Structural Analysis of Receptor Interactions and Structure-based Protein Engineering of Prenyltransferases

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Elena Ostertag

aus Sigmaringen

Tübingen

2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 21.06.2022

Dekan: Prof. Dr. Thilo Stehle
1. Berichterstatter: Prof. Dr. Thilo Stehle
2. Berichterstatterin: PD Dr. Martina Mühlenhoff

Abstract

This thesis combines three unrelated projects with diverse biological background. For all projects, the scientific questions were addressed by X-ray structure determination of the target molecules.

Since 2019, SARS-CoV-2 causes an ongoing pandemic with 460 million infections and 6 million deaths worldwide (date 15.03.2022) and up to now, all available vaccines offer no full, long-life protection against an infection with this virus. Nbs against the receptor binding domain (RBD) of SARS-CoV-2 were produced and the affinity to the RBD as well as the virus neutralisation capabilities were determined. In this work, two high-affinity binders, NM1226 and NM1230, were structurally investigated in complex with the RBD to analyse the binding properties on an atomic level. Both Nbs bind to the RBD with a non-overlapping interface. As a consequence, both Nbs utilise different strategies to prevent RBD interactions and therefore can block binding to the host cell entry receptor angiotensin converting enzyme 2 (ACE2). By combining these two Nbs, a new biparatopic nanobody (biNb) NM1267, which binds to the RBD with picomolar affinity, was created by our cooperation partner. Additionally, analysis of the binding ability to emerging SARS-CoV-2 variants revealed robust binding of the biNb to the RBD. An assay was established with the biNb to analyse the antibody composition of SARS-CoV-2 infected patients that allows to estimate the amount of neutralising antibodies in patients.

Secondary metabolites are chemically diverse, small molecules featuring antifungal, antibacterial or anti-inflammatory properties. Some of these compounds possess a prenyl entity, which often enhances their bioactivity. This modification is carried out by prenyltransferases in a chemo- and regioselectivity reaction. The selectivity of the enzymatic prenylation offers a substantial advantage compared to *de novo* synthesis, which makes these enzymes valuable targets for biotechnological applications. Over the last ten years, a subgroup of prenyltransferases, the dimethylallyl tryptophan synthases (DMATS) were analysed in detail with the aim of controlling the reaction mechanism and influencing the substrate specificity. The work presented here improved the general understanding of the enzyme family and provides the first 5-DMATS structure in complex with its natural substrates L-tryptophan and dimethylallyl pyrophosphate (DMAPP). In addition, a ligand-bound structure of a C-6 prenylating DMATS (6-DMATS) was obtained. By investigation of the structures, crucial residues for catalysis were determined in both enzymes. The regioselectivity of 6-DMATS could be switched towards a 5-DMATS by structure-based engineering of the enzyme. This principle could be applied to other prenyltransferases to specifically modify the enzymatic efficiency for biotechnological production of new compounds.

Abstract

Sialic acid-binding immunoglobulin-like lectins (Siglecs) are receptors that are mainly present on haematopoietic cells and recognise sialic acid residues on glycoproteins. Siglec-11 is expressed on macrophages and microglia and has, upon binding to α2,8-linked polysialic acid (polySia), an inhibitory regulatory role on immune cell activation. PolySia is found as rare decoration on glycoproteins mainly expressed during development, and in plastic regions of the brain of healthy adults. In addition, some pathogens, such as *Escherichia coli* serotype K1, and tumours present polySia on their surface to exploit the inhibitory effect on immune cell activation and evade the immune system. Compared to other Siglecs, whose specificity is mainly determined by the linkage and modification of the last few entities of the ligand glycan, Siglec-11 interaction with polySia differs in fundamental aspects. Only polySia with a degree of polymerisation larger than 20 has an anti-inflammatory effect on human macrophages. Therefore, other structurally characterised interactions of Siglecs with ligands offer no sufficient model to explain polySia engagement by Siglec-11. Siglec-11 is characterised by a N-terminal ligand binding V-set and three Ig-like C2-set domains. In this work, Siglec-11 variants with a varying number of extracellular domains were expressed and purified for structural studies and affinity determination to polySia. A structure of Siglec-11 V-set domain could be obtained at 2.15 Å resolution and lays the foundation for further structural studies with polySia.

Zusammenfassung

Diese Arbeit fasst drei Projekte mit unterschiedlichen biologischen Kontexten zusammen. Die Fragestellungen dieser Projekte wurden mittels Röntgenstrukturanalyse von Proteinkristallen adressiert. In den folgenden Absätzen werden daher die Ergebnisse dieser drei Projekte resümiert.

Das Virus "Severe Respiratory Syndrome Corona Virus 2" (SARS-CoV-2) ist für eine seit 2019 andauernde Pandemie mit 460 Millionen Infizierten und 6 Millionen Toten weltweit verantwortlich (Stand: 15.03.2022). Bis heute ist keine Impfung verfügbar, die einen vollständigen und andauernden Schutz gegen das Virus bietet. Es wurden Nanobodies (Nb) produziert, die an die Rezeptorbindedomäne (RBD) des SARS-CoV-2 binden. Deren Affinität sowie ihre SARS-CoV-2 neutralisierenden Eigenschaften wurden experimentell bestimmt. Zwei dieser Nb, NM1226 und NM1230, binden mit hoher Affinität an die RBD und ihre Bindungsstellen konnten mittels Strukturanalyse im Zuge dieser Arbeit auf atomarer Ebene charakterisiert werden. Dabei wurde gezeigt, dass beide Nbs die Interaktion des RBDs mit dem Wirtszellrezeptor, Angiotensin konvertierendes Enzym 2 (ACE2), unterbinden, wobei sich der Wirkungsmechanismus unterscheidet. Durch Fusion der beiden Nbs konnte unser Kooperationsparnter einen neuen biparatopen Nb (biNb) erzeugen, der eine pikomolare Affinität zur RBD aufweist und zuätzlich zuverlässig an SARS-CoV-2 Varianten bindet. Mit dem biNb wurde ein Test entwickelt, mit dem der Anteil an neutralisierenden Antikörpern von allen SARS-CoV-2-bindenden Antikörpern eines Patienten nach einer Infektion mit SARS-CoV-2 abgeschätzt werden kann.

Sekundärmetabolite sind kleine, chemisch diverse Moleküle, die unter anderem antimykotische, antibakterielle oder auch entzündungshemmende Wirkungen aufweisen, was sie besonders interessant für die medizinische Anwendung macht. Die Prenylierung ist eine häufige Modifikation dieser Verbindungen, die die Bioaktivität erhöhen kann und von Prenyltransferasen in einer regio- und chemoselektiven Reaktion katalysiert wird. In den letzten zehn Jahren wurde eine Untergruppe dieser Enzyme, die Dimethylallyltryptophan Synthasen (DMATS), intensiv erforscht, um die Parameter der Reaktions- und Substratspezifität zu ermitteln. In dieser Arbeit wurde die erste Komplexstruktur einer C-5 prenylierenden DMATS (5-DMATS) mit ihren natürlichen Substraten L-tryptophan und Dimethylallylpyrophsophat (DMAPP) ermittelt, welche das grundsätzliche Verständnis für die Enzymgruppe erweitert. Darüber hinaus wurde eine ligandgebundene Struktur einer C-6 prenylierenden DMATS (6-DMATS) bestimmt. Anhand der Strukturen konnten essentielle katalytische Aminosäuren bestimmt und basierend auf den gewonnenen Erkenntnissen konnte die Regioselektivität einer 6-DMATS zu einer 5-DMATS geändert werden. Dieses Prinzip kann auf andere Prenyltransferasen übertragen werden, um gezielt Enzyme für die biotechnologische Produktion von neuen Wirkstoffen anzupassen.

Zusammenfassung

Sialinsäure bindende Immunoglobulin-like Lektine (Siglecs) sind Rezeptoren, die Sialinsäurenverbindungen auf Glykoproteine erkennen und meist auf der Oberfläche von hämatopoetischen Zellen vorkommen. Siglec-11 ist ein Oberflächenrezeptor der von Makrophagen und Mikroglia exprimiert wird und, als Reaktion auf die Bindung zu polySia, hemmend auf die Signalkaskade der Immunzellaktivierung wirkt. α2,8-verknüpfte Polysialinsäure (polySia) ist eine vergleichsweise seltene Modifikation von Glykoproteinen, die vor allem während der Entwicklung, aber auch in plastischen Regionen des Gehirns von gesunden Erwachsenen auftritt. Darüber hinaus gibt es Pathogene, wie *Escherichia coli* Serotyp K1 und Tumore, welche polySia auf ihren Zelloberlfächen exprimieren, um die hemmende Wirkung auf das Immunsystems auszunutzen. Die Interaktion von Silgec-11 mit polySia ist untypisch. Die Ligandspezifität von anderen Siglecs ist meist durch die Verknüpfung und Modifikationen weniger Zuckerreste bestimmt. Im Gegensatz dazu wurde für polySia ein Polymerisationsgrad (DP) von mindestens 20 Zuckereinheiten ermittelt, damit eine hemmende Siglec-11 abhängige Wirkung in Makrophagen ausgelöst wird. Eine Analyse der verfügbaren strukturellen Daten anderer Siglecs und ihren Liganden ist daher unzureichend, um die Bindung von polySia zu Siglec-11 zu modellieren. Siglec-11 ist aus einer N-terminalen V-Set Domäne und drei Ig-like C2-Set Domänen aufgebaut. Für die strukturelle Charakterisierung und zur Bestimmung der Affinität der Interaktion von polySia und Siglec-11 wurden mehrere Varianten von Siglec-11 mit einer unterschiedlichen Anzahl von Extrazellulärendomänen produziert und gereinigt. Die Struktur der Siglec-11 V-set Domäne wurde mit einer Auflösung 2.15 Å bestimmt und bildet die Grundlage für weitere strukturelle Analysen von Siglec-11 im Komplex mit polySia.

Contents

List of Tables

List of Figures

List of Figures

Abbreviations

D/L-Trp	D/L-Tryptophan
3/6/8-PN	3/6/8-Prenylnaringenin
ABBA	$\alpha\beta\beta\alpha$ (Secondary Structure Elements)
ACE2	Angiotension Converting Enzyme 2
AD	Alzheimer's Disease
BDNF	Brain-derived Neurotrophic Factor
biNb	Biparatopic Nanobody
BLI	Biolayer Interferometry
bp	Base Pairs
CCR7	Chemokine Receptor Type 7
CD	Circular Dichroism
CD36	Cluster of Differentiation 36
cDNA	Complementary DNA
CDR	Complementary Determining Region
CH	Constant Region Heavy Chain
CL	Constant Region Light Chain
CMP	Cytidine Monophosphate
Con A	Concanavalin A
CoV	Corona Virus
COVID-19	Corona Virus Disease 2019
C-set	Constant-Set
DMAPP	Dimethylallyl Pyrophosphate
DMATS	Dimethylallyl Tryptophan Synthase
dNTP	Deoxynucleotide Triphosphate
DP	Degree of Polymerisation
DPP4	Dipeptidyl Peptidase 4
DTT	Dithiothreitol
E. coli	*Escherichia Coli*
ECD	Extracellular Domain
ER	Endoplasmic Reticulum
ERGIC	Endoplasmic Reticulum Golgi Intermediate Compartment
Fab	Antigen Binding Fragment
Fc	Fragment Crystallisable
FGF2	Fibroblast Growth Factor 2
FPP	Farnesyl Pyrophosphate
FR	Framework Region

Abbreviations

FS	Fine Screen
Fuc	Fucose
GAG	Glucosaminoglycan
GalNAc	N-Acetylgalactosamine
Gal	Galactose
GlcA	Glucoronic Acid
GlcNAc	N-Acetylglucosamine
Glc	Glucose
GnTI	N-Acetylglucosaminyltransferase I
GPI	Glycosylphosphatidylinositol
GPP	Geranyl Pyrophosphate
HDX-MS	Hydrogen Exchange Mass Spectrometry
HEK293 cells	Human Embryonic Kidney 293 Cells
HIV	Human Immunodeficiency Virus
HRP	Horseradish Peroxidase
IC_{50}	Half Maximal Inhibitory Concentration
IEX	Ion Exchange Chromatography
Ig	Immunoglobulin
IgG	Immunoglobulin Type G
IMAC	Immobilised Metal Affinity Chromatography
ITAM	Immunoreceptor Tyrosine-based Activation Motif
ITIM	Immunoreceptor Tyrosine-based Inhibitory Motif
I-type	Immunoglobulin-like
K_D	Dissociation Constant
KDN	Deaminoneuraminic Acid (2-Keto-3-Deoxy-D-Glycero-D-Galacto-Nononic Acid)
LB	Lysogeny Broth
mAB	Monoclonal Antibody
M. haemolytica	*Mannheimia Haemolytica*
MAG	Myelin-associated Glycoprotein
Man	Mannose
MERS	Middle East Respiratory Syndrome
MR	Molecular Replacement
MWCO	Molecular Weight Cut-Off
N. menigitidis	*Neisseria Meningitidis*
Nb	Nanobody
NCAM	Neuronal Cell Adhesion Molecule
NCS	Non-Crystallographic Symmetry
NET	Neutrophil Extracellular Trap
Neu	Neuraminic Acid

Neu5Ac	N-Acetylneuraminic Acid
Neu5Gc	N-Glycolylneuraminic Acid
NK cells	Natural Killer Cells
NMR	Nuclear Magnetic Resonance
NRP-2	Neuropilin 2
oligoSia	α2,8-linked Sialic Acid (Chain Length 2-7 Monomers)
ON	Over Night
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PEI	Polyethylenimine
polySia	α2,8-linked Sialic Acid (Chain Length > 7 Monomers)
PT	Prenyltransferase
RBD	Receptor Binding Domain
RBM	Receptor Binding Motif
RMSD	Root Mean Square Deviation
S1/2	Spike Protein 1/2
SARS	Severe Acute Respiratory Syndrom
SARS-CoV	Severe Acute Respiratory Syndrom Corona Virus
scFv	Single Chain Fragment Variable
SD75/200	Superdex 75/200
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SHP1/2	SH-Domain Containing Phosphatase 1 and 2
Siglec	Sialic Acid Binding Immunoglobulin-like Lectin
Siglec-11(4D/5D)	Siglec-11 4/5 Domain Variant
SLS	Swiss Light Source
SS	Secretion Signal
Syn-CAM-1	Synaptic Cell Adhesion Molecule 1
TAM	Tumour Associated Macrophages
TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch Virus
TLS	Translation/Libration/Screw
TMPRSS2	Transmembrane Serine Protease 2
VH	Variable Region Heavy Chain
VHH	Variable Region Heavy Chain Only Antibody
VL	Variable Region Light Chain
V-set	Variable-Set
WHO	Word Health Organisation
Xyl	Xylose

1 NeutrobodyPlex - Monitoring SARS-CoV-2 Neutralizing Immune Response Using Nanobodies

1.1 Introduction

1.1.1 Coronavirus

Coronaviruses (CoV) are positive-sense (+ sense) single stranded RNA viruses that belong to the order of *Nidovirales* and the family of *Coronaviridae*. Based on serological and genomic evidence, coronaviruses are classified into four genera (α-CoV, β-CoV, γ-CoV, δ-CoV). β-CoV are further distinguished in lineage A, B, C and D, whereas each subtype uses a different entry receptor to infiltrate host cells [1].

Coronaviridae were first identified in 1960 as pathogens causing a common cold [2]. CoV can cause respiratory and enteric disease and are not only found in humans but also in other mammals, especially bats, or in avian species [3]. So far, seven human pathogenic coronaviruses have been identified that predominantly attack the upper and lower respiratory tract. Four of those (HCoV-OC43, HCoV-NL63, HCoV-HKU1, HCoV-229E) cause a common cold, and three emerging viruses can provoke highly pathogenic respiratory infections [4]. One virus emerged in 2002, which belongs to the lineage B of *Coronaviridae* and causes a severe acute respiratory syndrome (SARS) [5]. The virus (SARS-CoV) was supposed to be a horse shoe bat pathogen, which was transmitted to human most likely via palm civets [6]. For infection, the virus uses the angiotensin converting enzyme 2 (ACE2) as its receptor to enter the cell. Since ACE2 is mainly present in the lung, gastrointestinal tract and kidney tissue, SARS-CoV-2 could be tracked in these organs [7, 8]. Human-to-human transmission takes place via small airborne droplets that occur during coughing or sneezing, but the virus can also be transmitted by personal contact. A mean incubation time of seven days was determined until infected individuals develop symptoms such as fever, dry cough, muscular pain and headache [9, 10]. During 2002-2004, around 8000 individuals were infected worldwide with a fatality rate of about 10%, but the epidemic could be contained with no more documented SARS-CoV infections after 2004 [11, 12, 13].

Middle east respiratory syndrome (MERS), caused by a novel CoV of lineage C of *Coronaviridae*, was first described 2012 in Saudi Arabia [14, 15]. The origin of MERS-CoV is not unequivocally determined, but there is strong evidence for a zoonotic transmission most probably from bats via dromedary camels to humans [16, 17]. In contrast to SARS-CoV, MERS-CoV did not spread rapidly and vanished again, but MERS-CoV cause

sporadic outbreaks with about 2300 cases in total and a fatality rate of 35% [16, 12]. Most of the virus transmissions were documented in a nosocomial environment, as MERS-CoV possesses a limited human-to-human transmission due to the attachment receptor tropism of dipeptidyl peptidase 4 (DPP4), which is mainly expressed in tissue of the lower respiratory tract [18, 19]. A higher risk for a severe course of the disease could be determined especially for elderly individuals with pre-existing illness [20, 19].

1.1.2 SARS-CoV-2

In December 2019, a cluster of patients were hospitalised with symptoms of atypical pneumonia in Wuhan (China). In January 2020, a novel CoV was identified to cause this new disease, which was determined to be a close relative of SARS-CoV, giving this virus the name SARS-CoV-2 and the corresponding disease COVID-19 (Coronavirus Disease 2019) [21, 22, 23]. The virus origin is not fully known yet, but the most plausible theory of transmission is from bats via an intermediate host to humans [24, 25]. SARS-CoV-like viruses with high sequence similarity to SARS-CoV-2 were identified in bats and first documented cases of COVID-19 can be tracked back to a seafood market in Wuhan [22, 25].

SARS-CoV-2 is a highly infectious virus with an airborne transmission via small droplets or aerosols that can persist in the air for several hours [26]. Similar to SARS-CoV, SARS-CoV-2 uses ACE2 as its attachment and entry receptor into the host cells, even though the spike proteins of both viruses share only a sequence similarity of 76% [27]. The highest viral load can be determined in the respiratory system, but SARS-CoV-2 could also be detected in other tissues such as the gastrointestinal tract of infected individuals [28]. Interestingly, infected persons can transmit the virus prior to developing symptoms, which facilitates viral spreading [29, 30]. The disease picture ranges from mild symptoms such as dry coughing, loss of smell and taste or fever to severe respiratory failure and the need to ventilate patients [31, 32]. Similar to SARS-CoV and MERS-CoV, a number of individuals carry the virus without developing any symptoms [33, 30]. Age and health condition of the individuals have a significant impact on disease progression, and elderly persons or patients with underlying medical conditions suffer disproportionally more often from a severe course of the disease [12, 34, 35]. In contrast to previous SARS-CoV and MERS-CoV epidemics, COVID-19 is a still ongoing pandemic, with 460 million infections and 6 million deaths worldwide (status: 15.03.22) [36]. Hence, SARS-CoV-2 possesses a higher transmission efficiency with a basic reproduction number of three, but the course of the disease is less lethal with a fatality rate of 0.4-5%, depending on the study and the number of detected asymptomatic cases [33, 37, 35].

Virus Architecture

CoV are enveloped viruses with a particle size of 80-120 nm and a spherical, pleomorphic shape (figure 1.1) [38, 39]. They represent the largest family of single stranded, + sense RNA viruses possessing a non-segmented genome of 30 kb in case of SARS-CoV-2 [40].

The genome is protected by a 5'-cap and a 3'-polyA tail and multiple open reading frames, encoding about 29 proteins, including four structural proteins: the spike protein, the envelope protein, the membrane protein and the nucleocapsid protein [41]. The RNA genome is bound by the nucleocapsid protein in a beads-on-a-string fashion to condense the genome and form a symmetric helical nucleocapsid (figure 1.1) [42]. The membrane protein, which is a glycoprotein embedded via three transmembrane domains, is the most abundant structural protein in the virus particle envelope [43]. The envelope protein, also a transmembrane protein, is only present in low amounts [44].
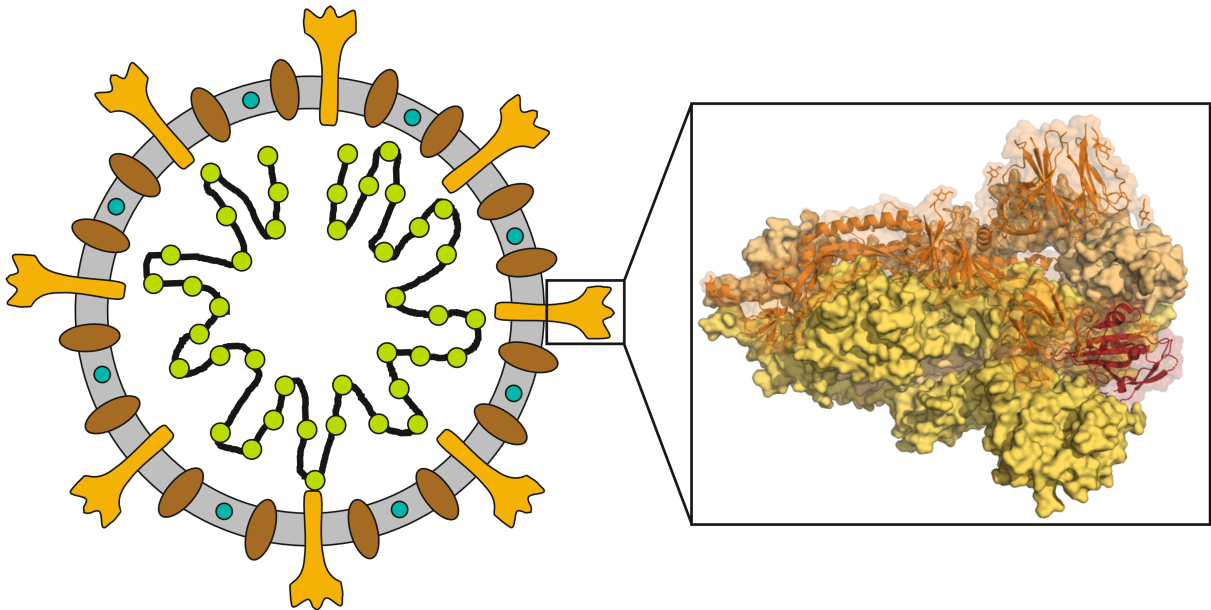
Figure 1.1: SARS-CoV-2 particle and the trimeric spike protein. Schematic representation of the viral envelope with the membrane protein (brown), the envelope protein (blue) and the spike protein (orange) embedded in the double membrane (grey). The RNA genome (black) is condensed by the nucleocapsid protein (green) in a bead-to-string fashion. The spike protein trimer is shown as surface representation in orange shades with one protomer drawn as cartoon and the included receptor binding domain (RBD) highlighted in dark red (pdb code: 6vxx).

The spike protein of SARS-CoV-2 is a homotrimeric attachment protein in the virus envelope that belongs to class I viral fusion proteins and enables attachment, as well as fusion of viral and host membranes. The spike monomer is post-translationally cleaved into S1 and S2 proteins by the furin protease of the host cell. The two functional subunits, S1 for receptor binding and S2 for membrane fusion, remain associated as one homotrimeric protein after cleavage [45]. The structure of S1 is dominated by β-sheets and provides the receptor binding domain (RBD), whereas S2 is predominantly based on α-helical structures including a transmembrane part and a short intra-viral part with a palmitoylated cysteine-rich region and a variable C-terminus [46, 47]. The RBD comprises the receptor binding motif (RBM), which mediates interactions with the entry receptor ACE2 in case of SARS-CoV-2 [47]. The diversity in the RBD, including the RBM, determines the host and tissue tropism of CoV-types. Here, it could be shown that the spike protein of SARS-CoV-2 can recognise ACE2 of other species such as pigs, dogs and cats, which enables a good virus transmission to other organisms [48]. RBDs of SARS-CoV-2 are

mainly in down conformation protecting the RBD from being recognised by the immune system. Statistically, only one RBD in the spike protein trimer is in up-conformation, which enables ACE2 binding (figure 1.2) [46]. This adaptation indicates a fine-tuning between masking of the RBM, a putative target for neutralisation by the humoral immune response, and presenting the RBM for attachment to host cells. Furthermore, spike proteins are highly glycoslyated, with 22 *N*-glycoslyation sites per protomer helping to mask parts of the spike protein from immune recognition [49].

(A)　　　　　　**(B)**　　　　　　**(C)**
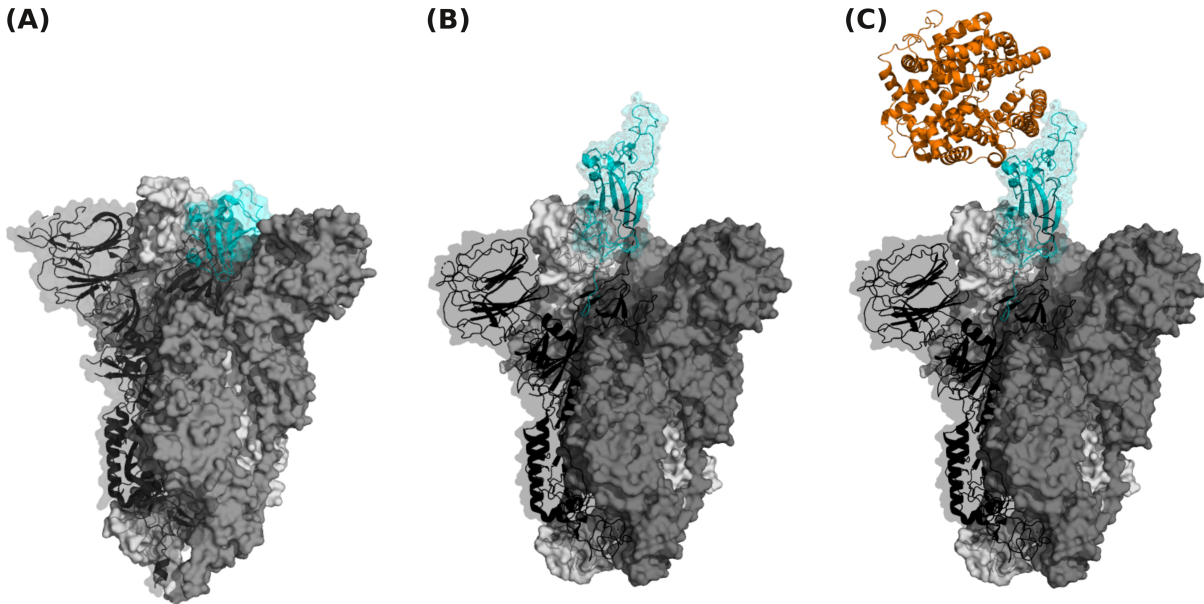
Figure 1.2: SARS-CoV-2 spike conformation to interact with ACE2. SARS-CoV-2 spike is depicted as grey surface representation with one protomer represented as cartoon and the corresponding RBD is highlighted in cyan. The spike is shown with all three RBDs **(A)** in down conformation (pdb code: 7ddd), and **(B)** one RBD in up conformation (pdb code: 6ack) **(C)** to interact with the ACE2 receptor (orange cartoon) (pdb code: 6ack).

Viral Life Cycle

The life cycle of SARS-CoV-2 starts with the attachment of the virus to a host cell. This is mainly achieved by interactions of the spike protein and the ACE2 receptor (figure 1.2) [50]. Additional interactions of the spike protein with heparan sulphate can further promote viral attachment [51]. After spike-ACE2 complex formation, the S2 protein of the spike trimer is cleaved by the transmembrane serine protease 2 (TMPRSS2) of the host cell. Cleaving off the spike initiates a conformational change, which enables the fusion of the host and viral membrane and the release of the RNA genome into cytoplasm [52, 45]. In the host cell, the $+$ sense viral genome is unpacked from the nucleocapsid by cellular proteases before a $-$ sense genome is transcribed, serving as template for further genome replication [53]. Newly synthesised $+$ sense RNA genome replica function either as templates for translation of viral proteins or are packed into new virions. In the first step of translation, two variants of polyproteins are produced and auto-proteolytic cleaved into nonstructural proteins [54]. These nonstructural proteins fulfil important functions, including RNA synthesis and processing for virus production, but also act as inhibitors for host cell immune response and translation of host proteins [55, 56]. In the following,

all structural proteins are produced in the endoplasmic reticulum (ER) of the host cell, except for the nucleocapsid protein, which is synthesised by free ribosomes [52]. New virus particles are assembled at the ER-Golgi intermediate compartment (ERGIC). The viral genome is condensed into the nucleocapsid and virus assembly is organised by the viral membrane protein by interacting with other structural proteins such as the spike protein [57]. The membrane protein also binds the nucleocapsid protein to recruit and stabilise the nucleocapsid protein-RNA complex [52]. Finally, membrane curvature, to form a viral particle, is introduced by membrane and envelope protein interactions. New virus particles are transported in large vesicles via the secretory pathways to the host cell membrane for exocytosis [58].

SARS-CoV-2 Variants

Table 1.1: SARS-CoV-2 variants. All variants, which were declared as variants of concern by the World Health Organisation (WHO) (date: 18.01.2021), are listed in the following [59]. (*) mark analysed antibodies, which were isolated from vaccinated individuals or patients that have been previously infected with another variant.

Name	Lineage	Variation in RBD of spike	Origin	Properties	Reference
Alpha	B.1.1.7	N501Y	United Kingdom	enhanced binding to ACE2	[60, 61]
Beta	B.1.351	K417N, E484K, N501Y	South Africa	enhanced binding to ACE2, partly antibody escape*	[62]
Gamma	P.1	K417T, E484K, N501Y	Brazil	enhanced binding to ACE2, partly antibody escape*	[63]
Delta	B.1.617.2	L452R, T478K	India	enhanced membrane fusion, partly antibody escape*	[64, 65]
Lambda	C.37	L452Q, F490S	Peru	antibody escape	[66]
Mu	B.1.621	R346K, E484K, N501Y	Columbia	enhanced binding to ACE2, partly antibody escape*	[67]
Omicron	B.1.1.529	G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y	Botswana	frequently observed antibody escape*	[68]

SARS-CoV-2, like other RNA viruses, possesses a high mutation rate [69]. Table 1.1 lists all variants of concern according to the World Health Organisation (WHO), with their mutations in the RBD and their changed properties. For all of these variants the mutations are not limited to the RBD, but are also present in other parts of the spike protein and other viral proteins. Each mutation contributes to variant survival, by either introducing structural rearrangements in the spike protein for a more efficient membrane fusion or enhancing ACE2 interaction. These variations increase infectivity and transmission, or enable a more efficient reproduction of the virus in the host cell [65]. The first described variant of SARS-CoV-2 comprises a point mutation (D614G) in the spike protein, causing one protomer per spike protein to be in up-conformation [70]. With this single point mutation, the virus variant shows an enhanced infectivity but also an enhanced neutralisation sensitivity and a higher transmissive phenotype *in vivo* [71, 72]. A vast number of variants are emerged so far, but only variations in the spike protein of a few examples will be discussed. A fast spreading and dominating variant was first identified in the United Kingdom and systematically named B.1.1.7, also called alpha-variant [60].

This SARS-CoV-2 type combines several mutations in the spike protein with one mutation being located in the RBD (N501Y) improving ACE2 interaction [60]. On the other hand, modification in the spike, especially in the RBD domain, reduces antibody binding. For the beta- and gamma-variant, viral escape from neutralising antibodies of vaccinated or postinfectional people are documented [73]. Transmission of the SARS-CoV-2 delta-variant is increased due to an enhanced membrane fusion potential. In addition, mutations in the spike protein can help the variant to evade the immune system of even recovered or vaccinated patients [74].

1.1.3 Immunoglobulins and Nanobodies

Immunoglobulins (Ig) or antibodies are proteins produced by the immune system to bind specifically against potential pathogenic structures, called antigens. The specificity of antibodies for a particular binding partner is often utilised in the field of diagnostics and imaging, but can also be used in biosensors for medicine and research [75, 76, 77]. Within the different Ig classes (IgA, IgG, IgM etc.), Ig type G (IgG) is the variant most commonly used as a biosensor. In the body, IgGs are mainly involved in the immune response of the adaptive immune system and can neutralise pathogens, recruit macrophages for phagocytosis, or activate other parts of the immune system such as the complementary system [78]. Additionally, camelids produce a special type of antibodies during the immune response, the so-called heavy-chain-only antibodies. This antibody variant is of great interest, as the antigen binding epitope can be expressed in *Escherichia coli* (*E. coli*) as a stable, single domain protein, the so-called nanobody (Nb) [79].

Production of Nanobodies

For the production of Nbs against a specific antigen, llamas or alpacas are immunised with the purified antigen, for example a protein. RNAs, coding for the antigen binding site of heavy-chain-only antibodies, are isolated from peripheral blood lymphocytes of the immunised animal and integrated as complementary DNA (cDNA) into the genome of phages or yeast to create a so-called phagmid library. Nbs against new antigens, or non-immunogenic targets such as RNA or DNA, can also be obtained by *in-silico* maturation of Nbs from either immune, naive or semisynthetic/synthetic Nb gene libraries [80]. Diverse sets of Nbs are obtained by complementary determining region (CDR) randomisation and error prone polymerase chain reaction (PCR) is the most popular method to randomly introduce mutations into the genetic information [81, 82]. From these libraries, high affinity binders are obtained by several rounds of phage- or yeast-display against an antigen [83].

Architecture of Antibodies and Nanobodies

IgG antibodies are large hetero complexes of around 150 kDa, composed of two heavy and two light chains. Each chain consists of several Ig-fold domains built up by two antiparallel β-sheets forming a β-sandwich [84]. The heavy chain is built up by three constant regions (CH1, CH2, CH3) and one variable region (VH), the light chain is smaller,
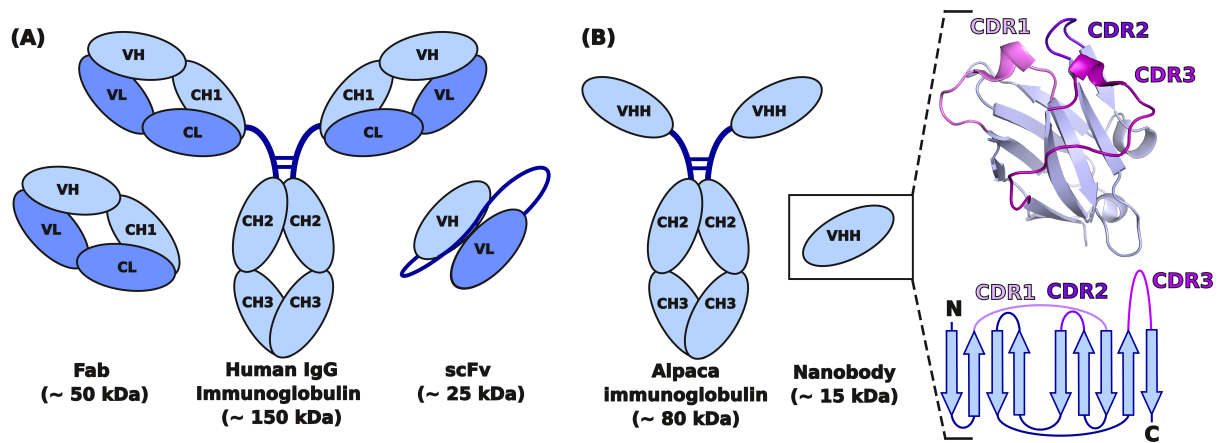
Figure 1.3: Representation of a human-derived IgG antibody, as well as a camelid-derived heavy-chain-only antibody and antigen-binding fragments thereof. (A) Schematic representation of human IgG antibody, fragment antigen binding (Fab), as well as single-chain fragment variable (scFv). For comparison, **(B)** a heavy chain only antibody, isolated from alpaca camelids, and a nanobody (Nb) are shown. Additionally, a Nb structure (pdb code: 5ivo) is represented as cartoon representation, as well as topology plot, with complementary determining region 1 (CDR1), CDR2 and CDR3 labelled and highlighted in light pink, purple and magenta, respectively. Heavy and light chains of the antibodies are coloured in dark and light blue, respectively. Disulphide bridges between the heavy chains are indicated by dark blue lines. The designation and molecular weight of the different antigen binding fragments is specified below.

with one constant (CL) and one variable (VL) region. IgG antibodies adopt a Y-shape with two arms [85]. Each arm is composed of a heterodimer of one heavy and one light chain, which is stabilised by a disulphide bridge (figure 1.3). Additionally, the two heavy chains dimerise via its CH2 and CH3 domains, again stabilised by two disulphide bridges. Each variable domain, VH and VL, of the heterodimeric antibody arm contains three hypervariable regions, CDR1, CDR2 and CDR3 [85]. The six CDRs in each arm of the antibody form the antigen binding site and determine antigen specificity. The two CH2 and CH3 domains, with a *N*-linked glycan present in each CH2 domain, build the conserved Fc-fragment of the antibody [86]. This Fc-fragment with its glycosylation is responsible for biological functions of the antibody, such as activation of the complement system, initiating phagocytosis or cell cytotoxicity during the immune response [85, 86].

For the application of antibodies as biosensors or for therapeutical usage, smaller antigen-binding fragments are favoured. One variant is the antigen binding fragment (Fab), which is derived from papain treated antibodies. Fabs are composed of VH and VL domains of the light and CH1 and CL domains of the heavy chain [87]. In addition, synthetically produced single-chain fragment variable (scFv) antibodies comprise only one VH and one VL domain, which are fused via a peptide linker, to form the minimal motif for antigen-recognition [88]. IgG, Fabs, as well as scFv, comprise an antigen binding epitope composed of two chains. Binding specificity is mainly determined by the six CDR regions of VH and VL and recognises predominantly flat epitopes [89].

In contrast, camelids produce smaller heavy-chain-only immunoglobulins, built up by two heavy chains forming a disulphide bridge stabilised homodimer with a size of around 80 kDa. Each chain is composed of three Ig-fold domains, with two constant regions (CH1 and CH2), as well as a variable region (VHH) (figure 1.3). The 15 kDa VHH-domain can be produced as single-domain Nb, which consists of two antiparallel β-sheets forming a β-sandwich. The antigen binding loops CDR1, CDR2 and CDR3 are located between the frame work regions (FR), which consists of β-strands. From the N-terminus, CDR1 is located between FR1 and FR2, which are formed by the first and second β-strands, respectively. In accordance, CDR2 is located between FR2 and FR3, comprising three β-strands, followed by CDR3 and a single β-strand of FR4 (figure 1.3) [90]. The antigen binding epitope of a Nb is composed of one VHH-domain with three CDR regions. Thereby, binding specificity is not exclusively determined by the CDRs in the VHH-domain, but also FR can play an important role in antigen recognition [91]. In contrast to antibodies, Nbs can bind in grooves or cavities [89].

Properties and Application

Antigen binding fragments such as Fab, scFv and Nb have different applications according to their properties. Nb, scFv and Fab can be cost-efficiently produced in *E. coli*, whereas IgG are mainly obtained from expensive mammalian cell culture [88, 92, 93]. In addition, small size Nb (~15 kDa) and scFv (~25 kDa) show an enhanced blood clearance and a deep tissue penetration, which makes them more beneficial compared to large antibodies (~150 kDa) or Fabs (~50 kDa) [94]. Nbs are beneficial for applications in different environments, especially in comparison to scFv, as they are often stable within broad pH ranges, in the presence of detergents or at high temperatures up to 90 °C and possess a long self-life time [95, 96].

1.2 Objective and Significance

In the current state of the COVID-19 pandemic, there is no efficient treatment or vaccine that offers a full protection against SARS-CoV-2 infections [97]. A recent study determined a significantly decreasing humoral immune response against SARS-CoV-2 over a period of six month after receiving a second vaccination dose [98]. As the protection against SARS-CoV-2 and the amount of neutralising antibodies correlate, there is not only a need for long-acting vaccination, but also for a diagnostic tool that monitors the immune response of recovered or vaccinated individuals to estimate their protection against SARS-CoV-2 [99]. In addition, new SARS-CoV-2 variants cause additionally danger for the success of the current vaccination campaign. To be able to control SARS-CoV-2 spreading, a more detail analysis of the immune status of recovered and vaccinated patients is needed to determine a time point at which the vaccination protection has to be renewed.

Alpaca-derived nanobodies against the RBD of SARS-CoV-2 were produced and biochemically characterised by our cooperation partner. In this work, structures of two Nbs, which both possess a high affinity and neutralisation capacity, were analysed in complex with the RBD. The obtained structures were aligned with the spike protein structure to reveal the neutralisation mechanism for each Nb. The structural findings were used by our cooperation partner in order to develop an efficient tool to monitor immune response of recovered patients against SARS-CoV-2 and variants thereof. This tool could help to monitor immune status of vaccinated and recovered individuals, as well their resistance of emerging SARS-CoV-2 variants over time.

1.3 Results and Discussion

In this section, the published results of SARS-CoV-2 specific nanobodies that can be used as potential new diagnostic tools to monitor the immune response of infected or vaccinated individuals are summarised. For details and references please consult appendix section A.2, which contains the publication as full text.

NeutrobodyPlex-Monitoring SARS-CoV-2 Neutralizing Immune Responses Using Nanobodies

Teresa R Wagner[*] (TRW), Elena Ostertag[*] (EO), Philipp D Kaiser (PDK), Marius Gramlich (MG), Natalia Ruetalo (NR), Daniel Junker (DJ), Julia Haering (JH), Bjoern Traenkle (BT), Matthias Becker (MB), Alex Dulovic (AD), Helen Schweizer (HS), Stefan Nueske (SN), Armin Scholz (AS), Anne Zeck (AZ), Katja Schenke-Layland (KSL), Annika Nelde (AN), Monika Strengert (MoS), Juliane S Walz (JSW), Georg Zocher (GZ), Thilo Stehle (TS), Michael Schindler (MiS), Nicole Schneiderhan-Marra (NSM) and Ulrich Rothbauer (UR)
EMBO Reports, 2021, 22:e52325.

[*] contributed equally.

Contribution

The study was designed by NSM, TRW, MB and UR and supervised by NSM and UR. Alpacas were immunised by HS, SN and AS for the production of Nbs, which were selected and biochemically characterised by PDK, BT and TRW. The multiplex binding assay was performed by JH, DJ and MB. HDX-MS experiments were performed and analysed by MG and AZ. For the functional analysis, patient samples were organised and provided by MoS, AN, JSW and KSL and the virus neutralisation assays was performed by NR and MiS. Data analysis and statistical analysis was carried out by TRW, MB, JH, MG, AZ, NR, MiS and UR. Structural studies of Nbs-RBD complexes was designed and performed by EO, GZ and TS. The manuscript draft was written by TRW, AD and UR, proof-read by all authors.

1.3.1 Biochemical Characterisation

Specific Nbs against SARS-CoV-2 were obtained by immunisation of an alpaca with purified SARS-CoV-2 RBD. 325 positive binders, out of 4×10^7 clones of the Nb phagemid library, were selected by two rounds of phage display and solid-phase phage ELISA against SARS-CoV-2 RBD. Sequencing of 72 clones revealed 10 unique Nb with highly diverse CDR3 regions and high affinity to the RBD with dissociation constants (K_D) of 1.4 to 53 nM, determined by biolayer-interferometry (BLI). Neutralisation capacity against SARS-CoV-2 infection was investigated with an ACE2 competition assay and an *in-vitro*

viral neutralisation test. Three Nbs (NM1226, NM1228 and NM1230) showed a high neutralisation capacity with half maximal inhibitory concentration (IC_{50}) values in the picomolar range for the ACE2 competition experiment and IC_{50} values in the nanomolar range in a viral neutralisation test. Binding epitopes of all Nbs were characterised by hydrogen exchange mass spectrometry (HDX-MS) and competitive BLI analysis. Four sets of Nb could be distinguished, while one Nb of each set can bind simultaneously with a Nb of another set.

1.3.2 Structural Investigation

NM1226 of set 1 and NM1230 of set 2 possess both high affinities to the RBD with K_D values of 3.7 nM and 8.2 nM, as well as a strong neutralisation potency with IC_{50} values of 15 nM and 37 nM, respectively. For a more detailed analysis of the binding properties and neutralisation ability, NM1226 and NM1230 were structurally investigated. A structure was obtained by X-ray crystallography for each Nb in complex with RBD at 2.3 Å and 2.9 Å resolution for NM1226 and NM1230, respectively. Comparison of both complexes revealed two distinct, non-overlapping binding epitopes on the RBD. The NM1226-RBD complex buries an interaction area of around 700Å^2, whereas major contacts of NM1226 to the RBD are formed via its CDR3 and CDR1. Beside one salt-bridge of D101$_{NM1226}$ to R408$_{RBD}$ and K378$_{RBD}$, various hydrogen bonds of side chains and backbone atoms in the CDR3 of NM1226 determine the binding to the RBD. The NM1230-RBD complex possesses a binding interface of approximately 830Å^2. Interactions of the NM1230 to the RBD are formed by CDR3, FR1, FR2 and FR3, while most interactions are contributed by CDR3, which binds in a groove on the RBD surface. The binding of NM1230 is determined by mainly polar interactions, but also one salt bridge of R98$_{NM1230}$ to E484$_{RBD}$ and π-π stacking interactions (W103$_{NM1230}$ to F490$_{RBD}$ and W100d$_{NM1230}$ to R489$_{RBD}$) contribute to the specificity towards the RBD of SARS-CoV-2.

1.3.3 Neutralising Effect and Viral Escape

For investigation of the neutralisation potency, both complexes were aligned with two cryo-EM structures of a spike trimer (pdb codes: 7kms, 6zxn) and a RBD bound to ACE2 (pdb code: 6m17). The comparison revealed different mechanisms how NM1226 and NM1230 can block the interaction of ACE2 with the RBD. NM1226 can only bind the RBD of a spike monomers in the up-conformation. The binding interface of NM1226 and ACE2 does not overlap, instead NM1226 blocks ACE2 binding by steric hindrance. The binding epitope of NM1226 on the RBD is less prone to mutation and this area has a higher sequence conservation in SARS-CoV-2 and SARS-CoV-1 than other Nb binding epitopes on the RBD [100]. Up to now, NM1226 showed a robust binding to all variants of concern, declared by the WHO, except omicron, which possesses mutations within this area (table 1.1). In contrast to NM1226, NM1230 is able to bind the RBD of a spike protein in up- as well as down-conformation and possesses a partly overlapping binding

epitope with ACE2. The overlap is limited to six residues in the binding site, which seems to be sufficient to block ACE2 binding on the virus. In addition to competition for the same binding epitope, NM1230, bound to RBD of one spike monomer, can block the ACE2 interaction at the neighbouring protomer by steric hindrance. In contrast to NM1226, the binding epitope of NM1230 is less conserved and each variant of concern possess at least one mutation in the NM1230 binding site (table 1.1).

For a stronger and more robust binding to the RBD of SARS-CoV-2 and variants thereof, NM1230 was fused with a glycine-serine-linker, composed of 20 amino acids, to NM1226 in a head-to-tail fashion to produce a biparatopic Nb (biNb), called NM1267. This linker enables the biNb to bridge the distance between RBDs of two neighbouring spike protomers of the same receptor. The new biNb show enhanced affinity to RBD for the unmodified SARS-CoV-2 and alpha-variant with K_D values in the picomolar range and low nanomolar range for SARS-CoV-2 beta-variant. Not only the affinity is improved, NM1267 possesses also competition ability to ACE2 and neutralisation potency with IC_{50} in the picomolar range for both assays.

1.3.4 Application

Several approaches are available to monitor the total amount of antibodies against SARS-CoV-2. However, most of these assays do not distinguish between total amount of spike protein binding antibodies and RBD binding antibodies with ACE2 inhibitory, and therefore neutralising ability. A new assay was developed to differentiate between the total amount of antibodies and neutralising proportion in patients, which could be also applied for vaccinated persons. For this, beads, coated with either RBD, S1 protein or spike protein, were pre-incubated with patient-derived antibodies. In the following, the binding of antibodies to RBD, S1 and spike was analysed by addition of NM1267, which supersede neutralising antibodies that possess an overlapping binding epitope on the RBD. In the sample with RBD-bound antibodies, a complete displacement could be detected. In contrast, antibodies, bound to spike-coated beads, were only partly displaced by the addition of NM1267. Based on this data, the proportion of RBD binding and potentially neutralising antibodies can be estimated in the total amount of SARS-CoV-2 specific antibodies.

1.4 Conclusion and Outlook

Neither the current available vaccines nor recovery from an infection offer a full and lifelong protection against SARS-CoV-2 and variants thereof. The assay, developed in this work, could be applied to monitor the antibody amount and composition of individuals after vaccination. The thereby obtained data would help to determine the time point at which an additional vaccination has to be delivered to refresh immune protection.

In addition, this assay can be adopted to new SARS-CoV-2 variants. For new emerging SARS-CoV-2 variants, the binding ability of NM1267 has to be investigated, before this assay can be used to diagnose protection of recovered and vaccinated individuals against new variants of SARS-CoV-2. In case of emerging SARS-CoV-2 variants that are not bound by NM1267, the existing SARS-CoV-2 phagemid library could be screened to identify Nbs that bind specifically or error prone PCR can be performed to obtain new Nbs that have the ability to bind. These newly obtained Nbs could then be used in the already established assay to evaluate the adaptive immune response of patients against SARS-CoV-2 variants.

2 Reprogramming Substrate and Catalytic Promiscuity of Tryptophan Prenyltransferases

2.1 Introduction

2.1.1 Secondary Metabolites

Secondary metabolites are small molecules with high chemical diversity such as polyketides, terpenes, quinones, non-ribosomal peptides, alkaloids and polyphenols (with flavonoids as a subgroup) [101, 102, 103, 104, 105, 106, 107]. In contrast to primary metabolites, which are crucial for a vital organism by recycling or synthesising amino acids for protein production or producing energy in glucose or fat metabolism, secondary metabolites are not essential for living. Instead, secondary metabolites help to protect the organism against pathogens and natural enemies, but their production is often accompanied by other advantages for the organism [108]. Secondary metabolites of plants, bacteria and fungi are chemical highly diverse and often possess pharmacological impact such as antifungal, anti-inflammatory, antibacterial and antiviral properties, as well as cytotoxic activity (figure 2.1). These compounds can be utilised for drug development, medical treatment and agriculture [109, 110]. A famous example is penicillin, which was isolated from *Penicillium* and combats bacteria to give the fungi a selection advantage for nutrient supply in the same environment. Today, this natural product is still used as antibiotic to treat bacterial infections in humans [111].

Application and Usage of Prenylated Secondary Metabolites

Some of these natural occurring secondary metabolites are prenylated, which not only increases the number of compounds in nature, but also enhances the biological activity by an increased lipophilic character that enables target binding (figure 2.1) [112]. For example, polyphenols that are commonly found in plants possess natural activity as antioxidants to reduce inflammatory reactions [113]. The number of plant families that possess prenylated polyphenols is limited, but plants that produce these compounds were often used in medicine. *Glycyrrhizae radix* is used in Chinese traditional medicine because of the constituent glycyrol (coumarin derivate) that possess anti-inflammatory properties, as well as antibacterial activity against *Streptococcus pyogenes*, *Haemophilus influenzae* and *Moraxella catarrhalis* [114]. Hop (*Humulus lupus* L.) is a well known natural alternative to hormone therapy for elderly women to treat menopausal symptoms [115]. There, several prenylated flavonoids like Xantohumol, 8-prenylnaringenin (8-PN) or 6-prenylnaringenin (6-PN) were found to act as phytoestrogen and can compensate for the
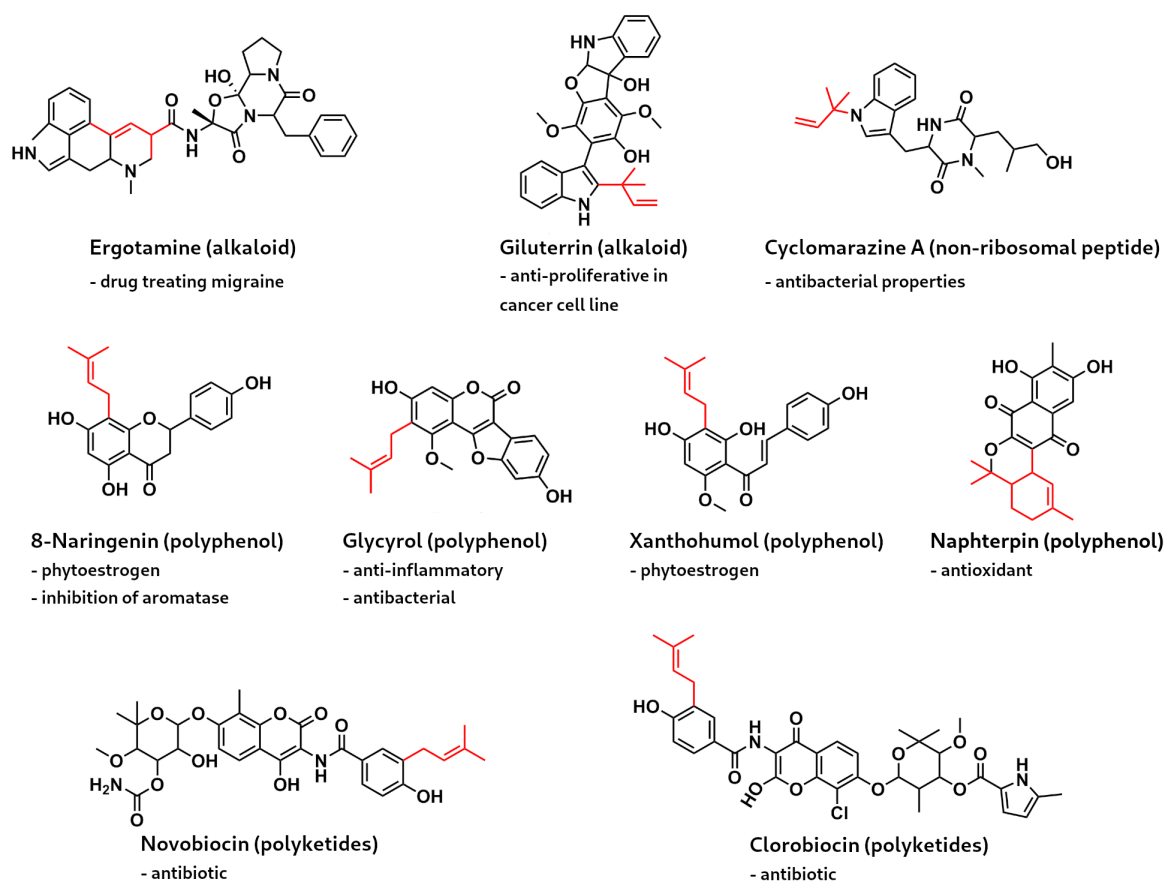
Figure 2.1: Selection of Secondary metabolites. Secondary metabolites with prenylation as part of the biosynthetic pathway and their pharmacological impact are represented. The integrated and in some cases modified prenyl moieties are highlighted in red.

lack of hormones. In addition, 8-PN inhibits aromatase activity, a risk factor for breast cancer after the menopause and has potential application to prevent osteoporosis and promote collagen synthesis in skin [116, 117, 118]. For 5-prenylisoflavanones, isolated from *Geoffroea decorticans*, antifungal activity was determined against a number of *Aspergillus* species with a potential application as biopesticides [119].

In addition, beneficial agents are not only found in plants, also components produced in bacteria and fungi are interesting candidates for medical applications. For example, Giluterrin is a prenylated indole alkaloid isolated from *Aspergillus terreus* P63, which has anti-proliferative activity on prostate and kidney cancer cell lines [120]. The antibiotic Novobiocin is produced by the actinobacterium *Streptomyces niveus* and is primarily active against gram-negative bacteria and an effective agent against methicillin-resistant *Staphylococcus aureus* [121]. Ergot alkaloids are a group of fungal secondary metabolites with a tetracyclic ergoline ring system, which makes them structurally similar to natural neurotransmitter such as noradrenaline, dopamine or serotonin [122]. Although, ergot alkaloids are normally not prenylated, prenylation of L-tryptophan (L-Trp) is the first step in the biosynthetic pathway [123, 124]. Ergonovine is an ergot alkaloid, which influences smooth muscle contractions in the uterus and is used as drug to treat heavy bleeding after giving birth [122]. Another example is Ergotamine, a useful drug to treat severe migraine

attacks [125]. Several prenylated alkaloids were identified from the marine-derived fungus *Aspergillus austroafricanus*, and some of these were determined to possess pro-angiogenic activity in zebra fish model experiments [126].

2.1.2 Chemical Synthesis and Limitations

Classical drug production is often a multistep synthesis, where prenylation of aromatic compounds is difficult. One problem are the unfavourable reaction conditions to activate the prenyl-donor and -acceptor. Often both substrates are coupled to a bromide or metal salt and prenylation reaction is carried out with palladium or ruthenium as catalyst, often at acidic pH, under anaerobic conditions and high or very low temperatures [127, 128, 129]. Another problem is the low regio- and chemoselectivity of the prenylation reaction, as in a Friedel-Craft alkylation the carbo-cation attacks atoms according to their electronegativity, predominantly position N-1 and C-3 (figure 2.2). Thereby, it was shown that N-1 is preferably prenylated in acidic milieu [128]. For prenylation of any other position, different strategies exists. On the one hand, electronegative groups are protected to shift the prenylation onto another position or an initial prenylation is carried out onto position N-1 or C-3, followed by different rearrangement reactions [130]. Furthermore, the carbon at position C-2 can be attacked selectively in a reaction with electron-deficient radicals [131]. Prenylation of positions at the benzene ring are challenging, especially position C-5 and C-6. Up to now, only small scale synthesis can be carried out [128].

In contrast to chemical synthesis with low specificity and yields, enzyme-catalysed prenylation of aromatic ring compounds are highly regio- and chemoselective, and can turnover up to 100% of the substrate [132]. This makes the enzymes an interesting alternative for a more efficient route to biologically active compounds.
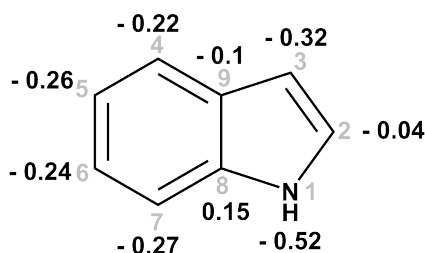
Figure 2.2: Electronegativity of each position in the indole ring. The number and electronegativity of each atom in the indole ring is represented in grey and black, respectively. The picture was adopted from Li et. al [131].

2.1.3 Prenyltransferases

As briefly described above, prenylation is often the initial step of larger biosynthetic pathways, which is carried out by enzymes encoded in a gene cluster in fungi, plants or bacteria [133]. Prenylation of aromatic natural products is carried out by prenyltransferases (PT), which are classified in different families; the membrane-associated, the so-called UbiA-type, and the soluble prenyltransferases. UbiA-type prenyltransferases have a broad

substrate repertoire including dolichols, ubiqinones, hemes, plant chlorophylls, vitamin E and K, structural lipids, as well as fungal merotropenoids [134]. The catalytic activity of these enzymes is dependent on divalent cations, bound by a conserved aspartate-rich motif (N/DXXDXXXD and DXXXD) [135].

In contrast, soluble prenyltranferases lacking an aspartate-rich motif and are metal-ion independent, with exceptions as some enzymes show an enhanced turnover in presence of Mg^{2+} or Ca^{2+} [136]. A collection of soluble prenyltranferases were structurally characterised and named ABBA-fold prenyltransferases according to their conserved three-dimensional overall fold build up by five repetitive secondary structure elements of two α-helices followed by two β-strands (or $α − β − β − α$) [137] (figure 2.3). The active site is located in the enzyme core, a β-barrel built up by ten anti-parallel β-strands, which is surrounded by solvent-exposed α-helices (figure 2.3). Based on sequence identity, the ABBA-fold prenyltransferases can be subdivided in two families; the CloQ/NphB group and dimethylallyl tryptophan synthase (DMATS) family [138]. The CloQ/NphB family is named after the first identified members CloQ and NphB. CloQ was characterised to catalyse the prenylation of 4-hydroxyphenylpyruvic acid in the biosynthesis of the antibiotic Clorobiocin, whereas NphB, found in *Streptomyces*, is part of the biosynthetic cluster of Naphterpin [139, 140]. Members of this enzyme family can be mainly found in bacteria and plants [105].

2.1.4 Dimethylallyl Tryptophan Synthases

Members of the DMATS family can be found in fungi as well as bacteria. In bacteria, the enzymes characterised so far are monomeric in solution. In contrast, fungi-derived DMATS often dimerise or tetramerise due to an additional α-helix at the N-terminus of the enzyme [124] (figure 2.3). Over the last two decades, several fungal and bacterial DMATS were biochemically and structurally characterised and the family was named after the preference to use dimethylallyl entities (C5) during catalysis [132]. DMATS have a higher catalytic activity for dimethylallyl entity and are also selective for their prenyl-acceptor substrate, which is limited to a distinct number of aromatic compounds like tyrosine, tryptophan or other indole ring containing aromatic compounds including cyclic dipeptides [134]. Single enzymes are named according to the associated product or biosynthetic pathway, from which they were isolated (e.g. SirD is part of the sirodesmin PL biosynthetic pathway). If the enzyme is not directly associated with a natural occurring secondary metabolite, the enzyme is named after the prenylation position in the indole ring (e.g. 5-DMATS attaches a dimethylallyl moiety onto position C-5 of the indole ring) [141, 142].

For L-Trp and derivatives thereof, enzymes for each position on the indole ring were biochemically, as well as structurally characterised except for prenylation of position C-5 (figure 2.5). Fungal- and bacterial-derived 5-DMATS were biochemical characterised and the first structure of a bacterial 5-DMATS was determined in the course of this work

(A)

(B)
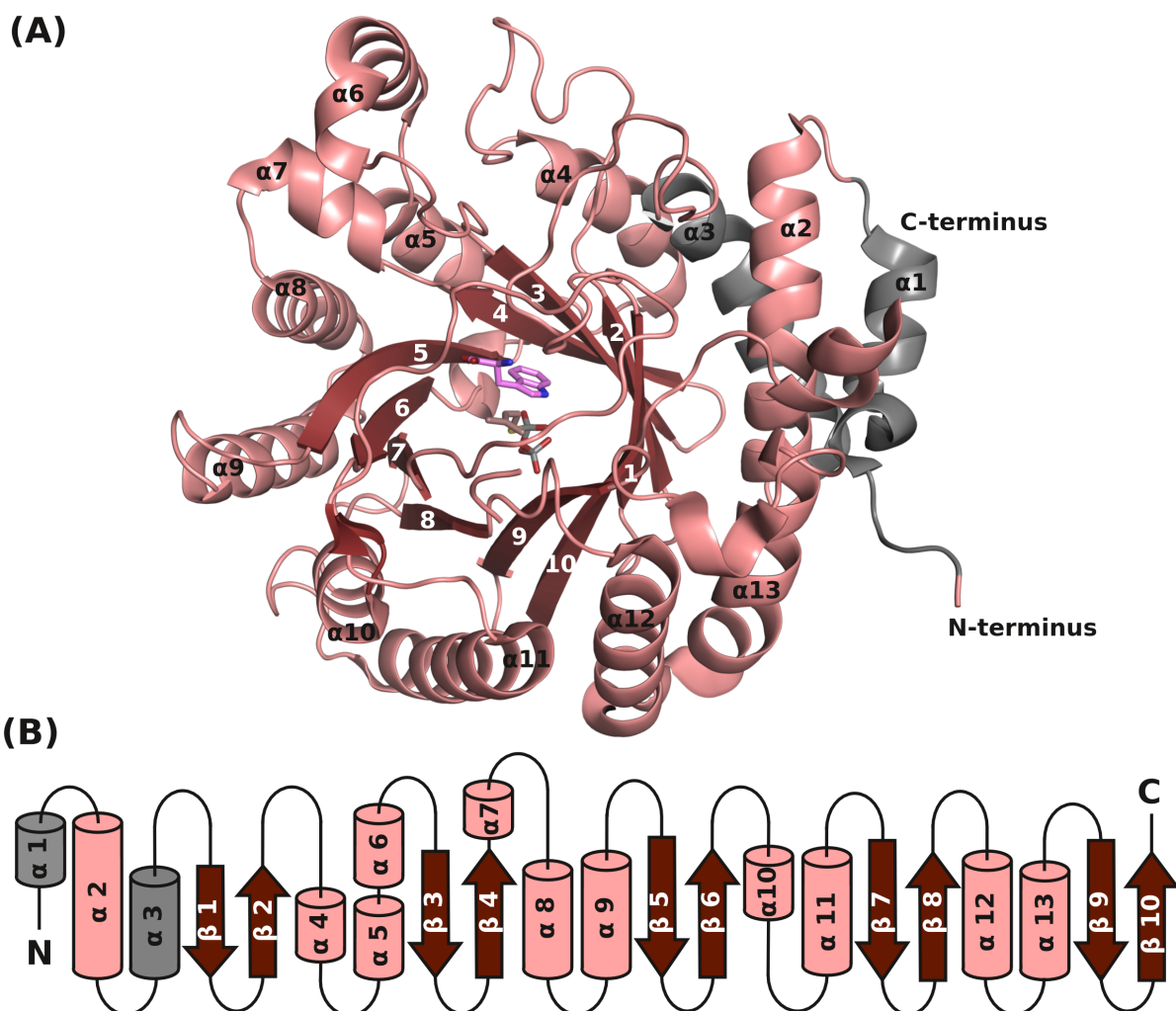
Figure 2.3: ABBA-fold prenyltransferase. (A) The crystal structure of FgaPT2 (pdb code: 3I4x) as representative of ABBA barrel-fold enzymes with a β-barrel build up by anti-parallel β-sheets shielded by solvent-exposed helices are represented in dark red and salmon, respectively [124]. The dimerising helices, present in the structure of fungal ABBA-fold prenyltransferases, are highlighted in grey. **(B)** The topology plots represents the corresponding repetitions of $\alpha, \alpha, \beta, \beta$ secondary structure elements.

[142]. Fungal-derived examples of DMATS that prenylate position N-1 to C-4 of the indole ring were structurally analysed. Thereof, CymD was found in *Salinispora arenicola* and prenylates the N-1 atom of the indole ring as part of the biosynthesis of anti-inflammatory Cyclomarine A and antibacterial component Cyclomarazine A [143]. The biosynthetic cluster of Tryptostatin B (cyclo-L-Trp-L-Pro) of *Aspergillus fumingatus* contains the PT FtmPT1, a C-2 prenylating enzyme [144]. Acteylaszonalenin, a mycotoxin isolated from *Neosartorya fischeri*, is prenylated by AnaPT at position C-3 of the indole ring and FgaPT2 was identified to transfer the dimethylallyl moiety onto the C-4 position of L-Trp as initial step of ergot alkaloid biosynthesis [124, 145]. In the literature, only structures of bacterial-derived DMATS are available for prenylation of position C-6 to C-7 of the indole ring. PriB from *Streptomyces* sp. RM-5-8 transfers dimethylallyl entities onto position C-6 and TleC, identified in *Streptomyces blastmyceticus*, acts as a 7-DMATS during Indolactam V synthesis [146, 147]. In addition, biochemical data are available for a 7-DMATS from *Aspergillus fumigatus* [148].

19

Substrate Binding

Prenyl entities of different sizes (n×C5) origin from the isoprenoid pathway as activated form attached to a pyrophosphate group e.g. C5: dimethylallyl pyrophosphate (DMAPP), C10: geranyl pyrophosphate (GPP), C15: farnesyl pyrophosphate (FPP) and so on [149]. The prenyl-acceptor is an aromatic compound that originates either from shikimate or polyketide pathway [150]. During catalysis, both substrates are bound in close proximity in the active site (figure 2.3). The prenyl-acceptor site is not conserved with the exception of a strictly conserved glutamate residue, which interacts with the nitrogen of the indole ring (figure 2.4) [124]. All other residues are variable, shaping the binding pocket to fit the site of the natural substrate. The prenyl-donor binding is conserved. The activating pyrophosphate group is hold in place by ionic interactions to positively charged residues (Arg, Lys), as well as hydrogen bonds, which are established by conserved tyrosines [124, 147]. The prenyl entity is located in a hydrophobic binding cavity and forms π-stacking interactions with the delocalised electrons in the aromatic ring system of the prenyl-acceptor (figure 2.4). These binding properties are conserved for bacterial- and fungal-derived DMATS, and the prenyl-donor cavity determines the prenyl entity size that can be transferred onto the aromatic ring system. For example, MpnD and TleC prenylate both the C-7 position of the indole ring of Indolactam V. In contrast to TleC, MpnD possesses a small prenyl-donor binding site and prefers activated C5-isoprenoid units (DMAPP) as substrate, whereas TleC showed a higher catalytic activity for the larger isoprenoid unit GPP [147].
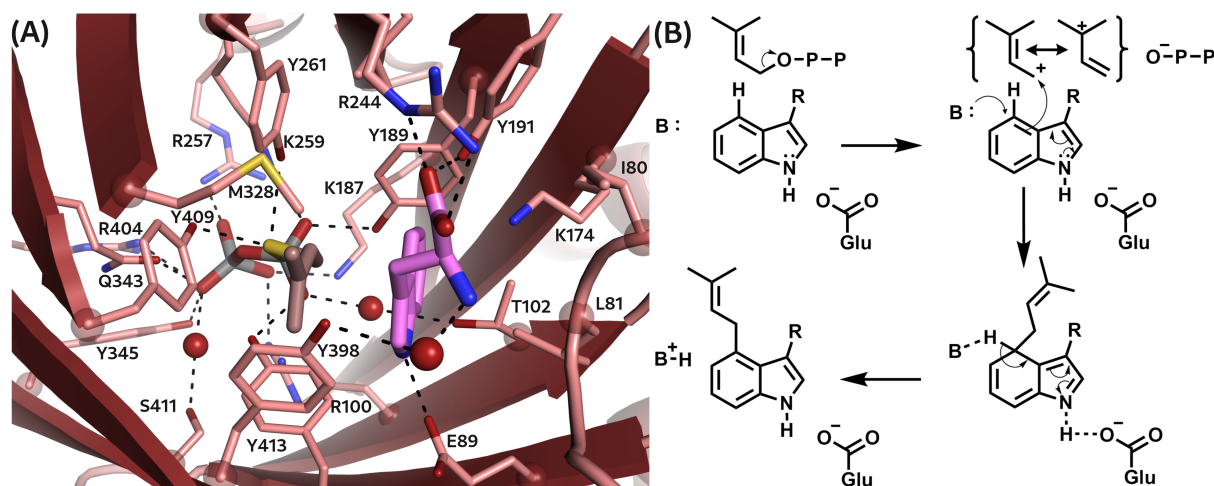
Figure 2.4: DMATS active site and reaction mechanism. (A) Binding of L-Trp (light pink) and DMASPP (brown), including all interacting residues. Amino acids and compounds are represented as sticks in the active site of FgaPT2 (pdb code: 3I4x) [124]. Hydrogen bonds and salt bridges are indicated as black dashed lines and coordinated waters are shown as red spheres. **(B)** The Friedel-Crafts alkylation reaction mechanism, proposed for DMATS, is represented with a basic residue (B) as catalytic base and a strictly conserved glutamate to enhance the catalytic efficiency.

The pyrophosphate group of the prenyl-donor and the amino- and carboxy-group of the prenyl-acceptor are well aligned in the structural alignment of the active sites of all available structures. In contrast, the prenyl entity of the prenyl-donor and the aromatic

ring of the prenyl-acceptor are both rotated by 90° in the active site of some enzymes. Based on these observation, enzymes can be split into two groups: one group contains the N-1, C-2, C-3 and C-4 prenlyating enzymes CymD, FtmPT1, AnaPT and FgaPT2, respectively, which were all isolated from fungi. The other group consists of the C-6 and C-7 prenlyating enzymes PriB and MpnD, respectively, which both originate from bacteria (figure 2.5). This roses the question, if the orientation of prenyl entity and indole ring in the active centre is determined by the origin of the enzyme (bacterial- or fungal-derived) or is determined by the position, where the the prenyl moiety is attached to.
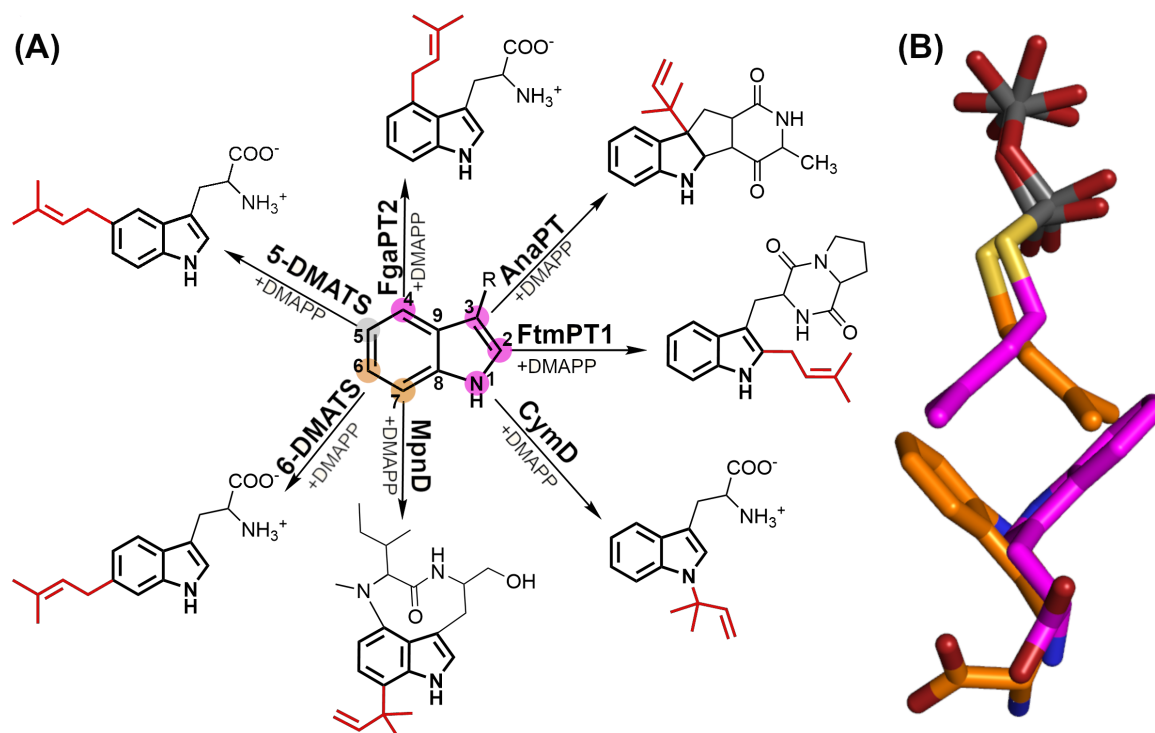
Figure 2.5: DMATS enzymes. (A) An indole ring including the numbering scheme is represented. Enzymes with their corresponding prenylation products and regioselectivity are listed. **(B)** The two distinct substrate binding modes for tryptophan and DMAPP, identified in the active site of structural characterised DMATS enzymes, are coloured according to the prenylation position in panel A on the basis of the substrate binding in FgaPT2 (orange) and PriB (magenta).

Reaction Mechanism

Prenyltransferases share a common catalysis mechanism to transfer prenyl moieties onto aromatic substrates. In comparison to the organic synthesis, the main advantage of the enzymatic reaction is the high regio- (determines the location of the prenylation) and chemoselectivity (determines the type of C-C bond, either secondary or quaternary). DMATS catalyse the transfer of a dimethylallyl moiety ($C5_n$) onto a defined position of the aromatic ring system in a Friedel-Crafts-like alkylation reaction: Here, a carbo-cation is formed by breaking the bond between pyrophosphate and C-1 of the prenyl-donor [151]. The nascent carbo-cation is protected against water or other solvent molecules by five conserved tyrosines, the so-called tyrosine shield [144]. The prenyl-acceptor, often a tryptophan derivative containing an indole ring, is activated by the strictly conserved

glutamate [144]. This glutamate binds to the hydrogen at the N-1 position to enhance the electronegativity of the indole ring. In the course of the reaction, the mesomeric stabilised carbo-cation can attack the aromatic ring system either via its C-1 (regular prenylation) or its C-3 (reverse prenylation) atom to form a sp^3-hybridised intermediate. For re-establishing the aromaticity of the ring system, a hydrogen is abstracted by the catalytic base [124] (figure 2.4).

The position that gets prenylated in the prenyl-acceptor substrate is dependent on: the electronegativity of the prenyl-acceptor atom, the distance to the carbo-cation, and the presence of a catalytic base to abstract the hydrogen and re-establish aromaticity of the ring system. For tyrosine prenylating enzymes such as SirD, the prenyl moiety is predominantly attached onto the oxygen of the hydroxyl group as a consequence of its electron negativity [152]. For FgaPT2, prenylation is carried out in regular manner onto position C-4 of the indole ring due to its proximity to C-1 of the prenyl moiety and a lysine residue that is located in close proximity to C-4 in the indole ring and acts as catalytic base [124]. For some enzymes, the distance of the carbo-cation at the C-1 or C-3 atoms (regular or reverse prenylation) to the prenylated position in the aromatic prenyl-acceptor substrate is not close enough for a direct Friedel-Craft alkylation reaction. In these cases, an initial Friedel-Crafts alkylation onto another position in the indole ring is proposed, followed by rearrangement reactions to shift the prenyl moiety to the final position in the indole ring, as shown for FtmPT1. The initial Friedel-Craft alkylation takes place as a reverse prenylation at position C-3 of the indole ring, followed by a Cope-rearrangement to form the product Brevianamide F, a regular C-2 prenylated compound [153].

2.1.5 Biotechnological Applications of Prenyltransferases

Some secondary metabolites, which are interesting as drugs, are still isolated from their host organisms. As already mentioned in section 2.1.1, prenylnaringenin (6-PN, 8-PN), produced in plants, possess properties interesting for medical usage. Production and isolation from plants is difficult, therefore a yeast system (*S cerevisiae*) was established to produce prenylnaringenin from its precursor L-phenylalanine. For the production of naringenin, genes for all six enzymes that are involved in the biosynthesis (phenylalanine ammonium lyase, *trans*-cinnamate 4-monooxygenase, cytochrome P450 reductase, 4-coumaric acid-CoA ligase, chalcone synthase, chalcone isomerase) were obtained from *Arabidopsis thaliana*. As plant prenyltransferases are poorly characterised, two fungal orthologous were used. 3-prenylnaringenin (3-PN) was obtained in presence of AnaPT *in vivo* and 8-PN, 6-PN and 3-PN could be produced by incubation with CdpNPT, an enzyme prenylating the N-1 position in the indole ring *in vitro* [154, 155].

For the production of aromatic compounds with new features, already existing enzymes can be used as tools in biotechnology to either catalyse prenylation of non-natural substrates or to create a new biosynthetic pathway by mixing enzymes of different gene clusters.

For example, 4,7-diprenylated indole derivatives were obtained by a tandem incubation with two prenyltransferases (FgaPT2 and 7-DMATS form *Aspergillus fumigatus*) *in-vitro* [156]. For biotechnological application, a non-ribosomal peptide synthetase (NRPS) from *Neosartorya fischeri*, which produces tryptophan-containing cyclic dipeptides, was successfully coexpressed with three aromatic prenyltransferases from different gene clusters of *Aspergillus nidulans* to obtain prenylated cyclic-dipeptide (L-Trp-L-Pro) with new properties [157]. In another approach, CdpNPT was used to prenylate Daptomycin, a last resort antibiotic. As more Daptomycin-resistant strains emerge, new Daptomycin analogues are needed. Although the regioselectivity of CdpNPT prenylation on the indole ring of Daptomycin is limited, the resulting products show enhanced antibacterial activity against Daptomycin-resistant bacteria strains [158]. For a specific production of Daptomycin analogues, the active site of CdpNPT could be modified to enhance the specificity of the reaction [158].

Engineering of Prenyltransferases

In the literature, several examples of modified prenyltransferases are described that show changed substrate specificity, and chemo- or regioselectivity of the prenylation reaction. For prenyl-donor specificity, structure analysis of several DMATS revealed a gate keeper residue that is responsible for the dimethylallyl preference [147]. For example, native FgaPT2 catalyses prenylation with DMAPP as the prenyl-donor, exclusively. By replacing methionine (M328) to a smaller residue such as cysteine, the enzyme catalyses the same reaction with GPP (figure 2.4) [159]. A similar behaviour was observed for TleC and MpnD, where wild type enzymes transfer preferentially GPP and DMAPP, respectively. These preferences could be swapped from GPP to DMAPP and vice versa by a single point mutation in each enzyme [147]. Not only the prenyl-donor specificity of enzymes can be modified, but also examples for changed prenyl-acceptor preference by mutagenesis have been described. FgaPT2 prenylates L-Trp as natural prenyl-acceptor substrate and cyclic dipeptides only in low amounts, and by a single point mutation the substrate turnover number could be enhanced for cyclic dipeptides by 76-fold [160]. The regio- and chemoselectivity of the prenylation reaction can be redirected by single mutations in the active site. For FgaPT2, deletion of the catalytic base, by replacement of the lysine to alanine, changed the regio- and chemoselectivity from a regular C-4 towards a reverse C-3 prenylating enzyme [161]. Modifying the active site of TleC at three positions partly shifted geranylation from reverse C-7 to regular C-5 on the indolactam V [147]. In addition, a single mutation in the active site of FtmPT1 changed the obtained compounds from regular C-2 prenylated to a mixture of regular C-2, regular and reverse C-3, and regular N-1 prenylated products [162].

5-DMATS from *Streptomyces coelicolor*

5-DMATS from *Streptomyces coelicolor* was identified by genome sequencing of the organism. The biochemical analysis of the putative enzyme revealed a L-Trp C-5 prenylating activity, which represents most probably the natural substrate of the enzyme [163]. The protein is presumably part of an alkaloid biosynthetic gene cluster, but the natural secondary metabolite has not yet been identified. Further biochemical studies classified the DMATS as regular C-6 and C-7 prenylating protein of L-Trp in presence of an unnatural prenyl-donor (2-pentenyl-PP) [142]. In addition, in the presence of tripeptides as prenyl-acceptor substrates, the enzyme showed reduced regioselectivity and transfers one prenyl moiety (monoprenylated) onto C-5, C-6 and C-7, as well as two prenyl moieties (diprenylated) onto position C-5 and C-6 on the indole ring in a regular manner [164].

6-DMATS from *Mircomonospora olivasterospora*

The putative gene (*MolI14.36*) was identified in *Micromonospora olivasterospora* as putative tryptophan prenyltransferase as part of a sequence homology search with known tryptophan prenyltransferases from *Streptomyces*. The enzyme selectively catalyses the transfer of dimethylallyl entities onto the C-6 position of L- and D-Trp, and in further studies with indole derivatives C-5 and C-7 prenylated products were observed. The DMATS shows a prenyl-donor specificity towards DMAPP, although a low GPP and other isoprenoid derivatives was observed. For GPP, the regioselectivity is preserved to C-6 in the indole ring, in contrast 2-terpen or benzyl-PP prenyl-donors are attached to position C-5, C-6 and C-7 [142, 165].

2.2 Objective and Significance

Over the last two decades, fungal- and bacterial-derived DMATS were biochemically and structurally analysed. Nevertheless, there are still open questions that need to be addressed in order to enhance our understanding of these enzymes and to improve their usability in biotechnology.

- What is the structure and the active site of a 5-DMATS?

- How is the regioselectivity of this enzyme determined, and which reaction mechanism underlies the C-5 prenylation reaction of a 5-DMATS?

- Which positions in the active site are crucial to specifically manipulate the substrate donor and acceptor preference?

- Is a structure-based modification possible to create an enzyme with new regio- and chemoselectivity?

These questions were addressed in this work in an X-ray crystallographic approach combined with structure-based site-directed mutagenesis. The ligand-bound and ligand-free structures of 5-DMATS from *Streptomyces coelicolor* were determined to analyse the binding mode of substrates in the active site and to identify crucial residues for enzyme-catalysed formation of 5-dimethylallyl-tryptophan. In addition, the structure of a bacterial 6-DMATS from *Micromonospora olivasterospora* was determined and compared to an already described 6-DMATS, named PriB, to identify the reaction mechanism of the enzyme and their crucial residues in the active site [146]. For biotechnological application of DMATS enzymes, not only the natural catalysis is of interest but also manipulation of reaction specificity would be beneficial in drug discovery and production to obtain new biological active compounds. Structure-based site-directed mutagenesis was performed to specifically manipulate the substrate preference as well as the regio- and chemoselectivity of the reaction. In Cooperation with the group of Professor Shu-Ming Li, activity assays were performed with the 5-DMATS and 6-DMATS variants to analyse their substrate preference, as well as the regio- and chemoselectivity of the catalysed reaction.

2.3 Results and Discussion

This section summarises the published results of the structural investigation and structure-based engineering of 5-DMATS from *Streptomyces coelicolor* and 6-DMATS from *Mircomonospora olivasterospora*. For details and references please consult appendix section A.2, which contains the publication as full text.

Reprogramming Substrate and Catalytic Promiscuity of Tryptophan Prenyltransferases

Elena Ostertag (EO), Liujuan Zheng (LZ), Karina Broger (KB), Thilo Stehle (TS), Shu-Ming Li (SML) and Georg Zocher (GZ)

Contribution

EO designed the study of 5-DMATS. Protein production, purification, crystallographic studies of 5-DMATS were carried out by Karina Broger in the course of her bachelor thesis. EO performed structure analysis of 5-DMATS, as well as the protein purification, crystallographic studies and structure analysis of the 6-DMATS. Activity mutants were designed by EO and GZ, cloned and purified by EO. Activity measurements were performed by LZ. The manuscript was written by EO, LZ, GZ and SML.

2.3.1 Structural Characterisation

In the course of this work, both enzymes were structurally characterised by X-ray crysallography. For 6-DMATS, the structural model was obtained at 1.6 Å resolution and additional positive electron density for L-Trp was determined in the active site, even so no L-Trp was present in any solution during purification or crystallisation. The liganded ternary complex was obtained at 1.7 Å resolution by soaking of unliganded crystals with the prenyl-acceptor L-Trp and the prenyl-donor analogue dimethylallyl thiopyrophosphate (DMASPP). For 5-DMATS, unliganded and liganded structures were determined at 1.5 Å and 1.7 Å resolution, respectively. The fold of the enzymes correspond to ABBA-fold prenyltransferases with a characteristic β-barrel core built by ten anti-parallel β-strands that is surrounded by ten solvent exposed helices. In the liganded structures of 5-DMATS und 6-DMATS, positive electron density could be observed for both substrates in the active site, which is located in the β-barrel centre.

In both enzymes, the orientation of indole ring and dimethylallyl moiety of L-Trp and DMASPP, respectively, is equivalent to previously described bacterial-derived ABBA-fold prenyltransferase examples, such as TleC and MnpD, and perpendicularly arranged to the bound compounds in the active centre of fungal-derived enzymes. Two factors could determine the binding mode: first, the origin of the enzyme or second, the prenylation

reaction that is catalysed by the enzyme. The available biochemical data supports the hypothesis that the binding mode is determined by the prenylation position. The catalysed prenlylations of both wild type enzymes are less regioselective in the presence of unnatural prenyl-donor and -acceptor substrates, with C-5 and C-7, but no N-1, C-2, C-3 or C-4 prenylated products [142, 165, 164]. This indicates that the substrate orientation determines the position, at which prenylation can take place. In one binding mode positions N-1, C-2, C3 and C4 are accessible, whereas in the other, prenylation of position C-5, C-6 and C-7 are achievable. Further investigation of this hypothesis and the question of how the prenyltransferase origin affects the substrate binding mode further experiments have to be conducted.

2.3.2 Catalysis and Crucial Amino Acids

The active sites of both enzymes were analysed to determine the catalytic base and mechanism of the prenylation reaction. In the active site of 6-DMATS, a histidine was identified to act as catalytic base in a classical Friedel-Craft alkylation reaction. This finding is in agreement with the structural analysis of PriB, another DMATS prenylating the C-6 position of the indole ring. In PriB, a histidine was proposed as catalytic base located on another position in the active site, but with a similar distance to the C-6 position of the indole ring. For 5-DMATS, no basic residue could be identified in the active site, but a glutamine was assigned to fulfil a crucial role in the catalytic mechanism by coordinating a water molecule that could act as catalytic base during turnover. The distance between C-1 position of prenyl-donor and C-5 position of prenyl-acceptor do not support a direct Friedel-Crafts alkylation. Instead, an initial Friedel-Crafts alkylation onto another position in the indole ring was proposed followed by a rearrangement reaction, such as a Cope-rearrangement, to obtain a regular C-5 prenylated L-Trp. The structural investigation of the active site together with the mutagenesis approach showed that both enzymes act with different catalytic mechanisms.

2.3.3 Structure-based Protein Engineering

Prenyltransferases are especially interesting as possible biotechnological tools in the synthesis of complex biological active compounds. Hence, a targeted structure-based engineering would be a favourable aim. To adjust the prenyl-donor preference, a gate keeper residue was already described for other prenyltransferases such as TleC, MnpD and FgaPT2 [147, 159]. In 5-DMATS and 6-DMATS, a leucine could be identified that controls the acceptance of GPP as prenyl-donor, by exchanging this residues to a smaller residue (alanine), the turn over for GPP could be significantly enhanced for both enzymes, 5-DMATS and 6-DMATS. Next, a structure-based mutagenesis of the active site was performed to redirect the regioselectivity. The activity of the 5-DMATS could be partly shifted towards a C-6 prenylating enzyme by mutation of the important glutamine residue and introduction of the catalytic histidine identified in 6-DMATS. For 6-DMATS, a

selective redirection to the C-5 position could be obtained by eliminating the initial catalytic base (histidine) and introducing the C-5 prenylation specific residue glutamine. In summary, the substrate preference and regioselectivity of ABBA-fold prenyltransferases is determined by single residues in the active site. Furthermore, this work demonstrated that a targeted modification is achievable to redirect the catalysed prenylation reaction to a desired position in the aromatic ring system.

2.4 Conclusion and Outlook

Friedel-Crafts alkylation is the reaction mechanism in 6-DMATS and was already described as the catalytic mechanism for several ABBA-fold prenyltransferases. For 6-DMATS, the mechanism and crucial residues are identified and a targeted manipulated of the reaction is possible. In contrast, the underlying mechanism for C-5 prenylation is estimated but not yet fully understood. Therefore, manipulation is more challenging and further structural investigation of C-5 prenylating enzymes is necessary to draw a full picture for the catalysis. Furthermore, analysis of fungal-derived 5-DMATS or 7-DMATS would be desirable to address the question of what determines the binding mode of the substrates in the active site. This information could be used for site-directed mutagenesis of prenyltransferases. By extension of structure-based mutagenesis studies in prenyltransferases, a better prediction could be achieved about which position influences the substrate specificity or regio- and chemoselectivity. For applications in biotechnology, the studies should be expanded to enzymes that catalyse prenylation of more complex prenyl-acceptor substrates such as cyclic-dipeptides.

The possibility to control the prenylation and by combining these enzymes into a new gene cluster, like already demonstrated in the biotechnological production of prenylated Naringenin, expands possible applications for these enzymes. Tuning the regio- and chemoselectivity as well as substrate specificity could be valuable in this context to obtain selective and efficient drug production pathway of complex chemical compounds that can not be synthesised by classical organic chemistry.

3 Characterisation of the Inhibitory Immune Receptor Siglec-11

3.1 Introduction

3.1.1 Glycocalyx and Sialic Acid

Each cell of the human body possesses a sugar coat, the so-called glycocalyx, which is essential for cell survival [166]. This coat consists of glycans that exhibit functions in signalling, as well as cell-cell interactions, such as adhesion or microbial attachment [167, 168, 169]. In glycans, monosaccharides with different anomeric states (α or β) are combined via different linkages and branches to diversify the sugar trees. In addition, modifications such as sulphation or phosphorylation are also found on the sugar entities [170, 171]. In the ER and Golgi network, these glycans are linked to proteins as post-translational modification or transferred to lipids to form glycoproteins, proteoglycans or glycolipids [172, 173, 174, 171, 175]. In glycoproteins, glycans are attached either via a threonine or a serine (*O*-glycosylation), or to an asparagine (*N*-glycosylation) side chain (figure 3.1) [175]. Proteoglycans are proteins with long unbranched negatively charged sugar polymers, which are linked via a serine (*O*-glycosylation) to the protein [176]. In the group of surface sugars, sialic acid is the collective name for the nine-carbon atom sugars (most prominent member: neuraminic acid), which is often attached as the last component on the nonreducing end of sugar trees [167]. Sialic acid is abundant in vertebrates, but not generally found in invertebrates, prokaryotes or plants, and only a few bacteria possess sialic acids in their capsular polysaccharides [177, 178, 179]. The family of sialic acid composes different backbone components such as neuraminic acid (Neu), N-acetylneuraminic acid (Neu5Ac), N-glycolylneuraminic acid (Neu5Gc) or deaminoneuraminic acid (KDN), which can be connected via different α-glyosidic linkages (2,3; 2,6; 2,8; 2,9) [180, 181, 182]. In humans, the most abundant form is the Neu5Ac variant, although Neu5Gc is commonly found in other mammals (figure 3.1) [183]. The lack of Neu5Gc in humans is the result of an inactivation mutation of the cytidine monophosphate-Neu5Ac (CMP-Neu5Ac) hydrolase gene, which converts the sialic acid donor CMP-Neu5Ac in CMP-Neu5Gc [184, 185, 186]. Chemical diversity is further increased by modifications at position C4, C7, C8 and C9, such as *O*-acetylation, *O*-phosphorylation, *O*-lactylation, *O*-methylation, *O*-glycosylation and *O*-sulphation [187, 188, 189, 190, 191, 192].
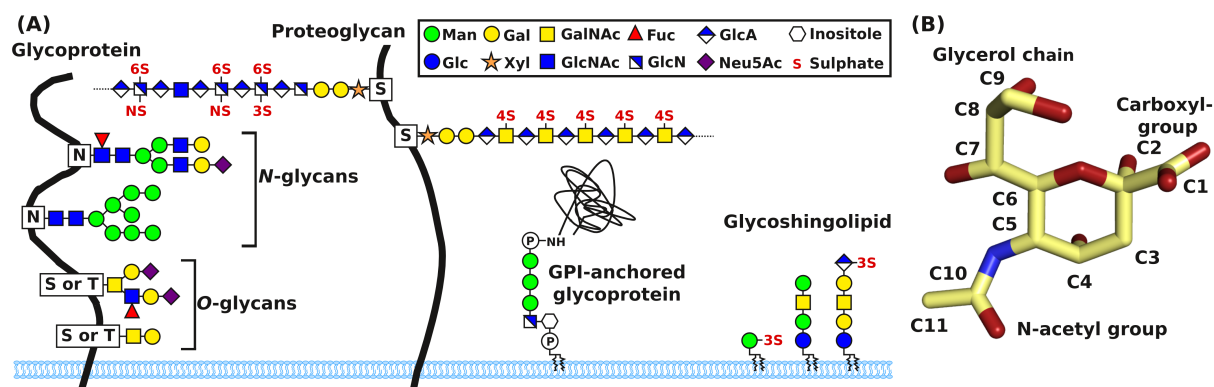
Figure 3.1: Major glycosylation types found in humans. (A) Glycans can be attached to glycoproteins, proteoglycans or lipids. In glycoproteins, *O*- and *N*-linked glycans are covalently linked to the oxygen of a threonine (T) or a serine (S) residue, or the nitrogen of an asparagine (N) residue in the protein, respectively. *N*-linked glycans form branched, heterogeneous structures with a conserved core of two N-acetylglucosamine (GlcNAc) and three mannose (Man) sugar moieties. *O*-glycans are even more diverse with mucin-type glycans, which comprise N-acetylgalactosamine (GalNAc) as the first sugar entity. *O*- and *N*-linked glycans contain various sugars, including glucose (Glc), galactose (Gal), fucose (Fuc) with modifications such as sulphation (S in red). Proteoglycans feature a long glycosaminoglycan (GAG) chain covalently linked via sugar chain of xylose (Xyl)-Gal-Gal-glucoronic acid (GlcA) on a serine of the protein and represents the major group of glycoproteins on the cell surface. GAGs are built up by repetitive disaccharide subunits with various compositions and degrees of sulphation. Glycosylphosphatidylinositol (GPI)-anchored glycoproteins are the major group of glycoconjugates. Their C-terminus is linked to a phosphoethanolamine, which is connected via a trimannosyl-nonacetylated glycosamine (Man3-GlcN) to a phosphatidylinositol in the membrane. Another class of glycoconjugates are glycosphingolipids, which are glycans attached to membrane lipids. This figure is adapted from [193] and the represented sugars in the glycans are drawn according to the symbol nomenclature for glycans [194]. **(B)** N-acetylneuraminic acid is the sialic acid form, which is predominantly found in humans and is represented including the numbering scheme and functional groups.

3.1.2 Polysialic Acid

Besides different glycosidic linkages and modification on sialic acids, further complexity arises by the presence of long sialic acid chains with different functions based on the degree of polymerisation (DP). Chain lengths of two to seven sialic acid monomers are referred to as oligosialic acid (oligoSia), longer chains are defined as polysialic acid (polySia). For mammals, only homopolymers of α2,8-linked Neu5Ac were described. Similar α2,8-linked polysialic acid structures are found in the capsule of some bacteria such as *E. coli* serotype K1, *Neisseria meningitidis* (*N. menigitidis*) serogroup B and *Mannheimia haemolytica* A2 (*M. haemolytica*) and are often referred to as colominic acid [195].

Polysialylation is a post-translational modification of proteins and lipids carried out in the Golgi-apparatus in humans. A group of sialyltransferases, the so-called ST8Sia family, are anchored in the membrane of the Golgi-apparatus and synthesise these oligo- and polyα2,8-linked sialic acid chains [196, 197, 198]. Thereof, ST8SiaIII is an oligosialyltransferase, whereas ST8SiaII and ST8SiaIV catalyse the polysialylation of a distinct repertoire of substrates [196]. The most commonly found and best studied substrate of ST8SiaII and ST8SiaIV is the neuronal receptor neural cell adhesion molecule (NCAM), from which all three isoforms (NCAM-120, NCAM-140, NCAM-18) exist as polysialylated and non-polysialylated form [199]. Neuropilin 2 (NRP-2), synaptic cell adhesion molecule 1

(Syn-CAM-1), CD36 a scavenger receptor in milk, C-C chemokine receptor type 7 (CCR7), as well as the α-subunit of the voltage sensitive sodium channel are also polysialylated by ST8SiaII and ST8SiaIV in humans [200, 201, 202, 203, 204]. It was shown that the amount of polysialylated proteins on the surface is tightly connected to the expression level of the polysialyltransferases [205]. The expression of polysialyltransferases themselves is regulated and ST8SiaII transcripts are mainly present in the hippocampus, the medulla oblongata and the putamen, and ST8SiaIV is prominent in the amylata, the subthalamic nucleus, the vertebral cortex and the occipital lobe of adult brains [206].

The conformation of polySia was investigated and up to now, a scientific debate is ongoing about the secondary and tertiary structure of this polymer. The results of one nuclear magnetic resonance (NMR) study indicate that polySia has a left-handed α-helical conformation with three to four monomers per turn and a pitch height of 9-11 Å [207]. But the data of another NMR study suggests a more relaxed helical conformation with six to nine monomers per turn [208]. In another approach, the structure of a polySia binding antibody (mAB735) was solved and the analysis, combined with NMR data, revealed a helical conformation with a pitch height of 36 Å and six monomers per turn for polySia (DP=10) [209]. In a more recent NMR study with oligoSia (DP=4), two models of a left-handed helix were predicted with most probably two, and less likely four, monomers per turn in solution [210]. By comparison of these studies, one can conclude that polySia can adopt a secondary structure with a high degree of flexibility that could be influenced by binding to interaction partners [211]. This transient secondary structure could explain the observation that different polySia-binding proteins need a certain DP number to engage the sugar chain. This theory is supported by several oligo- and polySia-binding antibodies, which either need a minimum DP number, such as mAB735 with DP>10, or a maximum number, with DP<4 in case of mAbA2B5, for binding [212, 209]. In addition, the DP dependent binding was also demonstrated for different receptors such as adenovirus 52 short fibre knob, which can bind DP=3 but exhibits higher affinity to DP>5 [213].
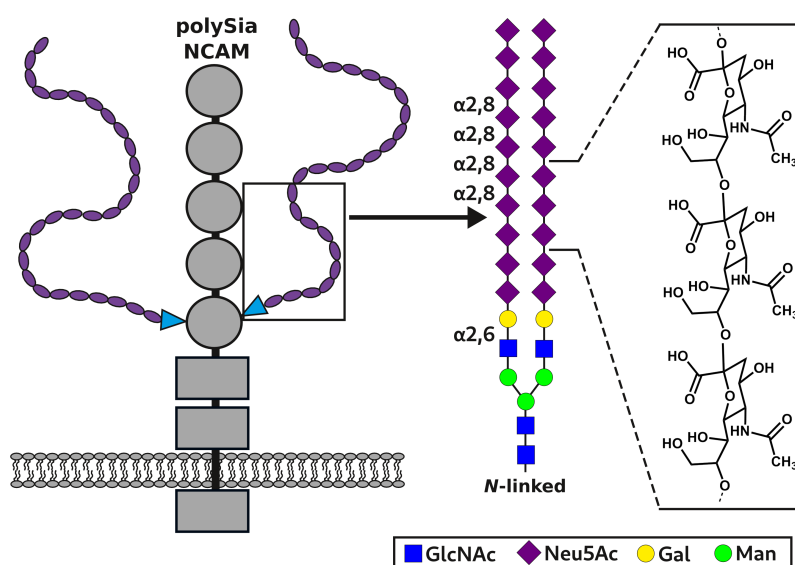
Figure 3.2: Schematic representation of polySia on the example of NCAM. Polysialylated NCAM is schematically represented. The composition of exemplary *N*-linked stem glycan is shown including two N-acetylglucosamine (GlcNAc) (blue squares), followed by three mannoses (Man) (green cycle) to split the glycan in two antenna with one GlcNAc and one galactose (Gal) (yellow cycle) on each site. On this stem glycan, a variable number of α2,8-linked sialic acid are attached. The chemical structure of the polySia is shown as a close-up on the right.

Physiological Function

In humans, oligoSia and polySia fulfil many functions in neuronal signalling and regulation of the immune system. Intensive research on polySia-NCAM revealed essential functions during embryonic development and a lack of ST8SiaII and ST8SiaIV is lethal for mouse off-springs [214]. PolySia impacts signalling and development processes in several ways. During embryonic-brain development, polySia-NCAM has the major role to mask NCAM and thereby ensures a tightly regulation of NCAM interactions. A loss of polySia leads to *trans*-interactions of NCAM at cell-cell contact sites, which reduces proliferation and triggers differentiation [215, 216, 217]. Polysialylation of NCAM disrupts fibroblast growth factor 2 (FGF2) receptor signalling, induced by heterophilic NCAM interaction, to control neurogenesis, cell growth and differentiation [218]. PolySia-NCAM is mainly found in brain tissue and is a key player in the regulation of brain development and control of cell differentiation with a critical role in axon outgrowth. Thereby, it was shown that the expression level of polySia-NCAM decreases post-natal, and NCAM in adults carries polySia only in brain regions with high plasticity such as hippocampus and olfactory bulb [219, 220]. The synaptic plasticity is especially important for memory formation and learning [221]. In the hypophysis, polySia is expressed as a reaction to changes in the ovarian cycle, lactation or dehydration and allows to establish new synapses [222]. A steric inhibition of cell-cell apposition is also part of the polySia modulation strategy. The membrane organisation is affected by polysialylation. PolySia-NCAM is mainly found in lipid rafts, whereas the non-polysialylated NCAM is present in both raft as well as non-raft fraction affecting NCAM signalling [223]. Besides the modulation of NCAM signalling, several examples for carrier independent functions of polySia are described. PolySia acts as a reservoir for small signalling molecules by interacting with brain-derived neurotrophic factor (BDNF), catecholamines, ions such as Ca^{2+} or fibroblast growth factor 2 (FGF2) in a DP dependent manner to modulate the activating and inhibiting signalling processes [224, 225, 226]. Here, polySia is predominantly, but not exclusively, found in neuronal tissue and has functions in protection, repair and regeneration of tissues. It was further discovered in lung, testis, placenta and milk [227, 228, 229].

PolySia regulates processes in the immune system and is widely distributed on the surface of immune cells, such as hematopoietic precursors, bone marrow-derived neutrophils, natural killer (NK) cells, monocytes and monocyte-derived macrophages [201, 230, 231]. In the innate immune system, polySia is involved in the regulation of neutrophil extracellular trap (NET) formation. The high number of histones (up to 70% of the NET proteins) that are released during NETosis can have side effects such as sepsis, autoimmune disease and infertility [232]. PolySia modulates this process to prevent NETosis induced tissue damage and fights pathogen invasion at the same time [233, 234]. Besides the function in NETosis, polySia has functions in maturation, activation and regulation of immune cells. For example on the surface of dendritic cells, it is involved in the activation of $CD4^+$ cells, but also present on the surface of $CD4^+$ cells itself. There, it has a potential

role in migration and maturation of T-cells in the thymus [201, 235, 236]. In addition, polySia acts as regulator of immune responses with an inhibitory effect on macrophage and microglia activation [237]. A reduced phagocytotic activity of macrophages was observed for bacteria with polySia on their cell surfaces [238]. This effect is based on polySia interactions with the sialic acid-binding immunoglobulin-like lectin-11 (Siglec-11) [237].

3.1.3 Siglec Family

Siglecs are the best explored subtype of immunoglobulin-like (I-type) lectins and belong to the type-1 membrane proteins [239, 240, 241]. These receptors are mostly found on the surfaces of haematopoietic cells with two exceptions: Siglec-4 (also known as myelin-associated glycoprotein, MAG) is present on the surface of Schwann cells and oligodendrocytes, and Siglec-6 was found on trophoblasts of placenta [242, 243]. Some of these receptors possess a highly restricted expression pattern. For example, Siglec-1 (also known as Sialoadhesin or CD169) is an adhesion molecule present on macrophages or activated monocytes and Siglec-2 (also known as CD22) is only found as an inhibitory receptor on surfaces of B cells [244, 245]. In contrast, Siglec-3 (CD33) is present on various cell-types such as macrophages, monocytes, microglia, granulocytes and myeloid progenitors, and Siglec-8 is found on eosinophils, alveolar macrophages and activated microglia [246]. In addition, cells can express more than one Siglec variant, as shown for B cells, that possess Siglec-2, as well as Siglec-10 (mouse Siglec-G) as inhibitory receptors [247, 248].
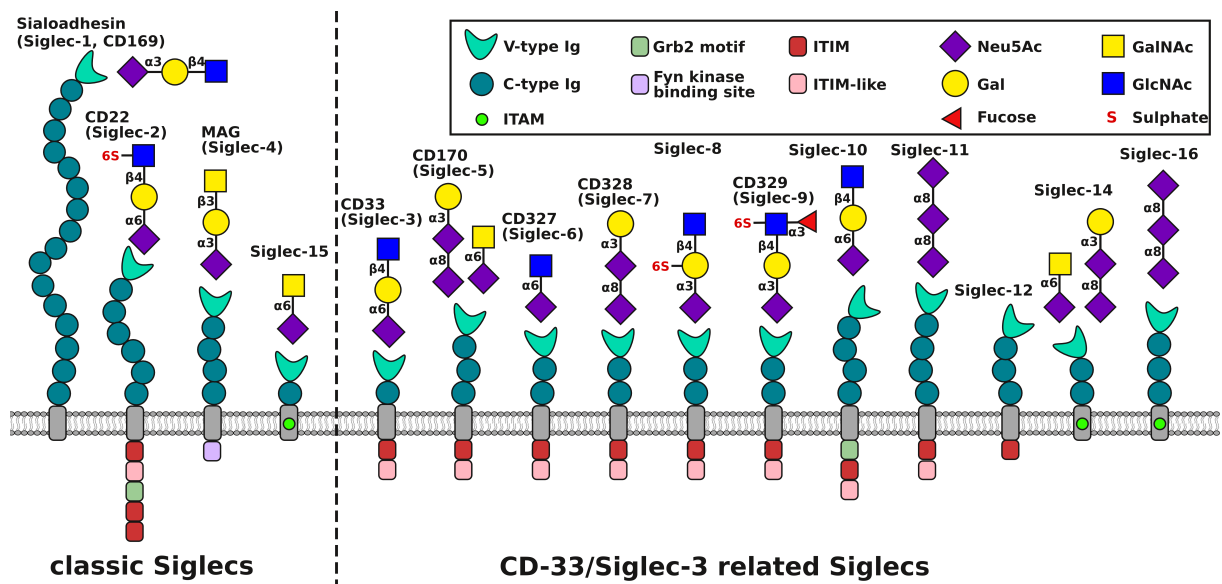
Figure 3.3: Schematic representation of the Siglec family members. The domain composition of all members of the classic and CD33-related Siglecs, found in humans, are shown with their ligand glycans. The glycans are represented according to the symbol nomenclature for glycans [194].

CD33-related Siglecs

The 15 Siglec members in humans can be subdivided in two subfamilies, the so-called classic Siglecs that are conserved among mammals comprising Siglec-1, Siglec-2, Siglec-4, Siglec-15 and the CD33-related Siglecs including Siglec-3 (also known as CD33), Siglec-5-11, Siglec-14 and Siglec-16 (figure 3.3). The CD33-related Siglecs of humans evolved rapidly and share a high sequence identity of 50-99% within the group, but low identity to possible orthologues in other mammals. Interestingly, the classical Siglecs are less conserved to their paralogs of the same subfamily but show a higher sequence identity towards orthologues of other organisms including several invertebrates and vertebrates, e.g. fishes [249, 250]. A good example for the rapid evolution in the CD33-related family is Siglec-5 (inhibitory) and Siglec-14 (activatory), both are very similar and are evolved, most likely, from each other to form counterpart in signalling [251, 249]. In mice, only nine members of the CD33-related Siglecs are known and these possess overlapping functions to their orthologous in humans, which makes it difficult to assign orthologues for humans and mice [252].

Physiological Function

Each Siglec preferentially binds a distinct sialylated glycan structure, which is present on the surface of mammalian cells. Exceptions are Siglec-12, which does not bind, and Siglec-13 and Siglec-17, which can only be found in other species due to gene inactivation in humans (figure 3.3) [253, 254]. Siglecs are low affinity binders and facilitate binding due to the high concentration of sialic acids on the cell surfaces of immune cells (up to $100\,\mathrm{mM}$ on the cell surface of B cells). Here, interactions on the same surface (*cis*-interactions) dominate over interactions with surface components of opponent cells (*trans*-interactions) [255]. The dominating *cis*-interactions downregulate but do not prevent binding of *trans*-ligands, which can occur dynamically in the presence of high *trans*-ligand concentrations. Siglecs recognise most of the glycans as 'self', and ligand interactions modulate the immune cell activity by either inhibiting the activatory receptor signalling cascade or sequestering the Siglec away from the activatory receptor [246]. For example, B-cell receptor signalling is essential for initiating a B-cell response, but the strength and signal initiation itself has to be tightly regulated to prevent hyperactive B-cells. This is ensured by the inhibitory receptors Siglec-2 and Siglec-10, and knock-out mutants of these receptors can lead to autoimmune reactions of B-cells [256]. The balance of *cis*- and *trans*-interactions is also influenced by the number of extracellular domains (ECD) of each Siglec. Siglec-1 establishes mainly *trans*-interactions due to its elongated ECD with 17 domains [257]. On the other hand, Siglec-3 or Siglec-8, which possess only two or three domains, respectively, interact predominantly in *cis* [258, 259]. Siglec signalling is additionally regulated by the dynamic number of receptors present on the immune cell surface. Most of the receptor cycle between the surface and intracellular endosomes or are taken up by endocytosis after multivalent ligand binding in an either cathrin dependent or independent pathway [260, 261, 262].

Pathogenic Role

Regulatory functions of Siglecs have a role in diseases such as Alzheimer's disease (AD) or in tumour progression, but are also exploited as viral receptor. In addition, bacteria such as *N. meningitidis* and *Haemophilus influenzae* are human pathogens that express glycans on their surfaces, which are recognised by Siglecs as 'self' to mask them from the immune system [263, 264]. Also tumours utilise this mechanism by decorating their surfaces with α2,8-linked sialic acids that are recognised by Siglec-7 and prevents clearance by NK cells [265]. Pancreatic ductal adenocarcinoma tumour cells possess an increased level of α2,3-linked sialic acids, which is recognised as ligands by Siglec-7 and Siglec-9 and not only downregulates the immune cell activity, additionally binding provokes a Siglec-9-mediated differentiation process from monocytes to immune suppressive tumour associated macrophages (TAM) [266]. In tumours, TAMs have a central role in tumour progression and metastasis and produce cytokines, chemokines and growth factors to create an immunosuppressive micro environments [267]. In addition, some viruses use Siglecs as receptors, such as human immune-deficiency virus (HIV), which binds Siglec-1 on monocytes via its glycoprotein gp120. These monocytes are often already infected with HIV but still bind virus particles to efficiently deliver them to target cells, enhancing virus distribution [268, 269].

Architecture and Domain Function

The ECD of Siglecs is composed of domains that belong to the immunoglobulin (Ig) superfamily, including a variable (V-set) Ig-like domain on the N-terminus, followed by a variable number (1-16) of constant (C-set) Ig-domains. The receptors span the membrane with a single α-helix and possess for signalling either one or more immunoreceptor tyrosine-based inhibitory motifs (ITIM) and ITIM-like motifs, or an activatory immunoreceptor tyrosine-based activation motif (ITAM) in the cytoplasm. The V-set comprises the sialic acid binding site and is structurally similar to the variable domain of antibodies, built up by two β-sheets, A(A')B(B')ED and C(C')FG(G'), that are linked by an intra-domain disulphide bridge (figure 3.4 A). The binding pocket is formed by F- and G-strands, as well as the CC'- and C'D-loops, with a conserved arginine residue that forms a salt bridge towards the carboxyl group of the sialic acid (figure 3.4 B) [270]. Furthermore, a conserved aromatic residue (mostly tryptophan) is present in all Siglecs and interacts with the glycerol side chain to further engage the sialic acid. The highly variable CC'-loop and residues of the G-strand are key determinants for the binding specificity [271]. Siglecs are low affinity binders, which engage their substrates with affinities in the high micromolar to low millimolare range. Thereby, each Siglec possesses a specificity profile for different linkages (mainly Neu5Acα2,3Gal, Neu5Acα2,6Gal and Neu5Acα2,8Neu5Ac) and modifications of the terminal sialic acid, but overlaps of these profiles are observed. Structural analysis revealed that V-set and ligand interactions are mainly determined by the sialic acid and the adjacent sugar monomer (figure 3.3 and 3.4 B) [272]. The C-type Ig-domains are subdivided into constant-1 (C1-set) and constant-2 (C2-set), both possess an Ig-like fold

and are built by two β-strands, but differ in the composition. C1-set is formed by ABED and CFG strands, whereas the C2-set topology possesses a strand composition of ABE and C(C')FG(G') (figure 3.4 A) [273]. The variable inter-domain linker length, the structural difference of the C1-set and C2-set and the variable number of C-type domains affect the flexibility of the ECD of each receptor. For CD22, a semi-rigid rod-like structure was determined that is beneficial for binding with flexible *cis-* and *trans-*ligands on the cell surfaces [274, 273].
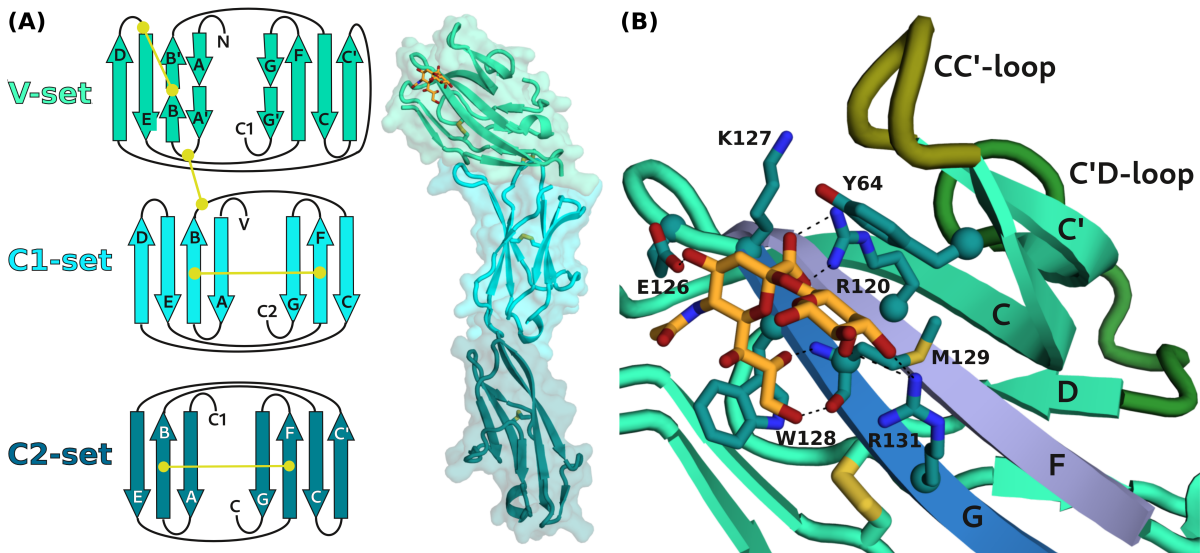
Figure 3.4: Crystallographic and topological strucutre of Siglec-2. (A) The structure and topology of the V-set (green), C1-set (cyan) and C2-set (dark blue) are exemplary shown for Siglec-2 (pdb code: 5vkm) [273]. The intra- and inter-domain disulphate bonds are indicated as yellow dots connected by lines. **(B)** A close-up view of the active site is represented with the engaged ligand (Neu5Acα2,6-Gal) in orange and all interacting residues, including the essential arginine residue (R120) in the F-strand (purple) and the conserved tryptophan (W128) of the G-strand represented in dark green. In addition, the CC' and C'D-loops are highlighted in brown green and dark green, respectively.

For signalling, most of the Siglecs contain either ITAM or ITIM/ITIM-like motifs in the cytoplasmic part of the receptor. In case of an ITAM motif, an activation signal is transduced into the cell, documented for Siglec-14, Siglec-15 and Siglec-16 [251, 275, 276]. The ITAM motifs comprise either a conserved tyrosine that is part of a consensus sequence of Y-x-x-I/L-x(6-8)-Y-x-x-I/L, whereas x stands for any amino acid, or have a positively charged transmembrane amino acid (lysine or arginine). In case of the positively charged transmembrane residue, the signal transduction is enabled by an ITAM-encoding adapter protein, such as DAP10 or DAP12, with a negatively charged residue in the transmembrane part. Siglec-14, Siglec-15 and Siglec-16 engage their ligands and an activating signal is transmitted via charge-charge interactions between the positively charged transmembrane residue of the Siglec and the aspartic acid of the ITAM-motif as part of DAP10/12 [251, 275, 276]. Downstream, tyrosine of DAP10/12 is phosphorylated by the Src kinase family and serves as a docking site for the SH2 domain of ZAP70 and Syk kinases, which initiate a signalling cascade (figure 3.5).

In contrast, other Siglecs contain one or two ITIM or ITIM-like motif to silence ITAM-dependent immune responses, as shown for Siglec-2, Sigelc-3, Siglec5, Siglec-6, Siglec-7, Siglec8, Siglec9, Siglec-10 and Siglec-11. The ITIM motif is a consensus sequence of (V/L/I/S)-x-Y-x-x-(L/V), whereas x means any amino acid, composing a tyrosine that becomes phosphorylated by Src kinase after ligand engagement. The phosphotyrosine provides a docking platform for the SH-domain containing phosphatases SHP1 and SHP2 to intervene in the activation of immune cells, by dephosphorylation of a kinase in the earliest step of the signal cascade [277].

3.1.4 Siglec-11

Siglec-11 is an inhibitory CD33-related receptor, which is also found in primates but not in rodents, and contains one ITIM and one ITIM-like motif in the cytoplasmic part of the receptor. In humans, Siglec-11 is expressed on the surface of macrophages and microglia, and downregulates immune cell signalling [278]. More recently, two different splice variants, Siglec-11(4D) and Siglec-11(5D), were found on the cell surfaces of microglia and macrophages, respectively. Both comprise one N-terminal V-set, but differ in the number of C2-sets with either three or four for the Siglec-11(4D) and Siglec-11(5D) variant, respectively [279]. It was shown that Siglec-11(4D) is able to bind to polySia with higher affinity than the Siglec-11(5D) variant [279]. Siglec-11 acts as the counterpart to the activatory receptor Siglec-16 [270]. Both receptors share a sequence identity of 99% for the two outermost ECD and both bind oligo- and poly-α2,8-linked sialic acid as ligands [280].

Siglec-11 binds exclusively α2,8-linked sialic acid with low affinity, unlike other α2,8-linked sialic acid binding Siglecs (Siglec-5, Siglec-7 and Siglec-14) that also bind α2,3- and α2,6-linkages [270]. *In-vitro* binding studies determined interaction of Siglec-11 and oligoSia with a DP number of 3, but in cell-based studies an average DP of 20 was defined as the physiological ligand [237]. In macrophages, soluble polySia with DP=20 could prevent an oxidative outburst, induced by neural debris or fibrillary amyloid-$\beta_{1\text{-}42}$, further polySia can neutralise specifically LPS-triggered phagocytosis, without affecting the basal phagocytotic activity of the cell [237]. For microglia, it was shown that Siglec-11 is controlled by *trans*-interactions and enables anti-inflammatory signalling, reduces phagocytosis and cytotoxicity for neuronal cells [280]. The conserved arginine, found in the binding site of other Siglecs, is also present in the V-set of Siglec-11, but does not seem to be essential for sialic acid binding. Mutation of this arginine to alanine or lysine reduces but not abolishes the binding, in contrast to Siglec-3, Siglec-7 and Siglec-9 [281, 282]. In addition, removal of the glycerol entity in sialic acid, which is normally engaged by an aromatic ring in the V-set of other Siglecs, does not influence the binding to Siglec-11. Together, these findings indicate that Siglec-11 ligand recognition differs from the mechanism revealed for other CD33-related Siglecs.
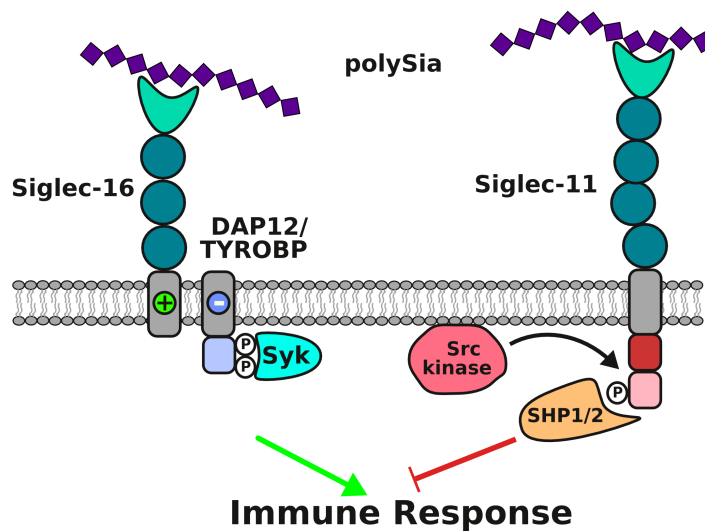
Figure 3.5: Siglec-11 and Siglec-16 antagonistic signalling cascade. Siglec-11 and Siglec-16 transfer signals via its cytoplasmic domains to modulate immune cell response to environmental triggers. Siglec-16 acts as an activatory receptor and initiates a signalling cascade via charge-charge interactions of its arginine and an aspartic acid of the DAP12/TYROBP, both located within the transmembrane domain. This causes the phosphorylation of cytoplasmic ITAM-motif in DAP12/TYROBP and initiates an activatory signalling cascade. On the other hand, Siglec-11 interaction with polySia prevents immune cell activation. The cytoplasmatic ITIM-motifs in Siglec-11 are phosphorylated by Src kinase upon ligand binding, which recruits and activates the phosphatases SHP1/2. These two phosphatases interrupt the activatory signalling cascade in the cells by dephosphorylation of the corresponding kinase.

For signal transduction, the ITIM-motifs of Siglec-11 are phosphorylated upon polySia binding to establish an interaction platform for the Src homology domain of SHP-1 and SHP-2. In the following, the two phosphatases, SHP-1 and SHP-2, can intervene in the activation cascade of macrophages and microglia (figure 3.5) [283, 270]. Besides the physiological role of Siglec-11 to inhibit inflammatory responses of macrophages and microglia, the downregulating property is exploited by pathogens and cancer cells. As already mentioned, some bacteria express α2,8-linked polySia on their capsule, which can interact with Siglec-11 to dampen inflammation processes and phagocytosis [283]. Not only pathogens are using the α2,8-linked sialic acid to hide from the immune system. Glioblastoma is an aggressive form of human cancer, which uses the downregulating properties of Siglecs to evade immune cells and promote tumour progression and metastasis. One strategy for this tumour is to decorate the cell surface with a dense coat of glycans including α2,8-linked polysialic acid [284]. As a consequence from the pathogenic role of polySia, the activatory Siglec-16 is evolutionary evolved from the Siglec-11 gene and recognises polySia as well [285]. This receptor acts as the counterpart and starts an activatory signalling cascade, via a positively charged residue in the transmembrane domain to engage the adapter protein DAP12, to initiate migration of the immune cell and phagocytosis (figure 3.5) [276].

3.2 Objective and Significance

PolySia is a tightly regulated and rare post-translational modification of distinct proteins, only present in specific tissues, and the amount of polySia differs during development. As already mentioned, polySia fulfils various functions, including a regulatory role on immune cell activation. This mechanism is not only used to protect healthy cells from the immune system, but is also exploited by pathogens, and some tumours are highly decorated with polySia to mask themself from the immune system. Besides the specific functions, the interaction of Siglec-11 with polySia is an interesting candidate for a detailed analysis, as it seems to follow other parameter than determined for ligand binding of other Siglecs. First, the binding pocket differs and residues crucial for ligand binding in other Siglecs fulfil no crucial role in Siglec-11, as Siglec-11 is still able to bind polySia in absence of the strictly conserved arginine [270]. Second, Siglec-11 ligand binding is not exclusively determined by the last two sugar moieties, instead it was demonstrated that an α2,8-linked sialic acid moiety of at least three monomers is necessary to establish an interactions to Siglec-11 and a chain length of 20 is needed for physiological activity [237]. A few open questions need to be addressed in order to understand the regulatory mechanism of polySia on immune cells via Siglec-11.

- How is polySia engaged by Siglec-11 V-set, and are there residues that can compensate for a mutation of the strictly conserved arginine in Siglec-11?

- Is the binding of polySia extended beyond the V-set with other ECDs involved?

- Is Siglec-11 monomer or dimer in solution, and in case of a dimer, is binding of polySia to both V-sets required to transduce signals into the cell?

- If Siglec-11 affinity is dependent on the DP of polySia, is there an optimal polySia chain length for binding?

For analysis of Siglec-11 and its interaction to polySia on molecular level, ECD of Siglec-11 needs to be produced as homogeneous, pure proteins in high amounts. In this work, Siglec-11 variants that differ in the number of ECDs were obtained to identify the oligomeric state of Siglec-11 in solution. First affinity measurements were performed to determine the affinity of Siglec-11 to polySia. In addition, the structure of Siglec-11 V-set domain was obtained and lays the foundation to investigate ligand binding on an atomic level.

3.3 Materials

3.3.1 Chemicals

Acrylamide-Bisacrylamide solution Rotiphorese® 30%(v/w) Gel	Roth
Agar	Roth
Agarose	Roth
Ampicillin	Sigma-Aldrich
Ammonium acetate	Sigma-Aldrich
Ammonium persulfate (APS)	Roth
Ammonium sulfate	Merck
L-Arginine	Roth
β-Mercaptoethanol	Sigma Aldrich
Bromphenol blue	Roth
Calcium chloride	Sigma-Aldrich
Colominic acid	Sigma Aldrich
dNTP	Genaxxon
Dithiothreitol (DTT)	Sigma-Aldrich
L-Glutamate	Sigma-Aldrich
L-Glycine	Sigma-Aldrich
HEPES	Roth
Instant blue®	Abcam
Imidazole	Roth
Isopropanol	Sigma Aldrich
Luria Both media (LB)	Roth
Magnesium chloride	Sigma-Aldrich
Manganese chloride	Sigma-Aldrich
Methanol	Merck
Methyl-α-D-mannopyranoside	Sigma-Aldrich
Milk powder	Roth
Polyethylene glycol (PEG) 400	Sigma-Aldrich
Polyethylene glycol (PEG) 4000	Sigma-Aldrich
Polyethylenimine (PEI)	Sigma-Aldrich
Rotiphorese® 10 × Running buffer	Roth
SDS	Roth
Sodium cacodylate	Sigma-Aldrich
Sodium chloride	Roth
Sodium citrate	Sigma-Aldrich
TEMED	Roth
Trizma base	Sigma Aldrich
Tryptone	Roth
Tween-20®	Sigma-Aldrich
Yeast extract	Roth
Zeocin™	Gibco™

3.3.2 Enzymes, Antibodies and Reagents

Antibodies and Reagents

ECL™ Western Blotting Detection Reagents	Sigma
Expifectamine	Thermo Fisher
GeneRuler 1 kb Plus DNA Ladder	Fisher Scientific
GlycoBuffer 2 (10×)	New England BioLabs

GlycoBuffer 3 (10×)	New England BioLabs
Goat α-mouse antibody	Novagen
Mouse α-his antibody	Novagen
Mouse α-c-myc antibody (E910)	Merck
Q5 reaction buffer	New England BioLabs
Page Ruler™ Prestained Protein Ladder	Fermentas
Page Ruler™ Unstained Protein Ladder	Fermentas
5 × buffer T4 DNA PNK	New England BioLabs

Enzymes

Dpn I	Fermentas
EndoH$_f$	New England BioLabs
In-Fusion HD® Enzyme Premix	Takara Bio Europe SAS
PNGaseF	New England BioLabs
TEV protease	in-house production
Q5® high-fidelity DNA Polymerase	New England BioLabs
T4 DNA ligase	New England BioLabs
T4 DNA PNK	New England BioLabs

Crystallisation Screens

Crystal screen	Hampton Research
Crystal screen 2	Hampton Research
JCSG Plus	Molecular Dimensions
Morpheus	Molecular Dimensions
PEG/Ion	Hampton Research
PEG/Ion 2	Hampton Research

3.3.3 Hardware

ÄKTA Prime Plus	GE Healthcare
ÄKTA Basic	Amersham Pharamcia Biotech
ÄKTA Express	GE Healthcare
ÄKTA Micro (Ettan LC)	GE Healthcare
Autoclave VX 95	Systenic
Centrifuge Haereus multifuge 1 L-R	Thermo scientific
SLC4000 rotor (Sorvall™ RC 6 Plus)	Thermo scientific
SS34 rotor (Sorvall™ RC 6 Plus)	Thermo scientific
Centrifuge 5415D	Eppendorf
ChemiDoc™ MP Imagine System	Bio-Rad
Cryoloops (0.1-0.4 mm)	Molecular Dimensions
Crystallisation Robot Gryphon	Art Robinson Instruments
Crystallisation Robot (Freedom EVO™)	Tecan
Digital Sonifier	Brandson
Heating block neoblock 1	neoLab
HERA cell 150	Kendro Laboratory Products
HisTrap FF Crude affinity column (5 mL)	Cytiva
HiTrap Con A 4B column (5 mL)	Cytiva
iCycler™ Thermal Cycler	Biorad
Incubation shaker	Infors HT
Infors HT celltron shaker	Infors HT
Microscope Leica MZ 16	Meyer Instruments
Microscope Leica MZ 95	Meyer Instruments
Microwave	Parasonic

Nanodrop ND-1000	Thermo Fisher Scientific
Nylon & pins	GE Healthcare
Milli-Q Merk	Millipore
Nalgene®Mr. Frosty	Sigma-Aldrich
pH meter BP11	Sartorius
Peristaltic pump	EconoPump BioRad
Primus96plus	MWG-Biotech AG
Pump P-500 (FPLC)	Pharmacia
Spectropolarimeter J-720	JASCO
Superdex™ 75 column (Hi-Load 16/600)	Cytiva
Superdex™ 200 column (Hi-Load 16/600)	Cytiva
Superdex™ 200 increase column (3.2/300)	Cytiva
Vacuum centrifuge SpeedVac Heto	Heto

3.3.4 Consumables

Amicon ULTRA (MWCO 3 kDa)	Millipore
Amicon ULTRA (MWCO 10 kDa)	Millipore
Blotting Paper(703)	VWR North america
Centrifugal filters (0.22 µm)	Corning
Corning® Erlenmeyer cell culture flask (125 mL, 250 mL, 500 mL)	Merck
Cover slides 12-542-C	Fischer Band
Crystalquick plate (96 deep-well plate)	Greiner Bio One (GBO)
Diamond pipet tips	Gilson
GenElute™ HP Plasmid Maxiprep Kit	Sigma-Aldrich
GenElute™ HP Plasmid Miniprep Kit	Sigma-Aldrich
GenElute™ PCR Clean-Up Kit	Sigma-Aldrich
In-Fusion HD® Cloning Kit	Takara Bio Europe SAS
Instant Blue™	Expedeon
Membrane filters (0.22 µm)	Millipore
PCR Micro Amp reaction tubes	Applied Biosystems
Pipet tips (different size)	Herbe Plus
Plastic cuvettes	Sarstedt
Plate HR3-140 (24 deep-well plate)	Hampton Research
Rothi®-PVDF Membrane (0.45 µm)	Roth
Reaction tubes (1.5, 2, 15, 50 mL)	Greiner Bio-one
Sensor Chip NTA	Cytiva
Specra/Por dialysis membrane (MWCO 25 kDa)	Spectrum Laboratories
Syringe-top filters (0.22 µm)	VWR International
Syringes (30 mL, 10 mL)	Greiner Bio-one

3.3.5 Software

Alphafold	EMBL-EBI, DeepMind [286, 287]
APBS Electrostatics Plugin	National Biomedical Computation Resource [288]
Blast	Basic Local Alignment Search Tool [289]
Chainsaw	Collaborative Computational Project number 4 7.1.002 Suite [290]
CCP4 (7.1.002 Suite)	Collaborative Computational Project number 4 7.1.002 Suite [291]
ClustalOmega	EMBL-EBI [292]

COOT (0.9.6.1)	[293]
I-TASSER	Zhang Lab [294]
NetNGlyc (1.0)	DTU Health Tech [295]
NetOGlyc (4.0)	DTU Health Tech [296, 297]
Phaser MR	Collaborative Computational Project number 4 7.1.002 Suite [298]
Phenix 1.16	[299]
PROSITE	Swiss institute of bioinformatics [300]
Protparam	Swiss institute of bioinformatics [301]
Pymol (version 2.5)	DeLano *Scientific* LLC [302]
REFMAC	Collaborative Computational Project number 4 7.1.002 Suite [303]
SMART	EMBL [304]
Swissmodel	Swiss institute of bioinformatics [305]
XDS, XSCALE,XDSCONV, XDSSTAT	[306]

3.3.6 Cells and Culture Medium

Mammalian Cells and Culture Medium

Cells

Expi293F™ GnTI- cells	Themo Fisher
Freestyle293F cells	Thermo Fisher

Medium

Expi293™ expression medium	Thermo Fisher
Freestyle™ expression medium	Thermo Fisher
Opti-MEM™ I reduced serum medium	Thermo Fisher
Opti-MEM SFM	Thermo Fisher

Bacteria Cultivation

Cells

DH5α

Table 3.1: LB-medium and LB-agar plates for bacteria cultivation.

	LB-medium	LB-agar
Tryptone	1% (w/v)	1%(w/v)
Yeast Extract	0.5% (w/v)	0.5%(w/v)
Agar		6%(w/v)
NaCl	1% (w/v)	1%(w/v)

3.3.7 Buffers

Table 3.2: Buffer used for the immobilised metal ion chromatography.

Buffers	His-A1 buffer	His-A2 buffer	His-B1 buffer	His-B2 buffer
Trizma base pH 7.4	50 mM	50 mM	50 mM	50 mM
NaCl	500 mM	500 mM	500 mM	500 mM
Glutamate		50 mM		50 mM
Arginine		50 mM		50 mM
Imidazole	5 mM	5 mM	500 mM	500 mM

Table 3.3: Composition of the buffers used for dialysis and size exclusion chromatography.

Buffers	Dialysis buffer	SEC buffer	Low-salt Buffer	High-salt Buffer
HEPES pH 7.4		20 mM		
Trizma base pH 7.4	50 mM		20 mM	20 mM
NaCl	150 mM	150 mM	50 mM	1 M

Table 3.4: Composition of the buffers used for Con A affinity chromatography.

Buffers	Con A binding Buffer	Con A elution Buffer
Trizma base pH 7.4	20 mM	20 mM
NaCl	500 mM	500 mM
$CaCl_2$	1 mM	1 mM
$MgCl_2$	1 mM	1 mM
$MnCl_2$	1 mM	1 mM
Methyl-α–D-mannopyranoside		12.5% (w/v)

Table 3.5: Composition of the western blot buffer and the TBS-T buffer. The buffers were used to perform a western blot by a semi-dry method.

Buffers	Western blot buffer	TBS-T buffer
Trizma base pH 7.4	25 mM	20 mM
Gycine	190 mM	
Methanol	20% (v/v)	
NaCl	150 mM	150 mM

Table 3.6: Composition of the surface plasmon resonance buffer solutions.

Buffers	Running	Activation	Washing	Regeneration
HEPES pH 7.4	20 mM	20 mM	20 mM	20 mM
NaCl	150 mM	150 mM	150 mM	150 mM
EDTA pH 7.4		50 mM	3 mM	350 mM
Tween-20	0.005% (v/v)	0.005% (v/v)	0.005% (v/v)	
$NiSO_4$		50 µM		

3.3.8 SDS-Acrylamide Gel and Sample Buffer

Table 3.7: SDS-acrylamide gels used for protein analysis. SDS-acrylamide separating gels were prepared with 12% and 15% acrylamide content for proteins with a molecular weight above 40 kDa and below 40 kDa, respectively.

	Separating gel 12%	Separating gel 15%	Stacking gel 3.85%
Rotiphorese® Gel 30%	6 mL	7.5 mL	1.3 mL
1.5 M Tris/HCl pH 6.8			2.5 mL
1.5 M Tris/HCl pH 8.8	3.75 mL	3.75 mL	
10% SDS (w/v)	150 µL	150 µL	100 µL
TEMED	10 µL	10 µL	7.5 µL
10% APS (w/v)	150 µL	150 µL	100 µL
H_2O	25 mL	17.5 mL	20.3 mL

Table 3.8: Sample buffer composition used for SDS-PAGE sample preparation.

	$4 \times$ Sample Buffer
Glycerol	30 mL
SDS (10%)	10 mL
EDTA (0.5 M, pH 8.0)	1.63 mL
β-Mercaptoethanol	4 mL
Bromphenol blue	20 mg
H_2O	13 mL

3.4 Methods

3.4.1 Molecular Biology

In-fusion® HD Cloning

For vector linearisation, a PCR was performed with primers complementary to the vector sequence. For amplification of the insert, primers were used that are complementary to the insert sequence and comprise additional 15 base pairs (bp) complementary to the vector sequence as suggested by the manufacture protocol of the In-fusion® HD cloning kit. Primers and templates are listed in appendix table A.1 and the PCR content is summarised in table 3.9. PCR was performed according to the scheme in table 3.10. PCR products were purified by gel electrophoresis and gel extraction from 0.5% and 1% agarose gel for vector and insert, respectively, according to the manufacture protocol of the GenElute™ PCR Clean-Up kit. For ligation, 100 ng insert, 100 ng vector and 2 µL In-fusion® HD Enzyme Premix were mixed in a total volume of 10 µL and incubated at 50 °C for 15 min, followed by transformation into chemical competent bacteria (DH5α).

Site-directed Mutagenesis

Primers were designed complementary to the sequence up and down stream of the region of interest, excluding the region to mutate or remove. For the insertion of point a mutation or small sequences up to 10 bp, overhangs were added to the 5'-end of the forward primer. All primers and templates are listed in appendix table A.1. The PCR reaction mix is listed in table 3.9 and the PCR was performed according to table 3.10. For the template digest, 1 µL *Dpn*I was added to the PCR product and incubated at 37 °C for 2 h. In the

next step, the blunt ends of the synthesised DNA were phosphorylated by mixing 2 µL PCR product with 0.5 µL 5× T4 DNA ligase buffer, 0.5 µL T4 DNA PNK and 2 µL water. The reaction mixture was incubated at 37 °C for 30 min. In the following, phosphorylated DNA was ligated at 22 °C for 2 h, by adding 0.5 µL 5×T4 DNA ligase buffer, 0.5 µL T4 DNA Ligase and 3 µL water. Up to 5 µL of this ligation mixture was used to transform 50 µL chemical competent bacteria (DH5α).

Table 3.9: PCR mix for In-fusion® HD cloning and site-directed mutagenesis. The complementary primers (appendix table A.1) include overhangs, complementary to the vector sequence, for In-fusion cloning and insertion of point mutation.

PCR mix	In-fusion cloning	site-directed mutagenesis
template (100 ng/µL)	0.5 µL	0.5 µL
fwd primer (10 mM)	0.3 µL	1 µL
rev primer (10 mM)	0.3 µL	1 µL
dNTP's		200 µM
5 × Q5 reaction buffer		10 µL
In-fusion x	12.5 µL	
Q5 polymerase		0.5 µL
H$_2$O	up to 25 µL	up to 50 µL

Table 3.10: PCR protocol for In-fusion cloning and site-directed mutagenesis. Annealing temperature are adjusted for each pair of primer and can be found in appendix table A.1.

	In-fusion cloning		site-directed mutagenesis	
initial denaturation			98 °C	30 s
35 cycles of				
denaturation	98 °C	10 s	98 °C	10 s
annealing	55 °C	15 s	50-72 °C	30 s
elongation	72 °C	30 s	72 °C	180 s
final elongation			72 °C	300 s
storage	4 °C	∞	4 °C	∞

Transformation of Chemical Competent Cells

Chemical competent *E. coli* DH5α (50 µL) were thawed on ice, before adding 1 µL plasmid DNA and incubation for 15 min. A heat-shock was performed to transform the chemical competent cells. For this, the cells were heated for 45 s to 42 °C, followed by an incubation for 2 min on ice. The cells were supplemented with 500 µL LB-medium and incubated at 37 °C and 180 rpm on an orbital shaker for 1 h. For selection of positive clones, 50 µL cell suspension were streaked onto an agar plate containing the selection marker corresponding to the encoded resistance of the plasmid (appendix table A.2) and the plates were incubated at 37 °C over-night (ON).

Plasmid Production and Purification

For production of plasmid DNA, clones were picked from the agar plate to inoculate 5 mL LB-media supplied with selection marker. Cells were cultivated at 37 °C for 16 h for a small scale (miniprep) and for 8 h for a large-scale culture (maxiprep). For a maxiprep, 250 µL of the pre-culture was used to inoculate 450 mL LB-media with selection marker and incubated at 37 °C for 16 h. On the following day, the cells were harvested by centrifugation

(10 min, 9173×g), the supernatant was discarded and plasmid DNA was extracted from the bacterial pellet according to the manufacture protocol of GenElute™ HP Plasmid Maxiprep Kit and GenElute™ HP Plasmid Miniprep Kit for the maxi- and miniprep, respectively.

Sequencing

The purified plasmid DNA (80-120 ng/μL) was sequenced with Sanger sequencing at Microsynth Seqlab (Göttingen, Germany) using sequencing primer provided by the company.

3.4.2 Cell Culture and Protein Production

If not stated differently, FreeStyle™ 293-F cells and Expi293F™ GnTI- cells, modified human embryonic kidney cells (HEK293F) cells, were cultivated in FreeStyle™ and Expi293™ expression medium, respectively (section 3.3.6). All following cultivation steps were carried out at 37 °C, 8% CO_2, 125 rpm in an Infors HT celltron shaker installed into a HERA cell 150 incubator.

Cryopreservation of Cells

Cultivated cells with a viability of >95% were spun down by centrifugation (300×g for 5 min), the supernatant was discarded and the cells were resuspended in pre-cooled medium supplemented with 10% (v/v) DMSO, to a final cell density of 1×10^7 cells/mL. The 1 mL cell vials, stored in Mr. Frosty®, were transferred to -20 °C for 3 h, followed by -80 °C, ON. For long-term storage, the vials were transferred to -180 °C.

Freestyle™ 293-F Cell

Cells (1×10^7), stored at -180 °C, were thawed at 37 °C and resuspended in 30 mL medium. Every three to four days, the cells were counted and split to a density of 0.1-0.2×10^6 cells/mL. For transfection, the culture, which was grown to a density of approx. 1.2×10^6 cells/mL, was diluted to 1.0×10^6 cells/mL. For each mL of culture to transfect, 1 μg DNA and 3 μg PEI were diluted separately in 200 μL Opti-MEM SFM, before both solutions were mixed and incubated at 37 °C for 20 min to form a lipid-DNA complex. Then, the transfection reagent mix was added to the cells.

Expi293F™ GnTI- Cell Cultivation and Transfection

For cultivation, 1.0×10^7 cells were thawed and resuspended in 30 mL medium and split to 0.3-0.4×10^6 cells/mL every three to four days. For transfection, cells were grown to a density of 5-6×10^6 cells/mL and diluted to 3×10^6 cells/mL. Transfection was performed according to the manufacture protocol for the Expifectamine transfection kit. DNA and Expifectamine were diluted in Opti-MEM™ I reduced serum medium according to the table in the Expi293™ cell guide and incubated for 5 min at RT. Both reagents were mixed and incubated for 15 min at RT, before Expifectamine/DNA mix was added to the cells. 16-22 h post transfection, Enhancer I and Enhancer II were added according the manufacture protocol and the cells were cultivated for additional three to seven days.

3.4.3 Protein Purification

If not stated differently, the following steps are all carried out at 4 °C with freshly filtered and degassed buffers that are listed in tables 3.2, 3.3, 3.4. The pH of the buffer was adjusted at 4 °C. Proteins, produced in Freestyle™ 293-F and Expi293F™ GnTI- cells (section 3.4.2) were secreted into the expression medium, due to the secretion signal at the N-terminus of the protein. For purification of the expressed protein, the medium was separated from the cells by centrifugation with a SS34 rotor (20318×g, 10 min).

Immobilised Metal Affinity Chromatography

The supernatant was supplemented with 5×His-A buffer stock solution to obtain a final concentration of 1×His-A buffer (table 3.2), prior to cyclically loading with a peristaltic pump onto a 5 mL Histrap crude column with a flow rate of 1 mL/min, ON. For analysis of the cell content, the cell pellet was resuspended in 1×His-A buffer and the cells were lysed by sonication with an amplitude of 20%, a pulse duration of 0.5 s and 2.5 s pause over a total pulse of 3 min. The insoluble part were spun down with a SS34 rotor (34541× g, 40 min) and the supernatant was loaded onto a HisTrap FF crude column as described above. On the next day, the column was installed into an Äkta Prime Plus system and washed with 10 column volume (CV) His-A buffer, followed by two washing steps with 10 CV of His-A buffer supplemented with 5% and then 10% His-B buffer (table 3.2). The protein was eluted with 60% His-B buffer and protein containing fractions were analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE).

Concanavalin A Affinity Chromatography

For Siglec-11$_{1D\Delta tag}$, the supernatant, obtained from the cell culture, was supplemented with 5× Con A-binding buffer to obtain a final concentration of 1× Con A-binding buffer. The protein was cyclically loaded via an peristaltic pump onto a HiTrap Con A (Concanavalin A) column with 0.5 mL/min for 16 h. For elution, the column was installed into an Äkta Prime Plus system and washed with 10 CV Con A binding buffer (table 3.4) to remove unspecifically-bound proteins. The target protein was eluted with 100% Con A elution buffer (table 3.4) in a stop-and-go fashion with a constant flow of 0.5 mL/min for 5 mL, intermittent by 10-20 min incubation steps.

Dialysis

Target protein containing fractions were pooled. IMAC-derived samples were dialysed against dialysis buffer to remove the imidazole in the protein sample (table 3.3). For Siglec-11$_{1D\Delta tag}$, samples were dialysed against a low-salt buffer to adjust the salt concentration in the buffer for ion exchange chromatography (IEX) (table 3.3).

Ion Exchange Chromatography

The dialysed Siglec-11$_{1D\Delta tag}$ was loaded via an Äkta Prime Plus system onto an 1 mL HiTrap SP column with a flow-rate of 0.4 mL/min. The column was washed with 10× low-salt buffer (table 3.6), followed by a gradient washing step from 64 mM to 100 mM

NaCl. The target protein was eluted with 240 mM NaCl, all remaining proteins, that were bound to the column, were eluted with 1 M NaCl.

Glycosidase and Protease Treatment

For crystallisation, the *N*-linked glycosylations on the protein surface were trimmed with EndoH$_f$ (6 U/mg protein) or removed with 1 μL PNGaseF at 8 °C ON, prior to size exclusion chromatography (SEC). In addition, deglycosylation of the protein was performed on an analytical scale. 15 μg dialysed protein was supplemented with 2 μL GlycoBuffer 2 and 2 μL PNGaseF, or GlycoBuffer 3 and 2 μL EndoH$_f$ in a volume of 20 μL. The reaction was incubated at either 4 °C or 20 °C, ON. The resulting samples were analysed by SDS-PAGE (section 3.4.4).

Tag-cleavage was performed simultaneously to dialysis. For this, dialysis buffer was supplemented with 2 mM DTT and 200 μg tobacco etch virus (TEV) protease was added to the protein solution and incubated ON at 8 °C.

Size Exclusion Chromatography

A SEC was performed to remove aggregated protein, EndoH$_f$ and further impurities. For this, the dialysed protein was concentrated to a volume of 2.5 mL with an Amicon ULTRA concentrator (molecular weight cut off 3 kDa and 10 kDa for Siglec-11$_{1D\Delta myc}$/Siglec-11$_{small}$ and Siglec-11$_{middle}$/Siglec-11$_{large}$, respectively). Depending on the protein size, the SEC was performed with SEC buffer with either an HiLoad Superdex 75 16/600 or an HiLoad Superdex 200 16/600 column installed on an Äkta Basic system.

3.4.4 Protein Analytics

SDS-PAGE

Samples, obtained during expression and purification, were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). For this, the samples (30 μL) were mixed with 10 μL 4×SDS sample buffer (table 3.7), vortexed and heated for 5 min at 95 °C. Insoluble particles were removed by centrifugation (5 min, 16000×g, RT) and 8-15 μL sample was loaded onto a 12-15% SDS-polyacrylamide gel (table 3.8), depending on the content. Electrophoresis was performed in 1×SDS-PAGE buffer applying 200 V for 60 min. For visualisation, the protein was stained with InstantBlue® Coomassie protein stain for 10 min.

Western Blot

A western blot was performed to identify the protein of interest in samples during expression and purification. The polyacrylamide gel, obtained in the SDS-PAGE (section 3.4.4), was blotted onto a PVDF membrane by the semi-dry transfer method. For this, four blotting paper, a PVDF membrane, the acrylamide gel and additional four blotting paper, were soaked with transfer buffer and stacked. The membrane was plotted with 20 V for 60 min and blocked with TBS-T buffer (table 3.5) supplemented with 5% (w/v) milk powder. In

the following, the membrane was decorated with primary antibodies (mouse α-c-myc or mouse α-His antibodies) diluted 1:5000 in TBS-T buffer, supplemented with 5% (w/v) milk powder, at 4 °C, ON. The next day, the membrane was washed four times with 10 mL TBS-T, before 5 mL of TBS-T buffer (supplemented with 5% (w/v) milk powder) with diluted (1:5000) secondary horseradish peroxidase (HRP)-associated goat α-mouse antibodies were added to the membrane and incubated for 1 h at RT. The membrane was finally washed with 50 mL TBS-T buffer for 40 min. For visualisation of the secondary antibodies by chemiluminescence, ECL™ western blotting developmental solution 1 and 2 were mixed in a 1:1 ratio and added to the membrane, before images were taken at 428 nm for 1-5 min in a ChemiDoc™ MP imager.

Circular Dichroism Spectroscopy

The total content of secondary structure elements in a protein can be determined by circular dichroism spectroscopy (CD). Minima at 208 nm and 222 nm in the spectra indicate an α-helical content in the protein, β-strands are represented with a maximum at 195 nm and a minimum at 225 nm, whereas random-coils show a minimum at 198 nm. The resulting CD spectrum represents the sum of all structural component in a protein. For measurements, the protein (2 mg/mL) was diluted in water to 0.15 mg/mL and the ellipticity was measured with an J-720 spectrometer seven times in the range of 195-250 nm. For determination of the background signal, caused by buffer components, the buffer was diluted in the same ratio than the protein and measured four times in the range of 195-250 nm.

Analytical Size Exclusion Chromatography

An analytical size exclusion chromatography run was performed to determine the oligomeric state, possible aggregates or instability of the protein sample. 30 μL of concentrated (1-8 mg/mL) and filtered (0.22 μm filter) protein were applied to a SD200 3.2/300 increase column installed into an Äkta Micro (Ettan LC) system. The SEC was performed with SEC buffer (table 3.2) and a flow rate of 0.05 mL/min. The protein was detected by UV absorbance at 280 nm and protein containing fractions were analysed by SDS-PAGE (section 3.4.4).

3.4.5 Affinity Measurements

Surface Plasmon Resonance

Protein (2.7 mg/mL) was diluted to 0.005 mg/mL with SPR running buffer (table 3.3) to capture onto a Ni-NTA sensor chip resulting in an immobilisation level of 330 RU. Initial binding experiments were performed at 25 °C with 2 mM colominic acid dissolved in SPR running buffer and a flow rate of 30 μL/s. The experiment was performed as following, the sensor chip was first activated with activation buffer (table 3.6) for 30 s, before Siglec-$11_{1D\Delta myc}$ in running buffer was applied two times for 30 s. The analyte, colominic acid, was injected for 60 s to determine association of the complex and dissociation was measured for 540 s. Finally, the sensor chip was regenerated with regeneration buffer (table 3.6).

3.4.6 Structure Determination

Crystallisation

Glycosylated Siglec-11$_{1D\Delta myc}$ with a concentration of 5.2 mg/mL and 7.9 mg/mL, and EndoH$_f$ treated Siglec-11$_{1D\Delta myc}$ with a concentration of 7.9 mg/mL and 10 mg/mL were filtered (0.22 μm filter) for crystallisation. In addition, Siglec-11$_{1D\Delta tag\Delta glyc}$ was concentrated to 7.0 mg/mL and filtered (0.22 μm filter) for crystallisation. Initial crystallisation trials were performed in 96-well plates by sitting drop vapour diffusion. The plates were set up with a Gryphon pipetting robot and commercially available screens (section 3.3.2). Screen solution (50 μL) was added in the reservoir, and a 400 nL drop, comprising a 1:1 mixture of the reservoir and protein solution, was pipetted into the well chamber. For crystal growth, the plates were either stored at 4 ℃ or 20 ℃. To obtain crystals measuring a size of < 50 μL, crystallisation conditions containing crystalline structures were further refined (section 3.3.2). Fine screens for Siglec-11$_{1D\Delta myc\Delta glyc}$ crystallisation were set up with a protein concentration of 10 mg/mL in the same procedure described above. For Siglec-11$_{1D\Delta tag\Delta glyc}$, a fine screen (appendix table A.3) of the JCSG+ screen condition 5 (1 M sodium citrate, 0.1 M sodium cacodylate pH 6.5) was set up with a protein concentration of 7 mg/mL and a drop size of 1μL in a 24-well plate by hanging drop vapour diffusion.

Crystal Harvesting, Soaking and Cryo-Protection

All crystals of Siglec-11$_{1D\Delta myc\Delta glyc}$ that were used for X-ray diffraction experiments were obtained from a fine screen (FS_CS_A2 condition B2 and C7). For cryoprotection, crystals were transferred into a fresh drop containing mother liquor supplemented with 20% (v/v) PEG 200 for 5 min. For the apo structure, Siglec-11$_{1D\Delta tag\Delta glyc}$ crystals, grown in the JCSG+ condition 5, were transferred into a fresh drop containing 1 M sodium citrate, 0.1 M sodium cacodylate pH 6.0 and 20% (v/v) glycerol for cryo-protection. For a ligand structure, crystals that were obtained from the fine screen of JCSG+ condition 5 were harvested and soaked with ligand. For this, the crystals were transferred into a substrate containing drop composed of 20 mM α2,8-linked oligosialic acid (DP=3), 1 M sodium citrate, 0.1 M sodium cacodylate pH 6.5 and incubated ON at 20 ℃. The next day, crystals were transferred into a fresh drop with 20 mM α2,8-linked oligosialic acid (DP=3), 1 M sodium citrate, 0.1 M sodium cacodylate pH 6.5 and 20% (v/v) glycerol for cryo-protection. After 5 min incubation, crystals were mounted into 0.05 - 0.2 mm loops and mother liquor was removed as much as possible, prior to flash freezing in liquid nitrogen.

Data Collection

All test shots and data sets of apo Siglec-11$_{1D\Delta myc\Delta glyc}$ crystals and Siglec-11$_{1D\Delta tag\Delta glyc}$ crystals soaked with oligoSia were collected at beam line X06DA at a wavelength of 1 Å. Data sets and test shots for apo Siglec-11$_{1D\Delta tag\Delta glyc}$ crystals were collected at the beamline X06SA of the Swiss Light Source (SLS) in Villigen (Switzerland). Siglec-11$_{1D\Delta myc\Delta glyc}$ apo crystals and Siglec-11$_{1D\Delta tag\Delta glyc}$ crystals soaked with oligoSia were tested with a beam

transmission of 100%, exposure time of 0.2 s per image and an increment of 90° between the single images to obtain spots on the a PILATUS 2M-F detector. For a data set, images of 240° were collected with an exposure time of 0.1 s per image and an increment of 0.1° between the single images. At beam line X06SA, a micro beam that possesses a higher beam intensity is used for data collection. Therefore, beam transmission was adjusted to 10%, to reduce radiation damage and two test images with an increment of 90° were collected on an EIGER X 16M detector and an exposure time of 0.2 s. For a data set, images of 360° were recorded with an increment of 0.1° and an exposure time of 0.1 s per image. Beam transmission was adjusted to 8% to further reduce radiation damage.

Data Processing

The obtained data was processed with the XDS package [306], which includes indexing, reducing and scaling of the data. In the first step, the background noise on the detector was defined to be able to distinguish between noise and weak spots on the detector. In the following, the position of each collected spot was specified and converted into a reciprocal lattice vector and possible primitive crystal lattices were extracted from this spot list. The intensity of each recorded spot, which corresponds to the amplitude of the diffracted and collected X-rays, was assigned in the integration step. In the correct step, values of the indexing and integration step were refined. The calculated unit cell lattice was used and possible point groups were tested to determine the space group of the data set. Reflections were merged according to the applied symmetry of the chosen space group. The resulting data was analysed according to the resolution dependent completeness of the data, the signal- to-noise ratio $(I/\sigma(I))$, $CC_{1/2}$ and R_{meas}. $CC_{1/2}$ is obtained by randomly splitting the data set in two halves and determining the Pearson correlation coefficient. Overall low R_{meas}, as well as high $I/\sigma(I)$ and $CC_{1/2}$ indicate good data quality. The resolution limit was determined for a data set based on $I/\sigma(I)$ and $CC_{1/2}$, which had to be above 1.0 and above 55%, respectively, and the completeness of the data had to be above 90% for the highest resolution shell. In the last step of processing, XDSCONV was used to convert the list of spots with their structure factor amplitudes into a new file format, the mtz format, which was used for the following structure determination and refinement procedure. At the same time, around 1500 spots were separated from the working set and marked in the new file as R-free flags for structure validation during refinement.

During collection of a 3-dimensional (3D) wave on a 2-dimensional (2D) detector the phase information is lost. Phases are needed to be able to convert the data collected in reciprocal space into an electron density map in real space from which the structure is built. These phases can be estimated by using an already solved structure as a search model for molecular replacement, performed with PHASER MR. Here, the structure of Siglec-5 was used as a model, as it was the structure with the highest sequence identity to Siglec-11. First, the sequence of Siglec-5 and Siglec-11 were aligned by BLAST and the model was adjusted for the molecular replacement search by trimming all side chains that were not identical to Siglec-11, using CHAINSAW. In addition, long loops, as well as extensions

at the N- and C-terminus were removed. The Matthews coefficient defines the median crystal volume per molecular weight of a protein and helps to estimate the solvent content of the crystal, as well as the number of molecules in the asymmetric unit (ASU). The ASU is the smallest element to describe the crystal content by applying crystallographic symmetry operations. The solvent content normally ranges between 40% and 70%. The information about the number of protomers per ASU, the molecular weight of the single protomer and the modified model structure were used for molecular replacement. The basis of this method is the Patterson function, which uses for its calculation the phase independent absolute square values of the measured intensities as this is proportional to the square of the wave amplitude. The resulting electron density map is of higher symmetry and describes all vectors between single atoms in the molecule. The Patterson map is calculated for the experimental data, as well as for the search model. Both maps are compared first in a 3D rotation search, followed by a 3D translation search by using the solutions obtained from the prior performed rotation search. A Z-score was calculated to estimate the quality of each possible rotation and translation search and a value above 6 indicated a possible solution. The phases of the best solution, found by phaser MR, were used to calculate an initial electron density map for the structure of Siglec-11.

Model Building and Refinement

The phases, used to determine the initial structure, are a rough estimation of the phases measured in the diffraction experiment. For reduction of the introduced model bias, simulated annealing combined with a rigid body refinement was performed by Phenix [299]. In the following, the new calculated electron density map was used to adjust the model, by elongation of the protein chain and addition of amino acid side chains, in a real space refinement carried out with COOT [293]. This new model was used in a reciprocal refinement against the experimental data, performed with REFMAC to adjust the phases and improve the electron density map [303]. The two refinement steps were carried out consecutively, until the protein chain with all amino acid side chains could be built, and water and other solvent molecules were added. Finally, an anisotropic B-factor refinement, using TLS refinement of REFMAC, was performed with three fragments per protein chain. B-factors are the temperature factors and describe the thermal movement and disorder of each atom in the protein. Each refinement round was validated by several factors, such as the root mean square deviations (RMSD) of the bond angles and bond lengths, as well as rotamer fit and Ramachandran plots. The relative difference between the calculated structure factors from the model and obtained structure factors from the measured data are indicated by the R-factors, R-work and R-free. The R-free value was calculated from the 1500 reflections that were separated in the XDSCONV step. The R-work value was calculated from the remaining the data set. Reflections that are used to calculate the R-free value are not used for the electron density map calculation and are therefore not influenced by the model. The value prevents over-fitting of the model and a good model should have R-values $< 30\%$ with a difference of $< 5\%$ between R-work and R-free.

3.5 Results

Contribution

The study was designed by EO. Design, purification and biochemical analysis of all constructs was performed by EO, except for Siglec-11$_{1D\Delta myc\_TEV}$ and Siglec-11$_{1D\Delta tag}$. For Siglec-11$_{1D\Delta myc\_TEV}$ and Siglec-11$_{1D\Delta tag}$, cloning and purification, as well as crystallisation of Siglec-11$_{1D\Delta tag}$ were carried out by Karina Broger (KB) in the course of her master thesis, which was supervised by EO. The structure was determined and analysed by KB and EO. Affinity measurements were planned and carried out by Katja Mindler and EO.

3.5.1 Construct Design

Constructs, used for Siglec-11 expression, are based on the protein sequence of the microglial Siglec-11(4D) splice variant (uniprot: Q96RL6), which exhibits higher affinity towards polySia than the macrophagal Siglec-11(5D) splice variant [279]. For construct design, the protein sequence was aligned with sequences of other, structurally known, Siglecs to determine the border of the domains (Siglec-3 (pdb code: 5ihb), Siglec-4 (pdb code: 5lf5), Siglec-5 (pdb code: 2zg2), Siglec-7 (pdb code: 2hrl) and Siglec-8 (pdb code: 2n7a)). The alignment was compared with results of the online domain prediction tools SMART and I-TASSER, as well as with a model structure of Siglec-11, which was predicted by SWISS-MODEL based on the structure of Siglec-4 (pdb code: 5lf5). Variants of the Siglec-11 were designed comprising all four (Siglec-11$_{large}$), the first three (Siglec-11$_{middle}$), two (Siglec-11$_{short}$) or only the first outermost ECD (Siglec-11$_{1D}$). For Siglec-11$_{large}$ and Siglec-11$_{middle}$, several variants were designed with different domain borders (appendix table A.2). The DNA sequence of the constructs were codon optimised for expression in mammalian cells. Only, the sequence of Siglec-11$_{large\ V4}$ is based on the mRNA sequence of Siglec-11(4D) (NCBI reference: NM_001135163.1).
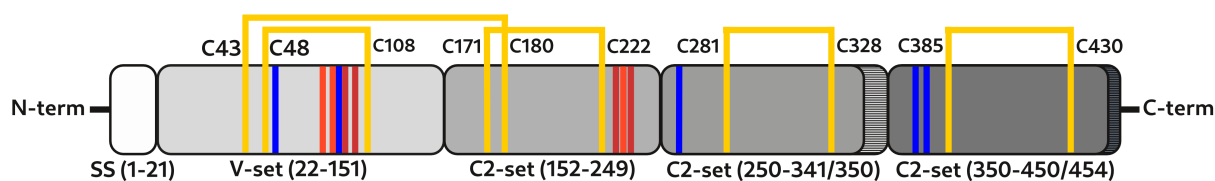

Figure 3.6: Ectodomain of Siglec-11 with highlighted disulphide bridges, *N*- and *O*-glycoslyation sites. The single domains, including the secretion signal (SS), of the Siglec-11 ectodomain are schematically represented as chosen for construct design. For Siglec-11$_{middle}$ and Siglec-11$_{large}$ the different domain border are indicated as hatched areas. Predicted disulphide bridges are represented as orange lines with the corresponding cysteines. In addition, five *N*-glycosylation sites at position N49, N84, N256, N360 and N369 are indicated in blue and eight *O*-glycoslyation sites at position T72, T82, S86, S92, S200, T201 and S202 are drawn in red (light red and dark red for threonine ans serine attached *O*-glycans), which were predicted by NetNGlyc and NetOGlyc, respectively.

For the ECD of Siglec-11, five possible *N*-linked and eight *O*-linked glycosylation sites were predicted by NetNGlyc and NetOGlyc, respectively, further five disulphide bounds were determined by PROSITE in Siglec-11 (figure 3.6)[295, 300].

3.5.2 Expression and Purification

Siglec-11$_{\text{large}}$

Siglec-11$_{\text{large V1}}$ contains a part of the transmembrane domain and was only used for cloning of the other Siglec-11 variants (appendix table A.1). Siglec-11$_{\text{large V2}}$ and Siglec-11$_{\text{large V3}}$ were expressed in Freestyle293F cells and each step of the purification was monitored by western blot or SDS-PAGE (appendix table A.2). The expression level of Siglec-11$_{\text{large V2}}$ was low and no protein could be determined in the medium by western blot analysis. The low expression level is most likely due to the missing Kozak sequence upstream to the start codon. Therefore, a Kozak sequence was added in Siglec-11$_{\text{large V3}}$ to increase the expression level.
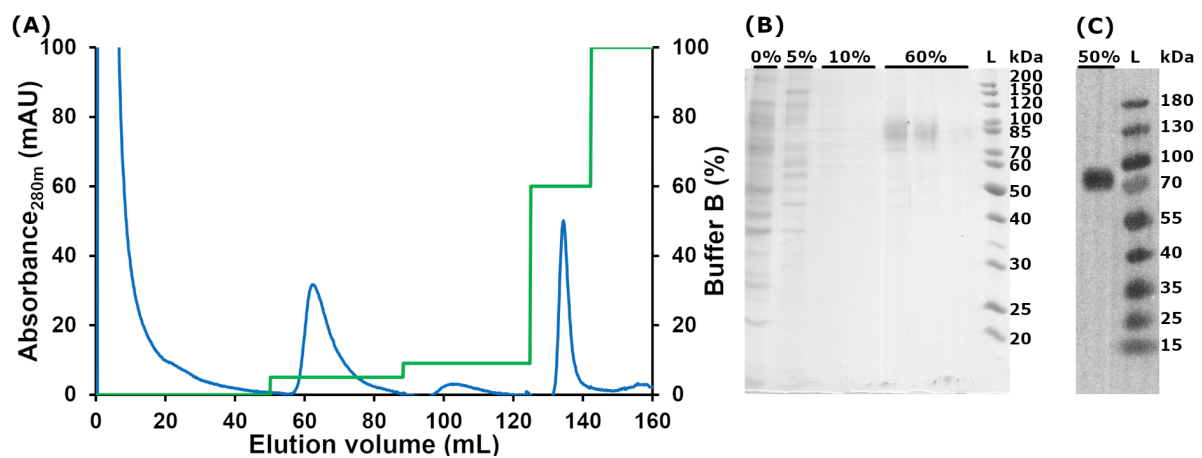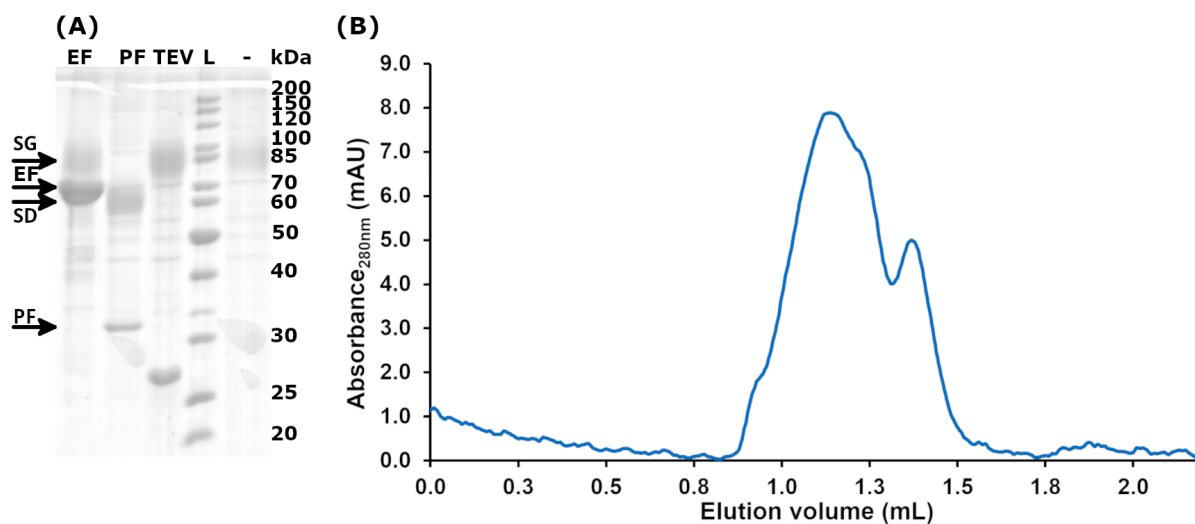

Figure 3.7: IMAC of Siglec-11$_{\text{large V3}}$ expressed in Freestyle293F cells. (A) Chromatogram with the **(B)** corresponding SDS-PAGE and **(C)** western blot are represented. In the chromatogram, UV absorption at 280 nm and the Buffer B concentration are drawn as a solid blue and a green line, respectively. The column was first washed with 30 mM (5% His-B1 buffer) and 55 mM imidazole (10% His-B1 buffer) to remove unspecific-bound protein, before the target protein was eluted with 300 mM imidazole (60% His-B1 buffer). Protein containing fractions were analysed by SDS-PAGE and presence of target protein was verified by western blot. For evaluation of the protein size, unstained protein ladder (L) or prestained protein ladder (L) was added to the SDS-PAGE or western blot, respectively. Target protein was detected with a primary α-c-myc antibody in the western blot.

The medium of Freestyle293F cells, transfected with Siglec-11$_{\text{large V3}}$, was harvested six to seven days after transfection by separating the cells from the medium via centrifugation. The target protein was not visible in the SDS-PAGE, but could be detected by western blot analysis in the medium, as well as the cell pellet (appendix figure A.3). The protein appears in the medium as a diffuse band of approx. 85 kDa and as a more prominent, defined band of approx. 65 kDa in the pellet. For affinity chromatography, the cell culture supernatant was loaded onto a HisTrap-column overnight and the protein elution was performed stepwise with 25 mM and 50 mM imidazole washing steps, before the protein was eluted with 300 mM imidazole (His-A1 and His-B1 buffer, table 3.2). In the SDS-PAGE, the target protein appears again as a diffuse band of higher molecular weight with minor impurities. Because the protein has a theoretical molecular weight of 54 kDa, the presence of Siglec-11$_{\text{large}}$ was additionally confirmed by western blot analysis (figure 3.7). The appearance as a diffuse band of higher molecular weight is probably due to diverse

glycans at the five predicted *N*-glycosylation sites in the protein (figure 3.6). To test this hypothesis, the protein was incubated with PNGaseF or EndoH$_f$ under non-denaturing conditions to remove either the *N*-linked glycans completely or trim the sugar tree to a single GlcNAc moiety, respectively. EndoH$_f$ had no effect on the appearance of the protein in the SDS-PAGE, in contrast, after treating the sample with PNGaseF the protein band shifted to approx. 60 kDa but stayed a diffuse band in the SDS-PAGE (figure 3.8). In addition, treatment with TEV protease did not cleave the affinity-tag at the C-terminus of the protein. Enzyme activity of EndoH$_f$ and TEV protease were both confirmed by personal communication.


Figure 3.8: Deglycosylation and analytical SEC of Siglec-11$_{large\,V3}$ expressed in Freestyle293F cells. Protein was concentrated to 0.5 mg/mL for analysis. **(A)** SDS-PAGE of Siglec-11$_{large\,V3}$ shows protein untreated (-), treated with EndoH$_f$ (EF), PNGaseF (PF) or TEV protease (TEV) under non-denaturing conditions overnight. Protein bands of EndoH$_f$ (EF), PNGaseF (PF), glycosylated (SD) and deglycosylated (SD) Siglec-11$_{large\,V3}$ are indicated by an arrow. For evaluation of the protein size, unstained protein ladder (L) was added to the SDS-PAGE. Target protein was detected with a primary α-c-myc antibody in the western blot. **(B)** The chromatogram of the analytical SEC shows that the protein is polydisperse and elutes as several overlapping peaks, starting with a small peak in the exclusion volume of the column (at 0.92 mL), followed by peaks eluted at 1.14 mL, 1.22 mL and 1.37 mL that corresponds to an apparent molecular weight of 520 kDa, 340 kDa and 140 kDa in solution, respectively. Apparent molecular weight of the proteins is based on the elution volume and was determined according to the calibration of the column (appendix figure A.2).

The pooled and dialysed fractions from the IMAC yielded 2 mg protein per litre culture. The protein was concentrated to 0.5 mg/mL and further analysed in a small scale analytical SEC (SD200 increase 3.2/300). There, the protein appears as several overlapping peaks with a small proportion eluting in the exclusion volume of the column. The elution volume of the peaks corresponds to higher oligomeric species with apparent molecular weights of 520 kDa, 340 kDa and 140 kDa (figure 3.8). Determination of the exact oligomeric state, based on the elution volume, is difficult, but Siglec-11$_{large\,V3}$ is apparently not monomeric in solution. As the protein is polydisperse and no single species could be separated by SEC, the samples could not be used for crystallisation or affinity measurements with polySia.

In another approach, Siglec-11$_{\text{large V3}}$ was extracted and purified from the cell pellet, as the main proportion of target protein was determined in the cell pellet by western blot analysis (appendix figure A.3). For this, the cells were resuspended in His-A1 buffer and lysed by sonication. After removing the insoluble parts by centrifugation, the protein was loaded onto a HisTrap column. The column was washed with His-A1 buffer and 50 mM imidazole and eluted with 500 mM imidazole. In the SDS-PAGE, proteins of different sizes were visible, but no prominent band of approx. 65 kDa could be determined in the IMAC elution fractions, even though protein is visible in the western blot (appendix figure A.6).

To achieve a more homogeneous glycoslyation pattern on the protein surface, Siglec-11$_{\text{large V3}}$ was expressed in Expi293F GnTI- cells. These cells have no N-acetylglucosaminyltransferase I (GnTI) activity and produce highly mannosylated, homogeneous *N*-linked glycans, which enable proper folding and processing in the ER and Golgi and can be enzymatically cleaved. The DNA was transfected with Expifectamine into Expi293F GnTI- cells. The expression was monitored by western blot and the amount of soluble protein in the medium decrease between day three and day six (appendix figure A.5). Therefore, the culture was harvested after three days of expression. Again, the protein could not be identified in the SDS-PAGE, but could be detected in a western blot, similar to the observation for Siglec-11$_{\text{large V3}}$ expression in Freestyle293F cells. The medium was loaded onto a HisTrap column and the protein was eluted stepwise with His-B2 buffer. Peak fractions were analysed by SDS-PAGE and the target protein was determined as the main band at a molecular weight of approx. 60 kDa in the elution fractions and confirmed by western blot analysis (appendix figure A.7). In addition to the target protein, the elution fractions contained more impurities in comparison to the IMAC of the expression in Freestyle293F cells (figure 3.7).
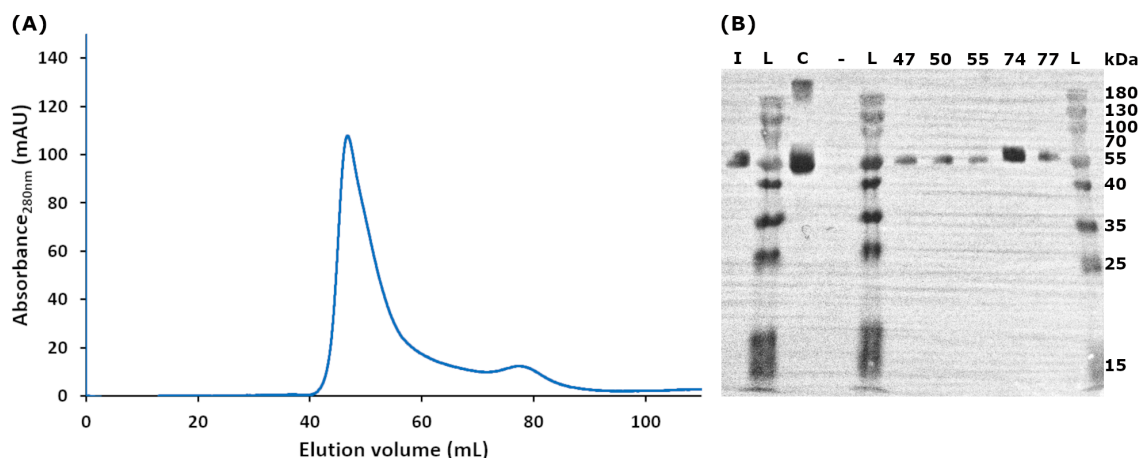
Figure 3.9: Preparative SEC of Siglec-11$_{\text{large V3}}$ expressed in Expi293F GnTI- cells. (A) SEC chromatogram of Siglec-11$_{\text{large V3}}$ and **(B)** the corresponding western blot are represented. A sample of the IMAC elution (I), the concentrated protein (C), which was injected into the SEC, as well as samples of the peak fractions at 47 mL (47), 50 mL (50), 55 mL (55), 74 mL (74) and 77 mL (77) were analysed by western blot. The target protein was detected with a primary α-c-myc antibody and the specificity of the antibody was evaluated with a negative control (-) consisting of purified Siglec-11$_{\text{1D}\Delta\text{myc}}$ protein, which is lacking a c-myc affinity-tag. For evaluation of the protein size, prestained protein ladder (L) was added to the western blot.

In the final purification step, a SEC was performed. The main proportion of the protein elutes in the exclusion volume of the column (volume 45 mL) with a fluent passage towards a small second peak at an elution volume of 76 mL. The protein amount in the elution fraction was not sufficient for analysis by SDS-PAGE. Based on western blot analysis, the target protein started to elute in the exclusion volume, but the main proportion of target protein was determined in the small peak at 74 mL (figure 3.9). Decreasing amounts of target protein in the expression medium over time and the presence of target protein in the exclusion volume of the SEC column indicate that Siglec-11$_{\text{large V3}}$ expressed with a modified glycosylation pattern (Expi293F GnTI- cells) is less stable than the protein expressed with the heterogeneous glycosylation pattern (Freestyle293F cells).
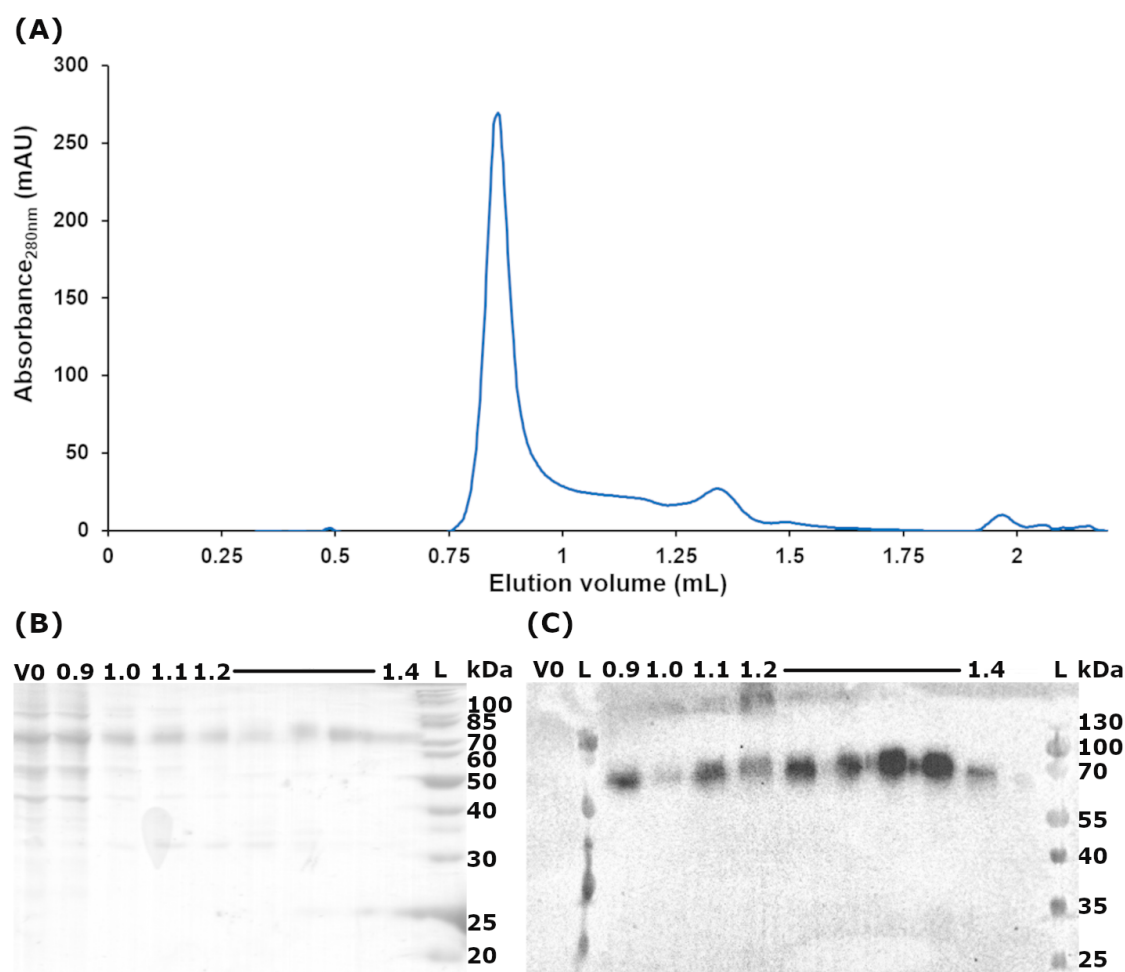

Figure 3.10: Analytical SEC of Siglec-11$_{\text{large V4}}$ expressed in Expi293F GnTI- cells. Protein was concentrated to 30 μL and subjected into an analytical SD200 increase 3.2/300 column. **(A)** Size exclusion chromatogram of Siglec-11$_{\text{large V3}}$, **(B)** the corresponding SDS-PAGE and **(C)** the western blot are represented. Several samples from the exclusion volume of the column (V0) at 0.82 mL to the second peak (1.2-1.4 mL) were analysed by SDS-PAGE and western blot. In the western blot, target protein was detected by chemiluminescence with a primary mouse α-c-myc antibody and a secondary HRP-associated goat α-mouse antibody. For evaluation of the protein size, unstained protein ladder (L) or a prestained protein ladder (L) was added to the SDS-PAGE or western blot, respectively.

The expression and purification protocol of Siglec-11$_{\text{large V3}}$ was applied for the production of Siglec$_{\text{large V4}}$, which is based on the original human-derived DNA sequence of the receptor. The IMAC chromatogram, SEC chromatogram and SDS-PAGE look similar to the results

obtained from Siglec-11$_{\text{large V3}}$ and the protein appears as an approx. 75 kDa protein in the SDS-PAGE (appendix figure A.8). Even with different apparent molecular weights in western blot and SDS-PAGE, analytical SEC showed a similar elution profile for Siglec-11$_{\text{large V3}}$ and Siglec-11$_{\text{large V4}}$, with the main protein fraction starting to elute as peak in the exclusion volume, then forming a shoulder that leads into a second small peak at 1.34 mL. Fractions of the second peak at 1.2-1.4 mL contained one major protein species with 75 kDa, which could be determined as the target protein in the western blot. In all samples, besides the exclusion volume of the column, the target protein could be detected by western blot (figure 3.10). Peak fractions (1.3-1.4 mL) were pooled and concentrated to reinject into the analytical SEC to validate a stable protein species. The protein elutes as a single peak at 1.3 mL, which corresponds to an apparent molecular weight of approx. 120 kDa and a dimer in solution (appendix figure A.9). The yield was not sufficient to determine the protein concentration or perform further experiments with the purified dimeric Siglec-11$_{\text{large V4}}$.
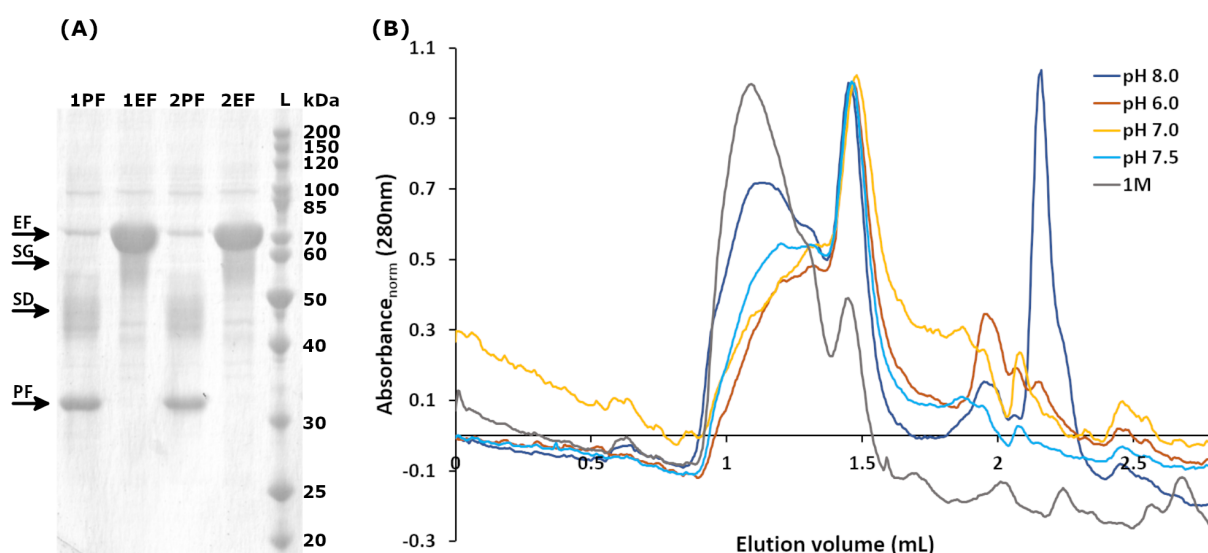
Siglec-11$_{\text{middle}}$

Siglec-11$_{\text{middle}}$ was produced in Freestyle293F and Expi293 GnTI- cells. In Freestyle293F cells, the Siglec-11$_{\text{middle}}$ constructs yielded similar results to the Siglec-11$_{\text{large}}$ construct. The Siglec-11$_{\text{middle V1}}$, which is lacking the Kozak sequence, could only be expressed in very low amounts detectable by western blot analysis after enrichment by IMAC (appendix figure A.10).
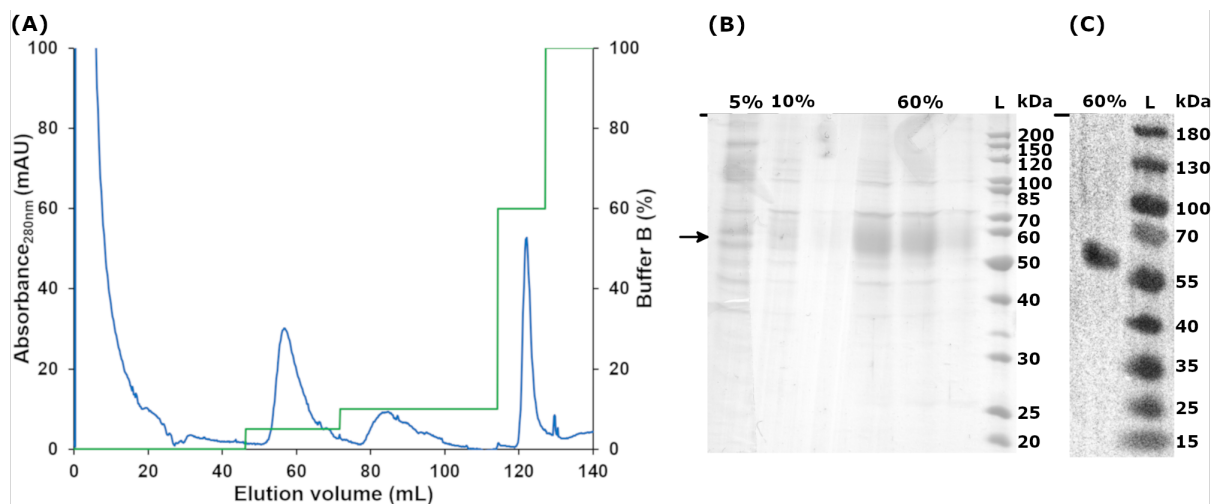
Figure 3.11: IMAC of Siglec-11$_{\text{middle V2}}$ expressed in Freestyle293F cells. **(A)** The chromatogram, **(B)** the corresponding SDS-PAGE and **(C)** the western blot of the IMAC are shown. In the chromatogram, UV absorption at 280 nm and the Buffer B concentration are drawn as solid blue and green line, respectively. The column was washed with 30 mM (5%) and 55 mM (10%) and the target protein was eluted with 255 mM imidazole (60%). Protein containing fractions were analysed by SDS-PAGE and presence of target protein was verified in the elution fraction by western blot. The band that corresponds to Siglec-11$_{\text{middle V2}}$ is indicated by an arrow in the SDS-PAGE. In the western blot, target protein was detected by chemiluminescence with a primary mouse α-c-myc antibody and a secondary HRP-associated goat α-mouse antibody. For evaluation of the protein size, unstained protein ladder (L) or prestained protein ladder (L) was added on the SDS-PAGE or western blot, respectively.

Expression and purification of Siglec-11$_{\text{middle V2}}$ and Siglec-11$_{\text{middle V3}}$ were very similar, therefore the purification is exemplary shown for Siglec-11$_{\text{middle V2}}$. Siglec-11$_{\text{middle V2}}$ was expressed for six days in Freestyle293F cells, before medium and cells were separated by centrifugation and analysed by western blot (appendix figure A.3). The medium contained low amounts of target protein, which appears as a diffuse band with a larger molecular mass of approx. 60 kDa, instead of the theoretical molecular weight of 43 kDa, similar to results observed for Siglec-11$_{\text{large V2}}$. Higher amount of protein, with a defined band and lower molecular weight (approx. 45 kDa) in the western blot, could be detected in the cell pellet (appendix figure A.3). The protein was purified from the medium by an IMAC and yielded approx. 0.1 mg protein per L culture with minor impurities still present. The corresponding SDS-PAGE again shows a diffuse band at 60 kDa for the target protein (figure 3.11). For deglycosylation, the target protein was treated with PNGaseF or EndoH$_{\text{f}}$. Only PNGaseF shifted the diffuse band to a lower molecular mass of approx. 50 kDa in the SDS-PAGE. EndoH$_{\text{f}}$ did not trim the N-linked glycans of the protein, as the band for the target protein did not shift to a smaller molecular weight in the SDS-PAGE (figure 3.12).

Figure 3.12: Deglycosylation and analytical SEC of Siglec-11$_{\text{middle V2}}$ expressed in Freestyle293F cells. (A) SDS-PAGE is represented with samples of Siglec-11$_{\text{middle V2}}$, which were incubated ON with EndoH$_{\text{f}}$ (EF) and PNGaseF (PF) at 4 °C (1EF and 1PF) and 20 °C (2EF and 2PF), respectively. In the SDS-PAGE, band of EndoH$_{\text{f}}$ (EF), PNGaseF (PF), as well as glycosylated (SG) and deglycosylated (SD) Siglec-11$_{\text{middle V2}}$ are indicated by an arrow. For evaluation of the protein size, unstained protein ladder (L) was added to the SDS-PAGE. **(B)** Chromatogram of the analytical SEC shows a polydisperse portein, which elutes as several overlapping peaks with a varying ratio dependent on pH (pH 6.0, 7.0, 7.5, 8.0) and salt concentration (0.15 M-1 M NaCl). Protein starts to elute in the exclusion volume of the column (at 0.92 mL), followed by peaks at 1.09 mL, 1.31 mL and 1.46 mL that corresponds to an apparent molecular weight of 670 kDa, 202 kDa and 90 kDa in solution, respectively. Apparent molecular weight of the proteins is based on the elution volume and was determined according to the calibration of the column (appendix figure A.2).

The protein was further characterised with analytical SEC. The protein elutes as several peaks with apparent molecular weights of approx. 670 kDa, 200 kDa and 90 kDa, which corresponds to different oligomeric states of the target protein. The peak with an apparent

molecular weight of 90 kDa could indicate a dimer in solution. Variation in pH (from pH 6 to 8) or salt concentration (up to 1 M NaCl) in the SEC buffer changed the ratio between the high molecular weight species and the dimer peak, but did not yield a monodisperse sample. Due to the low yield and inhomogeneity of the protein, Siglec-11$_{middle}$ could not be used for further experiments.

Siglec-11$_{short}$

Siglec-11$_{short}$ was expressed in Freestyle293F cells for three days. For Siglec-11$_{short\,V1}$, lacking the Kozak sequence, protein was only detectable in the western blot after enrichment by affinity chromatography (appendix figure A.10). Siglec-11$_{short\,V2}$, containing the Kozak sequence, showed a slightly increased expression level. After six days of expression, the protein was detectable by western blot analysis as a diffuse band at 40 kDa in the medium (appendix figure A.3). Higher amounts of protein were present as a defined band with a molecular weight of 37 kDa in the cell pellet sample (appendix figure A.3). Target protein was extracted from a 300 mL medium by IMAC. In the elution fractions, a small proportion of target protein, was eluted from the column together with impurities (figure 3.13). Due to the low yield and the high amount of insoluble protein in the cell pellet, the Siglec-11$_{short}$ construct comprising two ECDs was not further investigated.
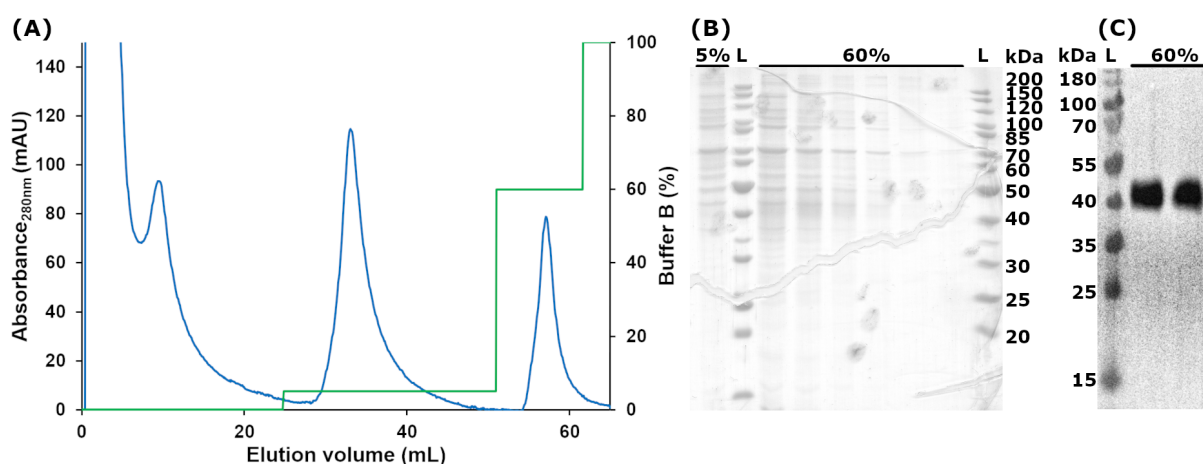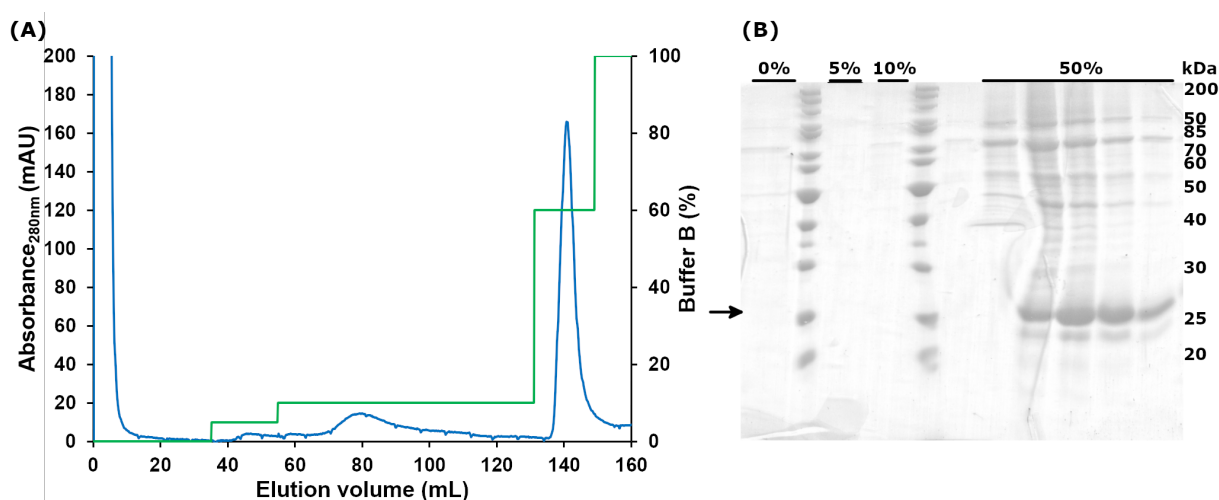

Figure 3.13: IMAC of Siglec-11$_{short\,V2}$ expressed in Freestyle293F cells. (A) The chromatogram, **(B)** the corresponding SDS-PAGE and **(C)** the western blot of the IMAC are represented. In the chromatogram, UV absorption at 280 nm and the Buffer B concentration are drawn as solid blue and green line, respectively. The column was washed with with 30 mM (5%) and target protein was eluted with 300 mM imidazole (60%). Peak fractions were analysed by SDS-PAGE and the target protein was verified in the elution fractions by western blot. In the western blot, target protein was detected by chemiluminescence with a primary mouse α-c-myc antibody and a secondary HRP-associated goat α-mouse antibody. For evaluation of the protein size, unstained protein ladder (L) or a prestained protein ladder (L) were added to the SDS-PAGE or western blot, respectively.

Siglec-11$_{1D}$

The domain border for the Siglec-11 V-set was determined based on the structure analysis of already crystallised Siglec homologous (Siglec-3 (pdb code:5ihb); Siglec-5 (pdb code: 2zg2); Siglec-7 (pdb code: 2hrl); Siglec-8 (pdb code: 2n7a)). Two new constructs were designed, one composing a c-myc- and a His-tag (Siglec-11$_{1D}$), the other has a His-tag (Siglec-11$_{1D\Delta myc}$) at the C-terminus of the protein. Expression of Siglec-11$_{large}$, Siglec-

11_{middle} and Siglec-11_{short} in Freestyle293F cells yielded heterogeneous, not removable glycosylation patterns that are not suitable for crystallisation experiments. Therefore, Siglec-11_{1D} and Siglec-$11_{1D\Delta myc}$, as the best crystallisation candidates, were only expressed in Expi293F GnTI- cells. The cells were transfected with Expifectamine and cultivated for five to seven days, and Siglec-$11_{1D\Delta myc}$ was detectable in the medium and pellet sample by SDS-PAGE and verified by western blot (appendix figure A.4). For purification, the cells were removed by centrifugation, the medium was supplemented with a final concentration of 1×HisA2-buffer and loaded onto a HisTrap column overnight. Unspecifically bound proteins were removed in two washing steps, before the target protein was eluted with 300 mM imidazole (HisB2-buffer). The protein composition of the elution was analysed by SDS-PAGE, and the target protein is visible at a molecular weight of 25 kDa together with various impurities with larger molecular weights in the elution peak (figure 3.14).

Figure 3.14: IMAC of Siglec-$11_{1D\Delta myc}$ expressed in Expi293F GnTI- cells. (A) The chromatogram and **(B)** the corresponding SDS-PAGE of the Siglec-$11_{1D\Delta myc}$ IMAC are represented. In the chromatogram, UV absorption at 280 nm and the Buffer B concentration are drawn as solid blue and green line, respectively. The column was first washed with 5 mM (0% HisA2-buffer), followed by two washing steps with 30 mM (5% HisB2-buffer) and 55 mM (10% HisB2-buffer) imidazole to remove unspecific-bound proteins, before the target protein was eluted with 300 mM imidazole (60%). Peak fractions were investigated by SDS-PAGE and unstained protein ladder (L) was added to the SDS-PAGE to evaluate the protein size. The band that corresponds to Siglec-$11_{1D\Delta myc}$ in the SDS-PAGE is indicated by an arrow.

The protein containing fractions were pooled and dialysed overnight (table 3.3). For crystallisation, glycans were trimmed by $EndoH_f$ simultaneously to the dialysis, which shifted the target protein from 25 kDa to approx. 18 kDa in the SDS-PAGE (figure 3.15). A SEC (with a SD200 column) was performed as the final purification step to remove the remaining impurities, the $EndoH_f$ and aggregated protein. Here, a prominent peak appears in the exclusion volume (42 mL) of the column containing mainly impurities but also small amounts of target protein (figure 3.15). $EndoH_f$ elutes as a second peak at 62 mL, before the target protein appears as symmetric peak at an elution volume of 68-87 mL without significant impurities and a yield of approx. 6 mg pure protein per L culture (figure 3.15). The stability of the glycosylated and deglycosylated ($EndoH_f$ treated) protein was analysed by analytical SEC after two weeks at 4 °C. The fresh and stored

samples are virtual identical fitting to a stable protein in solution. The elution volume of the glycosylated and deglyosylated Siglec-11$_{1D\Delta myc}$ corresponds to an apparent molecular weight of 23 kDa and 17 kDa, respectively (appendix figure A.14). In addition, the CD spectrum of Siglec-11$_{1D\Delta myc}$ was collected to verify the proper fold of the protein. The spectrum composes a maximum at 227 nm and minima at 215 nm and 233 nm, indicating β-sheets in the protein, as expected for an immunoglobulin-like fold (appendix figure A.16).
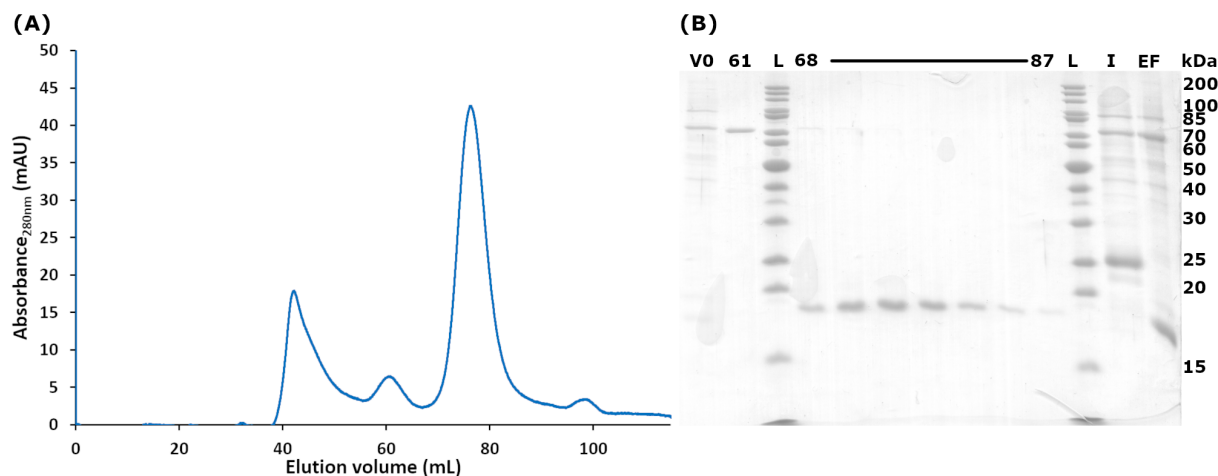

Figure 3.15: Deglycosylation and preparative SEC of Siglec-11$_{1D\Delta myc}$ expressed in Expi293F GnTI- cells. Protein was incubated with EndoH$_f$ ON at 8 °C to cleave the glycosylation, before the protein was concentrated to 2 mL and subjected into an SD200 16/600 column. **(A)** SEC chromatogram of Siglec-11$_{1D\Delta myc}$ and **(B)** the corresponding SDS-PAGE are represented. EndoH$_f$ treated sample (EF) was compared with IMAC elution (I) and samples of each peak in the SEC were analysed by SDS-PAGE (exclusion volume at 42 mL (V0), 61 mL (61) and from 68 mL (68) to 87 mL (87)). For evaluation of the protein size, unstained protein ladder (L) was added to the SDS-PAGE.

Because the affinity-tag could not be cleaved by TEV protease (appendix figure A.12), the construct design was reassessed, and one point mutation in the TEV protease cleavage site (Y to T) was determined to be the reason for issues with enzymatic cleavage of the tag. This point mutation was corrected by mutagenesis using PCR and the new Siglec-11$_{1D\Delta myc\_TEV}$ construct was transfected into Expi293F GnTI- cells. Very low protein expression could be detected by western blot analysis over time (appendix figure A.13). After five days, the medium was harvested and supplemented with His-A2 buffer for loading onto a HisTrap column over night (table 3.2). The IMAC was performed as described for Siglec-11$_{1D\Delta myc}$, but no target protein eluted from the column (appendix figure A.13).

In a new approach, Siglec-11$_{1D\Delta tag}$ was produced by the introduction of a stop codon at the C-terminus of the Siglec-11$_{1D}$ coding sequence. For protein production, Expi293F GnTI- cells were transfected with Siglec-11$_{1D\Delta tag}$ and cultivated for five days. A Con A column, which binds mannosylated glycan trees, was used to extract the target protein from the cell culture medium. For this, the medium was supplemented with Con A binding buffer to adjust the medium to the binding buffer composition of the affinity column (table 3.4). The target protein was eluted with high concentrations of mannose in the Con A elution buffer, but displacement of bound mannosylated glycans were not efficiently displaced from
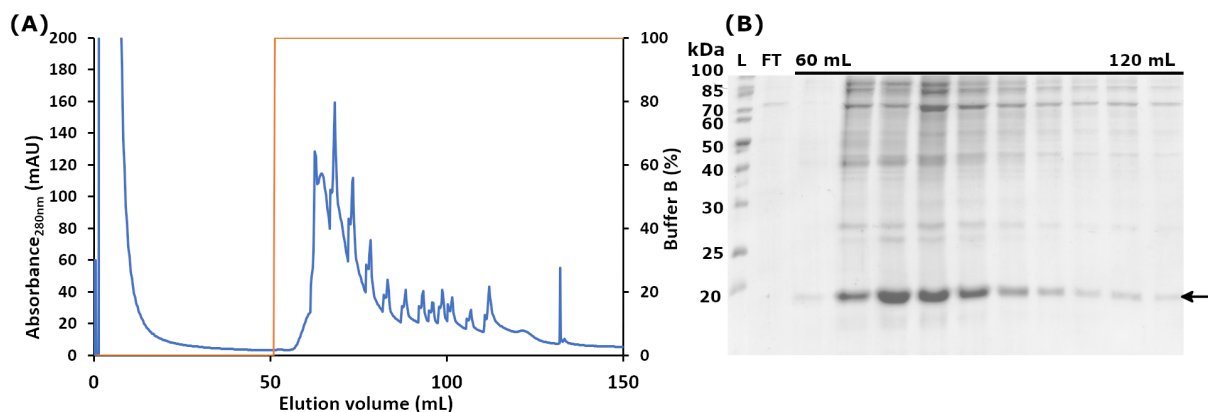
Figure 3.16: **Affinity chromatography of Siglec-11$_{1D\Delta tag}$ expressed in Expi293F GnTI- Cells.**
(A) Chromatogram of the Con A column with the **(B)** corresponding SDS-PAGE are represented. In the chromatogram, UV absorption at 280 nm and the Buffer B concentration are drawn as a solid blue and orange line, respectively. The column was washed with ConA-binding buffer and target protein was eluted with 12.5% (w/v) Mannose (100% Con-A-elution buffer). Protein composition of the column flow-through (FT) and elution fractions (60-120mL) were analysed by SDS-PAGE and unstained protein ladder (L) was added to the gel for evaluation of the protein sizes. The band that corresponds to Siglec-11$_{1D\Delta tag}$ is indicated by an arrow in the SDS-PAGE.

the column by soluble mannose in the elution buffer. Therefore, a stop-and-go procedure was performed to enhance displacement ability of the mannose and recover the target protein from the Con A column (figure 3.16). In the SDS-PAGE of the elution fractions, a dominant band at 20 kDa is visible, which corresponds to the target protein. In addition, two bands at 17 kDa and 27 kDa are present in the sample (figure 3.16).
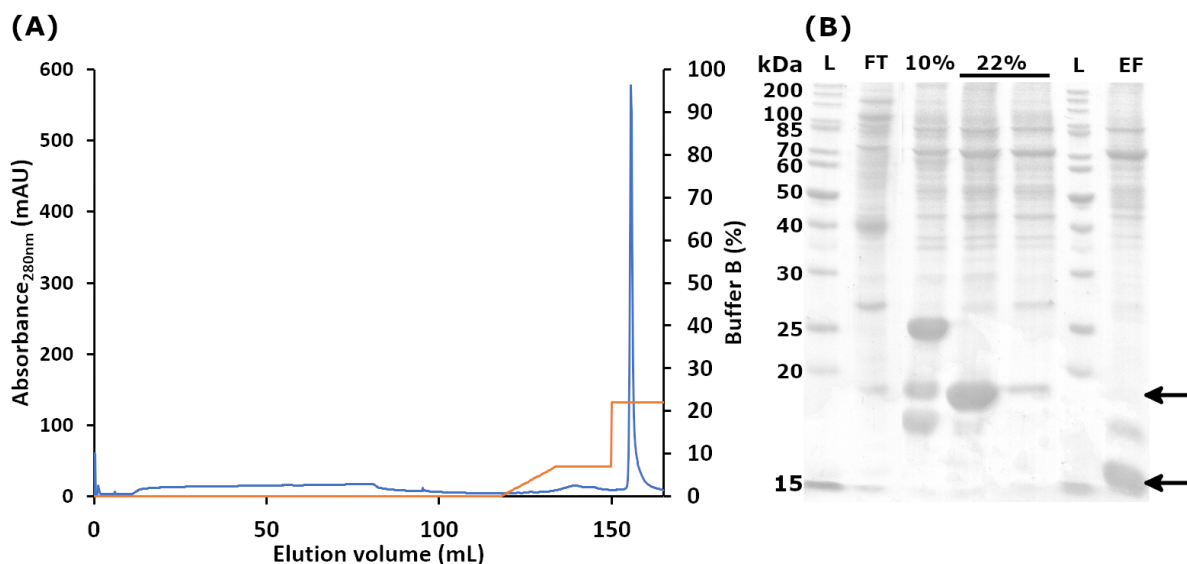
Figure 3.17: **Cation exchange chromatography of Siglec-11$_{1D\Delta tag}$ expressed in Expi293F GnTI- Cells. (A)** The chromatogram of the cation exchange-column with the **(B)** the corresponding SDS-PAGE are represented. In the chromatogram, UV absorption at 280 nm and the Buffer B concentration are drawn as a solid blue and orange line, respectively. The column was loaded with pooled and dialysed ConA-elution fractions, before washing with low-salt buffer and a gradient up to 100 mM NaCl. Target protein was eluted with 240 mM NaCl from the column. Protein composition of 10× concentrated flow-through (FT) and washing step (10%) samples, as well as of the elution sample (22%) were analysed by SDS-PAGE and unstained protein ladder (L) was added to the gel for evaluation of the protein sizes. The band that corresponds to Siglec-11$_{1D\Delta tag}$ is indicated by an arrow in the SDS-PAGE.

A cation exchange chromatography was performed to separate the 17 kDa and 27 kDa impurities from the target protein. The pooled elution fractions of the Con A column were dialysed against low-salt buffer (table 3.3) overnight and loaded onto an HiTrap SP FF column. The 17 kDa and 27 kDa impurities were separated with 100 mM NaCl and target protein was eluted with 240 mM NaCl (figure 3.17). Siglec-11$_{1D\Delta tag}$ containing fractions were pooled, EndoH$_f$ was added and the sample was incubated overnight at 8 °C to trim the glycan trees on the protein surface. The efficiency of EndoH$_f$ was analysed by SDS-PAGE and a shift from 20 kDa to 17 kDa is visible for the deglycosylated target protein (figure 3.17). For final purification, a SEC was performed on a SD75 16/600 column to remove aggregated protein, remaining impurities and EndoH$_f$. Siglec-11$_{1D\Delta tag\Delta glyc}$ elutes as symmetric peak at 78 mL and protein containing fraction (76-80 mL) were pooled and concentrated to 7.0 mg/mL for further experiments (figure 3.18). About 4.5 mg pure Siglec-11$_{1D\Delta tag\Delta glyc}$ could be obtained from L of cell culture.
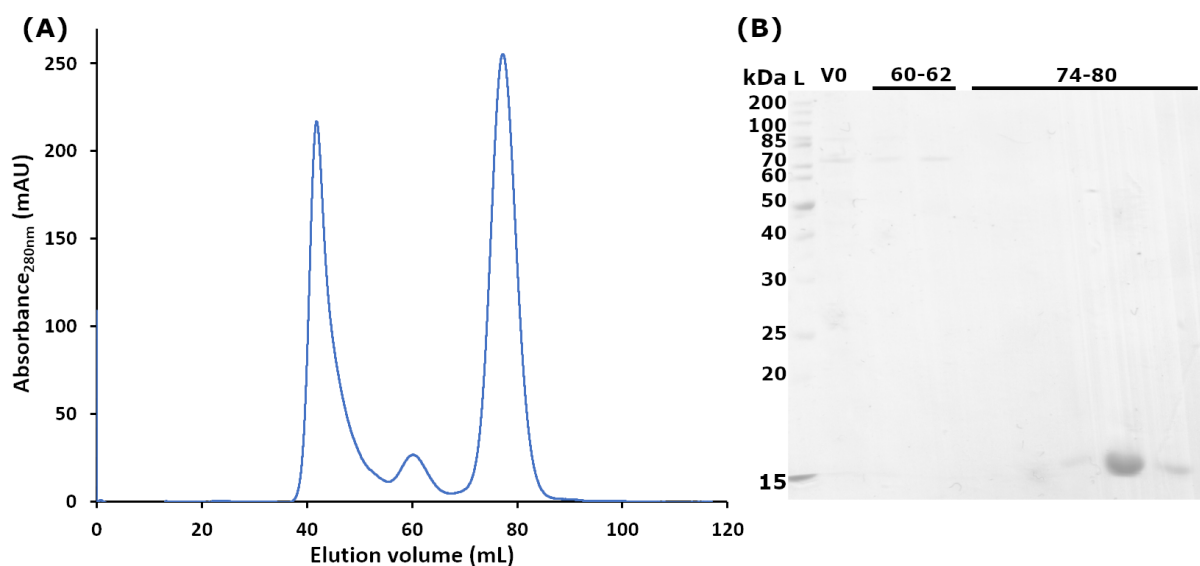
Figure 3.18: Preparative SEC of Siglec-11$_{1D\Delta tag\Delta glyc}$ expressed in Expi293F GnTI- Cells. **(A)** The SEC chromatogram of Siglec-11$_{1D\Delta tag\Delta glyc}$ with the **(B)** the corresponding SDS-PAGE are represented. Three peaks elute from the column: at the exclusion volume at 42 mL (V0), 60 mL (60-62) and 78 mL (74-80). Protein composition of all three peaks were analysed by SDS-PAGE and unstained protein ladder (L) was added for evaluation of the protein size.

The fold of Siglec-11$_{1D\Delta tag\Delta glyc}$ was analysed by CD spectroscopy and compared to the spectrum of Siglec-11$_{1D\Delta myc}$. The CD spectra of both proteins are representative for proteins composed of β-sheets with a maximum at 227 nm and a minimum at 233 nm (appendix figure A.16). The prominent minimum at 215 nm differs with a sharper dip for Siglec-11$_{1D\Delta myc}$ compared to the spectrum of Siglec-11$_{1D\Delta tag\Delta glyc}$. An analytical SEC (SD75 increase 3.2/300 column) was performed to determine the oligomeric state and the stability of Siglec-11$_{1D\Delta tag\Delta glyc}$. The protein elutes at 1.37 mL in a single symmetric peak. Based on the elution volume, an apparent molecular weight of 18 kDa was determined, which indicates a monomeric appearance in solution (appendix figure A.2). The protein is stable, as the chromatograms of the fresh and the stored (10 days at 4 °C) sample are identical (appendix figure A.15).

3.5.3 Affinity Measurements

Purified Siglec-11$_{1D\Delta myc}$ was used for affinity measurements. For SPR experiments, 330 RU of Siglec-11$_{1D\Delta myc}$ were immobilised onto a Ni-NTA sensor chip. After a stable baseline was obtained, analyte, 2 mM colominic acid (average DP=40, with average molecular weight of 30 kDa), was injected for 60 s to monitor the complex formation of Siglec-11$_{1D\Delta myc}$ and polySia. Application of analyte was stopped and running buffer was applied to determine the complex dissociation rate. No binding could be determined, as the response units did not increase during application of analyte. Instead, the baseline, obtained after immobilisation of Siglec-11$_{1D\Delta myc}$, was decreased by 40 RU after addition of colominic acid (appendix figure A.17). This indicates either loss of immobilised ligand during analyte application or unspecific binding of colominic acid on the Ni-NTA chip.

3.5.4 Crystallisation and X-Ray Diffraction

The His$_8$-tag of Siglec-11$_{1D\Delta myc}$ was not enzymatically cleavable. Still, first crystallisation trials were performed with the glycosylated and deglycosylated Siglec-11$_{1D\Delta myc}$ (5.22 mg/mL). No crystal growth could be observed in both cases, for the glycosylated sample 90% of crystallisation condition drops stayed clear, as for the deglycosylated protein around 40% showed precipitated protein or phase separation. Crystallisation trials with deglycosylated Siglec-11$_{1D\Delta myc\Delta glyc}$ at a concentration of 7.9 mg/mL yielded protein precipitation or phase separation for around 50% of the conditions. In three conditions (crystal screen 1+2 condition 9 and 76; JCSG+ condition 45), crystalline structures could be detected after three months at 4 °C (appendix figure A.18). The composition of these conditions were further refined (appendix table A.3) and crystallisation plates were set up with 10 mg/mL Siglec-11$_{1D\Delta myc\Delta glyc}$ at 4 °C. After four weeks, almond-shaped crystals appeared in the fine screen of crystal screen 1+2 condition 9 (appendix figure A.18). These crystals were flash-frozen in liquid nitrogen, after cryo-protection and tested at the SLS, but no diffraction could be detected.
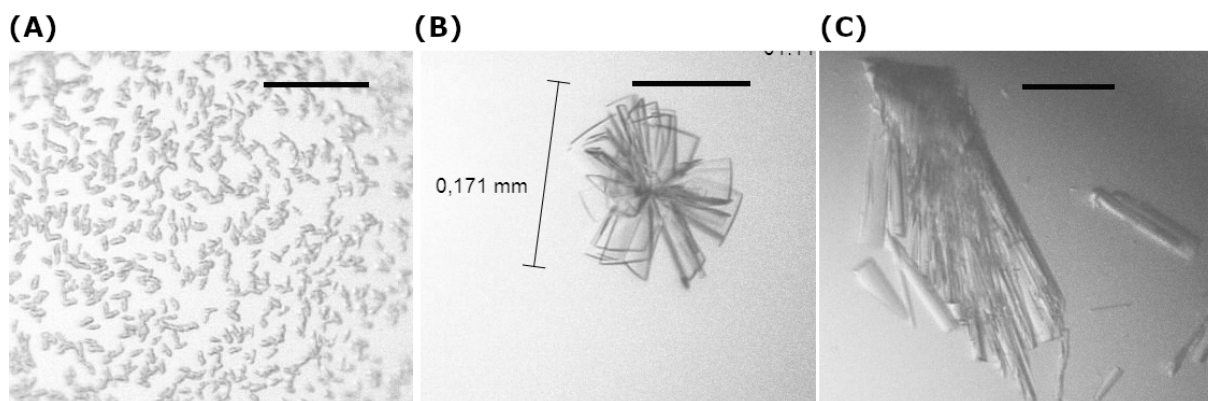
Figure 3.19: Siglec-11$_{1D\Delta tag\Delta glyc}$ crystals. (A) Crystalline structures are visible in condition 27 of the JCSG+ screen after one week at 20 °C. **(B)** Plate shape crystals appeared in condition 5 of JCSG+ containing 1 M sodium citrate and 0.1 M sodium cacodylate at pH 6.0 after one week at 20 °C. **(C)** Fine screen of JCSG+ screen condition 5 reproduce plate shaped crystals after two weeks at 20 °C. In each image a scale bar, measuring 100 µm, is shown as reference.

Siglec-11$_{1D\Delta myc\Delta glyc}$ crystallised but the internal crystal order was not sufficient for X-ray diffraction. One reason for the internal disorder in the crystal could be the extension at the C-terminus of the protein (His$_8$-tag and mutated TEV cleavage site) that is still present in the crystallised sample. Therefore, a new crystallisation trial was performed with Siglec-11$_{1D\Delta tag\Delta glyc}$, lacking any C-terminal extensions. Crystallisation plates were set up with a Gryphon robot with drops containing a 1:1 mix of protein (concentration of 7.0 mg/mL) and precipitation solutions from a commercially available crystallisation screen (JCSG+). After one week at 20 °C, plate-shaped crystals were obtained in a single condition of the JCSG+ screen (figure 3.19 B). The crystals were reproduced in a fine screen of this condition by hanging drop vapour diffusion and crystals of similar shape grew after one week at 20 °C (figure 3.19 C). Crystals from the initial crystallisation hit (condition 5 of the JCSG+ screen) were harvested and used for diffraction experiments. A data set was collected at the beam line X06SA of the SLS in Villigen. During data collection, the crystal suffered from radiation damage, therefore only images of the first 250 ° from the collected 360 ° could be used to solve the structure. The data set was processed in space group P2$_1$ to a resolution of 2.15 Å, with a signal-to-noise ratio of 1.53 and a completeness of 94.9% in the highest resolution shell (table 3.11).

To obtain a ligand structure of Siglec-11$_{1D\Delta myc\Delta glyc}$, crystals from the fine screen were soaked with 20 mM oligo-α2,8-linked sialic acid (DP3), dissolved in the crystallisation condition, overnight. The crystals were mounted in a loop and flash-frozen with liquid nitrogen for data collection at the beam line X06DA of the SLS. No radiation damage could be observed during data collection and images of 240 ° were processed and merged in space group P1. The crystal diffracted to a resolution of 2.75 Å with a signal-to-noise ratio of 1.18 and a CC$_{1/2}$ of 51.6% for the highest resolution shell (table 3.11).

3.5.5 Structural Investigation of Siglec-11$_{V\text{-set}}$

The Matthews coefficient was calculated to determine the solvent content and number of protomers in the ASU. For the Siglec-11$_{1D\Delta tag\Delta glyc}$ apo crystal, the highest Matthews probability was determined for four protomers per ASU and a solvent content of 48.4%. For solving the phase problem, molecular replacement was performed with a model based on the Siglec-5$_{V\text{-set}}$ structure (pdb code: 2zg2). The MR phaser run yielded an initial molecular replacement solution with Z-scores of 15.3, 16.1, 31.4 and 34.5. By visual inspection of the crystal packing, large holes could be determined between the placed protomers with significant amount of positive electron density in the difference density map. A second molecular replacement run was performed based on the initial solution with two additional protomers per ASU, resulting a single solution with Z-scores of 35.4 and 35.2, and an unusual dense crystal packing with a solvent content of only 23% (appendix figure A.20). The solvent content of protein crystals normally ranges between 40% and 70%. In the following, the Siglec-11$_{V\text{-set}}$ structure was adjusted using COOT for real space and REFMAC for reciprocal space refinement to R-work and R-free values of 21.7% and

Table 3.11: Siglec-11$_{1D\Delta tag\Delta glyc}$ data collection and refinement statistics. Summary of data quality and the current state of the refinements for Siglec-11$_{1D\Delta tag\Delta glyc}$ crystals. Values in parentheses are for the highest resolution shell.

	Siglec-11$_{1D\Delta tag\Delta glyc}$ apo	DP3-soaked
Data collection		
Beam line	X06SA	X06DA
Detector	EIGER 16M	PILATUS 2MF
Wavelength [Å]	0.999998	1.000023
Resolution range [Å]	46.67-2.15 (2.27-2.15)	43.21-2.75 (2.92-2.75)
Space group	P2$_1$	P1
Unit cell dimension		
a, b, c [Å]	59.8, 24.3, 223.2	23.5, 53.8, 72.2
α, β, γ [°]	90.0, 90.1, 90.0	89.9, 90.0, 90.1
Total reflection	159414 (22787)	21268 (3620)
Unique reflection	34246 (5323)	9005 (1477)
Multiplicity	4.2	2.4
Completeness [%]	98.1 (94.9)	97.7 (97.3)
I/σ(I)	6.79 (1.53)	8.73 (1.18)
R-meas [%]	17.4 (156.6)	8.8 (106.7)
CC1/2 [%]	99.4 (54.5)	99.7 (51.6)
Wilson B factor [Å^2]	46.2	76.9
Refinement		
R-work [%]	21.7	25.8
R-free [%]	26.3	30.7
Ramachandran		
favoured	98.1	86.6
outliers	0.27	2.8
RMSD		
bond lengths [Å]	0.011	0.011
bond angles [°]	1.61	1.58
Rotamer outliers	0.0	0.0
Clash score	6.5	17.0
B-factors		
Protein chain A/B/C/D	44.8/44.5/45.5/47.4	85.7/90.8/89.7/84.4
Protein chain E/F	44.9/48.9	
Solvent	42.4	

26.3 %, respectively. The final Siglec-11$_{V\text{-set}}$ structure possesses good geometry, with RMSD for bond lengths and bond angles of 0.011 Å and 1.61 °, respectively, and dihedral angles fitting to the Ramachandran plot (table 3.11).

All residues of the Siglec-11$_{V\text{-set}}$ (23-151) are visible in the electron density, except the N-terminal secretion signal, which was probably not part of the secreted protein. Siglec-11$_{V\text{-set}}$ is a single domain built up by eleven β-strands that form two antiparallel β-sheets. One sheet is composed of β-strands A(A')B(B')ED, the other one comprises β-strands C(C')FG(G'). A disulphide bridge between cysteine 48 and cysteine 108, each located in one of the two β-sheets, stabilises the β-sandwich. The β-strands are connected by loops, of which especially the ligand interacting CC'- and C'D-loop are elongated, and the EF-loop possesses a short α-helical turn within the loop. In addition, GlcNAc residues are visible at the two predicted *N*-linked glycosylation sites, one at asparagine 49 located within the B'-strand, the other one is present in the C'D-loop at asparagine 84 of Siglec-11$_{V\text{-set}}$ (figure 3.6 and figure 3.20). The *N*-linked glycan in the C'D-loop of the V-set is unique in the

group of the structural investigated Siglecs. In addition, an alignment of all CD33-related Siglec revealed that only Siglec-11 and Siglec-16 have the recognition sequence for *N*-linked glycosylation in the C'D-loop (appendix figure A.21).
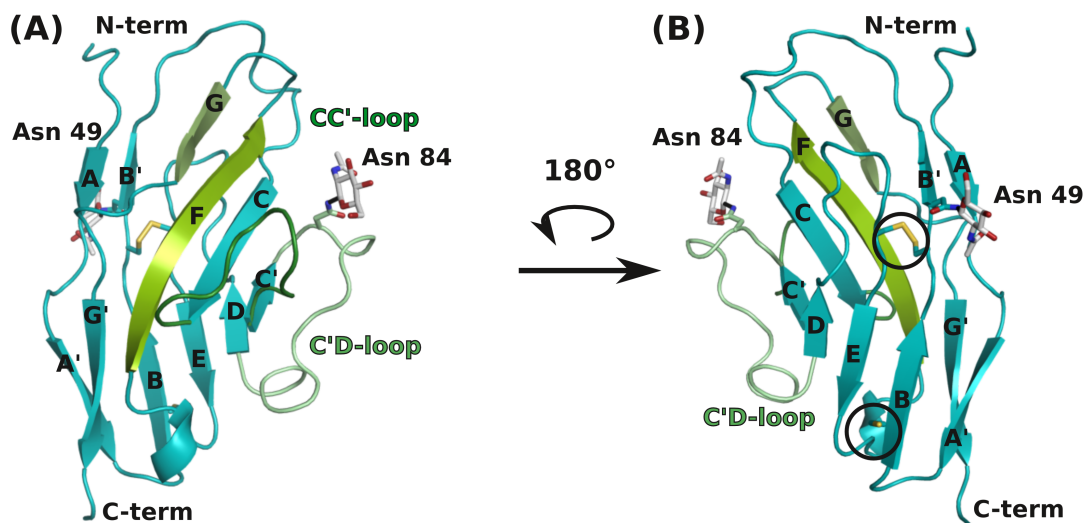
Figure 3.20: Structure of Siglec-11$_{V-set}$. Siglec-11$_{V-set}$ in **(A)** front view and **(B)** back view, vertical turned by 180°, is represented as cartoon in light blue, with the G- and F-strands highlighted in light green, as well as CC'- and C'D-loops are represented in dark green and mint, respectively. The two remaining GlcNAc residues of the *N*-linked glycans and the corresponding asparagines (N49 and N84) are drawn as sticks. In addition, cysteines (cys 48 and cys 108) forming an intra-domain disulphide bridge, as well as the single cysteine (cys 43), which forms an inter-domain disulphide bridge in the full-length Siglec-11 protein, are shown as light blue sticks and highlighted by a black circle. The β-strands, key loops, glycosylated asparagines, as well as N- and C-termini are labelled.

A structure-based alignment was performed with V-sets of other known Siglec structures (Siglec-1 (pdb code: 1od9), Siglec-3 (pdb code: 5ihb), Siglec-4 (pdb code: 5lfv), Siglec-5 (pdb code: 2zg2), Siglec-7 (pdb code:1o7s) and Siglec-8 (pdb code: 2n7a)). The overall fold in the Siglec V-sets is well conserved, with Cα-RMSDs ranging from 0.8 Å to 1.7 Å. Siglec-11 and Siglec-7 are most similar structures, with overall Cα-RMSD with 0.8 Å and a sequence identity of 47.7%. In general, Siglec-11$_{V-set}$ is more closely related to members of the CD33-related family (with Cα-RMSD ≤1.2 Å) than the classical Siglecs (Cα-RMSD ≥1.6 Å) (appendix table A.4). In the alignment, the β-sandwich core of the domain is rigid, but loops differ, even if they are not directly involved in ligand binding, such as B'C-loop (appendix figure A.19).

The Siglec-11$_{1D\Delta tag\Delta glyc}$ crystal, which was soaked with DP3, has an even more densely packed crystal lattice with a solvent content of 8.5% and four protomers in the ASU. Molecular replacement was performed with the Siglec-11$_{V-set}$ apo structure as the search model, yielding a single solution with Z-scores of 13.1, 11.2, 9.5 and 27.2. The structure was refined as described for the Siglec-11$_{V-set}$ apo structure. Refinement was stopped at R-work and R-free values of 25.6% and 30.5%, as no additional positive electron density could be observed in the difference electron density map of the ligand binding region (table 3.11). The packing in the crystal is exceptionally dense, with the drawback that soaking of ligands becomes difficult. In crystals with typical solvent contents of 40%-70%, small

molecules such as ligands or substrates can freely diffuse through pores in the crystal. The dense packing limits this diffusion and in addition, the available space in the ligand binding site is not sufficient for binding of oligoSia of DP=3.

As no Siglec-11$_{\text{V-set}}$ ligand structure could be obtained, the apo structure was compared to the structure of Siglec-7$_{\text{V-set}}$ in complex with GT1b (pdb code: 2hrl), a glycan with an α2,8-linked sialic acid on the nonreducing end. Both Siglec$_{\text{V-sets}}$ align well, with exceptions, the CC'-loop in Siglec-7$_{\text{V-set}}$ rearranges upon ligand binding and shows large differences to the CC'-loop in Siglec-11 with Cα-RMSD up to 5.0 Å (figure 3.21 A). It is similar for the C'D-loop forming a short α-helix in Siglec-7$_{\text{V-set}}$ even though, this loop is not directly involved in ligand engagement (figure 3.21 B). This rearrangements results in an overall Cα-RMSD of 2.1 Å between both structures. In the structure of Siglec-7 in complex with GT1b, not only the first sialic acid linkage determines the ligand specificity, also the other sugars establish interactions toward the V-set of Siglec-7 (figure 3.21 B). The first sialic acid of the ligand is hold in place by a conserved salt bridge between the carboxyl group and the conserved asparagine (R124), as well as hydrophobic interactions of the glycerol group and the tryptophan (W132) in the G-strand. In addition, several hydrogen bonds are established between side chain residues and the sugar entity (figure 3.21 B). The second sialic acid forms only one hydrogen bond with its acetyl group towards a lysine (K135). GalNAc (sugar 4) and Glc (sugar 5) form additional hydrogen bonds to amino acid side chains to stabilise the ligand, only for Gal (sugar 3) no direct interactions could be determined (figure 3.21 B) [307].
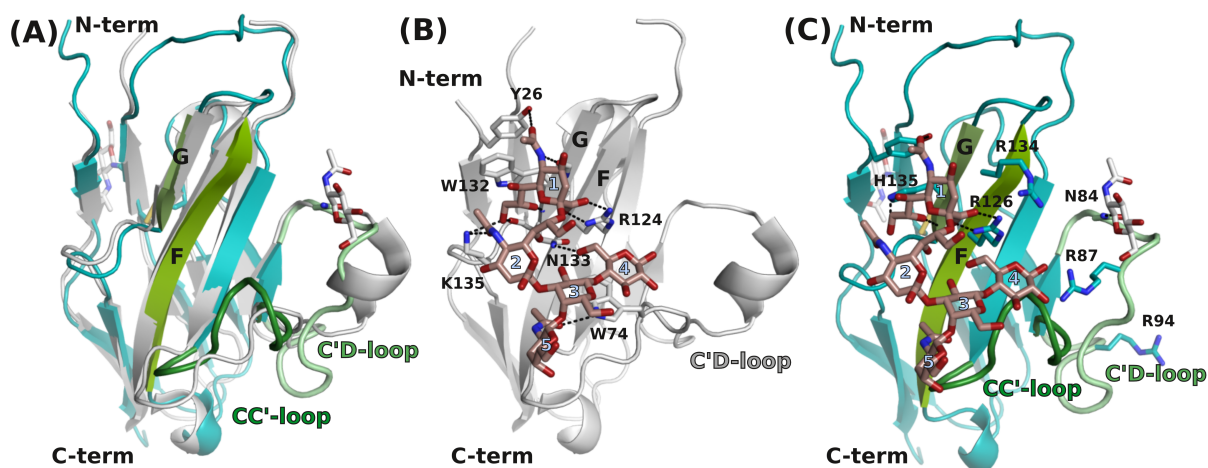
Figure 3.21: Alignment of Siglec-11$_{\text{V-set}}$ and Siglec-7$_{\text{V-set}}$ in complex with GT1b. Siglec-11$_{\text{V-set}}$ is shown in blue and green cartoon representation with its sugar drawn as grey sticks. Siglec-7$_{\text{V-set}}$ structure is represented as grey cartoon with the GT1b ligand depicted as light brown sticks. **(A)** Siglec-11$_{\text{V-set}}$ structure is aligned to Siglec-7$_{\text{V-set}}$-GT1b complex structure (pdb code:2hrl) [307]. **(B)** Siglec-7$_{\text{V-set}}$-GT1b complex is represented with GT1b (brown) and interacting residues (grey) drawn as sticks. **(C)** Siglec-11$_{\text{V-set}}$ is aligned to Siglec-7$_{\text{V-set}}$ in complex with GT1b to model the Siglec-11$_{\text{V-set}}$ (green and blue) GT1b (light brown) complex. The sugar moieties of GT1b and potentially interacting residues are numbered and represented as sticks. N- and C-terminus of Siglec-11$_{\text{V-set}}$ and Siglec-7$_{\text{V-set}}$ are indicated.

Siglec-11$_{vest}$ modelled with GT1b gives first insights into a possible α2,8-linked ligand interaction (figure 3.21 C). The conserved arginine (R 126) in the F-strand could form a salt bridge with the carboxyl group of the first sialic acid, and the glycerol group of the same sugar could be stabilised by interactions with histidine 135 of the G-strand and an additional hydrogen bond between the nitrogen of the histidine and the hydroxyl group in the glycerol entity. The CC'- and C'D-loop in Siglec-11$_{V-set}$ are too distant to establish any binding. In general, no interactions beyond the first sugar entity can be estimated. Siglec-11$_{V-set}$ could undergo an induced fit upon ligand binding to enable interactions with sugar entities beyond the first sialic acid. The current model is not sufficient to explain ligand engagement of Siglec-11$_{V-set}$, as at least the first two sugars are engaged by the binding domain of related Siglecs [308]. GT1b is a branched ligand with only one α2,8-linked sialic acids, in contrast to polySia, the natural ligand of Siglec-11, which is more flexible and can adopt different conformations upon binding [211]. One possible way for ligand engagement are electrostatic interactions of arginines or lysines in the V-set, as polySia is highly negatively charged (figure 3.21 C). Electrostatic surface potential analysis of Siglec-7$_{V-set}$ and Siglec-11$_{V-set}$ revealed no extended positive patches on the protein surface of Siglec-11$_{V-set}$ for polySia binding, instead both proteins have a positively charged patch of similar size but at different positions (figure 3.22). In Siglec-11, the CC'-loop forms a positive charged groove that is not present in Siglec-7. The current conformation of this groove is too narrow (≈ 3.5 Å) for polySia binding, but with small rearrangements in the CC'-loop the groove could provide sufficient space for engagement of linear α2,8-linked polysialic acid.
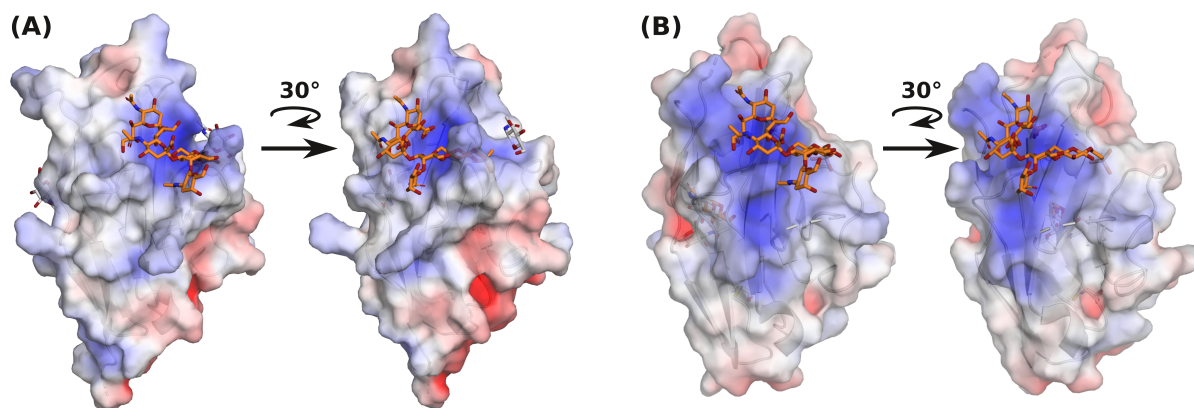
Figure 3.22: Electrostatic surface potential of Siglec-11$_{V-set}$ and Siglec-7$_{V-set}$ in complex with GT1b. For both structures an electrostatic surface potential was calculated using the APBS electrostatics plugin of pymol (radius=1.4, blue=-5 keV, white=0 keV, red=5keV) [288]. **(A)** Siglec-11$_{V-set}$ and **(B)** Siglec-7$_{V-set}$ are depicted as a semi-transparent surface with GT1b ligand as light brown sticks (pdb code:2hrl). Both structures are additionally shown with a turn by 30 °.

3.6 Discussion

For structural investigation, as well as to determine the affinity of Siglec-11 to polySia, high amounts of homogeneous protein are needed. Siglec-11 variants containing various numbers of ECDs reached only reasonable expression levels, if a Kozak sequence (CCACCATG) is present in the expression constructs. Even with the Kozak sequence, constructs with several domains show low expression compared to the one domain construct Siglec-11$_{1D}$. In Freestyle293F cells, all constructs are expressed in detectable amounts. Here, the protein levels are higher in the cell pellet than in the medium samples, but only protein from the medium could be purified. Siglec-11$_{large}$, Siglec-11$_{middle}$ and Siglec-11$_{short}$ appear as diffuse band in the SDS-PAGE, most probably due to their post-translational glycosylation. The heterogeneity of samples is additionally shown by an analytical SEC of Siglec-11$_{middle}$ and Siglec-11$_{large}$, where both Siglec-11 variants appear as several overlapping peaks with an elution volume that corresponds to higher oligomeric species of the proteins. Expi293F GnTI- cells produce protein with a more homogeneous appearance in the SDS-PAGE, but the expression level is lower and part of the protein eluted together with impurities in the exclusion volume of the SEC. Nevertheless, the purified Siglec-11$_{large}$ seems to be stable in solution and appears as a dimer in the analytical SEC.

Recently, Alpha-fold predicted the structure for the macrophagial 5(D)Siglec-11 variant [286]. The macrophagial 5(D)Siglec-11 variant has an identical protein sequence than the microglial four domain variant, but composes an additional domain between the third and last ECD of the microglial (four domain) variant that was used in this work. Based on the predicted structure, the fold of the ECD should not be disturbed by the chosen domain borders for Siglec-11$_{1D}$, Siglec-11$_{short}$, Siglec-11$_{middle\,V3}$, Siglec-11$_{large\,V2}$, Siglec-11$_{large\,V3}$ and Siglec-11$_{large\,V4}$, because all construct ends are located in the linker region between the predicted domains. Only Siglec-11$_{middle\,V2}$ construct ends within the last β-strand of the third ECD in the prediction (appendix figure A.22). Interestingly, the expression level and stability of Siglec-11$_{middle\,V2}$ and Siglec-11$_{middle\,V3}$ are comparable.

Siglec-11, obtained from Freestyle293F cells, can not be homogenised by deglycoslyation with PNGaseF or EndoH$_f$. Glycosylation could partly be removed by PNGaseF, as the corresponding band in the SDS-PAGE shifted to lower molecular weight but stays diffuse. In contrast, EndoH$_f$ treated samples did not shift in the SDS-PAGE. This can have two reasons. First, the asparagine, at which the *N*-linked glycan is attached on, is buried within the protein structure and the glycan is not accessible for the enzymes. Second, the glycans are complex and insensitive for enzymatic deglycosylation. *N*-linked glycans that are processed by Mannosidase II become insensitive to EndoH$_f$ [309]. In addition, fucose-containing glycans are inaccessible for PNGaseF [310, 311]. The diffuse band in the SDS-PAGE is most likely the result of remaining complex *N*-linked glycans and not the result of *O*-glycosylation, because purified Siglec-11, obtained from Expi293F GnTI- cells, appear as a defined band in the SDS-PAGE. Deletion of GnTI in Expi293F cells

results in homogeneous *N*-linked highmannose trees, but should not affect the *O*-linked glycoslyation of the protein. Additionally, C-mannosylation can be excluded, as the recognition sequence (W-x-x-W; x stands for any amino acid) is only present once in the fourth ECD of Siglec-11$_{large}$. This domain is not included in Siglec-11$_{middle}$, but also this variant appears as a heterogeneous protein, if expressed in Freestyle293F cells [312].

Siglec-11$_{1D\Delta myc}$ can be expressed and purified from Expi293F GnTI- cell culture in reasonable amounts (6 mg/L culture). Deglycosylation by EndoH$_f$ does not affect the stability of the protein and was required to obtain initial crystals. Unfortunately, these crystals show no diffraction in X-ray experiments. The Siglec-11$_{1D\Delta myc}$ construct was re-evaluated to obtain well ordered crystals, which are necessary for structure determination. Correction of the point mutation in the TEV protease cleavage site (Siglec-11$_{1D\Delta myc\_TEV}$) resulted in low protein expression in Expi293F GnTI- cells. The recognition sequence for glycoslyation of an asparagine is N-x-S/T, whereas x stands for any amino acid except proline. The present point mutation in the TEV protease cleavage site of Siglec-11$_{1D\Delta myc}$ yielded a N-L-T motif, which could act as an additional recognition sequence for *N*-glycosylation. This so introduced *N*-linked glycan could then stabilise the unstructured C-terminal extension comprising linker, TEV protease cleavage site and His-tag. By correcting the mutated TEV protease cleavage site, this artificial glycosylation site is removed, which could cause a direct degradation of the Siglec-11$_{1D\Delta myc\_TEV}$ after synthesis in the cell and could explain the low protein expression level. Alternatively, Siglec-11$_{1D\Delta tag}$ is lacking a C-terminal extension. This construct could be expressed and purified in reasonable amount (4.5 mg/L culture) and crystallisation experiments yielded crystals that diffract to 2.15 Å.

The V-set of Siglec-11 is built up by two antiparallel β-sheets forming a β-sandwich, a conserved fold for human Siglec V-sets. The crystals have an extraordinarily dense packing, with a solvent content of only 8.5 and 23% for the apo and ligand soaked crystal, respectively. The low solvent content makes soaking of the oligoSia ligand into the crystal impossible. As no Siglec-11 ligand structure is available, a structure-based alignment to Siglec-7 in complex with GT1b (a glycan containing α2,8-linked sialic acid) was performed to analyse possible interactions of Siglec-11 with the α2,8-linked sialic acid entity. Even though the Siglec-11$_{V-set}$ and Siglec-7$_{V-set}$ apo structures are very similar (Cα-RMSD of 0.8 Å), the modelled complex structure is not sufficient to explain binding of polySia beyond the first sialic acid entity of the glycan chain. Siglec-7 V-set rearranges upon GT1b binding, especially in the CC'-loop, which drives ligand specificity [307]. A similar structural rearrangement is possible in Siglec-11$_{V-set}$ to enable ligand binding. A electrostatic surface analysis revealed a positively charged groove formed by the CC'-loop, which could be large enough to bind the structurally flexible polySia chain after a small rearrangement of the loop. In addition, one major difference between Siglec-11$_{V-set}$ and all other structurally investigated Siglecs is a *N*-linked glycan in the C'D-loop, which is often involved in ligand binding. The glycan in this loop could influence the flexibility and conformation of the C'D-loop and play a role in polySia engagement.

Interactions of protein and glycans have often affinities in the high nanomolar to low millimolar range [313, 213]. As colominic acid is bound by the short fibre knob of Ad52 with an affinity in the low millimolar range (1.1 mM), a similar affinity is to be expected for the interaction of colominic acid with Siglec-11 [213]. To determine the affinity of Siglec-11 V-set to colominic acid, SPR experiments were performed with a Biacore X100 system, but no binding could be detected. This can have different reasons. First the experimental set-up is not suitable to determine low affinity interaction. Second, the affinity is lower than the detection limit of the used Biacore system, which is specified by the company with $1\,pM$ to $100\,\mu M$. Third, Siglec-11$_{large}$ was determined to be a dimer in solution, similar to human Siglec-3, Siglec-5, Siglec-8 and murine variant Siglec-E [314, 258]. This observation and the fact that Siglec-11 physiologically binds to polySia of average DP 20, suggests a cooperative binding of polySia by the two V-sets in the Siglec-11 dimer. Siglec-11$_{1D}$ appears as monomer in solution and therefore might no emulate the physiological condition. Another possibility is that Siglec-11 contains, besides the conserved and known Arg126-based binding site, an additional binding site in one of the C-sets of the ECD, that are not included in the Siglec-11$_{1D}$ protein. Further, the initial affinity studies were performed with high mannosylated Siglec-11$_{1D}$, which does not represent the physiological glycosylation of Siglec-11 expressed in microglia or macrophages. The composition and structure of the attached *N*-linked glycan can affect the flexibility and conformation of the C'D-loop, which is often involved in ligand engagement, in this protein family. In addition, carbohydrate-to-carbohydrate interactions could be formed between the *N*-linked glycan and the polySia ligand. In both cases, the expression system (the produced surface glycan) would affected the interaction of the protein with its ligand [315, 316].

3.7 Conclusion and Outlook

To investigate the the role of the Siglec-11 C2-sets in polySia binding, large amounts of pure and stable protein of Siglec-11 multi-domain variants are needed. Of all tested variants, the most promising candidate is Siglec-11$_{large}$, as it could be purified from Freestyle293F and Expi293 GnTI- cell culture. The purification protocol needs to be improved to obtain a less heterogeneous sample with only one oligomeric state. New constructs, containing more than one ECD with precisely defined domain borders, could be obtained by reinvestigation of the Alphafold Siglec-11 structure. Based on the predicted structure, Siglec-11$_{large}$ could be shortened by deletion of eight amino acids at the C-terminus (appendix figure A.22). In addition, the post-translational modification of proteins, produced in Freestyle293F cells, could be analysed by either enzymatic treatment or by mass-spectrometry analysis of the protein, as described by Morelle et al. [311]. If the remaining *N*-linked glycosylations are not accessible by PNGaseF in the folded protein, the protein could be treated under denaturing condition. In addition, possible fucose-containing glycans could be cleaved by PNGaseA [310, 311]. In case complex *N*-glycans are present on the protein surface, homogeneous glycosylated protein could be obtained by mutating the corresponding asparagine to a glutamine to prevent glycosylation at this position.

To obtain Siglec-11$_{V-set}$ stucture in complex with polySia, additional crystallisation experiments could be performed with Siglec-11$_{1D\Delta tag\Delta glyc}$ to find a new crystal form that possesses an higher solvent content. This crystals could then be soaked with oligo- or polySia. If the binding site for polySia is not accessible in the crystal or binding of the ligand introduces a structural rearrangement and damages the crystal lattice, a ligand-bound structure could be obtained by co-crystallisation of the protein with polySia of defined length. Besides Siglec-11$_{1D}$, multi-domain Siglec-11$_{large}$ could be structural investigated. The receptor dimerisation and the relative positions of the two V-sets in the dimer could give insights into a possible cooperative effect in polySia binding. Furthermore, an additional binding site within one of the C2-sets could be determined by co-crystallisation of polySia and Siglec-11$_{large}$.

A suitable method to determine Siglec-11 binding to polySia has to be established. For adenovirus 52 short fibre knob, low affinity interactions to polySia were measured by SPR experiments, with a more sensitive Biacore system (T200 system, detection limit 1 mM). The same system was used to analyse the binding of Siglec-3 to a glycan mimic called P22 possessing an affinity in the high micromolar range [313]. Either an equivalent system could be used or another more sensitive method, such as bio-layer interferometry (detection limit specified with 1 pM to 1 mM for state of the art systems), could be applied to determine binding of polySia to Siglec-11$_{1D\Delta myc}$. Moreover, monomeric Siglec-11$_{1D}$ is most probably not enough for polySia engagement. The effect of Siglec-11 dimerisation onto polySia binding could be analysed by a non physiological Siglec-11$_{1D}$ dimer. For this, Siglec-11$_{1D}$ could be fused to a GCN4-based leucine zipper at the C-terminus [317]. This motif is

commonly used with four repeats and forms a stable coil-coil structure. The affinity of monomeric and dimeric Siglec-11$_{1D}$ to polySia could be compared to investigate a possible cooperative effect in ligand binding. Furthermore, Siglec-11$_{1D}$, expressed in Freestyle293F cells, could be used for affinity experiments to investigate a possible role of a physiological glycan in the C'D-loop of the protein. Another attempt would be to investigate the binding of Siglec-16 to polySia. The first two ECDs of Siglec-11 and Siglec-16 are nearly identical (97% protein sequence identity) and binding properties of Siglec-16 to polySia in comparison to Siglec-11 could be analysed [285]. If the binding properties are exclusively dependent on the V-set or the first two ECD (V-set and first C2-set), both receptors should show similar affinity to polySia. For a deeper analysis of the physiological function of the Siglec-11 dimer, the effect of different oligo- and polySia chain lengths, as well as the number of ECD in Siglec-11 on the affinity could be investigated.

Bibliography

[1] Woo, P. C., Lau, S. K., Lam, C. S., Lau, C. C., *et al.* Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *Journal of Virology*, **86**(7): 3995–4008 (2012).

[2] Kendall, E., Bynoe, M., & Tyrrell, D. Virus isolations from common colds occurring in a residential school. *British Medical Journal*, **2**(5297): 82 (1962).

[3] Tortorici, M. A. & Veesler, D. Structural insights into coronavirus entry. *Advances in Virus Research*, **105**: 93–116 (2019).

[4] Myint, S. H. Human coronavirus infections. In The Coronaviridae, pp. 389–401. Springer (1995).

[5] Nicholls, J., Dong, X.-P., Jiang, G., & Peiris, M. SARS: clinical virology and pathogenesis. *Respirology*, **8**: S6–S8 (2003).

[6] Li, W., Shi, Z., Yu, M., Ren, W., *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science*, **310**(5748): 676–679 (2005).

[7] Ding, Y., He, L., Zhang, Q., Huang, Z., *et al.* Organ distribution of severe acute respiratory syndrome (SARS) associated coronavirus (SARS-CoV) in SARS patients: implications for pathogenesis and virus transmission pathways. *The Journal of Pathology*, **203**(2): 622–630 (2004).

[8] Li, W., Zhang, C., Sui, J., Kuhn, J. H., *et al.* Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *The EMBO Journal*, **24**(8): 1634–1643 (2005).

[9] Lei, H., Li, Y., Xiao, S., Lin, C.-H., *et al.* Routes of transmission of influenza A H1N1, SARS CoV, and norovirus in air cabin: comparative analyses. *Indoor Air*, **28**(3): 394–403 (2018).

[10] Srikantiah, P., Charles, M. D., Reagan, S., Clark, T. A., *et al.* SARS clinical features, United States, 2003. *Emerging Infectious Diseases*, **11**(1): 135 (2005).

[11] Bolles, M., Donaldson, E., & Baric, R. SARS-CoV and emergent coronaviruses: viral determinants of interspecies transmission. *Current Opinion in Virology*, **1**(6): 624–634 (2011).

Bibliography

[12] Bradley, B. T. & Bryan, A. Emerging respiratory infections: the infectious disease pathology of SARS, MERS, pandemic influenza, and *Legionella*. *Seminars in Diagnostic Pathology*, **36**(3): 152–159 (2019).

[13] Baric, R. S. SARS-CoV: lessons for global health. *Virus Research*, **133**(1): 1 (2008).

[14] Zaki, A. M., Van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine*, **367**(19): 1814–1820 (2012).

[15] Bermingham, A., Chand, M., Brown, C., Aarons, E., *et al.* Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Eurosurveillance*, **17**(40): 20290 (2012).

[16] Sabir, J. S., Lam, T. T.-Y., Ahmed, M. M., Li, L., *et al.* Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. *Science*, **351**(6268): 81–84 (2016).

[17] Banerjee, A., Kulcsar, K., Misra, V., Frieman, M., & Mossman, K. Bats and coronaviruses. *Viruses*, **11**(1): 41 (2019).

[18] Widagdo, W., Raj, V. S., Schipper, D., Kolijn, K., *et al.* Differential expression of the Middle East respiratory syndrome coronavirus receptor in the upper respiratory tracts of humans and dromedary camels. *Journal of Virology*, **90**(9): 4838–4842 (2016).

[19] Meyerholz, D. K., Lambertz, A. M., & McCray Jr, P. B. Dipeptidyl peptidase 4 distribution in the human respiratory tract: implications for the Middle East respiratory syndrome. *The American Journal of Pathology*, **186**(1): 78–86 (2016).

[20] Ahmed, A. E. Estimating survival rates in MERS-CoV patients 14 and 45 days after experiencing symptoms and determining the differences in survival rates by demographic data, disease characteristics and regions: a worldwide study. *Epidemiology & Infection*, **146**(4): 489–495 (2018).

[21] Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. A novel coronavirus outbreak of global health concern. *The Lancet*, **395**(10223): 470–473 (2020).

[22] Lu, R., Zhao, X., Li, J., Niu, P., *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, **395**(10224): 565–574 (2020).

[23] Group, T. I. C. S. *et al.* The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, **5**(4): 536 (2020).

[24] Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. The proximal origin of SARS-CoV-2. *Nature Medicine*, **26**(4): 450–452 (2020).

[25] Li, Q., Guan, X., Wu, P., Wang, X., *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, **382**(13): 1129–1207 (2020).

[26] Rabaan, A. A., Al-Ahmed, S. H., Al-Malkey, M. K., Alsubki, R. A., *et al.* Airborne transmission of SARS-CoV-2 is the dominant route of transmission: droplets and aerosols. *Le Infezioni in Medicina*, **29**(1): 10–19 (2021).

[27] Letko, M., Marzi, A., & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology*, **5**(4): 562–569 (2020).

[28] Deinhardt-Emmer, S., Wittschieber, D., Sanft, J., Kleemann, S., *et al.* Early postmortem mapping of SARS-CoV-2 RNA in patients with COVID-19 and the correlation with tissue damage. *Elife*, **10**: e60361 (2021).

[29] Bender, J. K., Brandl, M., Höhle, M., Buchholz, U., & Zeitlmann, N. Analysis of asymptomatic and presymptomatic transmission in SARS-CoV-2 outbreak, Germany, 2020. *Emerging Infectious Diseases*, **27**(4): 1159 (2021).

[30] Rasmussen, A. L. & Popescu, S. V. SARS-CoV-2 transmission without symptoms. *Science*, **371**(6535): 1206–1207 (2021).

[31] Spinato, G., Fabbris, C., Polesel, J., Cazzador, D., *et al.* Alterations in smell or taste in mildly symptomatic outpatients with SARS-CoV-2 infection. *JAMA*, **323**(20): 2089–2090 (2020).

[32] Grant, M. C., Geoghegan, L., Arbyn, M., Mohammed, Z., *et al.* The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): a systematic review and meta-analysis of 148 studies from 9 countries. *PloS One*, **15**(6): e0234765 (2020).

[33] Streeck, H., Schulte, B., Kümmerer, B. M., Richter, E., *et al.* Infection fatality rate of SARS-CoV2 in a super-spreading event in Germany. *Nature Communications*, **11**(1): 1–12 (2020).

[34] Hou, W., Zhang, W., Jin, R., Liang, L., *et al.* Risk factors for disease progression in hospitalized patients with COVID-19: a retrospective cohort study. *Infectious Diseases*, **52**(7): 498–505 (2020).

[35] Perez-Saez, J., Lauer, S. A., Kaiser, L., Regard, S., *et al.* Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland. *The Lancet Infectious Diseases*, **21**(4): e69–e70 (2021).

[36] Dong, E., Du, H., & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, **20**(5): 533–534 (2020).

[37] Erikstrup, C., Hother, C. E., Pedersen, O. B. V., Mølbak, K., *et al.* Estimation of SARS-CoV-2 infection fatality rate by real-time antibody screening of blood donors. *Clinical Infectious Diseases*, **72**(2): 249–253 (2021).

[38] Neuman, B. W., Adair, B. D., Yoshioka, C., Quispe, J. D., *et al.* Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. *Journal of Virology*, **80**(16): 7918–7928 (2006).

[39] Li, T., Li, Z., Deans, E. E., Mittler, E., *et al.* The shape of pleomorphic virions determines resistance to cell-entry pressure. *Nature Microbiology*, **6**(5): 617–629 (2021).

[40] Yadav, P. D., Potdar, V. A., Choudhary, M. L., Nyayanit, D. A., *et al.* Full-genome sequences of the first two SARS-CoV-2 viruses from India. *The Indian journal of Medical Research*, **151**(2-3): 200 (2020).

[41] Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J. W., *et al.* The architecture of SARS-CoV-2 transcriptome. *Cell*, **181**(4): 914–921 (2020).

[42] Cubuk, J., Alston, J. J., Incicco, J. J., Singh, S., *et al.* The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nature Communications*, **12**(1): 1–17 (2021).

[43] Mahtarin, R., Islam, S., Islam, M. J., Ullah, M. O., *et al.* Structure and dynamics of membrane protein in SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics*, pp. 1–14 (2020).

[44] Bianchi, M., Benvenuto, D., Giovanetti, M., Angeletti, S., *et al.* SARS-CoV-2 envelope and membrane proteins: structural differences linked to virus characteristics? *BioMed Research International*, **2020**(4389089): 1–6 (2020).

[45] Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., *et al.* SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, **181**(2): 271–280 (2020).

[46] Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., *et al.* Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, **181**(2): 281–292 (2020).

[47] Lan, J., Ge, J., Yu, J., Shan, S., *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, **581**(7807): 215–220 (2020).

[48] Hassani, A. & Khan, G. Human-animal interaction and the emergence of SARS-CoV-2. *JMIR Public Health and Surveillance*, **6**(4): e22117 (2020).

[49] Sanda, M., Morrison, L., & Goldman, R. N-and O-glycosylation of the SARS-CoV-2 spike protein. *Analytical Chemistry*, **93**(4): 2003–2009 (2021).

[50] Ou, X., Liu, Y., Lei, X., Li, P., *et al.* Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature Communications*, **11**(1): 1–12 (2020).

[51] Zhang, Q., Chen, C. Z., Swaroop, M., Xu, M., *et al.* Heparan sulfate assists SARS-CoV-2 in cell entry and can be targeted by approved drugs *in vitro*. *Cell Discovery*, **6**(1): 1–14 (2020).

[52] Vkovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology*, **19**(3): 155–170 (2021).

[53] Brant, A. C., Tian, W., Majerciak, V., Yang, W., & Zheng, Z.-M. SARS-CoV-2: from its discovery to genome structure, transcription, and replication. *Cell & Bioscience*, **11**(1): 1–17 (2021).

[54] Astuti, I. *et al.* Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, **14**(4): 407–412 (2020).

[55] Snijder, E., Decroly, E., & Ziebuhr, J. The nonstructural proteins directing coronavirus RNA synthesis and processing. *Advances in Virus Research*, **96**: 59–126 (2016).

[56] Burke, J. M., St Clair, L. A., Perera, R., & Parker, R. SARS-CoV-2 infection triggers widespread host mRNA decay leading to an mRNA export block. *RNA*, **27**(11): 1318–1329 (2021).

[57] Ye, Q., West, A. M., Silletti, S., & Corbett, K. D. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Science*, **29**(9): 1890–1901 (2020).

[58] Mendonça, L., Howe, A., Gilchrist, J. B., Sheng, Y., *et al.* Correlative multi-scale cryo-imaging unveils SARS-CoV-2 assembly and egress. *Nature Communications*, **12**(1): 1–10 (2021).

[59] World Health Organization. Tracking SARS-CoV-2 variants. Geneva, Switzerland: WHO; 2021 Nov 22. `https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/` (2021).

[60] Santos, J. & Passos, G. The high infectivity of SARS-CoV-2 B. 1.1. 7 is associated with increased interaction force between Spike-ACE2 caused by the viral N501Y mutation. . *Nature*, **587**: 340–341 (2020).

Bibliography

[61] Volz, E., Mishra, S., Chand, M., Barrett, J. C., *et al.* Transmission of SARS-CoV-2 lineage B. 1.1. 7 in England: Insights from linking epidemiological and genetic data. *MedRxiv*, pp. 2020–12 (2021).

[62] Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., *et al.* Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *MedRxiv* (2020).

[63] Zimerman, R. A., Ferrareze, P. A. G., Cadegiani, F. A., Wambier, C. G., *et al.* Comparative genomics and characterization of SARS-CoV-2 P. 1 (Gamma) Variant of Concern (VOC) from Amazonas, Brazil. *medRxiv* (2021).

[64] Mlcochova, P., Kemp, S. A., Dhar, M. S., Papa, G., *et al.* SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature*, **599**(7883): 1–6 (2021).

[65] Zhang, J., Xiao, T., Cai, Y., Lavine, C. L., *et al.* Membrane fusion and immune evasion by the spike protein of SARS-CoV-2 Delta variant. *Science*, **374**(6573): eabl9463 (2021).

[66] Romero, P. E., Dávila-Barclay, A., Salvatierra, G., González, L., *et al.* The emergence of SARS-CoV-2 variant lambda (C. 37) in South America. *Microbiology spectrum*, **9**(2): e00789–21 (2021).

[67] Uriu, K., Kimura, I., Shirakawa, K., Takaori-Kondo, A., *et al.* Neutralization of the SARS-CoV-2 Mu variant by convalescent and vaccine serum. *New England Journal of Medicine*, **385**(25): 2397–23399 (2021).

[68] Saxena, S. K., Kumar, S., Ansari, S., Paweska, J. T., *et al.* Characterization of the novel SARS-CoV-2 Omicron (B. 1.1. 529) Variant of Concern and its global perspective. *Journal of medical virology*, pp. 1–7 (2021).

[69] Khailany, R. A., Safdar, M., & Ozaslan, M. Genomic characterization of a novel SARS-CoV-2. *Gene Reports*, **19**(1): 100682 (2020).

[70] Yurkovetskiy, L., Wang, X., Pascal, K. E., Tomkins-Tinch, C., *et al.* Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell*, **183**(3): 739–751 (2020).

[71] Mansbach, R. A., Chakraborty, S., Nguyen, K., Montefiori, D. C., *et al.* The SARS-CoV-2 spike variant D614G favors an open conformational state. *Science Advances*, **7**(16): eabf3671 (2021).

[72] Zhou, B., Thao, T. T. N., Hoffmann, D., Taddeo, A., *et al.* SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature*, **592**(7852): 122–127 (2021).

[73] Hoffmann, M., Arora, P., Groß, R., Seidel, A., *et al.* SARS-CoV-2 variants B. 1.351 and P. 1 escape from neutralizing antibodies. *Cell*, **184**(9): 2384–2393 (2021).

[74] Vaughan, A. Delta to dominate world. *New Scientist*, **250**(3341): 9 (2021).

[75] Bush, D. L. & Rechnitz, G. A. Monoclonal antibody biosensor for antigen monitoring. *Analytical Letters*, **20**(11): 1781–1790 (1987).

[76] Uwamino, Y., Wakui, M., Aoki, W., Kurafuji, T., *et al.* Evaluation of the usability of various rapid antibody tests in the diagnostic application for COVID-19. *Annals of Clinical Biochemistry*, **58**(3): 174–180 (2021).

[77] Nozaki, T., Imai, R., Tanbo, M., Nagashima, R., *et al.* Dynamic organization of chromatin domains revealed by super-resolution live-cell imaging. *Molecular Cell*, **67**(2): 282–293 (2017).

[78] Vidarsson, G., Dekkers, G., & Rispens, T. IgG subclasses and allotypes: from structure to effector functions. *Frontiers in Immunology*, **5**: 520 (2014).

[79] Arbabi-Ghahroudi, M. Camelid single-domain antibodies: historical perspective and future outlook. *Frontiers in Immunology*, **8**: 1589 (2017).

[80] Cawez, F., Duray, E., Hu, Y., Vandenameele, J., *et al.* Combinatorial design of a nanobody that specifically targets structured RNAs. *Journal of Molecular Biology*, **430**(11): 1652–1670 (2018).

[81] Shahi, B., Mousavi Gargari, S., Rasooli, I., Rajabi Bazl, M., & Hoseinpoor, R. Random mutagenesis of Bo NT/EH c nanobody to construct a secondary phage-display library. *Journal of Applied Microbiology*, **117**(2): 528–536 (2014).

[82] 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. *Journal of Translational Medicine*, **12**(1): 1–12 (2014).

[83] Oloketuyi, S., Bernedo, R., Christmann, A., Borkowska, J., *et al.* Native llama nanobody library panning performed by phage and yeast display provides binders suitable for c-reactive protein detection. *Biosensors*, **11**(12): 496 (2021).

[84] Bork, P., Holm, L., & Sander, C. The immunoglobulin fold: structural classification, sequence patterns and common core. *Journal of Molecular Biology*, **242**(4): 309–320 (1994).

[85] Davies, D. R. & Chacko, S. Antibody structure. *Accounts of Chemical Research*, **26**(8): 421–427 (1993).

[86] Kiyoshi, M., Tsumoto, K., Ishii-Watabe, A., & Caaveiro, J. M. Glycosylation of IgG-Fc: a molecular perspective. *International immunology*, **29**(7): 311–317 (2017).

[87] Flanagan, R. J. & Jones, A. L. Fab antibody fragments. *Drug Safety*, **27**(14): 1115–1133 (2004).

[88] Miller, K. D., Weaver-Feldhaus, J., Gray, S. A., Siegel, R. W., & Feldhaus, M. J. Production, purification, and characterization of human scFv antibodies expressed in *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Escherichia coli*. *Protein Expression and Purification*, **42**(2): 255–267 (2005).

[89] Jovčevska, I. & Muyldermans, S. The therapeutic potential of nanobodies. *BioDrugs*, **34**(1): 11–26 (2020).

[90] Hendershot, L., Bole, D., Köhler, G., & Kearney, J. F. Assembly and secretion of heavy chains that do not associate posttranslationally with immunoglobulin heavy chain-binding protein. *Journal of Cell Biology*, **104**(3): 761–767 (1987).

[91] Mitchell, L. S. & Colwell, L. J. Comparative analysis of nanobody sequence and structure data. *Proteins: Structure, Function, and Bioinformatics*, **86**(7): 697–706 (2018).

[92] Birch, J. R. & Racher, A. J. Antibody production. *Advanced Drug Delivery Reviews*, **58**(5-6): 671–685 (2006).

[93] Fridy, P. C., Li, Y., Keegan, S., Thompson, M. K., *et al.* A robust pipeline for rapid production of versatile nanobody repertoires. *Nature Methods*, **11**(12): 1253–1260 (2014).

[94] Kolkman, J. A. & Law, D. A. Nanobodies–from llamas to therapeutic proteins. *Drug Discovery Today: Technologies*, **7**(2): e139–e146 (2010).

[95] Turner, K. B., Zabetakis, D., Goldman, E. R., & Anderson, G. P. Enhanced stabilization of a stable single domain antibody for SEB toxin by random mutagenesis and stringent selection. *Protein Engineering, Design & Selection*, **27**(3): 89–95 (2014).

[96] Quintero-Hernández, V., Juárez-González, V. R., Ortíz-León, M., Sánchez, R., *et al.* The change of the scFv into the Fab format improves the stability and *in vivo* toxin neutralization capacity of recombinant antibodies. *Molecular Immunology*, **44**(6): 1307–1315 (2007).

[97] Lazarevic, I., Pravica, V., Miljanovic, D., & Cupic, M. Immune evasion of SARS-CoV-2 emerging variants: What have we learnt so far? *Viruses*, **13**(7): 1192 (2021).

[98] Levin, E. G., Lustig, Y., Cohen, C., Fluss, R., *et al.* Waning immune humoral response to BNT162b2 Covid-19 vaccine over 6 months. *New England Journal of Medicine*, **385**(24): e84 (2021).

[99] Khoury, D. S., Cromer, D., Reynaldi, A., Schlub, T. E., *et al.* Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nature Medicine*, **27**: 1–7 (2021).

[100] Tang, Q., Owens, R. J., & Naismith, J. H. Structural biology of nanobodies against the spike protein of SARS-CoV-2. *Viruses*, **13**(11): 2214 (2021).

[101] Mapook, A., Macabeo, A. P. G., Thongbai, B., Hyde, K. D., & Stadler, M. Polyketide-derived secondary metabolites from a Dothideomycetes fungus, *Pseudopalawania siamensis* gen. et sp. nov.,(Muyocopronales) with antimicrobial and cytotoxic activities. *Biomolecules*, **10**(4): 569 (2020).

[102] Mazid, M., Khan, T., Mohammad, F., *et al.* Role of secondary metabolites in defense mechanisms of plants. *Biology and Medicine*, **3**(2): 232–249 (2011).

[103] Al Aboody, M. S. & Mickymaray, S. Anti-fungal efficacy and mechanisms of flavonoids. *Antibiotics*, **9**(2): 45 (2020).

[104] Silakowski, B., Kunze, B., & Müller, R. Multiple hybrid polyketide synthase/non-ribosomal peptide synthetase gene clusters in the myxobacterium *Stigmatella aurantiaca*. *Gene*, **275**(2): 233–240 (2001).

[105] Jakubczyk, D., Cheng, J. Z., & O'Connor, S. E. Biosynthesis of the ergot alkaloids. *Natural Product Reports*, **31**(10): 1328–1338 (2014).

[106] Mankelow, D. P. & Neilan, B. A. Non-ribosomal peptide antibiotics. *Expert Opinion on Therapeutic Patents*, **10**(10): 1583–1591 (2000).

[107] Sunassee, S. N. & Davies-Coleman, M. T. Cytotoxic and antioxidant marine prenylated quinones and hydroquinones. *Natural Product Reports*, **29**(5): 513–535 (2012).

[108] Demain, A. L. & Fang, A. The natural functions of secondary metabolites. *History of Modern Biotechnology I*, **96**: 1–39 (2000).

[109] Zhang, P., Yuan, X.-L., Du, Y.-M., Zhang, H.-B., *et al.* Angularly prenylated indole alkaloids with antimicrobial and insecticidal activities from an endophytic fungus *Fusarium sambucinum* TE-6L. *Journal of Agricultural and Food Chemistry*, **67**(43): 11994–12001 (2019).

[110] Keswani, C., Singh, H. B., García-Estrada, C., Caradus, J., *et al.* Antimicrobial secondary metabolites from agriculturally important bacteria as next-generation pesticides. *Applied Microbiology and Biotechnology*, **104**(3): 1013–1034 (2020).

[111] Fleming, A. Penicillin. *British Medical Journal*, **2**(4210): 386 (1941).

[112] Mukai, R. Prenylation enhances the biological activity of dietary flavonoids by altering their bioavailability. *Bioscience, Biotechnology, and Biochemistry*, **82**(2): 207–215 (2018).

[113] D Archivio, M., Filesi, C., Di Benedetto, R., Gargiulo, R., *et al.* Polyphenols, dietary sources and bioavailability. *Annali-Istituto Superiore di Sanita*, **43**(4): 348 (2007).

[114] Shin, E. M., Zhou, H. Y., Guo, L. Y., Kim, J. A., *et al.* Anti-inflammatory effects of glycyrol isolated from Glycyrrhiza uralensis in LPS-stimulated RAW264. 7 macrophages. *International Immunopharmacology*, **8**(11): 1524–1532 (2008).

[115] Aghamiri, V., Mirghafourvand, M., Mohammad-Alizadeh-Charandabi, S., & Nazemiyeh, H. The effect of Hop (*Humulus lupulus L.*) on early menopausal symptoms and hot flashes: A randomized placebo-controlled trial. *Complementary Therapies in Clinical Practice*, **23**: 130–135 (2016).

[116] Bolton, J. L., Dunlap, T. L., Hajirahimkhan, A., Mbachu, O., *et al.* The multiple biological targets of hops and bioactive compounds. *Chemical Research in Toxicology*, **32**(2): 222–233 (2019).

[117] Brown, K. A., Iyengar, N. M., Zhou, X. K., Gucalp, A., *et al.* Menopause is a determinant of breast aromatase expression and its associations with BMI, inflammation, and systemic markers. *The Journal of Clinical Endocrinology & Metabolism*, **102**(5): 1692–1701 (2017).

[118] Štulíková, K., Karabín, M., Nešpor, J., & Dostálek, P. Therapeutic perspectives of 8-prenylnaringenin, a potent phytoestrogen from hops. *Molecules*, **23**(3): 660 (2018).

[119] Quiroga, E. N., Sampietro, D. A., Sgariglia, M. A., Soberón, J. R., & Vattuone, M. A. Antimycotic activity of 5'-prenylisoflavanones of the plant *Geoffroea decorticans*, against *Aspergillus species*. *International Journal of Food Microbiology*, **132**(1): 42–46 (2009).

[120] Gubiani, J. R., Oliveira, M. C., Neponuceno, R. A., Camargo, M. J., *et al.* Cytotoxic prenylated indole alkaloid produced by the endophytic fungus *Aspergillus terreus* P63. *Phytochemistry Letters*, **32**: 162–167 (2019).

[121] Walsh, T., Standiford, H., Reboli, A., John, J., *et al.* Randomized double-blinded trial of rifampin with either novobiocin or trimethoprim-sulfamethoxazole against methicillin-resistant *Staphylococcus aureus* colonization: prevention of antimicrobial resistance and effect of host factors on outcome. *Antimicrobial Agents and Chemotherapy*, **37**(6): 1334–1342 (1993).

[122] Sharma, N., Sharma, V. K., Manikyam, H. K., & Krishna, A. B. Ergot alkaloids: a review on therapeutic applications. *European Journal of Medicinal Plants*, **14**(3): 1–17 (2016).

[123] Floss, H. G. Biosynthesis of ergot alkaloids and related compounds. *Tetrahedron*, **32**(8): 873–912 (1976).

[124] Metzger, U., Schall, C., Zocher, G., Unsöld, I., *et al.* The structure of dimethylallyl tryptophan synthase reveals a common architecture of aromatic prenyltransferases in fungi and bacteria. *Proceedings of the National Academy of Sciences*, **106**(34): 14309–14314 (2009).

[125] Schiff Jr, P. L. Ergot and its alkaloids. *American Journal of Pharmaceutical Education*, **70**(5) (2006).

[126] Li, P., Zhang, M., Li, H., Wang, R., *et al.* New prenylated indole homodimeric and pteridine alkaloids from the marine-derived fungus *Aspergillus austroafricanus* y32-2. *Marine Drugs*, **19**(2): 98 (2021).

[127] Gerbino, D. C., Mandolesi, S. D., Schmalz, H.-G., & Podesta, J. C. Introduction of allyl and prenyl side-chains into aromatic systems by suzuki cross-coupling reactions (2009).

[128] Lindel, T., Marsch, N., & Adla, S. K. Indole prenylation in alkaloid synthesis. *Alkaloid Synthesis*, **309**: 67–129 (2011).

[129] Kimura, M., Futamata, M., Mukai, R., & Tamaru, Y. Pd-catalyzed C3-selective allylation of indoles with allyl alcohols promoted by triethylborane. *Journal of the American Chemical Society*, **127**(13): 4592–4593 (2005).

[130] Odejinmi, S. I. & Wiemer, D. F. Application of benzyl protecting groups in the synthesis of prenylated aromatic compounds. *Tetrahedron Letters*, **46**(22): 3871–3874 (2005).

[131] Li, Y., Vaz, R. J., Olson, S. H., Munson, M., *et al.* Selectivity in the addition of electron deficient radicals to the C2 position of indoles. *European Journal of Organic Chemistry*, **2020**(36): 5828–5832 (2020).

[132] Fan, A., Winkelblech, J., & Li, S.-M. Impacts and perspectives of prenyltransferases of the DMATS superfamily for use in biotechnology. *Applied Microbiology and Biotechnology*, **99**(18): 7399–7415 (2015).

[133] Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*, **158**(2): 412–421 (2014).

[134] Chen, H.-P. & Abe, I. Microbial soluble aromatic prenyltransferases for engineered biosynthesis. *Synthetic and Systems Biotechnology*, **6**(2): 51–62 (2021).

[135] Mori, T. Enzymatic studies on aromatic prenyltransferases. *Journal of Natural Medicines*, **74**(3): 501–512 (2020).

[136] Winkelblech, J., Fan, A., & Li, S.-M. Prenyltransferases as key enzymes in primary and secondary metabolism. *Applied Microbiology and Biotechnology*, **99**(18): 7379–7397 (2015).

[137] Tello, M., Kuzuyama, T., Heide, L., Noel, J., & Richard, S. The ABBA family of aromatic prenyltransferases: broadening natural product diversity. *Cellular and Molecular Life Sciences: CMLS*, **65**(10): 1459 (2008).

[138] Bonitz, T., Alva, V., Saleh, O., Lupas, A. N., & Heide, L. Evolutionary relationships of microbial aromatic prenyltransferases. *PloS One*, **6**(11): e27336 (2011).

[139] Pojer, F., Wemakor, E., Kammerer, B., Chen, H., *et al.* CloQ, a prenyltransferase involved in clorobiocin biosynthesis. *Proceedings of the National Academy of Sciences*, **100**(5): 2316–2321 (2003).

[140] Metzger, U., Keller, S., Stevenson, C. E., Heide, L., & Lawson, D. M. Structure and mechanism of the magnesium-independent aromatic prenyltransferase CloQ from the clorobiocin biosynthetic pathway. *Journal of Molecular Biology*, **404**(4): 611–626 (2010).

[141] Rudolf, J. D. & Poulter, C. D. Tyrosine O-prenyltransferase SirD catalyzes S-, C-, and N-prenylations on tyrosine and tryptophan derivatives. *ACS Chemical Biology*, **8**(12): 2707–2714 (2013).

[142] Winkelblech, J., Liebhold, M., Gunera, J., Xie, X., *et al.* Tryptophan C5-, C6-and C7-prenylating enzymes displaying a preference for C-6 of the indole ring in the presence of unnatural dimethylallyl diphosphate analogues. *Advanced Synthesis & Catalysis*, **357**(5): 975–986 (2015).

[143] Roose, B. W. & Christianson, D. W. Structural basis of tryptophan reverse N-prenylation catalyzed by CymD. *Biochemistry*, **58**(30): 3232–3242 (2019).

[144] Jost, M., Zocher, G., Tarcz, S., Matuschek, M., *et al.* Structure- function analysis of an enzymatic prenyl transfer reaction identifies a reaction chamber with modifiable specificity. *Journal of the American Chemical Society*, **132**(50): 17849–17858 (2010).

[145] Yu, X., Zocher, G., Xie, X., Liebhold, M., *et al.* Catalytic mechanism of stereospecific formation of cis-configured prenylated pyrroloindoline diketopiperazines by indole prenyltransferases. *Chemistry & Biology*, **20**(12): 1492–1501 (2013).

[146] Elshahawi, S. I., Cao, H., Shaaban, K. A., Ponomareva, L. V., *et al.* Structure and specificity of a permissive bacterial C-prenyltransferase. *Nature Chemical Biology*, **13**(4): 366–368 (2017).

[147] Mori, T., Zhang, L., Awakawa, T., Hoshino, S., *et al.* Manipulation of prenylation reactions by structure-based engineering of bacterial indolactam prenyltransferases. *Nature Communications*, **7**(1): 1–11 (2016).

[148] Kremer, A., Westrich, L., & Li, S.-M. A 7-dimethylallyltryptophan synthase from *Aspergillus fumigatus*: overproduction, purification and biochemical characterization. *Microbiology*, **153**(10): 3409–3416 (2007).

[149] Chandran, S. S., Kealey, J. T., & Reeves, C. D. Microbial production of isoprenoids. *Process Biochemistry*, **46**(9): 1703–1710 (2011).

[150] Yazaki, K., Sasaki, K., & Tsurumaru, Y. Prenylation of aromatic compounds, a key diversification of plant secondary metabolites. *Phytochemistry*, **70**(15-16): 1739–1745 (2009).

[151] Luk, L. Y. & Tanner, M. E. Mechanism of dimethylallyltryptophan synthase: evidence for a dimethylallyl cation intermediate in an aromatic prenyltransferase reaction. *Journal of the American Chemical Society*, **131**(39): 13932–13933 (2009).

[152] Zou, H.-X., Xie, X., Zheng, X.-D., & Li, S.-M. The tyrosine O-prenyltransferase SirD catalyzes O-, N-, and C-prenylations. *Applied Microbiology and Biotechnology*, **89**(5): 1443–1451 (2011).

[153] Pan, L.-L., Yang, Y., & Merz Jr, K. M. Origin of product selectivity in a prenyl transfer reaction from the same intermediate: exploration of multiple FtmPT1-catalyzed prenyl transfer pathways. *Biochemistry*, **53**(38): 6126–6138 (2014).

[154] Isogai, S., Okahashi, N., Asama, R., Nakamura, T., *et al.* Synthetic production of prenylated naringenins in yeast using promiscuous microbial prenyltransferases. *Metabolic Engineering Communications*, **12**: e00169 (2021).

[155] Yin, W.-B., Ruan, H.-L., Westrich, L., Grundmann, A., & Li, S.-M. CdpNPT, an N-prenyltransferase from *Aspergillus fumigatus*: overproduction, purification and biochemical characterisation. *Chembiochem*, **8**(10): 1154–1161 (2007).

[156] Li, S.-M. Prenylated indole derivatives from fungi: structure diversity, biological activities, biosynthesis and chemoenzymatic synthesis. *Natural Product Reports*, **27**(1): 57–78 (2010).

[157] Wunsch, C., Mundt, K., & Li, S.-M. Targeted production of secondary metabolites by coexpression of non-ribosomal peptide synthetase and prenyltransferase genes in *Aspergillus*. *Applied Microbiology and Biotechnology*, **99**(10): 4213–4223 (2015).

[158] Gardner, E. D., Dimas, D. A., Finneran, M. C., Brown, S. M., *et al.* Indole C6 functionalization of Tryprostatin B using prenyltransferase CdpNPT. *Catalysts*, **10**(11): 1247 (2020).

Bibliography

[159] Mai, P., Zocher, G., Stehle, T., & Li, S.-M. Structure-based protein engineering enables prenyl donor switching of a fungal aromatic prenyltransferase. *Organic & Biomolecular Chemistry*, **16**(40): 7461–7469 (2018).

[160] Fan, A. & Li, S.-M. Saturation mutagenesis on Arg244 of the tryptophan C4-prenyltransferase FgaPT2 leads to enhanced catalytic ability and different preferences for tryptophan-containing cyclic dipeptides. *Applied Microbiology and Biotechnology*, **100**(12): 5389–5399 (2016).

[161] Luk, L. Y., Qian, Q., & Tanner, M. E. A Cope rearrangement in the reaction catalyzed by dimethylallyltryptophan synthase? *Journal of the American Chemical Society*, **133**(32): 12342–12345 (2011).

[162] Zhou, K., Zhao, W., Liu, X.-Q., & Li, S.-M. Saturation mutagenesis on Tyr205 of the cyclic dipeptide C2-prenyltransferase FtmPT1 results in mutants with strongly increased C3-prenylating activity. *Applied Microbiology and Biotechnology*, **100**(23): 9943–9953 (2016).

[163] Subramanian, S., Shen, X., Yuan, Q., & Yan, Y. Identification and biochemical characterization of a 5-dimethylallyl tryptophan synthase in *Streptomyces coelicolor* A3 (2). *Process Biochemistry*, **47**(9): 1419–1422 (2012).

[164] Mai, P., Zocher, G., Ludwig, L., Stehle, T., & Li, S.-M. Actions of tryptophan prenyltransferases toward fumiquinazolines and their potential application for the generation of prenylated derivatives by combining chemical and chemoenzymatic syntheses. *Advanced Synthesis & Catalysis*, **358**(10): 1639–1653 (2016).

[165] Winkelblech, J., Xie, X., & Li, S.-M. Characterisation of 6-DMATS Mo from *Micromonospora olivasterospora* leading to identification of the divergence in enantioselectivity, regioselectivity and multiple prenylation of tryptophan prenyltransferases. *Organic & Biomolecular Chemistry*, **14**(41): 9883–9895 (2016).

[166] Varki, A. Evolutionary forces shaping the Golgi glycosylation machinery: why cell surface glycans are universal to living cells. *Cold Spring Harbor Perspectives in Biology*, **3**(6): a005462 (2011).

[167] Varki, A. & Lowe, J. B. Biological roles of glycans. *Essentials of Glycobiology. 2nd edition*, **27**(1): 3–49 (2009).

[168] Möckl, L. The emerging role of the mammalian glycocalyx in functional membrane organization and immune system regulation. *Frontiers in Cell and Developmental Biology*, **8**(253): 1–14 (2020).

[169] Formosa-Dague, C., Castelain, M., Martin-Yken, H., Dunker, K., *et al.* The role of glycans in bacterial adhesion to mucosal surfaces: How can single-molecule techniques advance our understanding? *Microorganisms*, **6**(2): 39 (2018).

[170] Kleene, R. & Schachner, M. Glycans and neural cell interactions. *Nature Reviews Neuroscience*, **5**(3): 195–208 (2004).

[171] Cummings, R. D. The repertoire of glycan determinants in the human glycome. *Molecular BioSystems*, **5**(10): 1087–1104 (2009).

[172] Maccioni, H. J., Giraudo, C. G., & Daniotti, J. L. Understanding the stepwise synthesis of glycolipids. *Neurochemical Research*, **27**(7): 629–636 (2002).

[173] Fleischer, B., Fleischer, S., & Ozawa, H. Isolation and characterization of Golgi membranes from bovine liver. *The Journal of Cell Biology*, **43**(1): 59–79 (1969).

[174] Huang, S. & Wang, Y. Golgi structure formation, function, and post-translational modifications in mammalian cells. *F1000Research*, **6**(2050): 1–14 (2017).

[175] Marshall, R. Glycoproteins. *Annual Review of Biochemistry*, **41**(1): 673–702 (1972).

[176] Hassell, J. R., Kimura, J. H., & Hascall, V. C. Proteoglycan core protein families. *Annual Review of Biochemistry*, **55**(1): 539–567 (1986).

[177] Jennings, H., Katzenellenbogen, E., Lugowski, C., Michon, F., *et al.* Structure, conformation and immunology of sialic acid-containing polysaccharides of human pathogenic bacteria. *Pure and Applied Chemistry*, **56**(7): 893–905 (1984).

[178] Warren, L. The distribution of sialic acids in nature. *Comparative Biochemistry and Physiology*, **10**(2): 153–171 (1963).

[179] Corfield, A. P., Wember, M., Schauer, R., & Rott, R. The specificity of viral sialidases: The use of oligosaccharide substrates to probe enzymic characteristics and strain-specific differences. *European Journal of Biochemistry*, **124**(3): 521–525 (1982).

[180] Vliegenthart, J., Kamerling, J., Schauer, R., & Hotta, K. Identification of the sialic acids from egg jelly coat of the sea urchin pseudocentrotus depressus (Okayama). *Hoppe-Seyler's Zeitschrift für physiologische Chemie*, **361**: 1511–1516 (1980).

[181] Nadano, D., Iwasaki, M., Endo, S., Kitajima, K., *et al.* A naturally occurring deaminated neuraminic acid, 3-deoxy-D-glycero-D-galacto-nonulosonic acid (KDN). Its unique occurrence at the nonreducing ends of oligosialyl chains in polysialoglycoprotein of rainbow trout eggs. *Journal of Biological Chemistry*, **261**(25): 11550–11557 (1986).

[182] Varki, A. Diversity in the sialic acids. *Glycobiology*, **2**(1): 25 (1992).

[183] Schauer, R., Schoop, H. J., & Faillard, H. Zur Biosynthese der Glykolyl-Gruppe der N-Glykolyl-neuraminsäure. Die oxydative Umwandlung der N-Acetyl-Gruppe zur Glykolyl-Gruppe. *Biological Chemistry*, **349**(1): 645–652 (1968).

[184] Gagneux, P. Great apes and humans: Genetic differences. *eLS*, **18**(1): 2–13 (2006).

[185] Muchmore, E. A., Diaz, S., & Varki, A. A structural difference between the cell surfaces of humans and the great apes. *American Journal of Physical Anthropology*, **107**(2): 187–198 (1998).

[186] Suzuki, A. Genetic basis for the lack of N-glycolylneuraminic acid expression in human tissues and its implication to human evolution. *Proceedings of the Japan Academy, Series B*, **82**(3): 93–103 (2006).

[187] Mandal, C., Schwartz-Albiez, R., & Vlasak, R. Functions and biosynthesis of O-acetylated sialic acids. *SialoGlyco Chemistry and Biology I*, **366**: 1–30 (2012).

[188] Klein, A. & Roussel, P. O-acetylation of sialic acids. *Biochimie*, **80**(1): 49–57 (1998).

[189] Kochetkov, N., Smirnova, G., & Chekareva, N. Isolation and structural studies of a sulfated sialosphingolipid from the sea urchin *Echinocardium cordatom*. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*, **424**(2): 274–283 (1976).

[190] Smirnova, G., Kochetkov, N., & Sadovskaya, V. Gangliosides of the starfish *Aphelasterias japonica*, evidence for a new linkage between two N-glycolylneuraminic acid residues through the hydroxy group of the glycolic acid residue. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*, **920**(1): 47–55 (1987).

[191] Corfield, A., Wagner, S. a., Clamp, J. R., & Mountford, R. A. Demonstration of 9-0-lactyl-N-acetylneuraminic acid in human gastric aspirates. *Bid*, **100**(372): 379 (1987).

[192] Watson, D. R., Jourdian, G. W., & Roseman, S. The Sialic Acids: VIII. sialic acid 9-phosphate synthetase. *Journal of Biological Chemistry*, **241**(23): 5627–5636 (1966).

[193] Reily, C., Stewart, T. J., Renfrow, M. B., & Novak, J. Glycosylation in health and disease. *Nature Reviews Nephrology*, **15**(6): 346–366 (2019).

[194] Neelamegham, S., Aoki-Kinoshita, K., Bolton, E., Frank, M., *et al.* Updates to the symbol nomenclature for glycans guidelines. *Glycobiology*, **29**(9): 620–624 (2019).

[195] Adlam, C., Knights, J., Mugridge, A., Williams, J. M., & Lindon, J. Production of colominic acid by *Pasteurella haemolytica* serotype A2 organisms. *FEMS Microbiology Letters*, **42**(1): 23–25 (1987).

[196] Volkers, G., Worrall, L. J., Kwan, D. H., Yu, C.-C., *et al.* Structure of human ST8SiaIII sialyltransferase provides insight into cell-surface polysialylation. *Nature Structural & Molecular Biology*, **22**(8): 627–635 (2015).

[197] Roth, J., Taatjes, D. J., Lucocq, J. M., Weinstein, J., & Paulson, J. C. Demonstration of an extensive trans-tubular network continuous with the Golgi apparatus stack that may function in glycosylation. *Cell*, **43**(1): 287–295 (1985).

[198] Moremen, K. W. & Haltiwanger, R. S. Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nature Chemical Biology*, **15**(9): 853–864 (2019).

[199] Franceschini, I., Angata, K., Ong, E., Hong, A., *et al.* Polysialyltransferase ST8Sia II (STX) polysialylates all of the major isoforms of NCAM and facilitates neurite outgrowth. *Glycobiology*, **11**(3): 231–239 (2001).

[200] Kiermaier, E., Moussion, C., Veldkamp, C. T., Gerardy-Schahn, R., *et al.* Polysialylation controls dendritic cell trafficking by regulating chemokine recognition. *Science*, **351**(6269): 186–190 (2016).

[201] Curreli, S., Arany, Z., Gerardy-Schahn, R., Mann, D., & Stamatos, N. M. Polysialylated neuropilin-2 is expressed on the surface of human dendritic cells and modulates dendritic cell-T lymphocyte interactions. *Journal of Biological Chemistry*, **282**(42): 30346–30356 (2007).

[202] Galuska, S. P., Rollenhagen, M., Kaup, M., Eggers, K., *et al.* Synaptic cell adhesion molecule SynCAM 1 is a target for polysialylation in postnatal mouse brain. *Proceedings of the National Academy of Sciences*, **107**(22): 10250–10255 (2010).

[203] Yabe, U., Sato, C., Matsuda, T., & Kitajima, K. Polysialic acid in human milk: CD36 is a new member of mammalian polysialic acid-containing glycoprotein. *Journal of Biological Chemistry*, **278**(16): 13875–13880 (2003).

[204] James, W. M. & Agnew, W. S. Multiple oligosaccharide chains in the voltage-sensitive Na channel from *Electrophorus electricus*: evidence for α2, 8-linked polysialic acid. *Biochemical and Biophysical Research Communications*, **148**(2): 817–826 (1987).

[205] Oltmann-Norden, I., Galuska, S. P., Hildebrandt, H., Geyer, R., *et al.* Impact of the polysialyltransferases ST8SiaII and ST8SiaIV on polysialic acid synthesis during postnatal mouse brain development. *Journal of Biological Chemistry*, **283**(3): 1463–1471 (2008).

[206] Nakayama, J., Angata, K., Ong, E., Katsuyama, T., & Fukuda, M. Polysialic acid, a unique glycan that is developmentally regulated by two polysialyltransferases, PST and STX, in the central nervous system: from biosynthesis to function. *Pathology International*, **48**(9): 665–677 (1998).

[207] Yamasaki, R. & Bacon, B. Three-dimensional structural analysis of the group B polysaccharide of *Neisseria meningitidis* 6275 by two-dimensional NMR: the

polysaccharide is suggested to exist in helical conformations in solution. *Biochemistry*, **30**(3): 851–857 (1991).

[208] Brisson, J. R., Baumann, H., Imberty, A., Perez, S., & Jennings, H. J. Helical epitope of the group B meningococcal. α2,8-linked sialic acid polysaccharide. *Biochemistry*, **31**(21): 4996–5004 (1992).

[209] Evans, S., Sigurskjold, B., Jennings, H., Brisson, J.-R., *et al.* Evidence for the extended helical nature of polysaccharide epitopes. The 2.8Å resolution structure and thermodynamics of ligand binding of an antigen binding fragment specific for α2,8-poly (sialic acid). *Biochemistry*, **34**(20): 6737–6744 (1995).

[210] Battistel, M. D., Shangold, M., Trinh, L., Shiloach, J., & Freedberg, D. I. Evidence for helical structure in a tetramer of α2-8 sialic acid: Unveiling a structural antigen. *Journal of the American Chemical Society*, **134**(26): 10717–10720 (2012).

[211] Turupcu, A., Blaukopf, M., Kosma, P., & Oostenbrink, C. Molecular conformations of di-, tri-, and tetra-α2, 8-linked sialic acid from NMR spectroscopy and MD simulations. *International Journal of Molecular Sciences*, **21**(1): 30 (2020).

[212] Sato, C., Kitajima, K., Inoue, S., Seki, T., *et al.* Characterization of the antigenic specificity of four different anti-α2,8-linked polysialic acid antibodies using lipid-conjugated oligo/polysialic acids. *Journal of Biological Chemistry*, **270**(32): 18923–18928 (1995).

[213] Lenman, A., Liaci, A. M., Liu, Y., Frängsmyr, L., *et al.* Polysialic acid is a cellular receptor for human adenovirus 52. *Proceedings of the National Academy of Sciences*, **115**(18): E4264–E4273 (2018).

[214] Weinhold, B., Seidenfaden, R., Röckle, I., Mühlenhoff, M., *et al.* Genetic ablation of polysialic acid causes severe neurodevelopmental defects rescued by deletion of the neural cell adhesion molecule. *Journal of Biological Chemistry*, **280**(52): 42971–42977 (2005).

[215] Seidenfaden, R., Krauter, A., Schertzinger, F., Gerardy-Schahn, R., & Hildebrandt, H. Polysialic acid directs tumor cell growth by controlling heterophilic neural cell adhesion molecule interactions. *Molecular and Cellular Biology*, **23**(16): 5908 (2003).

[216] Seidenfaden, R., Krauter, A., & Hildebrandt, H. The neural cell adhesion molecule NCAM regulates neuritogenesis by multiple mechanisms of interaction. *Neurochemistry International*, **49**(1): 1–11 (2006).

[217] Suzuki, M., Angata, K., Nakayama, J., & Fukuda, M. Polysialic acid and mucin type O-glycans on the neural cell adhesion molecule differentially regulate myoblast fusion. *Journal of Biological Chemistry*, **278**(49): 49459–49468 (2003).

[218] Colley, K. J., Kitajima, K., & Sato, C. Polysialic acid: Biosynthesis, novel functions and applications. *Critical Reviews in Biochemistry and Molecular Biology*, **49**(6): 498–532 (2014).

[219] Röckle, I., Seidenfaden, R., Weinhold, B., Mühlenhoff, M., *et al.* Polysialic acid controls NCAM-induced differentiation of neuronal precursors into calretinin-positive olfactory bulb interneurons. *Developmental Neurobiology*, **68**(9): 1170–1184 (2008).

[220] Hildebrandt, H., Mühlenhoff, M., Weinhold, B., & Gerardy-Schahn, R. Dissecting polysialic acid and NCAM functions in brain development. *Journal of Neurochemistry*, **103**: 56–64 (2007).

[221] Nacher, J., Guirado, R., & Castillo-Gómez, E. Structural plasticity of interneurons in the adult brain: role of PSA-NCAM and implications for psychiatric disorders. *Neurochemical research*, **38**(6): 1122–1133 (2013).

[222] Rutishauser, U. Polysialic acid in the plasticity of the developing and adult vertebrate nervous system. *Nature Reviews Neuroscience*, **9**(1): 26–35 (2008).

[223] Sapoń, K., Janas, T., Sikorski, A. F., & Janas, T. Polysialic acid chains exhibit enhanced affinity for ordered regions of membranes. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, **1861**(1): 245–255 (2019).

[224] Kanato, Y., Kitajima, K., & Sato, C. Direct binding of polysialic acid to a brain-derived neurotrophic factor depends on the degree of polymerization. *Glycobiology*, **18**(12): 1044–1053 (2008).

[225] Ono, S., Hane, M., Kitajima, K., & Sato, C. Novel regulation of fibroblast growth factor 2 (FGF2)-mediated cell growth by polysialic acid. *Journal of Biological Chemistry*, **287**(6): 3710–3722 (2012).

[226] Shimoda, Y., Kitajima, K., Inoue, S., & Inoue, Y. Calcium ion binding of three different types of oligo/polysialic acids as studied by equilibrium dialysis and circular dichroic methods. *Biochemistry*, **33**(5): 1202–1208 (1994).

[227] Ulm, C., Saffarzadeh, M., Mahavadi, P., Müller, S., *et al.* Soluble polysialylated NCAM: a novel player of the innate immune system in the lung. *Cellular and Molecular Life Sciences*, **70**(19): 3695–3708 (2013).

[228] Hänsch, M., Simon, P., Schön, J., Kaese, M., *et al.* Polysialylation of NCAM correlates with onset and termination of seasonal spermatogenesis in roe deer. *Glycobiology*, **24**(6): 488–493 (2014).

[229] Hromatka, B. S., Drake, P. M., Kapidzic, M., Stolp, H., *et al.* Polysialic acid enhances the migration and invasion of human cytotrophoblasts. *Glycobiology*, **23**(5): 593–602 (2013).

[230] Werneburg, S., Buettner, F. F., Erben, L., Mathews, M., *et al.* Polysialylation and lipopolysaccharide-induced shedding of E-selectin ligand-1 and neuropilin-2 by microglia and THP-1 macrophages. *GLIA*, **64**(8): 1314–1330 (2016).

[231] Moebius, J. M., Widera, D., Schmitz, J., Kaltschmidt, C., & Piechaczek, C. Impact of polysialylated CD56 on natural killer cell cytotoxicity. *BMC Immunology*, **8**(1): 1–9 (2007).

[232] Hahn, S., Giaglis, S., Hoesli, I., & Hasler, P. Neutrophil NETs in reproduction: from infertility to preeclampsia and the possibility of fetal loss. *Frontiers in Immunology*, **3**: 362 (2012).

[233] Zlatina, K., Lütteke, T., & Galuska, S. P. Individual impact of distinct polysialic acid chain lengths on the cytotoxicity of histone H1, H2a, H2b, H3 and H4. *Polymers*, **9**(12): 720 (2017).

[234] Kühnle, A., Veelken, R., Galuska, C. E., Saftenberger, M., *et al.* Polysialic acid interacts with lactoferrin and supports its activity to inhibit the release of neutrophil extracellular traps. *Carbohydrate Polymers*, **208**: 32–41 (2019).

[235] Drake, P. M., Stock, C. M., Nathan, J. K., Gip, P., *et al.* Polysialic acid governs T-cell development by regulating progenitor access to the thymus. *Proceedings of the National Academy of Sciences*, **106**(29): 11995–12000 (2009).

[236] Villanueva-Cabello, T. M., Gutiérrez-Valenzuela, L. D., López-Guerrero, D. V., Cruz-Muñoz, M. E., *et al.* Polysialic acid is expressed in human naïve CD4$^+$ T cells and is involved in modulating activation. *Glycobiology*, **29**(7): 557–564 (2019).

[237] Shahraz, A., Kopatz, J., Mathy, R., Kappler, J., *et al.* Anti-inflammatory activity of low molecular weight polysialic acid on human macrophages. *Scientific Reports*, **5**(1): 1–17 (2015).

[238] Stamatos, N. M., Zhang, L., Jokilammi, A., Finne, J., *et al.* Changes in polysialic acid expression on myeloid cells during differentiation and recruitment to sites of inflammation: role in phagocytosis. *Glycobiology*, **24**(9): 864–879 (2014).

[239] Williams, A. F. & Barclay, A. N. The immunoglobulin superfamily-domains for cell surface recognition. *Annual Review of Immunology*, **6**(1): 381–405 (1988).

[240] Powell, L. D. & Varki, A. I-type lectins. *Journal of Biological Chemistry*, **270**(24): 14243–14246 (1995).

[241] Crocker, P. R., Paulson, J. C., & Varki, A. Siglecs and their roles in the immune system. *Nature Reviews Immunology*, **7**(4): 255–266 (2007).

[242] Trapp, B. D., Andrews, S. B., Cootauco, C., & Quarles, R. The myelin-associated glycoprotein is enriched in multivesicular bodies and periaxonal membranes of actively myelinating oligodendrocytes. *The Journal of Cell Biology*, **109**(5): 2417–2426 (1989).

[243] Brinkman-Van der Linden, E. C., Hurtado-Ziola, N., Hayakawa, T., Wiggleton, L., *et al.* Human-specific expression of Siglec-6 in the placenta. *Glycobiology*, **17**(9): 922–931 (2007).

[244] Nitschke, L., Carsetti, R., Ocker, B., Köhler, G., & Lamers, M. C. CD22 is a negative regulator of B-cell receptor signalling. *Current Biology*, **7**(2): 133–143 (1997).

[245] Crocker, P., Kelm, S., Dubois, C., Martin, B., *et al.* Purification and properties of sialoadhesin, a sialic acid-binding receptor of murine tissue macrophages. *The EMBO Journal*, **10**(7): 1661–1669 (1991).

[246] Macauley, M. S., Crocker, P. R., & Paulson, J. C. Siglec-mediated regulation of immune cell function in disease. *Nature Reviews Immunology*, **14**(10): 653–666 (2014).

[247] Nitschke, L. CD22 and Siglec-G: B-cell inhibitory receptors with distinct functions. *Immunological reviews*, **230**(1): 128–143 (2009).

[248] Yin, S.-S. & Gao, F.-H. Molecular mechanism of tumor cell immune escape mediated by CD24/Siglec-10. *Frontiers in Immunology*, **11**: 1324 (2020).

[249] Bornhöfft, K. F., Goldammer, T., Rebl, A., & Galuska, S. P. Siglecs: A journey through the evolution of sialic acid-binding immunoglobulin-type lectins. *Developmental & Comparative Immunology*, **86**: 219–231 (2018).

[250] Angata, T. Molecular diversity and evolution of the Siglec family of cell-surface lectins. *Molecular Diversity*, **10**(4): 555–566 (2006).

[251] Angata, T., Hayakawa, T., Yamanaka, M., Varki, A., & Nakamura, M. Discovery of Siglec-14, a novel sialic acid receptor undergoing concerted evolution with Siglec-5 in primates. *The FASEB Journal*, **20**(12): 1964–1973 (2006).

[252] Angata, T., Margulies, E. H., Green, E. D., & Varki, A. Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms. *Proceedings of the National Academy of Sciences*, **101**(36): 13251–13256 (2004).

[253] Mitra, N., Banda, K., Altheide, T. K., Schaffer, L., *et al.* Siglec-12, a human-specific segregating (pseudo) gene, encodes a signaling molecule expressed in prostate carcinomas. *Journal of Biological Chemistry*, **286**(26): 23003–23011 (2011).

[254] Wang, X., Mitra, N., Secundino, I., Banda, K., *et al.* Specific inactivation of two immunomodulatory Siglec genes during human evolution. *Proceedings of the National Academy of Sciences*, **109**(25): 9935–9940 (2012).

[255] Razi, N. & Varki, A. Masking and unmasking of the sialic acid-binding lectin activity of CD22 (Siglec-2) on B lymphocytes. *Proceedings of the National Academy of Sciences*, **95**(13): 7469–7474 (1998).

[256] Meyer, S. J., Linder, A. T., Brandl, C., & Nitschke, L. B cell Siglecs–news on signaling and its interplay with ligand binding. *Frontiers in Immunology*, **9**: 2820 (2018).

[257] Munday, J., Floyd, H., & Crocker, P. R. Sialic acid binding receptors (Siglecs) expressed by macrophages. *Journal of Leukocyte Biology*, **66**(5): 705–711 (1999).

[258] Floyd, H., Ni, J., Cornish, A. L., Zeng, Z., *et al.* Siglec-8: a novel eosinophil-specific member of the immunoglobulin superfamily. *Journal of Biological Chemistry*, **275**(2): 861–866 (2000).

[259] Son, M., Diamond, B., Volpe, B. T., Aranow, C. B., *et al.* Evidence for C1q-mediated crosslinking of CD33/LAIR-1 inhibitory immunoreceptors and biological control of CD33/LAIR-1 expression. *Scientific Reports*, **7**(1): 1–13 (2017).

[260] Delputte, P. L., Van Gorp, H., Favoreel, H. W., Hoebeke, I., *et al.* Porcine sialoadhesin (CD169/Siglec-1) is an endocytic receptor that allows targeted delivery of toxins and antigens to macrophages. *PloS One*, **6**(2): e16827 (2011).

[261] O'Reilly, M. K., Tian, H., & Paulson, J. C. CD22 is a recycling receptor that can shuttle cargo between the cell surface and endosomal compartments of B cells. *The Journal of Immunology*, **186**(3): 1554–1563 (2011).

[262] Winterstein, C., Trotter, J., & Krämer-Albers, E.-M. Distinct endocytic recycling of myelin proteins promotes oligodendroglial membrane remodeling. *Journal of Cell Science*, **121**(6): 834–842 (2008).

[263] Cao, H. & Crocker, P. R. Evolution of CD33-related siglecs: regulating host immune functions and escaping pathogen exploitation? *Immunology*, **132**(1): 18–26 (2011).

[264] Carlin, A. F., Uchiyama, S., Chang, Y.-C., Lewis, A. L., *et al.* Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage neutrophil Siglec-9 and dampen the innate immune response. *Blood*, **113**(14): 3333–3336 (2009).

[265] Gonzalez-Gil, A. & Schnaar, R. L. Siglec ligands. *Cells*, **10**(5): 1260 (2021).

[266] Rodriguez, E., Boelaars, K., Brown, K., Li, R. E., *et al.* Sialic acids in pancreatic cancer cells drive tumour-associated macrophage differentiation via the Siglec receptors Siglec-7 and Siglec-9. *Nature Communications*, **12**(1): 1–14 (2021).

[267] Siveen, K. & Kuttan, G. Role of macrophages in tumour progression. *Immunology Letters*, **123**(2): 97–102 (2009).

[268] Zou, Z., Chastain, A., Moir, S., Ford, J., *et al.* Siglecs facilitate HIV-1 infection of macrophages through adhesion with viral sialic acids. *PloS One*, **6**(9): e24559 (2011).

[269] Rempel, H., Calosing, C., Sun, B., & Pulliam, L. Sialoadhesin expressed on IFN-induced monocytes binds HIV-1 and enhances infectivity. *PloS One*, **3**(4): e1967 (2008).

[270] Angata, T., Kerr, S. C., Greaves, D. R., Varki, N. M., *et al.* Cloning and characterization of human Siglec-11: a recently evolved signaling molecule that can interact with SHP-1 and SHP-2 and is expressed by tissue macrophages, including brain microglia. *Journal of Biological Chemistry*, **277**(27): 24466–24474 (2002).

[271] Alphey, M. S., Attrill, H., Crocker, P. R., & van Aalten, D. M. High resolution crystal structures of Siglec-7: insights into ligand specificity in the Siglec family. *Journal of Biological Chemistry*, **278**(5): 3372–3377 (2003).

[272] Lenza, M. P., Atxabal, U., Oyenarte, I., Jiménez-Barbero, J., & Ereño-Orbea, J. Current status on therapeutic molecules targeting Siglec receptors. *Cells*, **9**(12): 2691 (2020).

[273] Ereño-Orbea, J., Sicard, T., Cui, H., Mazhab-Jafari, M. T., *et al.* Molecular basis of human CD22 function and therapeutic targeting. *Nature communications*, **8**(1): 1–11 (2017).

[274] Pronker, M. F., Lemstra, S., Snijder, J., Heck, A. J., *et al.* Structural basis of myelin-associated glycoprotein adhesion and signalling. *Nature Communications*, **7**(1): 1–13 (2016).

[275] Angata, T., Tabuchi, Y., Nakamura, K., & Nakamura, M. Siglec-15: an immune system Siglec conserved throughout vertebrate evolution. *Glycobiology*, **17**(8): 838–846 (2007).

[276] Cao, H., Lakner, U., de Bono, B., Traherne, J. A., *et al.* Siglec16 encodes a DAP12-associated receptor expressed in macrophages that evolved from its inhibitory counterpart Siglec11 and has functional and non-functional alleles in humans. *European Journal of Immunology*, **38**(8): 2303–2315 (2008).

[277] Ravetch, J. V. & Lanier, L. L. Immune inhibitory receptors. *Science*, **290**(5489): 84–89 (2000).

[278] Zhao, J., Brooks, D. M., & Lurie, D. I. Lipopolysaccharide-activated SHP-1-deficient motheaten microglia release increased nitric oxide, TNF-α, and IL-1β. *GLIA*, **53**(3): 304–312 (2006).

[279] Hane, M., Chen, D. Y., & Varki, A. Human-specific microglial Siglec-11 transcript variant has the potential to affect polysialic acid-mediated brain functions at a distance. *Glycobiology*, **31**(3): 231–242 (2021).

[280] Wang, Y. & Neumann, H. Alleviation of neurotoxicity by microglial human Siglec-11. *Journal of Neuroscience*, **30**(9): 3482–3488 (2010).

[281] Yoshimura, A., Hatanaka, R., Tanaka, H., Kitajima, K., & Sato, C. The conserved arginine residue in all Siglecs is essential for Siglec-7 binding to sialic acid. *Biochemical and Biophysical Research Communications*, **534**: 1069–1075 (2021).

[282] Angata, T. & Varki, A. Cloning, characterization, and phylogenetic analysis of Siglec-9, a new member of the CD33-related group of Siglecs: evidence for co-evolution with sialic acid synthesis pathways. *Journal of Biological Chemistry*, **275**(29): 22127–22135 (2000).

[283] Schwarz, F., Landig, C. S., Siddiqui, S., Secundino, I., *et al.* Paired Siglec receptors generate opposite inflammatory responses to a human-specific pathogen. *The EMBO journal*, **36**(6): 751–760 (2017).

[284] Wielgat, P., Rogowski, K., Niemirowicz-Laskowska, K., & Car, H. Sialic acid-Siglec axis as molecular checkpoints targeting of immune system: smart players in pathology and conventional therapy. *International Journal of Molecular Sciences*, **21**(12): 4361 (2020).

[285] Wang, X., Mitra, N., Cruz, P., Deng, L., *et al.* Evolution of Siglec-11 and Siglec-16 genes in hominins. *Molecular Biology and Evolution*, **29**(8): 2073–2086 (2012).

[286] Jumper, J., Evans, R., Pritzel, A., Green, T., *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**(7873): 583–589 (2021).

[287] Varadi, M., Anyango, S., Deshpande, M., Nair, S., *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, **50**(D1): D439–D444 (2022).

[288] Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, **98**(18): 10037–10041 (2001).

[289] Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., *et al.* Domain enhanced lookup time accelerated BLAST. *Biology Direct*, **7**(1): 1–14 (2012).

[290] Stein, N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography*, **41**(3): 641–643 (2008).

[291] Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, **67**(4): 235–242 (2011).

[292] Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, **7**(1): 539 (2011).

[293] Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. Features and development of COOT. *Acta Crystallographica Section D: Biological Crystallography*, **66**(4): 486–501 (2010).

[294] Yang, J. & Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Research*, **43**(W1): W174–W181 (2015).

[295] Gupta, R. & Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pacific Symposium on Biocomputing*, **7**: 310–22 (2001).

[296] Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., *et al.* NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate Journal*, **15**(2): 115–130 (1998).

[297] Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., *et al.* Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *The EMBO Journal*, **32**(10): 1478–1488 (2013).

[298] Read, R. J., Adams, P. D., & McCoy, A. J. Intensity statistics in the presence of translational noncrystallographic symmetry. *Acta Crystallographica Section D: Biological Crystallography*, **69**(2): 176–183 (2013).

[299] Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, **66**(2): 213–221 (2010).

[300] Sigrist, C. J. A., Castro, E., Cerutti, L., Cuche, B. A., *et al.* New and continuing developments at PROSITE. *Nucleic Acids Research*, **41**(D1): D344–D347 (2012).

[301] Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., *et al.* Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, **52**: 571–607 (2005).

[302] Schrödinger, L. & DeLano, W. Pymol version 2.4.0. `http://www.pymol.org/pymol` (2020).

Bibliography

[303] Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, **67**(4): 355–367 (2011).

[304] Letunic, I., Khedkar, S., & Bork, P. SMART: recent updates, new developments and status in 2020. *Nucleic Acids Research*, **49**(D1): D458–D460 (2021).

[305] Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, **42**(W1): W252–W258 (2014).

[306] Kabsch, W. XDS. *Acta Crystallographica Section D: Biological Crystallography*, **66**(2): 125–132 (2010).

[307] Attrill, H., Imamura, A., Sharma, R. S., Kiso, M., *et al.* Siglec-7 undergoes a major conformational change when complexed with the α 2, 8-disialylganglioside GT1b. *Journal of Biological Chemistry*, **281**(43): 32774–32783 (2006).

[308] Zhuravleva, M. A., Trandem, K., & Sun, P. D. Structural implications of Siglec-5-mediated sialoglycan recognition. *Journal of Molecular Biology*, **375**(2): 437–447 (2008).

[309] Moremen, K. W. & Touster, O. Biosynthesis and modification of Golgi mannosidase II in HeLa and 3T3 cells. *Journal of Biological Chemistry*, **260**(11): 6654–6662 (1985).

[310] Treter, V., Altmann, F., & März, L. Peptide-n4-(n-acetyl-β-glucosaminyl) asparagine amidase f cannot release glycans with fucose attached α1-3 to the asparagine-linked n-acetylglucosamine residue. *European Journal of Biochemistry*, **199**(3): 647–652 (1991).

[311] Morelle, W. & Michalski, J.-C. Analysis of protein glycosylation by mass spectrometry. *Nature Protocols*, **2**(7): 1585–1602 (2007).

[312] Krieg, J., Hartmann, S., Vicentini, A., Gläsner, W., *et al.* Recognition signal for C-mannosylation of Trp-7 in RNase 2 consists of sequence Trp-xx-Trp. *Molecular Biology of the Cell*, **9**(2): 301–309 (1998).

[313] Miles, L. A., Hermans, S. J., Crespi, G. A., Gooi, J. H., *et al.* Small molecule binding to Alzheimer risk factor CD33 promotes Aβ phagocytosis. *Iscience*, **19**: 110–118 (2019).

[314] Siddiqui, S., Schwarz, F., Springer, S., Khedri, Z., *et al.* Studies on the detection, expression, glycosylation, dimerization, and ligand binding properties of mouse Siglec-E. *Journal of Biological Chemistry*, **292**(3): 1029–1037 (2017).

[315] Handa, K., Takatani-Nakase, T., Larue, L., Stemmler, M. P., *et al.* Lex glycan mediates homotypic adhesion of embryonal cells independently from E-cadherin: A preliminary note. *Biochemical and Biophysical Research Communications*, **358**(1): 247–252 (2007).

[316] Spillings, B. L., Day, C. J., Garcia-Minambres, A., Aggarwal, A., *et al.* Host glycocalyx captures HIV proximal to the cell surface via oligomannose-GlcNAc glycan-glycan interactions to support viral entry. *Cell Reports*, **38**(5): 110296 (2022).

[317] Harbury, P. B., Zhang, T., Kim, P. S., & Alber, T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science*, **262**(5138): 1401–1407 (1993).

[318] Consortium, T. U. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, **49**(D1): D480–D489 (2021).

A Appendix

A.1 Siglec-11

A.1.1 Materials and Methods

Protein Sequence

```
        10         20         30         40         50         60
MYRMQLLSCI ALSLALVTNS ILNKDPSYSL QVQRQVPVPE GLCVIVSCNL SYPRDGWDES

        70         80         90        100        110        120
TAAYGYWFKG RTSPKTGAPV ATNNQSREVE MSTRDRFQLT GDPGKGSCSL VIRDAQREDE

       130        140        150        160        170        180
AWYFFRVERG SRVRHSFLSN AFFLKVTALT KKPDVYIPET LEPGQPVTVI CVFNWAFKKC

       190        200        210        220        230        240
PAPSFSWTGA ALSPRRTRPS TSHFSVLSFT PSPQDHDTDL TCHVDFSRKG VSAQRTVRLR

       250        260        270        280        290        300
VAYAPKDLII SISHDNTSAL ELQGNVIYLE VQKGQFLRLL CAADSQPPAT LSWVLQDRVL

       310        320        330        340        350        360
SSSHPWGPRT LGLELRGVRA GDSGRYTCRA ENRLGSQQQA LDLSVQYPPE NLRVMVSQAN

       370        380        390        400        410        420
RTVLENLGNG TSLPVLEGQS LRLVCVTHSS PPARLSWTRW GQTVGPSQPS DPGVLELPPI

       430        440        450        460        470        480
QMEHEGEFTC HAQHPLGSQH VSLSLSVHWK LEHGSGGSEN LYFQGGARGHP FEQKLISEED

LNMHTGHHHH HH
```

Figure A.1: Sequence Siglec-11. Protein sequence (green) of the Siglec-11$_{\text{large V2}}$. IL-2 based secretion signal (grey) was included on the N-terminus instead of the secretion signal encoded in the human genome. The protein sequence contains a TEV protease cleavage site (red), a c-myc detection tag (orange) and an His$_6$ affinity tag (brown).

Cloning

Table A.1: Primer sequences used for cloning. Template DNA and primer, with their annealing temperature are listed, which were used for In-fusion cloning and site-directed mutagenesis (SDM). Properties of the resulting constructs are listed in table A.2.

Construct	Template	Primer Sequence(5'-3')	T_M (°C)	Method
Siglec-11$_{\text{short V2}}$	pFUSE vector	TGAGTGCTAGCTGGCCAGAC	62.5	In-fusion
		CATGGTGGCCCTCCTTCG	60.7	
	pcDNA3.1 Siglec-11$_{\text{middle V1}}$	CTGGCCAGCTAGCACTCAATGGTGATGGTGATGATGACC	68.2	
		CGAAGGAGGGGCCACCATGTACAGGATGCAGCTGCTGTC	64.8	
Siglec-11$_{\text{middle V2}}$	pFUSE vector	TGAGTGCTAGCTGGCCAGAC	62.5	In-fusion
		CATGGTGGCCCCTCCTTCG	60.7	
	pcDNA3.1 Siglec-11$_{\text{middle V1}}$	CTGGCCAGCTAGCACTCAATGGTGATGGTGATGATGACC	59.6	
		CGAAGGAGGGGCCACCATGTACAGGATGCAGCTGCTGTC	64.8	
Siglec-11$_{\text{large V2}}$	pFUSE vector	TGAGTGCTAGCTGGCCAGAC	62.5	In-fusion
		CATGGTGGCCCCTCCTTCG	60.7	
	pcDNA3.1 Siglec-11$_{\text{large V1}}$	CGAAGGAGGGGCCACCATGTACAGGATGCAGCTGCTGTC	59.4	
		CGAAGGAGGGGCCACCATGTACAGGATGCAGCTGCTGTC	64.8	
Siglec-11$_{\text{1D}}$	Siglec-11$_{\text{middle V2}}$	CTTGTTCAGGGCGGTCAC	64.2	SDM
		CATCACCATCATCACCATCACCATTGAGTGCTA	67.1	
Siglec-11$_{\text{1D}\Delta\text{myc}}$	Siglec-11$_{\text{1D}}$	CTGAAAGGTCAGAATCTCGCTGC	66.0	SDM
		GAGAACCTGACCTTCCAGGG	65.3	
Siglec-11$_{\text{1D}\Delta\text{myc\_TEV}}$	Siglec-11$_{\text{1D}\Delta\text{myc}}$	GGCAGCGGAGAAATCTGACCTTTTC	66.0	SDM
		CTGAAAGTACAGATTCTCTCCGCTGCCAGAGC	70.0	
Siglec-11$_{\text{1D}\Delta\text{tag}}$	Siglec-11$_{\text{1D}\Delta\text{myc}}$	TGATCTGGCAGCGGAGG	65.0	SDM
		CTTGGTCAGGGCGGTCAC	66.0	
Siglec-11$_{\text{middle V3}}$	Siglec-11$_{\text{large V2}}$	GAGAACCTGACCTTCCAGGG	64.9	SDM
		GGGATAGTCCACGCTCAGATC	65.2	
Siglec-11$_{\text{large V3}}$	Siglec-11$_{\text{middle V2}}$	GAGAACCTGACCTTCCAGGG	64.9	SDM
		CATCACCATCATCACCATCACCATTGAGTGCTA	67.1	

Table A.2: Siglec-11 expression constructs. All expression constructs are listed with their size and vector backbone. All constructs comprimising the Kozak-sequence (CCGCC) in the vector, 3' of the start codon are labelled with a star (*). Each construct possesses either a Thrombin or TEV cleavage site and a His_6- or His_8-tag.

Construct	Length (amino acid)	Vector	Cleavage site & Tag	Size (kDa)
Siglec-11$_{1D}$	1-151	pFUSE*	TEV & c-myc-,His_8-tag	19.4
Siglec-11$_{1D\Delta myc}$	1-151	pFUSE*	TEV & His_8-tag	19.1
Siglec-11$_{1D\Delta myc\_TEV}$	1-151	pFUSE*	TEV_intract & His_8-tag	19.1
Siglec-11$_{1D\Delta tag}$	1-151	pFUSE*	-	17.1
Siglec-11$_{short V1}$	1-249	pcDNA3.1	TEV & c-myc-,His_6-tag	29.9
Siglec-11$_{short V2}$	1-249	pcDNA3.1*	TEV & c-myc-,His_6-tag	31.7
Siglec-11$_{middle V1}$	1-341	pcDNA3.1	TEV & c-myc-,His_8-tag	42.4
Siglec-11$_{middle V2}$	1-341	pFUSE*	TEV & c-myc-,His_6-tag	42.4
Siglec-11$_{middle V3}$	1-350	pFUSE*	TEV & c-myc-,His_8-tag	43.0
Siglec-11$_{large V1}$	1-481	pcDNA3.1	TEV & c-myc-,His_6-tag	56.9
Siglec-11$_{large V2}$	1-453	pcDNA3.1	TEV & c-myc-,His_6-tag	54.4
Siglec-11$_{large V3}$	1-453	pFUSE*	TEV & c-myc-,His_6-tag	54.4
Siglec-11$_{large V4}$	1-450	pcDNA3.4 TOPO*	Thrombin & His_6-tag	53.5

Calibration Line of the Analytical Size Exclusion
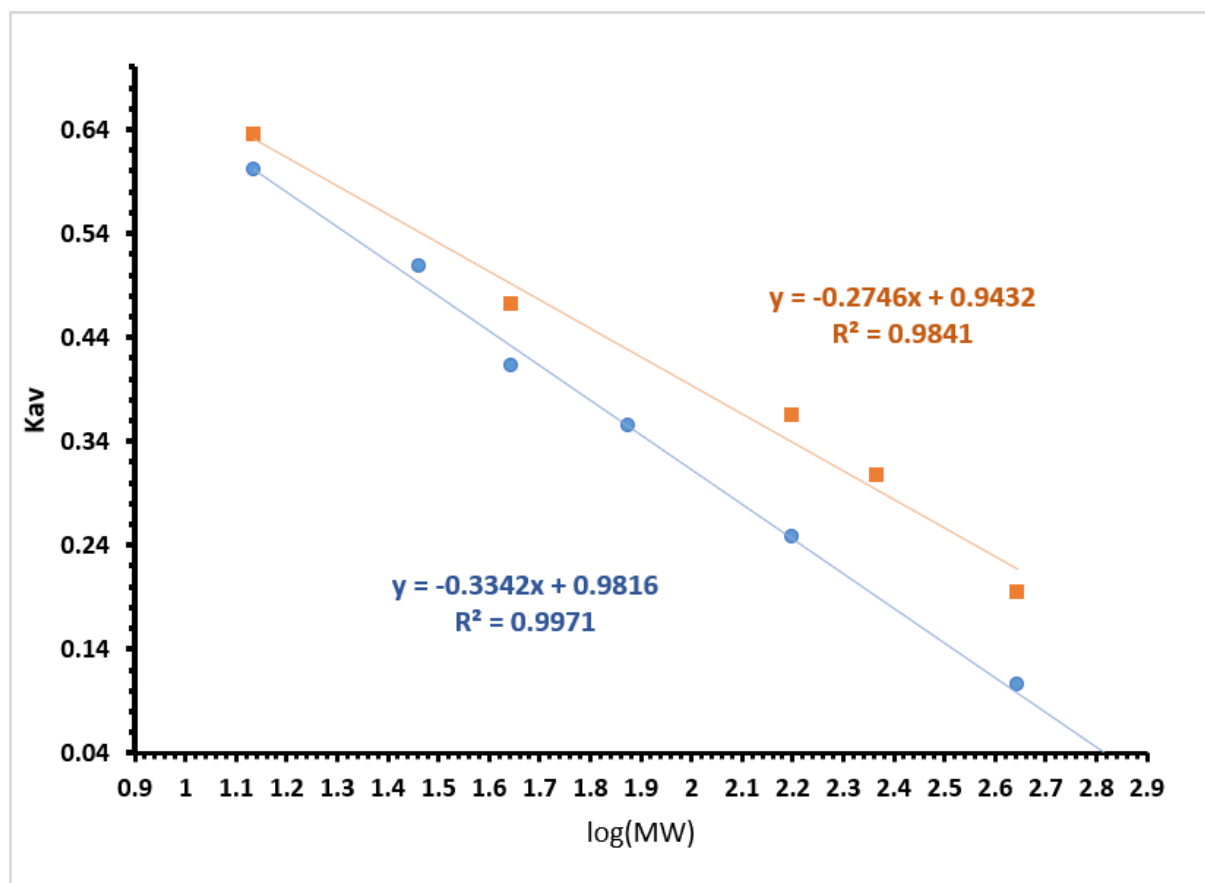
Figure A.2: Calibration line of the analytical SEC. For analysis of the apparent molecular weight of Siglec-11, the SD200 3.2/300 and SD75 3.2/300 increase columns were calibrated with seven protein of different sizes (Thyroglobulin (669 kDa); Ferrin (440 kDa), Aldolase (158 kDa), Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic anhydrase (29 kDa) and Ribonuclease A (13.7 kDa)) and blue dextran 2000 to determine the exclusion volume of the column (V_0). The K_{av} of each eluted protein was plotted against the logarithm of the molecular weight in blue and orange for SD200 and SD75 column, respectively.

Crystallisation Fine Screens

Table A.3: Fine screens of the initial crystallisation conditions. The concentration rang for the components in each fine screen is represented.

Fine Screen	Precipitant (%)	Salt (M)	Buffer (pH 0.1 M)
FS__CS__A2	PEG 4000 (24 - 34)	ammonium acetate (0.1 - 0.25)	sodium citrate tribasic (5.4 - 6.0)
FS__JCSG__E6	PEG 4000 (-)	ammonium sulfate (-)	
FS__JCSG+__E1		1 M sodium citrate,	0.1 M sodium cacodylate pH 6.0

A.1.2 Results

Expression and Purification
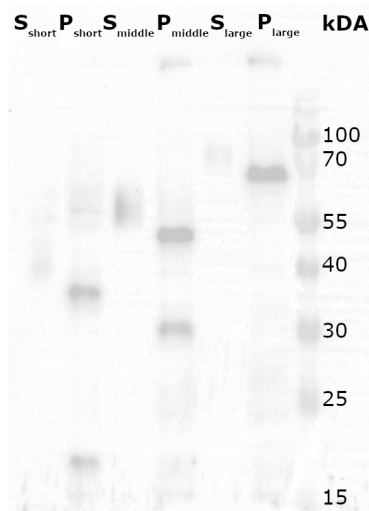
Figure A.3: Siglec-11 expression in Freestyle293F cells. Siglec-11$_{short}$,Siglec-11$_{middle}$ and Siglec-11$_{large\,V3}$ expression level in Freestyle293F cells was determined after six days by western blot analysis. The protein yield in the medium (S) and cell pellet (P) after centrifugation was identified with a primary mouse α-c-myc antibody and a secondary HRP-associated goat α-mouse antibody and the size is indicated by the pre-stained protein ladder (L).

Table A.4: Structural Homology of Siglecs. Structure-based alignment of V-sets from classical Siglecs (Siglec-1 and Siglec-4) and CD33-related Siglecs (Siglec-3, Siglec-5, Siglec-7, Siglec-8) are shown. Number of total residues in the single structures (No. AA), of aligned residues (No. aligned AA), overall Cα-RMSD and the sequence identity (Seq. ID) to Siglec-11 are listed. * Structure was solved by NMR and Cα-RMSD is calculated based on the 20 lowest energy conformations.

Siglec	No. of AA	No. of aligned AA	Cα-RMSD [Å]	Seq. ID [%]	pdb code
Siglec-7	113	109	0.8	47.7	1o7s
Siglec-3	213	119	1.1	45.4	5ihb
Siglec-5	214	117	1.2	50.4	2zg2
Siglec-8*	145	121	1.2	51.4	2n7a
Siglec-1	118	113	1.6	23.9	1od9
Siglec-4	307	118	1.7	21.2	5lfv

(A)

(B)
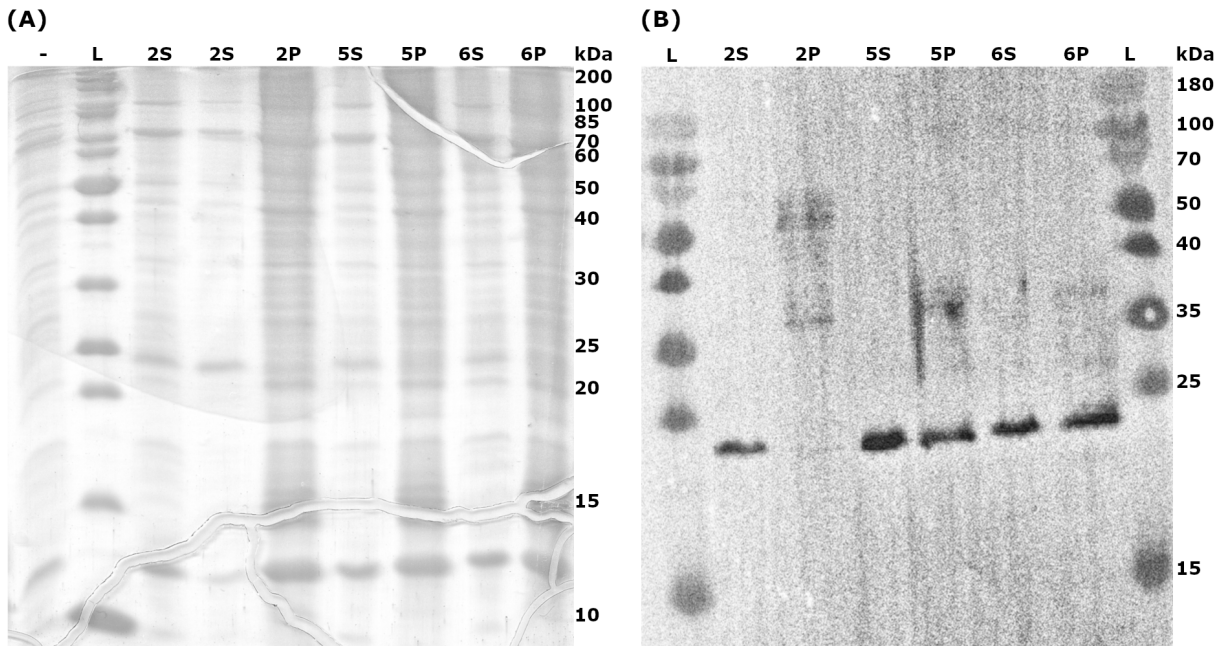
Figure A.4: Siglec-11$_{1D\Delta myc}$ expression in Expi293F GnTI- cells. Siglec-11$_{1D\Delta myc}$ expression level in the medium and cell pellet of Expi293F GnTI- cells was determined after two, five and six days by western blot analysis. The protein yield in the medium (2S, 5S, 6S) and the cell pellet (2P, 5P, 6P) after centrifugation was identified with primary mouse α-his antibody and secondary HRP-associated goat α-mouse antibody and the size is indicated by the pre-stained protein ladder (L).
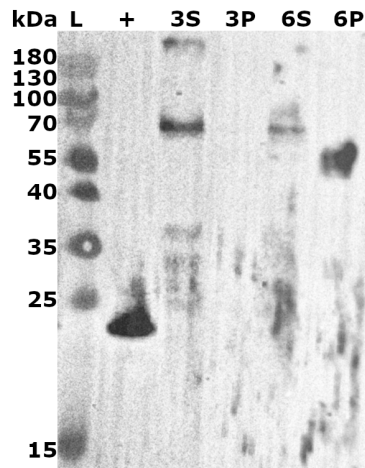
Figure A.5: Siglec-11$_{large\,V3}$ expression in Expi293F GnTI- cells. Siglec-11$_{large\,V3}$ expression level in the medium and cell pellet of Epi293Expi293F GnTI- were determined after three and six days by western blot analysis. The protein yield in the medium (3S, 6S) and cell pellet (3P, 6P) after centrifugation was identified with primary mouse α-his antibody and secondary HRP-associated goat α-mouse antibody, with Siglec-11$_{1D\Delta myc}$ as a positive control (+). The protein size is indicated by the pre-stained protein ladder (L).
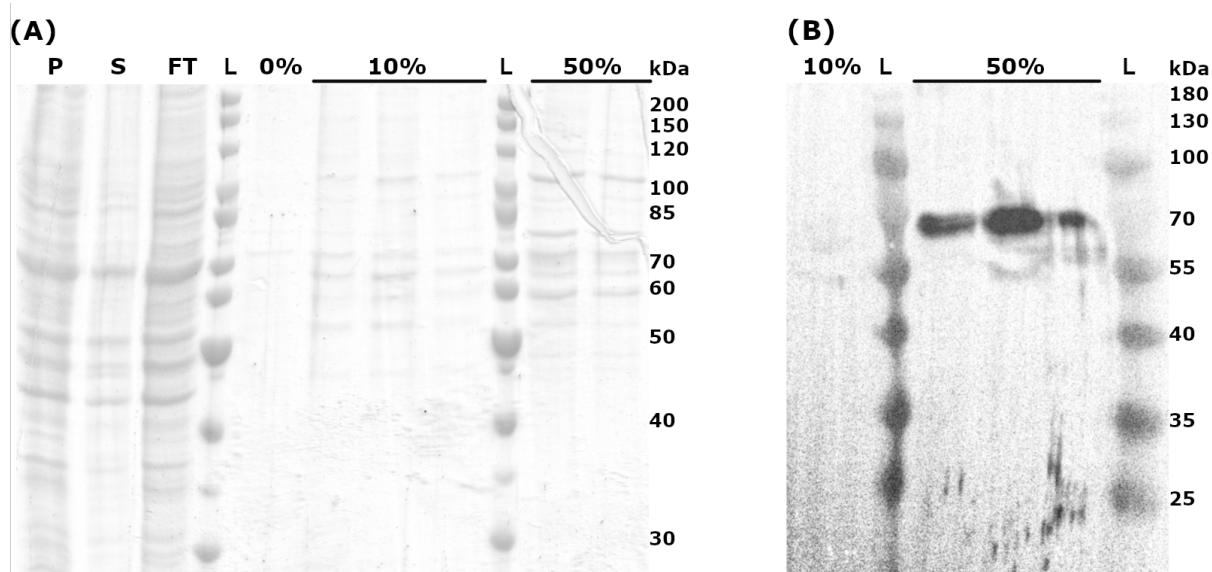
Figure A.6: IMAC of Siglec-11$_{\text{large V3}}$ from lysed Freestyle293F cells. (A) SDS-PAGE and, **(B)** the corresponding western blot of the Ni^{2+}-NTA affinity chromatography with a 5 mM (0%) and 55 mM imidazole (10% His-B1 buffer) washing steps to remove unspecific-bound protein, before the target protein was eluted with 250 mM imidazole (50% His-B1 buffer). Target protein was identified in the western blot with primary mouse α-his antibody and secondary HRP-associated goat α-mouse antibody. The protein size is indicated by an unstained and pre-stained protein ladder (L) in the SDS-PAGE and western blot, respectively.
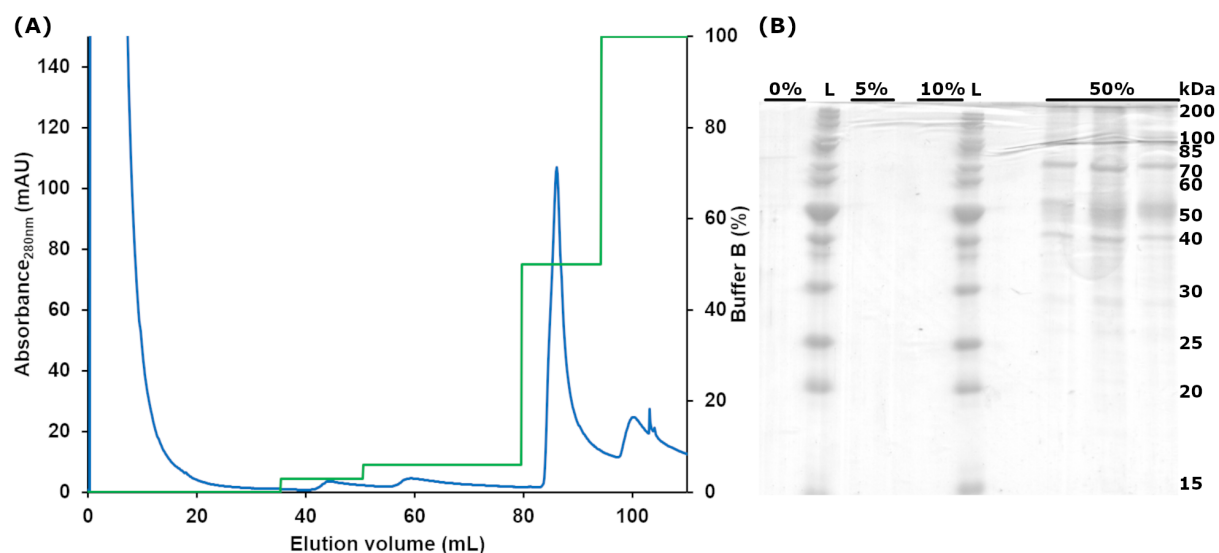
Figure A.7: IMAC of Siglec-11$_{\text{large V3}}$ expressed in Expi293 GnTI- cells (A) Chromatogram and **(B)** the corresponding SDS-PAGE of the Ni^{2+}-NTA affinity chromatography are represented. In the chromatogram, UV absorption at 280 nm and the Buffer B concentration are drawn as a solid blue and green line, respectively. The column was first washed with 5 mM (0%), 28 mM (5%) and 55 mM imidazole (10%) to remove unspecific-bound protein, before the target protein was eluted with 250 mM imidazole (50%). The protein size is indicated by the unstained protein ladder (L).
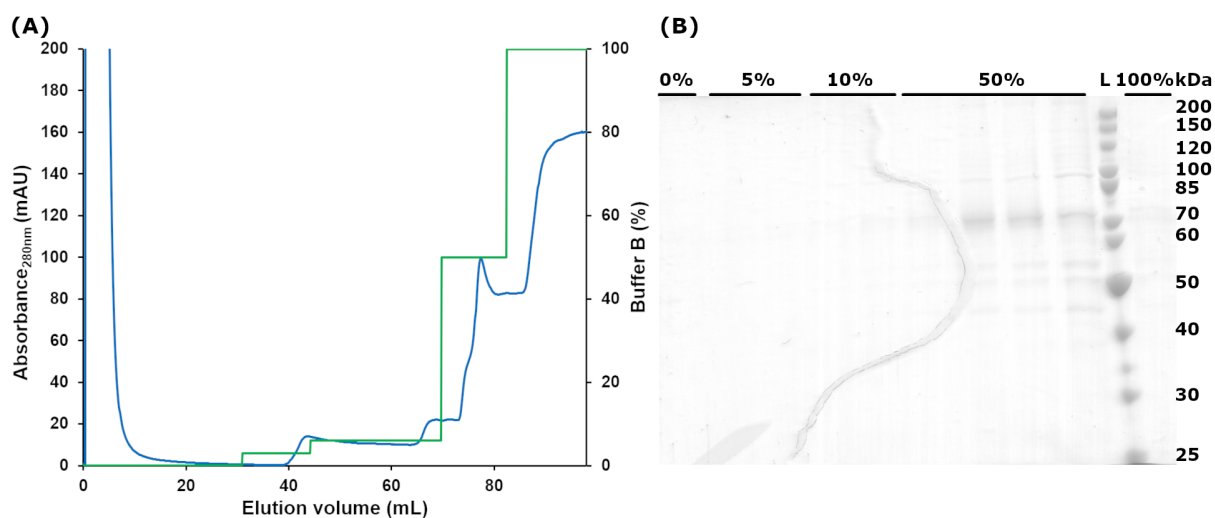
Figure A.8: IMAC of Siglec-11$_{\textbf{large V4}}$ expressed in Expi293 GnTI- cells. (A) Chromatogram and **(B)** the corresponding SDS-PAGE of the Ni^{2+}-NTA affinity chromatography are represented. In the chromatogram, UV absorption at 280 nm and the Buffer B concentration are drawn as a solid blue and green line, respectively. The column was first washed with 5 mM (0%), 28 mM imidazole (5%) and 55 mM imidazole (10%) to remove unspecific-bound protein, before the target protein was eluted with 250 mM imidazole (50%) and 500 mM imidazole (100%). The protein size is indicated by the unstained protein ladder (L).
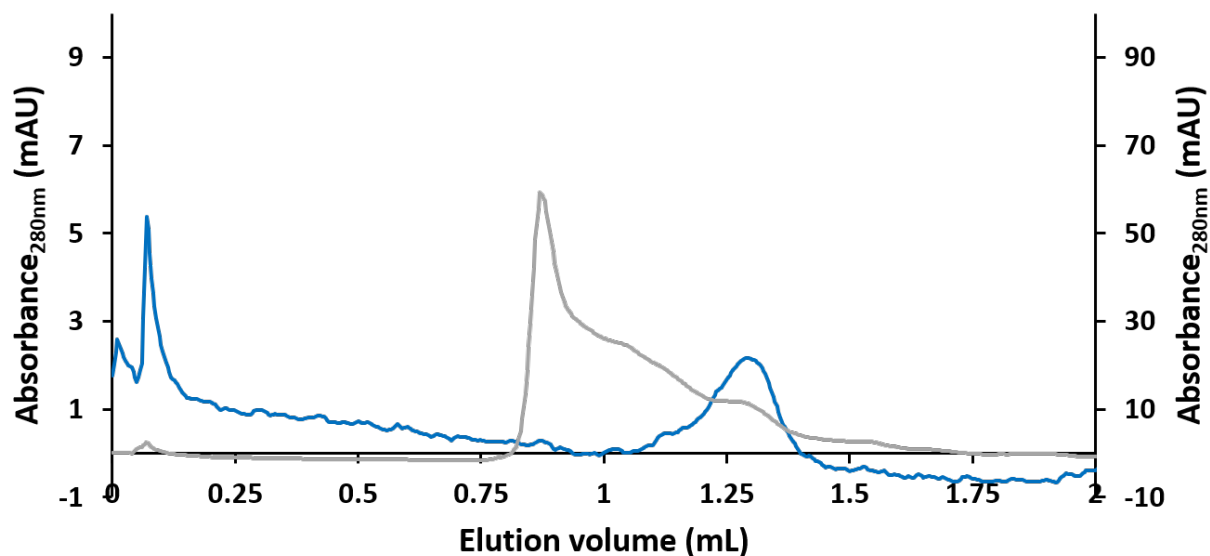
Figure A.9: Refocused analytical SEC of Siglec-11$_{\textbf{large V4}}$ expressed in Expi293F GnTI- cells. Dialysed protein from the IMAC was concentrated to 30 µL and injected into an analytical SD200 3.2/300 increase column (grey line). The peak at 1.3 mL was pooled (1.25-1-4 mL), concentrated and reinjected into the column. A single peak eluted at 1.3 mL (blue line), which corresponds to a molecular weight of approx. 120 kDa, based on the calibration line (figure A.2).
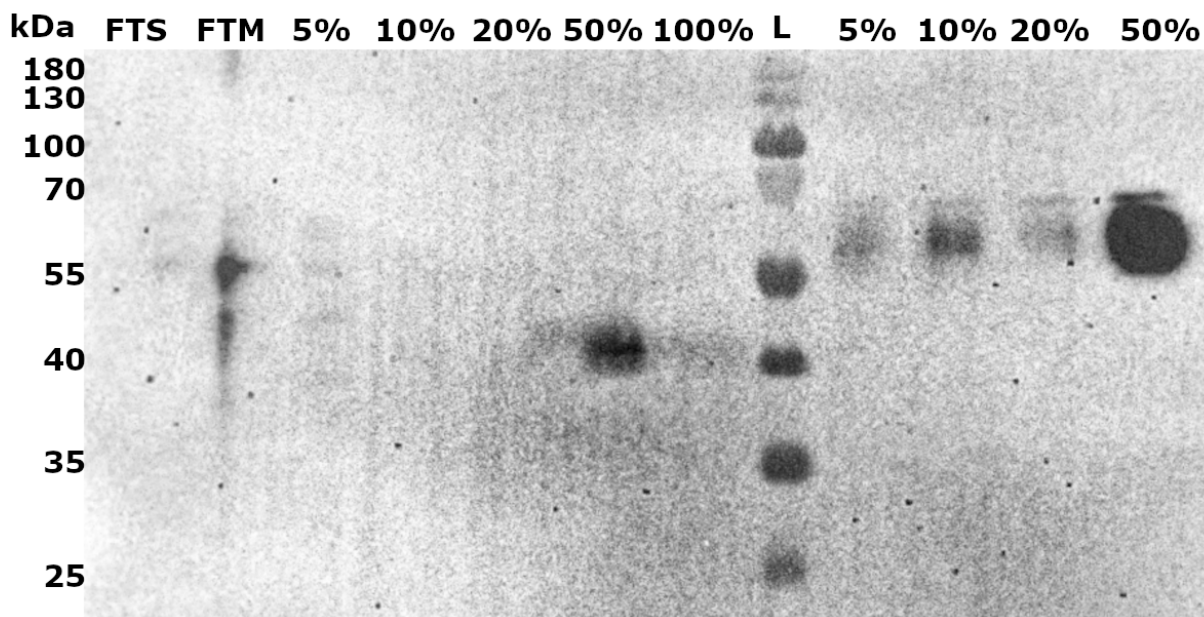
Figure A.10: IMAC of the Siglec-11middle V1 and Siglec-11short V1 expressed in Freestyle293F cells. The western blot of the Ni^{2+}-NTA affinity chromatography is represented with washing steps of 28 mM (5%), 55 mM (10%) and 105 mM imidazole (20%) to remove unspecific-bound protein, before the target protein was eluted with 250 mM imidazole (50%) and 500 mM imidazole (100%). Target protein was identified in the western blot with primary mouse α-his antibody and secondary HRP-associated goat α-mouse antibody. The protein size is indicated by an unstained protein ladder (L) in the western blot.
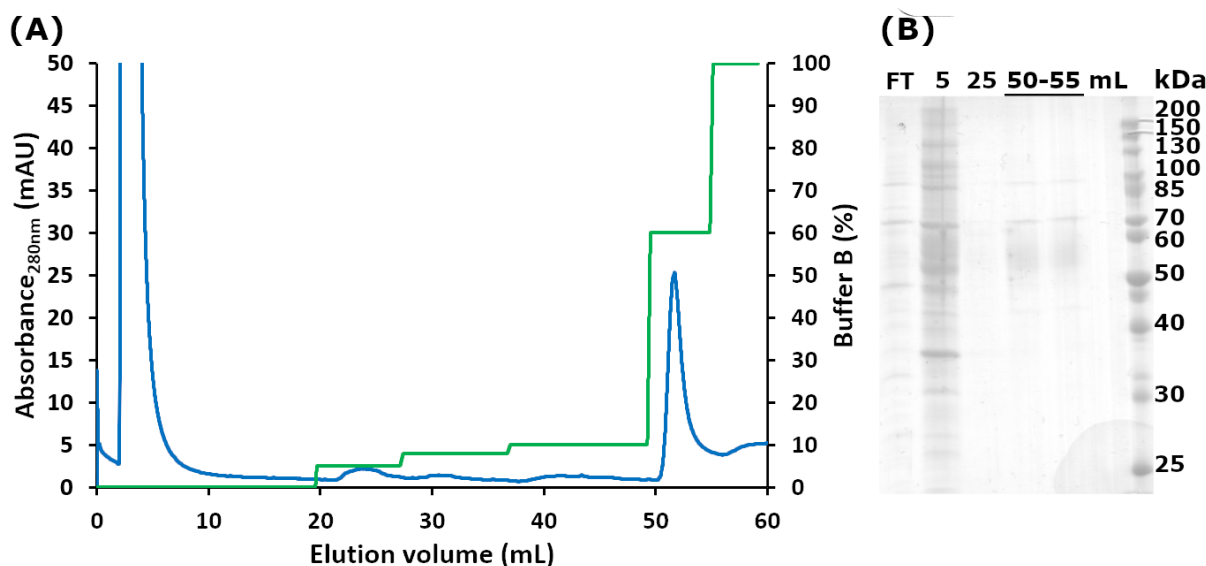
Figure A.11: IMAC of Siglec-11middle V2 expressed in Freestyle293F cells.(A) The Chromatogram, and **(B)** the corresponding SDS-PAGE are shown. In the chromatogram, UV absorption at 280 nm and the Buffer B concentration are drawn as solid blue and green line, respectively. The column was first washed with HisA1-buffer, followed by three washing steps with 30 mM (5% HisB1-buffer), 55 mM (10% HisB1-buffer) and 80 mM (15% HisB1-buffer) imidazole to remove unspecific-bound protein. Target protein was eluted with 305 mM imidazole (60% HisB1-buffer). Peak fractions were analysed by SDS-PAGE. For evaluation of the protein size, unstained protein ladder (L) and prestained protein ladder (L) was added to the SDS-PAGE and western blot, respectively.
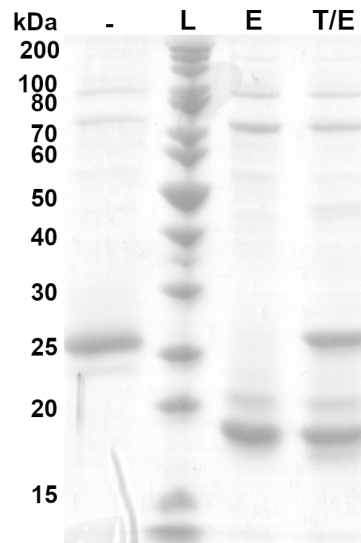
Figure A.12: SDS-PAGE of Siglec-11$_{1D\Delta myc}$ treated with EndoH*f* and TEV protease. Samples of untreated (-) and with EndoH$_f$ deglycosylated (E) Siglec-11$_{1D\Delta myc}$ were compared in the SDS-PAGE to a sample treated with Siglec-11$_{1D\Delta myc}$ and TEV protease (E/T). For evaluation of the protein size, unstained protein ladder (L) was added to the SDS-PAGE.
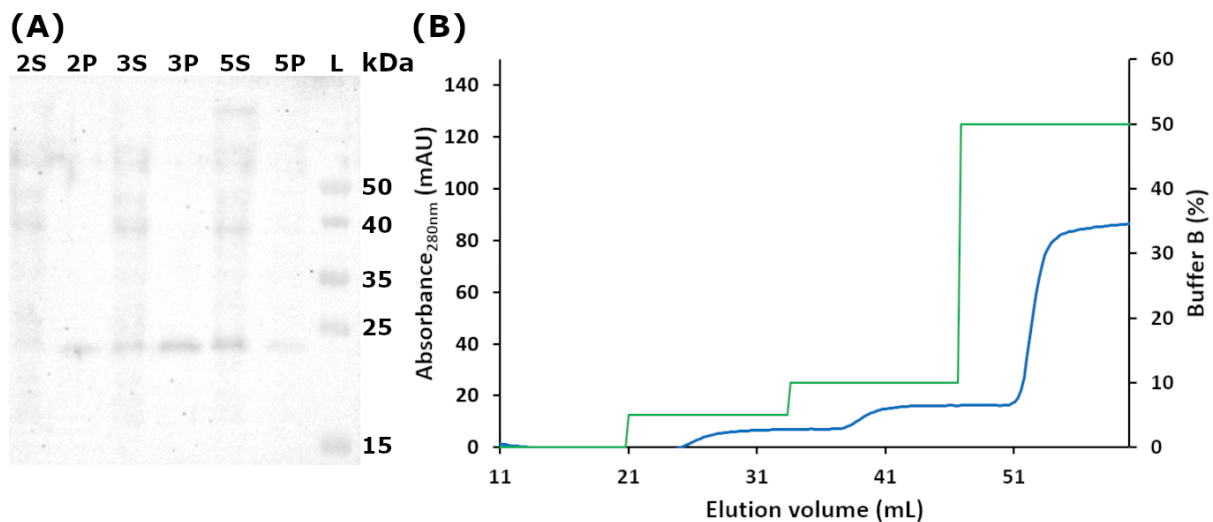
Figure A.13: **Siglec-11$_{1D\_TEV}$ expression in Expi293F GnTI- cells and IMAC .(A)** Siglec$_{1D\_TEV}$ expression level in medium and pellet were determined after two, three and five days. Protein amount in the medium (2S, 3S, 5S,) and cell pellet (2P, 3P, 5P) after centrifugation was determined by western blot. Target protein was identified with primary mouse α-his antibody and secondary HRP-associated goat α-mouse antibody and the protein size is indicated by an unstained protein ladder (L). **(B)** The Chromatogram of the IMAC is shown and the UV absorption at 280 nm and the Buffer B concentration are drawn as a solid blue and green line, respectively. The column was first washed with HisA2-buffer, followed by two washing steps with 30 mM (5% HisB2-buffer), 55 mM (10% HisB2-buffer) imidazole to remove unspecific-bound protein. No target protein eluted with 255 mM imidazole (50% HisB2-buffer).
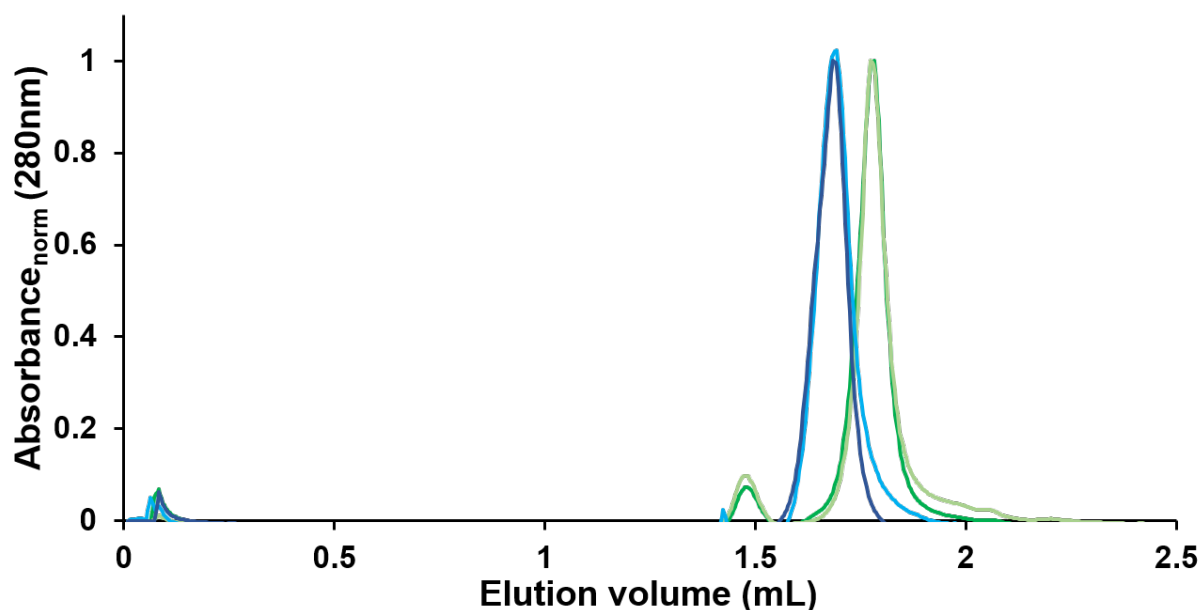
Figure A.14: Analytical SEC of glycosylated and deglycoslyated Siglec-11$_{1D\Delta myc}$ immediately after purification and after two weeks stored at 4°C. For each run, the UV absorbance was normalised to the peak maximum for better comparison. The chromatogram of glycosylated and deglycoslyated Siglec-11$_{1D\Delta myc}$ are shown in light blue and light green, respectively, the chromatograms for the stored samples are shown in dark blue and dark green, respectively. The small peak at 1.45 mL corresponds to the EndoH$_f$ used to deglycosylate Siglec-11$_{1D}$. Apparent molecular weight of the proteins is based on the elution volume and was determined according to the calibration of the column (figure A.2).
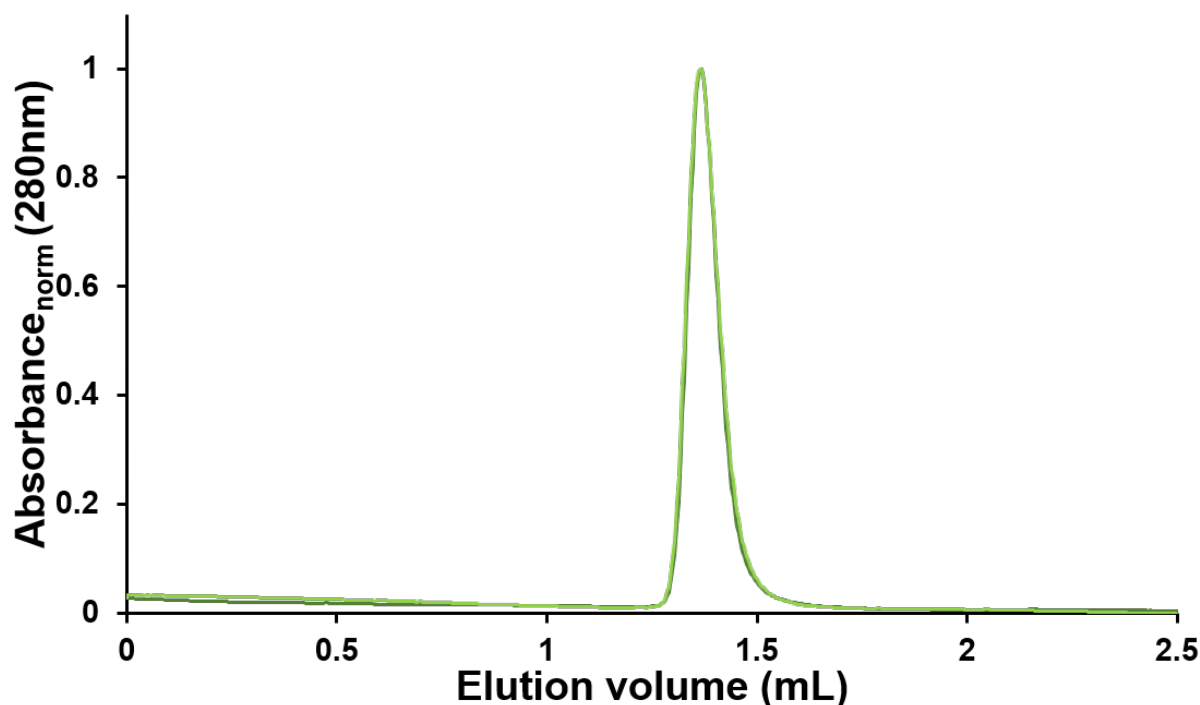

Figure A.15: Analytical SEC deglycoslyated Siglec-11$_{1D\Delta tag\Delta glyc}$ immediately after purification and after ten days stored at 4°C. For each run, the UV absorbance was normalised to the peak maximum for better comparison. The chromatogram of freshly purified and stored samples are shown in light and dark green, respectively. Apparent molecular weight of the proteins is based on the elution volume and was determined according to the calibration of the column (figure A.2).
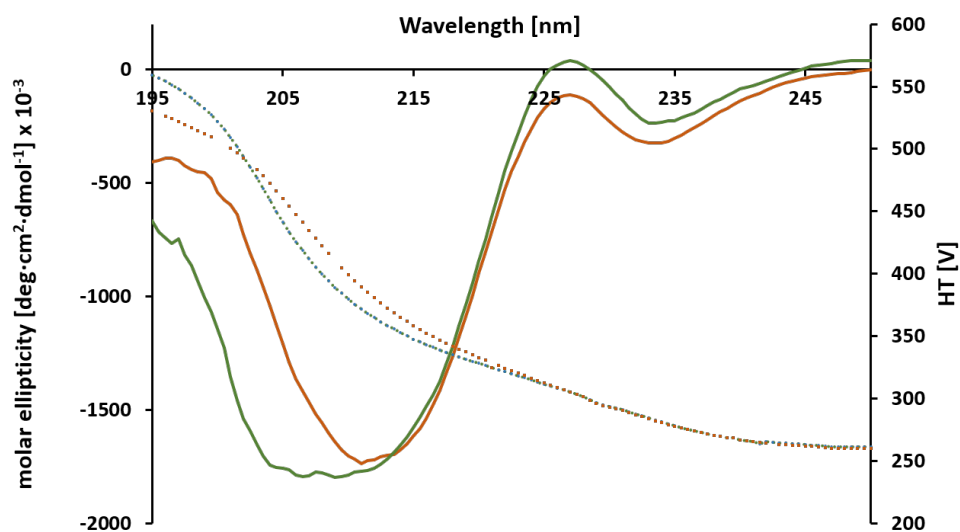
Figure A.16: CD Spectrum of Siglec-11$_{1D\Delta myc}$ and Siglec-11 $_{1D\Delta tag\Delta glyc}$. The ellipticity of Siglec-11 $_{1D\Delta myc}$ and Siglec-11 $_{1D\Delta tag\Delta glyc}$ are represented as a solid line in orange and green, respectively. The voltage of the photomultiplier is shown as a dashed line in the corresponding colour. Values above 800 V for the photomultiplier are not trustworthy.
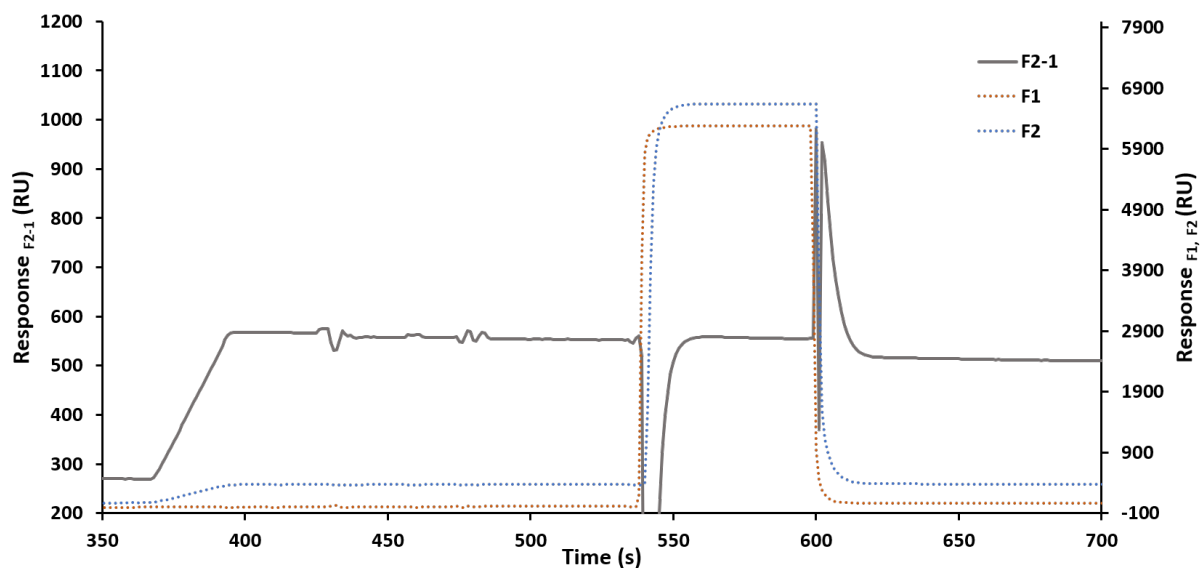

Figure A.17: SPR measurement of Siglec-11$_{1D\Delta myc}$ and colominic acid. 330 RU Siglec-11$_{1D\Delta myc}$ were immobilised as ligand on an Ni^{2+}-NTA chip and 2 mM colominic acid were applied for 60 s (time: 540-600 s). The sensogram of the reference (F1) and measuring (F2) cell are drawn as dashed line in blue and orange, respectively. The reference cell (F2-1) corrected SPR signal is shown as solid grey line.
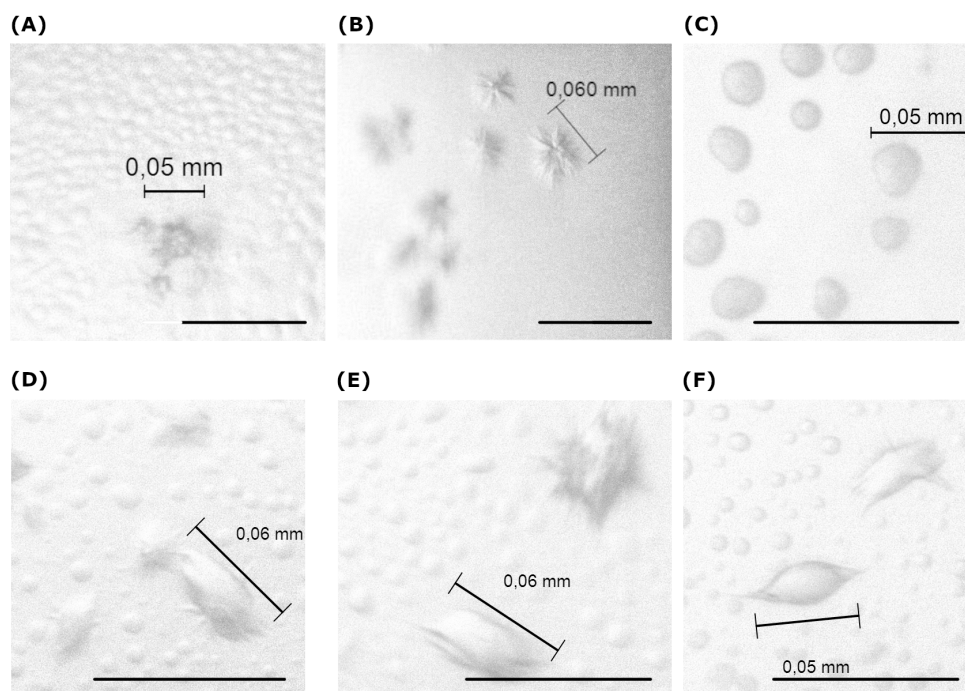
Figure A.18: Siglec-11$_{1D\Delta myc\Delta glyc}$ crystals. Images of the most promising conditions of commercial available screens and fine screens thereof. In commercial available screens, nucleation is visible in **(A)** condition 45 of JCSG+ with 0.17 M ammonium sulphate and 25%(w/v) PEG 4000; **(B)** condition 76 of crystal screen 1+2 with 1.6 M sodium citrate tribasic at pH 6.5 and **(C)** condition 9 of crystal screen 1+2 with 0.2 M ammonium acetate, 0.1 M sodium citrate pH 6.5 and 30%(w/v) PEG 4000. Only crystals appeared in the fine screen of crystal screen 1+2 condition 9 after 4 weeks. The most promising crystals were found in **(D)** condition 3 with 0.2 M ammonium acetate, 0.1 M sodium citrate tribasic pH 5.6, 24%(w/v) PEG 4000; **(E)** condition 10 with 0.15 M ammonium acetate, 0.1 M sodium citrate tribasic pH 5.4, 26%(w/v) PEG 4000 and **(F)** condition 51 with 0.2 M ammonium acetate, 0.1 M sodium citrate tribasic pH 5.6, 24%(w/v). In each image a scale bar measuring 100 μm is shown for reference.
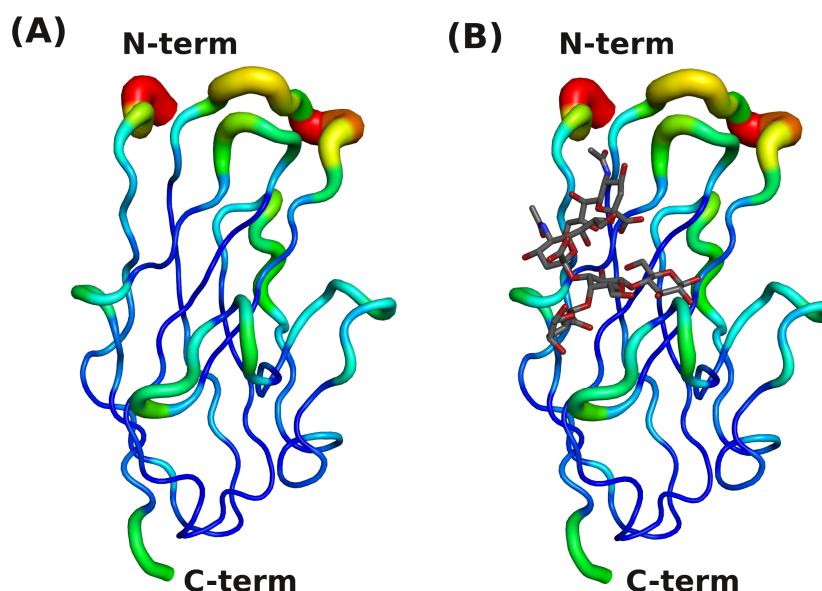
Figure A.19: Cα-RMSD of V-set domain of CD33-related Siglec. **(A)** RMSD between the Cα of Siglec-11$_{V\text{-set}}$ and structurally investigated CD33-related Siglecs are shown (table A.4). Small rmsd ($<$ 0.5 Å) are represented as blue cartoon and a thin diameter, large rmsd ($>$3.0 Å) are represented as red cartoon and a thick diameter. **(B)** The Siglec-11$_{V\text{-set}}$ structure with the modelled GT1b of Siglec-7 (pdb code: 2hrl).
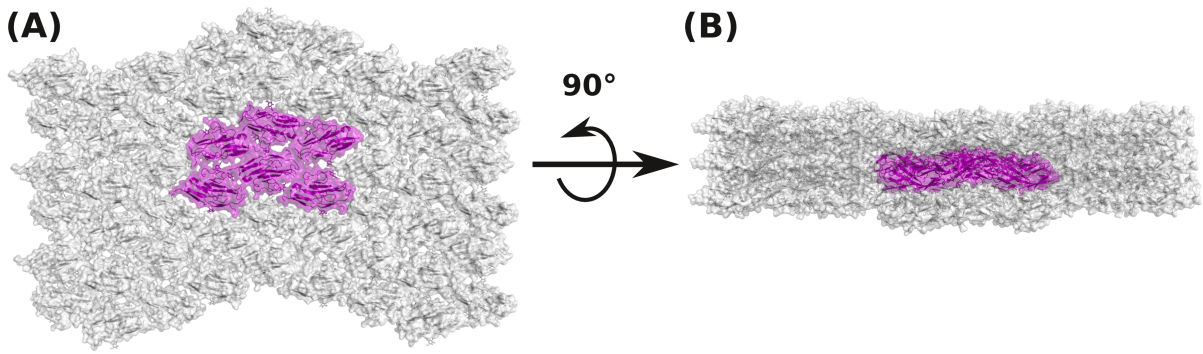
Figure A.20: Protein packing in the crystal. **(A)** Protomers of the ASU are shown in cartoon representation with a transparent surface in pink. The unit cell was extended (grey) to depict the dense packing of the crystal. **(B)** The crystal packing is additionally represented a 90 ∘ turn to illustrate all dimensions of the crystal.

```
sp|Q96RL6|Siglec-11    MVPGQAQPQSPEMLLLPLLLPVLGAGSL-----NKDPSYSLQVQRQVPVPEGLCVIVSCN    55
sp|A6NMB1|Siglec-16    ------------MLLLPLLLPVLGAGSL-----NKDPSYSLQVQRQVPVPEGLCVIVSCN    43
sp|Q08ET2|Siglec-14    --------------MLPLLLLPLLWGGSL-----QEKPVYELQVQKSVTVQEGLCVLVPCS    42
sp|Q96PQ1|Siglec-12    -----------MLLLLLLLPPLLCGRVGAK----EQKDYLLTMQKSVTVQEGLCVSVLCS    45
sp|Q96LC7|Siglec-10    --------------MLLPLLLSSLLGGSQ-----AMDGRFWIRVQESVMVPEGLCISVPCS    42
sp|Q9Y336|Siglec-9     --------------MLLLLLPLLWGRERAEGQTSK----LLTMQSSVTVQEGLCVHVPCS    42
sp|Q9NYZ4|Siglec-8     -----------MLLLLLLLLPLLWGTKGMEGDRQYGDGYLLQVQELVTVQEGLCVHVPCS    48
sp|Q9Y286|Siglec-7     -------------MLLLLLLLPLLWGRERVEGQKSNRKDYSLTMQSSVTVQEGMCVHVRCS    47
sp|O43699|Siglec-6     ---MQGAQEASASEMLPLLLPLLWAGALA-----QERRFQLEGPESLTVQEGLCVLVPCR    52
sp|O15389|Siglec-5     --------------MLPLLLLPLLWGGSL-----QEKPVYELQVQKSVTVQEGLCVLVPCS    42
sp|P20138|Siglec-3     -------------MPLLLLLPLLWAGAL-----AMDPNFWLQVQESVTVQEGLCVLVPCT    42
                                    **   *   .                  :        : * **:*: * *

sp|Q96RL6|Siglec-11    LSYPRDGWDEST-AAYGYWFKGRTSPKTGAPVATNNQSREVEMSTRDRFQLTGDPGKGSC   114
sp|A6NMB1|Siglec-16    LSYPRDGWDEST-AAYGYWFKGRTSPKTGAPVATNNQSREVAMSTRDRFQLTGDPGKGSC   102
sp|Q08ET2|Siglec-14    FSYPWRSWYS-SPPLYVYWFRDGEIPYYAEVVATNNPDRRVKPETQGRFRLLGDVQKKNC   101
sp|Q96PQ1|Siglec-12    FSYPQNGWTASD-PVHGYWFRAGDHVSRNIPVATNNPARAVQEETRDRFHLLGDPQNKDC   104
sp|Q96LC7|Siglec-10    FSYPRQDWTGST-PAYGYWFKAVTETTKGAPVATNHQSREVEMSTRGRFQLTGDPAKGNC   101
sp|Q9Y336|Siglec-9     FSYPSHGWIYPGPVVHGYWFREGANTDQDAPVATNNPARAVWEETRDRFHLLGDPHTKNC   102
sp|Q9NYZ4|Siglec-8     FSYPQDGWTDSD-PVHGYWFRAGDRPYQDAPVATNNPDREVQAETQGRFQLLGDIWSNDC   107
sp|Q9Y286|Siglec-7     FSYPVDSQTDSD-PVHGYWFRAGNDISWKAPVATNNPAWAVQEETRDRFHLLGDPQTKNC   106
sp|O43699|Siglec-6     LPTTLPASY----YGYGYWFLEGA----DVPVATNDPDEEVQEETRGRFHLLWDPRRKNC   104
sp|O15389|Siglec-5     FSYPWRSWYS-SPPLYVYWFRDGEIPYYAEVVATNNPDRRVKPETQGRFRLLGDVQKKNC   101
sp|P20138|Siglec-3     FFHPIPYYD-KNSPVHGYWFREGAIISRDSPVATNKLDQEVQEETQGRFRLLGDPSRNNC   101
                       :           :  ***        ****.      *   .*:.**:*   *      .*

sp|Q96RL6|Siglec-11    SLVIRDAQREDEAWYFFRVERGSRVRHSFL------SNAFFLKVTALTK-   157
sp|A6NMB1|Siglec-16    SLVIRDAQREDEAWYFFRVERGSRVRHSFL------SNAFFLKVTALTQK   146
sp|Q08ET2|Siglec-14    SLSIGDARMEDTGSYFFRVERGRDVKYSYQ------QNKLNLEVTALIEK   145
sp|Q96PQ1|Siglec-12    TLSIRDTRESDAGTYVFCVERGN-MKWNYK------YDQLSVNVTASQD-   146
sp|Q96LC7|Siglec-10    SLVIRDAQMQDESQYFFRVERGSYVRYNFM------NDGFFLKVTALTQK   145
sp|Q9Y336|Siglec-9     TLSIRDARRSDAGRYFFRMEKGS-IKWNYK------HHRLSVNVTALTHR   145
sp|Q9NYZ4|Siglec-8     SLSIRDARKRDKGSYFFRLERGS-MKWSYKSQLNYKTKQLSVFVTALTHR   156
sp|Q9Y286|Siglec-7     TLSIRDARMSDAGRYFFRMEKGN-IKWNYK------YDQLSVNVTALTHR   149
sp|O43699|Siglec-6     SLSIRDARRRDNAAYFFRLKSKW-MKYGYT------SSKLSVRVMALTHR   147
sp|O15389|Siglec-5     SLSIGDARMEDTGSYFFRVERGRDVKYSYQ------QNKLNLEVTALIEK   145
sp|P20138|Siglec-3     SLSIVDARRRDNGSYFFRMERGS-TKYSYK------SPQLSVHVTDLTH-   143
                       :* * *::  *  .*.* ::       :  :.          : : *    .
```

Figure A.21: Sequence Alignment of CD33-related Siglec V-sets. The sequence of the V-sets was obtained from the uniprot entry and aligned with ClustalOmega [318, 292]. Strictly conserved arginine and the conserved aromatic amino acid for sialic acid binding are highlighted in light and dark red, respectively, and cysteines are shown in orange. Glycosylated asparagines, observed in the corresponding structure, are represented in dark blue. For Siglec-16, highlighted asparagines are expected to be glycosylated based on the conserved sequence to Siglec-11.
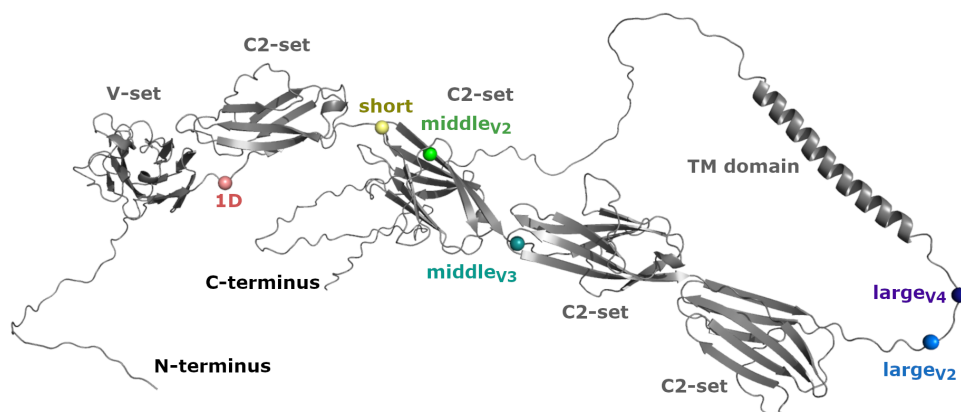
Figure A.22: Siglec-11 full-length structure predicted by Alpha-Fold 2.0, including the transmembrane and cytosolic ITIM domains. The domain boarder of each construct is indicated in the predicted structure of Siglec-11 (published online: 1.7.2021) [286]. The construct ends between two domains for Siglec-11$_{1D}$ (red), Siglec-11$_{short}$ (yellow), Siglec-11$_{middle\,V3}$ (dark green), Siglec-11$_{large\,V2}$ (dark blue) and Siglec-11$_{large\,V4}$ (light blue). Only Siglec-11$_{middle\,V2}$ (light green) seems to end within the last β-sheet, even though this construct showed similar expression level as Siglec-11$_{middle\,V3}$ in Freestyle293F cells.

A.2 List of Publications

$^+$ Authors contributed equally to this work.

Publications Discussed in this Thesis

Teresa R Wagner$^+$, **Elena Ostertag**$^+$, Philipp D Kaiser, Marius Gramlich, Natalia Ruetalo, Daniel Junker, Julia Haering, Bjoern Traekle, Matthias Becker, Alex Dulovic, Helen Schweizer, Stefan Nueske, Armin Scholz, Anne Zeck, Jatja Schenke-Layland, Annika Nelde, Monika Strengert, Juliane S Walz, Georg Zocher, Thilo Stehle, Michael Schindler, Nicole Schneiderhan-Marra and Ulrich Rothbauer. (2021) *NeurobodyPlex - monitoring SARS-CoV-2 neutralizing immune responses using nanobodies.* EMBO reports 22(5): e52325. doi:10.15252/embr.202052325. Published online: 27 April 2021

Elena Ostertag, Liujuan Zheng, Karina Broger, Thilo Stehle, Shu-Ming Li and Georg Zocher. (2021) *Reprogramming substrate and catalytic promiscuity of tryptophan prenyl-transferases.* JMB, 433(2): 166726. doi:10.1016/j.jmb.2020.11.025.
Published online: 26 November 2020 ©Elsevier Ldt.

Publications Not Discussed in this Thesis

Katja Mindler$^+$, **Elena Ostertag**$^+$ and Thilo Stehle. (2021) *The polyfunctional polysialic acid: a structural review.* Carbohydrate Research, 507:108376.
doi:10.1016/j.carres.2021.108376. Published online: 22 June 2021. ©Elsevier Ltd.

Erik Nöldeke, Lena M. Muckenfuss, Volker Niemann, Anna Müller, **Elena Störk**, Georg Zocher, Tanja Schneider and Thilo Stehle. (2018) *Structural basis of cell wall peptidoglycan amidation by the GatD/MurT complex of Staphylococcus aureus.* Scientific Reports, 8(1): 1-15. doi:10.1038/s41598-018-31098-x. Published online: 28 August 2018.

Article

EMBO *reports*

NeutrobodyPlex—monitoring SARS-CoV-2 neutralizing immune responses using nanobodies

Teresa R Wagner[1,2,†] (ID), Elena Ostertag[3,†], Philipp D Kaiser[2], Marius Gramlich[2], Natalia Ruetalo[4], Daniel Junker[2], Julia Haering[2], Bjoern Traenkle[2], Matthias Becker[2], Alex Dulovic[2], Helen Schweizer[5] (ID), Stefan Nueske[5], Armin Scholz[5], Anne Zeck[2], Katja Schenke-Layland[2,6,7,8], Annika Nelde[6,9,10], Monika Strengert[11,12], Juliane S Walz[6,9,10,13], Georg Zocher[3], Thilo Stehle[3,14], Michael Schindler[4], Nicole Schneiderhan-Marra[2] & Ulrich Rothbauer[1,2,6,*] (ID)

Abstract

In light of the COVID-19 pandemic, there is an ongoing need for diagnostic tools to monitor the immune status of large patient cohorts and the effectiveness of vaccination campaigns. Here, we present 11 unique nanobodies (Nbs) specific for the SARS-CoV-2 spike receptor-binding domain (RBD), of which 8 Nbs potently inhibit the interaction of RBD with angiotensin-converting enzyme 2 (ACE2) as the major viral docking site. Following detailed epitope mapping and structural analysis, we select two inhibitory Nbs, one of which binds an epitope inside and one of which binds an epitope outside the RBD:ACE2 interface. Based on these, we generate a biparatopic nanobody (bipNb) with viral neutralization efficacy in the picomolar range. Using bipNb as a surrogate, we establish a competitive multiplex binding assay ("NeutrobodyPlex") for detailed analysis of the presence and performance of neutralizing RBD-binding antibodies in serum of convalescent or vaccinated patients. We demonstrate that NeutrobodyPlex enables high-throughput screening and detailed analysis of neutralizing immune responses in infected or vaccinated individuals, to monitor immune status or to guide vaccine design.

Introduction

As of March 2021, the COVID-19 pandemic has led to the deaths of more than 2.8 million people worldwide and continues to cause severe lockdowns and dramatic economic losses. The emergence and spread of new mutants pose additional risk to current vaccination campaigns (Wang *et al*, 2021). The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as causative agent of the disease expresses a surface spike glycoprotein (spike), which consists of two subunits S1 and S2 (Wrapp *et al*, 2020b; Zhu *et al*, 2020). For viral infection, the receptor-binding domain (RBD), located within S1 interacts with the angiotensin-converting enzyme 2 (ACE2) expressed on human epithelial cells of the respiratory tract (Tai *et al*, 2020; Wrapp *et al*, 2020b; Yan *et al*, 2020). In-depth analysis revealed the presence of spike-specific neutralizing antibodies (NAbs) in convalescent individuals, which were shown to inhibit viral uptake by various mechanisms (Rogers *et al*, 2020; Tortorici *et al*, 2020). In this context, a constantly growing number of NAbs specifically targeting the RBD have been described, underlining the importance of blocking the RBD:ACE2 interaction site for the

1 Pharmaceutical Biotechnology, Eberhard Karls University, Tuebingen, Germany
2 Natural and Medical Sciences Institute, University of Tuebingen, Reutlingen, Germany
3 Interfaculty Institute of Biochemistry, Eberhard Karls University, Tuebingen, Germany
4 Institute for Medical Virology and Epidemiology of Viral Diseases, University Hospital Tuebingen, Tuebingen, Germany
5 Livestock Center of the Faculty of Veterinary Medicine, Ludwig Maximilians University, Oberschleissheim, Germany
6 Cluster of Excellence iFIT (EXC2180) "Image-Guided and Functionally Instructed Tumor Therapies", Eberhard Karls University, Tuebingen, Germany
7 Department of Women's Health, Research Institute for Women's Health, Eberhard Karls University, Tuebingen, Germany
8 Department of Medicine/Cardiology, Cardiovascular Research Laboratories, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA
9 Clinical Collaboration Unit Translational Immunology, German Cancer Consortium (DKTK), Department of Internal Medicine, University Hospital Tuebingen, Tuebingen, Germany
10 Institute for Cell Biology, Department of Immunology, Eberhard Karls University, Tuebingen, Germany
11 Department of Epidemiology, Helmholtz Centre for Infection Research, Braunschweig, Germany
12 TWINCORE GmbH, Centre for Experimental and Clinical Infection Research, A Joint venture of the Hannover Medical School and the Helmholtz Centre for Infection Research, Hannover, Germany
13 Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology and Robert Bosch Center for Tumor Disease, RBCT, Stuttgart, Germany
14 Vanderbilt University School of Medicine, Nashville, TN, USA
*Corresponding author. Tel: +49 7121 51530-415; Fax: +49 7121 51530-816; E-mail: ulrich.rothbauer@uni-tuebingen.de
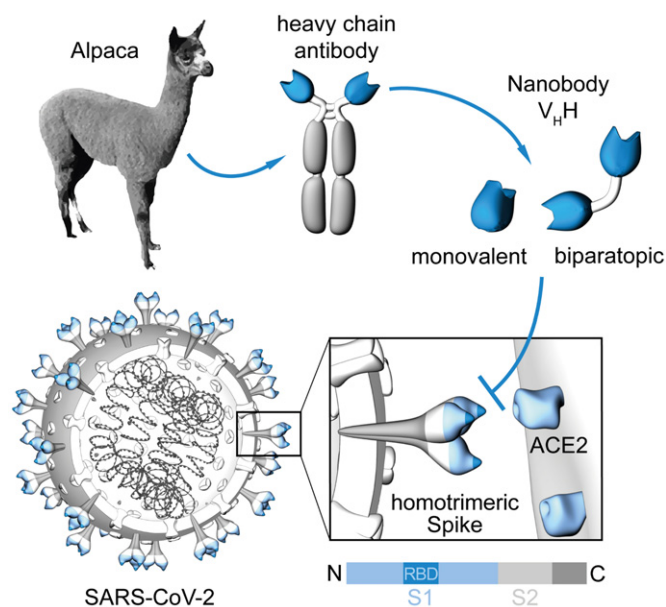†These authors contributed equally to this work

development of a protective immune response (Brouwer *et al*, 2020; Cao *et al*, 2020; Ju *et al*, 2020; Robbiani *et al*, 2020; Shi *et al*, 2020; Tai *et al*, 2020; Yu *et al*, 2020). To date, numerous NAbs are in preclinical or clinical development for prophylactic and therapeutic options, providing immediate protection against SARS-CoV-2 infection (reviewed in Jiang *et al*, 2020; Zohar & Alter, 2020).

Promising alternatives to conventional antibodies (IgGs) are single-domain antibodies (nanobodies, Nbs) derived from the heavy-chain antibodies of camelids (Fig 1) (Muyldermans, 2013). Due to their small size and compact fold, Nbs show high chemical stability, solubility, and fast tissue penetration. Nbs can be efficiently selected against different epitopes on the same antigen and converted into multivalent formats (Muyldermans, 2013). The potential of Nbs to address SARS-CoV-2 has been impressively demonstrated by the recent identification of several RBD-specific Nbs from naïve/ synthetic libraries (Chi *et al*, 2020a; Custodio *et al*, 2020; Huo *et al*, 2020; preprint: Schoof *et al*, 2020; preprint: Walter *et al*, 2020; preprint: Ahmad *et al*, 2021) or immunized animals (Chi *et al*, 2020a; preprint: Esparza *et al*, 2020; preprint: Gai *et al*, 2020; Hanke *et al*, 2020; preprint: Nieto *et al*, 2020; Wrapp *et al*, 2020a; Xiang *et al*, 2020; Koenig *et al*, 2021). Some of the identified Nbs show a high viral neutralizing potency proposed by blocking the RBD:ACE2 interface (Custodio *et al*, 2020; Hanke *et al*, 2020; Koenig *et al*, 2021), activation of the SARS-CoV-2 fusion machinery (Koenig *et al*, 2021), or induction of an inactive spike conformation (preprint: Schoof *et al*, 2020).

Since the pandemic outbreak, multiple serological SARS-CoV-2 assays have been established to monitor seroconversion in individuals and estimate the level of endemic infection in the general population. However, most available serological tests measure the full immune response and can therefore not differentiate between total binding and neutralizing antibodies (Amanat *et al*, 2020; preprint: Lassaunière *et al*, 2020; Robbiani *et al*, 2020; preprint: Roxhed *et al*, 2020; Stadlbauer *et al*, 2020; Tang *et al*, 2020; Becker *et al*, 2021). Detection of the latter is still mostly performed by conventional virus neutralization tests (VNTs), which are both time consuming (2–4 days) and require work with infectious SARS-CoV-2 virions in a specialized biosafety level 3 (BSL3) facility (Muruato *et al*, 2020; Scholer *et al*, 2020).

To overcome these limitations, we aimed to employ Nbs as antibody surrogates and developed a competitive binding approach to screen for neutralizing antibodies on a high-throughput basis in samples from patients or vaccinated individuals. Here, we describe the selection of 11 unique Nbs derived from an alpaca immunized with glycosylated SARS-CoV-2 RBD. Employing a multiplex *in vitro* binding assay, we identified 8 Nbs that effectively block the interaction between RBD, S1, and homotrimeric spike protein (spike) with ACE2 and neutralize SARS-CoV-2 infection in a human cell line. Based on a detailed epitope mapping and structural analysis of RBD: Nb complexes, we selected two of the most potent Nbs simultaneously targeting different epitopes within the RBD and generated a biparatopic Nb (bipNb). The bipNb represents a potent antibody surrogate with IC_{50} values in the low picomolar range and exhibits substantially improved binding affinities. Notably, by addressing at least one conserved epitope outside the RBD:ACE2 interface, the bipNb is still capable to bind recently described RBD mutants derived from strains B.1.1.7 (UK) and B.1.351 (South Africa). To monitor the presence and performance of neutralizing antibodies addressing the RBD:ACE2 interface in convalescent patient samples, we implemented the bipNb in a competitive multiplex binding assay, termed NeutrobodyPlex. Based on the data presented, the NeutrobodyPlex provides a versatile high-throughput approach to screen for a neutralizing immune response in infected or vaccinated individuals, helping to monitor immune status of large populations, to determine the success of vaccination campaigns and to guide vaccine design.

Results

Selection of SARS-CoV-2-specific Nbs

To generate Nbs directed against the RBD of SARS-CoV-2, we immunized an alpaca (*Vicugna pacos*) with purified RBD (Amanat *et al*, 2020) and established a Nb phagemid library comprising $\sim 4 \times 10^7$ clones representing the full repertoire of variable heavy chains of heavy-chain antibodies (V_HHs or Nbs). After two rounds of phage display on passively adsorbed or biotinylated RBD immobilized on streptavidin plates, we analyzed 492 individual clones in a solid-phase phage ELISA and identified 325 positive binders. Sequence analysis of 72 clones revealed 11 unique Nbs, which cluster into eight families with highly diverse complementarity determining regions (CDR) 3 (Fig 2A, Appendix Table S1). Individual Nbs were produced and purified from *Escherichia coli* (*E. coli*) (Fig 2B), and affinity measurements revealed K_D values ranging from ~ 1.4 to ~ 53 nM indicating the selection of 10 high-affinity monovalent binders. NM1225, that displayed a binding affinity in the micromolar range, was not

Figure 1. Generation of nanobodies blocking the SARS-CoV-2 RBD:ACE2 interface.

Nanobodies (Nbs) are genetically engineered from heavy-chain only antibodies of alpacas. The interaction between the SARS-CoV-2 homotrimeric spike protein and angiotensin-converting enzyme (ACE) 2 can be blocked by receptor-binding domain (RBD)-specific Nbs in the monovalent or biparatopic format.

considered for further analysis (Fig 2C, Appendix Fig S1). Next, we analyzed whether selected Nbs can block the interaction between RBD, S1, or spike of SARS-CoV-2 to human ACE2. To this end, we utilized a multiplex ACE2 competition assay employing the respective SARS-CoV-2 antigens coupled to paramagnetic beads (MagPlex Microspheres) (Becker *et al*, 2021) and incubated them with biotinylated ACE2 and dilutions of purified Nbs, before measuring residual binding of ACE2 via streptavidin-PE conjugate. As controls, we included a non-specific Nb (GFP-Nb, negative control) and two inhibiting mouse antibodies (Gorshkov *et al*, 2020) as positive controls. Data obtained through this multiplex binding assay showed that 8 out of 10 analyzed Nbs inhibit ACE2 binding to all tested SARS-CoV-2 antigens (Fig EV1). Notably, IC_{50} values obtained for the most potent inhibitory Nbs NM1228 (0.5 nM), NM1226 (0.82 nM) and NM1230 (2.12 nM) are highly comparable to IC_{50} values measured for the mouse IgGs (MM43: 0.38 nM; MM57: 3.22 nM).

Nanobodies show a high potency in neutralizing SARS-CoV-2

Next, a set of Nbs representing the full diversity according to the CDR3 region and highest affinities were examined for their potential to inhibit viral infection. For a viral neutralization test (VNT), human Caco-2 cells were co-incubated with the SARS-CoV-2-mNG infectious clone and serial dilutions of NM1223, NM1224, NM1226, NM1228, NM1230 or GFP-Nb as negative control. 48 h post-infection neutralization potency was determined via automated fluorescence-microscopy of fixed and nuclear-stained cells. The infection rate, normalized to a non-treated control was plotted and IC_{50} values were determined via sigmoidal inhibition curve fits. NM1226 and NM1228 showed a strong neutralization potency with IC_{50} values of ~ 15 and ~ 7 nM followed by NM1230 (~ 37 nM) and NM1224 (~ 256 nM) (Appendix Fig S2). As expected from our biochemical analysis, NM1223 was not found to neutralize SARS-CoV-2 (Appendix Fig S2). Overall, these findings are highly consistent with the results obtained from the multiplex ACE2 competition assay, thus demonstrating high potencies of ACE2-blocking Nbs to neutralize viral infection.

Epitope identification

To identify the relative location of Nb epitopes within the RBD, we firstly performed epitope binning experiments using biolayer inter-ferometry (BLI). For this, sensors, pre-coated with biotinylated RBD, were loaded with a first Nb followed by a short dissociation step and subsequent loading of a second Nb (Appendix Fig S3A). As expected, Nbs displaying similar CDR3 sequences (NM1221, NM1222, and NM1230, Nb-Set2) were unable to bind
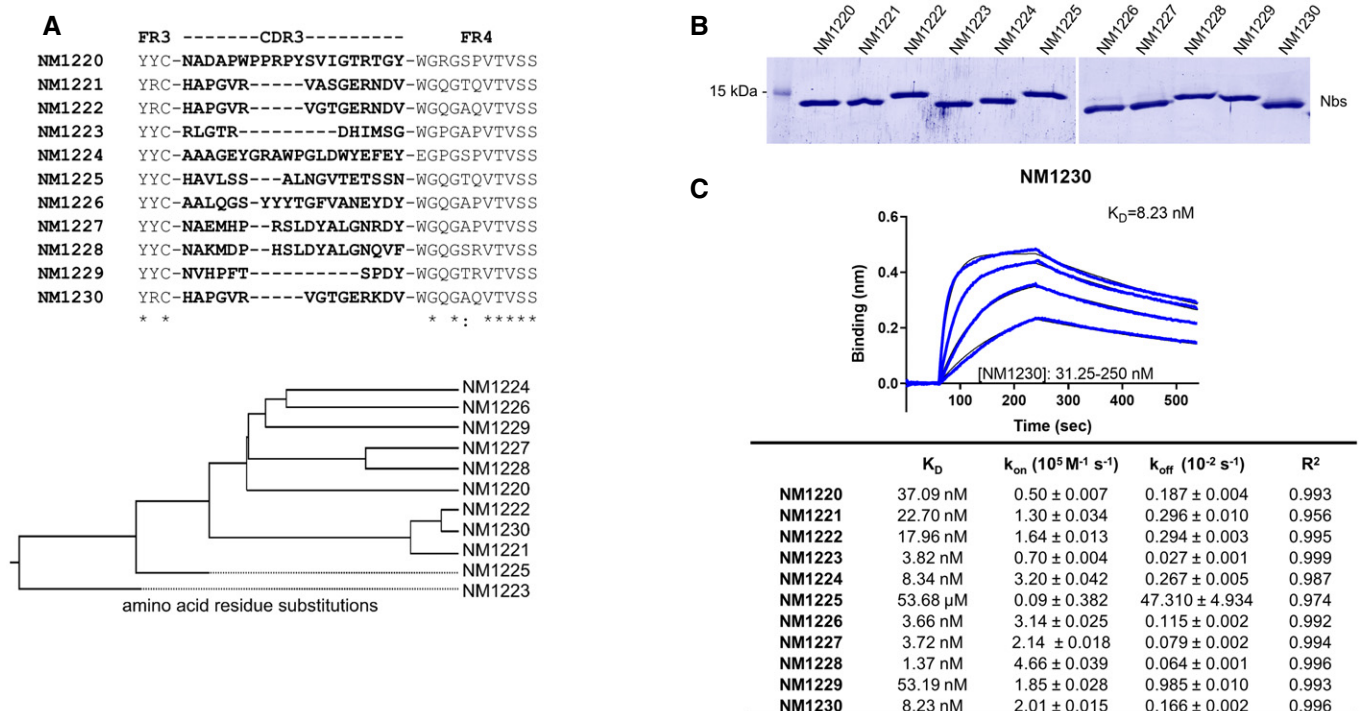
Figure 2. Biochemical characterization of RBD-specific Nbs.

A Amino acid sequences of the complementarity determining region (CDR) 3 from unique Nbs selected after two rounds of biopanning are listed (upper panel). Phylogenetic tree based on a ClustalW alignment of the Nb sequences is shown (lower panel).

B Recombinant expression and purification of Nbs using immobilized metal affinity chromatography (IMAC) and size exclusion chromatography (SEC). Coomassie staining of 2 μg of purified Nbs is shown.

C For biolayer interferometry (BLI)-based affinity measurements, biotinylated RBD was immobilized on streptavidin biosensors. Kinetic measurements were performed by using four concentrations of purified Nbs ranging from 15.6 nM to 2 μM. As an example, the sensogram of NM1230 at indicated concentrations is shown (upper panel). The table summarizes affinities (K_D), association (k_{on}), and dissociation constants (k_{off}) determined for individual Nbs (lower panel).

simultaneously as they recognize identical or highly similar epitopes. Interestingly, we noticed that Nbs with highly diverse CDR3s such as NM1228, NM1226, NM1227, and NM1229 could not bind simultaneously, suggesting that these Nbs also recognize similar or overlapping epitopes. Consequently, we clustered these diverse Nbs into Nb-Set1. In total, we identified five distinct Nbs-Sets, comprising at least one candidate targeting a different epitope within the RBD compared with any member of a different Nb-Set (Appendix Fig S3B).

For detailed epitope mapping, we performed Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS) using the most potent inhibitory Nbs selected from the different Nb-Sets and determined their binding regions within the RBD in relation to previously described ACE2 interaction sites (Fig EV2A, Appendix Fig S4) (Lan *et al*, 2020; Yan *et al*, 2020). Both members of Nb-Set1, NM1226 and NM1228, interacted with the RBD at the back/ lower right site (Back View, Fig EV2B and C). The highest exchange protection for NM1226 was found in the amino acid (aa) region $N_{RBD}370 - L_{RBD}387$ (Fig EV2B). This region is also covered by NM1228, which displayed additional binding to $Y_{RBD}489 - S_{RBD}514$, being part of the RBD:ACE2 interface (Fig EV2C). NM1230 (Nb-Set2) shows the highest protection in the region of $C_{RBD}432 - L_{RBD}452$ covering two amino acids involved in ACE2 binding ($G_{RBD}446$, $Y_{RBD}449$). A second protected region was found covering $N_{RBD}487 - G_{RBD}496$, which overlaps with the RBD:ACE2 interface (Fig EV2D). In accordance with our binning studies, the main epitope regions differ considerably from both Nb-Sets. As expected, NM1221 and NM1222 (both Nb-Set2) addressed similar RBD regions compared with NM1230 (Fig EV2E and F) while NM1224 (Nb-Set4) showed an interaction distinct from all other Nbs, covering both its main binding region located at the lower right side (Fig EV2G, Front View) and residues in the RBD:ACE2 interface (Fig EV2G, Front View, upper left corner). Notably, the non-inhibitory NM1223 (Nb-Set3) was shown not to contact residues involved in the RBD:ACE2 interface but rather binds to the opposite side (Fig EV2H, Front View).

Crystal structure of RBD:Nb complexes

To obtain a deeper insight into the binding mode of the most potent inhibitory Nbs, we analyzed the crystal structures of NM1226 and NM1230 in complex with RBD (Fig 3A). All Nb amino acid residues listed below follow the numbering according to Kabat (Kabat *et al*, 1991). The RBD:NM1226 complex structure was solved at a resolution of 2.3 Å (Appendix Table S2). NM1226 was found to interact with various RBD regions (aa 369–384 and aa 504–508) as well as $D_{RBD}405$ and $R_{RBD}408$. Major contacts of NM1226 to RBD are established by residues of the CDR3 ($L_{NM1226}95$, $Q_{NM1226}96$, $G_{NM1226}97$, $S_{NM1226}98$, $Y_{NM1226}99$, $Y_{NM1226}100$, $Y_{NM1226}100a$, $V_{NM1226}100e$, $N_{NM1226}100g$, $E_{NM1226}100h$, $Y_{NM1226}100i$, $D_{NM1226}101$) and CDR1 ($Y_{NM1226}30$). Beside a salt-bridge ($D_{NM1226}101$ to $R_{RBD}408$ and $K_{RBD}378$), various hydrogen bonds were found at contact sites (Fig 3B). In total, the interface of NM1226 buries an area of 698 Å^2. The RBD:NM1226 complex superimposes well onto spike protein (pdb code: 7KMS) with a Cα-rmsd derivation of 1.7 Å for all three RBD and shows that binding of NM1226 takes place exclusively in the "up" conformation of the spike protein (Fig EV3A and B) (Zhou *et al*, 2020). Although its epitope does not overlap with the binding site of ACE2, our structural data suggest that binding of NM1226

prohibits ACE2 recruitment by steric collision, which would explain its neutralizing effect (Fig EV3A and B).

For the RBD:NM1230 complex, we determined the structure at a resolution of 2.9 Å (Appendix Table S2). In accordance with our HDX-MS analysis, NM1230 binds a distinct three-dimensional epitope located at the opposite site of RBD (Fig 3A). Major contacts were formed by residues of the CDR3 ($P_{NM1230}95$, $R_{NM1230}98$, $T_{NM1230}100a$, $E_{NM1230}100c$, $R_{NM1230}100d$, $K_{NM1230}100e$, $D_{NM1230}101$, $V_{NM1230}102$, $W_{NM1230}103$). In contrast to NM1226, also residues located in the framework regions (FR) including FR1 ($Q_{NM1230}1$), FR2 ($N_{NM1230}35$, $Y_{NM1230}37$, $Q_{NM1230}39$, $G_{NM1230}42$, $K_{NM1230}43$, $A_{NM1230}44$, $L_{NM1230}45$ and $L_{NM1230}47$, and $A_{NM1230}49$), and FR3 ($E_{NM1230}60$) are involved in binding. Consequently, specific binding of NM1230 is achieved by a combination of the CDR3 and residues of the framework regions as previously observed for other Nbs (Kirchhofer *et al*, 2010). On the RBD site, NM1230 interacts with $Y_{RBD}351$ and the large loop region ranging from aa 437–503 (Fig 3C). The RBD:NM1230 interface is mainly formed by polar contacts and one salt-bridge ($R_{NM1230}98$ to $E_{RBD}484$), but also π-π stackings ($W_{NM1230}103$ to $F490_{RBD}$, $R_{NM1230}100d$ to $Y_{RBD}489$) are present (Fig 3C). In total, it buries a surface area of ~ 830 Å^2. The alignment (Cα-rmsd deviation of 1.2 Å for all three RBD) and comparison to a recently reported structure of the SARS-CoV-2 spike protein bound to neutralizing Nb-Ty1 (pdb code: 6ZXN) (Hanke *et al*, 2020) revealed that NM1230 is able to bind in "up" as well as in "down" conformation (Fig EV3A and B; Appendix Fig S5A).

Next, we compared the RBD:NM1230 structure with the recently reported RBD:ACE2 receptor complex (pdb code:6M17) (Yan *et al*, 2020) to structurally validate the neutralizing potential of NM1230. Closer inspection of the binding site (Appendix Fig S5B and C) showed that NM1230 partially overlaps with the ACE2-binding interface and that the neutralizing effect can be isolated to a limited set of residues within the RBD ($K_{RBD}417$, $Y_{RBD}449$, $F_{RBD}456$, $Y_{RBD}489$, $Q_{RBD}493$, $S_{RBD}494$) (Fig 3C, Appendix Fig S5C). However, NM1230 does not only block binding to ACE2 via its interaction with the RBD on the same protomer (Appendix Fig S5A, clash I), but also impairs ACE2 binding through steric collision on the neighboring RBD (Appendix Fig S5A, clash II). Based upon our structural data, we propose that NM1230-mediated blocking of two out of three RBDs would suffice to abolish ACE2 binding to a trimeric spike molecule.

Considering viral mutagenesis of SARS-CoV-2 within the RBD and associated occurrence of new super spreading strains (preprint: Davies *et al*, 2020; preprint: Tegally *et al*, 2020; Davies *et al*, 2021; Lauring & Hodcroft, 2021; preprint: Walker *et al*, 2021), we analyzed the impact of mutated RBD positions in the recently described SARS-CoV-2 strains B.1.1.7 (UK) and B.1.351 (South Africa) on binding to both neutralizing Nbs NM1226 and NM1230. From our structural data, we assumed that the common amino acid exchange on position 501 (N > Y) has no influence on Nb binding, whereas exchanges on position 417 (K > N) and 484 (E > K), both present in B.1.351, might only affect binding of NM1230 (Appendix Fig S6). In fact, when determining affinities of both Nbs for the RBD mutants derived from B.1.1.7 or B.1.351 by BLI, we obtained similar affinities compared with RBD$_{wt}$ for NM1226 (K_D = ~ 5.4 nM for RBD$_{B.1.1.7}$; K_D = ~ 5.5 nM for RBD$_{B.1.351}$) (Appendix Fig S7A). Interestingly, while affinity of NM1230 to RBD$_{B.1.1.7}$ was not affected (K_D = ~ 10 nM), it decreased for RBD$_{B.1.351}$ (K_D = ~ 26 nM) (Appendix Fig S7B). In accordance with
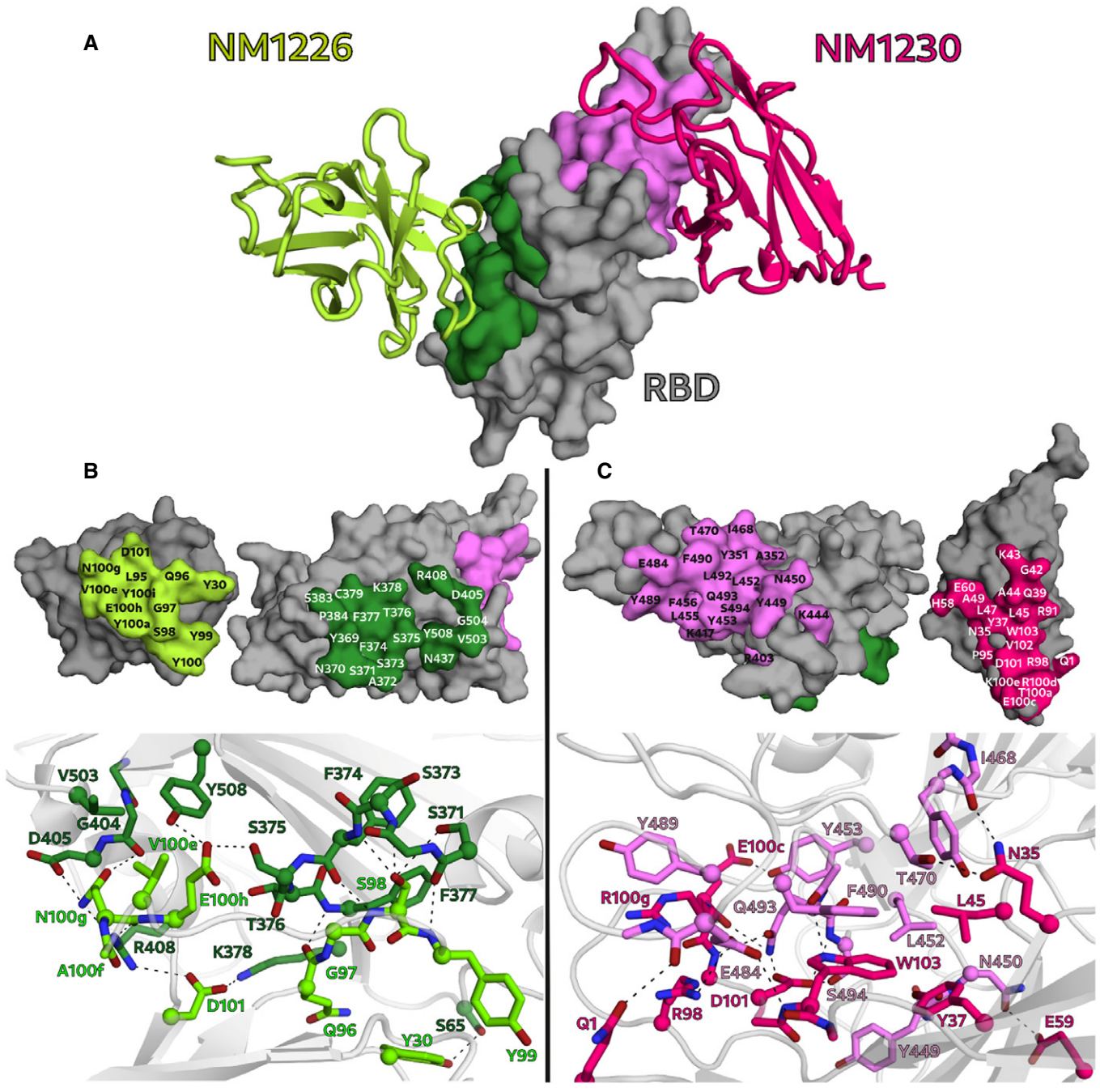
Figure 3. NM1226 and NM1230 binding epitopes on RBD.

A   Overview of the binding epitopes of NM1226 (light green) on RBD (dark green) and NM1230 (magenta) on RBD (light pink), respectively.

B, C The binding interfaces of NM1226 (B) and NM1230 (C) with the RBD are shown as surface representation (upper panels) and in a close-up view as balls and sticks with direct hydrogen bonds and salt bridges indicated as dotted lines (lower panels). All residues involved in contact formation at a distance cutoff of < 4 Å are listed and labeled. A close-up view of both interfaces reports a detailed interaction map of both contact areas.

our structural epitope analysis, the reduced affinity is most likely due to the loss of the charged interaction between $R_{NM1230}98$ to $E_{RBD}484$. Also, an additional influence by mutation $K_{RBD}417{>}N$ cannot be excluded, since position 417 is part of the NM1230 binding epitope but does not form direct contact to NM1230.

**Generation of a biparatopic Nb with improved efficacies**

Having identified potent Nbs with strong neutralizing characteristics, we proposed that candidates derived from Nb-Set1 (NM1226) and Nb-Set2 (NM1230) might act synergistically. To examine this,

we genetically fused the coding sequences of NM1230 and NM1226 head-to-tail thereby inserting a flexible Gly-Ser linker ($(G_4S)_4$) of 20 amino acids and generated a biparatopic Nb (bipNb) named NM1267. Following production and purification from mammalian cells (Fig 4A), we determined the affinity and analyzed its performance in our multiplex ACE2 competition assay and in the VNT. Compared with the monovalent format, the bipNb showed considerably improved affinities not only for RBD_{wt} ($K_D = \sim 0.5$ nM) but also for RBD mutants ($K_D = \sim 0.6$ nM for $RBD_{B.1.1.7}$; $K_D = \sim 1.15$ nM for $RBD_{B.1.351}$) (Fig 4B and C; Appendix Fig S7C). In line with these findings, it revealed an outstanding inhibition of ACE2 binding to RBD, S1, and spike with an IC_{50} in the low picomolar range (Fig 4B and D). Additionally, we observed an increased potency for viral neutralization indicated by an IC_{50} of ~ 0.9 nM (Fig 4B and E), demonstrating that the bipNb NM1267 embodies a substantially refined tool which is beneficial for viral neutralization and competitive binding studies.

NeutrobodyPlex—using Nbs to determine a SARS-CoV-2 neutralizing immune response

Currently available serological assays provide data on the presence and distribution of antibody subtypes against different SARS-CoV-2 antigens within serum samples of infected and recovered SARS-CoV-2 patients (Amanat *et al*, 2020; preprint: Lassaunière *et al*, 2020; Robbiani *et al*, 2020; preprint: Roxhed *et al*, 2020; Stadlbauer *et al*, 2020; Becker *et al*, 2021; Fink *et al*, 2021). However, they do not differentiate between total and neutralizing RBD-binding antibodies, which sterically inhibit viral entry via ACE2 (Ju *et al*, 2020; Rogers *et al*, 2020; Tang *et al*, 2020). To address this important issue, our bipNb NM1267 might be a suitable surrogate to monitor the emergence and presence of neutralizing antibodies in serum samples of patients. We speculated that NM1267 specifically displaces such antibodies from the RBD:ACE2 interface, which can be monitored as a declining IgG signal (Fig 5A). To test this hypothesis, we first co-incubated antigen-coated beads comprising RBD, S1, or spike with a dilution series of NM1267 and a well-characterized NAb (clone REGN10933), targeting an epitope within the RBD:ACE2 interface (Hansen *et al*, 2020; Jiang *et al*, 2020). As control, we applied the IgG 4A8 (anti-Spike NTD) which was shown to bind an epitope outside the RBD (Chi *et al*, 2020b). Antigen-bound IgGs in the presence of bipNb were detected via an anti-human IgG-PE as mean fluorescent intensities (MFI). Upon addition of increasing concentrations of NM1267, a distinct displacement of REGN10933 was observed (Appendix Fig S8A and B). In contrast, binding of 4A8 was not affected in the presence of NM1267 (Appendix Fig S8C and D), thus proofing the suitability of the bipNb as potent surrogate to displace NAbs targeting the RBD:ACE2 interface. In a next step, we generated a high-throughput competitive binding assay, termed NeutrobodyPlex, by implementing the NM1267 in a recently
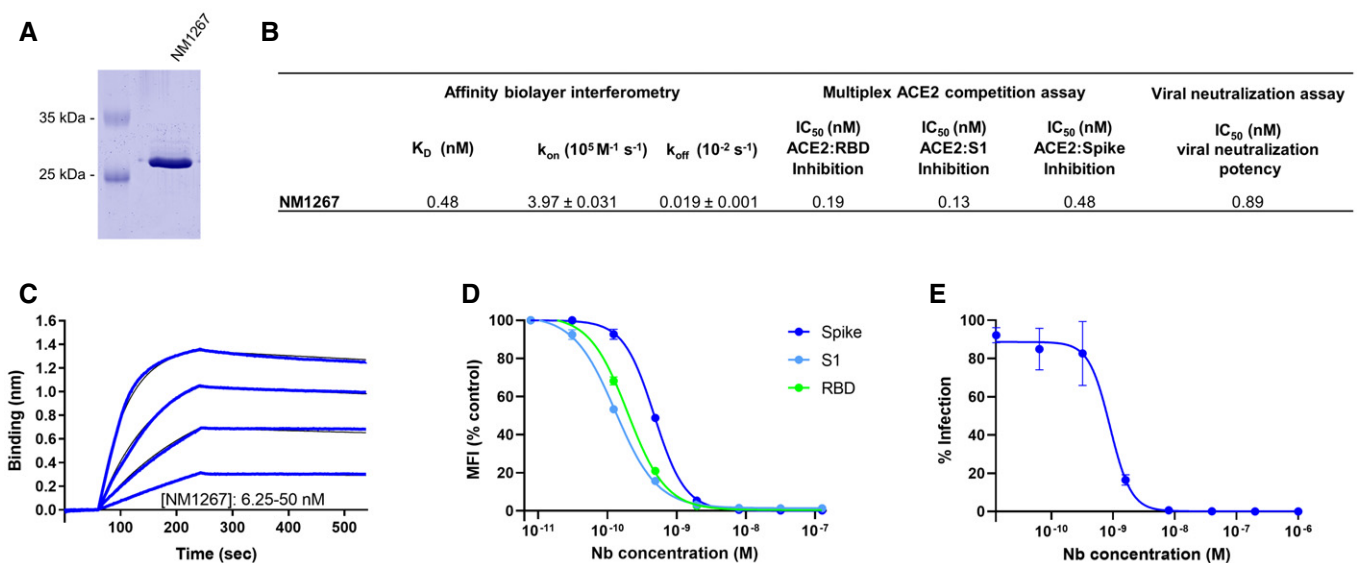
Figure 4. Biparatopic NM1267 competes with ACE2 and neutralizes SARS-CoV-2 infection.

A Coomassie staining of 1 μg of purified biparatopic Nb NM1267.

B Table summarizing the affinity (K_D), association (k_{on}), and dissociation constants (k_{off}) determined by BLI and IC_{50} values of the multiplex ACE2 competition assay and virus neutralization assay obtained for NM1267.

C Sensogram of affinity measurements via BLI using four concentrations (6.25–50 nM) of NM1267.

D Results from multiplex ACE2 competition assay are shown for the three spike-derived antigens: RBD, S1-domain (S1), and homotrimeric spike (Spike). Color-coded beads coated with the respective antigens were co-incubated with biotinylated ACE2 and dilution series of NM1267 (8 pM to 126 nM) followed by measuring residual binding of ACE2. MFI signals were normalized to the maximum detectable signal per antigen given by the ACE2-only control. IC_{50} values were calculated from a four-parametric sigmoidal model. Data are presented as mean ± s.d. of three technical replicates.

E Neutralization potency of NM1267 was analyzed in Caco-2 cells using the SARS-CoV-2-mNG infectious clones. Infection rate normalized to virus-only infection control is illustrated as percent of infection (% Infection). IC_{50} value was calculated from a four-parametric sigmoidal model, and data are presented as mean ± s.e.m. of three biological replicates ($n = 3$).
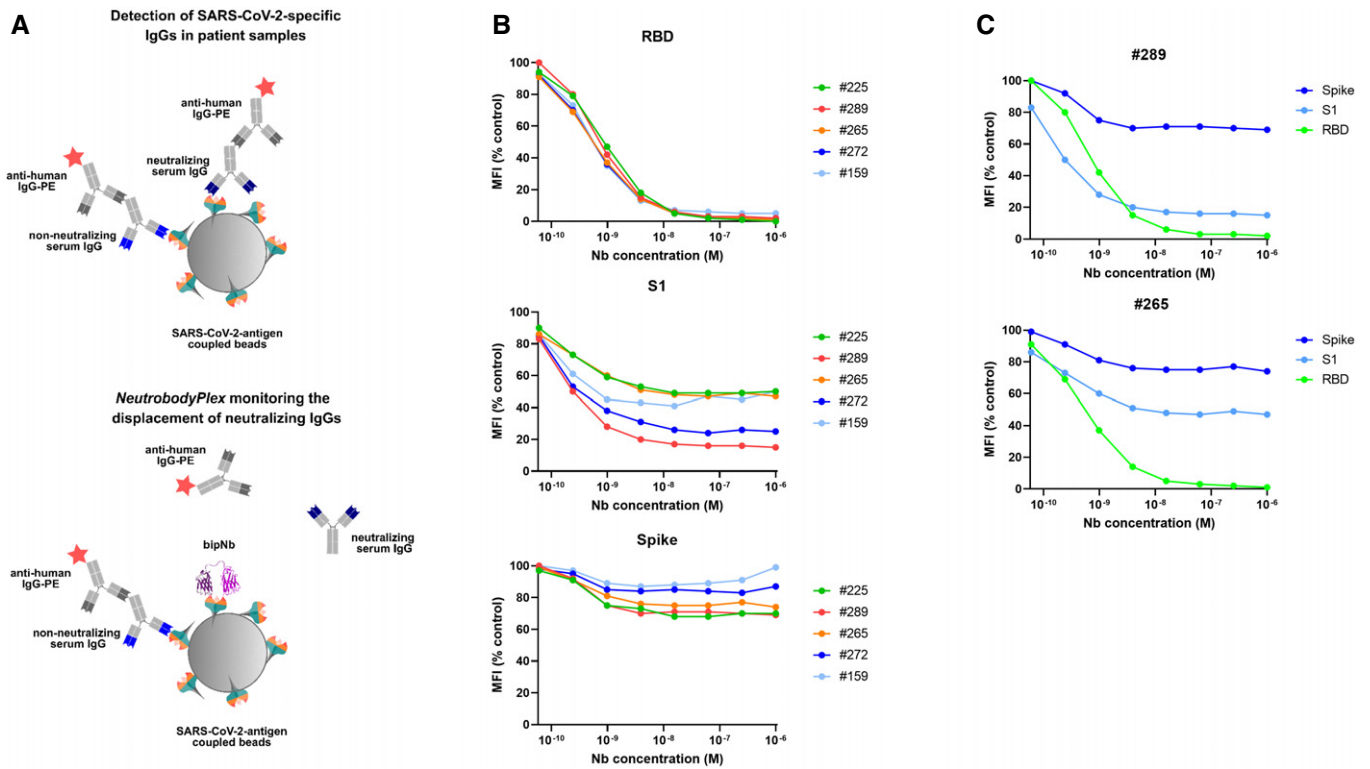
Figure 5. NeutrobodyPlex: multiplex competitive binding assay to monitor a neutralizing immune response in patients.

A   Schematic illustration of the NeutrobodyPlex. The displacement of serum-derived neutralizing IgGs binding to SARS-CoV-2 antigens upon addition of bipNb is measured. In presence of neutralizing IgGs, the fluorescent signal from anti-human-IgG-PE is inversely proportional to the applied bipNb concentration.

B   For the NeutrobodyPlex, antigen-coated beads comprising RBD, S1, or spike were co-incubated with serum samples from five patients and a dilution series of NM1267 (1 µM to 6 pM) (*n* = 1). Mean fluorescent intensities (MFI) derived from antigen-bound IgGs in the presence of bipNb normalized to the MFI values of IgGs in the serum-only samples, illustrated as MFI (% control), are shown.

C   For two patient-derived serum samples (#289, #265), differences in the NM1267-mediated displacement of IgGs binding the three spike-derived antigens (RBD, S1, spike) are shown.

developed, multiplexed serological assay (Becker *et al*, 2021). Initially, serum samples from five patients were screened. When analyzing IgG binding to RBD, we detected a complete displacement in the presence of ~ 63 nM NM1267. Similarly, a distinct signal reduction for S1 binding IgGs became observable reaching a plateau upon addition of ~ 63 nM NM1267 (Fig 5B, Appendix Table S3). Notably, we observed only minor signal reduction when analyzing spike-binding IgGs (Fig 5B) indicating that the majority of serum IgGs bound this large antigen at epitopes beyond the RBD:ACE2 interaction site (preprint: Heffron *et al*, 2020). From these data, we concluded that all five tested individuals comprise a substantial proportion of neutralizing IgGs that could be detected by competitive bipNb binding using RBD or S1 as antigens. To demonstrate that the NeutrobodyPlex can determine the presence of neutralizing IgGs at detailed resolution, we highlight the results of monitoring the displacement of S1-binding IgGs in two selected patient samples #289 and #265. Upon addition of increasing concentrations of NM1267, we observed a prominent signal reduction of ~ 85% in patient #289, whereas samples of patient #265 only revealed ~ 53% displacement of S1-binding IgGs (Fig 5C).

Next, we compared the NeutrobodyPlex on RBD with the cell-based VNT by analyzing a set of 18 serum samples from

convalescent SARS-CoV-2 patients, collected on days 19–57 following a positive PCR test result, and four control samples from healthy donors. To detect differences within the overall immune response and to qualify the neutralizing capacity of the serum samples, we performed the NeutrobodyPlex using two concentrations of NM1267 previously shown to completely (1 µM) or partly (1 nM) displace IgGs binding the RBD:ACE2 interface (Fig 5B). Both assays demonstrated the presence of neutralizing antibodies in all serum samples from convalescent individuals, when using the 1:40 serum dilution in the VNT or when adding 1 µM NM1267 in the NeutrobodyPlex (Fig 6A and B). Notably, neutralizing IgGs were not detected in any of the control serum samples (Appendix Table S4). Further analysis of the displacement of RBD-binding IgGs using the lower NM1267 concentration (1 nM) revealed significant differences in individual neutralization capacities. Whereas some patient samples contained potent neutralizing IgGs that could displace NM1267 upon binding to the RBD:ACE2 interface (Fig 6A, high MFI (% control), light squares), continuous bipNb-mediated displacement of IgGs was detectable in other samples (Fig 6A, low MFI (% control), dark squares), indicating the presence of IgGs with lower neutralizing potency. In parallel, VNT considering the full range of serum dilutions also showed significant differences in individual neutralizing
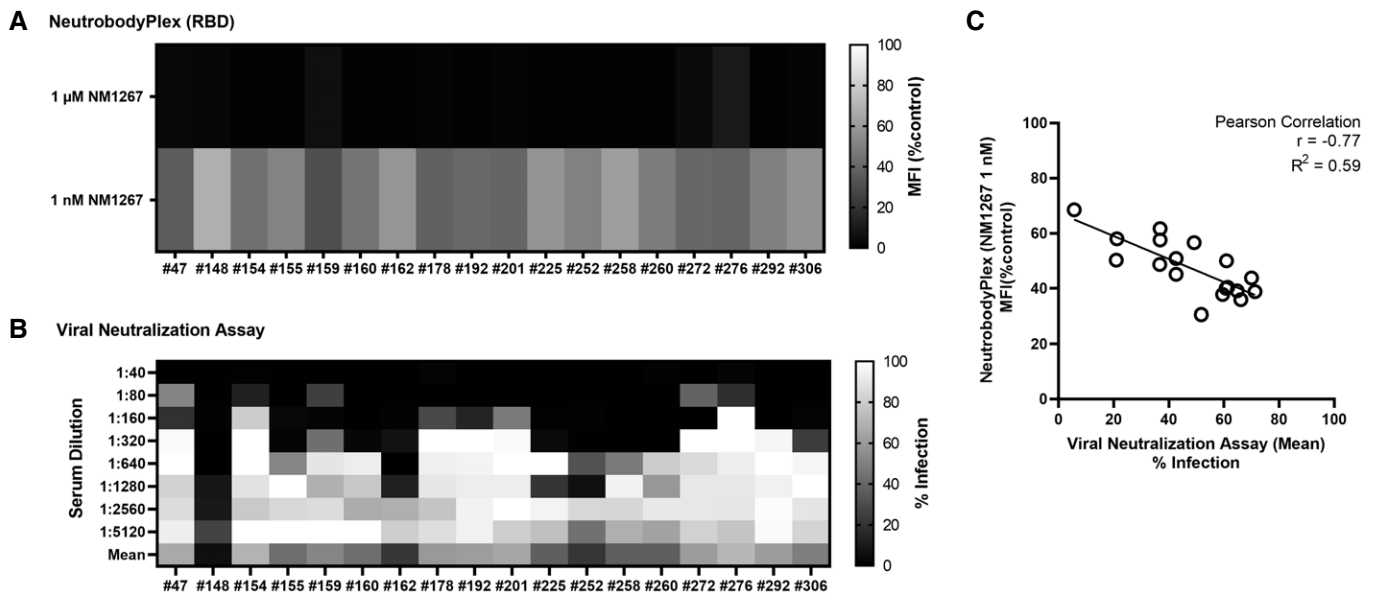
Figure 6. Specificity of the NeutrobodyPlex in determining individual neutralizing immune responses.

Serum samples of 18 convalescent SARS-CoV-2-infected individuals were analyzed using the NeutrobodyPlex and the virus neutralization assay.

A   For the NeutrobodyPlex on RBD, two concentrations of NM1267 (1 μM and 1 nM) were applied (*n* = 1). Light colored squares (high MFI (%control)) are indicative for IgGs outcompeting NM1267 from the RBD:ACE2 interface; dark colored squares (low MFI (%control)) show a continuous displacement of IgGs from serum samples in the presence of NM1267.
B   For the virus neutralization assay, serial dilutions of the serum samples (1:40 – 1:5,120) were applied (*n* = 1). Dark colored squares are indicative for a low infection level (low % Infection); light colored squares show a high infection rate (high % Infection).
C   The mean percent of infection (% Infection) derived from all individual serum dilutions obtained by the virus neutralization assay was calculated and plotted against the respective MFI (% control) obtained from the NeutrobodyPlex on RBD in the presence of NM1267 (1 nM). The Pearson correlation determined a negative correlation with $r = -0.77$ and $R^2 = 0.59$ ($P < 0.05$).

potency (Fig 6B). To confirm the validity of the NeutrobodyPlex, we calculated the mean percent of infection (% Infection) of all individual serum dilutions and plotted these values against the respective MFI (% control) obtained from the NeutrobodyPlex (Fig 6C). The observed negative correlation (high MFI (% control) vs. low mean % Infection) strongly suggests that the NeutrobodyPlex allows the detection of individual neutralizing immune responses and assessment of their potency.

For final validation, we screened a serum sample set of 112 convalescent SARS-CoV-2-infected and eight uninfected individuals in the NeutrobodyPlex. In addition to RBD, S1, and spike, we included the S2 domain (S2) and the nucleocapsid (N) of SARS-CoV-2 to monitor total immune response. Incubation with the higher amount of NM1267 (1 μM) revealed the presence of neutralizing IgGs in all serum samples from convalescent patients as shown by an effective replacement of IgGs from RBD, S1, and partially from spike, but not from S2 or N (Fig EV4A). In line with findings from the initial patient sample set, differences in the potencies of neutralizing IgGs present in the individual serum samples became detectable upon addition of the lower concentration of NM1267 (1 nM) (Fig EV4B). As observed previously, non-infected individuals did not display any detectable antibody signal for the presented antigens. We further analyzed, if total level of SARS-CoV-2-specific IgGs correlates with the amount of neutralizing IgGs. As a marker for total IgGs we defined MFI values measured for spike-binding IgGs (Fig 7A and B) and N-binding IgGs (Fig 7C and D) and plotted them against normalized MFI values from IgGs binding to RBD in the

presence of both concentrations of NM1267 (1 μM, 1 nM (MFI RBD (%control)) (Fig 7A–D). Overall, we found high variability of total IgGs (MFI: ∼ 3,500 – ∼ 50,000) in tested serum samples. Upon addition of the higher concentration of NM1267 (1 μM), we observed a complete displacement of IgGs binding the RBD:ACE2 interface independent of the amount of total IgG (Fig 7A and C), whereas upon addition of the lower amount of bipNb, substantial differences of the neutralizing capacity of individual patient samples became detectable (Fig 7B and D). Specifically, using the NeutrobodyPlex, we identified several individuals who have low levels of SARS-CoV-2-specific antibodies overall but have strong neutralizing IgGs (Fig 7B and D highlighted with light-blue background). From these data, it can be concluded that the appearance of neutralizing IgGs after infection with SARS-CoV-2 does not correlate with an overall high SARS-CoV-2-specific antibody titer. However, an increased likelihood for a higher neutralizing capacity in patients showing an overall high SARS-CoV-2-specific IgG level also cannot be excluded. In summary, these results demonstrate how the NeutrobodyPlex can not only provide detailed information on the presence of neutralizing antibodies in patient samples, but also allow qualitative and quantitative assessment of the individual immune response.

Discussion

In this study, we identified a set of 11 high-affinity RBD-specific Nbs derived from an immunized animal. The high success rate in
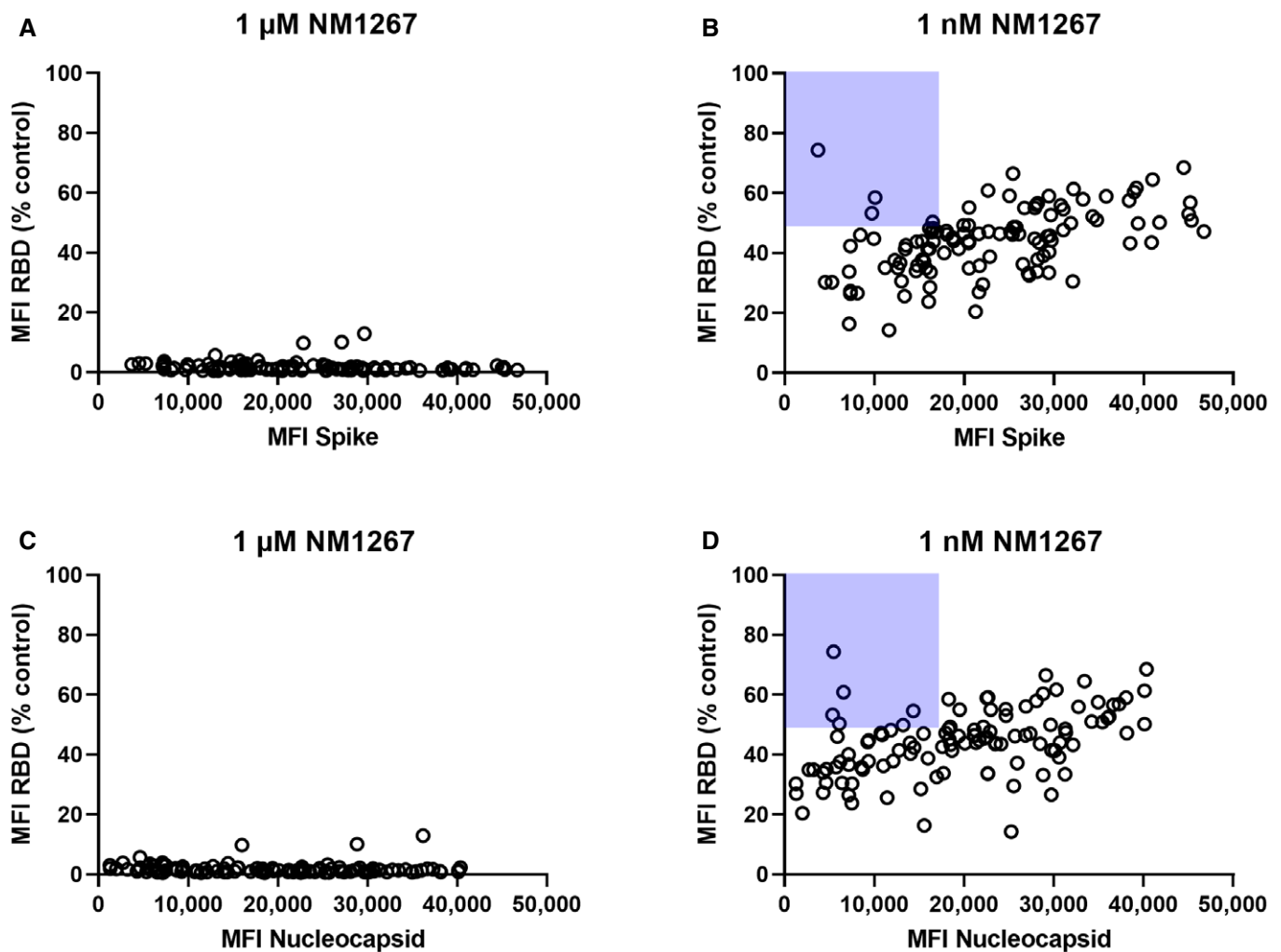
Figure 7. NeutrobodyPlex enables a differentiated analysis on neutralizing IgGs compared with total SARS-CoV-2 binding IgGs in individuals.

To investigate neutralizing capacities in relation to total levels of SARS-CoV-2-specific IgGs in individuals, serum samples of 112 convalescent SARS-CoV-2 infected individuals were analyzed using the NeutrobodyPlex (*n* = 1). Individuals displaying strong neutralizing IgGs but showed an overall low level of SARS-CoV-2-specific antibodies are highlighted with light-blue background.

A, B Total IgGs derived from spike-binding IgGs were plotted against normalized MFI values from IgGs binding to RBD (MFI RBD (%control)) in the presence of both concentrations of NM1267 (1 µM; 1 nM).

C, D Total IgGs derived from N-binding IgGs were plotted against normalized MFI values from IgGs binding to RBD (MFI RBD (%control)) in the presence of both concentrations of NM1267 (1 µM; 1 nM).

identifying well-functioning Nbs is consistent with recently reported findings describing an ever-growing list of such binders derived from either immunized (Chi *et al,* 2020a; preprint: Esparza *et al,* 2020; preprint: Gai *et al,* 2020; Hanke *et al,* 2020; preprint: Nieto *et al,* 2020; Wrapp *et al,* 2020a; Xiang *et al,* 2020; Koenig *et al,* 2021) or naïve/ synthetic Nb gene libraries (Chi *et al,* 2020a; Custodio *et al,* 2020; Huo *et al,* 2020; preprint: Schoof *et al,* 2020; preprint: Walter *et al,* 2020; preprint: Ahmad *et al,* 2021). By establishing a workflow involving a multiplexed ACE2 competition assay followed by a cell-based VNT, we were able to select high-affinity binding candidates that blocked the interaction between RBD and ACE2 in the context of various SARS-CoV-2 spike-derived antigens and additionally exhibited high viral neutralizing potency. Detailed epitope analysis based on HDX-MS and structural analysis of RBD:

Nb complexes for two of the most potent binders further revealed the molecular basis of the observed neutralizing effect. Both structurally characterized Nbs (NM1226 and NM1230) bind to opposite sides of the RBD. Similar to the recently reported strong neutralizing Nb-Ty1, which displays a partially overlapping binding epitope (Hanke *et al,* 2020), NM1230 can bind to the "down" and "up" form of the spike protomer, thereby inhibiting ACE2 interaction either by steric collisions or due to the overlapping interface (Appendix Fig S5). In contrast, NM1226 binds only the "up" conformation on RBD. As the RBD:ACE2 interface is not addressed by NM1226, its neutralizing effect can most likely be explained due to steric collision. Moreover, NM1226 seems to be more robust against mutations within the RBD:ACE2 interface. Indeed, our comprehensive analyses showed that both Nbs are able to bind the

RBD variant of B.1.1.7 (UK) with similar affinities to RBD_{wt}. In comparison, binding of NM1226 to RBD of the more virulent SARS-CoV-2 variant B.1.351 (South Africa) was unaffected, whereas NM1230 showed reduced affinity. To combine the advantages of both single Nbs, add avidity to the construct, and reduce the possibility of viral escape by targeting two independent epitopes, we generated the biparatopic Nb NM1267, which performed excellently in our multiplexed ACE2 competition assay and VNT with IC_{50} values in the picomolar range. In addition, NM1267 revealed high binding affinities to RBDs either derived from B.1.1.7 (UK) or B.1.351 (South Africa) from which we postulated that NM1267 represents a substantially refined tool with highly versatile binding properties.

The ongoing pandemic and the emergence of new mutations leading to more infectious viral strains make the development of improved diagnostic tools and therapies essential. In this context, several studies have shown that humoral immune responses resulting in antibodies directed against the interaction site of RBD and the ACE2 receptor exposed on epithelial cells of the human respiratory tract lead to a strong viral neutralizing effect (Brouwer *et al,* 2020; Cao *et al,* 2020; Chi *et al,* 2020b; Jiang *et al,* 2020; Ju *et al,* 2020; Robbiani *et al,* 2020; Shi *et al,* 2020; Tai *et al,* 2020; Yu *et al,* 2020; Zohar & Alter, 2020). Consequently, there is a keen demand to expand currently applied serological diagnostics toward a more differentiated analysis to specifically monitor and classify immune responses. Therefore, we extended our previously described multiplex serological assay (Becker *et al,* 2021) and implemented NM1267 as a potent surrogate, which specifically displaces IgGs from binding to the RBD:ACE2 interface to monitor the presence and emergence of NAbs in serum samples from convalescent and vaccinated individuals. The NeutrobodyPlex was not only able to successfully detect the overall presence of NAbs targeting the RBD:ACE2 interface, but also to categorize them according to their potency. In addition, it allows comparison between neutralizing IgGs and the total level of SARS-CoV-2 antibodies, which is an important parameter, e.g., for determining the success of a vaccine. However, the NeutrobodyPlex approach is limited in that only NAbs that preferably bind the RBD:ACE2 interaction site can be detected. While such NAbs have been proposed as a milestone for the development of a protective immune response against SARS-CoV-2 (Brouwer *et al,* 2020; Cao *et al,* 2020; Ju *et al,* 2020; Robbiani *et al,* 2020; Shi *et al,* 2020; Tai *et al,* 2020; Yu *et al,* 2020), several NAbs have also been identified that bind outside the RBD:ACE2 interface (Chi *et al,* 2020b; Liu *et al,* 2020), which would be missed by the NeutrobodyPlex.

Compared with recently described assays that directly measure antibody-mediated displacement of ACE2 to determine a neutralizing immune response (preprint: Byrnes *et al,* 2020; Tan *et al,* 2020; Walker *et al,* 2020), a much lower binding affinity of ~ 15–30 nM of RBD to ACE2 must be considered. Consequently, neutralizing immune responses may be detected even in the presence of weak RBD-binding antibodies, which could lead to a false statement of immune protection. Additionally, the usage of small-sized, Nb-derived surrogates further lowers the possibility of a non-targeted and non-reproducible displacement of ACE2, e.g., mediated by steric inhibition and dimerization effects derived from non-specifically binding IgGs. Finally, the multiplex design of the NeutrobodyPlex allows rapid expansion to different antibody isotypes and additional antigens to include, e.g., new relevant RBD mutants that are of importance for advanced vaccination strategies.

To our knowledge, the NeutrobodyPlex demonstrates for the first time a multiplex, antigen-resolved analysis of the presence of neutralizing human IgGs in convalescent individuals suffering from SARS-CoV-2 infection. Compared with other assays, the NeutrobodyPlex can be performed in a fully automated, high-throughput manner and can be readily used for large cohort screening. Since it requires only non-living and non-infectious viral material, costs and safety conditions can be significantly reduced (Muruato *et al,* 2020; Scholer *et al,* 2020; Tan *et al,* 2020). In addition, the NeutrobodyPlex is highly sensitive, as low dilutions of serum (tested dilution: 1:400) are sufficient for analysis, significantly reducing patient material compared with standard assays. However, in its current state, important parameters such as specificity and linearity remain to be defined and more detailed information about the binding properties of neutralizing IgGs are needed. Thus, we have initiated a collaboration with the University Hospital Tuebingen to test large cohorts of convalescent and vaccinated individuals with the NeutrobodyPlex in comparison with RBD ELISA assays and VNTs.

In summary, we envision that the NeutrobodyPlex will provide unique opportunities for detailed classification of individual immune status with respect to the development of protective antibodies. In addition, the NeutrobodyPlex could represent a valuable approach to monitor the efficiency of currently starting vaccination campaigns and their long-term efficacy.

Materials and Methods

Expression constructs

For bacterial expression of Nbs, sequences were cloned into the pHEN6 vector (Arbabi Ghahroudi *et al,* 1997), thereby adding a C-terminal His_6-Tag for IMAC purification as described previously (Rothbauer *et al,* 2008; Kirchhofer *et al,* 2010). To generate the biparatopic Nb NM1267 from NM1230 and NM1226, cDNAs were amplified by PCR using forward primer 1230Nfor 5´- GGA CGT CTC AAC TCT CAA GTG CAG CTG GTG GAG TC - 3´ and reverse primer 1230Nrev 5´- CAC CAC CGC CAG ATC CAC CGC CAC CTG ATC CTC CGC CTC CTG AGG ACA CGG TGA CCT GGG CCC - 3´ for NM1230 Nb, and forward primer 1226Cfor 5´- GGT GGA TCT GGC GGT GGT GGA AGT GGT GGC GGA GGT AGT CAG GTG CAG CTG GTG GAA T - 3´ and reverse primer 1226Crev 5´- GGG GAA TTC AGT GAT GGT GAT GGT GGT GTG AGG ACA CGG TGA CCG GGG CC - 3´ for NM1226 Nb. Next PCR products from Nbs NM1230 and NM1226 were genetically fused via an internal $(G_4S)_4$–linker by fusion PCR using forward primer 1230Nfor and reverse primer 1226Crev. Full-length cDNA was cloned into Esp3I and EcoRI site of pCDNA3.4 expression vector with N-terminal signal peptide (MGWTLVFLFLLSVTAGVHS) for secretory pathway that comprises Esp3I site.

The pCAGGS plasmids encoding the stabilized homotrimeric spike protein and the receptor-binding domain (RBD) of SARS-CoV-2 were kindly provided by F. Krammer (Amanat *et al,* 2020). The cDNA encoding the S1 domain (aa 1 - 681) of the SARS-CoV-2 spike protein was obtained by PCR amplification using the forward primer S1_CoV2-for 5´- CTT CTG GCG TGT GAC CGG - 3´ and reverse

primer S1_CoV2-rev 5′ - GTT GCG GCC GCT TAG TGG TGG TGG TGG TGG TGG GGG CTG TTT GTC TGT GTC TG - 3′ and the full-length SARS-CoV-2 spike cDNA as template and cloned into the XbaI/ NotI-digested backbone of the pCAGGS vector, thereby adding a C-terminal His$_6$-tag. The optimized sequence of the full-length nucleocapsid protein of SARS-CoV-2 with an N-terminal hexahistidine (His$_6$)-tag was obtained by DNA synthesis (Thermo Fisher Scientific) (GenBank accession numbers QHD43423.2) and cloned into the bacterial expression vector pRSET2b (Thermo Fisher Scientific) via NdeI/HindIII restriction sites. RBDs of virus mutants were generated by PCR amplification of fragments from wild-type DNA template followed by fusion PCRs to introduce described mutation N501Y for B.1.1.7 (RBD$_{B.1.1.7}$) and additional mutations K417N and E484K for B.1.351 (RBD$_{B.1.351}$) (https://covariants.org/shared-mutations). Forward primer RBDfor 5′- ATA TCT AGA GCC ACC ATG TTC GTG TTT CTG G - 3′ and reverse primer N501Yrev 5′- CCA CGC CAT ATG TGG GCT GAA AGC CGT AG - 3′ were used for amplification of fragment 1, forward primer N501Yfor 5′- GGC TTT CAG CCC ACA TAT GGC GTG GGC TAT CAG C - 3′ and reverse primer RBDrev 5′- AAG ATC TGC TAG CTC GAG TCG C - 3′ were used for amplification of fragment 2. Both fragments containing an overlap sequence at the 3` and 5` end were fused by an additional PCR using forward primer RBDfor and RBDrev. Based on cDNA for RBD$_{B.1.1.7}$, additional mutations of B.1.351 were introduced by PCR amplification of three fragments using forward primer RBDfor and reverse primer K417Nrev 5′- GTT GTA GTC GGC GAT GTT GCC TGT CTG TCC AGG G - 3′, forward primer K417Nfor 5′- GAC AGA CAG GCA ACA TCG CCG ACT ACA ACT ACA AGC - 3′ and reverse primer E484Krev 5′- GCA GTT GAA GCC TTT CAC GCC GTT ACA AGG GGT - 3′, forward primer E484Kfor 5′- GTA ACG GCG TGA AAG GCT TCA ACT GCT ACT CCC - 3′ and reverse primer RBDrev. Amplified fragments were assembled by subsequent fusion PCR using forward primer RBDfor and RBDrev. DNA coding for mutant RBDs (amino acids 319–541 of respective spike proteins) was cloned into Esp3I and EcoRI site of pCDNA3.4 expression vector as described for NM1267. All expression constructs were verified by sequence analysis.

Nb libraries

Alpaca immunizations with purified RBD and Nb library construction were carried out as described previously (Rothbauer *et al*, 2006). Animal immunization was approved by the government of Upper Bavaria (Permit number: 55.2-1-54-2532.0-80-14). In brief, nine weeks after immunization of an animal (*Vicugna pacos*) with C-terminal histidine-tagged RBD (RBD-His$_6$), ~ 100 ml blood was collected and lymphocytes were isolated by Ficoll gradient centrifugation using the Lymphocyte Separation Medium (PAA Laboratories GmbH). Total RNA was extracted using TRIzol (Life Technologies) and mRNA was reverse transcribed to cDNA using a First-Strand cDNA Synthesis Kit (GE Healthcare). The Nb repertoire was isolated in three subsequent PCRs using the following primer combinations (i) CALL001 (5′-GTC CTG GCT GCT CTT CTA CA A GG-3′) and CALL002 (5′-GGT ACG TGC TGT TGA ACT GTT CC-3′); (ii) forward primer set FR1-1, FR1-2, FR1-3, FR1-4 (5′-CAT GGC NSA NGT GCA GCT GGT GGA NTC NGG NGG-3′, 5′-CAT GGC NSA NGT GCA GCT GCA GGA NTC NGG NGG-3′, 5′-CAT GGC NSA NGT GCA GCT GGT GGA NAG YGG NGG-3′, 5′-CAT GGC NSA NGT GCA

GCT GCA GGA NAG YGG NGG-3′) and reverse primer CALL002 and (3) forward primer FR1-ext1 and FR1-ext2 (5′-GTA GGC CCA GCC GGC CAT GGC NSA NGT GCA GCT GGT GG-3`, 5`-GTA GGC CCA GCC GGC CAT GGC NSA NGT GCA GCT GCA GGA-3′ A-) and reverse primer set FR4-1, FR4-2, FR4-3, FR4-4, FR4-5 and FR4-6 (5`-GAT GCG GCC GCN GAN GAN ACG GTG ACC NGN RYN CC-3′. 5`-GAT GCG GCC GCN GAN GAN ACG GTG ACC NGN GAN CC-3′. 5`-GAT GCG GCC GCN GAN GAN ACG GTG ACC NGR CTN CC-3′. 5`-GAT GCG GCC GCR CTN GAN ACG GTG ACC NGN RYN CC-3′. 5`-GAT GCG GCC GCR CTN GAN ACG GTG ACC NGN GAN CC-3′. 5`-GAT GCG GCC GCR CTN GAN ACG GTG ACC NGR CTN CC-3′) introducing SfiI and NotI restriction sites. The Nb library was subcloned into the SfiI/ NotI sites of the pHEN4 phagemid vector (Arbabi Ghahroudi *et al*, 1997).

Nb screening

For the selection of RBD-specific Nbs, two consecutive phage enrichment rounds were performed. TG1 cells containing the "immune"-library in pHEN4 were infected with the M13K07 helper phage, hence the V$_H$H domains were presented superficially on phages. For each round, 1×10^{11} phages of the "immune"-library were applied on RBD either directly coated on immunotubes (10 µg/ml) or biotinylated RBD (5 µg/ml) immobilized on 96-well plates pre-coated with Streptavidin. In each selection round, extensive blocking of antigen and phages was performed by using 5% milk or BSA in PBS-T and with increasing panning round, PBS-T washing stringency was intensified. Bound phages were eluted in 100 mM triethylamin, TEA (pH 10.0), followed by immediate neutralization with 1 M Tris/HCl (pH 7.4). For phage preparation for following rounds, exponentially growing TG1 cells were infected and spread on selection plates. Antigen-specific enrichment for each round was monitored by comparing colony number of antigen vs. no antigen selection. Following panning 492 individual clones of the second selection round were screened by standard Phage-ELISA procedures using a horseradish peroxidase-labeled anti-M13 monoclonal antibody (GE Healthcare).

Protein expression and purification

RBD-specific Nbs were expressed and purified as previously described (Rothbauer *et al*, 2008; Kirchhofer *et al*, 2010). For the expression of SARS-CoV-2 proteins (RBD, stabilized homotrimeric spike, and S1 domain), Expi293 cells were used (Stadlbauer *et al*, 2020) and the biparatopic Nb NM1267 was expressed using the ExpiCHO system. The full-length nucleocapsid protein (N) was expressed using E.coli BL21 (DE3) cells and purification was performed as described previously (Becker *et al*, 2021). The S2 ecto-domain of the SARS-CoV-2 spike protein (aa 686–1213) was purchased from Sino Biological (cat # 40590, lot # LC14MC3007). For quality control, all purified proteins were analyzed via SDS–PAGE according to standard procedures. For immunoblotting, proteins were transferred on nitrocellulose membranes (Bio-Rad Laboratories) and detection was performed using anti-His antibody (Penta-His Antibody, #34660, Qiagen) followed by donkey-anti-mouse antibody labeled with AlexaFluor647 (Invitrogen) using a Typhoon Trio scanner (GE Healthcare, Freiburg, Germany; excitation 633 nm, emission filter settings 670 nm BP 30).

Biophysical biolayer interferometry (BLI)

To analyze the binding affinity of purified Nbs toward RBD, biolayer interferometry (BLItz, ForteBio) was performed as per the manufacturer's protocols. Briefly, biotinylated RBD was immobilized on single-use high-precision streptavidin biosensors (SAX). Depending on the affinity of the RBD-Nb interaction, an appropriate concentration range (15.6 nM to 2 µM) of Nbs was used. For each run, four different Nb concentrations were measured as well as a reference run using PBS instead of Nb in the association step. As negative control, GFP-Nb (500 nM) was applied in the binding studies. Global fits were determined using the BLItzPro software and the global dissociation constant (K_D) was calculated.

Bead-based multiplex ACE2 competition assay

Purified RBD, S1 domain, and homotrimeric spike of SARS-CoV-2 were covalently immobilized on spectrally distinct populations of carboxylated paramagnetic beads (MagPlex Microspheres, Luminex Corporation, Austin, TX) using 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC)/ sulfo-N-hydroxysuccinimide (sNHS) chemistry as described by Becker *et al* (2021). Briefly, beads were activated in activation buffer containing 5 mg/ml EDC and 5 mg/ml sNHS. For immobilization, activated beads and respective antigens were incubated in coupling buffer (500 mM MES, pH 5.0, 0.005% (v/v) Triton X-100) for 2 h at 21°C and individual bead populations were combined into a bead mix. For the bead-based ACE2 competition assay, Nbs were incubated with the bead mix (containing beads coupled with SARS-CoV-2 homotrimeric spike, RBD, and S1 proteins) and biotinylated ACE2 (Sino Biological), which competes for the binding of SARS-CoV-2 spike-derived antigens. Single Nbs or bipNbs were pre-diluted to a concentration of 6.3 µmol/L per Nb in assay buffer. Afterward, a fourfold dilution series was made over eight steps in assay buffer containing 160 ng/ml biotinylated ACE2. Subsequently, every dilution was transferred the same volume of bead mix in a 96-well half-area plate. The plate was incubated for 2 h at 21°C, shaking at 750 rpm. Beads were washed using a microplate washer (Biotek 405TS, Biotek Instruments GmbH) to remove unbound ACE2 or Nbs. R-phycoerythrin (PE)-labeled streptavidin was used to detect biotinylated ACE2. Measurements were performed with a FLEXMAP 3D instrument using the xPONENT Software version 4.3 (settings: sample size: 80 µl, 50 events, Gate: 7,500 – 15,000, Reporter Gain: Standard PMT).

NeutrobodyPlex: Bead-based multiplex neutralizing antibody detection assay

Based on the recently described automatable multiplex immunoassay (Becker *et al*, 2021), the NeutrobodyPlex was developed and similar assay conditions were applied. For the detection of neutralizing serum antibodies, the bead mix containing beads coupled with purified RBD (receptor-binding domain), S1 (S1 domain), spike (homotrimeric spike), S2 (S2 domain), or N (nucleocapsid) of SARS-CoV-2 was incubated with bipNb NM1267 (concentrations ranging from 1 µM to 6 pM) and purified NAbs (0.08 nM, REGN10933 Cell Sciences; 4A8 ProteoGenix) or serum samples of convalescent SARS-CoV-2 patients and healthy donors at a 1:400 dilution. As positive control and maximal signal detection per sample, serum

only was included. Bound serum IgGs were detected via PE-labeled anti-human-IF(ab')$_2$ Fragment (#109-116-098 Jackson ImmunoResearch) as previously described (Becker *et al,* 2021).

Hydrogen-Deuterium exchange

RBD deuteration kinetics and epitope elucidation

RBD (5 µl, 73 µM) was either incubated with PBS or RBD-specific Nbs (2.5 µl, 2.5 mg/ml in PBS) at 25°C for 10 min. Deuterium exchange of the pre-incubated nanobody-antigen complex was initiated by dilution with 67.5 µl PBS (150 mM NaCl, pH 7.4) prepared with D$_2$O and incubation for 5 and 50 min, respectively, at 25°C. To ensure a minimum of 90% of complex formation, the molar ratio of antigen to Nbs was calculated as previously described(Kochert *et al,* 2018), using the affinity constants of 1.37 nM (NM1228), 3.66 nM (NM1226), 3.82 nM (NM1223), 8.23 nM (NM1230), and 8.34 nM (NM1224) (pre-determined by BLI analysis). The final D$_2$O concentration was 90%. After 5 and 50 min at 25°C, aliquots of 15 µl were taken and quenched by adding 15 µl ice-cold quenching solution (0.2 M TCEP with 1.5% formic acid and 4 M guanidine HCl in 100 mM ammonium formate solution pH 2.2) resulting in a final pH of 2.5. Quenched samples were immediately snap-frozen. The immobilized pepsin was prepared by adding 60 µl of 50% slurry (in ammonium formate solution pH 2.5) to a tube and dried by centrifugation at 1,000 *g* for 3 min at 0°C and discarding the supernatant. Before injection, aliquots were thawed and added to the dried pepsin beads. Proteolysis was performed for 2 min in a water ice bath followed by filtration using a 22 µm filter and centrifugation at 1,000 *g* for 30 s at 0°C. Samples were immediately injected into a LC-MS system. Undeuterated control samples were prepared under the same conditions using H$_2$O instead of D$_2$O. The same protocol was applied for the Nbs without addition of RBD as well to create a list of peptic peptides. The HDX experiments of the RBD-Nb-complex were performed in triplicates. The back-exchange of the method as estimated using a standard peptide mixture of 14 synthetic peptides was 24%.

Chromatography and mass spectrometry

HDX samples were analyzed on a LC-MS system comprised of RSLC pumps (UltiMate 3000 RSLCnano, Thermo Fisher Scientific, Dreieeich, Germany), a chilling device for chromatography (MéCour Temperature Control, Groveland, MA, USA), and a mass spectrometer Q Exactive (Thermo Fisher Scientific, Dreieich, Germany). The chilling device contained the LC column (ACQUITY BEH C18, 1.7 µm, 300 Å, 1 mm × 50 mm (Waters GmbH, Eschborn, Germany)), a cooling loop for HPLC solvents, a sample loop, and the injection valve and kept all components at 0°C. Samples were analyzed using a two-step 20 min linear gradient with a flow rate of 50 µl/min. Solvent A was 0.1% (v/v) formic acid, and solvent B was 80% acetonitrile (v/v) with 0.1% formic acid (v/v). After 3 min desalting at 10% B, a 9 min linear gradient from 10 to 25% B was applied followed by an 8 min linear gradient from 25 to 68.8%. Experiments were performed using a Q Exactive (Thermo Fisher Scientific, Dreieich, Germany) with 70,000 resolutions instrument configurations as follows: sheath gas flow rate of 25; aux gas flow rate of 5; S-lens RF level of 50, spray voltage of 3.5 kV, and a capillary temperature of 300°C.

HDX data analysis

A peptic peptide list containing peptide sequence, retention time, and charge state was generated in a preliminary LC-MS/MS experiment. The peptides were identified by exact mass and their fragment ion spectrum using protein database searches by Proteome Discoverer v2.1.0.81 (Thermo Fisher Scientific, Dreieich, Germany) and implemented SEQUEST HT search engine. The protein database contained the RBD and the pepsin sequences. Precursor and fragments mass tolerance were set to 6 ppm and 0.05 Da, respectively. No enzyme selectivity was applied; however, identified peptides were manually evaluated to exclude peptides originated through cleavage after arginine, histidine, lysine, proline, and the residue after proline (Hamuro & Coales, 2018). FDR was estimated using q-values calculated by Percolator and only peptides with high-confidence identification (q-value $\leq$ 0.01) were included to the list. Peptides with overlapping mass, retention time, and charge in Nb and antigen digest, were manually removed. The deuterated samples were recorded in MS mode only, and the generated peptide list was imported into HDExaminer v2.5.0 (Sierra Analytics, Modesto, CA, USA). Deuterium uptake was calculated using the increase in the centroid mass of the deuterated peptides. HDX could be followed for 79% of the RBD amino acid sequence. The calculated percentage deuterium uptake of each peptide between RBD-Nb and RBD-only were compared. Any peptide with uptake reduction of 5% or greater upon Nb binding was considered as protected.

Crystallization and structural analysis

Production of the RBD domain and complex formation with Nbs for crystallization

The RBD domain was produced and purified as described previously (Stadlbauer *et al*, 2020). As a variation to the established protocol, the production of the RBD (residues 319–541) was performed in the Expi293F™ GntI- expression system (Thermo Fisher Scientific, Dreieich, Germany). Expi293F™ GntI- cells were cultivated (37°C, 125 rpm, 8% (v/v) CO_2) to a density of 5.5×10^6 cells/ml. The cells were diluted with Expi293F expression medium to a density of 3.0×10^6 cells/ml, followed by transfection of RBD plasmid (1 µg/ml cell culture) with Expifectamine (Thermo Fisher Scientific) dissolved in Opti-MEM medium (Thermo Fisher Scientific), according the manufacturer's instructions. 20 h post-transfection, the transfection enhancers were added as documented in the Expi293F™ GntI- cells manufacturer's instructions. The cell suspension was cultivated for 5 days (37°C, 125 rpm, 8% (v/v) CO_2) and centrifuged (4°C, 23,900 *g*, 20 min) to clarify the supernatant. The supernatant was supplemented with His-A buffer stock solution (final concentration in the medium: 20 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole, pH 7.4), before the solution was applied to a HisTrap FF crude column (GE Healthcare). The column was extensively washed with His-buffer-A (20 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole, pH 7.4) and 50 mM imidazole before the RBD was eluted with 280 mM imidazole from the column. The RBD was dialyzed against PBS and concentrated to 2 mg/ml. The nanobody complex was formed by mixing the RBD with the purified NM1226 or NM1230 in a molar ratio of 1:1.1, followed by incubation for 3 h at 4°C.

For crystallization, the complexes were treated with Endo H_f (New England Biolabs) to truncate the oligosaccharide chain on the RBD. Therefore, Endo H_f (300 U per mg RBD) was added to the RBD:Nb complex and incubated for 2 days at 8°C. Endo H_f was removed by passing the sample through a MBP-Trap column (GE Healthcare). Finally, a size exclusion chromatography using a SD200 16/60 column (GE Healthcare) exchanged the buffer (20 mM HEPES, 150 mM NaCl, pH 7.4) and separated the RBD:Nb complexes from aggregates and nanobody excess.

Crystallization

The RBD:NM1230 and RBD:NM1226 complexes were concentrated to 23.2 mg/ml and 29.6 mg/ml, respectively, prior to crystallization. Initial crystallization screening was performed on an ART Robbins Gryphon crystallization robot with placing 400 nL drop RBD:Nb and mixed in a 1:1 ratio with the reservoir solution. For RBD:NM1230, initial crystals appeared overnight in crystallization buffer (200 mM $MgCl_2$, 20% (w/v) PEG 3350, pH 5.9) at 20°C and grew to a final size of $30 \times 30 \times 120$ µm^3 within 4 days. After 8 weeks, RBD:NM1226 complex crystallized in a precipitation solution (50 mM K_2HPO_4, 20% (w/v) PEG 8000, pH 5.0) at 4°C. The crystals were harvested and frozen in liquid nitrogen until data collection.

Structure determination and refinement

Data collection was performed on beamline X06SA at the Swiss Light Source (Villigen, Switzerland). For data reduction, the XDS package was used (Kabsch, 2010) and the resulting dataset was analyzed by XDSSTAT (Diederichs, 2006) to check for radiation damage. For RBD:1226 crystals, a dataset with 2.3 Å resolution was obtained and processed in space group I4$_1$ containing one complex in the asymmetric unit. The tetragonal crystals of RBD:NM1230 diffracted to 2.9 Å and were processed in space group P4$_3$2$_1$2 containing two copies of the complex. The structures were solved by molecular replacement using PHASER (McCoy, 2007) and CHAINSAW (Stein, 2008) modified templates of the RBD domain (pdb code: 6Z1Z) and a structure homologue of the nanobody (pdb code: 6XC4). Initial refinement involved simulated annealing as implemented in PHENIX (Adams *et al*, 2010) to reduce model bias. Further refinement was done in a cyclic procedure of reciprocal space refinement as implemented in REFMAC5 (Vagin *et al*, 2004) and real space corrections using COOT (Emsley *et al*, 2010). Several cycles of refinement yielded to a model with good stereochemistry and acceptable R-factors with R_{work}/ R_{free} of 26.7% / 30.7% and 19.0% / 22.2% for RBD:NM1230 and RBD:NM1226 complex, respectively (Appendix Table S2). The structure was validated with MOLPROBITY (Williams *et al*, 2018) prior deposition to the protein data bank (pdb code: 7NKT and 7B27 for RBD:NM1226 and RBD:NM1230, respectively). Figures were generated with PYMOL (SCHRODINGER, L. L. C. The PyMOL molecular graphics system. Version, 2010, 1. Jg., Nr. 5, S. 0.) and structure comparison was performed with DALI (Holm, 2019).

Cell culture

Caco-2 (Human Colorectal adenocarcinoma, ATCC HTB-37) cells were cultured at 37°C with 5% CO_2 in DMEM containing 10% FCS, 2 mM l-glutamine, 100 µg/ml penicillin-streptomycin, and 1% NEAA. The cell line was tested negative for mycoplasma using the

PCR mycoplasma kit Venor GeM Classic (Minerva Biolabs) and the Taq DNA Polymerase (Minerva Biolabs).

Viruses

All experiments associated with the SARS-CoV-2 virus were conducted in Biosafety Level 3 laboratory. The recombinant infectious SARS-CoV-2 clone expressing mNeonGreen (icSARS-CoV-2-mNG) (PMID: 32289263) was obtained from the World Reference Center for Emerging Viruses and Arboviruses (WRCEVA) at the UTMB (University of Texas Medical Branch). To generate icSARS-CoV-2-mNG stocks, 200.000 Caco-2 cells were infected with 50 µl of virus in a 6-well plate; the supernatant was harvested 48 hpi, centrifuged, and stored at $-80°C$. For MOI determination, a titration using serial dilutions of the mNeonGreen (icSARS-CoV-2-mNG) was conducted. The number of infectious virus particles per ml was calculated as the (MOI × cell number)/ (infection volume), where $MOI = -\ln(1 - \text{infection rate})$.

Virus neutralization assay

For neutralization experiments, 1×10^4 Caco-2 cells/well were seeded in 96-well plates the day before infection in media containing 5% FCS. Caco-2 cells were co-incubated with the SARS-CoV-2 strain icSARS-CoV-2-mNG at a MOI = 1.1 and Nbs or serum samples in serial dilutions in the indicated concentrations. 48 hpi cells were fixed with 2% PFA and stained with Hoechst33342 (1 µg/ml final concentration) for 10 min at 37°C. The staining solution was removed and exchanged for PBS. For quantification of infection rates, images were taken with the Cytation3 (Biotek) and Hoechst+ and mNG+ cells were automatically counted by the Gen5 Software (Biotek). Infection rate was determined by dividing the number of infected cells through total cell count per condition. Data were normalized to respective virus-only infection control. Inhibitory concentration 50 (IC_{50}) was calculated as the half-maximal inhibitory dose using 4-parameter nonlinear regression (GraphPad Prism).

Patient samples

112 Serum samples of convalescent SARS-CoV-2-infected and eight uninfected individuals were analyzed in the course of this study. All samples used were de-identified and pre-existing. Ethical consent was granted from the Ethics Commission of the University of Tuebingen under the votum 179/2020/BO2. Samples were classified as SARS-CoV-2 infected, based upon a self-reported positive SARS-CoV-2 RT–PCR result.

Analyses and statistics

Graph preparation and statistical analysis were performed using the GraphPad Prism Software (Version 9.0.0).

Data availability

Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession code 7NKT (RBD:NM1226 complex, www.rcsb.org/structure/unreleased/7NKT) and 7B27 (RBD:NM1230 complex, www.rcsb.org/structure/unreleased/7B27). Protein sequences of all nanobodies are listed in Appendix Table S1. All data that support the findings of this study are available from the corresponding authors upon reasonable request.

Author contributions
Study design: NSM, TRW, MB, UR; Nb selection and biochemical characterization: PDK, BT, TRW; Immunization of the animal: HS, SN, AS; Multiplex binding assay: JH, DJ, MB; HDX-MS experiments: MG, AZ; Organization and providing patient samples: MoS, AN, JSW, KSL; Designing and performing crystallization studies: EO, GZ, TS; Virus neutralization assays: NR, MiS; Data analysis and statistical analysis: TRW, MB, JH, MG, AZ, NR, MiS, UR; Manuscript drafting: TRW, AD, UR; Study supervision: NSM, UR; Manuscript reading: All authors.

Conflict of interest
T.R.W., P.K., N.S.M., and U.R. are named as inventors on a patent application (EP 20 197 031.6) claiming the use of the described Nanobodies for diagnosis and therapeutics filed by the Natural and Medical Sciences Institute. The other authors declare no competing interest.

References

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW *et al* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66: 213–221

Ahmad J, Jiang J, Boyd LF, Natarajan K, Margulies DH (2021) Synthetic nanobody–SARS-CoV-2 receptor-binding domain structures identify distinct epitopes. *bioRxiv* https://doi.org/10.1101/2021.01.27.428466 [PREPRINT]

Amanat F, Stadlbauer D, Strohmeier S, Nguyen THO, Chromikova V, McMahon M, Jiang K, Arunkumar GA, Jurczyszak D, Polanco J *et al* (2020) A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat Med* 26: 1033–1036

Arbabi Ghahroudi M, Desmyter A, Wyns L, Hamers R, Muyldermans S (1997) Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. *FEBS Lett* 414: 521–526

Becker M, Strengert M, Junker D, Kaiser PD, Kerrinnes T, Traenkle B, Dinter H, Häring J, Ghozzi S, Zeck A *et al* (2021) Exploring beyond clinical routine SARS-CoV-2 serology using MultiCoV-Ab to evaluate endemic coronavirus cross-reactivity. *Nat Commun* 12: 1152

Brouwer PJM, Caniels TG, van der Straten K, Snitselaar JL, Aldon Y, Bangaru S, Torres JL, Okba NMA, Claireaux M, Kerster G *et al* (2020) Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* 369: 643–650

Byrnes JR, Zhou XX, Lui I, Elledge SK, Glasgow JE, Lim SA, Loudermilk R, Chiu CY, Wilson MR, Leung KK (2020) A SARS-CoV-2 serological assay to determine the presence of blocking antibodies that compete for human ACE2 binding. *medRxiv* https://doi.org/10.1101/2020.05.27.20114652 [PREPRINT]

Cao Y, Su B, Guo X, Sun W, Deng Y, Bao L, Zhu Q, Zhang X, Zheng Y, Geng C *et al* (2020) Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* 182: 73–84 e16

Chi X, Liu X, Wang C, Zhang X, Li X, Hou J, Ren L, Jin Q, Wang J, Yang W (2020a) Humanized single domain antibodies neutralize SARS-CoV-2 by targeting the spike receptor binding domain. *Nat Commun* 11: 4528

Chi X, Yan R, Zhang J, Zhang G, Zhang Y, Hao M, Zhang Z, Fan P, Dong Y, Yang Y *et al* (2020b) A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* 369: 650–655

Custódio TF, Das H, Sheward DJ, Hanke L, Pazicky S, Pieprzyk J, Sorgenfrei M, Schroer MA, Gruzinov AY, Jeffries CM *et al* (2020) Selection, biophysical and structural analysis of synthetic nanobodies that effectively neutralize SARS-CoV-2. *Nat Commun* 11: 5588

Davies NG, Barnard RC, Jarvis CI, Kucharski AJ, Munday J, Pearson CAB, Russell TW, Tully DC, Abbott S, Gimma A *et al* (2020) Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England. *medRxiv* https://doi.org/10.1101/2020.12.24.20248822 [PREPRINT]

Davies NG, Jarvis CI, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH (2021) Increased hazard of death in community-tested cases of SARS-CoV-2 variant of concern 202012/01. *medRxiv* https://doi.org/10.1101/2021.02.01.21250959 [PREPRINT]

Diederichs K (2006) Some aspects of quantitative analysis and correction of radiation damage. *Acta Crystallogr D Biol Crystallogr* 62: 96–101

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486–501

Esparza TJ, Martin NP, Anderson GP, Goldman ER, Brody DL (2020) High affinity nanobodies block SARS-CoV-2 spike receptor binding domain interaction with human angiotensin converting enzyme. *bioRxiv* https://doi.org/10.1101/2020.07.24.219857 [PREPRINT]

Fink S, Ruoff F, Stahl A, Becker M, Kaiser P, Traenkle B, Junker D, Weise F, Ruetalo N, Horber S *et al* (2021) Multiplexed serum antibody screening platform using virus extracts from endemic coronaviridae and SARS-CoV-2. *ACS Infect Dis* https://doi.org/10.1021/acsinfecdis.0c00725

Gai J, Ma L, Li G, Zhu M, Qiao P, Li X, Zhang H, Zhang Y, Chen Y, Gong R *et al* (2020) A potent neutralizing nanobody against SARS-CoV-2 with inhaled delivery potential. *bioRxiv* https://doi.org/10.1101/2020.08.09.242867 [PREPRINT]

Gorshkov K, Susumu K, Chen J, Xu M, Pradhan M, Zhu W, Hu X, Breger JC, Wolak M, Oh E (2020) Quantum dot-conjugated SARS-CoV-2 spike pseudo-virions enable tracking of angiotensin converting enzyme 2 binding and endocytosis. *ACS Nano* 14: 12234–12247

Hamuro Y, Coales SJ (2018) Optimization of feasibility stage for hydrogen/ deuterium exchange mass spectrometry. *J Am Soc Mass Spectrom* 29: 623–629

Hanke L, Vidakovics Perez L, Sheward DJ, Das H, Schulte T, Moliner-Morro A, Corcoran M, Achour A, Karlsson Hedestam GB, Hällberg BM *et al* (2020) An alpaca nanobody neutralizes SARS-CoV-2 by blocking receptor interaction. *Nat Commun* 11: 4420

Hansen J, Baum A, Pascal KE, Russo V, Giordano S, Wloga E, Fulton BO, Yan Y, Koon K, Patel K *et al* (2020) Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* 369: 1010–1014

Heffron AS, McIlwain SJ, Baker DA, Amjadi MF, Khullar S, Sethi AK, Shelef MA, O'Connor DH, Ong IM (2020) The landscape of antibody binding to SARS-CoV-2. *bioRxiv* https://doi.org/10.1101/2020.10.10.334292 [PREPRINT]

Holm L (2019) Benchmarking fold detection by DaliLite vol 5. *Bioinformatics* 35: 5326–5327

Huo J, Le Bas A, Ruza RR, Duyvesteyn HME, Mikolajek H, Malinauskas T, Tan TK, Rijal P, Dumoux M, Ward PN *et al* (2020) Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2. *Nat Struct Mol Biol* 9: 846–854

Jiang S, Zhang X, Yang Y, Hotez PJ, Du L (2020) Neutralizing antibodies for the treatment of COVID-19. *Nat Biomed Eng* 4: 1134–1139

Ju B, Zhang Qi, Ge J, Wang R, Sun J, Ge X, Yu J, Shan S, Zhou B, Song S *et al* (2020) Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* 584: 115–119

Kabat E, Wu T, Perry H, Gottesman K, Foeller C (1991) *Sequences of proteins of immunological interest (Fifth)*, Washington, DC: US Department of Public Health Service, National Institutes of Health

Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* 66: 133–144

Kirchhofer A, Helma J, Schmidthals K, Frauer C, Cui S, Karcher A, Pellis M, Muyldermans S, Casas-Delucchi CS, Cardoso MC *et al* (2010) Modulation of protein properties in living cells using nanobodies. *Nat Struct Mol Biol* 17: 133–138

Kochert BA, Iacob RE, Wales TE, Makriyannis A, Engen JR (2018) Hydrogen-deuterium exchange mass spectrometry to study protein complexes. *Methods Mol Biol* 1764: 153–171

Koenig PA, Das H, Liu H, Kummerer BM, Gohr FN, Jenster LM, Schiffelers LDJ, Tesfamariam YM, Uchima M, Wuerth JD *et al* (2021) Structure-guided multivalent nanobodies block SARS-CoV-2 infection and suppress mutational escape. *Science* 371: eabe6230

Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Qi, Shi X, Wang Q, Zhang L *et al* (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581: 215–220

Lassaunière R, Frische A, Harboe ZB, Nielsen AC, Fomsgaard A, Krogfelt KA, Jørgensen CS (2020) Evaluation of nine commercial SARS-CoV-2 immunoassays. *medRxiv* https://doi.org/10.1101/2020.04.09.20056325 [PREPRINT]

Lauring AS, Hodcroft EB (2021) Genetic variants of SARS-CoV-2-what do they mean? *JAMA* 325: 529–531

Liu L, Wang P, Nair MS, Yu J, Rapp M, Wang Q, Luo Y, Chan J-W, Sahi V, Figueroa A *et al* (2020) Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* 584: 450–456

McCoy AJ (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* 63: 32–41

Muruato AE, Fontes-Garfias CR, Ren P, Garcia-Blanco MA, Menachery VD, Xie X, Shi PY (2020) A high-throughput neutralizing antibody assay for COVID-19 diagnosis and vaccine evaluation. *Nat Commun* 11: 4059

Muyldermans S (2013) Nanobodies: natural single-domain antibodies. *Annu Rev Biochem* 82: 775–797

Nieto GV, Jara R, Himelreichs J, Salinas C, Pinto T, Cheuquemilla Y, Margolles Y, del Rey NLG, Miranda-Chacon Z, Cuevas A *et al* (2020) Fast isolation of sub-nanomolar affinity alpaca nanobody against the Spike RBD of SARS-CoV-2 by combining bacterial display and a simple single-step density gradient selection. *bioRxiv* https://doi.org/10.1101/2020.06.09.137935 [PREPRINT]

Robbiani DF, Gaebler C, Muecksch F, Lorenzi JCC, Wang Z, Cho A, Agudelo M, Barnes CO, Gazumyan A, Finkin S *et al* (2020) Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* 584: 437–442

Rogers TF, Zhao F, Huang D, Beutler N, Burns A, He W-T, Limbo O, Smith C, Song Ge, Woehl J *et al* (2020) Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science* 369: 956−963

Rothbauer U, Zolghadr K, Tillib S, Nowak D, Schermelleh L, Gahl A, Backmann N, Conrath K, Muyldermans S, Cardoso MC *et al* (2006) Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat Methods* 3: 887−889

Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, Leonhardt H (2008) A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics* 7: 282−289

Roxhed N, Bendes A, Dale M, Mattsson C, Hanke L, Dodig-Crnkovic T, Christian M, Meineke B, Elsasser S, Andrell J *et al* (2020) A translational multiplex serology approach to profile the prevalence of anti-SARS-CoV-2 antibodies in home-sampled blood. *medRxiv* https://doi.org/10.1101/2020.07.01.20143966 [PREPRINT]

Scholer L, Le-Trilling VTK, Eilbrecht M, Mennerich D, Anastasiou OE, Krawczyk A, Herrmann A, Dittmer U, Trilling M (2020) A novel in-cell ELISA assay allows rapid and automated quantification of SARS-CoV-2 to analyze neutralizing antibodies and antiviral compounds. *Front Immunol* 11: 573526

Schoof M, Faust B, Saunders RA, Sangwan S, Rezelj V, Hoppe N, Boone M, Billesbolle CB, Zimanyi M, Deshpande I *et al* (2020) An ultra-high affinity synthetic nanobody blocks SARS-CoV-2 infection by locking Spike into an inactive conformation. *bioRxiv* https://doi.org/10.1101/2020.08.08.238469 [PREPRINT]

Shi R, Shan C, Duan X, Chen Z, Liu P, Song J, Song T, Bi X, Han C, Wu L *et al* (2020) A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* 584: 120−124

Stadlbauer D, Amanat F, Chromikova V, Jiang K, Strohmeier S, Arunkumar GA, Tan J, Bhavsar D, Capuano C, Kirkpatrick E *et al* (2020) SARS-CoV-2 seroconversion in humans: a detailed protocol for a serological assay, antigen production, and test setup. *Curr Protoc Microbiol* 57: e100

Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J Appl Crystallogr* 41: 641−643

Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, Zhou Y, Du L (2020) Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol* 17: 613−620

Tan CW, Chia WN, Qin X, Liu P, Chen M-C, Tiu C, Hu Z, Chen V-W, Young BE, Sia WR *et al* (2020) A SARS-CoV-2 surrogate virus neutralization test based on antibody-mediated blockage of ACE2-spike protein-protein interaction. *Nat Biotechnol* 38: 1073−1078

Tang MS, Case JB, Franks CE, Chen RE, Anderson NW, Henderson JP, Diamond MS, Gronowski AM, Farnsworth CW (2020) Association between SARS-CoV-2 neutralizing antibodies and commercial serological assays. *Clin Chem* 66: 1538−1547

Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay S, San EJ, Msomi N *et al* (2020) Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* https://doi.org/10.1101/2020.12.21.20248640 [PREPRINT]

Tortorici MA, Beltramello M, Lempp FA, Pinto D, Dang HV, Rosen LE, McCallum M, Bowen J, Minola A, Jaconi S *et al* (2020) Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. *Science* 370: 950−957

Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* 60: 2184−2195

Walker SN, Chokkalingam N, Reuschel EL, Purwar M, Xu Z, Gary EN, Kim KY, Helble M, Schultheis K, Walters J *et al* (2020) SARS-CoV-2 assays to detect functional antibody responses that block ACE2 recognition in vaccinated animals and infected patients. *J Clin Microbiol* 58: e01533-20

Walker AS, Vihta K-D, Gethings O, Pritchard E, Jones J, House T, Bell I, Bell JI, Newton JN, Farrar J *et al* (2021) Increased infections, but not viral burden, with a new SARS-CoV-2 variant. *medRxiv* https://doi.org/10.1101/2021.01.13.21249721 [PREPRINT]

Walter JD, Hutter CAJ, Zimmermann I, Wyss M, Egloff P, Sorgenfrei M, Hürlimann LM, Gonda I, Meier G, Remm S *et al* (2020) Sybodies targeting the SARS-CoV-2 receptor-binding domain. *bioRxiv* https://doi.org/10.1101/2020.04.16.045419 [PREPRINT]

Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, Schaefer-Babajew D, Cipolla M, Gaebler C, Lieberman JA *et al* (2021) mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* https://doi.org/10.1038/s41586-021-03324-6

Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB *et al* (2018) MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci* 27: 293−315

Wrapp D, De Vlieger D, Corbett KS, Torres GM, Wang N, Van Breedam W, Roose K, van Schie L, VIB-CMB COVID-19 Response Team, Hoffmann M *et al* (2020a) Structural basis for potent neutralization of betacoronaviruses by single-domain camelid antibodies. *Cell* 181: 1004−1015

Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS (2020b) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367: 1260−1263

Xiang Y, Nambulli S, Xiao Z, Liu H, Sang Z, Duprex WP, Schneidman-Duhovny D, Zhang C, Shi Y (2020) Versatile and multivalent nanobodies efficiently neutralize SARS-CoV-2. *Science* 370: 1479−1484

Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q (2020) Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367: 1444−1448

Yu F, Xiang R, Deng X, Wang L, Yu Z, Tian S, Liang R, Li Y, Ying T, Jiang S (2020) Receptor-binding domain-specific human neutralizing monoclonal antibodies against SARS-CoV and SARS-CoV-2. *Signal Transduct Target Ther* 5: 212

Zhou T, Tsybovsky Y, Gorman J, Rapp M, Cerutti G, Chuang GY, Katsamba PS, Sampson JM, Schon A, Bimela J *et al* (2020) Cryo-EM structures of SARS-CoV-2 spike without and with ACE2 reveal a pH-dependent switch to mediate endosomal positioning of receptor-binding domains. *Cell Host Microbe* 28: 867−879

Zhu Na, Zhang D, Wang W, Li X, Yang Bo, Song J, Zhao X, Huang B, Shi W, Lu R *et al* (2020) A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 382: 727−733

Zohar T, Alter G (2020) Dissecting antibody-mediated protection against SARS-CoV-2. *Nat Rev Immunol* 20: 392−394

Reprogramming Substrate and Catalytic Promiscuity of Tryptophan Prenyltransferases

Elena Ostertag [1], Liujuan Zheng [2], Karina Broger [1], Thilo Stehle [1], Shu-Ming Li [2]* and Georg Zocher [1]*

1 - *Interfaculty Institute of Biochemistry,* University of Tübingen, 72076 Tübingen, Germany
2 - *Institute of Pharmaceutical Biology and Biotechnology,* Fachbereich Pharmacy, University of Marburg, 35037 Marburg, Germany

Correspondence to Shu-Ming Li and Georg Zocher: shuming.li@staff.uni-marburg.de *(S.-M. Li),* georg.zocher@uni-tuebingen.de *(G. Zocher)*
https://doi.org/10.1016/j.jmb.2020.11.025
Edited by Dan Tawfik

Abstract

Prenylation is a process widely prevalent in primary and secondary metabolism, contributing to functionality and chemical diversity in natural systems. Due to their high regio- and chemoselectivities, prenyltransferases are also valuable tools for creation of new compounds by chemoenzymatic synthesis and synthetic biology. Over the last ten years, biochemical and structural investigations shed light on the mechanism and key residues that control the catalytic process, but to date crucial information on how certain prenyltransferases control regioselectivity and chemoselectivity is still lacking. Here, we advance a general understanding of the enzyme family by contributing the first structure of a tryptophan C5-prenyltransferase 5-DMATS. Additinally, the structure of a bacterial tryptophan C6-prenyltransferase 6-DMATS was solved. Analysis and comparison of both substrate-bound complexes led to the identification of key residues for catalysis. Next, site-directed mutagenesis was successfully implemented to not only modify the prenyl donor specificity but also to redirect the prenylation, thereby switching the regioselectivity of 6-DMATS to that of 5-DMATS. The general strategy of structure-guided protein engineering should be applicable to other related prenyltransferases, thus enabling the production of novel prenylated compounds.

Introduction

Prenylation of naturally occurring compounds contributes to chemical diversity[1,2] and enhances the therapeutic potential in drug development and medical treatment.[3–5] For example, notoamides A-D from *Aspergillus* sp. and prenylated flavonoids from *Morus alba* show cytotoxic activity,[6,7] while certain prenylated flavonoids from *Glycyrrhiza lepidota* have antimicrobial properties.[8] It has been shown that prenylation modulates the biological and pharmacological impact of target compounds.[9–11] As an example, cochliodinol and iso-cochliodinol were isolated from the fungus *Chaetomium* sp. and showed cytotoxic activity against L5178Y mouse lymphoma cells. Remarkably, the activity of cochliodinol ($EC_{50} = 7.0$ μg/mL) was ten times higher than for isocochliodinol ($EC_{50} = 71.1$ μg/mL), although the two compounds differ only in the prenylation position of the indole ring.[12]

Recently, a cell-free platform for the prenylation of cannabinoids that uses an enzymatic toolkit for the production of cannabinoid precursors have been established, demonstrating the potential of chemoenzymatic synthesis.[13] Prenyltransferases

play a crucial role in this process because of their regio- and chemoselective catalysis that is hardly achievable by conventional organic synthesis. The chemical prenylation of aromatic ring systems is often restricted to specific ring atoms. For example, tryptophan can be chemically prenylated predominantly at positions N1, C2, and C3 of the indole ring.[14,15] In contrast, enzymatic reactions catalysed by a set of prenyltransferases enable prenylations at additional positions of the aromatic ring system.[16] Nevertheless, an increased diversity of available compounds would be desirable to broaden the product spectrum and obtain a general application toolkit.[17] A crucial step in this progress lies in a better understanding of how prenyltransferases act and how they can be modulated.

Most of the known prenylated compounds originate from fungi, bacteria and plants, and represent a large group of secondary metabolites. They fulfil various functions in the natural environment like communication with and adaptation to the environment, as well as defence against other organisms.[18,19] The biosynthesis of those natural compounds is a multistep process. A class of prenyltransferases, the so-called ABBA prenyltransferases (ABBA-PT's),[20] is involved in the formation of a large number of these compounds. These enzymes represent a subgroup of aromatic prenyltransferases, catalysing the prenyl transfer of an activated prenyl donor such as dimethylallyl diphosphate (DMAPP) or geranyl diphosphate (GPP) to an aromatic acceptor, e.g. tryptophan.[21] Each enzyme is specific with regard to its substrates and possesses its own regio- (location of the prenylation) and chemoselectivity (type of the C-C bond, secondary or quaternary), rendering these enzymes particularly interesting for synthetic chemistry. The prenylation process itself follows a Friedel-Crafts reaction. The prenyl moiety ($n \times C_5$) of an activated prenyl donor such as DMAPP is cleaved by its stabilising active site environment to produce an intermediate carbocation, followed by the electrophilic attack onto the aromatic acceptor.[22] Due to the resonance stabilisation of the charged intermediate, the attack might be initiated by the C1 atom (regular prenylation) or the C3 atom (reverse prenylation), which establishes the chemoselectivity of the reaction[23] (Figure 1). For example, FgaPT2, which is part of the ergot alkaloid biosynthesis in *Aspergillus fumigatus*, catalyses the transfer of a dimethylallyl moiety to the C4 position of the indole ring in a regular fashion.[24] In contrast, FgaPT1 of the same organism prenylates the natural substrate fumigaclavine A in a reverse manner onto the C2 position of the indole ring.[25] The ABBA-PTs can be subdivided into two clusters based on their substrate discrimination and sequence identity.[26] The CloQ/NphB family is able to transfer different sized prenyl donors onto the aromatic ring system.[27] The dimethylallyltryptophan synthases (DMATS) mainly use DMAPP as a prenyl source. A deeper understanding of the regioselectivity, as well as the determinants for the substrate specificities of different prenyltransferases, is important for engineering these enzymes in order to utilise them in biotechnological drug production. Knowing the precise enzymatic mechanism can enable the design of enzymes that carry out a desired reaction in a highly specific manner. Successful structure-based enzyme engineering of this enzyme class has been carried out for several ABBA prenyltransferases.[28–30] Over the years, several DMATS structures were solved. Structural data for regular prenylation of tryptophan and derivatives at position C2 (FtmPT1), C4 (FgaPT2), C6 (PriB), as well as for reverse prenylation at position C3 (AnaPT), C7 (MpnD) and N1 (CymD) are available (Figure 1)[31,28,24,32–39]. To date, a C5 prenylating DMATS has been biochemical characterised, but structural information about this class of enzymes has been lacking.[40]

Two bacterial DMATSs, 5-DMATS from *Streptomyces coelicolor* and 6-DMATS from *Micromonospora olivasterospora*, have been characterised biochemically with purified recombinant proteins.[41,42] Both enzymes catalyse a regiospecific tryptophan prenylation at positions C5 and C6, respectively. But to date, the secondary metabolites, in which they are part of the synthesis, are unknown. It has been reported that 5-DMATS from *S. colicoelor* also accepts unnatural alkyl diphosphates such as methylallyl, 2-pentenyl and benzyl diphosphate as donors, but with significantly lower regioselectivity than with its natural substrate DMAPP.[41] In these cases, C5, C6 and C7 served as attachment positions of the alkyl or benzyl moiety. 6-DMATS can also use GPP for prenylation, but with a significantly lower product yield than with DMAPP (about 10% compared to that of DMAPP).[42]

In this study, the crystal structure of this 5-DMATS is presented, as the first representative of a C5-prenylating DMATS. Furthermore, the structure of 6-DMATS from *Micromonospora olivasterospora* was characterised and compared to 5-DMATS to identify the differences crucial for the regioselectivities in both reactions (Figure 1). Based on the structural data, catalytic bases for each reaction mechanism were determined and mutants were produced to validate our findings. Finally, we set out to modify the regioselectivity of both enzymes to convert a C6 prenylating enzyme to a C5 specific variant and vice versa. In addition, a 6-DMATS variant with a substrate preference towards GPP instead of DMAPP was produced.

Results

Structure determination of 5-DMATS and 6-DMATS

The recombinantly produced proteins were purified using a two-step protocol employing Ni-
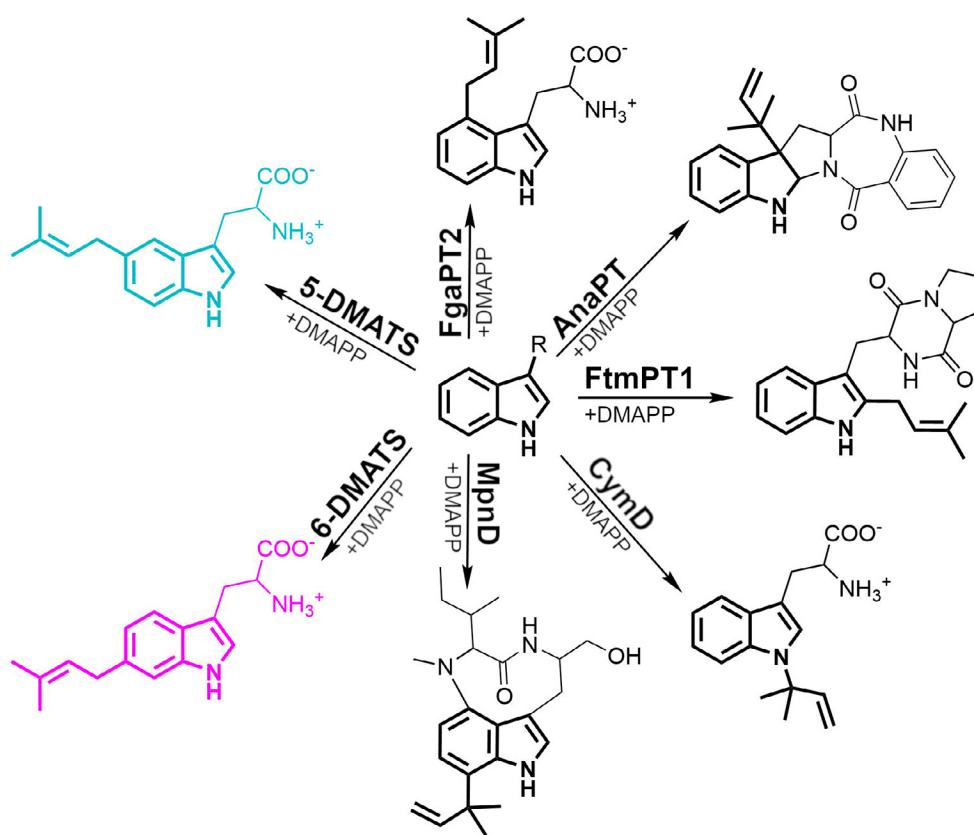
Figure 1. Overview of enzymatic prenylation reactions at the indole ring. The prenylations at each position of the indole ring are plotted together with the corresponding prenyltransferase and its prenylation product. CymD from *Salinispora arenicola*, AnaPT from *Neosartorya fischeri*, and TleC from *Streptomyces blastmyceticus* prenylate the indole ring in a reverse fashion,[56,34] FtmPT1 and FgaPT2 from *Aspergillus fumigatus* catalyse the reaction in a regular manner.[31,24] The prenylation product of 5-DMATS from *Streptomyces coelicolor* and 6-DMATS from *Micromonospora olivasterospora*, which are discussed in this paper, are represented in cyan and pink, respectively.

NTA affinity chromatography followed by size exclusion chromatography (SEC). Analytical SEC established an apparent molecular weight of 30 ± 8 kDa and 32 ± 8 kDa for 5-DMATS and 6-DMATS, respectively, indicating the presence of monomeric molecules in solution (Figure S7). The crystal structures were solved by molecular replacement using the homologous structure of PriB[33] as search model. Although L-Trp was not included in any of the buffers used for protein preparation, crystallisation or crystal handling, the structure of 6-DMATS contained substantial positive electron density in the active site that was unequivocally determined to be L-Trp on the basis of the diffraction data with a resolution of 1.65 Å (Figure S1). Therefore, the structure is referred to as 6-DMATS^{trp}. In contrast, the structure of the unliganded 5-DMATS, which was solved at a resolution of 1.5 Å, did not show any molecules bound to the active site and is referred to as 5-DMATS^{apo}. Ternary complexes containing L-Trp and DMASPP (a less-reactive homologue for the natural substrate DMAPP) were obtained at 1.7 Å resolution for 6-DMATS (6-DMATS^{ter}) and 5-

DMATS (5-DMATS^{ter}). All structures were refined to good model quality and provide a solid basis to understand ligand binding and catalysis. Data collection and refinement statistics are summarised in the supporting information (Table S1).

Overall 3D structure

As expected, both 5-DMATS^{apo} and 6-DMATS^{trp} fold into single domains that display the characteristics of the so-called ABBA barrel fold.[20] A structurally well-conserved set of ten antiparallel β-strands and ten α-helices forms a repetitive pattern of αββα secondary structure elements (Figure S2). The ten β-strands build a central β-barrel that houses the solvent-filled reaction chamber. This barrel is surrounded by ten solvent exposed α-helices. The barrel is open on both sites, top and bottom, to allow for substrate binding at the prenyl acceptor and donor sites. An in-depth structural comparison of 5-DMATS^{apo} and 6-DMATS^{trp} using DALI[43] revealed highest similarity to other members of the ABBA-PT family, and yielded nine hits with Cα-rmsd values below 3.5 Å

(Tables S3 and S4). The closest structural relative of both 5-DMATS^{apo} and 6-DMATS^{trp} is another bacterial 6-DMATS from *Streptomyces* sp. RM-5-8. PriB (Cα-rmsd: 1.4 Å and 1.8 Å for 5-DMATS^{apo} and 6-DMATS^{trp}, respectively), although its sequence identity to 5-DMATS and 6-DMATS is 62% and 42%, respectively. In general, 5-DMATS and 6-DMATS are structurally more closely related to other members of the DMATS family (Cα-rmsd < 3.5 Å), for example TleC from *Streptomyces blastmyceticus* and FgaPT2 from *Aspergillus fumigatus*, than to enzymes of the CloQ/NphB family (Cα-rmsd > 3.5 Å). A Cα-rmsd comparison of 5-DMATS^{apo} and 6-DMATS^{trp} to the closest structural homologues (Cα-rmsd < 3.5 Å) resulted in a Cα-rmsd significant smaller for the β-barrel region than for the flanking α-helices and the loop regions of the structure (Figure S3). This finding agrees with the observation that the location of the active centre is inside the rigid β-barrel and the analysed enzymes catalyse similar reactions, and hence use similar substrates.

The binding of the prenyl donor

The electron density clearly defines the location and orientation of the substrates L-Trp and DMASPP inside the barrel for both ternary complexes at the expected active site position for prenyl donor and acceptor binding. It is noteworthy that the electron density is well defined even for the dimethylallyl entity in both structures, allowing to unequivocally place the prenyl donor (Figure 2). To bind DMASPP, charge-charge interactions are formed to compensate for the negative charge of the diphosphate group. Additionally, hydrogen bonds are formed by tyrosines as well as by coordinated water molecules. A schematic mode of binding is depicted in Figure 3.

Although DMASPP is very similarly placed in the active site of both enzymes, an obvious difference of DMASPP binding is a shift of about 1 Å (coordinate error: 0.23 Å) in 6-DMATS, pushing the prenyl donor slightly deeper inside the cavity (Figure 4). This is mainly caused by R209 and a reorientation of the loop attached to this residue during prenyl donor binding (Figure S6). For both enzymes, it is noteworthy that the dimethylallyl entity is positioned almost perfectly below the benzene ring of tryptophan, whereas the heterocyclic ring moiety of the aromatic substrate is less accessible for an electrophilic attack, in agreement with the observation that neither N1, C2, nor C3 prenylation was observed for both enzymes (Figure S4). For 6-DMATS, the prenyl moiety of the DMASPP is coordinated by stacking interactions toward the indole ring of L-Trp, with a distance of 3.8 Å between the C1 atom of the dimethylallyl moiety and C6 of the indole ring. This is the shortest distance of a C-atom of the prenyl donor towards its corresponding prenyl

acceptor, and it is comparable to the distances reported in FgaPT2 between C4 of the indole ring and C1 of the dimethylallyl group and PriB between C6 of the indole ring and C1 of the dimethylallyl with 3.8 Å and 3.6 Å, respectively. The DMASPP prenyl moiety in 5-DMATS forms hydrophobic interactions with L81 and Y151. In contrast to 6-DMATS, the distance of the C1 atom of the dimethylallyl moiety towards the C5 of the indole ring is unusually long (5.3 Å), which makes a direct regular prenylation at this position unlikely. Distances of the C1 and C3 of the dimethylallyl moiety to the C7 position of the indole ring with 4.2 Å and 3.7 Å, as well as 4.4 Å and 4.1 Å to the C6 position of the acceptor, respectively, support an indirect prenylation (Figure S4). We speculate that the product formation happens via a reverse prenylated intermediate at position C6 or C7 of the indole ring (Figure S11).

Common features of prenyl acceptor binding

Both enzymes bind L-Trp in a virtually identical conformation, which is similar to the binding mode described for PriB. This binding mode was first identified in TleC[37] and is distinct from what was observed for FgaPT2 and other members of the fungal prenyltransferases. In contrast to FgaPT2, substrate binding is characterised by an almost 90° rotation of both sp^2-hybridised substrate planes. L-Trp is held in place by hydrogen bonds of the amino groups to backbone oxygens, as well as to coordinated waters. In addition, the carboxyl group is bound via salt bridges and hydrogen bonds. The indole ring is held in place via ionic interactions of N1 towards strictly conserved glutamate residues and by hydrophobic interactions, as well as π-stacking of the indole ring (Figures 2 and 3). Highly conserved residues are present for the pyrophosphate as well as for the tyrosine shield to protect the carbocation during reaction (Figures 2 and 3). Residues of the L-Trp binding site are less conserved between members of the DMATS family. This fits well with the diverse prenyl acceptor preferences of these enzymes.

Structural rearrangements during ligand binding

The alignment of the unliganded and liganded structures of 5-DMATS (5-DMATS^{apo} and 5-DMATS^{ter}) and 6-DMATS (6-DMATS^{trp} and 6-DMATS^{ter}) shows they are virtual identical with an overall Cα-rmsd values of 0.4 Å and 0.8 Å, respectively (Figure S5 and S6). In the active site of 5-DMATS, arginine 252 and 278, as well as the backbone oxygen F56 are reoriented during L-Trp binding. In addition, the side chains of two arginines (R202, R338) and two lysines (K149, K204) are reorganised to facilitate the interaction with the DMASPP. For 6-DMATS^{ter}, differences in
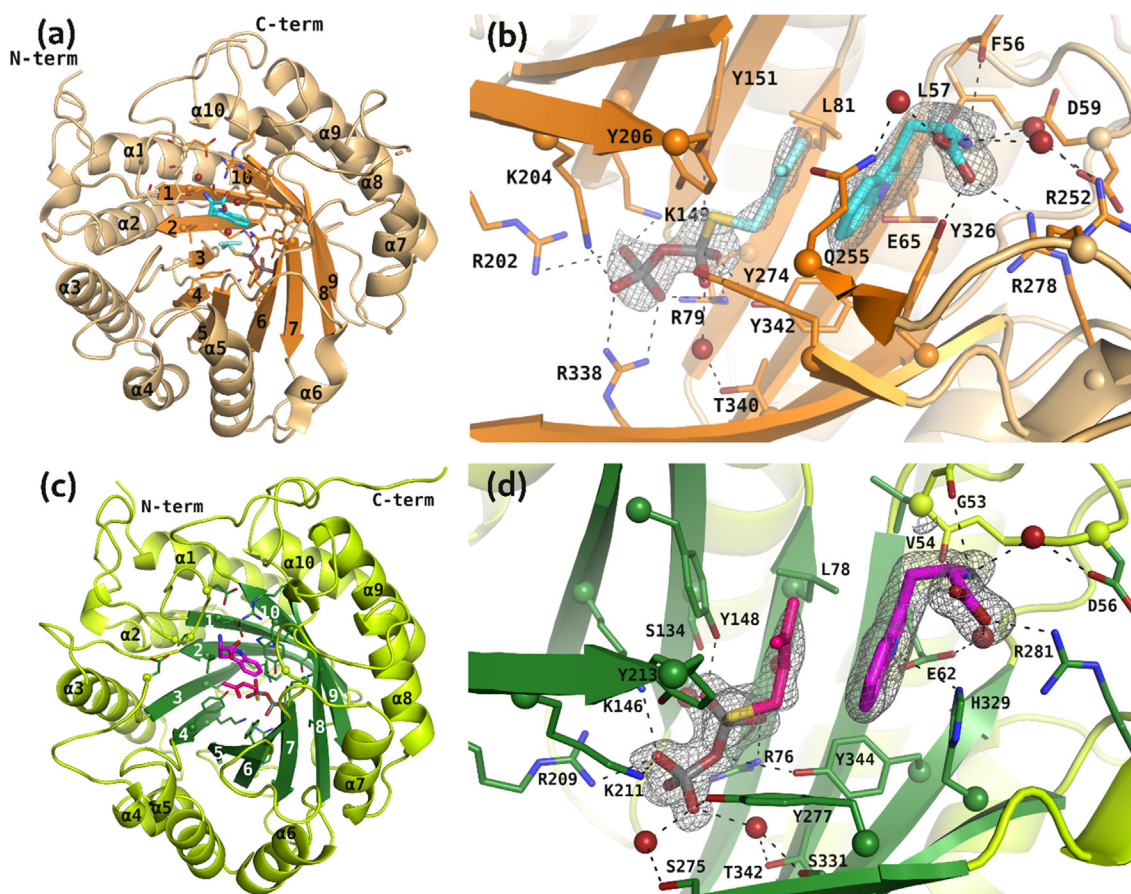
Figure 2. Structures of 5-DMATS and 6-DMATS. The overall fold of the substrate bound structures of 5-DMATS (a) and 6-DMATS (c) are depicted in cartoon representation, with α-helices, and loops coloured in light orange and light green, while the β-strands are dyed in dark orange and dark green for 5-DMATS and 6-DMATS, respectively. The ligands DMASPP and L-Trp (light blue in 5-DMATS and pink in 6-DMATS) and the interacting residues are shown as stick representation (orange and dark green for 5-DMATS and 6-DMATS, respectively). The close-up views of the active sites of 5-DMATS (b) and 6-DMATS (d), show a simulated-annealing (F_o-F_c)-difference omit map (grey mesh) for the ligands at a contour level of 3 σ. Hydrogen bonds and ionic interactions between ligands and interacting residues are depicted with dotted black lines. Water molecules that interact with the ligands are represented as red spheres.

backbone conformation can be tracked down to five loop regions (residue 67-73, 137-141, 200-211, 263-273, 333-340), which are also partly involved in crystal contacts and are close to the DMASPP binding site. These rearrangements enable R209 to enter the active site and are accompanied with the adjustment of K146 and K211 to bind the pyrophosphate of DMASPP. In addition, four tyrosines (Y344, Y148, Y277, Y213) and L78 are slightly reoriented to fit in the prenyl donor. Both L-Trp binding sites of 6-DMATS[trp] and 6-DMATS[ter] provide an identically bound L-Trp, hence no rearrangement of the interacting residues could be observed.

Comparison of the 6-DMATS structure with PriB

A comparison of 6-DMATS with PriB, the only other known 6-DMATS structure to date, shows that while the active sites of both ternary complexes are highly similar, there are three major differences. Residues V259, Y277 and H329 of 6-DMATS are replaced by L293, H312 and Y364 in PriB. All three residues are located around the benzene ring of L-Trp (Figure 4). The side chains of the aromatic residues are positioned in a manner that one aromatic system is able to perform π-stacking with the prenyl acceptor, whereas the second aromatic residue is orientated almost perfectly perpendicular to the indole ring system. The amino acid exchange of the tyrosine-histidine pair is particularly interesting as H312 of PriB is the only possible residue positioned in a way that is suitable to act as catalytic base in that enzyme.[33] Based on these observations and the close proximity of H329 to C6 of L-Trp, we hypothesise that H329 is the catalytic base of 6-DMATS and hence controls the regioselectivity of the reaction by product
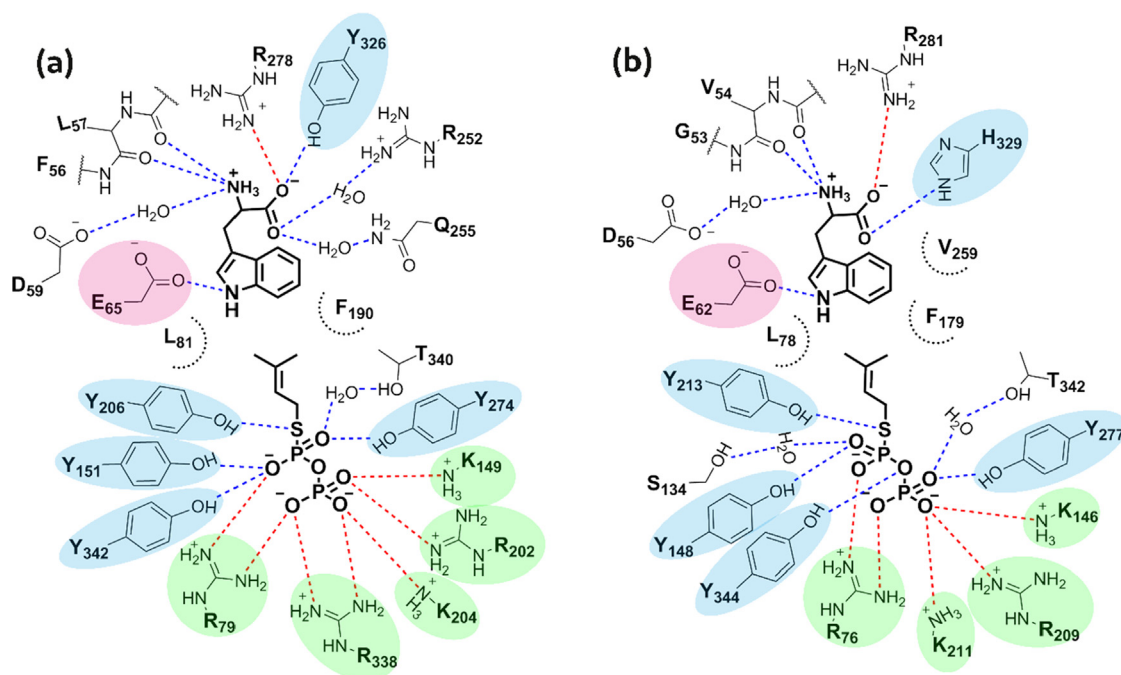
Figure 3. Interactions of the DMASPP and L-Trp in the active sites of 5-DMATS (a) and 6-DMATS (b). Hydrogen bonds and charge-charge interactions are shown as blue and red dotted lines, respectively. Strictly conserved glutamates are highlighted in red. Conserved residues of the diphosphate binding site, as well as the tyrosine shield are highlighted in blue and green, respectively. R333 in 6-DMATS is the residue corresponding to R338 in 5-DMATS, but forms no interactions with the ligand.

development control. The intermediate is partially stabilised by the location of the base that abstracts the hydrogen to regain the aromaticity of the indole ring. The mutation of V259 (6-DMATS) to L293 (PriB) in both proteins is accompanied with a subtle shift of the prenyl acceptor by steric repulsion of the larger leucine side chain. Although both enzymes catalyse the same reaction, they differ from each other in the active sites. It is plausible that the positions exchanged from H312/Y364 in PriB to Y277/H329 in 6-DMATS are a result of evolution to adjust for the prenyl acceptor shift maintaining the regioselectivity by controlling the position of the base. To test this hypothesis, the side chain properties were investigated by site-directed mutagenesis. To verify the function of H329 as the crucial base for the reaction mechanism, a knock down mutant was produced by replacing H329 with tyrosine. Tyrosine would still be able to bind L-Trp by π-stacking but cannot deprotonate the Wheland intermediate in the course of the reaction. In addition, a double mutant that partially mimics the active site of PriB was established by mutating H329 to tyrosine and Y277 to histidine. As expected, the 6-DMATS$_{H329Y}$ variant lost almost completely its catalytic activity, with a total product yield of only 7% compared to the wild type protein, clearly indicating that H329 serves as the catalytic base. The enzymatic turnover is slightly improved in variant 6-DMATS$_{H329Y,Y277H}$ with a total product yield of

16% and 38% of the catalytic efficiency, compared to the 6-DMATS wild type (Figure 5, Table 2).

Comparison of the 5-DMATS structure with that of 6-DMATS

5-DMATS and 6-DMATS bind L-Trp very similarly, with a C_α-rms deviation of 0.4 Å for the indole ring when the enzymes are superposed onto each other (Figure 4). An in-depth comparison of the active site of these enzymes revealed two major changes. V259 and the catalytic base H329 of 6-DMATS are replaced by residues Q255 and Y326 in 5-DMATS. Taking into account that the dimethylallyl entity possesses larger distances to the benzene ring atoms in 5-DMATS as discussed above, two major questions arose. Firstly, how is the regioselectivity controlled by 5-DMATS and secondly, what is the catalytic base of 5-DMATS as an obvious candidate is missing in the active site. At first glance, no obvious reason could be determined why prenylation takes place at position C5, neither because of steric restrictions nor because of other directing groups in the active site. As both questions are related, we set out to perform the most obvious replacements in the active site by mutating amino acids Y326 and Q255. We hypothesised through structural comparison that proton abstraction, to regain
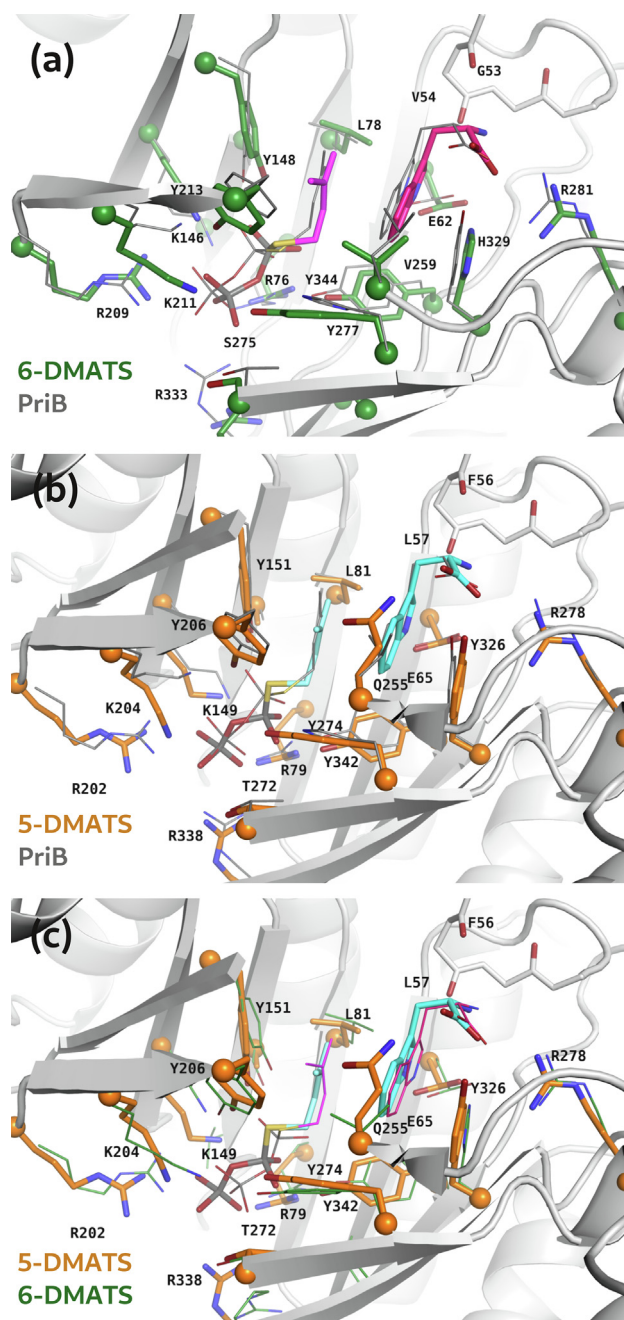
Figure 4. Active site Cα-rmsd alignment of PriB, 5-DMATS and 6-DMATS structures. Close-up views are represented for PriB aligned to (a) 6-DMATS and (b) 5-DMATS, as well as the alignment of (c) 5-DMATS to 6-DMATS. Ligands and residues are represented in lines and sticks, for PriB (grey) and 5-DMATS (ligands in light blue, residues in orange), respectively. 6-DMATS ligands (pink) and interacting residues (dark green) are represented as sticks in panel (a) and as lines for panel (c). The residue labels correspond to the residues of 6-DMATS (a) and 5-DMATS (b-c), respectively.

aromaticity for product formation, is caused by a water molecule either attached to the side chain of Q255 or Y326. The hydroxyl group of Y326 could be excluded as crucial for catalysis, as mutation Y326F has no substantial effect on enzyme kinetics (Table 1) and product yield (Figure 5). In contrast, replacing Q255 with a hydrophobic site chain (Q255V) or shorter polar side chain (Q255N) decreased the catalytic efficiency three-fold, which is caused by a nearly ten-fold reduced catalytic turnover (Table 1). Both mutations are not suitable to establish hydrogen bond interactions thereby reducing the affinity on attaching a water molecule at that position. Polar substitutions of a similar size, such as Q255E or Q255H reduced the catalytic turnover to 63% and

Table 1 Conversion of L-Trp by 5-DMATS and mutants in the presence of DMAPP or GPP. The corresponding kinetic curves are represented in Figure S9. **2** = 5-DMA-L-Trp, **3** = 6-DMA-L-Trp, **4** = 5,6-di-DMA-L-Trp, **5** = 5,7-di-DMA-L-Trp, **7** = 5-Granyl-L-Trp

Product yields with DMAPP				Kinetic parameters with DMAPP		
	2	3	4 and 5	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})
5-DMATS$_{WT}$	82 ± 3%	–	8.0 ± 1.5%	0.46 ± 0.09	0.44 ± 0.02	960
5-DMATS$_{L81A}$	51 ± 4%	–	–	–	–	–
5-DMATS$_{Q255V}$	20.8 ± 1.1%	–	–	0.04 ± 0.01	0.007 ± 0.001	175
5-DMATS$_{Q255N}$	21.1 ± 0.5%	–	35.1 ± 1.6%	0.15 ± 0.01	0.05 ± 0.01	330
5-DMATS$_{Q255E}$	51.7 ± 0.1%	–	18.1 ± 1.3%	0.57 ± 0.04	0.28 ± 0.01	490
5-DMATS$_{Q255H}$	55.1 ± 0.3%	–	17.3 ± 0.1%	0.54 ± 0.06	0.26 ± 0.01	480
5-DMATS$_{Y326F}$	75.4 ± 1.9%	–	5.7 ± 0.3%	0.53 ± 0.03	0.49 ± 0.01	810
5-DMATS$_{Y326H}$	50.9 ± 0.1 %	–	–	0.14 ± 0.02	0.20 ± 0.01	1430
5-DMATS$_{Q255V, Y326H}$	28.8 ± 0.3%	11.6 ± 0.6%	27 ± 3%	0.74 ± 0.07	0.29 ± 0.01	390
5-DMATS$_{Q255N, Y326H}$	26.4 ± 1.1%	8.1 ± 0.6%	33 ± 8%	0.14 ± 0.02	0.08 ± 0.01	570
Product yields with GPP				Kinetic parameters with GPP		
	7					
5-DMATS$_{L81A}$	28.1 ± 1.3 %		0.80 ± 0.14		0.04 ± 0.01	50
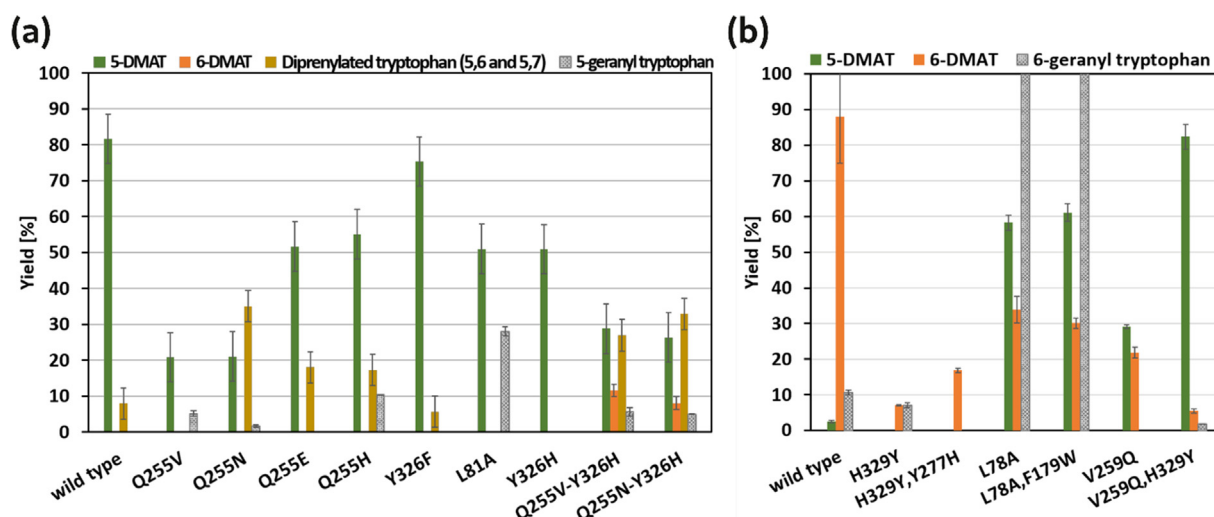
Figure 5. Product yields of the conversion of 5-DMATS (a) and 6-DMATS (b) variants. Original data are represented in Figure S7 and Table 1.

59%, respectively, resulting in a catalytic efficacy of 50% compared to the wild type level (Figure 5, Table 1). These results indicate that residue Q255 is important for catalysis but not essential.

Modifying both enzymes

Having identified the key residues for catalysis of 5-DMATS and 6-DMATS, a structure comparison of the two enzymes with PriB, FgaPT2, FtmPT1, AnaPT and TleC was performed. The aim was to adjust the catalytic properties of both 5-DMATS and 6-DMATS through targeted engineering of the binding sites. Our experiments had two aims. Our first intention was to alter the prenyl donor specificity to achieve an acceptance for larger compounds such as GPP. In a second and more

challenging approach, the regioselectivities of the two enzymes should be changed, i.e. to switch a C5-prenylating enzyme to a C6-prenylating enzyme and vice versa.

In contrast to 5-DMATS, 6-DMATS is capable of accepting GPP as prenyl donor, but with low product yield (Figure 5, Tables 1 and 2). Recently, gatekeeper residues were identified in the enzymes TleC and MpnD.[37] These residues control the size of the prenyl donor that can be used for prenylation. A similar gatekeeper residue was identified in FgaPT2[44] and showed that replacement of M328 with smaller amino acids enhances the catalysis for molecules such as GPP. Based on our structure analysis, L81 and L78 were identified as potential gatekeeper residue for 5-DMATS and 6-DMATS, respectively. In

Table 2 Conversion of L-Trp by 6-DMATS and mutants in the presence of DMAPP or GPP. The corresponding kinetic curves are represented in Figure S10. **2** = 5-DMA-L-Trp, **3** = 6-DMA-L-Trp, **6** = 6-Granyl-L-Trp

	Product yields with DMAPP		Kinetic parameters with DMAPP		
	2	3	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})
6-DMATS$_{WT}$	2.5 ± 0.4%	88 ± 13%	0.12 ± 0.01	0.19 ± 0.01	1580
6-DMATS$_{L78A}$	58 ± 2%	34 ± 4%	0.051 ± 0.007	0.10 ± 0.01	1960
6-DMATS$_{V259Q}$			0.29 ± 0.03	0.41 ± 0.01	1410
6-DMATS$_{H329Y}$	–	7.1 ± 0.2%	–	–	–
6-DMATS$_{H329Y,Y277H}$	–	16.9 ± 0.6%	0.012 ± 0.003	0.007 ± 0.001	600
6-DMATS$_{L78A,F197W}$	61 ± 2%	30.1 ± 1.4%	0.061 ± 0.008	0.09 ± 0.01	1475
6-DMATS$_{V259Q,H329Y}$	82 ± 3%	5.5 ± 0.6%	0.13 ± 0.01	0.18 ± 0.02	1380
	Product yields with GPP		Kinetic parameters with GPP		
	6				
6-DMATS$_{L78A}$	100%		0.047 ± 0.008	0.13 ± 0.01	2760
6-DMATS$_{L78A,F197W}$	100%		0.043 ± 0.005	0.12 ± 0.01	2790

addition to L81/L78, the prenyl donor is flanked from all sites by rigid aromatic systems to envelope the carbocation. L81/L78 were replaced with a small alanine residue and the product yield of the GPP conversion was analysed. For 6-DMATS$_{L78A}$ a substantial enhancement in GPP acceptance could be achieved, increasing the product yield from 10% to 100% (Figure 5, Table 2, Figure S16) with a comparable catalytic efficiency like the wild type with its natural substrate DMAPP. Finally, we were able to transform a GPP inactive enzyme into a GPP accepting variant by the creation of 5-DMATS$_{L81A}$, although with low catalytic efficiency compared to the wild type reaction (Table 1, Figure S15).

Next, we set out to modify the sequence specificity of 5-DMATS with the aim to redirect the prenylation to position C6. Wild type 5-DMATS is strictly selective for prenylation at position C5. In a first attempt, Y326 was exchanged to histidine to mimic the catalytic base H329 of 6-DMATS, but this did not change the regioselectivity of 5-DMATS. However, the introduced charge resulted in a three-fold decreased K_M value, which enhanced the catalytic efficiency of the enzyme (Table 1). The second attempt aimed to abolish the C5-selectivity by replacing Q255 with a valine or asparagine. Hence, double mutants 5-DMATS$_{Q255V,Y326H}$ and 5-DMATS$_{Q255N,Y326H}$ were produced that successfully introduced a C6-prenylation, although the selectivity could not be switched. Both enzymes produce C5- and C6-prenylated L-Trp in a ratio of 5:2, but with different catalytic efficiency. While the 5-DMATS$_{Q255V,Y326H}$ possesses a three-fold increased K_M value but gained only 18% of the catalytic turnover, 5-DMATS$_{Q255N,Y326H}$ had a decreased K_M value and turnover number, compared to the wild type protein. Interestingly, mutation of Q255 increased the ability for diprenylation reactions at C5 and C6 as well as C5 and C7 (Figure 5, Table 1, Figure S14).

To change the regioselectivity of 6-DMATS toward the C5 position, we initially followed the idea of steric restraints and aimed for a slight reorientation of the indole ring in the reaction chamber. To achieve this, the active site was enlarged by reducing the gatekeeper residue L78 at one site and shrinking the chamber at the opposite position (F197). Variant 6-DMATS$_{L78A,F197W}$ indeed switched the specificity of the 6-DMATS to a very promiscuous 5-DMATS possessing a C5- to C6- prenylation ratio of 2:1 with a catalytic efficiency comparable to the wild type enzyme. Surprisingly, the 6-DMATS$_{L78A}$ also possesses the ability to produce 5-dimethylallyl tryptophan but with a lower specificity towards the C5. It should be mentioned that the regioselectivity for both mutants could not be switched for GPP as exclusively C6-prenylation is observed for that substrate. Obviously, the regioselectivity depends also on the size of the prenyl donor and the enzyme adjusts the conformation on the basis of the available space.

As the specificity obtained for 6-DMATS$_{L78A,F197W}$ was not satisfying, the strategy was changed away from directed steric pressure and aimed to replace the position of the residue important for catalysis. The catalytic residue should stabilise the sp^3-hybridised intermediate and redirection of this residue should change the regioselectivity by product development control. Therefore, point mutation V259Q was introduced, as this position is critical for the regioselectivitiy in 5-DMATS. The turnover number doubles through this replacement and the resulting product was a 1:1 mixture of C5 and C6 prenylated L-Trp. For higher regioselectivity of the reaction, a 6-DMATS$_{V259Q,H329Y}$ mutant was generated to abolish the C6-prenylation activity. This variant showed the power of structure-guided protein design, as we were able to establish a 6-DMATS variant with C5 regioselectivity. For 6-DMATS$_{V259Q,H329Y}$ a C6 to C5 ratio of approx. 1:15 without a significant decrease in the catalytic

efficiency was observed (Figure 5, Table 2, Figures S12 and S13).

Discussion

The structural data suggested that the two DMATSs analysed here operate with different catalytic mechanisms. Analysis of the ternary complex of 6-DMATS revealed that the enzyme follows a classical Friedel-Crafts-Alkylation (Figure 6), as also reported for the 4-DMATS FgaPT2.[32] The nascent carbocation attaches the prenyl acceptor at position C6 and aromaticity is then re-established by the proton abstraction from the catalytic base H329. The regioselectivity in the enzyme is achieved by the positioning of the substrates but also by the location of the base H329 that reduces the activation energy for a C6 prenylation.

In contrast, the architecture of the active site of 5-DMATS differs from that of 6-DMATS. The distances of the reacting atoms in 5-DMATS are too far away for a direct attack on C1 of DMASPP to establish a regular prenylated 5-dimethylallyltryptophan (Figure S4), in contrast the C3 of DMASPP is in reasonable distance with 4.1 Å and 3.7 Å toward the C6 and C7 of the indole ring, respectively. We hypothesise that the initial reaction step involves a reverse prenylation either at C6 or C7 of L-Trp (Figure 6), followed by one or more rearrangement steps (Figure S11). A similar rearrangement reaction is proposed for FtmPT1.[14] This enzyme lacks a catalytic base in the active site, which would facilitate the deprotonation at a particular site. Q255 was experimentally determined to play an important role for the catalytic activity of 5-DMATS properly by binding a water molecule that we propose as catalytic base. This water molecule is additionally basified by the interaction with the carboxyl group of the substrate L-Trp. Furthermore, a direct interaction of the proton in the sp^3-hybridised intermediate of L-Trp with a surrounding water cannot be ruled out.

For directly reacting PTs like FgaPT2[32] or the 6-DMATS enzymes[33] a catalytic base could be assigned and experimentally determined. The importance of the catalytic base for regioselectivity was experimentally confirmed for FgaPT2 by site-directed mutagenesis. The exchange of K174 in that enzyme resulted in a predominantly reverse C3 prenylated L-Trp.[45] This study shows that the position of the catalytic base (H329) is more important for the regioselectivity than the steric restraints introduced by site-directed mutagenesis. This is emphasised by our regioselectivity switch variants of 6-DMATS. Introducing a residue that is important for C5 prenylation yielded a 1:1 ratio of C5 to C6 prenlyated products. This could be further improved by removal of the former catalytic base, which established a redirection of catalysis towards the C5 position. By applying of this concept, the 6-DMATS could be transformed to a regular C5-prenylating enzyme.

In summary, the first crystal structures of 5-DMATS in the ligand-free and substrate bound
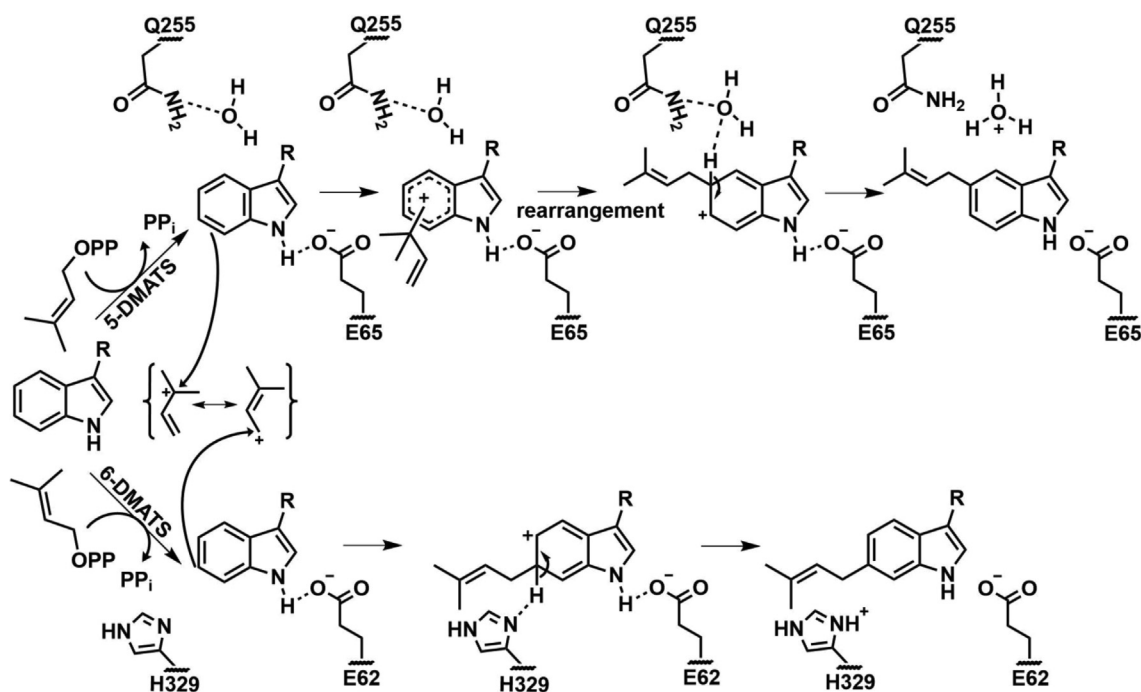
Figure 6. Proposed reaction mechanisms of 5-DMATS (top) and 6-DMATS (bottom). The rearrangement intermediates of 5-DMATS are not listed due to the lack of experimental data. A water molecule was proposed as catalytic base in the reaction cycle of 5-DMATS. Crucial residues are labelled in both cases.

form were determined and compared to the structures of 6-DMATS. Based on the ternary complex structures, key residues were identified and confirmed the gatekeeper postulate[37] by creation of 5-DMATS$_{L81A}$ and 6-DMATS$_{L78A}$ to allow for much higher GPP turnover compared to the wild type enzymes. The regioselectivity of both enzymes was modified, although with less success for 5-DMATS as we do not fully understand the reaction mechanism yet and therefore taking control of it is hard to put in practice. This clearly shows that more research on ABBA prenyltransferases is needed to predict changes in the active site and successfully design novel variants for biochemical applications. Nevertheless, our ability to switch the enzyme from a 6-DMATS to a 5-DMATS shows the potential for structure-guided protein design in this class of enzymes. Furthermore, significantly increased product yields of diprenylated derivatives were observed for 5-DMATS$_{Q255N}$, 5-DMATS$_{Q255E}$, 5-DMATS$_{Q255H}$, 5-DMATS$_{Q255V,Y326H}$ and 5-DMATS$_{Q255N,Y326H}$ mutants, which might be worth more detailed investigation in future experiments.

Experimental Procedure

Protein production and purification

Proteins were produced in *E. coli* BL21 (DE3) cells harbouring either the gene of the 6-DMATS protein (aa 1 to 358) from *Micromonospora olivasterospora* (pET28a-6DMATS) (*MolI14.36*) or the 5-DMATS (aa 11 to 366) from *Streptomyces coelicolor* (pET28a-5DMATS) (*FHV93_101284*) with a thrombin cleavage site and His$_8$-tag in a pET28a vector. The cells were grown in LB-medium containing kanamycin (40 μg/mL) at 37 °C to an OD$_{600}$ of 0.4. The temperature was decreased to 18 °C (6-DMATS) or 25 °C (5-DMATS) and the expression was induced with 0.5 mM IPTG at OD$_{600}$ of 0.8. After 16 h of protein production, the cells were harvested by centrifugation (9200g, 15 min, 4 °C). The bacteria pellet was resuspended in His-A buffer (50 mM Trizma base, pH 8.0 for 6-DMATS and pH 9.0 for 5-DMATS, 500 mM NaCl, 10% (v/v) Glycerol, 25 mM Imidazole), lysed by sonication and the cell debris was separated by centrifugation (34,500g, 45 min, 4 °C). The supernatant containing the recombinant His$_8$-tagged fusion protein was loaded onto a HisTrap FF crude column (5 mL, GE Healthcare) and eluted with 205 mM Imidazole. The affinity-tag was removed by thrombin cleavage (0.5 U/mg protein, ON, 4 °C), followed by size exclusion chromatography (SEC) with a Superdex 200 16/60 (GE Healthcare) column (6-DMATS (25 mM HEPES, pH 8.0, 150 mM NaCl); 5-DMATS (25 mM glycine, pH 9.5, 150 mM NaCl)). The protein quality was analysed by circular dichroism

spectroscopy (Figure S7) and analytical size exclusion chromatography with a Superdex 200 Increase 3.2/300 column (6-DMATS (25 mM HEPES, pH 8.0, 150 mM NaCl); 5-DMATS (25 mM glycine pH 9.5, 150 mM NaCl)).

Crystallisation

Crystals were obtained by sitting drop vapour diffusion method. To obtain 6-DMATS crystals, the protein (7.4 mg/mL) was mixed in a 1:1 ratio with precipitant (200 mM CaAc$_2$, 26% PEG 3350) and stored 14 days at 4 °C. Crystals of 5-DMATS were grown over night at 4 °C by mixing 9.5 mg/mL protein with an equal amount of precipitant (150 mM Bis-Tris, pH 5.5, 20% PEG 3350).

Data collection and structure determination

Crystals of unliganded 5-DMATS and 6-DMATS were flash frozen in liquid nitrogen after preincubation with crystallisation buffer containing 30% (v/v) PEG 400 to achieve cryo-protection (5-DMATS^{apo}; 6-DMATS^{trp}). Ligand bound structures (5-DMATS^{ter}; 6-DMATS^{ter}) were obtained by soaking with L-Trp (10 mM) and DMASPP (10 mM) for 0.5–2 h at 4 °C. Data sets of 5-DMATS^{apo}, 5-DMATS^{ter} and 6-DMATS^{trp} were collected at the Swiss Light Source beamline X06DA in Villingen, Switzerland. The data set of 6-DMATS^{ter} was collected at beamline X06SA in Villigen, Switzerland. Experimental diffraction data were reduced by the XDS package.[46] Approximately 1500 reflections were marked and kept separated for validation of the model quality. Initial protein phases were established via molecular replacement (MR) using PHASER[47] and a CHAINSAW[48] modified model of PriB (PDB code 5INJ) as input template. Initial refinement of the MR models involved simulated-annealing refinement as implemented in PHENIX.[49] Structures were refined by several cycles of reciprocal space refinement using REFMAC,[50] implemented in CCP4 package, followed by real-space corrections using COOT.[51] The final refinement steps include TLS parameterisation as obtained by TLSMD.[52] Model validation was performed using MOLPROBITY.[53] Figures were generated with PYMOL.[54] Coordinates for 5-DMATS^{apo}, 5-DMATS^{ter}, 6-DMATS^{trp} and 6-DMATS^{ter}, were deposited with PDB codes 6ZS0, 6ZRZ, 6ZRY and 6ZRX, respectively.

Mutagenesis

Site-directed mutagenesis was performed on the pET28a-6DMATS and pET28-5DMATS constructs to introduce mutations in the active centre with mutagenesis primer (Table S2). The mutagenesis PCR was carried out according to the manufacturer's protocol for the Q5 Polymerase

(NEB), followed by a template digest with 1 μl DpnI (NEB) at 37 °C for 2 h. To ligate the linear PCR product, the PCR product was phosphorylated with T4 Kinase (NEB) at 37 °C for 30 min, followed by the addition of T4 ligase (NEB). The mutants were produced and purified as described above.

Enzyme assays and monitoring the product formation

For determination of relative enzyme activities, the reaction mixtures (50 μL) containing 50 mM Tris–HCl (pH 7.5), 5 mM $CaCl_2$, 1 mM L-Trp, 1 mM DMAPP or GPP, and 10 μg of the purified recombinant protein were incubated at 37 °C for 2 h. The reactions were terminated by addition of 50 μL methanol. After centrifugation at 16700xg for 30 min, 20 μL were analysed on an Agilent HPLC 1260 system with an Agilent Eclipse XDB-C18 column (5 μm, 4.6 × 150 mm). Separation was performed by using a linear gradient from 5 to 100% acetonitrile in water over 40 min and a flow rate of 1 mL/min. Detection was carried out on a photo diode array detector and absorptions at 277 nm were monitored.

Isolation and structure elucidation of the enzyme products

For structure elucidation, the enzyme assays of 5-DMATS and 6-DMATS mutants were carried out in large-scale (10 mL) and the products were isolated on the same HPLC equipped with a semi-preparative Multospher 120 RP-18 column (5 μm, 10 × 250 mm). Separation was performed by using a linear gradient from 30 to 50% over 30 min and a flow rate of 2 mL/min. The isolated products were subjected to NMR analysis on a JEOL ECA-500 spectrometer (JEOL, Akishima, Tokyo, Japan). Interpretation of the ^{1}H NMR spectra and comparison with those published previously confirmed **2**, **3**, **4**, **5**, **6** and **7** to be 5-dimethylallyl-, 6-dimethylallyl-, 5, 6-didimethylallyl-, 5, 7-didimethylallyl- and 6-geranyltryptophan and 5-geranyltryptophan, respectively (Tables S5–S7, Figures S12–S16).[40,42,55]

Determination of kinetic parameters

Enzyme assays for determination of the kinetic parameters (50 μL) contained 1 mM DMAPP or GPP, 5 mM $CaCl_2$, 2 μg protein and L-Trp at final concentrations ranging from 0.02 to 5 mM. The incubations were carried out at 37 °C for 30 min. In the reactions of 5-DMATS$_{L81A}$ and 5-DMATS$_{Q255V}$, 20 and 10 μg protein were used. After addition of MeOH and centrifugation, the supernatants were analysed on HPLC. Kinetic parameters were determined by nonlinear regression using the software GraphPad Prism 6.0 via Michaelis-Menten equation.

Accession numbers

The atomic coordinates and structure factors (accession numbers: 6ZS0, 6ZRZ, 6ZRY and 6ZRX) have been deposited in the Protein Data Bank (https://www.pdb.org/).

CRediT authorship contribution statement

Elena Ostertag: Conceptualization, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization, Supervision. **Liujuan Zheng:** Conceptualization, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Karina Broger:** Investigation, Writing - original draft, Writing - review & editing. **Thilo Stehle:** Conceptualization, Formal analysis, Resources, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition. **Shu-Ming Li:** Conceptualization, Formal analysis, Resources, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition. **Georg Zocher:** Conceptualization, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Supervision.

Acknowledgments

Funding and additional information

Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2020.11.025.

References

1. Sunassee, S.N., Davies-Coleman, M.T., (2012). Cytotoxic and antioxidant marine prenylated quinones and hydroquinones. *Nature Prod. Rep.*, **29**, 513–535.

2. Demain, A., (1999). Pharmaceutically active secondary metabolites of microorganisms. *Appl. Microbiol. Biotechnol.*, **52**, 455–463.

3. Bourgaud, F., Gravot, A., Milesi, S., Gontier, E., (2001). Production of plant secondary metabolites: a historical perspective. *Plant Sci.*, **161**, 839–851.

4. Wätjen, W., Weber, N., Lou, Y.-J., Wang, Z.-Q., Chovolou, Y., Kampkötter, A., Kahl, R., Proksch, P., (2007). Prenylation enhances cytotoxicity of apigenin and liquiritigenin in rat H4IIE hepatoma and C6 glioma cells. *Food Chem. Toxicol.*, **45**, 119–124.

5. Miranda, C., Stevens, J., Helmrich, A., Henderson, M., Rodriguez, R., Yang, Y.-H., Deinzer, M., Barnes, D., et al., (1999). Antiproliferative and cytotoxic effects of prenylated flavonoids from hops (*Humulus lupulus*) in human cancer cell lines. *Food Chem. Toxicol.*, **37**, 271–285.

6. Kato, H., Yoshida, T., Tokue, T., Nojiri, Y., Hirota, H., Ohta, T., Williams, R.M., Tsukamoto, S., (2007). Notoamides A-D: prenylated indole alkaloids isolated from a marine-derived fungus, Aspergillus sp. *Angew. Chem. Int. Ed.*, **46**, 2254–2256.

7. Dat, N.T., Binh, P.T.X., Van Minh, C., Huong, H.T., Lee, J.J., (2010). Cytotoxic prenylated flavonoids from *Morus alba*. *Fitoterapia*, **81**, 1224–1227.

8. Mitscher, L.A., Rao, G.R., Khanna, I., Veysoglu, T., Drake, S., (1983). Antimicrobial agents from higher plants: prenylated flavonoids and other phenols from Glycyrrhiza lepidota. *Phytochemistry*, **22**, 573–576.

9. Botta, B., Vitali, A., Menendez, P., Misiti, D., Monache, G.D., (2005). Prenylated flavonoids: pharmacology and biotechnology. *Curr. Med. Chem.*, **12**, 713–739.

10. Habtemariam, S., (2012). The anti-obesity potential of sigmoidin A. *Pharm. Biol.*, **50**, 1519–1522.

11. Wollinsky, B., Ludwig, L., Hamacher, A., Yu, X., Kassack, M.U., Li, S.-M., (2012). Prenylation at the indole ring leads to a significant increase of cytotoxicity of tryptophan-containing cyclic dipeptides. *Bioorg. Med. Chem. Letters*, **22**, 3866–3869.

12. Debbab, A., Aly, A.H., Edrada-Ebel, R., Müller, W.E., Mosaddak, M., Hakikj, A., Ebel, R., Proksch, P., (2009). Bioactive secondary metabolites from the endophytic fungus *Chaetomium sp.* isolated from *Salvia officinalis* growing in Morocco. *Biotechnol. Agron. Soc. Environ.*, **13**, 229–234.

13. Valliere, M.A., Korman, T.P., Woodall, N.B., Khitrov, G.A., Taylor, R.E., Baker, D., Bowie, J.U., (2019). A cell-free platform for the prenylation of natural products and application to cannabinoid production. *Nature Commun.*, **10**, 1–9.

14. Bocchi, V., Casnati, G., Dossena, A., Villani, F., (1976). Synthesis of N-alkylindoles using tetraalkylammonium salt catalysis. *Synthesis*, **1976**, 414–416.

15. Kimura, M., Futamata, M., Mukai, R., Tamaru, Y., (2005). Pd-catalyzed C3-selective allylation of indoles with allyl alcohols promoted by triethylborane. *J. Am. Chem. Soc.*, **127**, 4592–4593.

16. Botta, B., Delle Monache, G., Menendez, P., Boffi, A., (2005). Novel prenyltransferase enzymes as a tool for flavonoid prenylation. *Trends Pharmacol. Sci.*, **26**, 606–608.

17. Nguyen, P.H., Nguyen, T.N.A., Kang, K.W., Ndinteh, D.T., Mbafor, J.T., Kim, Y.R., Oh, W.K., (2010). Prenylated pterocarpans as bacterial neuraminidase inhibitors. *Biorg. Med. Chem.*, **18**, 3335–3344.

18. Yazaki, K., Sasaki, K., Tsurumaru, Y., (2009). Prenylation of aromatic compounds, a key diversification of plant secondary metabolites. *Phytochemistry*, **70**, 1739–1745.

19. Demain, A.L., Fang, A., (2000). The Natural Functions of Secondary Metabolites. History of Modern Biotechnology I. Springer, pp. 1–39.

20. Kuzuyama, T., Noel, J.P., Richard, S.B., (2005). Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. *Nature*, **435**, 983.

21. Winkelblech, J., Fan, A., Li, S.-M., (2015). Prenyltransferases as key enzymes in primary and secondary metabolism. *Appl. Microbiol. Biotechnol.*, **99**, 7379–7397.

22. Brandt, W., Bräuer, L., Günnewich, N., Kufka, J., Rausch, F., Schulze, D., Schulze, E., Weber, R., et al., (2009). Molecular and structural basis of metabolic diversity mediated by prenyldiphosphate converting enzymes. *Phytochemistry*, **70**, 1758–1775.

23. Heide, L., (2009). Prenyl transfer to aromatic substrates: genetics and enzymology. *Curr. Opin. Chem. Biol.*, **13**, 171–179.

24. Unsöld, I.A., Li, S.-M., (2005). Overproduction, purification and characterization of FgaPT2, a dimethylallyltryptophan synthase from *Aspergillus fumigatus*. *Microbiology*, **151**, 1499–1505.

25. Unsöld, I.A., Li, S.M., (2006). Reverse prenyltransferase in the biosynthesis of fumigaclavine C in Aspergillus fumigatus: gene expression, purification, and characterization of fumigaclavine C synthase FGAPT1. *ChemBioChem*, **7**, 158–164.

26. Bonitz, T., Alva, V., Saleh, O., Lupas, A.N., Heide, L., (2011). Evolutionary relationships of microbial aromatic prenyltransferases. *PLoS One*, **6**

27. Metzger, U., Keller, S., Stevenson, C.E., Heide, L., Lawson, D.M., (2010). Structure and mechanism of the magnesium-independent aromatic prenyltransferase CloQ from the clorobiocin biosynthetic pathway. *J. Mol. Biol.*, **404**, 611–626.

28. Jost, M., Zocher, G., Tarcz, S., Matuschek, M., Xie, X., Li, S.-M., Stehle, T., (2010). Structure− function analysis of an enzymatic prenyl transfer reaction identifies a reaction chamber with modifiable specificity. *J. Am. Chem. Soc.*, **132**, 17849–17858.

29. Fan, A., Li, S.-M., (2016). Saturation mutagenesis on Arg244 of the tryptophan C4-prenyltransferase FgaPT2 leads to enhanced catalytic ability and different preferences for tryptophan-containing cyclic dipeptides. *Appl. Microbiol. Biotechnol.*, **100**, 5389–5399.

30. Liao, G., Mai, P., Fan, J., Zocher, G., Stehle, T., Li, S.-M., (2018). Complete decoration of the indolyl residue in cyclo-L-Trp-L-Trp with geranyl moieties by using engineered dimethylallyl transferases. *Org. Letters*, **20**, 7201–7205.

31. Grundmann, A., Li, S.-M., (2005). Overproduction, purification and characterization of FtmPT1, a brevianamide F prenyltransferase from Aspergillus fumigatus. *Microbiology*, **151**, 2199–2207.

32. Metzger, U., Schall, C., Zocher, G., Unsöld, I., Stec, E., Li, S.-M., Heide, L., Stehle, T., (2009). The structure of dimethylallyl tryptophan synthase reveals a common architecture of aromatic prenyltransferases in fungi and bacteria. *PNAS*, **106**, 14309–14314.

33. Elshahawi, S.I., Cao, H., Shaaban, K.A., Ponomareva, L. V., Subramanian, T., Farman, M.L., Spielmann, H.P., Phillips Jr, G.N., et al., (2017). Structure and specificity of a permissive bacterial C-prenyltransferase. *Nature Chem. Biol.*, **13**, 366.

34. Yin, W.-B., Grundmann, A., Cheng, J., Li, S.-M., (2009). Acetylaszonalenin biosynthesis in Neosartorya fischeri identification of the biosynthetic gene cluster by genomic mining and functional proof of the genes by biochemical investigation. *J. Biol. Chem.*, **284**, 100–109.

35. Yu, X., Zocher, G., Xie, X., Liebhold, M., Schütz, S., Stehle, T., Li, S.-M., (2013). Catalytic mechanism of stereospecific formation of cis-configured prenylated pyrroloindoline diketopiperazines by indole prenyltransferases. *Chem. Biol.*, **20**, 1492–1501.

36. Ma, J., Zuo, D., Song, Y., Wang, B., Huang, H., Yao, Y., Li, W., Zhang, S., et al., (2012). Characterization of a single gene cluster responsible for methylpendolmycin and pendolmycin biosynthesis in the deep sea bacterium Marinactinospora thermotolerans. *ChemBioChem*, **13**, 547–552.

37. Mori, T., Zhang, L., Awakawa, T., Hoshino, S., Okada, M., Morita, H., Abe, I., (2016). Manipulation of prenylation reactions by structure-based engineering of bacterial indolactam prenyltransferases. *Nature Commun.*, **7**, 10849.

38. Schultz, A.W., Lewis, C.A., Luzung, M.R., Baran, P.S., Moore, B.S., (2010). Functional characterization of the cyclomarin/cyclomarazine prenyltransferase CymD directs the biosynthesis of unnatural cyclic peptides. *J. Nat. Prod.*, **73**, 373–377.

39. Roose, B.W., Christianson, D.W., (2019). Structural basis of tryptophan reverse N-prenylation catalyzed by CymD. *Biochemistry*, **58**, 3232–3242.

40. Yu, X., Liu, Y., Xie, X., Zheng, X.-D., Li, S.-M., (2012). Biochemical characterization of indole prenyltransferases filling the last gap of prenylation positions by A 5-dimethylallyltryptophan synthase from *Aspergillus Clavatus. J. Biol. Chem.*, **287**, 1371–1380.

41. Winkelblech, J., Liebhold, M., Gunera, J., Xie, X., Kolb, P., Li, S.M., (2015). Tryptophan C5-, C6-and C7-prenylating enzymes displaying a preference for C-6 of the indole ring in the presence of unnatural dimethylallyl diphosphate analogues. *Adv. Synth. Catal.*, **357**, 975–986.

42. Winkelblech, J., Xie, X., Li, S.-M., (2016). Characterisation of 6-DMATS Mo from Micromonospora olivasterospora leading to identification of the divergence in enantioselectivity, regioselectivity and multiple prenylation of tryptophan prenyltransferases. *Org. Biomol. Chem.*, **14**, 9883–9895.

43. Holm, L., (2019). Benchmarking fold detection by DaliLite v. 5. *Bioinformatics*, **35**, 5326–5327.

44. Mai, P., Zocher, G., Stehle, T., Li, S.-M., (2018). Structure-based protein engineering enables prenyl donor switching of a fungal aromatic prenyltransferase. *Org. Biomol. Chem.*, **16**, 7461–7469.

45. Luk, L.Y., Qian, Q., Tanner, M.E., (2011). A cope rearrangement in the reaction catalyzed by dimethylallyltryptophan synthase?. *J. Am. Chem. Soc.*, **133**, 12342–12345.

46. Kabsch, W., (2010). Xds. *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, **66**, 125–132.

47. McCoy, A.J., (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, **63**, 32–41.

48. Stein, N., (2008). CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.*, **41**, 641–643.

49. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., et al., (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, **66**, 213–221.

50. Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., Murshudov, G.N., (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, **60**, 2184–2195.

51. Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., (2010). Features and development of Coot. *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, **66**, 486–501.

52. Painter, J., Merritt, E.A., (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, **62**, 439–450.

53. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., et al., (2018). MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.*, **27**, 293–315.

54. Schrodinger, L. (2008). The PyMOL Molecular Graphics System Version 1.2 r3pre.

55. Winkelblech, J., Li, S.M., (2014). Biochemical investigations of two 6-DMATS enzymes from *Streptomyces* reveal new features of L-tryptophan prenyltransferases. *ChemBioChem*, **15**, 1030–1039.

56. Schultz, A.W., Oh, D.-C., Carney, J.R., Williamson, R.T., Udwary, D.W., Jensen, P.R., Gould, S.J., Fenical, W., et al., (2008). Biosynthesis and structures of cyclomarins and cyclomarazines, prenylated cyclic peptides of marine actinobacterial origin. *J. Am. Chem. Soc.*, **130**, 4507–4516.

Acknowledgments

... meinen fleißigen Korrektur-Lesern, Georg, Michael, Katja und Nina, für ihr aufmerksames Lesen und die Liebe zum Detail.

... meinen Freunden Magda, Anja, Johanna, Katrin, Julia, Thomas, Jonas, Mario und Carsten. Unsere Freundschaft ist unverändert, egal wie oft wir uns sehen oder wo wir gerade auf der Welt verstreut sind. Ich bin froh, euch als als Freunde zu haben.

... meiner Familie, die mich nicht nur wo immer möglich unterstützt hat, sondern auch einen wundervollen Ausgleich zum wissenschaftlichen Alltag bildet.

... Simon. Ohne deine Unterstützung und dein Verständnis wäre Vieles nicht möglich gewesen. Ich freue mich, mein Leben mit dir teilen zu können.