules on the cell surface could also interfere with the Ab and glycan binding. In line with this, the difference between GB11 and 1116-NS-19-9 binding, even though significant, is also lower on cells than it was with glycan array (**Figure 18**).

Altogether, HA8 binding to synthetic glycan was much higher than 1116-NS-19-9 and comparable to GB11, while its binding to native sLe A was drastically lower than both. GB11, on the other hand binds significantly better than 1116-NS-19-9, not only to synthetic (more than 24-fold), but also to native cell surface-immobilised sLe A (more than 3-fold).

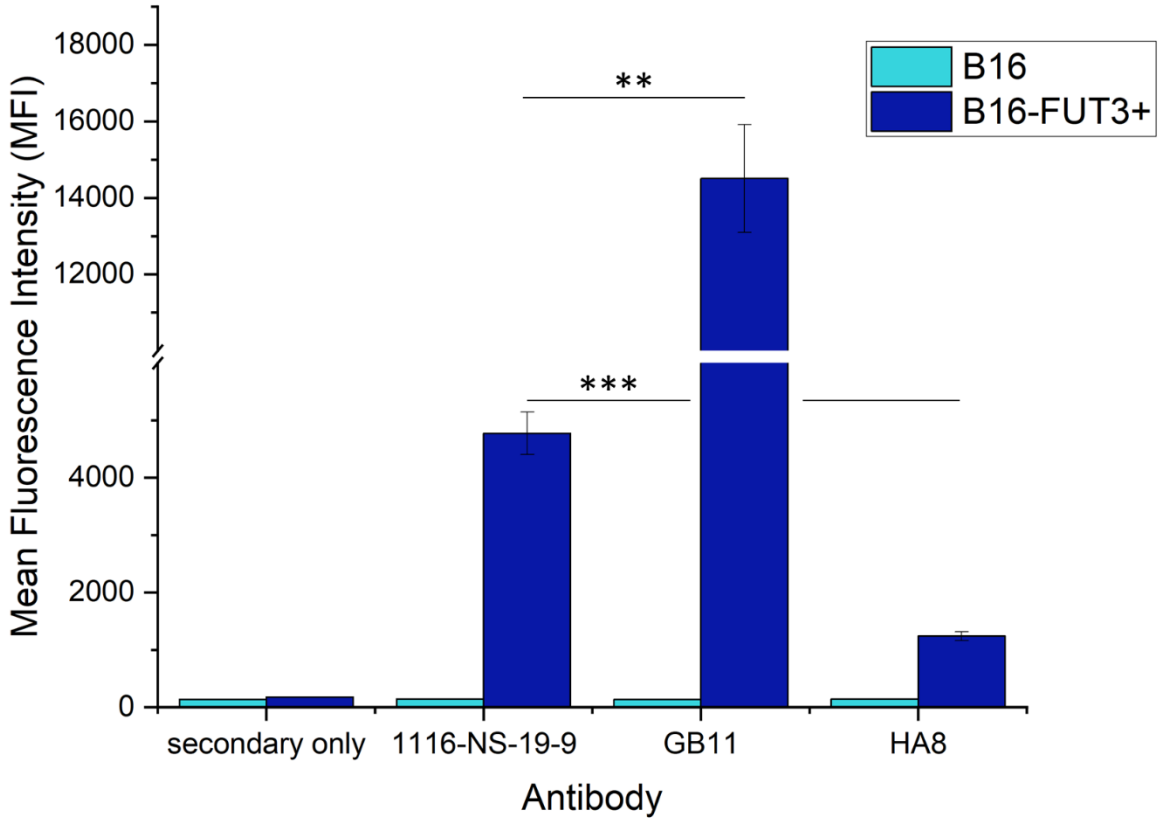


Figure 21: Cell binding analysis of anti-sLe A Abs. *The turquoise and blue colours represent B16 and B16-FUT3⁺ cells, respectively, as shown in the figure legend. MFI was calculated from three independent experiments and compared by two-sample t-test with (**) $p < 0.01$ and (***) $p < 0.001$. Error bars represent SEM.*

3.1.6. Immunohistochemistry (IHC)

To further establish the potential of using GB11 and HA8 in clinic, one of the questions addressed in this thesis for this project was whether GB11 and HA8 would bind to sLe A in human tissue for diagnosis of sLe A-expressing cancers as their commercial counterpart 1116-NS-19-9 is used. To this end, our collaborator carried out the IHC experiments (**Figure 22**). The method is described in the supplementary information (**section 5.1.2**).

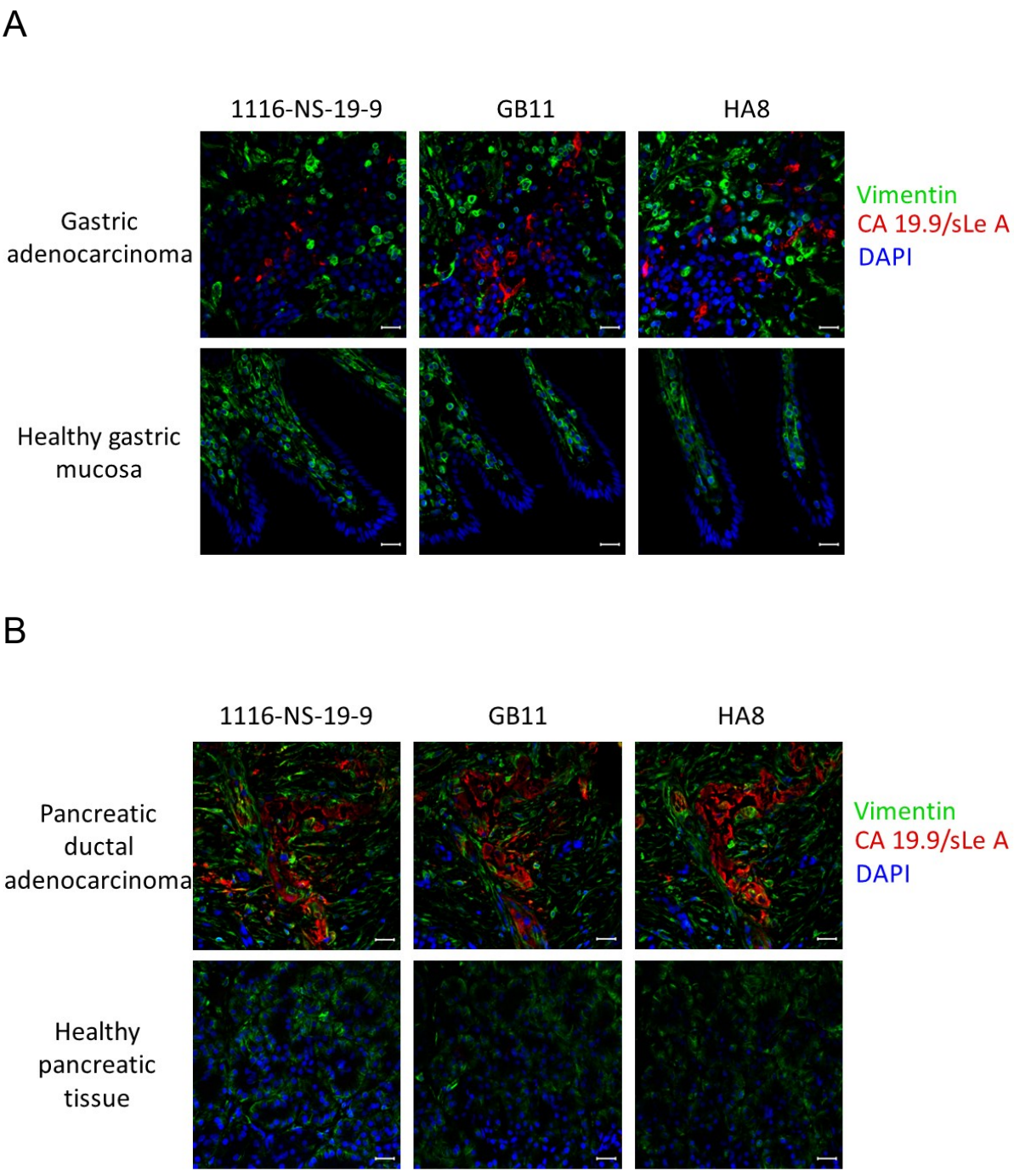


Figure 22: IHC with 1116-NS-19-9 (left), GB11 (centre), and HA8 (right) on healthy and carcinoma pancreatic and gastric mucosa tissues. ***A.*** *top: IHC of gastric adenocarcinoma; bottom: healthy gastric mucosa.* ***B.*** *top: IHC of pancreatic ductal adenocarcinoma; bottom: healthy pancreatic tissue. Cell nuclei (DAPI, blue), Vimentin (green) and sLe A (red) stained with α-sLe A mAb. Scale bar: 20 µm.*

Both GB11 and HA8 identified sLe A in pancreatic ductal adenocarcinoma as well as gastric adenocarcinoma similar to 1116-NS-19-9. Interestingly, no binding to healthy gastric mucosa or pancreatic tissue was seen showing that the Abs are specific for sLe A binding. Therefore, both Abs could be used in the future for diagnosis of

pancreatic cancer. However, with all results combined, GB11 presents a more viable option for an anti-CA 19-9/sLe A Ab that binds highly specifically and even more strongly to the glycan than the commercially used counterpart, 1116-NS-19-9.

Further, I wanted to study the structure of GB11 and its interaction with sLe A and compare it with that of 1116-NS-19-9 in order to test the hypothesis that the more hydrophobic residues of GB11 allowed it to be a stronger candidate. To this end, I obtained the crystal structure of Fab-GB11 and analysed it with and without the ligand.

3.1.7. Crystallisation

As sLe A has a relatively small size (819 Da) and hydrophilic nature, it may be considered as a suboptimal immunogen, so that recognition by Abs would be expected to be suboptimal and restricted to few contacts only[203]. The crystal structure of Fab-1116-NS-19-9 has already been studied in an unbound state, as well as in complex with sLe A[203]. I wanted to structurally analyse the binding of sLe A to Fab-GB11 and Fab-HA8 and compare them to 1116-NS-19-9. While attempts to crystallise Fab-HA8 were unsuccessful, Fab-GB11 formed diffraction-quality crystals. The crystals have been subsequently soaked with sLe A to obtain a complex between the Ab and the ligand. **Figure 23** shows the crystal structure of GB11 at a resolution of 1.86 Å (ice blue) in an overlay with 1116-NS-19-9 (gold) to show the differences between the structures. As anticipated, the Fabs of GB11 and 1116-NS-19-9 do not differ significantly in conformation, however, differences in the CDRH2 can be seen (**Figure 23B**) at positions 54 and 56 where Tyr and Ile replaced the charged residues Lys and Asn of 1116-NS-19-9, respectively. The hydrophobic amino acids may provide additional hydrophobic interactions in the binding pocket allowing for better binding to native as well as synthetic immobilised glycan. Furthermore, the co-crystal structure of Fab-1116-NS-19-9[203] (pdb: 6XTG) shows GlcNAc and Fuc lie in the binding pocket close to the amino acid positions 54 and 56. In case of GB11, the hydrophobicity of Tyr and Ile may allow for better interaction with the glycan, in particular with additional residues extending the glycan at the reducing end of GlcNAc. Additionally, unlike 1116-NS-19-9, GB11 has a Val at position 100 instead Thr (**Figure 17**). Val at this position is one of the seven core mutations Borenstein-Katz *et al.* introduced to improve the binding of 1116-NS-19-9 to sLe A[203]. The amino acid lies close to CDRH3 which interacts with the sialic acid and GlcNAc via Arg95 (**Figure 24**) and could influence

binding by affecting the binding pocket's conformation or flexibility. This may explain the observed higher binding affinity of GB11 to sLe A. This is further supported by the fact that the replacement of Thr by Val introduces a hydrophobic residue in a hydrophobic environment, which should be energetically favoured in terms of packing interactions.

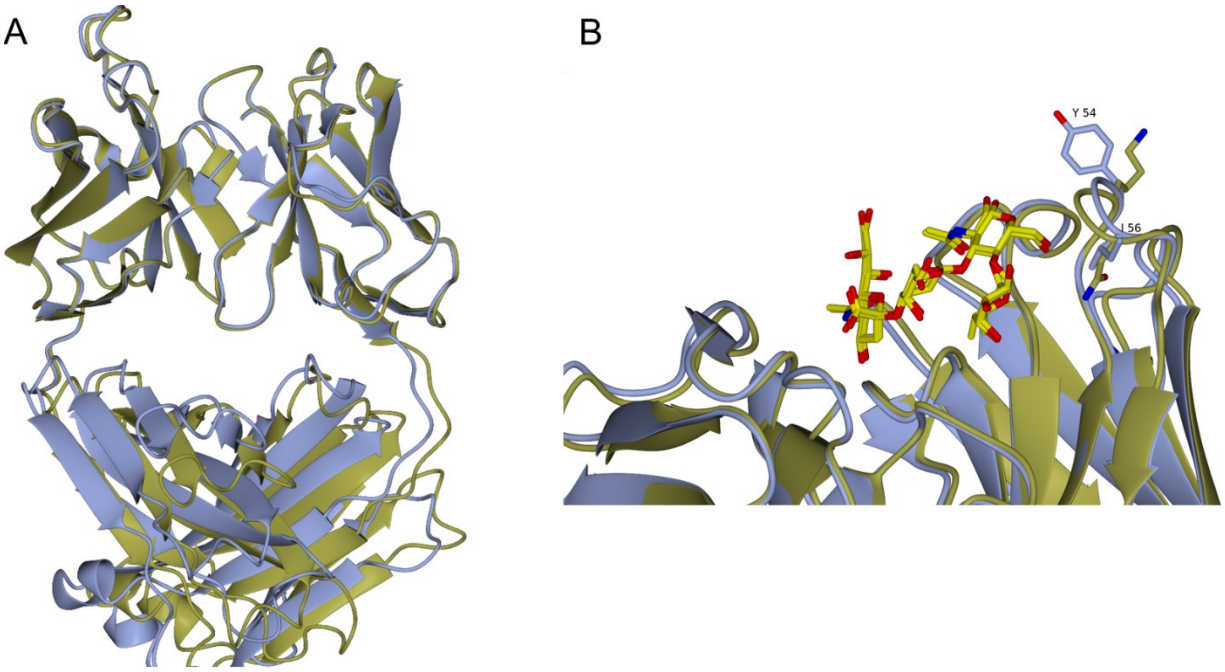


Figure 23: Crystal structure comparison of GB11 with 1116-NS-19-9. A. *Overlay of Fab-Gb11 (ice-blue) and Fab-1116-NS-9-9 (gold).* ***B.*** *CDRH2 of both Abs with their differences at positions 54 and 56 in the binding pocket are shown. In GB11, Tyr replaces Lys at position 54 and Ile replaces Asn at position 56. sLe A in the binding pocket is shown. The glycan has the same orientation for its interaction with both Abs. Figure created by Michael Krummhaar.*

Crystals soaked with sLe A diffracted only to a resolution of 2.96 Å. The interactions between the amino acids and glycans could, therefore, not be determined with precision compared to 1116-NS-19-9. In contrast to the ligand free form and 1116-NS-19-9, 4 molecules are in the asymmetric unit and clear difference density of the glycan could be observed in each Fab monomer. A final model comprising the four Fab fragments and the respective ligands could be built. All four sLe A ligands superpose well with each other and with 1116-NS-19-9. In both Abs, all CDRs except CDRL1 interact with sLe A (**Figure 24**). Also, in both cases most of the hydrogen bonds are preserved or a substitute hydrogen bond is formed. For example, the indole nitrogen of Trp33 interacts with GlcNAc acetyl oxygen instead of a water mediated hydrogen bond of OH6 of GlcNAc and Asn52A in 1116-NS-19-9. In general, water mediated

hydrogen bonds are not observed due to the low resolution of the GB11 structure. In both Abs, sialic acid is held in place via three direct hydrogen bonds. We observed an additional direct hydrogen bond between Fuc and Arg96, which is water mediated in case of 1116-NS-19-9. A crystal structure at higher resolution might reveal further interactions similar to 1116-NS-19-9. Furthermore, other methods such as saturation transfer difference-nuclear magnetic resonance (STD-NMR) should be obtained to have a clearer picture and compare the interactions of the two Abs with sLe A.

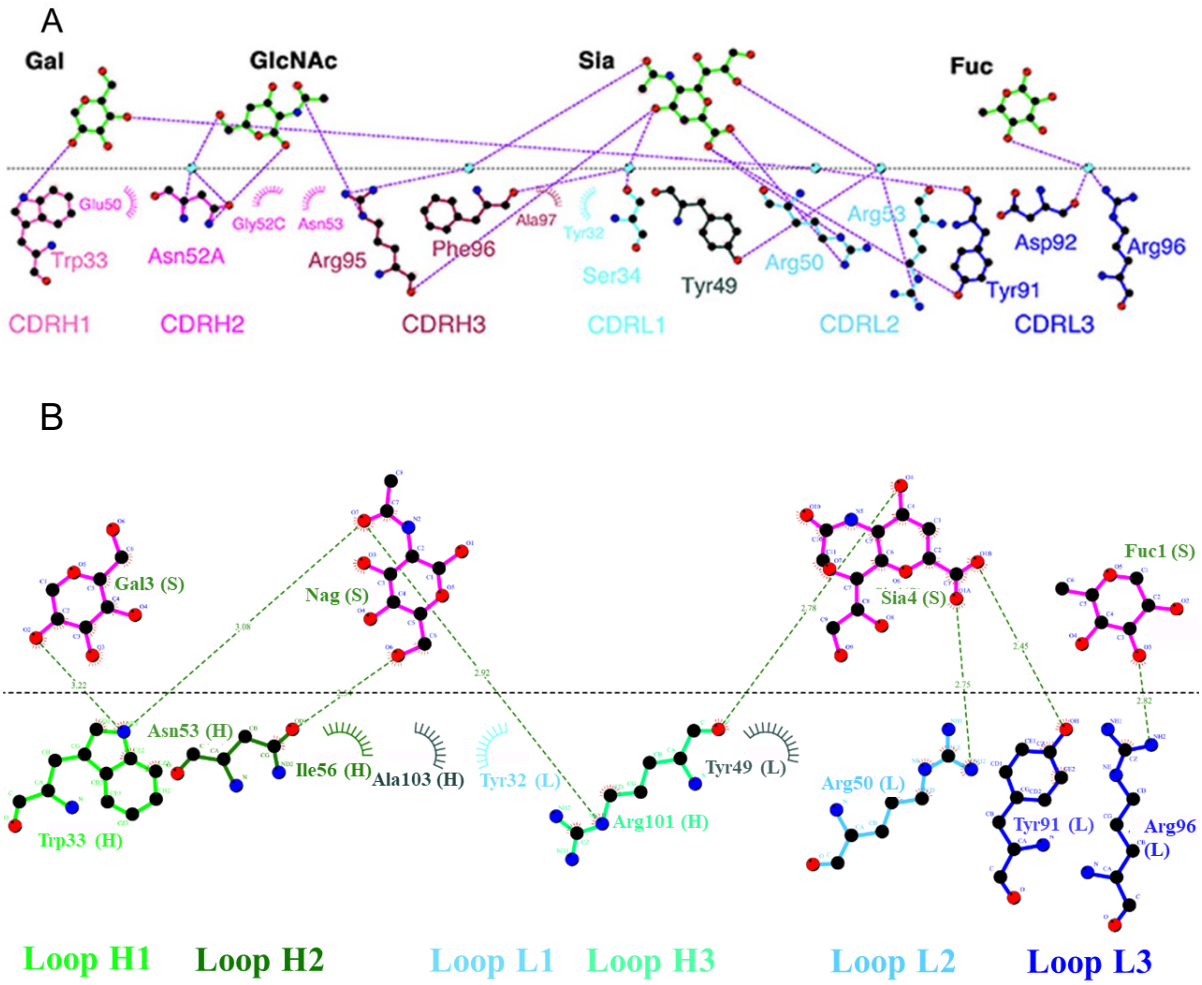


Figure 24: Ab-Fab interactions with sLe A as deduced from crystal structures. A. *Interactions of amino acids in CDRs of heavy and light chains of 1116-NS-19-9 with sLe A components. Figure taken from Borenstein-Katz et al.*[203]. ***B.*** *Interactions of amino acids in CDRs of heavy and light chains of GB11. Figure created by Dr. Christian Roth. In case of both 1116-NS-19-9 and GB11, all CDRs except CDRL1 are seen to interact with sLe A.*

While it could not yet entirely be confirmed that the higher binding affinity of GB11 for sLe A compared to 1116-NS-19-9 is due to a more hydrophobic CDRH2, the close

position of amino acids Tyr54 and Ile56 to the GlcNAc and Fuc in the binding pocket make it highly likely. Similarly, a more hydrophobic CDRH2 in HA8 than 1116-NS-19-9 was also seen (Trp instead of Lys54 and Val instead of Gly55; **Figure 17**). Furthermore, a more hydrophobic Val100 close to CDRH3 may also have affected the conformation of the binding pocket such that it enhanced the binding of GB11 and sLe A. HA8 also has Val100 in CDRH3 instead of Gly. This could be the reason why HA8 also binds synthetic glycan more strongly than 1116-NS-19-9. The few amino acid differences in the core region (**Figure 17**) could also be affecting the binding pocket rigidity as it has been shown before that core mutations can improve binding affinity to epitope[203].

In conclusion, I was able to show that GB11 and HA8 are highly specific binders for sLe A. Both Abs bind synthetic sLe A more strongly than 1116-NS-19-9 and recognise sLe A in human cancer specimens similar to 1116-NS-19-9. GB11 also binds native sLe A more strongly than 1116-NS-19-9. Its high specificity and affinity could help in the future to decrease false positive or negative results, which are a bottleneck for Abs with broader specificities[81].

3.2. Globo-H in breast cancer

Nbs targeting Globo-H had previously been generated by our group following immunisation of alpacas with synthetic Globo-H (**Figure 9**)[291]. A nanobody library was established through next generation sequencing and translated *in silico* into corresponding protein sequences. In parallel, immobilised Globo-H was used to isolate Globo-H specific hcAbs from the alpaca serum. Through analysis with LC-MS/MS, purified hcAbs were compared with the established Nb library to obtain 36 potential candidates for Globo-H targeting Nbs[291]. As part of this thesis, one of my aims was to filter the Nbs through a selection process that included binding to cells expressing Globo-H as well as other cells expressing lower molecular weight Globo-family members only (GB2, GB3, and GB4; **Figure 6**) to fish out the best Globo-H binder.

3.2.1. Identifying Globo-H binding Nanobodies

MCF-7 are GB5 and Globo-H expressing epithelial breast mammary gland cells from a White female patient in her late 60s who had metastatic adenocarcinoma[82,298,299]. For the initial round of screening for their binding to Globo-H, Nbs were tested on MCF-7 via flow cytometry and confocal microscopy to determine binding to native Globo-H. **Figure 25** show results from nanobody selection on MCF-7 with flow cytometry.

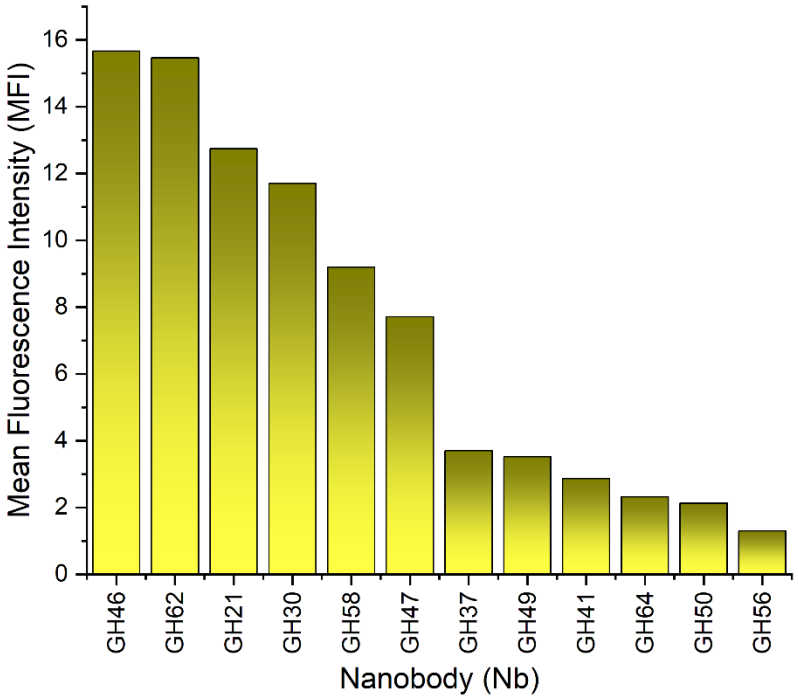


Figure 25: Nanobody selection for binding to Globo-H on MCF-7 cells. *MFI was measured with flow cytometry. This data has been published in Khilji et al. 2022*[291] *and the figure has been modified here for the thesis.*

12 of the 36 Nbs showed binding to MCF-7 cells in flow cytometry. As a threshold for binding, a cut-off of MFI was taken as low as 10, which left only the four, GH46, GH62, GH21, and GH30, as good binders[291]. Of these, the best binder was GH46. Along with GH46, GH62, GH21, GH30, GH58, GH47, GH41, and GH64 are poorly expressing proteins with very low end yield (data not shown). However, due to its superior binding, GH46 was further characterised (data for GH46 is discussed in detail later). Some of the lower yet better expressing binders were also studied further. Confocal microscopy was also carried out and results of some of the binders are shown in **Figure 26**. GH61 did not bind to MCF-7 cells in the flow cytometry analysis and was taken as a non-Globo-H and non-GB5-binding control.

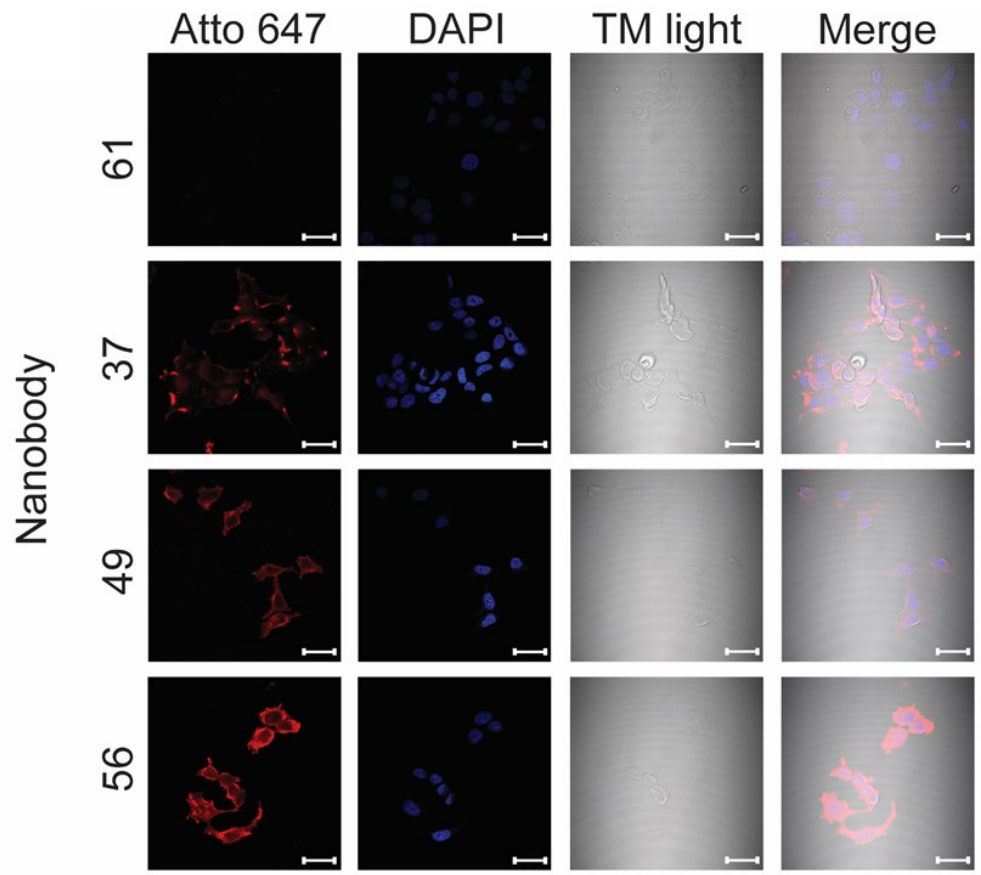


Figure 26: Immunostaining of MCF-7 cells with GH56, 49, 37, and 61, listed in decreasing intensity of mean fluorescence in flow cytometry as seen in figure 24. *Nanobodies are in red and DAPI staining is in blue. Scale bar = 5 µm. GH56 shows more binding than the others. GH61, in accordance with flow cytometry, shows no binding and has been used here as an example of a non-binding control. GH49 also shows low binding to MCF-7 in confocal microscopy, as is seen in flow cytometry. This data has been published in Khilji et al. 2022* [291] *and figure has been modified for the thesis.*

Interestingly, while GH56 shows low binding in flow cytometry, it shows better binding than GH37 and GH49 in confocal microscopy. However, it was the worst binder in flow cytometry with living cells in comparison to fixed cells in the microscopy and, therefore, not pursued further.

Because GH37 is a relatively stable Nb compared to most other Nbs in the repertoire and it showed binding, albeit low, to MCF-7 cells, I wanted to test its binding to Globo-H further. To this end, STD-NMR was carried out by our collaborator (**Figure 27**). The method details are described in **section 5.1.3** as part of the supplementary information.

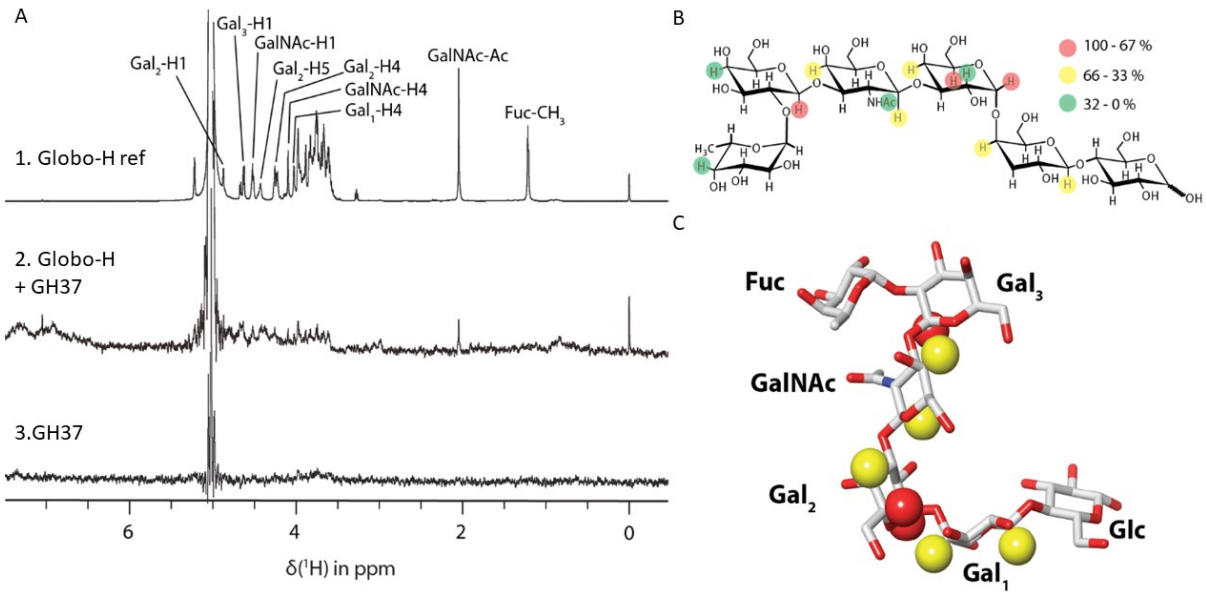


Figure 27: STD-NMR of GH37 with Globo-H. A. *STD-NMR spectrum. 1. Globo-H reference shows isolated signals of sugars in Globo-H. 2. STD spectrum of Globo-H with GH37. 3. GH37 shows peaks of nanobody only.* ***B.*** *and* ***C.*** *Red indicates glycan area that shows the highest binding (100-67%), yellow indicates intermediate binding (66-33%), and green indicates low or no binding (32-0%). From this it can be deduced that GH37 binds GB5.*

GH37 interacts with Gal2 (66-33%), Gal3 (100-67%), Gal5 (100-67%) (labelled Gal_1, Gal_2, and Gal_3 in figure, respectively) as well as GalNAc (66-33%). Combined, the results from the cell binding assays and STD-NMR show that GH37 interacts with the sugars of GB3 and GB4, but it requires the end terminal Gal5 to bind. Therefore, GH37 binds GB5. As GB5 expression on MCF-7 cells is similar to that of Globo-H (~91.7%)[82], the low binding of GH37 to MCF-7 (**Figure 25** and **Figure 26**) could be either because it is a specific yet weak binder of GB5 or because it was generated against synthetic glycan, and the orientation of native glycans on cells make binding difficult. Either way, as GB5 is also a target candidate for cancer immunotherapy [82,232], GH37 could be of great interest and should be studied further. It also shows that even if many of the Nbs did not show good binding on MCF-7 cells, there is a chance they probably do bind to parts of the Globo-H moiety. However, as the thesis was focused on finding a Globo-H-specific binder, GH37 was not pursued further.

Nbs were also tested for binding on HEK293 cells that express GB2, GB3, and GB4 but not GB5 and Globo-H[300] to exclude binding to the lower molecular weight globo-family glycans. Results of weaker MCF-7-binders, including GH37, are shown in **Figure 28**, while results of GH46 will follow later. GH37 did not bind HEK293 cells, which corroborates the STD-NMR results that GH37 binds GB5 instead.

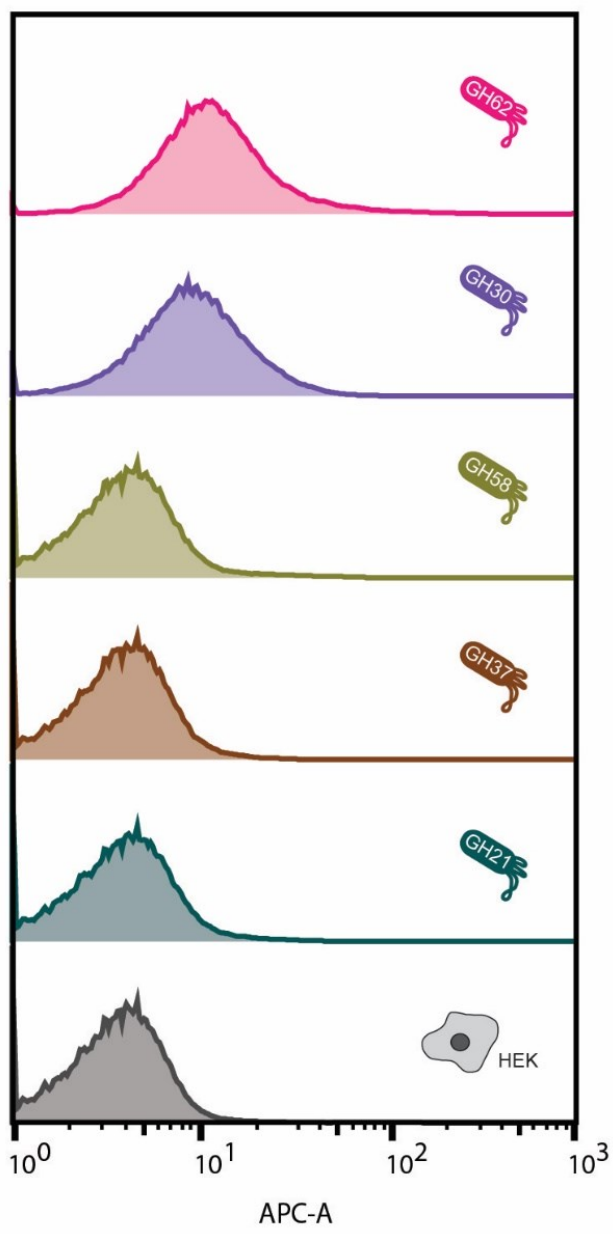


Figure 28: Weak MCF-7-binding nanobodies tested on HEK293 cells for binding to lower molecular weight Globo-family members including GB2, GB3, and GB4 in flow cytometry. *GH62 and GH30 show binding to HEK293 cells and, therefore, probably have GB2, GB3, or GB4 as minimal binding epitope. Gray: HEK293 cells with secondary only; Pink: GH62; Violet: GH30; Olive: GH58; Brown: GH37; Sea green: GH21. This data has been published in Khilji et al. 2022* [291] *and the figure has been modified here for the thesis.*

Only 2 (GH62 and GH30) of the 5 lower binders showed binding to both MCF-7 (**Figure 25**) and HEK293 (**Figure 28**). This suggested that their minimal binding epitope is a lower molecular weight globo-family member. While this is again of potential interest, the scope of this thesis was to characterise a binder for Globo-H, therefore, these Nbs were also not pursued further. In conclusion, out of 36 Nbs, only 2 (GH46 (**see section**

3.2.2) and GH21) remained after the screening as potential globo-H binders. While both Nbs are poorly expressing proteins, GH46 expresses better than GH21 (with higher end yield, data not shown) as well as shows higher binding for MCF-7 cells (**Figure 25**). GH46 binding to breast cancer cell lines was, therefore, further analysed.

3.2.2. GH46 binds Globo-H expressing cells but not lower molecular weight Globo-family members

GH46 also showed binding on MCF-7 in flow cytometry as well as confocal microscopy where it was compared to the positive control (**Figure 29A and B**), VK9, the commercially used IgG3 Globo-H-binder[239]. Additionally, I excluded non-specific binding of the Atto-647 anti-His secondary Ab, by showing concentration-dependent increase of GH46 binding to MCF-7 cells (**Figure 29C**).

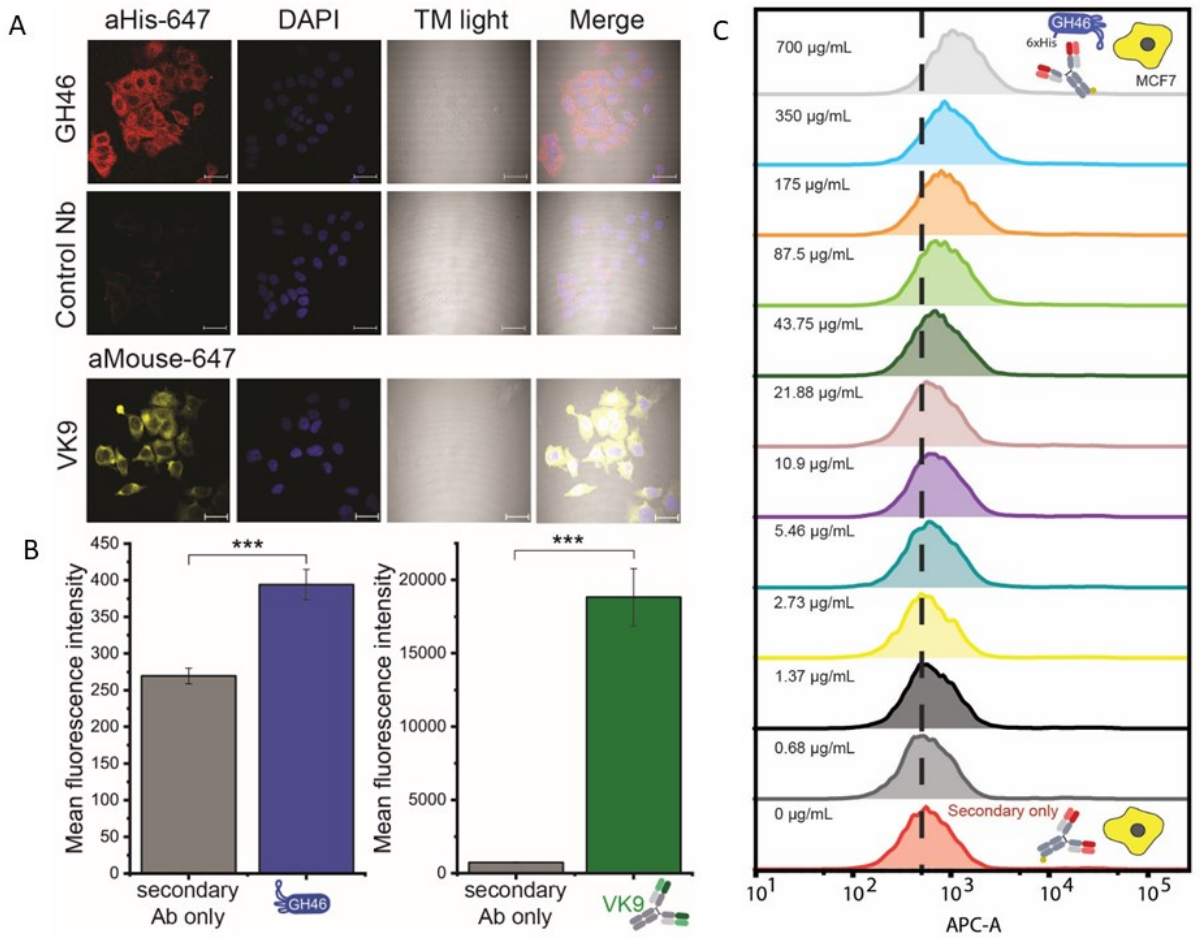


Figure 29: GH46 binding on MCF-7 cells. A. *Immunostaining was carried out with GH46 (red). VK9 (yellow) and non-binder Nb served as positive and negative controls. DAPI (blue) (scale bar = 5 µm). A different pseudocolour was used for VK9 to emphasize the distinct display settings as well as the fact that two different secondary Abs were used for GH46 and VK9. As negative control, a non-binder Nb was taken.* ***B.*** *Flow cytometry binding assay. Left panel: staining was carried out with anti-6xHis-Atto647N Ab in the absence (gray) or presence (blue) of GH46. Right panel: staining was carried out with anti-mouse-IgG-635 Ab in the absence (gray) or presence (green) of VK9. Y axis: Mean fluorescence intensity. Values represent mean ± SEM (n = 7). Differences were tested for significance via t-test. (***) represents p < 0.001.* ***C.*** *GH46 to MCF7 cells is specific and increases with Nb concentration. Concentration of GH46 is given. anti-6xHis-Atto647N Ab was used as secondary. This data has been published in Khilji et al. 2022* [291] *and the figure has been modified here for the thesis.*

Even though GH46 is a weak binder compared to VK9, it shows significant binding ($p < 0.001$ in t-test) to native Globo-H on MCF-7 cells (**Figure 29B**). Binding to lower molecular weight globo-family members was also excluded by testing GH46 on HEK293 and non-tumourigenic breast cell line MCF-10A (**Figure 30**), which like HEK293 also only expresses lower molecular weight Globo-family glycans and, therefore, serves as a good control[298]. VK9 was used as negative control.

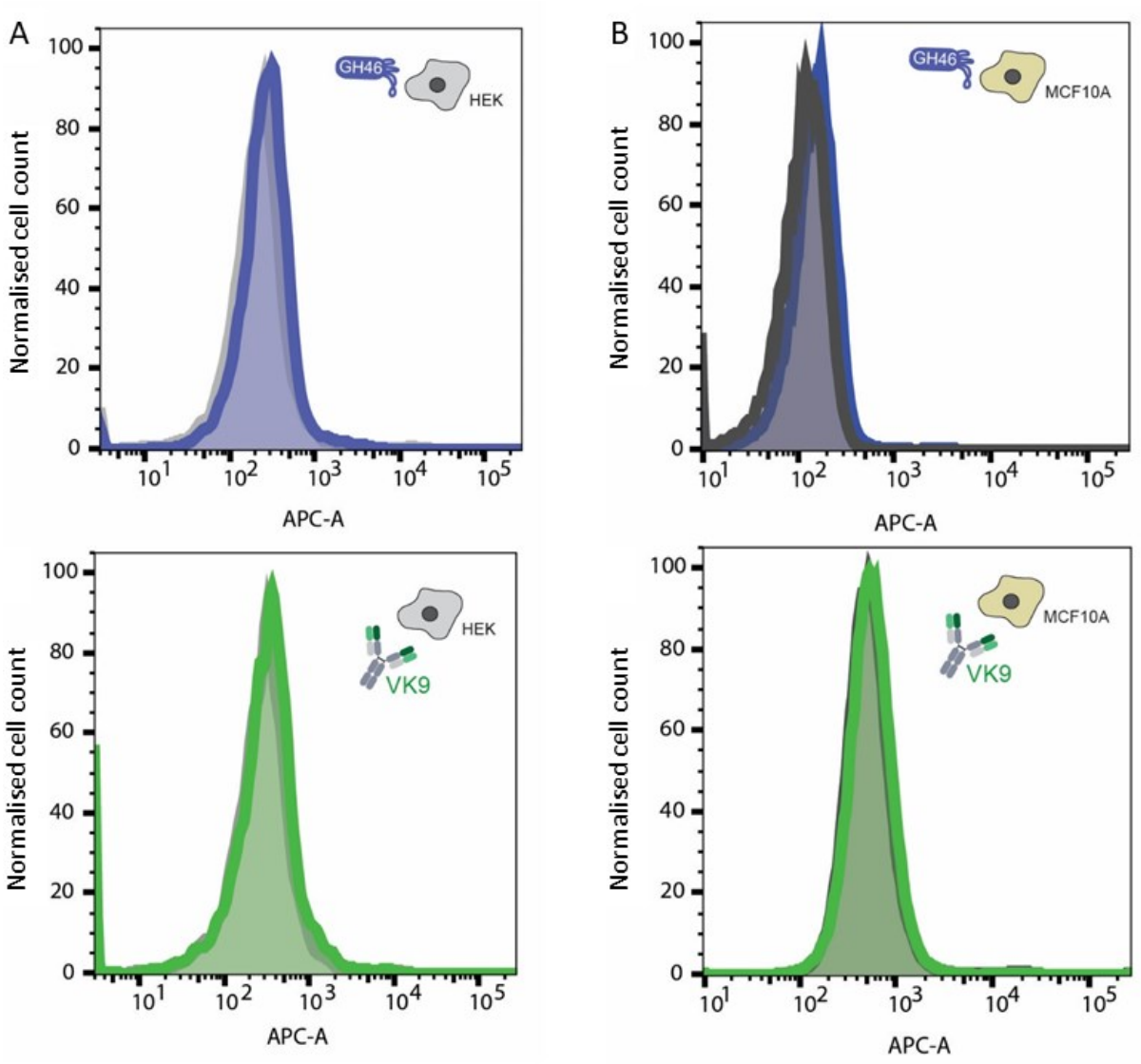


Figure 30: GH46 and VK9 binding to non-Globo-H and -GB5 expressing* cells.** ***A. *GH46 and VK9 binding to HEK293 cells (left panel).* ***B.*** *GH46 and VK9 binding to MCF-10A (right panel). GH46 is shown in blue and VK9 is shown in green, while secondaries anti-6xHis-Atto647N Ab and anti-mouse-IgG-635 Ab for Nb and VK9, respectively, are shown in gray. Both proteins do not bind to the cell lines showing their minimal binding epitope is not a lower molecular Globo-family member (GB2, GB3, or GB4) that is expressed on these cells. This data has been published in Khilji et al. 2022* [291] *and the figure has been modified here for the thesis.*

With this I established that GH46 is the best binder for Globo-H expressing MCF-7 cells. Next, I wanted to confirm GH46 binds to Globo-H by carrying out binding competition assays with Globo-H binder, VK9.

3.2.3. GH46 and VK9 binding competition assay

To test whether VK9, as a known binder of Globo-H, can inhibit binding of GH46 to MCF-7 cells, I co-incubated the IgG3 and Nb with the cells (**Figure 31**).

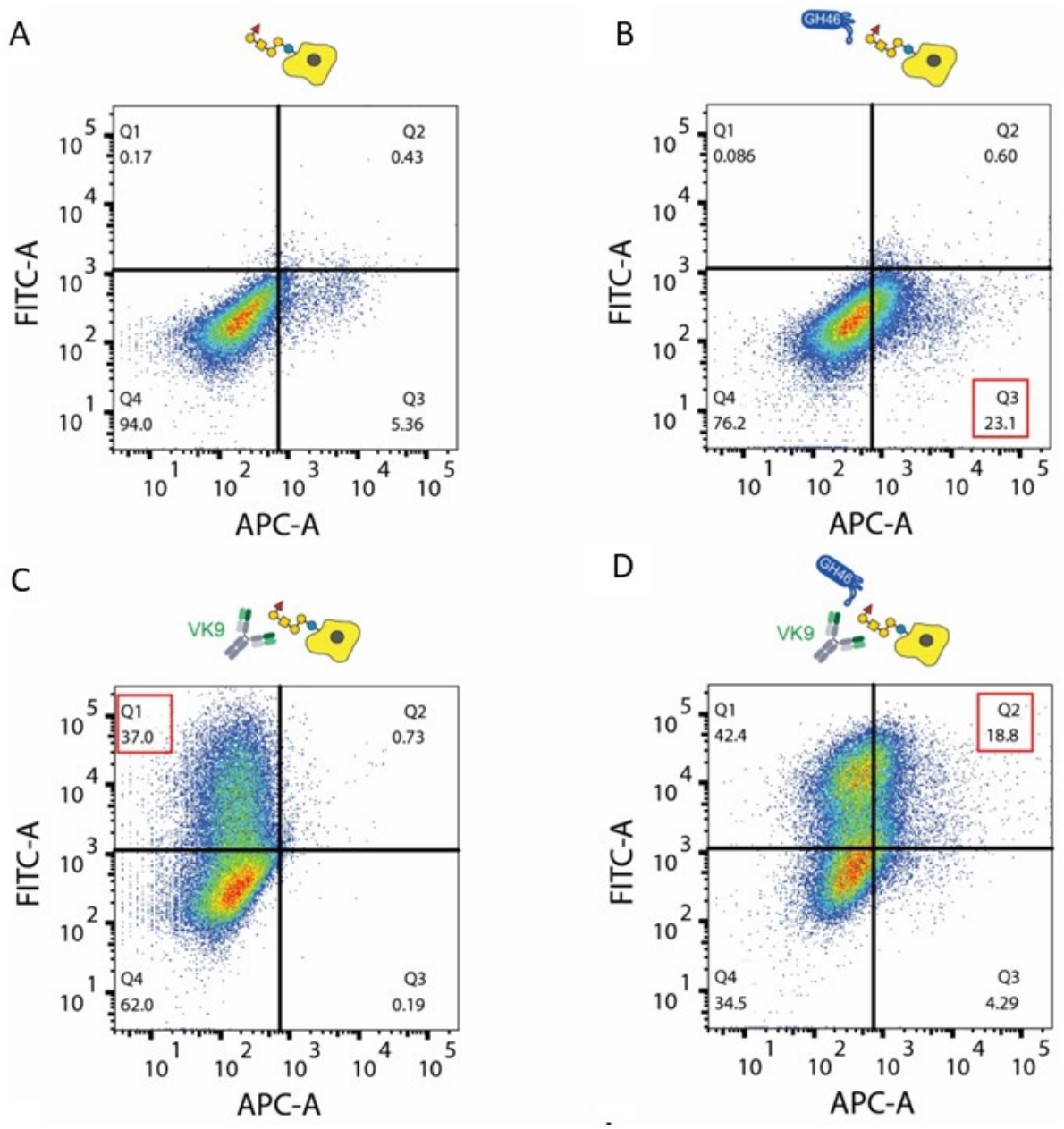


Figure 31: Co-incubation of MCF7 cells with GH46 and VK9. *Cells were simultaneously incubated with GH46 and VK9 and binding was detected with secondary Abs in the APC channel (x-axis) and FITC channel (y-axis), respectively.* ***A.*** *Control with anti-6xHis-Atto647N Ab and anti-mouse-IgG3 gamma 3-FITC Ab only.* ***B.*** *Incubation with GH46 only.* ***C.*** *Incubation with VK9 only.* ***D.*** *Co-incubation with VK9 and GH46. The co-stained cells shifted into the Q2 quadrant, showing simultaneous binding of both proteins. This data has been published in Khilji et al. 2022* [291] *and the figure has been modified here for the thesis.*

VK9 did not inhibit GH46 binding to MCF-7 cells (data shown with manufacturer's recommended concentration of 5 µg/mL for VK9 as mentioned in the Materials and Methods section). Experiments repeated with a high concentration of 20 µg/mL yielded the same result (data not shown). This was an interesting finding as previously it has

been shown that the minimal binding epitope of VK9 is the tetrasaccharide Fuc-Gal-GalNAc-Gal in which Fuc may be essential for binding[149], though detailed structural analysis of VK9 binding to Globo-H is not available. Therefore, it could be that GH46 binding was not inhibited because GH46 binds to Globo-H in a different manner. Therefore, I decided to chart out the minimal epitope of GH46 through binding assays with synthetic Globo family members. Along with this, I also wanted to determine the binding affinity of GH46 to Globo-H.

3.2.4. Charting minimal binding epitope of GH46 and determining its binding affinity with Microscale Thermophoresis (MST)

Based on the breast cancer cell binding assays, the minimal binding epitope of GH46 was expected to be either Globo-H or GB5. Glycan arrays were carried out to test this, however, even with different concentrations and a range of altered conditions, GH46 binding could not be detected on glycan array. This could be because it has low binding affinity and is washed away in the multiple washing steps of the experiment. Therefore, MST was employed instead: GH46 was incubated with Globo-H, GB5, GB4, and GB3 and the binding analysed. The binding affinities of GH46 interactions were also determined (**Figure 32**).

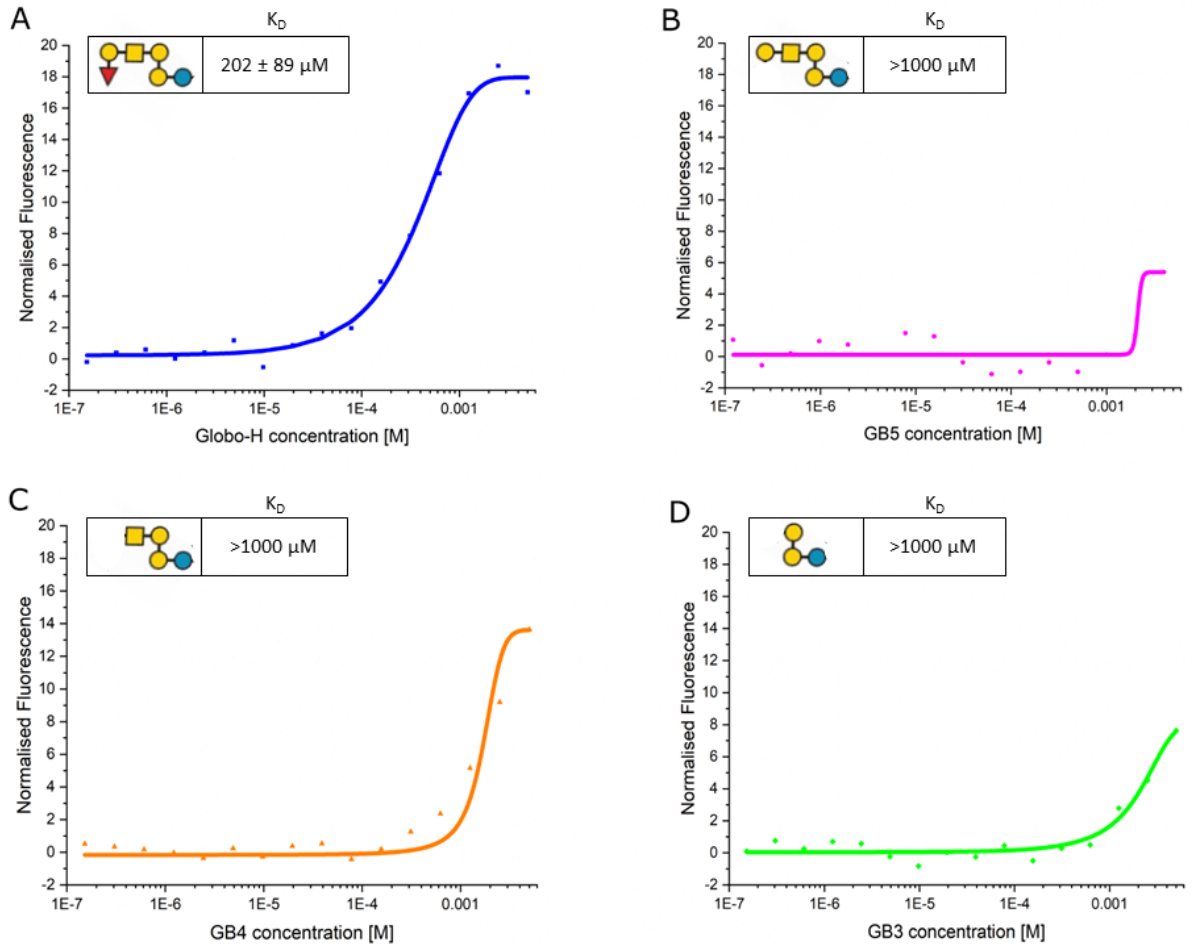


Figure 32: Measured in-solution K_D of GH46 and synthetic glycans (table) in MST affinity curves. ***A.*** *Globo-H (blue).* ***B.*** *GB5 (pink).* ***C.*** *GB4 (orange).* ***D.*** *GB3 (green). Values represent mean ± SEM (n ≥ 3). GH46 binding affinity to GB3, GB4, and GB5 is >1000 µM, while the binding affinity to Globo-H lies at 202 ± 89 µM. The plots were generated in OriginPro 2021b using the ‚dose response' parameters. This data has been published in Khilji et al. 2022*[291] *and the figure has been modified here for the thesis.*

GH46 binds to Globo-H with a high micromolar affinity of 202 ± 89 µM but does not bind to other lower molecular weight Globo-family members. This indicates that Fuc is important for the binding. To validate this data, STD-NMR of GH46 was performed with Globo-H and GB5 by our collaborator (**Figure 33**). The method employed for the STD-NMR is described in the supplementary information (**section 5.1.3**).

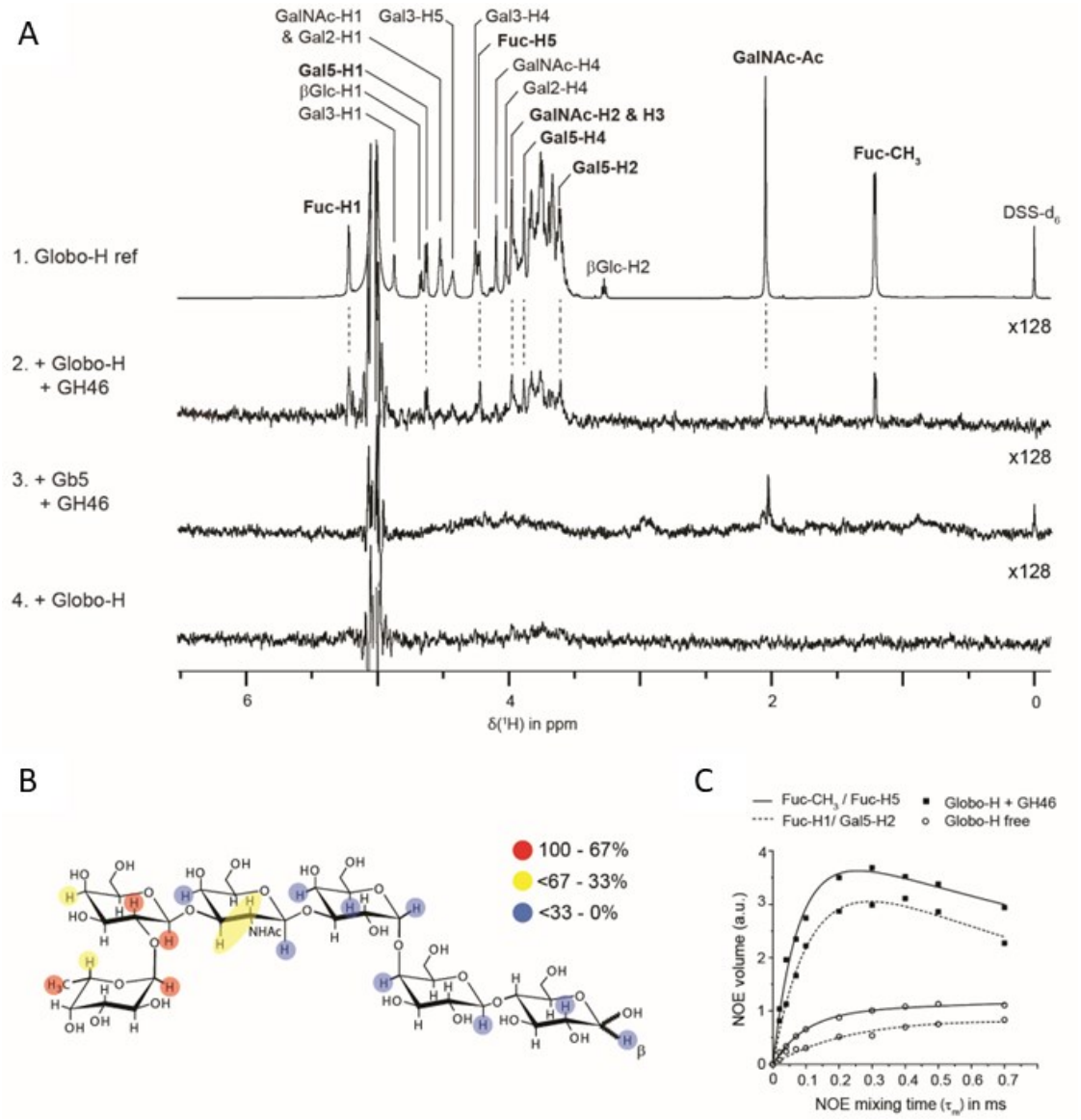


Figure 33: STD-NMR of GH46 interaction with Globo-H and GB5. A. *1. Globo-H reference spectrum. Isolated signals showing magnetisation transfer are highlighted in bold. 2. STD spectrum of 4 mM Globo-H with 60 µM GH46. 3. STD spectrum of 4 mM GB5 with 110 µM GH46. (4) STD spectrum of 4 mM Globo-H. So that Fuc and Gal3 anomeric protons could be observed, residual water signal was not suppressed.* ***B.*** *Binding epitope of Globo-H bound to GH46 with Fuc and Gal5 protons along with some interactions from GalNAc.* ***C.*** *Nuclear Overhauser Effect (NOE) build-up curves of Globo-H in presence and absence of GH46 against mixing time [ms] at 600 MHz and 298 K. Curves represent intra- (Fuc-CH3/Fuc-H5) and inter-pyranose (Fuc-H1/Gal5-H2) connectivities. This data has been published in Khilji et al. 2022*[291] *and the figure has been modified here for the thesis.*

STD-NMR is known for studying interactions in the micromolar range[301], which was in accordance with the MST results. Due to signal overlap, only the STD amplification factors for a subset of Globo-H proton signals could be measured (**Figure 33A**), however it could be established that interaction between GH46 and Globo-H is mediated by the

Fucα(1–2)Galβ moiety (**Figure 33B**). In comparison, no STD signals were observed with GB5 and GH46, which also corroborates the MST results suggesting that Fuc is required for GH46 binding. Additionally, Globo-H transferred (tr)NOEs signals were measured [302] with and without GH46 that showed that the NOE build-up rates are significantly higher with GH46 present further confirming GH46-Globo-H interaction (**Figure 33C**).

Taken together, the data from the GH46 and VK9 competition assay, MST, and STD-NMR suggest that even though VK9 does not inhibit GH46 binding to native Globo-H, GH46 binds to the Fucα(1–2)Galβ epitope in Globo-H and Fuc is essential for the binding. Further studies are required for a detailed binding analysis of VK9 to Globo-H and its exact epitope needs to be determined to establish differences between GH46 and VK9 binding.

Next, I aimed to functionalise the Globo-H binding GH46 in different ways.

3.2.5. Functionalisation and multimerisation of GH46

Various attempts of functionalisation of GH46 were carried out in which cloning was the main method of choice: green fluorescence protein (GFP from *Aequorea victoria),* AfcA catalytic domain only (synthesized from α-1,2-L-fucosidase of *Bifidobacterium bifidum;* plasmid was kindly gifted by Prof. Dr. Takane Katayama (Kyoto University, Kyoto, Japan)), or lytic peptide tachyplesin I (from *Tachypleus tridentatus* (peptide sequence: KWFRVYRGIYR)) were successfully cloned into the GH46 plasmid and DNA sequences were confirmed (Eurofins Genomics); however, the test expressions showed no protein expression for the lytic peptide fusion. Soluble GH46-GFP bound to MCF-7 cells even less than negative control of GFP alone and did not bind to synthetic Globo-H at all showing that GH46 lost its binding ability. Even with a molecular weight of ~108 kDa, GH46-AfcA catalytic domain had high expression in bacterial expression cells and large amounts of soluble and highly pure protein were obtained, however GH46-fucosidase-catalytic domain bound neither to cells nor to synthetic glycan also. The reason for this could be the bigger sizes of the fused proteins compared to the Nb as well as the high micromolar affinity of GH46 to start with. AfcA catalytic domain also contains three cysteines that could have formed disulphide bonds with cysteines of GH46 and disrupted its conformation, hence, causing it to lose its binding ability. Similarly, GFP also contains 2 cysteines in the sequence. For the lytic

peptide fusion, the plasmids were then ordered commercially. This included a plasmid for GH46 trimer fused with lytic peptide. While these fusions also did not yield any protein either, the plasmid with GH46 trimer was used to clone a stop codon before the lytic peptide and generate a trimer of the Nb instead. One of the reasons the lytic peptide fusion did not work could be that the lytic peptide is lethal for the bacteria and, hence, is not produced or kills the bacteria that produce it. However, this mini project was not pursued further, and expression was not tested in other expression systems.
Dimerisation of GH46 was also attempted by fusing the Nb with obligate dimerization domains p50 (a subunit of NF-κB1 with molecular weight of 13 kDa[303]) or Hsp70-interacting protein dimerisation domain (Hip-DD with a molecular weight of 5 kDa[304]). Plasmids with dimerisation domains were kindly gifted by Prof. Sutapa Chakrabarti, PhD, (Department of Biology, Chemistry and Pharmacy, Freie Universität Berlin). Here also, the cloning and protein expression were successful, even though the constructs were prone to degradation. However, the constructs did not bind MCF-7 cells, either. Again, the reason could be mismatched disulfide bonds with cysteines in the fused proteins as Hip-DD has 1 cysteine and p50 has 2 cysteines that could have changed the conformation of the Nb. Table shows a compilation of the various attempts of functionalisation of GH46.

Table 2: Functionalisation attempts for GH46 through cloning.

Protein/domain added	**Source**	**Successful expression**	**Construct binds well to MCF-7 cells**
GFP	*Aequorea victoria*	Yes	No
AfcA (α-1,2-L-fucosidase) catalytic domain	*Bifidobacterium bifidum*	Yes	No
lytic peptide Tachyplesin I	*Tachypleus tridentatus*	No	-
p50	subunit of NF-κB1	Yes	No
Hip-DD (Hsp70-interacting protein)	-	Yes	No

GH46-Trimer (see below)	Commercially ordered bacterial expression plasmid	Yes	Yes

After having established that GH46 is a Globo-H binder albeit with high micromolar affinity, one of the next aims that I had was to functionalise GH46 to improve its binding affinity. The single-domain characteristic of Nbs allows for relatively easy modifications for functionalisation. Nbs have been "mixed and matched" before to obtain a single construct binding multiple targets (multispecific), identical targets (multivalent), or different epitopes in the same target[284]. Avidity effects can be increased with multimerisation[273,285]. GH46-trimer was successfully cloned and expressed, however the trimeric Nb was not soluble and instead ended in inclusion bodies. The protein purification and solubilisation are described in **section 2.2.3.**

Similar to GH46, GH46-trimer binding in glycan array could not be detected, again probably due to low binding affinity. However, binding was detected in MCF-7 cell binding assays with immunostaining and flow cytometry (**Figure 34** and **Figure 35**). As previously the other GH46 fusions did not bind well to MCF-7 cells, the binding of GH46-trimer indicates proper folding and conformation of the three Nb monomers in the construct.

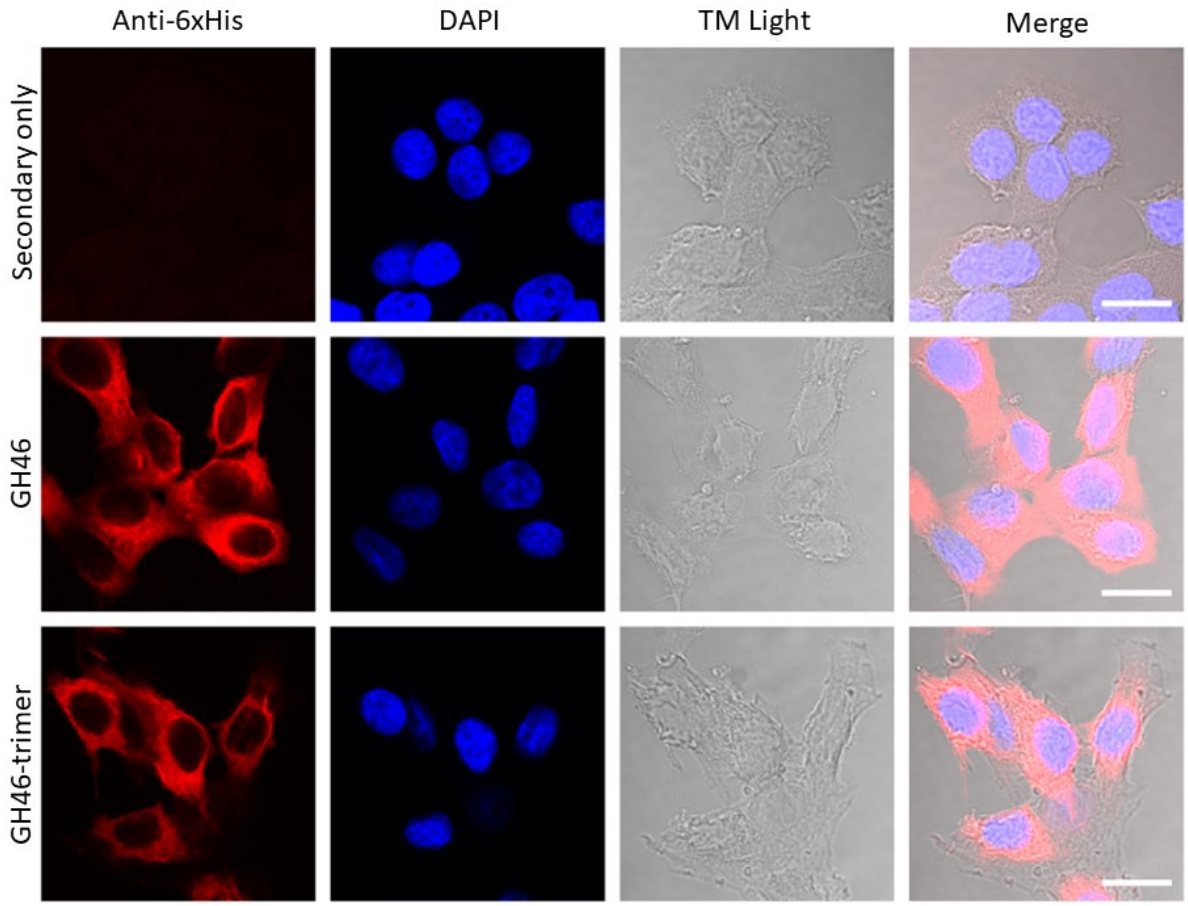


Figure 34: Immunostaining of MCF-7 cells with GH46 and GH46-trimer. *Anti-6xHis-Atto 647 was used as secondary Ab. Secondary Ab alone was used as negative control. Brightness and contrast were similarly adjusted for all images for a comparison of binding. Red: protein binding on cells; Blue: DAPI. Scale = 20 μM.*

GH46 and GH46-trimer showed similar intensity of binding in immunostaining of fixed cells, however, flow cytometry showed improved binding with multimer on living cells suggesting binding avidity was increased (**Figure 35**). I tested a range of concentrations of GH46, GH46-trimer, and VK9 on MCF-7 cells to calculate and compare the apparent K_D for cell binding.

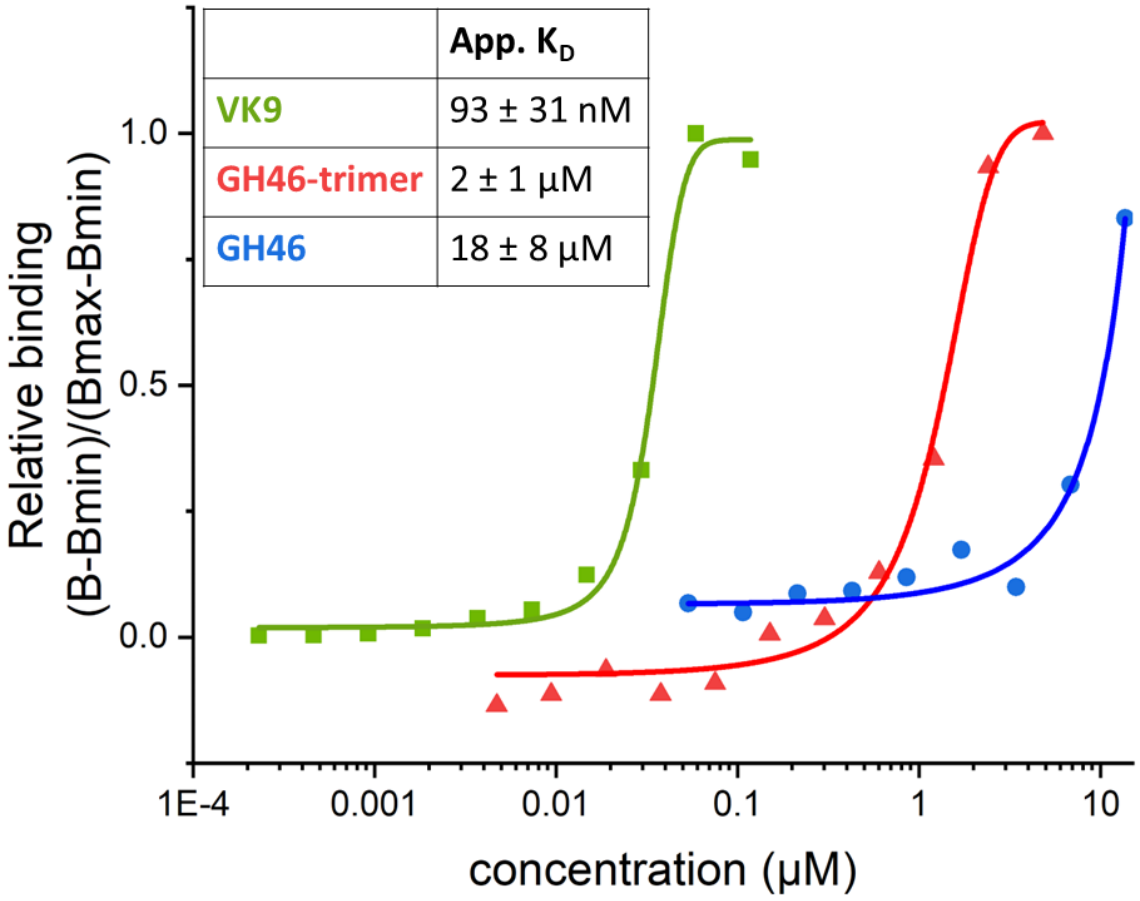


Figure 35: Relative MCF-7 binding of GH46, GH46-trimer, and VK9 over concentration increase. *Representative affinity curves are shown for GH46 (blue), GH46-trimer (red), and VK9 (green). Table shows calculated apparent K_D. Values represent mean ± SEM (n ≥ 3). This data has been published in Khilji et al. 2022*[291] *and the figure has been modified here for the thesis.*

The apparent K_D of GH46-trimer (2 ± 1 µM) for binding to native Globo-H on MCF-7 cells was 9-fold better than that of GH46 (18 ± 8 µM). This confirms that at a molecular weight approximately one-third of a conventional IgG (~49 kDa *vs.* ~150 kDa), the multivalent GH46-trimer improved in avidity compared to the monovalent GH46. This can be advantageous as it has been shown before that HER2-specific Nb-dimers can improve apparent binding compared to the monomers as well as penetrate tumour more homogeneously than a conventional IgG due to their smaller size[273]. Hence, due to the single-domain nature of Nbs and generally low protein-glycan binding affinity[305], multimerisation offers a viable approach in boosting avidity of glycan-binding Nbs.

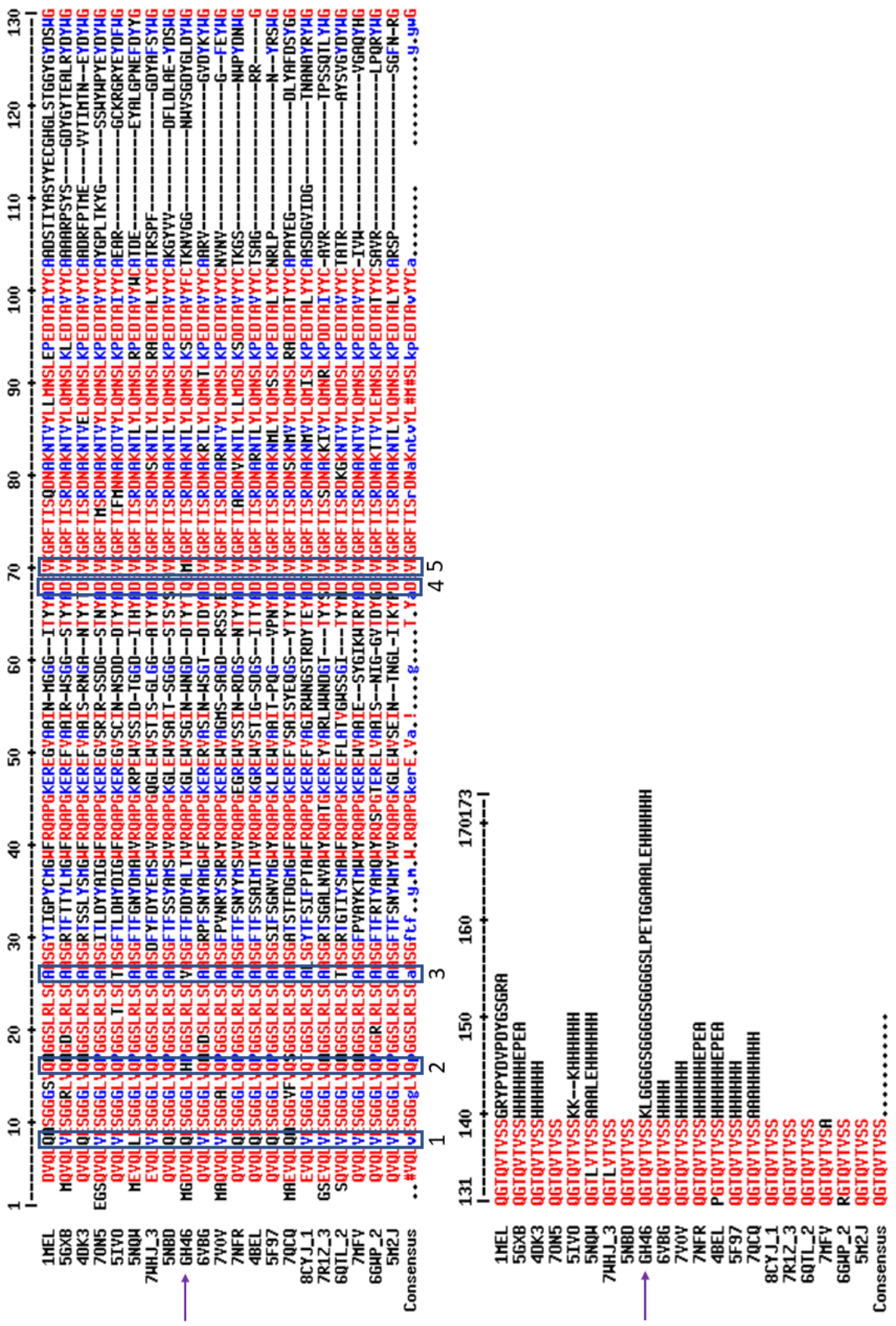
As mentioned previously, GH46 is a poorly stable protein. I next aimed to improve the solubility, thermal stability, and yield of the protein.

3.2.6. Improving solubility, stability, and yield of GH46

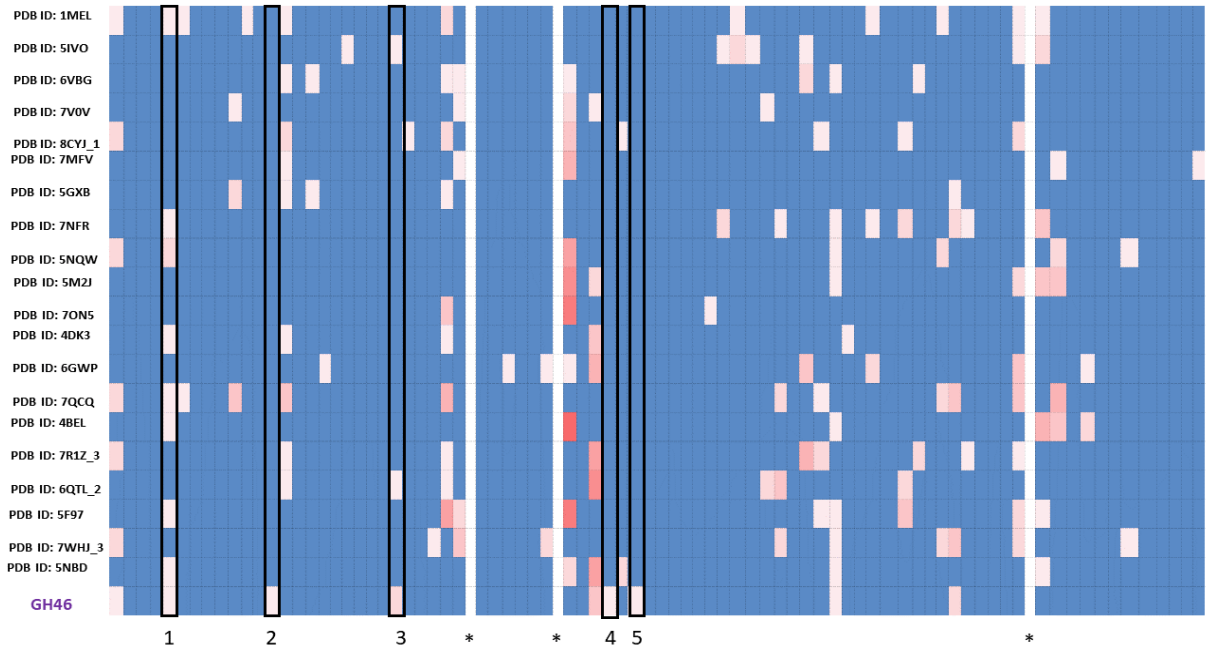
GH46 tends to aggregate easily and cannot be concentrated above 1 mg/mL without losing all the protein. Its average end yield is ~0.8 mg/1L of expression bacterial culture. To improve this, alterations in various conditions were studied: 1) different culture media including LB, TB, and magic medium; 2) different temperatures for initial lag phase growth ranging between RT and 37°C; 3) different concentrations of IPTG; 4) different temperatures for protein expressing log phase ranging between 12 and 30°C; 5) Ni^{2+}-purification at RT or 4°C; 6) cloning with a GST-tag for solubilisation[306]. The pH used for Ni^{2+}-purification in buffers in lab lies at 8 and PBS pH 7.4 was used for SEC. As the theoretical pI of GH46 lies at 5.91 (ProtParam tool, Expasy), these pH values should theoretically also not cause aggregation. However, GH46 solubility and stability remain an issue. To improve on this, I analysed the amino acid sequence of GH46 in comparison to those of 20 randomly selected Protein Data Bank (PDB) published Nbs as they already have crystal structures available, suggesting they are highly stable (**Figure 36A**). My hypothesis was that the highly stable Nbs may have common amino acids in the FRs that could help stabilise the Nb, but not affect its structure and the binding to its epitope. Nbs that were modified in any manner were not considered. In case more than one Nb chain was found in the PDB entry, I selected a chain at random too. **Figure 36B** shows a heat map of the differences in amino acids in the FR region sequences only of 20 Nbs as well as GH46. I identified 5 amino acids common in the core of highly stable Nbs, which GH46 did not have (marked as 1-5 in **Figure 36** as well as in following text). I studied the mutations one after the other to pinpoint which ones would be of high importance to the stability of a Nb.

I selected three amino acids in GH46 FR1 and two in FR3: Gln5Val (1), His13Gln (2), Val23Ala (3), Gln62Asp (4), and Met64Val (5) (**Figure 36**). 13-His (2), 62-Gln (4), and 64-Met (5) are unanimously Gln, Asp, and Val in stable Nbs, respectively, while 23-Val (3) is mostly found as Ala and less frequently as Thr (in 2/20 Nbs). 5-Gln (1) was either similarly Gln (in 7/20 Nbs) or Val (in 12/20 Nbs or Leu in case of 1/20 Nbs).

A



B



C

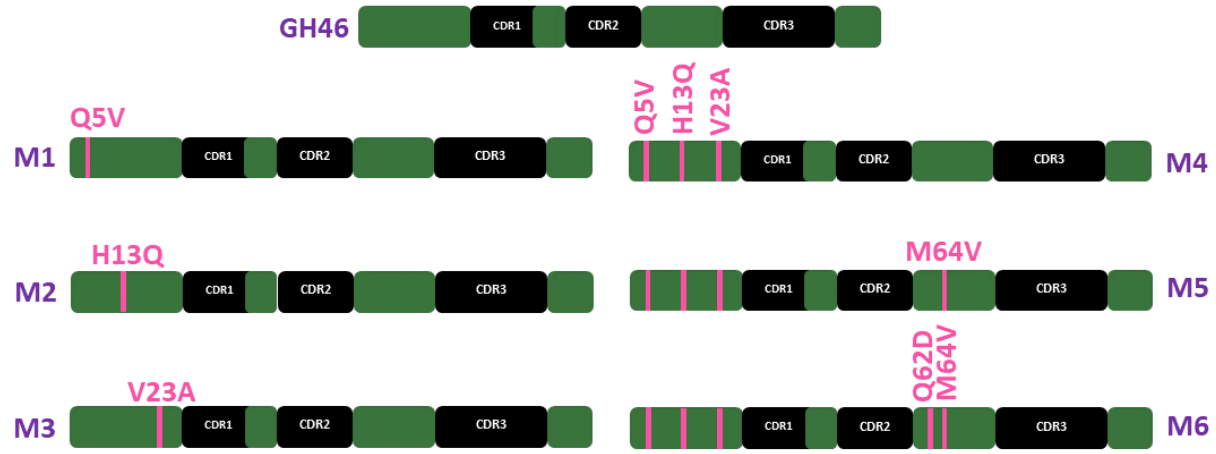


Figure 36: Identifying amino acids in stable Nbs that are not present in GH46. A. *Comparison of the complete sequences of 20 stable Nbs entered into the PDB Bank with GH46 (marked with a purple arrow). PDB IDs are provided. The amino acids marked with black boxes and numbers were mutated to study their effects on Nb stability. Figure was made using the Multalin interface page*[307]. ***B.*** *Heat map for differences at each amino acid position in the Nb FR sequences. Differences in FR1-4 were analysed. * = CDRs1-3. The amino acids mutated in FR1 and 3 are marked with black boxes and numbers.* ***C.*** *Schematic representation of GH46 and the FR cloned mutants. Mutated amino acids are specified along with the names of the mutants, M1-6.*

As mutant 1 (M1: GH46-Gln5Val (1)), M2 (GH46-His13Gln (2)), and M3 (GH46-Val23Ala (3)) gave lower end yield (data not shown) than GH46, they were not pursued further. M4 (GH46-Gln5Val (1), His13Gln (2), and Val23Ala (3)) had all three mutations

in FR1 (**Figure 36**). With analytical SEC, the yields of the proteins (M4, M5, and M6) were compared with GH46 (**Figure 37**).

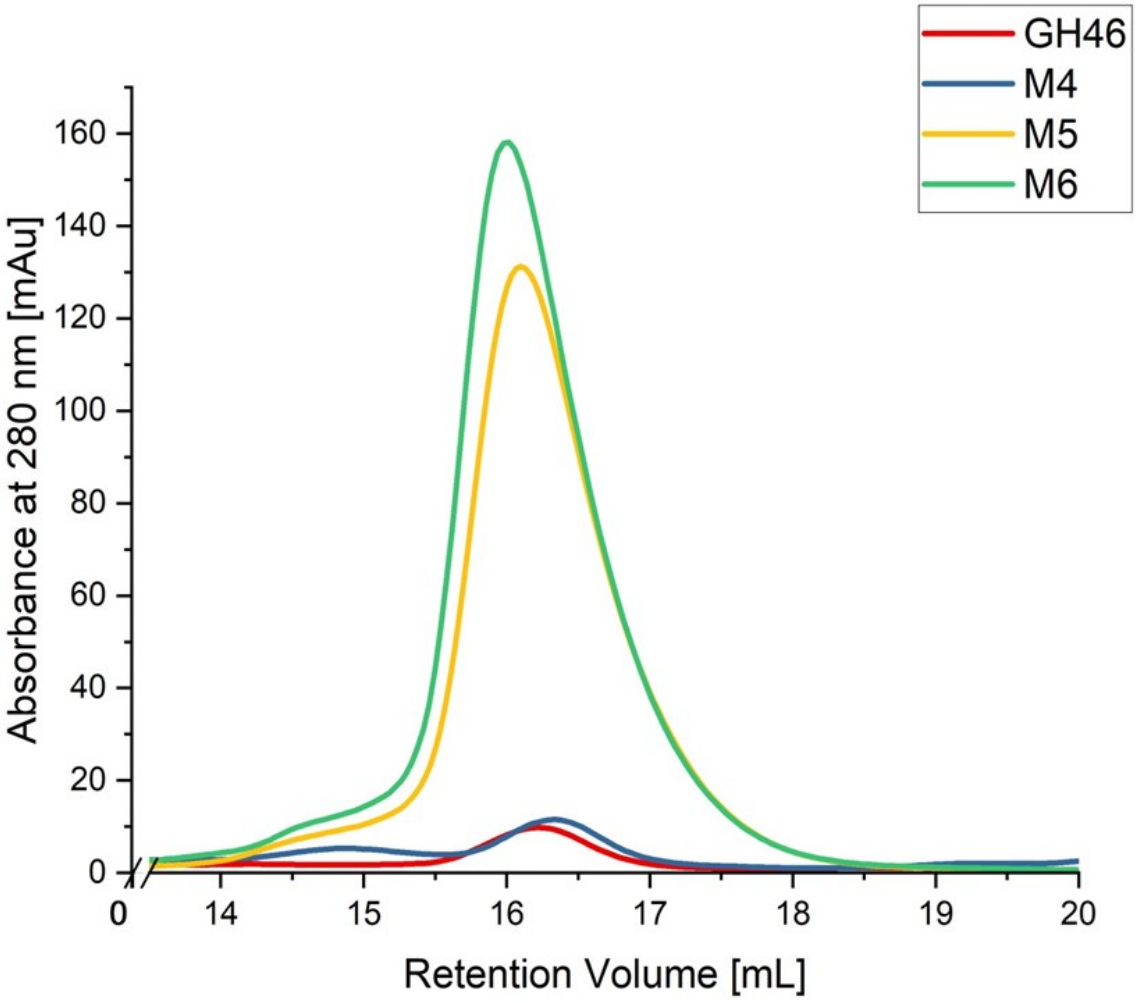


Figure 37: Analytical SEC of M4, M5, and M6 in comparison to GH46. *Results are from purification of proteins from 100 mL bacterial expression culture. Chromatograms have been overlaid for comparison. GH46: red; M4: blue; M5: yellow; M6: green.*

M4 had similar yield to that of GH46 (~13 mAu), but the protein aggregated when concentrated after SEC and there was no protein left for further analysis. M5 is a combination of four mutations: Gln5Val (1), His13Gln (2), Val23Ala (3), and Met64Val (5). The protein yield and purity increased dramatically (~10-fold with ~135 mAu), which was fascinating as it was owing to just one additional mutation. M6 includes all four mutations as well Gln62Asp (4). Here, the protein had even higher yield than M5, which was measured at ~160 mAu (**Table 3**).

Next, DSF was carried out to measure the unfolding of the proteins in correlation with temperature increase to study thermostability and, hence, aggregation behaviour of the mutants in comparison to GH46 (**Figure 38**).

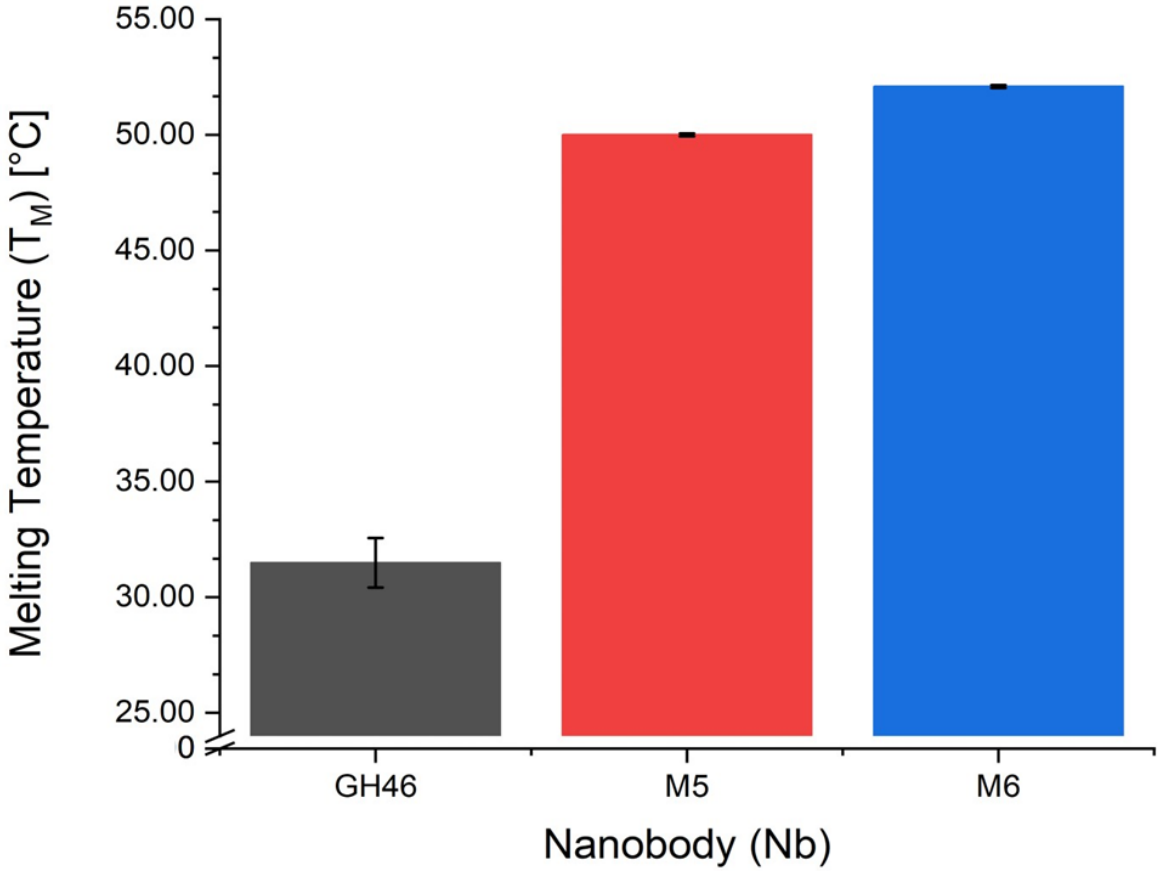


Figure 38: T_M of GH46, M5, and M6 from DSF studies carried out with SYPRO Orange dye. *Fluorescence was measured over temperature increase. Values represent mean ± SEM (n = 3). GH46: gray; M5: red; M6: blue.*

GH46 displayed the lowest T_M at ~31 °C after which it starts to irreversibly unfold, which would mean that the Nb cannot be used in the future in any clinical studies as it does not survive at physiological temperature (37°C). On the other hand, both M5 and M6 have comparatively high T_Ms of 50 °C and 52 °C, respectively, and, hence, would be stable at physiological conditions.

Gln5Val has previously been shown to increase Nb thermostability[279], however FR1 mutations Gln5Val (1), His13Gln (2), and Val23Ala (3) did not improve the solubility, and therefore thermostability, of proteins when taken single or in combination (M1-4). Interestingly, the polar, non-charged amino acid Gln-13 (2) is ubiquitous in stably expressed Nbs but mutating the positively charged His to Gln also did not help improve the Nb solubility on its own (M2) and neither did changing the more hydrophobic amino acid with a lesser hydrophobic amino acid in M3 (Val23Ala (3)).

The FR3 mutations (Met64Val (5) alone in M5 and in combination with Gln62Asp (4) in M6), however, not only improved solubility and protein yield (more than 10-fold,

Figure 37), but also thermostability of the Nb ($\Delta T_M \approx 20$ °C, **Figure 38**). In stable PDB Nbs, Gln62 and Met64 are unanimously Asp and Val, respectively (**Figure 36B**). Additionally, the relatively higher number of negatively charged conserved residues of FR3 that helps with the solubility of the V_{HH} domain has also been shown to include the conserved Asp62, while Val64 was also mostly conserved among Nbs[308]. This iterates the importance of FR3 in Nb conformation and stability.

Next step was to test M5 and M6 for binding to native Globo-H on MCF-7 cells as well as to HEK293 cells to rule out binding to lower molecular weight Globo-family members (**Figure 39**).

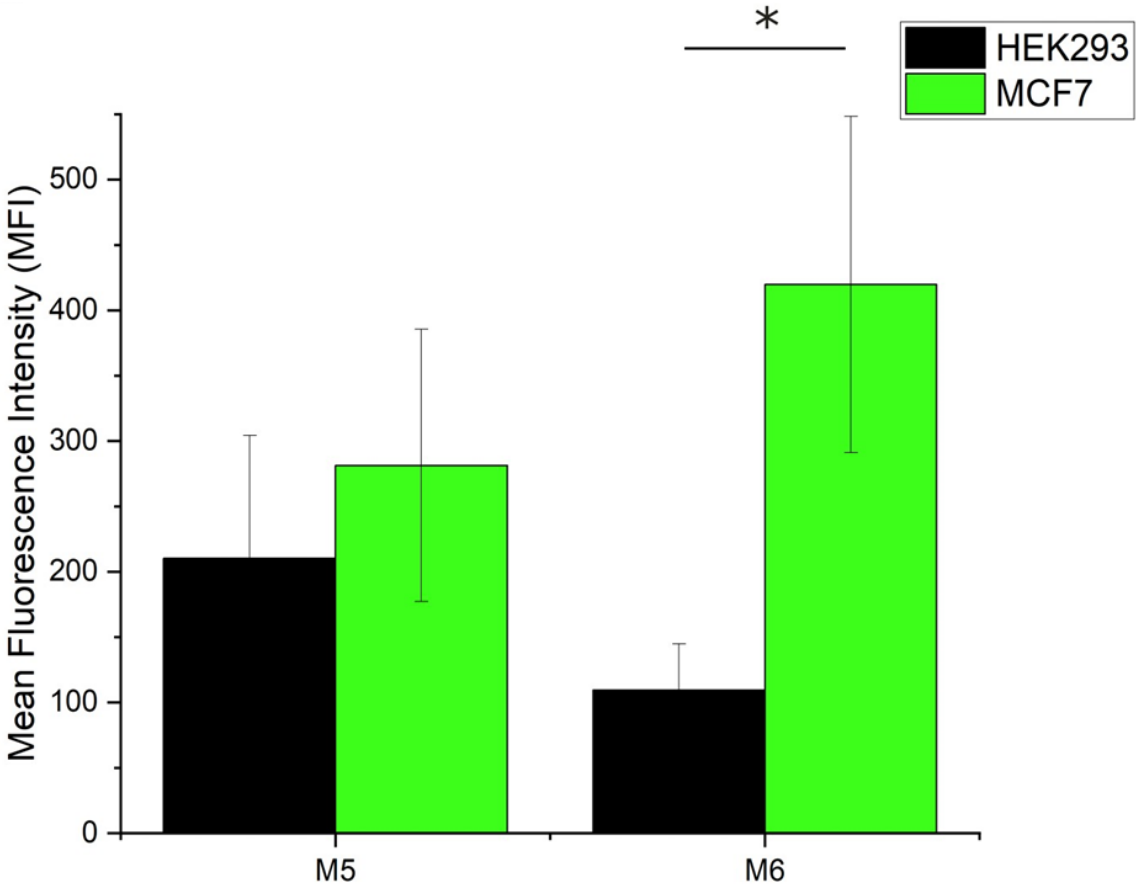


Figure 39: MFI of MCF7 (green) or HEK293 (black) cells with M5 and M6. *Assays were carried out at physiological temperature (37°C). Values represent mean ± SEM (n = 3). Statistical analysis was performed with unpaired t-test. *: $p < 0.05$.*

M5 and M6 both bound to MCF-7, however M5 binding was not significant when compared to HEK293 cells. This could probably be because of slight conformational changes of the binding pocket caused by Met64Val mutation alone that may have affected the specificity of the Nb for Globo-H, even though Val64 is a conserved residue in the FR3. Interestingly, K65 has been shown before to have the propensity

to interact with Nb epitopes[308]. If this is the case for GH46 also, the close proximity of Met64Val could have influenced this interaction. M6, however, retains the specificity of the parent GH46 as it bound significantly better to MCF7 than HEK293 ($p<0.05$). Therefore, if Met64Val does negatively affect binding of GH46 to Globo-H, the effect is probably minimal. Taken altogether, FR 1 and 3 mutations do not affect the binding site of the Nb and only help stabilise the protein. **Table 3** compiles the results of mutations, yield, T_M, and cell binding assays for M1-6.

Table 3: Results of analytical SEC, DSF, and cell binding assays for M1-6.

Nanobody	**Mutations**	**Absorbance at 280 nm**	**T_M**	**Significant native Globo-H binding**
M1	Gln5Val (marked as 1 in **Figure 36**)	Lower than GH46	-	-
M2	His13Gln (2)	Lower than GH46	-	-
M3	Val23Ala (3)	Lower than GH46	-	-
M4 (aggregated and could not be analysed)	Gln5Val (1), His13Gln (2), and Val23Ala (3)	Similar to GH46 - ~13 mAu	-	-
M5	Gln5Val (1), His13Gln (2), and Val23Ala (3), and Met64Val (5)	~135 mAu	50°C	No
M6	Gln5Val (1), His13Gln (2), and Val23Ala (3), Gln62Asp (4), and Met64Val (5)	~160 mAu	52°C	Yes

With M6, I was able to generate a FR-mutant of GH46, which contains the same CDR regions and, therefore binding site of Globo-H, but is a more stable protein. M6 also

retained the significant binding ability of parent GH46 to Globo-H-expressing MCF-7 cells.

GH46 binds specifically to Globo-H (**Figure 32** and **Figure 33**), however the K_D of the interaction lies around 202 ± 89 µM. Next, I wanted to apply a similar strategy and introduce mutations in the CDR regions to hopefully obtain a mutant that is a better binder of Globo-H than GH46, however retains the stability of M6 as well as GH46's specificity for Globo-H.

3.2.7. Improving binding affinity of GH46

C. Soliman *et al.* 2020 provide a crystal structure of anti-SSEA-4 chimeric Ab ch28/11 Fab with SSEA-4. The Fab-V_H CDRH3 101-Gly interacts with Gal3 and 102-Tyr holds down hydrophobic GalNAc to tightly keep SSEA4 in its binding pocket. Additionally, 56Ser and 58Asn of CDRH2 also interact with Gal3[246]. Globo-H and SSEA-4 differ only in their terminal sugars, Fuc and sialic acid, respectively (**Figure 6**). As discussed before, GH46 binding to Globo-H is Fuc-dependent, while Gal5 is also involved so I wanted to see if I could introduce amino acids into GH46 that would also interact with the other sugar residues in Globo-H and, therefore, enhance the binding of the V_{HH} and glycan (**Figure 32** and **Figure 33**)[291]. I compared the amino acid sequences of V_H-ch28/11 and V_{HH}, GH46 (**Figure 40**).

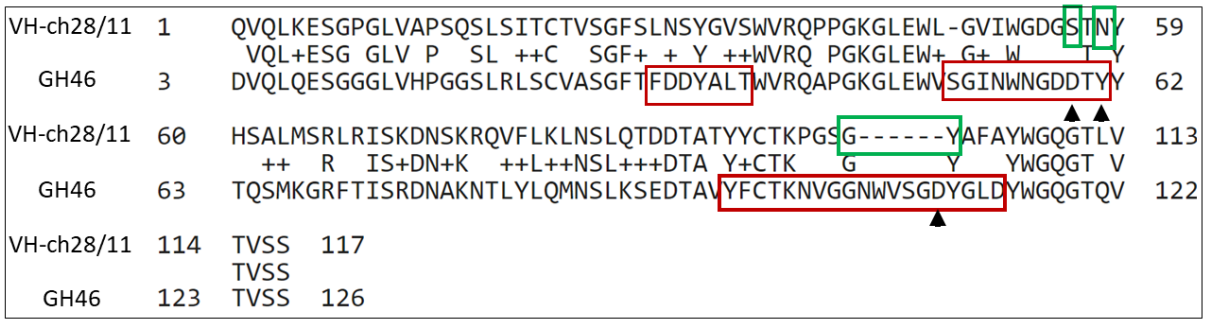


Figure 40: Amino acid sequence comparison of V_H-ch28/11 and the VHH, GH46 using Protein BLAST[296]. *The red boxes mark the three CDRs of GH46 and the arrows show where mutations (Asp59Ser, Tyr61Asn, and deletion of Asp110) in CDRs 2 and 3 were introduced. The green boxes mark the amino acids in CDR2 and 3 of V_H-ch28/11 involved in binding with Gal3 and GalNAc in the GB5 structure common to both SSEA-4 (ch28/11's target antigen) and Globo-H.* ***Figure 6*** *shows the structures of GB5, Globo-H, and SSEA-4.*

Similar to in V_H-ch28/11, Gly and Tyr are present in GH46 CDR3 at the corresponding positions (109 and 111 in the Nb sequence, respectively) and could be prominent in

GH46 interaction with Globo-H, however they are separated by the negatively charged Asp110 that could be weakening hydrophobic interactions. Also, instead of Ser and Asn in CDR2, GH46 has Asp and Tyr in the corresponding positions, respectively. I decided to test whether Globo-H binding could be enhanced by introducing the amino acids involved in Gal3 and GalNAc binding into the M6 (**Figure 41**) CDRs. In CDR3 Asp110 was deleted, while in CDR2 Asp59 and Tyr61 were substituted with Ser and Asn, respectively. Hence, three further mutants were generated that included the previous 5 mutations in the FRs: M7 also has Asp110 deleted in CDR3, M8 only has Asp59Ser and Tyr61Asn substitution point mutations in CDR2, and M9 contains all three mutations (**Figure 41**).

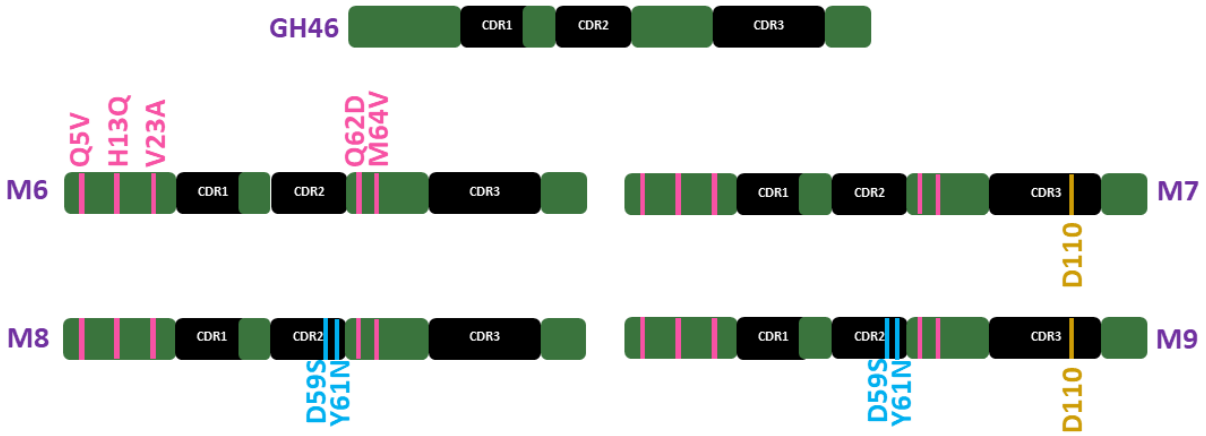


Figure 41: ***Schematic representation of CDR mutants M7-9.*** *Mutated amino acids are specified along with the names of the mutants. For M7, Asp110 was deleted, for M8 Asp59Ser and Tyr61Asn substitutions were carried out, and M9 contained all three mutations. Schematic representation of GH46 and M6 for comparison are shown.*

M7, M8, and M9 were studied in analytical SEC to determine their end yield from 100 mL bacterial cultures (**Figure 42**).

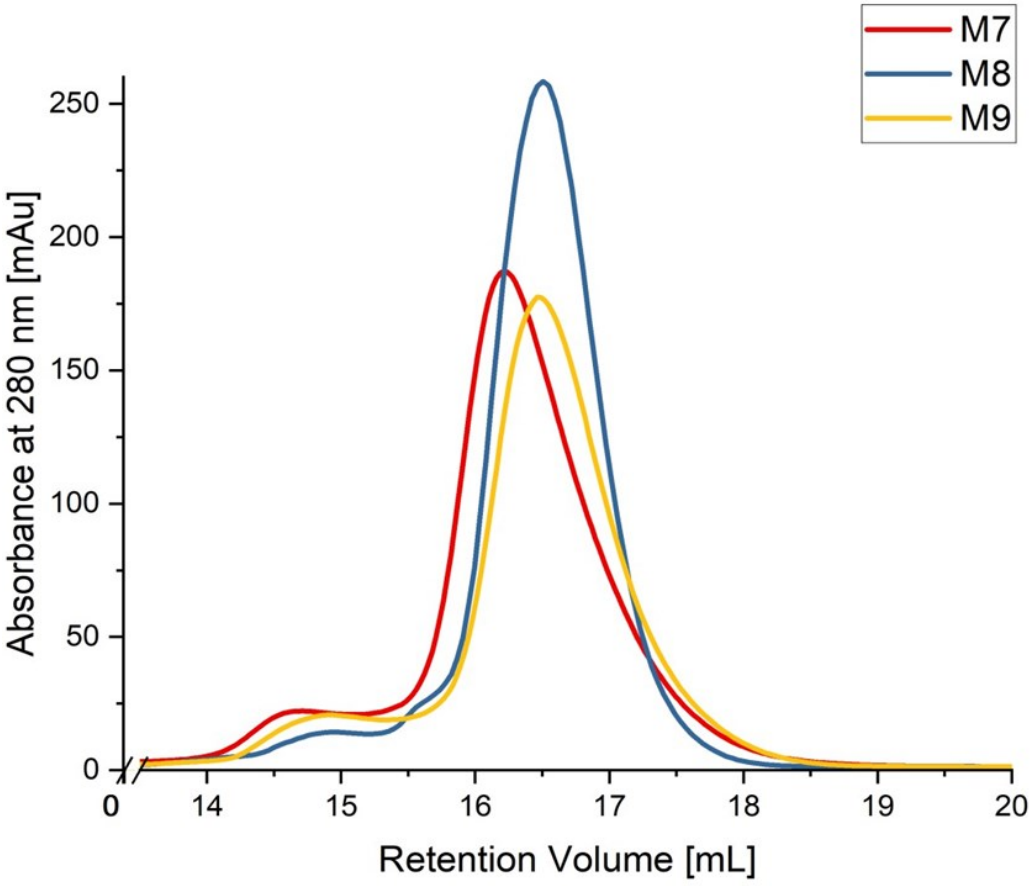


Figure 42: Analytical SEC of M7, M8, and M9. *Results are from purification of proteins from 100 mL bacterial expression culture. Analytical S200 increase 10/300 column was used. M7: red; M8: blue; M9: yellow. While M7 and M9 show similar yields of ~190 and 180 mAu, respectively, M8 has a higher end yield of ~260 mAu.*

M7, M8, and M9 yields were measured at ~190, ~260, and ~180 mAu, respectively. These are approximately 16-, 22-, and 15-fold higher than GH46, respectively as well as even higher values than those of M5 and M6 (~135 and ~160 mAu, respectively; **Figure 37** and **Table 3**). Therefore, high amounts of highly soluble and pure protein could be obtained. In fact, M9 was measured at ~80 mg/mL end yield, which is almost 100-

fold higher than the parent GH46 and could be concentrated up to 20 mg/mL without precipitation in comparison to only ~0.7 mg/mL of GH46.

Next, for M7-9 also, I assessed the thermostability through DSF and compared their T_Ms (**Figure 43**).

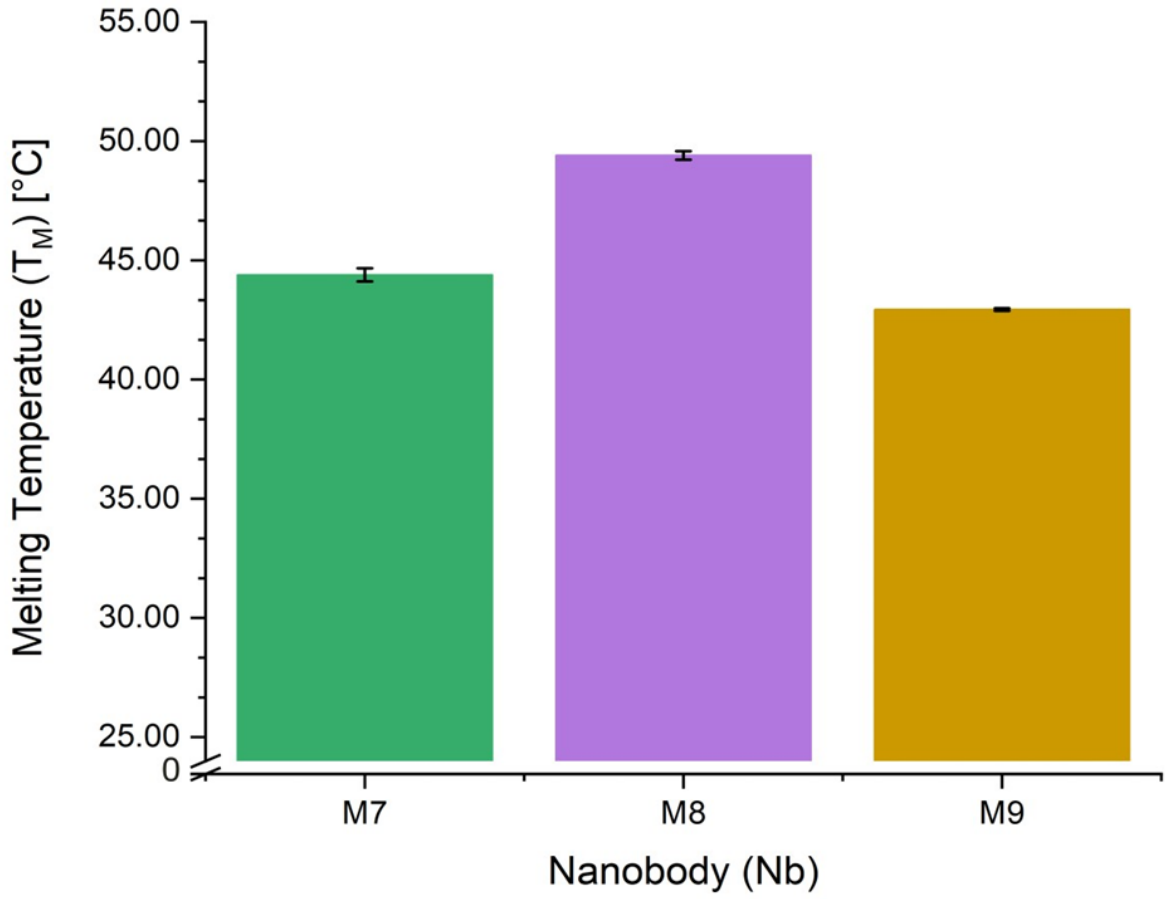


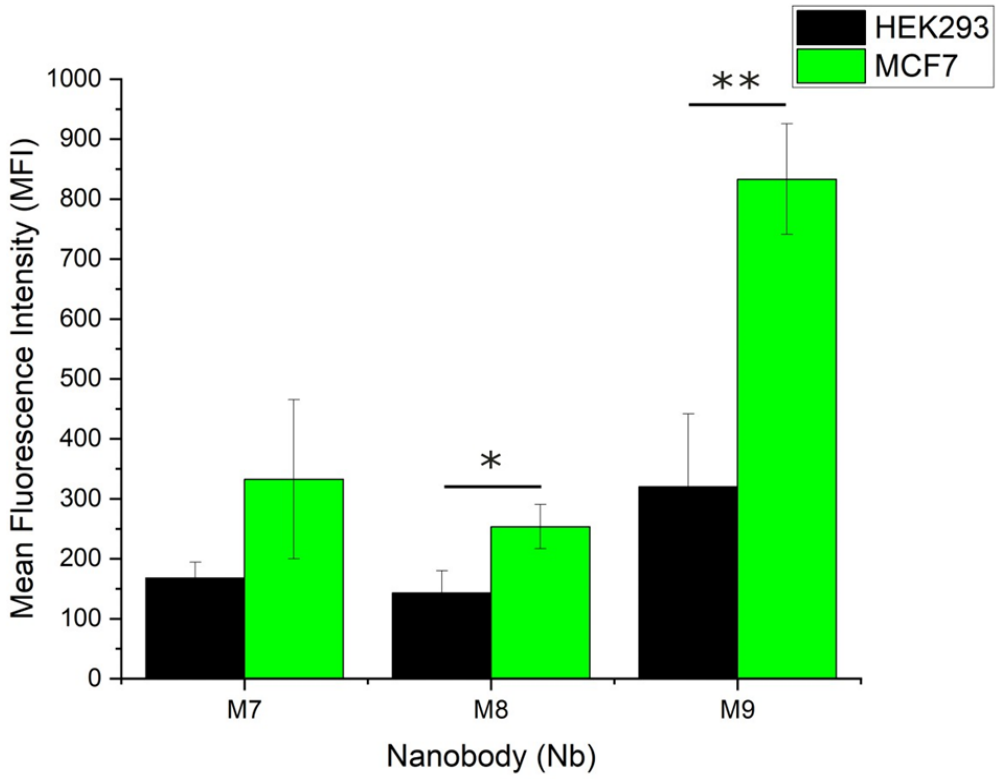
***Figure 43: Melting temperature of M7-9 from DSF studies carried out SYPRO Orange dye**. Fluorescence was measured over temperature increase. Values represent mean ± SEM (n = 3).*

While the T_M of M8 was ~50 °C, M7 and M9 had lowered T_Ms of ~45°C and ~43°C, respectively. The T_Ms were, however, still sufficiently higher than that of GH46 (31 °C, **Figure 38**). Combined, the analytical SEC yield determination and thermostability

assays suggest vastly improved solubility, yield, and stability of mutants M7-9 in comparison to parent GH46.

Following this, I analysed the binding of the mutants on Globo-H expressing MCF-7 cells and compared it to that of HEK293 cells (**Figure 44**).

A



B

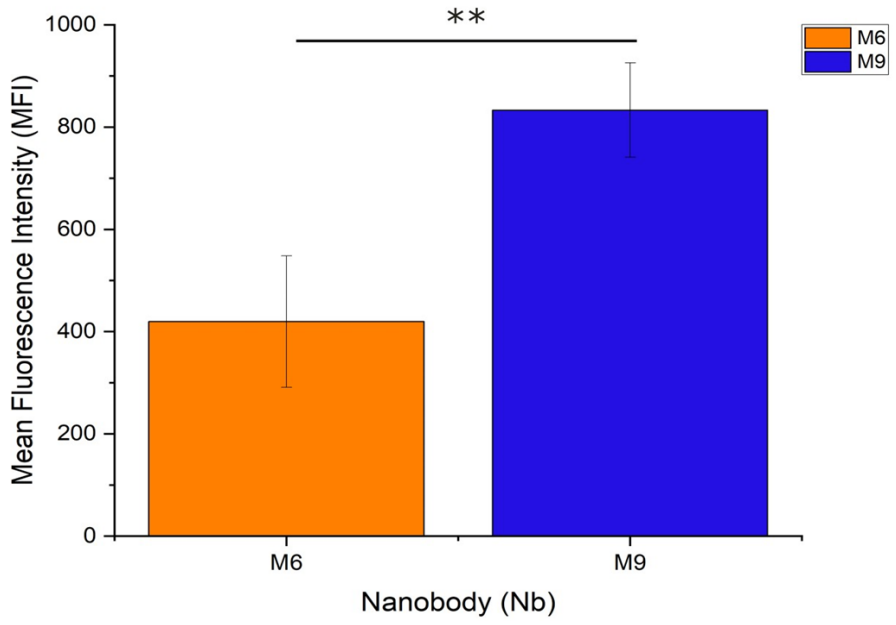


Figure 44: Cell binding comparison of Nbs on MCF-7 and HEK293 cells. A. *MFI of MCF-7 (green) or HEK293 (black) cells with M7, M8, and M9. Values represent mean ± SEM (n = 3). Statistical analysis was performed with unpaired t-test. *: $p < 0.05$; **: $p < 0.01$.* ***B.*** *Comparison of MFI of M9 to M6 on MCF-*

*7 cells. Values represent mean ± SEM (n = 3). Statistical analysis was performed with unpaired t-test. **: p < 0.01. Assays were carried out at physiological temperature (37°C). M6: orange; M9: blue.*

While all three mutants bound to MCF-7 cells, M8 ($p < 0.05$) and M9 ($p < 0.01$) bound to MCF-7 significantly higher than to HEK293 cells. As both M8 and M9 contain substitutions of Asp59Ser and Tyr61Asn, which M7 lacks, the mutations could be driving the significantly higher binding. Also, the combination of all three mutations, Asp59Ser, Tyr61Asn, and D110 deletion, in M9 gives more than three times higher binding for MCF-7 cells than M8. This shows that even though the deletion of D110 alone (M7) does not improve binding to cells on its own, it still plays an important role in Globo-H and M9 interactions, allowing M9 to be a much stronger binder.

M6 shares the same CDR sequences as the parent GH46, and only contains mutations in the FRs (**Figure 36** and **Table 3**). Similar to GH46, M6 also binds MCF-7 cells (**Figure 39**). Therefore, it can be taken as a more stable version of GH46. Hence, to maintain physiological temperature for all assays, M6 was used to represent GH46 in comparing the binding of M9 on MCF-7 cells. M9 binds ~2-fold better than M6 ($p \leq 0.01$) to native Globo-H displayed on MCF-7 cells showing better affinity for Globo-H (**Figure 44B**). Therefore, taken together, the CDR2 (Asp59Ser and Tyr61Asn) and 3 (D110 deletion) mutations improve the binding affinity for Globo-H, while maintaining the high stability and high yield characteristics of M6. **Table 4** compiles the results of yield, T_M, and cell binding analyses for M7-9.

Table 4: Results of analytical SEC, DSF, and cell binding assays for M7-9.

Nanobody	**CDR Mutations** (all 3 mutants contain FR mutations Gln5Val (1), His13Gln (2), and Val23Ala (3), Gln62Asp (4), and Met64Val (5) from M6 (**Figure 36C** and **Table 3**))	**Absorbance at 280 nm**	**T_M**	**Significant native Globo-H binding**
M7	Asp110 deletion (**Figure 41**)	~190 mAu	45°C	No
M8	Asp59Ser and Tyr61Asn	~260 mAu	50°C	Yes
M9	Asp59Ser, Tyr61Asn, and D110 deletion	~180 mAu	43°C	Yes

For my next and final step, I wanted to test whether M9 retains the specificity of the parent GH46 for Globo-H binding and does not bind to another target on MCF-7 cells. Additionally, I wanted to determine the binding affinity of M9 to Globo-H. For this I employed MST again (**Figure 45**).

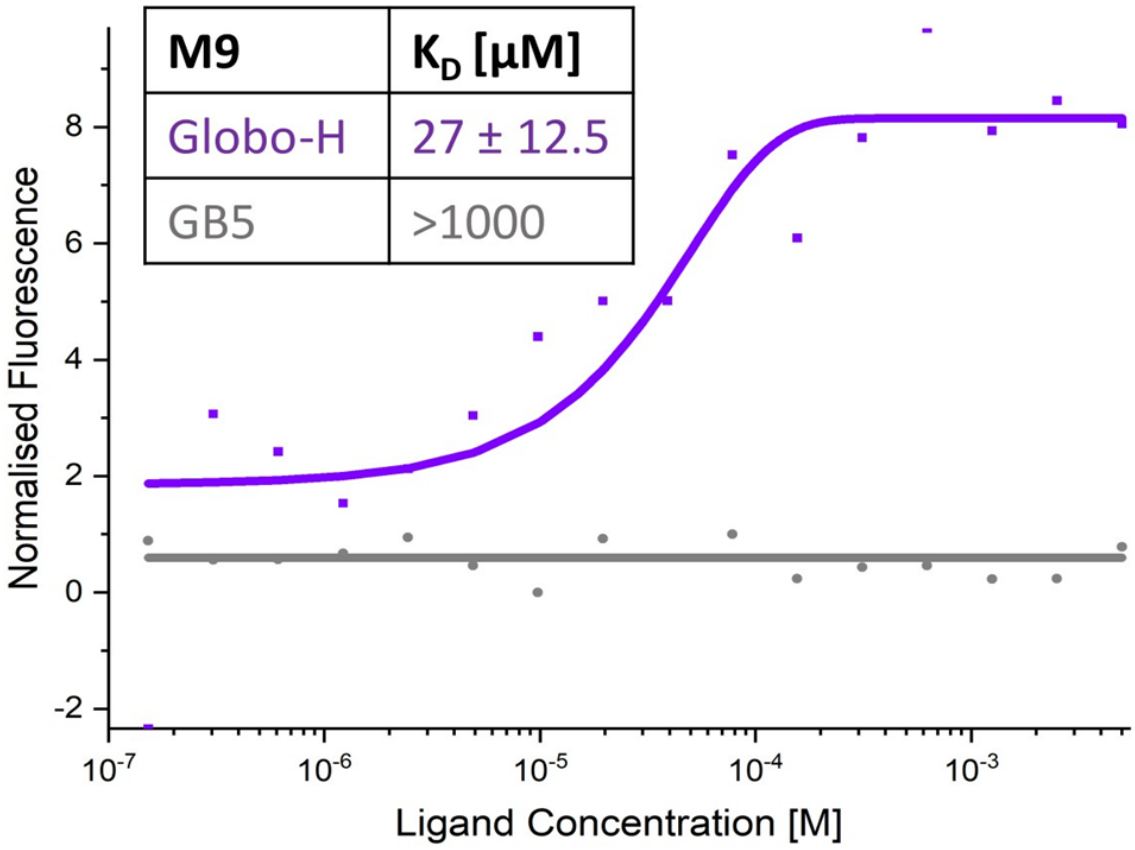


Figure 45: Measured in-solution K_D of M9 with synthetic Globo-H and GB5. *Table shows the determined K_D values. The comparative affinity curves for GB5 (gray) and Globo-H (violet) with MST are shown. Values represent mean ± SEM (n ≥ 3). The plots were generated in OriginPro 2021b using the ‚dose response' parameters.*

Cross-reactivity of TACA-targeting Abs to other structurally similar glycans is a general issue faced in the field[309], therefore, it was also important to determine whether M9 binds to other Globo-family members. Interestingly, even though I introduced mutations for better binding to lower globo series members also (CDR mutations were aimed for better binding with GalNAc and Gal3 in the Globo-H structure (**Figure 6** and **Figure 41**)), M9 seems to retain the inherited GH46 specificity for Globo-H as it only binds to Globo-H and not GB5 in MST. Furthermore, the determined K_D of M9 lies at 27 ± 12.5 µM, which is more than 7x better than K_D of GH46 (202 ± 89 µM (**Figure 32**)

[291]). The higher binding affinity shown also correlates with the results from the cell assay (**Figure 44**).

Hence, I was able to develop M9, which is a mutant of the parent GH46 containing 8 mutations, of which 5 are in the FRs and 3 in the CDRs of the protein. This study showed that the FR3 mutations (Gln62Asp and Met64Val) greatly improved the yield and thermostability of the Nb, while the CDR 2 and 3 mutations (Asp59Ser and Tyr61Asn, and D110 deletion, respectively) significantly improved the binding affinity of the Nb towards Globo-H.

4. Conclusion and Outlook

Despite numerous studies in the field of cancer research, cancer affects millions of people every year and portrays the second-highest death cause worldwide. Due to their high expression level and specificity for cancer cells, TACAs are important therapy targets in the fight against cancer. However, immunity against TACAs is difficult to develop due to their poor immunogenic nature. Also, development of Abs against TACAs faces many issues including low immunogenicity and low and/or impure amounts of native glycans for immunisation. Using synthetic glycan conjugated to an immunogenic protein carrier could be a solution for this. With this thesis, I show proof that Abs and Nbs produced in this way are highly specific for their target and bind not only to synthetic glycans but to their native counterparts as well.

This thesis is based on two examples of TACAs of great importance, sLeA and Globo-H, that were studied in pancreatic and breast cancer models, respectively. From the immunisation with synthetic glycans, Abs were produced. In the thesis, two highly specific Abs for sLeA and a novel Nb binder for Globo-H were characterised.

Sialyl Le A is the only FDA-approved biomarker for pancreatic cancer. Also, there are only two Abs, 5B1 and 1116-NS-19-9, mostly used in the clinic to target the glycan for therapy and diagnosis, respectively. Pancreatic cancer is often diagnosed at a later stage leading to a poor prognosis of the disease. Therefore, better diagnostic tools for earlier detection are needed. Also, other CA 19-9 targeting Abs have previously been seen to have broader specificities and, therefore, could lead to false results[81]. Consequently, there is a high demand for the clinically used Abs and a dire need for more Abs in the market that could provide high specificity and affinity for CA 19-9 and be used in clinic as well as research. I analysed two Abs, GB11 and HA8, generated against synthetic sLe A, and compared them to 1116-NS-19-9. I showed that all three Abs have high amino acid Fab sequence similarity, which could be due to germline Abs produced against sLe A by the murine immune system already highly selective in nature and leading to affinity-matured Abs with highly similar sequences. All Abs are also highly specific for sLe A with no cross-reaction to other glycans structurally close to sLe A.

Both GB11 and HA8 bind stronger to synthetic glycan than 1116-NS-19-9. This could be because GB11 and HA8 were generated against synthetic sLe A, while 1116-NS-19-9 was generated against native sLe A on cell. A more hydrophobic CDRH2 is seen

in GB11 (Tyr54 and Ile56 instead of Lys and Asn, respectively) and HA8 (Trp54 and Val 55 instead of Lys and Gly, respectively) compared to 1116-NS-19-9. Additionally, the more hydrophobic Val100 in both GB11 and HA8 near CDRH3 instead of Thr in 1116-NS-19-9 is also present. Combined, the three more hydrophobic residues in both GB11 and HA8 could have been energetically favoured compared to 1116-NS-19-9 in terms of packing interactions in the hydrophobic environment as they either lie in the binding domain (GB11-Trp54 and Ile56 and HA8-Trp54 and Val55 (CDRH2)) or close to it (Val100-CDRH3 for both Abs). Val100Thr is one of the seven core mutations carried out by Borenstein-Katz *et al.* to improve the affinity of 1116-NS-19-9 from 14.7 to 1.7 µM to sLe A[203]. In comparison, GB11 and HA8 boast similar affinity values of 3.2 ± 0.87 µM and 2.9 ± 0.47 µM in SPR and 1.9 ± 0.4 µM and 2.2 ± 0.3 µM in ITC, respectively.

GB11 is even a better binder for native CA 19-9 than 1116-NS-19-9 as well as HA8. Because here HA8 bound less strongly than 1116-NS-19-9, more differences between the binding of the three Abs to sLe A could be present that could not be elucidated yet. So far, limited information could be gained from the crystal structure of GB11 with sLe A due to low resolution, even though some hydrogen bonds similar to 1116-NS-19-9-sLe A interaction were detected. A higher resolution crystal structure is required for a better understanding. Moreover, other methods should also be utilised in this respect. Currently, STD-NMR analyses are being carried out by Anika Freitag for all three Abs against sLe A. This could provide an even better picture for comparison of their interactions in the future. The few differences in the core sequences of the three Abs could have also caused differences in their binding to sLe A. It would, therefore, be interesting in the future to study these differences further and chart out which core amino acids possibly enhance the binding of GB11 to sLe A. Additionally, it would also be interesting to introduce core mutations in HA8 with a method like the automated 'AbLift'[310] to potentially improve its binding to native CA 19-9.

GB11 and HA8 could also detect sLe A highly specifically in human pancreatic and gastric adenocarcinoma tissues showing their clinical relevance in diagnosis of sLe A expressing cancers. Taken together, I have shown with this work that GB11 is a significantly stronger binder of sLe A than the commercially used 1116-NS-19-9, is highly specific for the glycan and could, therefore, be used for diagnosis of pancreatic and gastric cancers. Hence, it could pave way to a better clinical diagnosis in the future. In the next steps, animal studies should also be carried out to test the Abs in

mammalian systems for diagnosis of pancreatic cancer to further assess their potential in clinic.

Nbs have various advantages over Abs including that they are one-tenth the size of a conventional Ab which makes them suitable for targeting solid tumours[311]. With a longer CDR3, Nbs access dense tumour microenvironments more easily and homogeneously distribute in the cancerous tissue[273]. Additionally, the renal clearance for Nbs is rapid, which is advantageous for diagnostics and imaging[273,274]. Furthermore, the smaller size and easy bacterial expression give Nb the advantage over Ab that they can be easily engineered into multifunctional domains[312].

However, a Nb targeting TACAs had not been produced before. From a previously established library of Nbs against Globo-H, I determined the best binder for the glycan[291]. Even though GH46 binds with high micromolar affinity, it shows high specificity for the glycan suggesting that immunisation of alpacas with pure synthetic glycans lead to specifically binding Nbs. Moreover, I was able to functionalise the Nb by constructing a trimer. Multimerisation can be advantageous in boosting avidity of glycan-binding Nbs as protein-glycan binding affinity tend to generally be low[305] and Nbs are single domains with three CDRs only in comparison to six CDRs of Abs[312]. Indeed, GH46-trimer showed improved apparent K_D (9-fold) on breast cancer cells emphasising the benefit of higher avidity of protein[291]. As the trimer is only about one-third the size of an IgG, it still has the advantage of smaller size and easy bacterial expression. The trimer could be further functionalised in the future for applications such as drug delivery, Ab recruitment to cells, and optical imaging of breast cancer, among others[312,313].

GH46 and GH46-trimer, however, are not very stable protein constructs. In fact, GH46-trimer is found in cells only in inclusion bodies and needs to be purified and solubilised accordingly. Therefore, I aimed to improve the stability of the GH46 by using point mutations after comparing GH46 amino acid sequence with 20 stable Nbs from the PDB. I generated mutant M6 with 3 FR1 and 2 FR3 mutations. The Nb boasts more than 10-fold higher protein yield showing much greater solubility than the parent Nb, GH46. The mutant also has ~20°C higher T_M, reiterating the stability of the protein, while also retaining the GH46 binding specificity for Globo-H. With this, I could appreciate that FR3 plays a substantial role in the conformation and, therefore, stability of Nbs and that there are two amino acids, in particular (Asp62 and Val64) that can stabilise Nbs dramatically. However, because only one Nb was studied in this thesis,

future studies should be carried out on other less stable Nbs to confirm its relevance to Nbs in general. I was able to show, however, that identifying and carrying out mutations in comparison to stable Nbs found in the PDB is a viable approach to improving the physical properties of poorly stable Nbs.

I also compared the GH46 to anti-SSEA-4 Ab-V_H and added three site-directed mutations in the CDRs (two in CDR2 and one in CDR3) to improve the binding affinity of the Nb to Globo-H by 7-fold. M9 is, therefore, a good binder of Globo-H and retains the high specificity of the parent GH46 as well as the high stability of M6. However, the M9-glycan binding specificity analysis was only carried out with MST so far and additional studies are required, such as with STD-NMR or glycan array, to support the results. Additionally, methods such as autonomous hypermutation yeast surface display could also be employed to further improve on binding, as it has been shown to muster functional affinity improvement of the mutant roughly 20-fold over the parent Nb[314]. Due to its high stability and better affinity, M9 could also be functionalised in various ways (as mentioned above), including being engineered into CAR T cells for immunotherapy of breast cancer cells[315]. Therefore, I envisage numerous potential applications for M9 in immunotherapy of breast cancer in the future.

5. Supplementary Information

5.1. Methods by collaborators

5.1.1. Crystallography

Diffraction data were collected and analysed by Michael Krummhaar (AG Roth, MPIKG), who also generated the figures. Diffraction data were collected at Berlin BESSY II, beamline 14-2 at 100 K and processed with Xia2/DIALS[316–319]. To solve the apo structure of GB11 molecular replacement was carried out with coordinates of 1116-NS-19-9 Fab fragment (PDB file ID: 6XTG) using MR-Phaser[320]. To solve the GB11 - Sialyl Lewis A complex holo structure, the refined structure of free GB11 was used as search model. Refining of structures was carried out using Refmac5[321] followed by iterative model building cycles using Coot[322]. Sugars restraints were created using Privateer[323]. Once refinement statistics converged, PDBredo was used with paired refinement[324] to determine whether data of higher resolution shells should be used. The resulting model was once more refined using Coot and Refmac5 as described previously. Figures were created using CCP4mg.

5.1.2. Immunohistochemistry (IHC) staining

Immunohistochemistry staining was carried out and analysed by Janine Arndt as part of a collaboration with Dr. med. Carsten Kamphues, Chirurgisches Forschungslabor, Klinik für Allgemein- und Viszeralchirurgie, Charité - Universitätsmedizin Berlin, Campus Benjamin Franklin, Germany. Surgically resected tissue was fixed in neutral buffered 4% formaldehyde for 16-24 h and processed for paraffin embedding. After deparaffinisation of 4 µm tissue sections, antigen unmasking was performed using 10 mM sodium citrate buffer (pH 6.0) with 0.05% Tween for 60 min at 90°C. Paraffin sections were permeabilised with 0.5% Triton X-100 for 10 min and blocked with 5% normal goat serum + 1% BSA for 60-120 min. Incubation with primary Abs was performed overnight at 4°C in blocking solution using 1116-NS-19-9 in a dilution of 1:100 as suggested by the manufacturer (Thermo Fisher #MA5-12421) or GB11 or HA8 at 5 µg/mL. A rabbit mAb against Vimentin in a dilution of 1:250 (clone EPR3776,

Abcam #ab9254) was also used simultaneously. Once washed with PBS, sections were incubated for 60-90 min at 37°C in 1% BSA with 4′,6-diamidino-2-phenylindole (DAPI) at a dilution of 1:1,000 and the secondary Abs goat anti-mouse Alexa Fluor™ 594 (1:250**,** Thermo Fisher #A-11005) as well as goat anti-rabbit Alexa Fluor™ 488 (1:500, Abcam #ab150081). The paraffin sections were embedded in ProTaqs® MountFluor and analysed using the confocal Laser Scanning Microscope 510 META (Zeiss).

5.1.3. Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) experiments were performed and analysed by Onur Turak and Dr. Alvaro Mallagaray (Institute of Chemistry and Metabolomics, Center of Structural and Cell Biology in Medicine, University of Lübeck, 23562 Lübeck, Germany) as part of a collaboration. The data was acquired on a Bruker Avance III HD 600 MHz NMR spectrometer equipped with a TCI cryogenic probe. For the processing of NMR spectra TopSpin 4.0.6 was used, while, for the quantification of TRNOE build-up rates, peak volumes were extracted using CCPNMR Analysis 2.4.2 software suite. Chemical shifts were measured relative to 3-(trimethylsilyl)-1-propanesulfonic acid-d_6 (DSS-d_6).

STD-NMR[325] spectra were acquired at 277 K using Gaussian-shaped radio frequency pulses for 4 s. A 45 dB attenuation was taken, which resulted in a 380.0 flip angle. To attenuate protein signals, a 30 ms spinlock filter was applied before acquisition. The acquisition time was 1.95 s along with an additional relaxation delay of 5 s. On- and off-resonances were set at 7 ppm and 200 ppm, respectively, and a total of 6000 scans were carried out. For the ligand epitope determination, STD amplification factors (AF) were calculated using Equation 1:

$$STD - AF = \frac{I_0 - I_{sat}}{I_0}, \qquad \text{Eq. 2}$$

where I_0 and I_{sat} are the signal intensities in the off- and on-resonance spectra, respectively. The largest STD effect (anomeric proton H1 of the fucose unit) was taken as a reference for 100% and all STD-AFs were normalized accordingly. NMR samples containing either 4 mM Globo-H or 4 mM GB5 in the presence or absence of 60 µM or 110 µM Nb, respectively, were prepared in STD-NMR buffer (40 mM sodium

phosphate buffer pH 8.0, 150 mM NaCl, 200 µM DSS-d_6 and 0.02 % NaN_3 in D_2O (>99.9 %)).

For the determination of transferred NOE build-up rates, experiments were recorded with a total of 8 scans of 1024 points and 128 t1 increments and a relaxation delay of 2.3 s. The frequency offset and spectral width were set at 4.70 ppm and 10.0 ppm for both dimensions. Residual water signal was suppressed via excitation sculpting. Zero filling to 256 points in F1 were used for data processing to give a final 1024 · 256 matrix. The optimal conditions for the transferred NOESY measurements were determined: a) At 298 K free Globo-H shows minimised NOESY cross-peaks relative to the complex and GH46 **remains** stable during NMR spectra acquisition. Therefore, experiments were carried out at 298K. b) The GH46:Globo-H molar ratio ranged from 1:4.5 to 1:16.4[326]. c) Mixing times (t_m) for the experiments were taken as 20, 40, 70, 100, 200, 300, 400, 500 and 700 ms. The bound state showed faster rate of build up for the pairs (Fuc-CH_3/Fuc-H5 and Fuc-H1/Gal5-H2) in comparison to the free state[327]. NMR samples either comprised of 3 mM Globo-H and 180 µM GH46 (16.4:1 ratio) or 20 mM Globo-H in the STD-NMR buffer.

6. References

1. Wild, C. P., Weiderpass, E. & Stewart, B. W. *World Cancer report 2020: Cancer Research for Cancer Precention*. *Cancer Control* vol. 199 (2020).
2. World Cancer Day: closing the care gap. https://www.who.int/news/item/03-02-2022-world-cancer-day-closing-the-care-gap.
3. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394–424 (2018).
4. Maino Vieytes, C. A., Taha, H. M., Burton-Obanla, A. A., Douglas, K. G. & Arthur, A. E. Carbohydrate Nutrition and the Risk of Cancer. *Current Nutrition Reports* vol. 8 230–239 Preprint at https://doi.org/10.1007/s13668-019-0264-3 (2019).
5. Tian, T., Olson, S., Whitacre, J. M. & Harding, A. The origins of cancer robustness and evolvability. *Integrative Biology* **3**, 17–30 (2011).
6. Chen, D. S. & Mellman, I. Oncology meets immunology: The cancer-immunity cycle. *Immunity* vol. 39 1–10 Preprint at https://doi.org/10.1016/j.immuni.2013.07.012 (2013).
7. Spiering, M. J. Primer on the immune system. *Alcohol Research: Current Reviews* vol. 37 171 Preprint at (2015).
8. Schlake, T., Thess, A., Thran, M. & Jordan, I. mRNA as novel technology for passive immunotherapy. *Cellular and Molecular Life Sciences* **76**, 301–328 (2019).
9. Saxena, M., van der Burg, S. H., Melief, C. J. M. & Bhardwaj, N. Therapeutic cancer vaccines. *Nature Reviews Cancer* vol. 21 360–378 Preprint at https://doi.org/10.1038/s41568-021-00346-0 (2021).
10. Büll, C., den Brok, M. H. & Adema, G. J. Sweet escape: Sialic acids in tumor immune evasion. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1846**, 238–246 (2014).
11. Stanczak, M. A. *et al.* Self-associated molecular patterns mediate cancer immune evasion by engaging Siglecs on T cells. *J Clin Invest* **128**, 4912–4923 (2018).
12. van Houtum, E. J. H., Büll, C., Cornelissen, L. A. M. & Adema, G. J. Siglec Signaling in the Tumor Microenvironment. *Front Immunol* **12**, (2021).
13. Varki, A. PAMPs, DAMPs and SAMPs: Host Glycans are Self-Associated Molecular Patterns, but subject to Microbial Molecular Mimicry. *The FASEB Journal* **35**, (2021).
14. Vansteenkiste, J. F. *et al.* Efficacy of the MAGE-A3 cancer immunotherapeutic as adjuvant therapy in patients with resected MAGE-A3-positive non-small-cell lung cancer (MAGRIT): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol* **17**, 822–835 (2016).
15. Giaccone, G. *et al.* A phase III study of belagenpumatucel-L, an allogeneic tumour cell vaccine, as maintenance therapy for non-small cell lung cancer. *Eur J Cancer* **51**, 2321–2329 (2015).
16. Butts, C. *et al.* Tecemotide (L-BLP25) versus placebo after chemoradiotherapy for stage III non-small-cell lung cancer (START): A randomised, double-blind, phase 3 trial. *Lancet Oncol* **15**, 59–68 (2014).
17. Rini, B. I. *et al.* IMA901, a multipeptide cancer vaccine, plus sunitinib versus sunitinib alone, as first-line therapy for advanced or metastatic renal cell carcinoma (IMPRINT):

a multicentre, open-label, randomised, controlled, phase 3 trial. *Lancet Oncol* **17**, 1599–1611 (2016).
18. Middleton, G. *et al.* Gemcitabine and capecitabine with or without telomerase peptide vaccine GV1001 in patients with locally advanced or metastatic pancreatic cancer (TeloVac): An open-label, randomised, phase 3 trial. *Lancet Oncol* **15**, 829–840 (2014).
19. Lawson, D. H. *et al.* Randomized, placebo-controlled, phase III trial of yeast- derIVed granulocyte-macrophage colony-stimulating factor (GM-CSF) versus peptide vaccination versus GM-CSF plus peptide vaccination versus placebo in patients with no evidence of disease after complete surgical resection of locally advanced and/or stage IV melanoma: A trial of the eastern cooperatIVe oncology groupâ€"American college of radiology imaging network cancer research group (E4697). *Journal of Clinical Oncology* **33**, 4066–4076 (2015).
20. Cheever, M. A. & Higano, C. S. PROVENGE (sipuleucel-T) in prostate cancer: The first FDA-approved therapeutic cancer vaccine. *Clinical Cancer Research* vol. 17 3520–3526 Preprint at https://doi.org/10.1158/1078-0432.CCR-10-3126 (2011).
21. Lamm, D. L. & Morales, A. A BCG success story: From prevention of tuberculosis to optimal bladder cancer treatment. *Vaccine* **39**, 7308–7318 (2021).
22. Ferrucci, P. F., Pala, L., Conforti, F. & Cocorocchio, E. Talimogene laherparepvec (T-vec): An intralesional cancer immunotherapy for advanced melanoma. *Cancers* vol. 13 1–14 Preprint at https://doi.org/10.3390/cancers13061383 (2021).
23. McKee, A. S., Munks, M. W. & Marrack, P. How Do Adjuvants Work? Important Considerations for New Generation Adjuvants. *Immunity* vol. 27 687–690 Preprint at https://doi.org/10.1016/j.immuni.2007.11.003 (2007).
24. Pulendran, B., S. Arunachalam, P. & O'Hagan, D. T. Emerging concepts in the science of vaccine adjuvants. *Nature Reviews Drug Discovery* vol. 20 454–475 Preprint at https://doi.org/10.1038/s41573-021-00163-y (2021).
25. Costa, A. F., Campos, D., Reis, C. A. & Gomes, C. Targeting Glycosylation: A New Road for Cancer Drug Discovery. *Trends in Cancer* vol. 6 757–766 Preprint at https://doi.org/10.1016/j.trecan.2020.04.002 (2020).
26. Cox, T. R. The matrix in cancer. *Nature Reviews Cancer* vol. 21 217–238 Preprint at https://doi.org/10.1038/s41568-020-00329-7 (2021).
27. Mancini, M. L. & Sonis, S. T. Mechanisms of cellular fibrosis associated with cancer regimen-related toxicities. *Frontiers in Pharmacology* vol. 5 MAR Preprint at https://doi.org/10.3389/fphar.2014.00051 (2014).
28. Lippi, G. & Mattiuzzi, C. The global burden of pancreatic cancer. *Archives of Medical Science* **16**, 820–824 (2020).
29. Haeberle, L. & Esposito, I. Pathology of pancreatic cancer. *Translational Gastroenterology and Hepatology* vol. 4 Preprint at https://doi.org/10.21037/tgh.2019.06.02 (2019).
30. Ryan, D. P., Hong, T. S. & Bardeesy, N. Pancreatic Adenocarcinoma. *New England Journal of Medicine* **371**, 1039–1049 (2014).
31. Tonini, V. & Zanni, M. Pancreatic cancer in 2021: What you need to know to win. *World Journal of Gastroenterology* vol. 27 5851–5889 Preprint at https://doi.org/10.3748/wjg.v27.i35.5851 (2021).
32. Kleeff, J. *et al.* Pancreatic cancer. *Nat Rev Dis Primers* **2**, (2016).

33. Krebs - Pancreatic cancer. https://www.krebsdaten.de/Krebs/EN/Content/Cancer_sites/Pancreatic_cancer/pancreatic_cancer_node.html.
34. Ilic, M. & Ilic, I. Epidemiology of pancreatic cancer. *World Journal of Gastroenterology* vol. 22 9694–9705 Preprint at https://doi.org/10.3748/wjg.v22.i44.9694 (2016).
35. Yu, J., Yang, X., He, W. & Ye, W. Burden of pancreatic cancer along with attributable risk factors in Europe between 1990 and 2019, and projections until 2039. *Int J Cancer* **149**, 993–1001 (2021).
36. Yachida, S. & Lacobuzio-Donahue, C. A. The pathology and genetics of metastatic pancreatic cancer. *Archives of Pathology and Laboratory Medicine* vol. 133 413–422 Preprint at https://doi.org/10.5858/133.3.413 (2009).
37. Lee, E. S. & Lee, J. M. Imaging diagnosis of pancreatic cancer: A state-of-the-art review. *World J Gastroenterol* **20**, 7864–7877 (2014).
38. Garces-Descovich, A., Beker, K., Jaramillo-Cardoso, A., James Moser, A. & Mortele, K. J. Applicability of current NCCN Guidelines for pancreatic adenocarcinoma resectability: analysis and pitfalls. *Abdominal Radiology* **43**, 314–322 (2018).
39. Motosugi, U. *et al.* Detection of pancreatic carcinoma and liver metastases with gadoxetic acid-enhanced MR imaging: Comparison with contrast-enhanced multi-detector row CT. *Radiology* **260**, 446–453 (2011).
40. Schmocker, R. K. *et al.* Utilization of preoperative endoscopic ultrasound for pancreatic adenocarcinoma. *HPB* **19**, 465–472 (2017).
41. Niederau, C. & Grendell, J. H. Diagnosis of pancreatic carcinoma. Imaging techniques and tumor markers. *Pancreas* **7**, 66–86 (1992).
42. Vozzo, C. F. & Sanaka, M. R. Endoscopic Management of Pancreaticobiliary Disease. *Surgical Clinics of North America* vol. 100 1151–1168 Preprint at https://doi.org/10.1016/j.suc.2020.08.006 (2020).
43. Alauddin, M. & Palatis, L. Current and Future Trends in Early Detection of Pancreatic Cancer: Molecular Targets and PET Probes. *Curr Med Chem* **22**, 3370–3389 (2015).
44. Pisters, P. W. T., Lee, J. E., Vauthey, J. N., Charnsangavej, C. & Evans, D. B. Laparoscopy in the staging of pancreatic cancer. *British Journal of Surgery* vol. 88 325–337 Preprint at https://doi.org/10.1046/j.1365-2168.2001.01695.x (2001).
45. Fong, Z. V. *et al.* Reappraisal of Staging Laparoscopy for Patients with Pancreatic Adenocarcinoma: A Contemporary Analysis of 1001 Patients. *Ann Surg Oncol* **24**, 3203–3211 (2017).
46. Zhang, L., Sanagapalli, S. & Stoita, A. Challenges in diagnosis of pancreatic cancer. *World Journal of Gastroenterology* vol. 24 2047–2060 Preprint at https://doi.org/10.3748/wjg.v24.i19.2047 (2018).
47. Clarke, D. L., Clarke, B. A., Thomson, S. R., Garden, O. J. & Lazarus, N. G. The role of preoperative biopsy in pancreatic cancer. *HPB* vol. 6 144–153 Preprint at https://doi.org/10.1080/13651820410030862 (2004).
48. Matthaei, H., Schulick, R. D., Hruban, R. H. & Maitra, A. Cystic precursors to invasive pancreatic cancer. *Nature Reviews Gastroenterology and Hepatology* vol. 8 141–150 Preprint at https://doi.org/10.1038/nrgastro.2011.2 (2011).
49. Radon, T. P. *et al.* Identification of a three-biomarker panel in urine for early detection of pancreatic adenocarcinoma. *Clinical Cancer Research* **21**, 3512–3521 (2015).

50. Brezgyte, G., Shah, V., Jach, D. & Crnogorac-jurcevic, T. Non-invasive biomarkers for earlier detection of pancreatic cancer—a comprehensive review. *Cancers* vol. 13 Preprint at https://doi.org/10.3390/cancers13112722 (2021).
51. Zhang, L. *et al.* Salivary Transcriptomic Biomarkers for Detection of Resectable Pancreatic Cancer. *Gastroenterology* **138**, 949 (2010).
52. Xie, Z. *et al.* Salivary microRNAs show potential as a noninvasive biomarker for detecting resectable pancreatic cancer. *Cancer Prevention Research* **8**, 165–173 (2015).
53. Satoh, K. Molecular approaches using body fluid for the early detection of pancreatic cancer. *Diagnostics* vol. 11 Preprint at https://doi.org/10.3390/diagnostics11020375 (2021).
54. Ishige, F. *et al.* MIR1246 in body fluids as a biomarker for pancreatic cancer. *Sci Rep* **10**, (2020).
55. Kanda, M. *et al.* Mutant TP53 in duodenal samples of pancreatic juice from patients with pancreatic cancer or high-grade dysplasia. *Clinical Gastroenterology and Hepatology* **11**, 719 (2013).
56. Singhi, A. D. *et al.* Preoperative next-generation sequencing of pancreatic cyst fluid is highly accurate in cyst classification and detection of advanced neoplasia. *Gut* **67**, 2131 (2017).
57. Kanda, M. *et al.* Mutant GNAS detected in duodenal collections of secretin-stimulated pancreatic juice indicates the presence or emergence of pancreatic cysts. *Gut* **62**, 1024–1033 (2013).
58. Haab, B. B. *et al.* Glycosylation variants of mucins and CEACAMs as candidate biomarkers for the diagnosis of pancreatic cystic neoplasms. *Ann Surg* **251**, 937–945 (2010).
59. Maker, A. v. *et al.* Cyst fluid interleukin-1β (IL1β) levels predict the risk of carcinoma in intraductal papillary mucinous neoplasms of the pancreas. *Clinical Cancer Research* **17**, 1502–1508 (2011).
60. Permuth, J. B. *et al.* Linc-ing Circulating Long Non-coding RNAs to the Diagnosis and Malignant Prediction of Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Sci Rep* **7**, (2017).
61. Ding, J. *et al.* Identification of key lncRNAs in the tumorigenesis of intraductal pancreatic mucinous neoplasm by coexpression network analysis. *Cancer Med* **9**, 3840–3851 (2020).
62. Duell, E. J. *et al.* Plasma microRNAs as biomarkers of pancreatic cancer risk in a prospective cohort study. *Int J Cancer* **141**, 905–915 (2017).
63. Bloomston, M. *et al.* MicroRNA expression patterns to differentiate pancreatic adenocarcinoma from normal pancreas and chronic pancreatitis. *J Am Med Assoc* **297**, 1901–1908 (2007).
64. Cirmena, G. *et al.* Assessment of circulating nucleic acids in cancer: From current status to future perspectives and potential clinical applications. *Cancers* vol. 13 Preprint at https://doi.org/10.3390/cancers13143460 (2021).
65. Shen, S. Y. *et al.* Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* **563**, 579–583 (2018).
66. Eissa, M. A. L. *et al.* Promoter methylation of ADAMTS1 and BNC1 as potential biomarkers for early detection of pancreatic cancer in blood. *Clin Epigenetics* **11**, 59 (2019).

67. Luang, S., Teeravirote, K., Saentaweesuk, W., Ma-In, P. & Silsirivanit, A. Carbohydrate Antigen 50: Values for Diagnosis and Prognostic Prediction of Intrahepatic Cholangiocarcinoma. *Medicina (B Aires)* **56**, 1–9 (2020).
68. Holdenrieder, S. & Stieber, P. Carbohydrate antigen 50. 521–521 (2019) doi:10.1007/978-3-662-48986-4_667.
69. Holdenrieder, S. & Stieber, P. Carbohydrate antigen 72-4. 522–522 (2019) doi:10.1007/978-3-662-48986-4_669.
70. Nlllan, C. M., Haglund', C., Lundin', J., Kuusela2, P. & Roberts', P. J. 492 CA 242, a new tumour marker for pancreatic cancer: a comparison with CA 19-9, CA 50 and CEA. **7**, 487 (1994).
71. Szajda, S. D., Waszkiewicz, N., Chojnowska, S. & Zwierz, K. Carbohydrate markers of pancreatic cancer. *Biochem Soc Trans* **39**, 340–343 (2011).
72. White, B., Patterson, M., Karnwal, S. & Brooks, C. L. Crystal Structure of a Human MUC16 SEA Domain Reveals Insight Into the Nature of the CA125 Tumour Marker. *Proteins* **90**, 1210 (2022).
73. Scarà, S., Bottoni, P. & Scatena, R. CA 19-9: Biochemical and clinical aspects. in *Advances in Experimental Medicine and Biology* vol. 867 247–260 (Springer New York LLC, 2015).
74. Satake, K. & Takeuchi, T. Comparison of CA19-9 with other Tumor markers in the diagnosis of cancer of the Pancreas. *Pancreas* **9**, 720–724 (1994).
75. Bussom, S. & Saif, M. W. Methods and rationale for the early detection of pancreatic cancer. Highlights from the '2010 ASCO Gastrointestinal Cancers Symposium'. Orlando, FL, USA. January 22-24, 2010. *JOP* **11**, 128–30 (2010).
76. Tang, H. *et al.* Glycans Related to the CA19-9 Antigen Are Increased in Distinct Subsets of Pancreatic Cancers and Improve Diagnostic Accuracy Over CA19-9. *Cell Mol Gastroenterol Hepatol* **2**, 210 (2016).
77. Staal, B. *et al.* The Stra plasma biomarker: Blinded validation of improved accuracy over CA19-9 in pancreatic cancer diagnosis. *Clinical Cancer Research* **25**, 2745–2754 (2019).
78. Barnett, D. *et al.* The CA19-9 and Sialyl-TRA Antigens Define Separate Subpopulations of Pancreatic Cancer Cells. *Scientific Reports 2017 7:1* **7**, 1–13 (2017).
79. Gao, C. F. *et al.* Detection of chemotherapy-resistant pancreatic cancer using a glycan biomarker, sTRA. *Clinical Cancer Research* **27**, 226–236 (2021).
80. Liu, J. *et al.* Combination of plasma microRNAs with serum CA19-9 for early detection of pancreatic cancer. *Int J Cancer* **131**, 683–691 (2012).
81. Partyka, K., Maupin, K. A., Brand, R. E. & Haab, B. B. Diverse monoclonal antibodies against the CA 19-9 antigen show variation in binding specificity with consequences for clinical interpretation. *Proteomics* **12**, 2212–2220 (2012).
82. Chang, W.-W. *et al.* Expression of Globo H and SSEA3 in breast cancer stem cells and the involvement of fucosyl transferases 1 and 2 in Globo H synthesis. *Proceedings of the National Academy of Sciences* **105**, 11667–11672 (2008).
83. Gottschling, S. *et al.* Stage-specific embryonic antigen-4 is expressed in basaloid lung cancer and associated with poor prognosis. *Eur Respir J* **41**, 656–663 (2013).
84. Hung, T.-C., Lin, C.-W., Hsu, T.-L., Wu, C.-Y. & Wong, C.-H. Investigation of SSEA-4 Binding Protein in Breast Cancer Cells. *J Am Chem Soc* **135**, 5934–5937 (2013).
85. Lukong, K. E. Understanding breast cancer – The long and winding road. *BBA Clin* **7**, 64–77 (2017).

86. Conde, C. & Arnoldo Nunes Miranda, F. de. Sentimentos vivenciados por mulheres ao receberem o diagnóstico de câncer de mama Related papers Feelings and Expect at ions of Women Diagnosed wit h Breast Cancer: A Reflect ion Sent iment …. doi:10.5205/1981-8963-v12i01a23520p102-111-2018.
87. Arnold, M. *et al.* Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast* **66**, 15–23 (2022).
88. Silvestri, V., Nicolussi, A., Pensabene, M., von Arx, C. & de Laurentiis, M. Male Breast Cancer: From Molecular Genetics to Clinical Management. *Cancers 2022, Vol. 14, Page 2006* **14**, 2006 (2022).
89. Beral, V. *et al.* Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. *The Lancet* **358**, 1389–1399 (2001).
90. Wang, S. M. A global perspective on the ethnic-specific BRCA variation and its implication in clinical application. *Journal of the National Cancer Center* (2022) doi:10.1016/J.JNCC.2022.12.001.
91. Wilkinson, L. & Gathani, T. Understanding breast cancer as a global health concern. *https://doi.org/10.1259/bjr.20211033* **95**, 20211033 (2021).
92. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209–249 (2021).
93. Breast cancer statistics | Cancer Research UK. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer.
94. Heer, E. *et al.* Global burden and trends in premenopausal and postmenopausal breast cancer: a population-based study. *Lancet Glob Health* **8**, e1027–e1037 (2020).
95. Tarighati, E., Keivan, H. & Mahani, H. A review of prognostic and predictive biomarkers in breast cancer. *Clinical and Experimental Medicine 2021* 1–16 (2022) doi:10.1007/S10238-021-00781-1.
96. Dass, S. A. *et al.* Triple Negative Breast Cancer: A Review of Present and Future Diagnostic Modalities. *Medicina 2021, Vol. 57, Page 62* **57**, 62 (2021).
97. Ellis, M. J. *et al.* Outcome Prediction for Estrogen Receptor–Positive Breast Cancer Based on Postneoadjuvant Endocrine Therapy Tumor Characteristics. *JNCI: Journal of the National Cancer Institute* **100**, 1380–1388 (2008).
98. Fragomeni, S. M., Sciallis, A. & Jeruss, J. S. Molecular Subtypes and Local-Regional Control of Breast Cancer. *Surgical Oncology Clinics* **27**, 95–120 (2018).
99. Nounou, M. I. *et al.* Breast Cancer: Conventional Diagnosis and Treatment Modalities and Recent Patents and Technologies Supplementary Issue: Targeted Therapies in Breast Cancer Treatment. doi:10.4137/BCBCr.s29420.
100. Sauter, E. R. Reliable Biomarkers to Identify New and Recurrent Cancer. *Eur J Breast Health* **13**, 162 (2017).
101. Mcdonald, E. S., Clark, A. S., Tchou, J., Zhang, P. & Freedman, G. M. Clinical Diagnosis and Management of Breast Cancer. *J Nucl Med* **57**, 9–16 (2016).
102. Desreux, J. A. C. Breast cancer screening in young women. *European Journal of Obstetrics & Gynecology and Reproductive Biology* **230**, 208–211 (2018).
103. Berg, W. A. *et al.* Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA* **299**, 2151–2163 (2008).

104. Riedl, C. C. *et al.* Triple-modality screening trial for familial breast cancer underlines the importance of magnetic resonance imaging and questions the role of mammography and ultrasound regardless of patient mutation status, age, and breast density. *J Clin Oncol* **33**, 1128–1135 (2015).
105. He, Z. *et al.* A review on methods for diagnosis of breast cancer cells and tissues. *Cell Prolif* **53**, (2020).
106. Santiago, L., Adrada, B. E., Huang, M. L., Wei, W. & Candelaria, R. P. Breast cancer neoplastic seeding in the setting of image-guided needle biopsies of the breast. *Breast Cancer Res Treat* **166**, 29–39 (2017).
107. Jahanbin, B., Soleimani, V. & Azmoudeh-Ardalan, F. Displaced Epithelium in Breast Pathology: A Review. *Archives of Breast Cancer* 150–158 (2018) doi:10.32768/ABC.201854150-158.
108. Yip, C. H. & Rhodes, A. Estrogen and progesterone receptors in breast cancer. *http://dx.doi.org/10.2217/fon.14.110* **10**, 2293–2301 (2014).
109. Canas-Marques, R. & Schnitt, S. J. E-cadherin immunohistochemistry in breast pathology: uses and pitfalls. *Histopathology* **68**, 57–69 (2016).
110. Maeda, I. *et al.* Effectiveness of computer-aided diagnosis (CADx) of breast pathology using immunohistochemistry results of core needle biopsy samples for synaptophysin, oestrogen receptor and CK14/p63 for classification of epithelial proliferative lesions of the breast. *J Clin Pathol* **70**, 1057–1062 (2017).
111. Hanley, K. Z., Birdsong, G. G., Cohen, C. & Siddiqui, M. T. Immunohistochemical detection of estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 expression in breast carcinomas: comparison on cell block, needle-core, and tissue block preparations. *Cancer* **117**, 279–288 (2009).
112. Jahr, S. *et al.* DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res* **61**, 1659–1665 (2001).
113. Gasch, C. *et al.* Heterogeneity of epidermal growth factor receptor status and mutations of KRAS/PIK3CA in circulating tumor cells of patients with colorectal cancer. *Clin Chem* **59**, 252–260 (2013).
114. Alimirzaie, S., Bagherzadeh, M. & Akbari, M. R. Liquid biopsy in breast cancer: A comprehensive review. *Clin Genet* **95**, 643–660 (2019).
115. Lee, J. S., Magbanua, M. J. M. & Park, J. W. Circulating tumor cells in breast cancer: applications in personalized medicine. *Breast Cancer Research and Treatment 2016 160:3* **160**, 411–424 (2016).
116. Andree, K. C., van Dalum, G. & Terstappen, L. W. M. M. Challenges in circulating tumor cell detection by the CellSearch system. *Mol Oncol* **10**, 395–407 (2016).
117. Lee, J. *et al.* Circulating Tumor DNA in a Breast Cancer Patient's Plasma Represents Driver Alterations in the Tumor Tissue. *Genomics Inform* **15**, 48–50 (2017).
118. Guttery, D. S. *et al.* Noninvasive Detection of Activating Estrogen Receptor 1 (ESR1) Mutations in Estrogen Receptor–Positive Metastatic Breast Cancer. *Clin Chem* **61**, 974–982 (2015).
119. Trejo-Becerril, C. *et al.* Cancer Progression Mediated by Horizontal Gene Transfer in an In Vivo Model. *PLoS One* **7**, e52754 (2012).
120. Chang, Y. *et al.* Review of the clinical applications and technological advances of circulating tumor DNA in cancer monitoring. *Ther Clin Risk Manag* **13**, 1363 (2017).

121. Eigeliene, N., Saarenheimo, J. & Jekunen, A. Potential of Liquid Biopsies for Breast Cancer Screening, Diagnosis, and Response to Treatment. *Oncology* **96**, 115–124 (2019).
122. Cohen, J. D. *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science (1979)* **359**, 926–930 (2018).
123. Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, (2014).
124. Li, J. *et al.* Non-Invasive Biomarkers for Early Detection of Breast Cancer. *Cancers 2020, Vol. 12, Page 2767* **12**, 2767 (2020).
125. Kosaka, N. *et al.* Exploiting the message from cancer: the diagnostic value of extracellular vesicles for clinical applications. *Experimental & Molecular Medicine 2019 51:3* **51**, 1–9 (2019).
126. Sadovska, L., Eglitis, J. & Line, A. Extracellular Vesicles as Biomarkers and Therapeutic Targets in Breast Cancer. *Anticancer Res* **35**, 6379–6390 (2015).
127. Aatonen, M. T. *et al.* Isolation and characterization of platelet-derived extracellular vesicles. *J Extracell Vesicles* **3**, (2014).
128. Royo, F., Théry, C., Falcón-Pérez, J. M., Nieuwland, R. & Witwer, K. W. Methods for Separation and Characterization of Extracellular Vesicles: Results of a Worldwide Survey Performed by the ISEV Rigor and Standardization Subcommittee. *Cells 2020, Vol. 9, Page 1955* **9**, 1955 (2020).
129. Punnonen, K., Hietanen, E., Auvinen, O. & Punnonen, R. Phospholipids and fatty acids in breast cancer tissue. *J Cancer Res Clin Oncol* **115**, 575–578 (1989).
130. April Haridatt Mistry, D. & French, P. W. Circulating Phospholipids as Biomarkers of Breast Cancer: A Review. doi:10.4137/BCBCr.s40693.
131. Kim, H., Min, H. K., Kong, G. & Moon, M. H. Quantitative analysis of phosphatidylcholines and phosphatidylethanolamines in urine of patients with breast cancer by nanoflow liquid chromatography/tandem mass spectrometry. *Anal Bioanal Chem* **393**, 1649–1656 (2009).
132. Erbes, T. *et al.* Feasibility of urinary microRNA detection in breast cancer patients and its potential as an innovative non-invasive biomarker. *BMC Cancer* **15**, 1–9 (2015).
133. Hirschfeld, M. *et al.* Urinary Exosomal MicroRNAs as Potential Non-invasive Biomarkers in Breast Cancer Detection. *Molecular Diagnosis & Therapy 2020 24:2* **24**, 215–232 (2020).
134. Ando, W. *et al.* Novel breast cancer screening: combined expression of miR-21 and MMP-1 in urinary exosomes detects 95% of breast cancer without metastasis. *Sci Rep* **9**, 13595–13595 (2019).
135. Dinges, S. S. *et al.* Cancer metabolomic markers in urine: evidence, techniques and recommendations. *Nature Reviews Urology 2019 16:6* **16**, 339–362 (2019).
136. Rauf, F., Anderson, K. S. & La-Baer, J. L. Autoantibodies in early detection of breast cancer. *Cancer Epidemiology Biomarkers and Prevention* **29**, 2475–2485 (2020).
137. Temme, J. S., Butler, D. L. & Gildersleeve, J. C. Anti-glycan antibodies: roles in human disease. *Biochemical Journal* **478**, 1485–1509 (2021).
138. Turnbull, A. R. *et al.* Autoantibodies in early breast cancer: a stage-related phenomenon? *Br J Cancer* **38**, 461 (1978).
139. Suzuki, H., Graziano, D. F., McKolanis, J. & Finn, O. J. T Cell–Dependent Antibody Responses against Aberrantly Expressed Cyclin B1 Protein in Patients with Cancer and Premalignant Disease. *Clinical Cancer Research* **11**, 1521–1526 (2005).

140. Lu, H. *et al.* Evaluation of known oncoantibodies, HER2, p53, and cyclin B1, in prediagnostic breast cancer sera. *Cancer Prevention Research* **5**, 1036–1043 (2012).
141. Anderson, K. S. & LaBaer, J. The sentinel within: Exploiting the immune system for cancer biomarkers. *J Proteome Res* **4**, 1123–1133 (2005).
142. Dudas, S. P., Chatterjee, M. & Tainsky, M. A. Usage of cancer associated autoantibodies in the detection of disease. *Cancer Biomarkers* **6**, 257–270 (2010).
143. Kathrikolly, T., Nair, S. N., Mathew, A., Saxena, P. P. U. & Nair, S. Can serum autoantibodies be a potential early detection biomarker for breast cancer in women? A diagnostic test accuracy review and meta-analysis. *Syst Rev* **11**, 1–12 (2022).
144. Wu, M., Mao, C., Chen, Q., Cu, X. W. & Zhang, W. S. Serum p53 protein and anti-p53 antibodies are associated with increased cancer risk: A case-control study of 569 patients and 879 healthy controls. *Mol Biol Rep* **37**, 339–343 (2010).
145. Lu, H. *et al.* Evaluation of known oncoantibodies, HER2, p53, and cyclin B1, in prediagnostic breast cancer sera. *Cancer Prevention Research* **5**, 1036–1043 (2012).
146. Blixt, O. *et al.* Autoantibodies to aberrantly glycosylated MUC1 in early stage breast cancer are associated with a better prognosis. *Breast Cancer Research* **13**, 1–16 (2011).
147. Hamrita, B. *et al.* Identification of tumor antigens that elicit a humoral immune response in breast cancer patients' sera by serological proteome analysis (SERPA). *Clinica Chimica Acta* **393**, 95–102 (2008).
148. Katayama, H. *et al.* An autoimmune response signature associated with the development of triple-negative breast cancer reflects disease pathogenesis. *Cancer Res* **75**, 3246–3254 (2015).
149. Wang, C.-C. *et al.* Glycan microarray of Globo H and related structures for quantitative analysis of breast cancer. *Proceedings of the National Academy of Sciences* **105**, 11661–11666 (2008).
150. Lacombe, J. *et al.* Identification and validation of new autoantibodies for the diagnosis of DCIS and node negative early-stage breast cancers. *Int J Cancer* **132**, 1105–1113 (2013).
151. Oloomi, M., Moazzezy, N. & Bouzari, S. Comparing blood versus tissue-based biomarkers expression in breast cancer patients. *Heliyon* **6**, e03728 (2020).
152. Hou, M. F. *et al.* Evaluation of serum CA27.29, CA15-3 and CEA in patients with breast cancer. *Kaohsiung J Med Sci* **15**, 520–528 (1999).
153. Peric, L. *et al.* Glycosylation Alterations in Cancer Cells, Prognostic Value of Glycan Biomarkers and Their Potential as Novel Therapeutic Targets in Breast Cancer. *Biomedicines 2022, Vol. 10, Page 3265* **10**, 3265 (2022).
154. Pinto, R. *et al.* Identification of new cancer biomarkers based on aberrant mucin glycoforms by in situ proximity ligation. *J Cell Mol Med* **16**, 1474–1484 (2012).
155. Joshi, H. J. *et al.* SnapShot: O-Glycosylation Pathways across Kingdoms. *Cell* vol. 172 632-632.e2 Preprint at https://doi.org/10.1016/j.cell.2018.01.016 (2018).
156. Schwarz, F. & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Current Opinion in Structural Biology* vol. 21 576–582 Preprint at https://doi.org/10.1016/j.sbi.2011.08.005 (2011).
157. Valverde, P., Ardá, A., Reichardt, N. C., Jiménez-Barbero, J. & Gimeno, A. Glycans in drug discovery. *MedChemComm* vol. 10 1678–1691 Preprint at https://doi.org/10.1039/c9md00292h (2019).

158. Varki, A. & Gagneux, P. *Biological Functions of Glycans*. *Essentials of Glycobiology* (2015).
159. Stowell, S. R., Ju, T. & Cummings, R. D. Protein glycosylation in cancer. *Annual Review of Pathology: Mechanisms of Disease* **10**, 473–510 (2015).
160. Hakomori, S. Tumor-associated carbohydrate antigens. *Annual review of immunology* vol. 2 103–126 Preprint at https://doi.org/10.1146/annurev.iy.02.040184.000535 (1984).
161. Oliveira-Ferrer, L., Legler, K. & Milde-Langosch, K. Role of protein glycosylation in cancer metastasis. *Seminars in Cancer Biology* vol. 44 141–152 Preprint at https://doi.org/10.1016/j.semcancer.2017.03.002 (2017).
162. An, G. *et al.* Increased susceptibility to colitis and colorectal tumors in mice lacking core 3-derived O-glycans. *Journal of Experimental Medicine* **204**, 1417–1429 (2007).
163. da Costa, V. & Freire, T. Advances in the Immunomodulatory Properties of Glycoantigens in Cancer. *Cancers 2022, Vol. 14, Page 1854* **14**, 1854 (2022).
164. Veillon, L., Fakih, C., Abou-El-Hassan, H., Kobeissy, F. & Mechref, Y. Glycosylation Changes in Brain Cancer. *ACS Chem Neurosci* **9**, 51–72 (2018).
165. Yamamoto, E. *et al.* Expression of N-acetylglucosaminyltransferase V in endometrial cancer correlates with poor prognosis. *British Journal of Cancer 2007 97:11* **97**, 1538–1544 (2007).
166. Yamamoto, H. *et al.* Beta1,6-N-acetylglucosamine-bearing N-glycans in human gliomas: implications for a role in regulating invasivity. *Cancer Res* **60**, 134–142 (2000).
167. Kudelka, M. R., Ju, T., Heimburg-Molinaro, J. & Cummings, R. D. Simple Sugars to Complex Disease—Mucin-Type O-Glycans in Cancer. *Adv Cancer Res* **126**, 53–135 (2015).
168. Beckwith, D. M. & Cudic, M. Tumor-associated O-glycans of MUC1: Carriers of the glyco-code and targets for cancer vaccine design. *Semin Immunol* **47**, 101389 (2020).
169. Gubbels, J. A. A. *et al.* MUC16 provides immune protection by inhibiting synapse formation between NK and ovarian tumor cells. *Mol Cancer* **9**, 1–14 (2010).
170. Kopitz, J. Lipid glycosylation: a primer for histochemists and cell biologists. *Histochemistry and Cell Biology 2016 147:2* **147**, 175–198 (2016).
171. Mantuano, N. R., Natoli, M., Zippelius, A. & Läubli, H. Tumor-associated carbohydrates and immunomodulatory lectins as targets for cancer immunotherapy. *J Immunother Cancer* **8**, 1222 (2020).
172. Breimer, M. E., Säljö, K., Barone, A. & Teneberg, S. Glycosphingolipids of human embryonic stem cells. *Glycoconj J* **34**, 713–723 (2017).
173. Ohtsubo, K. & Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **126**, 855–867 (2006).
174. Prydz, K. Determinants of Glycosaminoglycan (GAG) Structure. *Biomolecules 2015, Vol. 5, Pages 2003-2022* **5**, 2003–2022 (2015).
175. Ghiselli, G. Drug-Mediated Regulation of Glycosaminoglycan Biosynthesis. *Med Res Rev* **37**, 1051–1094 (2017).
176. Reily, C., Stewart, T. J., Renfrow, M. B. & Novak, J. Glycosylation in health and disease. *Nature Reviews Nephrology 2019 15:6* **15**, 346–366 (2019).
177. Sanderson, R. D. Heparan sulfate proteoglycans in invasion and metastasis. *Semin Cell Dev Biol* **12**, 89–98 (2001).

178. Yu, A. L. *et al.* Long-term follow-up of a phase III study of ch14.18 (dinutuximab) + cytokine immunotherapy in children with high-risk neuroblastoma: COG study ANBL0032. *Clinical Cancer Research* **27**, 2179–2189 (2021).
179. Ploessl, C., Pan, A., Maples, K. T. & Lowe, D. K. Dinutuximab: An Anti-GD2 Monoclonal Antibody for High-Risk Neuroblastoma. *Ann Pharmacother* **50**, 416–422 (2016).
180. Mueller, B. M., Romerdahl, C. A., Gillies, S. D. & Reisfeld, R. A. Enhancement of antibody-dependent cytotoxicity with a chimeric anti-GD2 antibody. *The Journal of Immunology* **144**, 1382–1386 (1990).
181. Yu, A. L. *et al.* Anti-GD2 Antibody with GM-CSF, Interleukin-2, and Isotretinoin for Neuroblastoma. *New England Journal of Medicine* **363**, 1324–1334 (2010).
182. Riley, R. S., June, C. H., Langer, R. & Mitchell, M. J. Delivery technologies for cancer immunotherapy. *Nature Reviews Drug Discovery 2018 18:3* **18**, 175–196 (2019).
183. Craddock, J. A. *et al.* Enhanced Tumor Trafficking of GD2 Chimeric Antigen Receptor T Cells by Expression of the Chemokine Receptor CCR2b. *J Immunother* **33**, 780 (2010).
184. Bocca, P. *et al.* Bevacizumab-mediated tumor vasculature remodelling improves tumor infiltration and antitumor efficacy of GD2-CAR T cells in a human neuroblastoma preclinical model. *Oncoimmunology* **7**, (2018).
185. Rashidijahanabad, Z. & Huang, X. Recent advances in tumor associated carbohydrate antigen based chimeric antigen receptor T cells and bispecific antibodies for anti-cancer immunotherapy. *Semin Immunol* **47**, 101390 (2020).
186. Wilkie, S. *et al.* Retargeting of Human T Cells to Tumor-Associated MUC1: The Evolution of a Chimeric Antigen Receptor. *The Journal of Immunology* **180**, 4901–4909 (2008).
187. Herlyn, M., Steplewski, Z., Herlyn, D. & Koprowski, H. Colorectal carcinoma-specific antigen: Detection by means of monoclonal antibodies. *Proc Natl Acad Sci U S A* **76**, 1438–1442 (1979).
188. Edrecolomab Overview - Creative Biolabs. https://www.creativebiolabs.net/edrecolomab-overview.htm.
189. Koprowski, H., Herlyn, M., Steplewski, Z. & Sears, H. F. Specific antigen in serum of patients with colon carcinoma. *Science (1979)* **212**, 53–55 (1981).
190. Lee, T., Teng, T. Z. J. & Shelat, V. G. Carbohydrate antigen 19-9 — tumor marker: Past, present, and future. *World J Gastrointest Surg* **12**, 468–490 (2020).
191. Magnani, J. L. *et al.* A monosialoganglioside is a monoclonal antibody-defined antigen of colon carcinoma. *Science (1979)* **212**, 55–56 (1981).
192. Ugorski, M. & Laskowska, A. Sialyl Lewisa: A tumor-associated carbohydrate antigen involved in adhesion and metastatic potential of cancer cells. *Acta Biochimica Polonica* vol. 49 303–311 Preprint at https://doi.org/10.18388/abp.2002_3788 (2002).
193. Engle, D. D. *et al.* The glycan CA19-9 promotes pancreatitis and pancreatic cancer in mice. *Science (1979)* **364**, 1156–1162 (2019).
194. Magnani, J. L., Ginsburg, V., Steplewski, Z. & Koprowsk, H. Identification of The Gastrointestinal and Pancreatic Cancer-Associated Antigen Detected by Monoclonal Antibody 19–9 in the Sera of Patients as a Mucin. *Cancer Res* **43**, 5489–5492 (1983).
195. Schmiegel, W. H., Greten, H. & Thiele, H. G. Characterization of CA 19–9 Bearing Mucins as Physiological Exocrine Pancreatic Secretion Products. *Cancer Res* **46**, 3605–3607 (1986).

196. Yue, T. *et al.* Identification of blood-protein carriers of the CA 19-9 antigen and characterization of prevalence in pancreatic diseases. *Proteomics* **11**, 3665–3674 (2011).
197. Blanas, A., Sahasrabudhe, N. M., Rodríguez, E., van Kooyk, Y. & van Vliet, S. J. Fucosylated antigens in cancer: An alliance toward tumor progression, metastasis, and resistance to chemotherapy. *Frontiers in Oncology* vol. 8 39 Preprint at https://doi.org/10.3389/fonc.2018.00039 (2018).
198. Stowell, C. P. & Stowell, S. R. Biologic roles of the <scp>ABH</scp> and Lewis histo-blood group antigens Part I: infection and immunity. *Vox Sang* **114**, 426–442 (2019).
199. Thornton, N. M. & Grimsley, S. P. Clinical significance of antibodies to antigens in the ABO, MNS, P1PK, Rh, Lutheran, Kell, Lewis, Duffy, Kidd, Diego, Yt, and Xg blood group systems. *Immunohematology* vol. 35 95–101 Preprint at https://doi.org/10.21307/IMMUNOHEMATOLOGY-2020-021 (2020).
200. Wang, G., Ge, Z., Rasko, D. A. & Taylor, D. E. Lewis antigens in Helicobacter pylori: biosynthesis and phase variation. *Mol Microbiol* **36**, 1187–1196 (2000).
201. Ma, Z. *et al.* Sialyl Lewis X mediates interleukin-1 beta-induced trophoblast adhesion to endometrial cells during human embryo implantation. *Biol Reprod* (2023) doi:10.1093/BIOLRE/IOAD007.
202. Kannagi, R. Carbohydrate antigen sialyl Lewis a - Its pathophysiological significance and induction mechanism in cancer progression. *Chang Gung Medical Journal* vol. 30 189–209 Preprint at (2007).
203. Borenstein-Katz, A. *et al.* Biomolecular Recognition of the Glycan Neoantigen CA19-9 by Distinct Antibodies. *J Mol Biol* **433**, 167099 (2021).
204. Norman, K. E., Anderson, G. P., Kolb, H. C., Ley, K. & Ernst, B. Sialyl Lewis(x) (sLe(x)) and an sLe(x) mimetic, CGP69669A, disrupt E- selectin-dependent leukocyte rolling in vivo. *Blood* **91**, 475–483 (1998).
205. Bussard, K. M., Mutkus, L., Stumpf, K., Gomez-Manzano, C. & Marini, F. C. Tumor-associated stromal cells as key contributors to the tumor microenvironment. *Breast Cancer Research* vol. 18 Preprint at https://doi.org/10.1186/s13058-016-0740-2 (2016).
206. Peixoto, A., Relvas-Santos, M., Azevedo, R., Lara Santos, L. & Ferreira, J. A. Protein glycosylation and tumor microenvironment alterations driving cancer hallmarks. *Front Oncol* **9**, 380 (2019).
207. Miyazaki, K. *et al.* Colonic Epithelial Cells Express Specific Ligands for Mucosal Macrophage Immunosuppressive Receptors Siglec-7 and -9. *The Journal of Immunology* **188**, 4690–4700 (2012).
208. Shiao, S. L., Preethi Ganesan, A., Rugo, H. S. & Coussens, L. M. Immune microenvironments in solid tumors: New targets for therapy. *Genes Dev* **25**, 2559–2572 (2011).
209. Huang, Z. & Liu, F. Diagnostic value of serum carbohydrate antigen 19-9 in pancreatic cancer: a meta-analysis. *Tumor Biology* **35**, 7459–7465 (2014).
210. Gul, K. *et al.* CA 19-9 level in patients with type 2 diabetes mellitus and its relation to the metabolic control and microvascular complications. *American Journal of the Medical Sciences* **341**, 28–32 (2011).
211. Ballehaninna, U. K. & Chamberlain, R. S. The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: An evidence based appraisal. *J Gastrointest Oncol* **3**, 105–119 (2012).

212. Sawada, R. *et al.* Human Monoclonal Antibodies to Sialyl-Lewis a (CA19.9) with Potent CDC, ADCC and Anti-Tumor Activity. *Clin Cancer Res* **17**, 1024 (2011).
213. Yamashita, T. *et al.* A vital role for glycosphingolipid synthesis during development and differentiation. *Proceedings of the National Academy of Sciences* **96**, 9142–9147 (1999).
214. Sud, M. *et al.* LMSD: LIPID MAPS structure database. *Nucleic Acids Res* **35**, D527–D532 (2007).
215. D'Angelo, G., Capasso, S., Sticco, L. & Russo, D. Glycosphingolipids: synthesis and functions. *FEBS J* **280**, 6338–6353 (2013).
216. Jacob, F. *et al.* Transition of mesenchymal and epithelial cancer cells depends on a1-4 galactosyltransferase-mediated glycosphingolipids. *Cancer Res* **78**, 2952–2965 (2018).
217. Russo, D., Capolupo, L., Loomba, J. S., Sticco, L. & D'Angelo, G. Glycosphingolipid metabolism in cell fate specification. *J Cell Sci* **131**, (2018).
218. Liang, Y. J. *et al.* Switching of the core structures of glycosphingolipids from globo- and lacto- to ganglio-series upon human embryonic stem cell differentiation. *Proc Natl Acad Sci U S A* **107**, 22564–22569 (2010).
219. Jennemann, R. & Gröne, H.-J. Cell-specific in vivo functions of glycosphingolipids: Lessons from genetic deletions of enzymes involved in glycosphingolipid synthesis. *Prog Lipid Res* **52**, 231–248 (2013).
220. Suila, H. *et al.* Are globoseries glycosphingolipids SSEA-3 and -4 markers for stem cells derived from human umbilical cord blood? *J Mol Cell Biol* **3**, 99–107 (2011).
221. Hakomori, S. Tumor-Associated Carbohydrate Antigens. *Annu Rev Immunol* **2**, 103–126 (1984).
222. Wiels, J., Fellous, M. & Tursz, T. Monoclonal antibody against a Burkitt lymphoma-associated antigen. *Proceedings of the National Academy of Sciences* **78**, 6485–6488 (1981).
223. Nudelman, E. *et al.* A Glycolipid Antigen Associated with Burkitt Lymphoma Defined by a Monoclonal Antibody. *Science (1979)* **220**, 509–511 (1983).
224. Distler, U. *et al.* Shiga Toxin Receptor Gb3Cer/CD77: Tumor-Association and Promising Therapeutic Target in Pancreas and Colon Cancer. *PLoS One* **4**, e6813 (2009).
225. Geyer, P. E. *et al.* Gastric Adenocarcinomas Express the Glycosphingolipid Gb3/CD77: Targeting of Gastric Cancer Cells with Shiga Toxin B-Subunit. *Mol Cancer Ther* **15**, 1008–1017 (2016).
226. Eierhoff, T. *et al.* A lipid zipper triggers bacterial invasion. *Proceedings of the National Academy of Sciences* **111**, 12895–12900 (2014).
227. Naiki, M. & Marcus, D. M. Immunochemical study of the human blood group P_1, P, and P^k glycosphingolipid antigens. *Biochemistry* **14**, 4837–4841 (1975).
228. Suzuki, A., Kundu, S. K. & Marcus, D. M. An improved technique for separation of neutral glycosphingolipids by high-performance liquid chromatography. *J Lipid Res* **21**, 473–477 (1980).
229. Park, S.-Y., Kwak, C.-Y., Shayman, J. A. & Kim, J. H. Globoside promotes activation of ERK by interaction with the epidermal growth factor receptor. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1820**, 1141–1148 (2012).
230. Song, Y., Withers, D. A. & Hakomori, S. Globoside-dependent Adhesion of Human Embryonal Carcinoma Cells, Based on Carbohydrate-Carbohydrate Interaction, Initiates Signal Transduction and Induces Enhanced Activity of Transcription Factors AP1 and CREB. *Journal of Biological Chemistry* **273**, 2517–2525 (1998).

231. Brimble, S. N. *et al.* The Cell Surface Glycosphingolipids SSEA-3 and SSEA-4 Are Not Essential for Human ESC Pluripotency. *Stem Cells* **25**, 54–62 (2007).
232. Haraguchi, N. SSEA-3 as a novel amplifying cancer cell surface marker in colorectal cancers. *Int J Oncol* (2012) doi:10.3892/ijo.2012.1713.
233. Chuang, P.-K. *et al.* Signaling pathway of globo-series glycosphingolipids and β1,3-galactosyltransferase V (β3GalT5) in breast cancer. *Proceedings of the National Academy of Sciences* **116**, 3518–3523 (2019).
234. Zhang, S. *et al.* Expression of potential target antigens for immunotherapy on primary and metastatic prostate cancers. *Clin Cancer Res* **4**, 295–302 (1998).
235. Yu, J., Hung, J., Wang, S., Cheng, J. & Yu, A. L. Targeting glycosphingolipids for cancer immunotherapy. *FEBS Lett* **594**, 3602–3618 (2020).
236. Kannagi, R. *et al.* New globoseries glycosphingolipids in human teratocarcinoma reactive with the monoclonal antibody directed to a developmentally regulated antigen, stage-specific embryonic antigen 3. *Journal of Biological Chemistry* **258**, 8934–8942 (1983).
237. Zhang, S. *et al.* Selection of tumor antigens as targets for immune attack using immunohistochemistry: II. Blood group-related antigens. *Int J Cancer* **73**, 50–56 (1997).
238. Mènard, S., Tagliabue, E., Canevari, S., Fossati, G. & Colnaghi, M. I. Generation of monoclonal antibodies reacting with normal and cancer cells of human breast. *Cancer Res* **43**, 1295–300 (1983).
239. Kudryashov, V. *et al.* Characterization of a mouse monoclonal IgG3 antibody to the tumor-associated globo H structure produced by immunization with a synthetic glycoconjugate. *Glycoconj J* **15**, 243–249 (1998).
240. Cheng, J. Y. *et al.* Globo-H ceramide shed from cancer cells triggers translin-associated factor X-dependent angiogenesis. *Cancer Res* **74**, 6856–6866 (2014).
241. Tsai, Y.-C. *et al.* A prevalent cancer associated glycan, globo H ceramide, induces immunosuppression by reducing Notch1 signaling. *J Cancer Sci Ther* **5**, 264–270 (2013).
242. Huang, C.-S. *et al.* Globo H-KLH vaccine adagloxad simolenin (OBI-822)/OBI-821 in patients with metastatic breast cancer: phase II randomized, placebo-controlled study. *J Immunother Cancer* **8**, e000342 (2020).
243. Huang, C.-S. *et al.* Randomized phase II/III trial of active immunotherapy with OPT-822/OPT-821 in patients with metastatic breast cancer. *Journal of Clinical Oncology* (2016) doi:10.1200/jco.2016.34.15_suppl.1003.
244. This Study is to Evaluate Safe and Effective Treatment Dose of OBI-888 in Patients With Locally Advanced or Metastatic Solid Tumors. - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT03573544.
245. Phase 1/2 Study of OBI-999 in Patients With Advanced Solid Tumors - Full Text View - ClinicalTrials.gov. https://www.clinicaltrials.gov/ct2/show/NCT04084366.
246. Soliman, C. *et al.* The terminal sialic acid of stage-specific embryonic antigen-4 has a crucial role in binding to a cancer-targeting antibody. *J Biol Chem* **295**, 1009–1020 (2020).
247. Kaneko, E. & Niwa, R. Optimizing therapeutic antibody function: Progress with fc domain engineering. *BioDrugs* vol. 25 1–11 Preprint at https://doi.org/10.2165/11537830-000000000-00000 (2011).

248. Diniz, F. *et al.* Glycans as Targets for Drug Delivery in Cancer. *Cancers* vol. 14 Preprint at https://doi.org/10.3390/cancers14040911 (2022).
249. Köhler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497 (1975).
250. Mitra, S. & Tomar, P. C. Hybridoma technology; advancements, clinical significance, and future aspects. *Journal of Genetic Engineering and Biotechnology* vol. 19 1–12 Preprint at https://doi.org/10.1186/s43141-021-00264-6 (2021).
251. Yu, J., Song, Y. & Tian, W. How to select IgG subclasses in developing anti-tumor therapeutic antibodies. *Journal of Hematology and Oncology* vol. 13 45 Preprint at https://doi.org/10.1186/s13045-020-00876-4 (2020).
252. Mullard, A. FDA approves 100th monoclonal antibody product. *Nature reviews. Drug discovery* vol. 20 491–495 Preprint at https://doi.org/10.1038/d41573-021-00079-7 (2021).
253. Sterner, E., Flanagan, N. & Gildersleeve, J. C. Perspectives on Anti-Glycan Antibodies Gleaned from Development of a Community Resource Database. *ACS Chem Biol* **11**, 1773–1783 (2016).
254. Sharma, P., Hu-Lieskovan, S., Wargo, J. A. & Ribas, A. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell* **168**, 707–723 (2017).
255. Female Breast Cancer Subtypes — Cancer Stat Facts. https://seer.cancer.gov/statfacts/html/breast-subtypes.html.
256. Hakomori, S. I. Tumor-associated carbohydrate antigens defining tumor malignancy: Basis for development of anti-cancer vaccines. *Adv Exp Med Biol* **491**, 369–402 (2001).
257. Itzkowitz, S. H. *et al.* Expression of Tn, sialosyl-Tn, and T antigens in human colon cancer. *Cancer Res* **49**, 197–204 (1989).
258. Goletz, S. *et al.* Thomsen-friedenreich antigen: Thé hiddeń tumor antigen. *Adv Exp Med Biol* **535**, 147–162 (2002).
259. Feng, D., Shaikh, A. S. & Wang, F. Recent Advance in Tumor-associated Carbohydrate Antigens (TACAs)-based Antitumor Vaccines. *ACS Chemical Biology* vol. 11 850–863 Preprint at https://doi.org/10.1021/acschembio.6b00084 (2016).
260. Hole, N. & Stern, P. L. A 72 kD trophoblast glycoprotein defined by a monoclonal antibody. *Br J Cancer* **57**, 239–246 (1988).
261. Koscielak, J., Hakomori, S. & Jeanloz, R. W. Glycolipid antigen and its antibody. *Immunochemistry* **5**, 441–455 (1968).
262. Pardo-Vargas, A., Delbianco, M. & Seeberger, P. H. Automated glycan assembly as an enabling technology. *Current Opinion in Chemical Biology* vol. 46 48–55 Preprint at https://doi.org/10.1016/j.cbpa.2018.04.007 (2018).
263. Broecker, F., Anish, C. & Seeberger, P. H. Generation of monoclonal antibodies against defined oligosaccharide antigens. in *Methods in Molecular Biology* vol. 1331 57–80 (Humana Press Inc., 2015).
264. Avci, F. Y., Li, X., Tsuji, M. & Kasper, D. L. A mechanism for glycoconjugate vaccine activation of the adaptive immune system and its implications for vaccine design. *Nat Med* **17**, 1602–1609 (2011).
265. Ishioka, G. Y. *et al.* MHC interaction and T cell recognition of carbohydrates and glycopeptides. *The Journal of Immunology* **148**, (1992).
266. Giannini, G., Rappuoli, R. & Ratti, G. The amino-acid sequence of two non-toxic mutants of diphtheria toxin: CRM45 and CRM197. *Nucleic Acids Res* **12**, 4063–4069 (1984).

267. Kaplonek, P. Improving the Immunoprotective Effect of Carbohydrate Vaccine Against Bacterial Pneumonia. (2019) doi:10.17169/REFUBIUM-27044.
268. Hamers-Casterman, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446–448 (1993).
269. Maass, D. R., Sepulveda, J., Pernthaner, A. & Shoemaker, C. B. Alpaca (Lama pacos) as a convenient source of recombinant camelid heavy chain antibodies (VHHs). *J Immunol Methods* **324**, 13–25 (2007).
270. Muyldermans, S. Nanobodies: Natural Single-Domain Antibodies. *Annu Rev Biochem* **82**, 775–797 (2013).
271. Huston, J. S. *et al.* Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli. *Proceedings of the National Academy of Sciences* **85**, 5879–5883 (1988).
272. Harmsen, M. M. & de Haard, H. J. Properties, production, and applications of camelid single-domain antibody fragments. *Applied Microbiology and Biotechnology* vol. 77 13–22 Preprint at https://doi.org/10.1007/s00253-007-1142-2 (2007).
273. Debie, P. *et al.* Size and affinity kinetics of nanobodies influence targeting and penetration of solid tumours. *J Control Release* **317**, 34–42 (2020).
274. Bannas, P. *et al.* In vivo near-infrared fluorescence targeting of T cells: comparison of nanobodies and conventional monoclonal antibodies. *Contrast Media Mol Imaging* **9**, 135–142 (2014).
275. Hoefman, S., Ottevaere, I., Baumeister, J. & Sargentini-Maier, M. L. Pre-Clinical Intravenous Serum Pharmacokinetics of Albumin Binding and Non-Half-Life Extended Nanobodies®. *Antibodies 2015, Vol. 4, Pages 141-156* **4**, 141–156 (2015).
276. Jovčevska, I. & Muyldermans, S. The Therapeutic Potential of Nanobodies. *BioDrugs* **34**, (123AD).
277. Schumacher, D., Helma, J., Schneider, A. F. L., Leonhardt, H. & Hackenberger, C. P. R. Nanobodies: Chemical Functionalization Strategies and Intracellular Applications. *Angewandte Chemie International Edition* **57**, 2314–2333 (2018).
278. Goldman, E. R., Liu, J. L., Zabetakis, D. & Anderson, G. P. Enhancing stability of camelid and shark single domain antibodies: An overview. *Front Immunol* **8**, 865 (2017).
279. Kunz, P. *et al.* Exploiting sequence and stability information for directing nanobody stability engineering. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1861**, 2196–2205 (2017).
280. Vu, K. B., Ghahroudi, M. A., Wyns, L. & Muyldermans, S. Comparison of llama VH sequences from conventional and heavy chain antibodies. *Mol Immunol* **34**, 1121–1131 (1997).
281. Govaert, J. *et al.* Dual beneficial effect of interloop disulfide bond for single domain antibody fragments. *Journal of Biological Chemistry* **287**, 1970–1979 (2012).
282. Monegal, A. *et al.* Immunological applications of single-domain llama recombinant antibodies isolated from a naïve library. *Protein Engineering, Design and Selection* **22**, 273–280 (2009).
283. Yan, J., Li, G., Hu, Y., Ou, W. & Wan, Y. Construction of a synthetic phage-displayed Nanobody library with CDR3 regions randomized by trinucleotide cassettes for diagnostic applications. *J Transl Med* **12**, 343–343 (2014).
284. Kolkman, J. A. & Law, D. A. Nanobodies – from llamas to therapeutic proteins. *Drug Discov Today Technol* **7**, e139–e146 (2010).

285. Iqbal, U. *et al.* Kinetic analysis of novel mono- and multivalent VHH-fragments and their application for molecular imaging of brain tumours. *Br J Pharmacol* **160**, 1016–1028 (2010).
286. Palomo, C. *et al.* Trivalency of a nanobody specific for the human respiratory syncytial virus fusion glycoprotein drastically enhances virus neutralization and impacts escape mutant selection. *Antimicrob Agents Chemother* **60**, 6498–6509 (2016).
287. Huet, H. A. *et al.* Multivalent nanobodies targeting death receptor 5 elicit superior tumor cell killing through efficient caspase induction. *MAbs* **6**, 1560–1570 (2014).
288. Stijlemans, B. *et al.* Efficient Targeting of Conserved Cryptic Epitopes of Infectious Agents by Single Domain Antibodies: AFRICAN TRYPANOSOMES AS PARADIGM *. *Journal of Biological Chemistry* **279**, 1256–1261 (2004).
289. Ebrahimizadeh, W. *et al.* Isolation and characterization of protective anti-LPS nanobody against V. cholerae O1 recognizing Inaba and Ogawa serotypes. *Appl Microbiol Biotechnol* **97**, 4457–4466 (2013).
290. Behar, G. *et al.* Llama single-domain antibodies directed against nonconventional epitopes of tumor-associated carcinoembryonic antigen absent from nonspecific cross-reacting antigen. *FEBS Journal* **276**, 3881–3893 (2009).
291. Khilji, S. K. *et al.* Generation of glycan-specific nanobodies. *Cell Chem Biol* 1–9 (2022) doi:10.1016/j.chembiol.2022.05.007.
292. Geissner, A. *et al.* Microbe-focused glycan array screening platform. *Proc Natl Acad Sci U S A* **116**, 1958–1967 (2019).
293. Weitzenfeld, P., Bournazos, S. & Ravetch, J. V. Antibodies targeting sialyl Lewis A mediate tumor clearance through distinct effector pathways. *J Clin Invest* **129**, 3952–3962 (2019).
294. Heckman, K. L. & Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc* **2**, 924–932 (2007).
295. Xu, L., Song, X. & Jia, L. A camelid nanobody against EGFR was easily obtained through refolding of inclusion body expressed in Escherichia coli. *Biotechnol Appl Biochem* **64**, 895–901 (2017).
296. Protein BLAST: Align two or more sequences using BLAST. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&LINK_LOC=blasttab.
297. Delaitsch, A. T. *et al.* Selective Recognition of Carbohydrate Antigens by Germline Antibodies Isolated from AID Knockout Mice. *J Am Chem Soc* **144**, 4925–4941 (2022).
298. Huang, X., Schurman, N., Handa, K. & Hakomori, S. Functional role of glycosphingolipids in contact inhibition of growth in a human mammary epithelial cell line. *FEBS Lett* **591**, 1918–1928 (2017).
299. MCF7 - HTB-22 | ATCC. https://www.atcc.org/products/htb-22.
300. Fujitani, N. *et al.* Total cellular glycomics allows characterizing cells and streamlining the discovery process for cellular biomarkers. *Proc Natl Acad Sci U S A* (2013) doi:10.1073/pnas.1214233110.
301. Meyer, B. & Peters, T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed Engl* **42**, 864–890 (2003).
302. Kumar, A., Wagner, G., Wüthrich, K., Kumar, A. & Ernst, R. R. Buildup Rates of the Nuclear Overhauser Effect Measured by Two-Dimensional Proton Magnetic Resonance Spectroscopy: Implications for Studies of Protein Conformation. *J Am Chem Soc* **103**, 3654–3658 (1981).

303. Ghosh, G., van Duyne, G., Ghosh, S. & Sigler, P. B. Structure of NF-κB p50 homodimer bound to a κB site. *Nature 1995 373:6512* **373**, 303–310 (1995).
304. Li, Z., Hartl, F. U. & Bracher, A. Structure and function of Hip, an attenuator of the Hsp70 chaperone cycle. *Nature Structural & Molecular Biology 2013 20:8* **20**, 929–935 (2013).
305. Collins, B. E. & Paulson, J. C. Cell surface biology mediated by low affinity multivalent protein-glycan interactions. *Curr Opin Chem Biol* **8**, 617–625 (2004).
306. Schäfer, F., Seip, N., Maertens, B., Block, H. & Kubicek, J. Purification of GST-Tagged Proteins. *Methods Enzymol* **559**, 127–139 (2015).
307. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881–10890 (1988).
308. Mitchell, L. S. & Colwell, L. J. Comparative analysis of nanobody sequence and structure data. *Proteins: Structure, Function, and Bioinformatics* **86**, 697–706 (2018).
309. Lapidus, S. *et al.* Plasmodium infection is associated with cross-reactive antibodies to carbohydrate epitopes on the SARS-CoV-2 Spike protein. *Scientific Reports 2022 12:1* **12**, 1–16 (2022).
310. Warszawski, S. *et al.* Optimizing antibody affinity and stability by the automated design of the variable light-heavy chain interfaces. *PLoS Comput Biol* **15**, e1007207 (2019).
311. Kijanka, M., Dorresteijn, B., Oliveira, S. & van Bergen En Henegouwen, P. M. P. Nanobody-based cancer therapy of solid tumors. *Nanomedicine* **10**, 161–174 (2015).
312. Liu, M., Li, L., Jin, D. & Liu, Y. Nanobody-A versatile tool for cancer diagnosis and therapeutics. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **13**, e1697–e1697 (2021).
313. Wang, J. *et al.* Research Progress and Applications of Multivalent, Multispecific and Modified Nanobodies for Disease Treatment. *Front Immunol* **12**, 6013 (2022).
314. Wellner, A. *et al.* Rapid generation of potent antibodies by autonomous hypermutation in yeast. *Nature Chemical Biology 2021 17:10* **17**, 1057–1064 (2021).
315. Bao, C. *et al.* The Application of Nanobody in CAR-T Therapy. *Biomolecules 2021, Vol. 11, Page 238* **11**, 238 (2021).
316. Beilsten-Edmands, J. *et al.* Scaling diffraction data in the DIALS software package: algorithms and new approaches for multi-crystal scaling. *Acta Crystallogr D Struct Biol* **76**, 385 (2020).
317. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186–190 (2010).
318. Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol* **74**, 85 (2018).
319. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242 (2011).
320. McCoy, A. J. *et al.* Phaser crystallographic software. *urn:issn:0021-8898* **40**, 658–674 (2007).
321. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *urn:issn:0907-4449* **67**, 355–367 (2011).
322. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).
323. Agirre, J. *et al.* Privateer: software for the conformational validation of carbohydrate structures. *Nat Struct Mol Biol* **22**, 833–834 (2015).

324. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220 (2014).
325. Mayer, M. & Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J Am Chem Soc* (2001) doi:10.1021/ja0100120.
326. Angulo, J. *et al.* Blood group B galactosyltransferase: Insights into substrate binding from NMR experiments. *J Am Chem Soc* (2006) doi:10.1021/ja063550r.
327. Kumar, A., Wagner, G., Wüthrich, K., Kumar, A. & Ernst, R. R. Buildup Rates of the Nuclear Overhauser Effect Measured by Two-Dimensional Proton Magnetic Resonance Spectroscopy: Implications for Studies of Protein Conformation. *J Am Chem Soc* (1981) doi:10.1021/ja00403a008.

7. Scientific publications and conferences

Publications:

1. Khilji SK*, Goerdeler F*, Frensemeier K, Warschkau D, Lühle J, Fandi Z, Schirmeister F, Chen ZA, Turak O, Mallagaray A, Boerno S, Timmermann B, Rappsilber J, Seeberger PH, Moscovitz O. *Generation of glycan-specific nanobodies.* Cell Chem. Biol. (2022) 1–9. doi:10.1016/j.chembiol.2022.05.007 (*equal contribution)
2. Ben Ami Pilo H*, Khilji S K*, Lühle J*, Biskup K, Levy Gal B, Rosenhek Goldian I, Alfandari D, Revach OY, Kiper E, Morandi MI, Rotkopf R, Porat Z, Blanchard V, Seeberger PH, Regev-Rudzki N, Moscovitz O. *Sialylated N-glycans mediate monocyte uptake of extracellular vesicles secreted from Plasmodium falciparum-infected red blood cells.* Journal of Extracellular Biology (2022), 1(2): e33, doi: 10.1002/jex2.33 (*equal contribution)
3. Gowravaram M, Schwarz J, Khilji SK, Urlaub H, Chakrabarti S. *Insights into the assembly and architecture of a Staufen-mediated mRNA decay (SMD)-competent mRNP*. Nat Commun **10**, 5054 (2019). doi: 10.1038/s41467-019-13080-x
4. Alagesan K, Khilji SK, Kolarich D. *It is all about the solvent: on the importance of the mobile phase for ZIC-HILIC glycopeptide enrichment.* Anal Bioanal Chem **409**, 529–538 (2017). doi: 10.1007/s00216-016-0051-6

Conference:

1. Nanobodies (Hybrid 2nd edition) -Brussels, Belgium, 2021 (poster)