Helix Shuffling in Staphylococcal Protein A – Design, Selection and Characterization of Novel Domain B Variants

Dissertation

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

submitted to the Faculty of Natural Sciences I – Life Sciences –
of Martin-Luther-Universität Halle-Wittenberg

by
Hanna Bobolowski

2022

The work presented in the dissertation was conducted as industrial PhD between April 2018 and Mai 2022 at Navigo Proteins GmbH under the supervision of PD. Dr. Hauke Lilie at the Institute for Biochemistry and Biotechnology at the Martin-Luther-Universität Halle-Wittenberg.

1st Reviewer: PD Dr. Hauke Lilie

2nd Reviewer: Prof. Dr. Frank Bordusa

3rd Reviewer: Prof. Dr. Reinhard Sterner

Date of defense: 26. 09. 2023

Table of contents

Abbreviations

°C	Degree Celsius
Ω	Ohm
Å	Ångstrøm, 1 Å = 10^{-10} m
A	Optical absorption
ABD	Albumin-binding domain
AIC	Affinity chromatography
AEBSF	4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride
ATP	Adenosine triphosphate
a.u.	Arbitrary unit
BLI	Biolayer interferometry
bp	Base pair
BSA	Bovine serum albumin
CAP	Catabolite activator protein
CD	Circular dichroism
cDNA	Complementary DNA
CDR	Complementarity determining region
CH	Constant domain of heavy chain
CHO	Chinese hamster ovary cells
CID	Clone identification number
CL	Constant domain of light chain
CSA	Chemical shift anisotropy
CV	Column volume
Da	Dalton
ddH$_2$O	Double-distilled water
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
DTT	Dithiothreitol
EB	Elution buffer
E. coli	*Escherichia coli*
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EF	End fragment
ELISA	Enzyme-linked immunosorbent assay
GFP	Green fluorescent protein
F	Farad
Fab	Fragment antigen-binding
Fc	Fragment crystallizable
FDA	Food and drug administration
FI	Fluorescence intensity
g	Acceleration
GdnHCl	Guanidine hydrochloride

h	Hour
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
HRP	Horseradish peroxidase
HSQC	Heteronuclear single quantum coherence
HTRF	Homogeneous time-resolved fluorescence
HTS	High-throughput screening
IF	Intermediate fragment
IgG	Immunoglobulin G
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KIH	Knobs-into-holes
k_{on}	Association constant rate
k_{off}	Dissociation constant rate
K_D	Dissociation constant [M]
LB	Lysogeny broth
LIP	Loops in proteins
M	Molar
MALS	Multiangle light scattering
mAU	Milli absorption units
MD	Molecular dynamics
MES	2-(N-morpholino)ethanesulfonic acid
mHz	Megahertz
min	Minutes
MOE	Molecular operating environment
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	Messenger RNA
MRE	Mean residue ellipticity
MW	Molecular weight
MWCO	Molecular weight cut-off
NGS	Next-generation sequencing
NHS	N-hydroxysuccinimide
Ni-NTA	Nickel nitriloacetic acid agarose
nm	Nanometer
NMR	Nuclear magnetic resonance
OD	Optical density
oePCR	Overlap extension PCR
p. a.	Pro analysis
PACB	Protein A cyclic shuffled domain B
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase-chain reaction
PDB	Protein data bank
POI	Protein of interest

ppm	Parts per million
RBD	Receptor binding domain
RD	Ribosome display
RF	Randomized fragment
RNA	Ribonucleic acid
Rop	Repressor of primer protein
rpm	Rounds per minute
RU	Response unit
SARS	Severe acute respiratory syndrome
S. aureus	*Staphylococcus aureus*
scPCR	Single colony PCR
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SF	Start fragment
sfGFP	Super folder green fluorescent protein
SOC	Super optimal broth with catabolite repression
SPR	Surface plasmon resonance
SR	Selection round
TAE	Tris-acetate EDTA
TB	Terrific broth
TCEP	Tris(2-carboxyethyl)phosphine
TNFα	Tumor necrosis factor alpha
TNFR1	Tumor necrosis factor receptor 1
TOCSY	Total correlation spectroscopy
Tris	Tris(hydroxymethyl)aminomethane
TROSY	Transverse relaxation optimized spectroscopy
TVMV	Tobacco vein mottling virus
U	Unit
UV	Ultraviolet
V	Volt
VH	Variable domain of heavy chain
Vis	Visible
v/v	Volume per volume
vWf	Von Willebrand factor
WG	Working group
wt	Wild type
w/v	Weight per volume

Abstract

Small and highly stable domains of Protein A, an IgG-binding surface protein originally found in the cell wall of *Staphylococcus aureus* (*S. aureus*), are attractive scaffolds for the generation of engineered binding proteins. For example, proteins obtained by rational design are used as ligands for chromatographic purification of antibodies and Fc-fusions. Along this line, the aim of the present work is to generate IgG-binding variants of domain B by helix shuffling that are suitable as affinity ligands. First, the influence of loops on structure and function of domain B is investigated, since the considered helix shuffling strategy also involves insertion of alternative loops. For this, the loops are partially or completely substituted for $(GS)_n$ linkers. Variants generated with this rational engineering approach show a lower affinity for Fc and to some extent a lower proportion of α-helical structures compared to wild type domain B. Nevertheless, both loops of domain B can be substituted to a certain extent without destroying the functionality and structure of the protein.

Second, the helices in the three-helix bundle are swapped by helix shuffling and two alternative loops with several randomized positions are inserted to connect the rearranged structural elements and generate diversity. DNA libraries are constructed and IgG-binding variants are selected by ribosome display. Next, high-throughput screening is performed to identify individual IgG-binding variants. To characterize the function and structure of the engineered binding proteins, 24 selected variants are produced on larger scale. All analyzed variants show a decreased affinity for Fc compared to wild type domain B, mainly caused by a higher dissociation rate. An NMR titration experiment shows that 81.8 % of the residues involved in Fc binding correspond to the Fc-binding sites of wild type domain B. Interestingly, all variants show a significant decrease in thermal stability (T_m ~31 °C), although α-helical content is comparable to wild type control. CID 25, an IgG-binding variant ($K_D = 52$ nM) that exhibits a proportion of α-helical structures ≥ 70 %, is selected for a detailed characterization. To investigate structural transitions and folding mechanisms of CID 25, thermal-induced unfolding and refolding of the protein are analyzed by DSC. In contrast to wild type domain B, CID 25 shows a non-two state folding and irreversible refolding. Next, the structure of CID 25 is investigated experimentally by NMR and *in silico* by the AI-based network AlphaFold. NMR experiments reveal a mainly right-handed α-helical secondary structure of CID 25. All three expected helices inspired by wild type domain B are experimentally verified. Indeed, the helices are arranged in a compact bundle with domain B-like fold as demonstrated by 3D structure prediction using AlphaFold. Compared to theory, the helix 3/2 segment looks structurally as expected, whereas the helix 2/1 segment differ from expectations.

Finally, the suitability of CID 25 as ligand for affinity chromatography is investigated. CID 25 is coupled on the surface of a column matrix and purification of hIgG1 from cell extracts is demonstrated. The CID 25 column exhibits lower binding capacity and lifetime compared to matrix functionalized with wild type domain B. An elution study demonstrates that bound hIgG1 can be eluted at pH 4.8 from the CID 25 column, which is milder compared to wild type domain B (pH 3.7). Taken together, this initial engineering approach offers high technical potential and is suitable for the generation of IgG-binding proteins structural and functional related to domain B. However, further engineering is necessary to improve the stability and affinity of these binding proteins.

Zusammenfassung

Protein A ist ein IgG-bindendes Oberflächenprotein, welches in der Zellwand von *Staphylococcus aureus* (*S. aureus*) vorkommt. Aufgrund ihrer Eigenschaften werden einzelne Domänen des Protein A als Gerüstproteine zur Generierung von alternativen Bindeproteinen verwendet. Die durch rationales Design erzeugten rekombinanten Proteine werden beispielsweise als Liganden für die chromatographische Reinigung von Antikörpern und Fc-Fusionen eingesetzt. Das Ziel der vorliegenden Arbeit ist es, durch Helix-Shuffling in Domäne B IgG-bindende Varianten zu generieren, welche als Affinitätsliganden genutzt werden können. Da die in Betracht gezogene Helix-Shuffling-Strategie die Insertion alternativer Loops vorsieht, wird zunächst der Einfluss der natürlichen Domäne B Loops auf die Struktur und Funktion des Proteins untersucht. Dazu werden die Loops teilweise oder vollständig mit $(GS)_n$-Linkern substituiert. Die mit diesem rationalen Engineering-Ansatz generierten Varianten zeigen eine verringerte Affinität zu Fc und einen teilweise reduzierten Anteil an α-helikalen Strukturen im Vergleich zu Wildtyp Domäne B. Jedoch können beide Loops der Domäne B in gewissem Maß substituiert werden, ohne die Funktionalität und Struktur des Proteins zu zerstören.

Anschließend wird durch Helix-Shuffling die Reihenfolge der Helices im Drei-Helix-Bündel geändert. Zwei alternative Loops mit mehreren randomisierten Positionen werden eingefügt, um die neu angeordneten Strukturelemente zu verbinden und Diversität zu generieren. Vier DNA-Bibliotheken werden konstruiert und IgG-bindende Varianten durch Ribosomen-Display selektiert. Zur Identifizierung von individuellen IgG-bindenden Varianten wird ein Hochdurchsatz-Screening durchgeführt. Um die Funktion und Struktur der erhaltenen Bindeproteine zu charakterisieren, werden 24 repräsentative Varianten für eine Produktion in größerem Maßstab ausgewählt. Alle analysierten Varianten zeigen eine verringerte Affinität zu Fc verglichen zu Wildtyp Domäne B, was hauptsächlich auf eine erhöhte Dissoziationsrate zurückzuführen ist. Ein NMR-Titrationsexperiment zeigt, dass 81,8 % der Reste, die an der Fc-Bindung beteiligt sind, den Fc-Bindesstellen von Wildtyp Domäne B entsprechen. Interessanterweise zeigen alle Domäne B Varianten eine signifikante Abnahme der thermischen Stabilität (T_m ~31 °C), obwohl der Anteil an α-helikalen Strukturen vergleichbar zur Wildtyp-Kontrolle ist. Die Variante CID 25, welche eine nanomolare Affinität zu Fc ($K_D = 52$ nM) und einen Anteil an α-helikalen Strukturen ≥ 70 % aufweist, wird für eine detaillierte Charakterisierung ausgewählt. Um Faltungsmechanismen und strukturelle Übergänge von CID 25 zu untersuchen, werden thermisch induzierte Entfaltung und Rückfaltung des Proteins mittels DSC analysiert. Im Gegensatz zu Wildtyp Domäne B ist die Faltung von CID 25 nicht mit einem Zwei-Zustandsmodell beschreibbar, und die Variante zeigt darüber hinaus eine irreversible Rückfaltung. Anschließend wird die Struktur von CID 25 experimentell mittels NMR und *in silico* mittels KI-basiertem Netzwerk AlphaFold untersucht. Die NMR-Experimente ergeben eine überwiegend rechtshändige α-helikale Sekundärstruktur von CID 25. Alle drei erwarteten Helices, welche durch Wildtyp Domäne B inspiriert sind, können experimentell verifiziert werden. Die 3D-Strukturvorhersage mittels AlphaFold zeigt, dass die drei Helices tatsächlich in einem kompakten Bündel mit Domäne B-ähnlicher Faltung angeordnet sind. Im Vergleich zur Theorie sieht das Helix 3/2-Segment strukturell wie erwartet aus, während das Helix 2/1-Segments von den Erwartungen abweicht.

Schließlich wird die Eignung von CID 25 als Ligand einer Affinitätschromatographie untersuchen. Dazu wird CID 25 auf die Oberfläche einer Säulenmatrix gekoppelt und die Reinigung von hIgG1 aus Zellextrakten demonstriert. Die CID 25-Säule zeigt eine reduzierte Bindekapazität und Lebensdauer im Vergleich zu Matrix, welche mit Wildtyp Domäne B funktionalisiert ist. Jedoch konnte gebundenes hIgG1 schon bei einem pH-Wert von 4,8 von der CID 25-Säule eluiert werden, was verglichen zu Wildtyp Domäne B (pH 3,7) deutlich milder ist. Zusammengefasst bietet dieser erste Engineering-Ansatz ein hohes technisches Potenzial und eignet sich für die Generierung von IgG-bindenden Proteinen, welche strukturell und funktionell mit Domäne B verwandt sind. Es ist jedoch weiteres Proteinengineering erforderlich, um die Stabilität und Affinität dieser neuartigen Domäne B Varianten zu verbessern.

1 Introduction

Certain proteins catalyze chemical reactions in the cell (enzymes) and others give the cell a specific structure (structural proteins). Motor proteins ensure movements within the cell and receptor proteins bind various chemical messengers. However, most proteins are involved in the regulation of other proteins and their biosynthesis (Litwack, 2021). Thus, proteins can mediate a wide variety of different functions. One of their main roles is to recognize and bind other molecules. Hormones must bind to their receptors to exert their effects; other proteins bind to proteins to determine their translocation or to mark them for degradation. After binding, proteins can change the bound molecule, adjust their structural conformation, or even trigger subsequent reactions by encountering a third molecule. To this end, each protein has a distinctive surface profile shaped by the side chains of its amino acids. Due to the enormous combinatorial diversity of the amino acid residues of the protein sequence, the surface can assume almost any conceivable shape suitable for substance recognition and binding (Müller-Esterl, 2018).

Antibodies, proteins in the service of the immune system, represent a special case. The immunoglobulins of the IgG class, for example, consist of two light and two heavy chains with hypervariable loops representing the binding surface (Heinrich *et al.*, 2021). Due to their diversity, these proteins can bind to almost any molecular structure, making it accessible to the immune system. About 30 years ago, the immune system represented the only source of molecular diversity from which specific binding proteins could be obtained. Since the invention of the hybridoma technology (Kohler & Milstein, 1975) which allows the production of monoclonal antibodies, the favorable binding properties of antibodies can be successfully exploited in research and therapy. The first therapeutic antibodies were approved in 1986 followed by the introduction of a wide range of antibodies and antibody derivatives for the treatment of inflammatory diseases and cancer (Kaplon *et al.*, 2020). For decades, antibodies have become essential in almost all fields of life sciences. However, despite all the potential applications of antibodies as specific binding proteins this technology has drawbacks, mainly due to the complicated architecture of the molecules. As multidomain proteins of about 150 kDa, the correct conformation of antibodies depends on the formation of disulfide bonds and posttranslational modifications such as glycosylation. Therefore, a production in complex eukaryotic expression systems is required. In addition, antibodies are susceptible to degradation, aggregation, and undesired modifications such as deamidation (Yan *et al.*, 2018). Moreover, the intensive research on antibodies has led to a confusing patent situation that limits further research and commercialization (Gebauer & Skerra, 2020).

Combinatorial protein engineering in combination with generation of synthetic libraries and development of various selection systems has enabled the design of artificial binding proteins with desired properties. Thus, more than 50 so-called alternative scaffold proteins have been described that do not belong to the class of immunoglobulins (Pankaj, 2020). These scaffolds form an extremely divergent group of binding proteins that differ in their origin, size, and structural topology. The term "scaffold" already contains some essential properties for the generation of alternative binding proteins. Ideally, a scaffold consists of a single polypeptide chain, has a size of 6-30 kDa and exhibits high intrinsic thermal stability. A certain proportion of the protein can be modified without affecting its intrinsic properties (Zajc *et al.*, 2020).

Moreover, the absence of internal cysteines facilitates their cytoplasmatic production in bacterial expression systems. Because natively folded scaffold proteins can be produced by renouncing complex *in vitro* refolding processes, their industrial scale applications have significantly increased. Due to their small size and simplified manufacturability, these binding proteins offer advantages in therapy and diagnostics, and are suitable as ligands for chromatographic applications (Gebauer & Skerra, 2020). Bacterial surface proteins and individual domains derived therefrom have been identified which bind some immunoglobulin (Ig) families and isotypes. For example, Protein A from *S. aureus* binds human IgG1, 2, and 4 with an affinity of $K_D = 10^{-7}$-10^{-8} M via interaction with the Fc region of the antibody. In addition, protein G from the genus *Streptococcus* recognizes all human IgG subtypes, whereas protein L from *Peptostreptococcus magnus* is restricted to those IgGs that contain kappa light chains (Mouratou *et al.*, 2015). Today, these small (5-60 kDa) and highly stable proteins ($T_m > 70$ °C), mostly improved by protein engineering, are gaining attention for a variety of biotechnological and medical applications. In methods such as ELISA and immunoprecipitation, antibodies can be captured or detected by Protein A/G. In addition, Protein A is widely used for the efficient chromatographic purification of immune complexes and serum immunoglobulins applied in pharmaceutical approaches (Rigi *et al.*, 2019).

1.1 Protein A

1.1.1 Structural properties of Protein A

The B domain of Protein A, a 42 kDa multi-domain surface protein originally located in the cell wall of *S. aureus*, is the main subject of this work. In 1958, Jensen observed the binding of a bacterial antigen extracted from *S. aureus*, so-called antigen A, to rabbit and human serum antibodies and described the molecule as a polysaccharide (Jensen, 1958). Jensen's assumption was quickly refuted by data, showing that antigen A is a surface protein in the bacterial cell wall (Löfkvist & Sjöquist, 1962). The Bergen group described the antigen as "Protein A" to clearly distinguish the molecule from other staphylococcal proteins and polysaccharides.
The entire protein consists of a likely disordered C-terminal half that anchors Protein A to the extracellular surface via a LPXTG motif, and a N-terminal region comprising five highly stable IgG-binding domains (E, D, A, B, and C) (Figure 1). The conserved sequence KADNKF connects all domains, except E with D, where KADAQQNKF forms the highly flexible linker (Deis *et al.*, 2014). The domains are characterized by a three-helix bundle topology with a sequence homology between 74 % and 91 % compared to domain A. Each domain folds in a thermodynamically uncoupled manner from the others, whereas the thermal stability increases towards domain C. For example, domain B rapidly unfolds and refolds (~70 times per second) in a two-state model, which has also been demonstrated for the other domains (Myers and Oas, 2001). As a pathogen, *S. aureus* uses Protein A in combination with a broad range of other proteins and surface factors to ensure its survival during host infection by binding of host antibodies. After specific binding to Fc and Fab fragments of antibodies was demonstrated, Protein A was used as scaffold for the generation of engineered binding proteins. The first commercial Protein A was released in 1978 (Ey *et al.*, 1978) and subsequently a variety of engineered affinity ligands based on Protein A were developed (Rigi *et al.*, 2019).
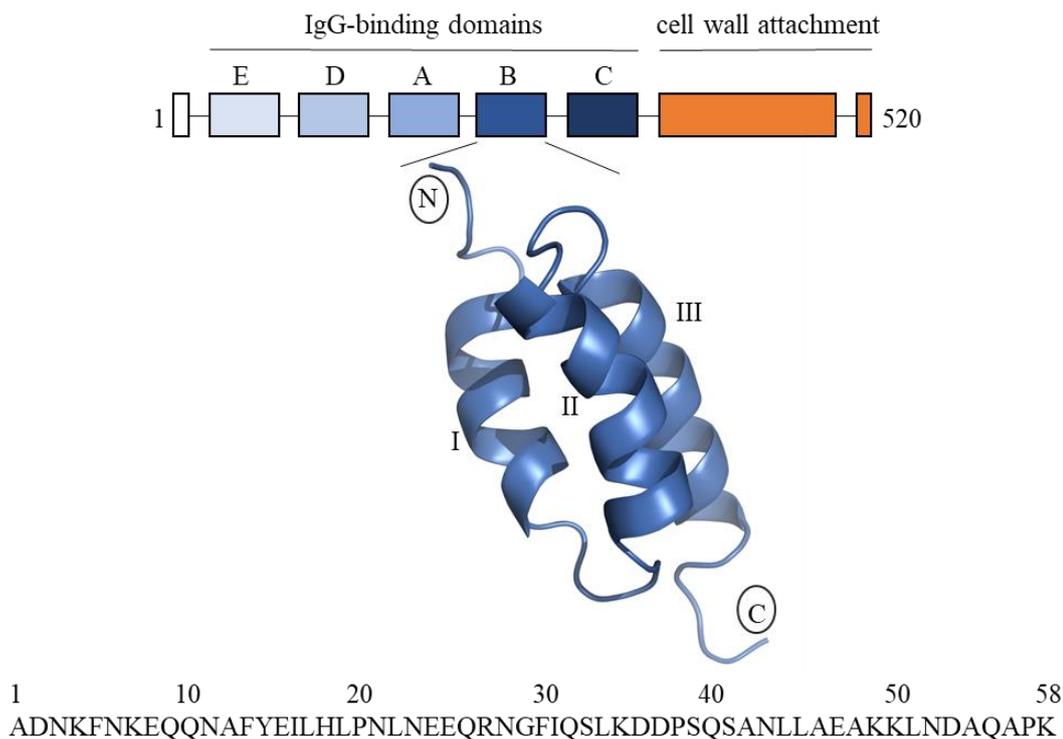
Figure 1: Staphylococcal Protein A and the NMR structure of its domain B (PDB: 1BDD).
The figure shows the structure of Protein A found in *S. aureus* with its five IgG-binding domains (A, B, C, D, and E) colored in different shades of blue. The tertiary structure of wild type domain B (PDB: 1BDD) is shown in which the three helices (I-III) and the termini (N, C) are labeled. The amino acid sequence of domain B is given.

The 3D structure determination of Protein A in the early 1980s enabled the targeted improvement of the already advantageous protein properties. Currently, several tertiary structures of individual wild type domains are known, some of which form complexes with antibody fragments. Deisenhofer made the start by resolving the structure of domain B in complex with the human Fc fragment (PDB: 1FC2) refined at 2.9 °A resolution (Deisenhofer, 1981). However, since this crystal structure lacked coordinates for most of the helix 3 residues, further efforts were made by Gouda *et al.* (Gouda *et al.*, 1992) that led to the complete structure determination of domain B (PDB: 1BDD). The tertiary structure of domain B is characterized by three amphipathic helices connected by two highly conserved loops that form a small bundle of ~6.6 kDa (Figure 1). The helices are arranged in an antiparallel orientation which is typical for most three-helix bundle proteins. A hydrophobic core is formed based on Crick's "knobs-into-holes" (KIH) packing model in which the hydrophobic amino acids "knobs" of one helix bury themselves into the "holes" between the hydrophobic knobs of their partner helix (Crick, 1953). The three-helix bundle is stabilized by a variety of non-covalent interactions between amino acids in α-helical regions of the protein. Furthermore, amino acids of the two loops are highly involved in formation of van-der-Waals interactions and hydrogen bonds (Shao & Gao, 2011). Due to their structural properties, all Protein A domains possess high thermal and chemical stability. Under physiological conditions, the proteins unfold at a temperature of ~70 °C and are resistant to treatment with acidic solutions (pH 3.0) (Minakuchi *et al.*, 2013).

1.1.2 The interaction of Protein A and antibodies

During host infection, Protein A has a wide range of different functions that require the binding of many targets. Tumor necrosis factor receptor 1 (TNFR1) is one of these targets. TNFR1 binds to Protein A residues in helix 1 and helix 2 which mimics TNFα activation of epithelial cells and leads to inflammation. In addition, Protein A binds the A1 domain of von Willebrand factor (vWf), allowing *S. aureus* to adhere to surfaces. Protein A also mediates immune evasion by binding of host antibodies. Each domain of Protein A can bind both Fc and Fab fragments, with an affinity 32-fold higher for Fc (10-30 nM) than for Fab. Hence, binding of the Fc subunit to host phagocyte Fc receptor is inhibited, protecting *S. aureus* from phagocytosis. It is assumed that binding of Protein A to Fab during pathogen infection contributes to selection of B cell antigen receptors encoded for VH3 and suppression of B lymphocytes (Deis *et al.*, 2015).

Antibodies belong to the class of globulins and are synthesized in vertebrates as an important part of the humoral immune response. The success story of antibodies began with the observation that blood of bacterial infected animals showed a destructive effect on bacteria in cultures (Hankin, 1890). Antibodies can mediate an immunogenic effector function by highly specific binding of foreign target molecules known as antigens. Antibodies can not only block antigens, but also activate the complement system and initiate phagocytosis. Five globulin isoforms can be found in mammalian cells, each with an individual function (Chiu *et al.*, 2019). For initial defense of pathogens, IgAs are mainly secreted as homodimers in the gastrointestinal tract, the eyes, and the mucous membranes of the respiratory tract. IgD can be detected in small amounts in the blood or the lymph. Located on the membrane of naïve B cells, IgDs act as antigen receptors. IgEs are involved in parasite defense, whereas IgMs are the first immunoglobulins produced upon activation of the complement system. IgGs are formed in a delayed defense phase and belong to the secondary immune response (Müller-Esterl, 2018).

Today, full-length antibodies can be produced in eukaryotic cell lines such as Chinese hamster ovary (CHO) cells. In addition, the immune systems of mammals such as mice, sheep and goats are valuable sources of antibodies. Thus, a mixture of polyclonal antibodies with different affinity or specificity for different epitopes can be produced. The hybridoma technology can be used to produce monoclonal antibodies (Kohler & Milstein, 1975). Functional antibody fragments can be generated by using proteases such as pepsin and papain (Deveuve *et al.*, 2020). For example, small Fab fragments enable better tissue penetration in various *in vivo* applications and improve coupling density when immobilized on solid matrices in technical applications (Bates & Power, 2019). Today, more than 70 therapeutic antibodies have been approved by the FDA as biopharmaceuticals, making the monoclonal antibody market valued at approximately $150 billion (Lu *et al.*, 2020). Currently, antibodies can be used to treat diseases such as rheumatoid arthritis, osteoporosis, breast cancer and psoriasis. Various preparative chromatographic methods can be used to purify antibodies, although the majority of all approved monoclonal antibodies is purified by a Protein A chromatography step (Bolton & Mehta, 2016).

Antibodies consist of two identical light (L-chain, approximately 25 kDa) and heavy polypeptide chains (H chain, approximately 50 kDa) arranged in a Y-shape (Figure 2). The individual chains consists of different immunoglobulin domains, each containing about 110 amino acids in a predominantly antiparallel β-sheet structure. A centrally located intramolecular

disulfide bridge stabilizes the individual domains and an intermolecular disulfide bond connects a light chain with a heavy chain, respectively. The light chains consist of one variable (VL) and one constant domain (CL), whereas the heavy chains consist of one variable domain (VH) and several constant domains (CH). Between the CH1 and the CH2 domains is a hinge region where the heavy chains are connected by two to three intermolecular disulfide bridges depending on the antibody class described above (class A, D and G: CH1-3, class E and M: CH1-4).
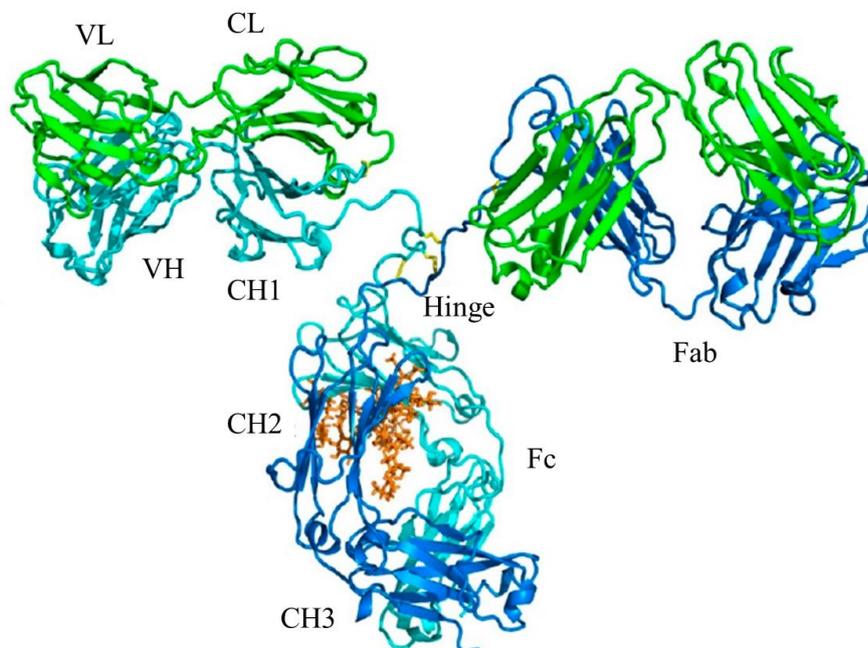
Figure 2: Structure of an immunoglobulin G (PDB: 1IGT) as ribbon representation (Chiu *et al.*, 2019). Antibodies consist of two identical light (green) and heavy chains (cyan and blue) connected by disulfide bridges (yellow). The IgG type consist of three constant domains (CH), two of which (CH2 and CH3) form the Fc fragment which is post translationally glycosylated (orange). The hinge region connects the Fc with the Fab fragment. The Fab fragment is formed by the constant domain of the light chain (CL), the variable domains (VL and VH), and the heavy chain. Three hypervariable loops (complementarity determining regions, CDRs) are located on each variable domain of the light and the heavy chains.

According to the structure solved by Deisenhofer *et al*. (PDB: 1FC2), domain B binds to the hinge region between the CH2 and CH3 domains of the Fc region. Several amino acids in the helix 1/2 interface of domain B contact with the residues S254, Q311, L432, and N434 located in the Fc region of the antibody. More precisely, F5, Q9, Q10, N11, F13, Y14, L17 (helix 1) and N28, I31, Q32, K35 (helix 2) of Protein A domains are involved in Fc binding (Figure 3). Mutagenesis experiments confirmed the findings obtained from the crystal structure 1FC2. A two-helix derivative of domain B still bound Fc with an affinity comparable to that of the wild type and, in addition, mutation of key residues including Q9 and Q10 decreased the affinity of domain B for Fc (Kim *et al.*, 2010). Upon Fc binding, the residues in the helix 1/2 interface undergo extensive conformational changes. Comparison of the solution structure of domain C (PDB: 4NPD) with that in the Fc complex (PDB: 4WWI) reveals a significant variation in the angle between helix 1 and helix 2 leading to suppression of structural heterogeneity upon Fc binding (Deis *et al.*, 2015). In addition to Fc binding, all Protein A domains also bind the Fab region of antibodies via the helix 2/3 interface as clearly shown by the 3D structure determination of domain D in complex with a Fab fragment (PDB: 1DEE) (Graille *et al.*, 2000). Residues Q26, G29, F30, Q32, S33, D36 (helix 2) and N43, E47, L51 (helix 3) contact the VH

region and form the interface with ~32-fold weaker affinity compared to Fc binding (Figure 3).
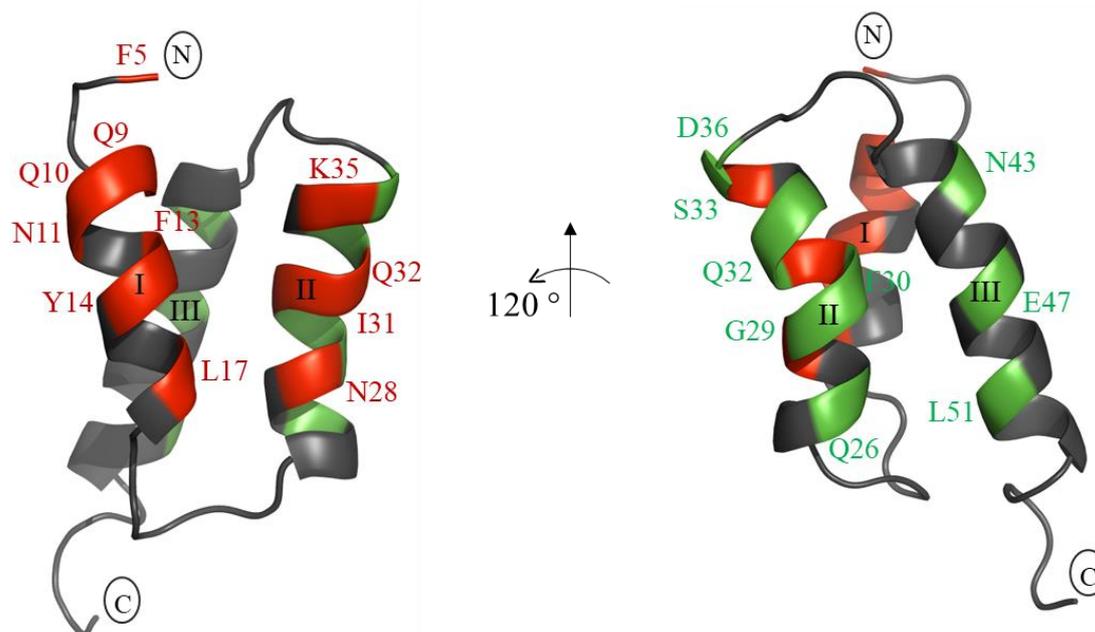
Figure 3: Interaction sites of domain B (PDB: 1BDD) with Fc and Fab fragments.
Starting from the N-terminus, the three helices of domain B (PDB: 1BDD) are numbered I-III. Residues colored in red are involved in Fc binding and are located in helix 1 (F5, Q9, Q10, N11, F13, Y14, and L17) and helix 2 (N28, I31, Q32, and K35). Fab binding is mediated by the green colored residues in helix 2 (Q26, G29, F30, Q32, S33, and D36) and helix 3 (N43, E47, and L51). The figure was created as a cartoon representation using the software PyMOL.

1.2 Protein A as scaffold for the generation of engineered binding proteins

Due to its beneficial properties, Protein A and its individual domains have already been successfully used as scaffold for the generation of binding proteins derived therefrom. Proteins that bind to various targets could be obtained by rational engineering approaches and empirical diversification followed by selection via ribosome and phage display. In early engineering efforts, the primary sequence of Protein A served as basis for improvements of chemical stability. Domain B served as template for the development of the synthetic Z domain by targeted mutation of the amino acids A1V and G29A, making the protein less sensitive to hydroxylamine treatment (Nilsson *et al.*, 1987). A side effect was that the G29A substitution simultaneously removed the alkali-sensitive N28-G29 motif. Since asparagines are prone to deamidation and backbone cleavage which are accelerated at high pH, protein engineering efforts have focused on substituting these residues to obtain variants that are stable under alkaline pH conditions (Linhult *et al.*, 2004). Since affinity ligands are exposed to extremely harsh pH conditions during target elution (pH ≤ 3) and column cleaning (pH ≥ 10), the desire for highly stable proteins has led to the generation of various engineered binding proteins based on domain Z. For example, MabSelect ligands (Cytiva) exhibit exceptional chemical stability and an improved tolerance for harsh cleaning conditions, allowing the reuse of the expensive resins in further purification cycles. Starting from domain Z, 13 solvent-exposed residues in helices 1 and 2 which form the Fc binding region were optimized by mutagenesis. Based on the generated scaffold, so-called Affibody, engineered binding proteins against various other targets were obtained from library selections using phage and ribosome display (Ståhl *et al.*,

2017). The anti-HER2 Affibody ABY-025 with an affinity of $K_D = 22$ pM is the most popular therapeutical molecule for breast cancer diagnosis (Alhuseinalkhudhur *et al.*, 2020). In addition, Affibodies can be used as ligands for the affinity chromatographic purification of various targets such as antibodies, apolipoproteins, Taq polymerases, recombinant human factor VIII, and RSV-G (Ståhl *et al.*, 2017).

A few efforts focused on loop engineering of individual wild type domains and domain Z in which Gülich *et al.* substituted and even extended the loop between helix 2 and helix 3 for multiple glycine residues. Although the domain was still functional, a lower affinity for IgG was observed, allowing a mild elution of antibodies at pH 4.5 from chromatographic columns (Gülich *et al.*, 2000). Another engineering approach was used by Kanje *et al.* by inserting an EF-hand loop between the helices 2 and 3 of domain Z to enable calcium-dependent binding of antibodies. A protein library with several randomized positions was generated and a descendant of domain Z (Z_{Ca}) that showed calcium-dependent binding of antibodies was obtained from phage display selection (Kanje *et al.*, 2018).

Inspired by domain B, Reiersen and Rees studied so-called minidomains that consist of two helices (helices 1 and 2) stabilized by a disulfide bond. In addition, the natural loop between the two helices was substituted for an elastin turn, resulting in thermo- and salt-responsible proteins (Reiersen & Rees, 1999). At room temperature, these proteins bind IgG, whereas elution is induced by a decrease in temperature or salt concentration. However, these ligands are currently not widely used in industry. Similar efforts were made by minimizing the size of domain Z while retaining its function (Kanje *et al.*, 2020). First, the helix bundle was truncated by deletion of helix 3, resulting in a 105-fold decrease in affinity for Fc. However, by use of library design and phage display techniques, affinity was recaptured and a new protein with 12 amino acid substitutions was generated (Z34C). In a final minimization step, even helix 2 of Z34C was removed, resulting in a peptide with 14 amino acids (LH1). A covalent linker was engineered and binding to Fc with millimolar affinity was observed.

The mosaic technology developed by Navigo Proteins provides another engineering approach. The amino acid sequences of all five Protein A domains and various other known staphylococcal proteins were aligned *in silico* and statistically fragmented. By reassembling the fragments, a novel IgG-binding protein scaffold was generated that allows further modifications by site-specific randomization of the amino acid sequence. This commercial approach enables the generation of ligands for affinity chromatographic purification of antibodies and a variety of non-IgG targets. Theses artificial binding proteins are characterized by exceptional caustic stability and high dynamic binding capacities. The commercially available ligand NGL-Impact A (Repligen) can be used for the capturing of monoclonal antibodies with a DBC_{10} of ~80 g/L and a caustic stability of > 95 % after treatment with 0.5 M NaOH for at least 50 h Another commercially available ligand NGL-Impact A HipH (Repligen) allows the purification of even pH-sensitive antibodies and antibody fragments (Knick *et al.*, 2022). The intrinsic Fc binding of these scaffolds can be deleted by single point mutations within the first two helices of the three-helix bundle, generating new specific binding sites for different targets. The ligand NGL-COVID-19 Spike Protein (Repligen) with nanomolar affinity for the spike protein of the SARS-CoV-2 virus, discovered by phage display selection, was developed for the chromatographic purification of protein-based coronavirus vaccines (Kahl *et al.*, 2021). All examples presented underline the potential of individual Protein A domains to serve as scaffolds

for the generation of engineered binding proteins. Due to their adaptability to their respective tasks and their remarkable chemical and thermal stability, these binding proteins represent a valuable molecular tool for research, therapy, and industry.

1.3 Generation of engineered binding proteins

Based on a structure-giving scaffold, engineered binding proteins can be generated by diversification of solvent-exposed amino acid residues followed by selection of variants with desired property profile (Ku & Schultz, 1995). Basically, the goal is to achieve the largest possible potential binding surface while maintaining the native structure, thermal stability, and solubility of the scaffold. The selection of suitable randomization positions of the scaffold can be based on different criteria. If the functionality of the scaffold should be retained, it makes sense to use residues or structural elements that are not directly involved in the interaction with the natural binding partner. Residues close to those that form the natural binding surface can also be used for randomization to specifically increase the affinity for the target. In addition to substitutions, multiple insertions or deletions can also be a part of the strategy to generate diversity (Hey *et al.*, 2005) (Schilling *et al.*, 2014). In addition, existing interaction sites of the scaffold can be used as starting point for the generation of a *de novo* paratope as already demonstrated for engineered binding proteins such as Affibodies (Zajc *et al.*, 2020). Positions in loops or secondary structure elements of the scaffold can also be randomized to generate binding surfaces (Gebauer & Skerra, 2020). Even structures with unknown interaction partners have been successfully used for the generation of a *de novo* paratope by rational selection of randomized regions. In particular, computational design of scaffold proteins has led to the successful generation of binding proteins against targets such as IgG, interleukins, and cofactors (Korendovych & DeGrado, 2020). *In silico* generated data and consensus studies of scaffold-related proteins allow the selection of amino acid positions with high substitution tolerance.

Based on the scaffold and the randomization strategy, combinatorial libraries are prepared at DNA level. Ideally, with a functional library it is possible to generate binding proteins against many different target molecules. Variability within the library allows the right binders to be selected for the respective targets. To ensure high quality, DNA libraries should only contain mutations at predefined positions and all nucleotides should be statistically evenly distributed. The symmetrical distribution of codons can be realized by using primers with NNK motifs in which all four nucleotides occur at N positions, whereas at K positions only guanine and thymidine are represented. In such a library, the number of redundant codons is reduced to 32. Another strategy is the use of trinucleotide-based systems (Popova *et al.*, 2015). By using prefabricated trinucleotides in DNA synthesis, sequences coding for arbitrary amino acids can be produced site-specifically in defined ratios (Virnekas *et al.*, 1994). In addition to avoiding undesired amino acids at certain positions in the translated library variants, stop codons and reading frame shifts can also be eliminated to increase the functionality of the libraries. Thus, the theoretical diversity of the libraries depends on the number of randomized positions and the allowed codons. For example, if all 20 amino acids are allowed for mutation, a scaffold with 7 randomized positions results in a library size of 1.28×10^9 variants at protein level. Hence, an appropriate selection process is an important requirement for the successful enrichment of target-specific variants (Kunamneni *et al.*, 2020). For an enrichment of variants with the desired

properties, the phenotypes of the individual library molecules are linked to their corresponding genotypes (display). The formed complexes are exposed to immobilized or immobilizable targets and binding library variants are enriched by removing the unbound ones. This process, called biopanning, can be realized in different ways. Established selection techniques such as ribosome (Hanes & Plückthun, 1997) (Plückthun, 2012), mRNA (Roberts & Szostak, 1997) (Barendt *et al.*, 2013) and CIS display (Odegrip *et al.*, 2004) are based on direct molecular coupling of protein variants with corresponding genetic information. Other display techniques include the exposure of protein or peptide variants on the surface of bacteriophages (Smith, 1985) (Friedeman *et al.*, 2007), yeast or bacterial cells (Boder & Wittrup, 1997) (Francisco *et al.*, 1993), and mammalian cells (Bowers *et al.*, 2014).

In particular, cell-free technologies such as ribosome display offer the advantage of being entirely *in vitro* and independent of the transformation efficiencies of living cells (Figure 4). This allows the use of comparably large libraries of up to 10^{14} variants, whereas the typical transformation efficiencies of *Escherichia coli* (*E. coli*) would limit the libraries to sizes of about 10^{10} variants (Amstutz *et al.*, 2001). Ribosome display was first described by Mattheakis and further developed by Plückthun (Mattheakis *et al.*, 1994) (Hanes & Plückthun, 1997). In ribosome display, the coupling of the genotype and phenotype of the library variants is realized by stabilization of ternary complexes consisting of ribosome, mRNA, and corresponding nascent protein formed during translation. The DNA coding for the protein variants is genetically fused with a short spacer and an anchor sequence (stalling) and is subsequently transcribed *in vitro* (Kunamneni *et al.*, 2020). During translation, anchor sequences such as the SecM sequence prevent peptide bond formation within the peptidyl transferase center of the ribosomes (Gumbart *et al.*, 2012). The absence of stop codons within the sequence avoids the dissociation of ternary complexes consisting of mRNA, translated protein and ribosome. The spacer sequence allows the translated protein to completely exit the ribosomal exit tunnel and the ternary complexes can be used directly for the selection process.

A successful selection strategy includes not only the presentation of a defined target amount but also the number and duration of wash steps and a defined method for obtaining the genetic information of the enriched variants. In addition, various techniques are available to select variants with regard to affinity and epitope specificity (Alfaleh *et al.*, 2020). For example, after the variants are bound to immobilized targets, non-immobilizable proteins can be added (Dreier & Plückthun, 2018). Depending on their dissociation rates, the library variants form a new equilibrium. By removing the subsequently added targets before an equilibrium is reached, variants with a slow dissociation are favored (off-rate selection). To favor specific epitopes, targets can be preincubated with masking elements such antibodies, binding proteins, peptides, or ligands (Ditzel, 2002). In specific cases, adjustment of buffer conditions or addition of detergents may also be necessary for a functional conformation of targets or library variants. After selection, the genetic information of bound protein variants can be obtained as mRNA using various methods. A distinction between the direct release of genetic information and the release of ternary complexes followed by RNA extraction can be made. The ternary complexes can be dissociated by chelation of Mg^{2+} ions required to maintain stability (Guo *et al.*, 2011). The uncoupling of the phenotype from the genotype of the variants allows the reuse of the entire genetic material while the phenotypes remain bound to the targets. In addition, competitive elution methods can also be used to dissociate ternary complexes and obtain variants that bind

specific epitopes (Zahnd *et al.*, 2007). After elution, the mRNA is converted to complementary DNA (cDNA) by reverse transcription (RT-PCR). This represents the starting point for a further selection round. In addition to target binding, other protein properties can also be addressed by adjusting the ribosome display. Methods for improving thermal stability or affinity of library variants have been described in literature (Buchanan, 2012). Moreover, additional diversity can be generated by error-prone PCR or gene shuffling between two rounds of selection (Hanes *et al.*, 2000). However, it should be noted that the non-covalent coupling of genotypes and phenotypes as well as the weak thermal stability of ternary complexes can limit the wide application range of this display method.
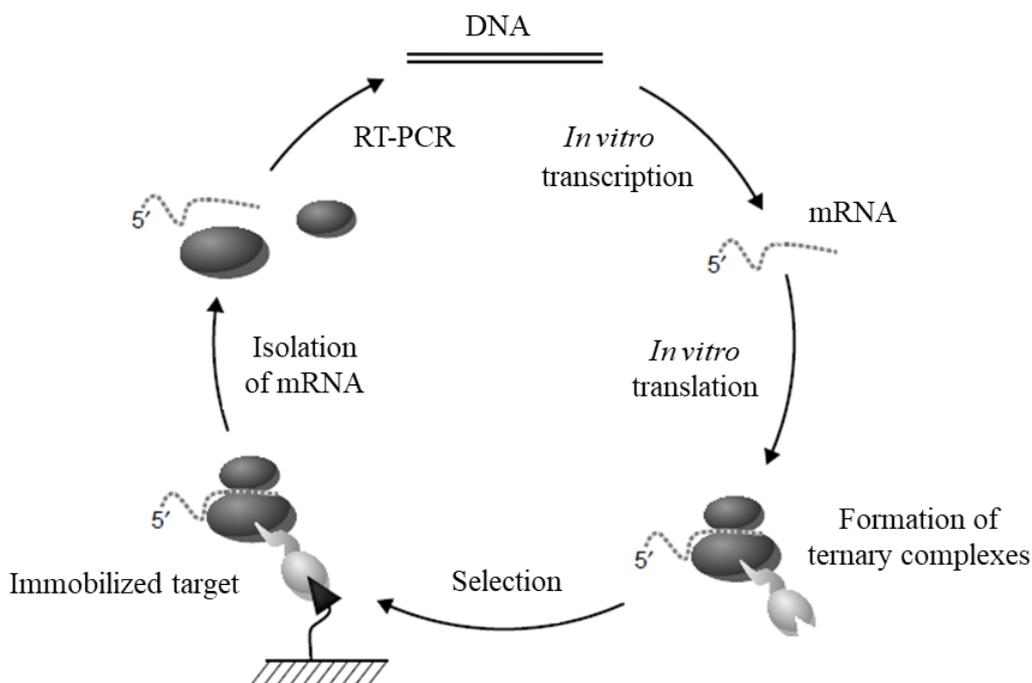
Figure 4: Schematic illustration of ribosome display (Amstutz *et al.*, 2001).
First, the DNA library is transcribed and translated *in vitro*. Ternary complexes are formed consisting of mRNA, ribosome, and nascent protein. The formed ternary complexes are selected versus immobilized targets and mRNA of bound complexes is eluted and isolated. To reuse the library in a further selection round, eluted mRNA is transcribed into complementary DNA (cDNA) by reverse transcription (RT-PCR).

1.4 Aim of this thesis

The B domain of Protein A is the main object of investigation in this work. First, the influence of loops on folding and function of domain B is studied to determine the accessibility of these structural elements for engineering. For this, loops of domain B are partially or completely substituted for $(GS)_n$ linkers and variants obtained from this rational engineering approach are produced and characterized by CD and SPR.

Second, IgG-binding variants of domain B are generated applying helix shuffling and loop substitution. For this, the order of the helices 1 and 3 is swapped and two alternative loops with several randomization positions are inserted to generate diversity and connect the rearranged helices of the bundle (Figure 5). Next, DNA libraries are generated and IgG-specific proteins enriched by ribosome display. The obtained protein pools are analyzed by ELISA-based screening and individual variants are produced on larger scale to characterize their function and structure by ELISA, SPR, DSC, CD, and NMR.

To investigate the suitability of the novel IgG-binding proteins as affinity ligands, one representative variant is coupled on the surface of a column matrix. Subsequently, capturing of hIgG1 from different cell extracts is analyzed and binding capacity as well as elution pH of the ligand determined.
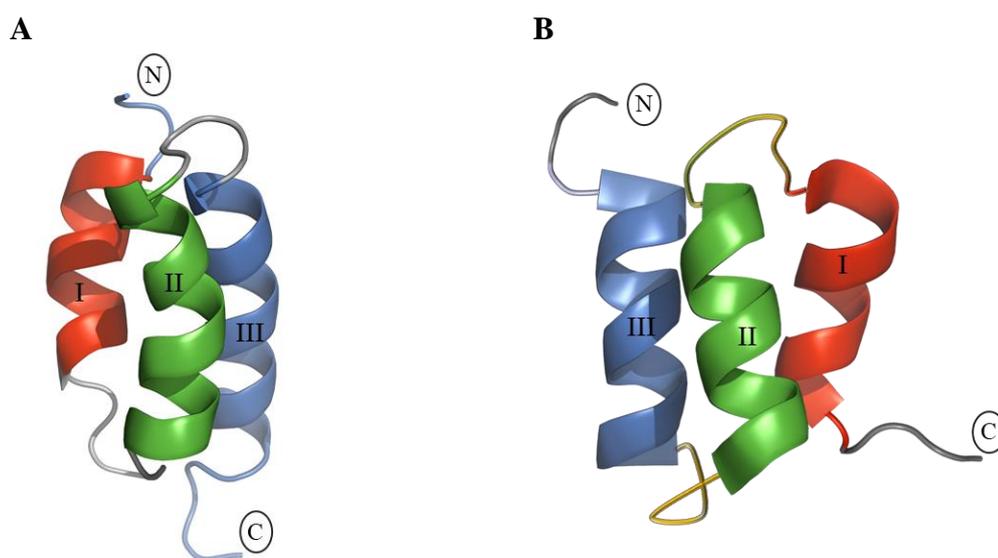
Figure 5: Helix shuffling strategy for the generation of novel domain B variants.
(**A**) Structural elements of wild type domain B (PDB: 1BDD) in their original order. Starting from the N-terminus (N), the three helices are numbered as follows: helix 1 (I, red), helix 2 (II, green), and helix 3 (III, blue). Helix connecting loops are colored gray. (**B**) Applied helix shuffling strategy based on domain B. Compared to wild type domain B, the order of the helices 1 and 3 is swapped. Alternative loops (yellow) with several randomization positions are inserted to form a three-helix bundle and generate diversity. The figure was created as cartoon representation using the software PyMOL.

2 Material and methods

2.1 Material

2.1.1 Chemicals

All chemicals used in this work had at least p. a. purity, unless explicitly stated, and are summarized in Table 1.

Table 1: List of chemicals

Chemical	Manufacturer
Acetic acid	VWR (Darmstadt, GER)
Acetone	Carl Roth (Karlsruhe, GER)
AEBSF	AppliChem (Darmstadt, GER)
Agar agar	Carl Roth (Karlsruhe, GER)
Agarose	Biozym (Hessisch Oldendorf, GER)
Agarose UltraPure	Thermo Scientific (Schwerte, GER)
Albumin fraction V	AppliChem (Darmstadt, GER)
Ammonium-^{14}N chloride	Carl Roth (Karlsruhe, GER)
Ammonium-^{15}N chloride	CortecNet (Les Ulis, FRA)
Ammonium sulphate	Carl Roth (Karlsruhe, GER)
Antifoam SE-15	Sigma Aldrich (Steinheim, GER)
Ampicillin sodium salt	Carl Roth (Karlsruhe, GER)
Arginine monohydrochloride	VWR (Darmstadt, GER)
ATP	NEB (Frankfurt am Main, GER)
Bis-Tris	Carl Roth (Karlsruhe, GER)
Bromophenol blue	VWR (Darmstadt, GER)
Calcium chloride	Carl Roth (Karlsruhe, GER)
Citric acid	VWR (Darmstadt, GER)
Cobalt(II) chloride hexahydrate	VWR (Darmstadt, GER)
Coomassie brilliant blue R 250	VWR (Darmstadt, GER)
Copper(II) sulphate pentahydrate	Carl Roth (Karlsruhe, GER)
Deuterium oxide	CortecNet (Les Ulis, FRA)
D-glucose-^{13}C6	CortecNet (Les Ulis, FRA)
D-glucose monohydrate	Carl Roth (Karlsruhe, GER)
Disodium hydrogen phosphate	VWR (Darmstadt, GER)
DTT	AppliChem (Darmstadt, GER)
EDTA	VWR (Darmstadt, GER)
Ethanol	Carl Roth (Karlsruhe, GER)
Guanidinium hydrochloride	NIGU (Waldkraiburg, GER)
Glycerol	Carl Roth (Karlsruhe, GER)
Glycine	AppliChem (Darmstadt, GER)

Chemical	Manufacturer
Heparin	Sigma Aldrich (Steinheim, GER)
HEPES	AppliChem (Darmstadt, GER)
Hydrochloric acid	VWR (Darmstadt, GER)
Imidazole	VWR (Darmstadt, GER)
IPTG	Carl Roth (Karlsruhe, GER)
Iron(III) chloride hexahydrate	Carl Roth (Karlsruhe, GER)
Kanamycin sulphate	Carl Roth (Karlsruhe, GER)
Lactose	Sigma Aldrich (Steinheim, GER)
Magnesium chloride	VWR (Darmstadt, GER)
Magnesium sulphate	Carl Roth (Karlsruhe, GER)
Manganese sulphate monohydrate	Carl Roth (Karlsruhe, GER)
MES	Sigma Aldrich (Steinheim, GER)
MOPS	Carl Roth (Karlsruhe, GER)
Nickel(II) chloride hexahydrate	Carl Roth (Karlsruhe, GER)
Peptone from casein	Carl Roth (Karlsruhe, GER)
Potassium chloride	Carl Roth (Karlsruhe, GER)
Potassium dihydrogen phosphate	Carl Roth (Karlsruhe, GER)
Potassium hydrochloride	Carl Roth (Karlsruhe, GER)
SDS	Carl Roth (Karlsruhe, GER)
Sodium acetate trihydrate	Carl Roth (Karlsruhe, GER)
Sodium azide	Carl Roth (Karlsruhe, GER)
Sodium chloride	Carl Roth (Karlsruhe, GER)
Sodium hydroxide	Carl Roth (Karlsruhe, GER)
Sulfuric acid	Carl Roth (Karlsruhe, GER)
Thiamine hydrochloride	Carl Roth (Karlsruhe, GER)
Tris	Carl Roth (Karlsruhe, GER)
Tween 20	VWR (Darmstadt, GER)
Urea	Carl Roth (Karlsruhe, GER)
Water (RNase/DNase-free)	VWR (Darmstadt, GER)
Water (HiPerSolv for HPLC)	VWR (Darmstadt, GER)
Yeast extract	Carl Roth (Karlsruhe, GER)
Zinc sulphate heptahydrate	Carl Roth (Karlsruhe, GER)

2.1.2 Consumables

Table 2: List of consumables

Material	Type	Manufacturer
96-deepwell plate	MASTERBLOCK, PP, V-bottom	Greiner (Frickenhausen, GER)
96-deepwell plate	KingFisher Flex, PP, V-bottom	Fisher Scientific (Schwerte, GER)
96-well plate	NUNC, PP, fritted	Fisher Scientific (Schwerte, GER)
96-well plate	NUNC, PS, F-bottom	Fisher Scientific (Schwerte, GER)

Material	Type	Manufacturer
384-well plate	PP, V-bottom	Greiner (Frickenhausen, GER)
384-well plate	Black, PS, high binding, F-bottom	Greiner (Frickenhausen, GER)
384-well plate	Black, PS, F-bottom	Greiner (Frickenhausen, GER)
Centrifugal filter unit	Amicon Ultra, MWCO 3 kDa	Fisher Scientific (Schwerte, GER)
CM5 chip	High-capacity amine	Bruker Daltonik (Bremen, GER)
Autoclavable bag	200 mm×300 mm	Schubert (Leipzig, GER)
Biosensor	Octet Ni-NTA	ForteBio (Göttingen, GER)
Culture tube	17 mm×100 mm, 14 mL	Schubert (Leipzig, GER)
Centrifuge tube	50-500 mL	Beckman Coulter (Krefeld, GER)
Comb	KingFisher 12 tip comb	Fisher Scientific (Schwerte, GER)
Crimp neck vial	1.5 mL	VWR (Darmstadt, GER)
Crimp cap	Aluminum with central hole	VWR (Darmstadt, GER)
Cuvette	BRAND UV cuvette micro, Suprasil 110-QS	VWR (Darmstadt, GER) Hellma Optik (Jena, GER)
Dynabeads	Streptavidin M270, Neutravidin M270	Invitrogen (Darmstadt, GER)
Foil for microtiter plates	Air pore (breathable), Tape pad (unbreathable)	QIAGEN (Hilden, GER)
Gene pulse cuvette	0.1 cm for electroporation	VWR (Darmstadt, GER)
Glove	Nitrile	Schubert (Leipzig, GER)
Membrane disc filter	Supor-200, 0.2 μm, Ø 47 mm	VWR (Darmstadt, GER)
NMR tube	5 mm	New Era Enterprise Inc. (Vineland, NY, USA)
PCR cap	Domed 10×12 stripes	VWR (Darmstadt, GER)
PCR plate	48-well and 96-well	Fisher Scientific (Schwerte, GER)
Petri dish	PS, Ø 92 mm×16 mm	Schubert (Leipzig, GER)
PhyNexus tips	200 μL	PhyNexus (San Jose, CA, USA)
Pipette tips	0.1-10 μL, 2-200 μL, 50-1,000 μL	VWR (Darmstadt, GER)
Plastic vial	7 mm, 15 mm	GE Healthcare (Solingen, GER)
Plate lid	Clear polystyrene	Greiner (Frickenhausen, GER)
QTray	Polystyrene with cover	MolDev (Biberach, GER)
Reaction tube	Protein LoBind 0.5-5 mL, DNA LoBind 0.5-5 mL	VWR (Darmstadt, GER)
Reagent reservoir	100 mL single use	VWR (Darmstadt, GER)
Rubber cap	Type 3/5	GE Healthcare (Solingen, GER)
SDS-PAGE gel	NuPAGE Novex 4-12 % Bis-Tris gel, 10-17 wells	Invitrogen (Darmstadt, GER)
Serological pipette	5-50 mL, sterile	Schubert (Leipzig, GER)

Material	Type	Manufacturer
Shaking flask	0.25-1.25 L	VWR (Darmstadt, GER)
	Ultra Yield 2.5 L	Biotrend (Cologne, GER)
Spreader	L-shaped, disposable	VWR (Darmstadt, GER)
Syringe filter	0.22-0.45 µm (PES membrane)	Schubert (Leipzig, GER)
Syringe	1-5 mL, sterile, single use	Schubert (Leipzig, GER)
Weighing pan	ROTILABO white, 330 mL	Carl Roth (Karlsruhe, GER)

2.1.3 Buffer and solutions

After complete dissolution of the solids, the pH was adjusted with NaOH or HCl at room temperature. The maximum deviation was +/- 0.1 pH units. All buffers and solutions used for chromatography were filtered before use (pore size 0.2 µm) and stored at room temperature. Buffers used for ribosome display were prepared with RNase/DNase-free water (VWR) according to RNase-free guidelines.

Table 3: List of stock solutions

Buffer	Composition	pH value
5x SDS-PAGE loading buffer	65 % Glycerol (v/v)	8.0
	250 mM Tris	
	5 % SDS (w/v)	
	0.12 % Bromophenol blue (v/v)	
10x PBS	1.37 M NaCl	6.7
	27 mM KCl	
	80 mM Na_2HPO_4	
	20 mM KH_2PO_4	
20x MES	1 M MES monohydrate	7.4
	1 M Tris	
	20 mM EDTA	
	2 % SDS (w/v)	
50x TAE buffer	2 M Tris	8.5
	1 M Acetic acid	
	125 mM Na-EDTA	

Table 4: List of buffers used for ribosome display

Buffer	Composition	pH value
EB buffer	50 mM Tris acetate	7.5
	150 mM NaCl	
	25 mM EDTA	
	50 µg/mL *S. cerevisiae* RNA	

Buffer	Composition	pH value
Stop buffer	50 mM Tris acetate 150 mM NaCl 50 mM Magnesium acetate 0.5 % BSA (w/v) 2.5 mg/mL Heparin	7.5
TBST buffer	150 mM NaCl 50 mM Tris 10 % Tween 20 (v/v)	7.5
WBT buffer	50 mM Tris acetate 150 mM NaCl 50 mM Magnesium acetate 10 % Tween 20 (v/v)	7.5

Table 5: List of buffers used for chromatography

Buffer	Composition	pH value
AIC buffer A	1x PBS 0.1 % Tween 20 (v/v)	7.4
AIC buffer B	100 mM Citrate	3.0
AIC buffer C	1 M Tris	10.0
IMAC buffer A	50 mM NaH_2PO_4 300 mM NaCl 20 mM Imidazole 1 mM DTT	7.4
IMAC buffer B	50 mM NaH_2PO_4 300 mM NaCl 500 mM Imidazole 1 mM DTT	7.4

Table 6: List of other buffers and solutions

Buffer	Composition	pH value
1x PBS	1:10 dilution of 10x PBS	7.4
BLI buffer	1x PBS 0.02 % Tween 20 (v/v) 0.1 % BSA (w/v) 20 mM Imidazole	7.4
Blocking buffer (ELISA)	1x PBS 3 % BSA (w/v) 0.5 % Tween 20 (v/v)	7.4
CD buffer	20 mM NaH_2PO_4	7.4
Destaining solution	10 % Ethanol (v/v) 10 % Acetic acid (v/v)	

Buffer	Composition	pH value
Lysis buffer	1x PBS or AIC buffer A	7.4
	250 µg/mL Lysozyme	
	1 mM AEBSF	
	2.5 mM MgCl$_2$/MgSO$_4$	
	20 U/µL Benzonase	
SPR buffer	1x PBS	7.4
	0.05 % Tween 20 (v/v)	
Staining solution	40 % Ethanol (v/v)	
	10 % Acetic acid (v/v)	
	0.1 % Coomassie R 250 (w/v)	
ELISA buffer A	1x PBS	7.4
	0.1 % Tween 20 (v/v)	
ELISA buffer B	1x PBS	7.4

2.1.4 Materials used for chromatography

Table 7: List of materials used for chromatography

Column	Size	Manufacturer
HiLoad 16/60 Superdex 75 prep grade	120 mL	GE Healthcare (Munich, GER)
HiLoad 16/60 Superdex 200 prep grade	120 mL	GE Healthcare (Munich, GER)
HiLoad 26/60 Superdex 75 prep grade	320 mL	GE Healthcare (Munich, GER)
HiLoad 26/60 Superdex 200 prep grade	320 mL	GE Healthcare (Munich, GER)
HiTrap Desalting	5 mL	GE Healthcare (Munich, GER)
HiTrap NHS-Activated HP	1 mL	Cytiva (Munich, GER)
HisTrap HP	5 mL	GE Healthcare (Munich, GER)
IgG Sepharose 6 Fast Flow	50 mL	GE Healthcare (Munich, GER)
Ni Sepharose 6 Fast Flow	50 mL	GE Healthcare (Munich, GER)
Sephadex G-25	50 mL	Cytiva (Munich, GER)
Tricorn 5/150 GL Superdex Increase 75 prep grade	3 mL	GE Healthcare (Munich, GER)
Tricorn 5/150 GL Superdex Increase 200 prep grade	3 mL	GE Healthcare (Munich, GER)

2.1.5 Standards, markers and kits

Standards, markers and kits were used according to the manufacturer's instructions unless otherwise stated.

Table 8: List of standards and markers

Material	Manufacturer
100 bp DNA Ladder	NEB (Frankfurt am Main, GER)
Gel Filtration Calibration Standard	GE Healthcare (Munich, GER)

Material	Manufacturer
Gel Filtration Standard (HPLC)	Bio-Rad (Hercules, CA, USA)
PageRuler Unstained Broad Range Protein	Fisher Scientific (Schwerte, GER)
Quick-Load 2-Log DNA Ladder (0.1-10 kb)	NEB (Frankfurt am Main, GER)
Quick-Load 1 kb DNA Ladder	NEB (Frankfurt am Main, GER)

Table 9: List of kits and ready-to-use solutions

Material	Manufacturer
Coomassie Plus Protein Assay Kit	Pierce (Rockford, IL, USA)
DNA Clean & Concentrator-5	Zymo Research (Freiburg, GER)
E. coli S30 Extract System for Linear Templates	Promega (Walldorf, GER)
EZ-Link Sulfo-NHS-LC-Biotinylation Kit	Fisher Scientific (Schwerte, GER)
Gel Loading Dye Purple (6x)	NEB (Frankfurt am Main, GER)
GelRed	VWR (Darmstadt, GER)
HiScribe T7 High Yield RNA Synthesis Kit	NEB (Frankfurt am Main, GER)
KAPA HiFi HotStart ReadyMix	KapaBiosystems (Wilmington, DE, USA)
NHS/EDC Coupling Kit	Bruker Daltonik (Bremen, GER)
NuPAGE MES SDS Running Buffer (20x)	Invitrogen (Carlsbad, CA, USA)
QIAprep Spin Miniprep Kit	QIAGEN (Hilden, GER)
Quick Ligation Kit	NEB (Frankfurt am Main, GER)
RNA Clean & Concentrator-5	Zymo Research (Freiburg, GER)
RNA Clean & Concentrator-25	Zymo Research (Freiburg, GER)
RNA from baker's yeast (*S. cerevisiae*)	Sigma Aldrich (Steinheim, GER)
SYBR Safe DNA Gel Stain	Fisher Scientific (Schwerte, GER)
TMB PLUS 2	Biotrend (Cologne, GER)
Zymoclean Gel DNA Recovery Kit	Zymo Research (Freiburg, GER)
ZymoPURE Plasmid Maxiprep Kit	Zymo Research (Freiburg, GER)

2.1.6 Primers and plasmids

All oligonucleotides were purchased from Microsynth Seqlab GmbH (Göttingen) in HPLC purity. A detailed overview of the plasmids used in this work can be found in the appendix (Figure A 1).

Table 10: List of primers

Primer	Sequence (5'-3')
Domain_B_fw	TAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACA TATGGCCGATAACAAATTCAACAAAGAACAG
Domain_B_rev	CTTGCTTCTGGCCACCAGATCCTTTCGGTGCCTGTGCATC
F1complete_fw	CATACGAAATTAATACGACTCACTATAGGGAGACCACA ACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAA GGAGATATACATATG

Primer	Sequence (5'-3')
nativeB_pNP-013_fw	TATAGGTCTCGGAGCATGGCCGATAACAAATTCAACAAAGAACAG
nativeB_pNP-013_rev	TATAGGTCTCGAACATTTCGGTGCCTGTGCATC
PAB2_BsaI_fw	GGTCTCAAATGCATGCATGCCAACAACGCCCATGCCAACAACGCCAATGAAGAACAGCGTAATGG
PAB2_BsaI_rev	GGTCTCTGCACTGGCGTTGTTGGCATGGGCGTTGTTGGCATGATCATCTTTCAGGCTCTGG
PACB_pNP-013_fw	TATAGGTCTCGGAGCATGGCCGATAACAAATTCAGC
PACB_pNP-013_rev	TATAGGTCTCGAACATTTCGGCAGATGCAG
PACB_3xVNN_fw	GCAAATCTGCTGGCCGAAGCAAAAAAACTGAATGATGCACAGVNNVNNVNNGAAGAACAGCGTAATGG
PACB_4xVNN_fw	GCAAATCTGCTGGCCGAAGCAAAAAAACTGAATGATGCACAGVNNVNNVNNVNNGAAGAACAGCGTAATGG
PACB_5xVNN_fw	GCAAATCTGCTGGCCGAAGCAAAAAAACTGAATGATGCACAGVNNVNNVNNVNNVNNGAAGAACAGCGTAATGG
PACB_6xVNN_rev	ATAGAAGGCGTTCTGCTGTTCTTTGTTNNBNNBNNBNNBNNBNNBATCATCTTTCAGGCTCTG
PACB_7xVNN_rev	ATAGAAGGCGTTCTGCTGTTCTTTGTTNNBNNBNNBNNBNNBNNBNNBATCATCTTTCAGGCTCTG
PACB_8xVNN_rev	ATAGAAGGCGTTCTGCTGTTCTTTGTTNNBNNBNNBNNBNNBNNBNNBNNBATCATCTTTCAGGCTCTG
PACB_fw2	TGTTTAACTTTAAGAAGGAGATATACATATGGCCGATAACAAATTCAGCCAGAGCGCAAATCTGCTGGCC
PACB_rev2	CTTGCTTCTGGCCACCAGATCCTTTCGGCAGATGCAGAATTTCATAGAAGGCGTTCTGCTG
PACB_ol_2H_rev	GCTCTGGATAAAACCATTACGCTG
T7 Promoter	TAATACGACTCACTATAGGG
T7 Terminator	GCTAGTTATTGCTCAGCGG
tolA-stem-rev	CCGCACACCAGTAAGGTGTGCGGTTTCAGTTGCCGCTTTCTTTCTTGC

Table 11: List of plasmids

Plasmid	Description	Reference
pNP-001	Expression vector, kanamycin resistance, T7 promoter	Navigo Proteins (Halle, GER)
pNP-002	Modified pNP-001, sequence for C-terminal Strep-tag II, kanamycin resistance, T7 promoter	Navigo Proteins (Halle, GER)
pNP-013	Modified pNP-001, sequence for N-terminal sfGFP, His10-tag, TVMV cleavage site, kanamycin resistance, T7 promoter	Navigo Proteins (Halle, GER)
pPR-IBA1b-tolA1	Plasmid template for tolA spacer, ampicillin resistance, T7 promoter	IBA (St. Louis, MO, USA)

2.1.7 Enzymes, inhibitors and antibodies

Enzymes, antibodies and inhibitors were used according to the manufacturer's instructions unless otherwise stated. Prior to use, the activity of recombinant TVMV protease was analyzed with already known proteins that possess a TVMV cleavage site (provided by Dr. Anja Katzschmann, Navigo Proteins).

Table 12: List of enzymes and inhibitors

Enzyme	Concentration	Manufacturer
Alkaline Phosphatase (CIP)	10 U/µL	NEB (Frankfurt am Main, GER)
Benzonase Nuclease	250 U/µL	VWR (Darmstadt, GER)
BsaI-HFv2	20 U/µL	NEB (Frankfurt am Main, GER)
DNase I	2 U/µL	Thermo Scientific (Schwerte, GER)
Lysozyme	30 U/µg	VWR (Darmstadt, GER)
OneTaq Quick-Load Master Mix	-	NEB (Frankfurt am Main, GER)
Phusion HF DNA Polymerase	2 U/µL	NEB (Frankfurt am Main, GER)
Q5 DNA Polymerase	2 U/µL	NEB (Frankfurt am Main, GER)
RNasin Ribonuclease Inhibitor	40 U/µL	Promega (Walldorf, GER)
Strep-Tactin-HRP	1,000 U/mg	IBA (Göttingen, GER)
T4 DNA Ligase	3 U/µL	Promega (Walldorf, GER)
TVMV protease (recombinant)	100 U/µL	Navigo Proteins (Halle, GER)
Transcriptor Reverse Transcriptase	20 U/µL	Roche (Penzberg, GER)

Table 13: List of antibodies and antibody fragments

Antibody	Description	Manufacturer
Avelumab-Fab	Papain cleaved	Navigo Proteins (Halle, GER)
Cetuximab (IgG1)	Erbitux	Sigma Aldrich (Steinheim, GER)
IgG1-Fc	Human	Navigo Proteins (Halle, GER)

2.1.8 Microorganisms

In this work, *E. coli* bacteria were grown for the expression of recombinant genes. Table 14 gives an overview of the used bacterial strains.

Table 14: List of bacterial strains

Strain	Derivate	Genotype	Reference
BL21(DE3)	B	*lon-11 Δ(ompT-nfrA)885 Δ(galM-ybhJ)884, λDE3 [lacI, lacUV5-T7 gene 1, ind1, sam7, nin5] Δ46, [mal⁺] $_{K-12}(λ^S)$, hsdS10*	Navigo Proteins (Halle, GER)
XL2Blue	K12	*endA1 gyrA96(nalᴿ) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB⁺ lacI�q Δ(lacZ)M15 Amy Cmᴿ] hsdR17(r$_K^-$ m$_K^+$)*	Agilent Technologies (Santa Clara, CA, USA)

2.1.9 Antibiotics and culture media

All culture media and components were sterilized by autoclaving at 121 °C for 20 min. Heat-labile components were sterile filtered and added to the cooled medium.

Table 15: List of antibiotics

Antibiotic	Stock solution	End concentration
Ampicillin	100 mg/mL	100 µg/mL
Kanamycin	50 mg/mL	50 µg/mL

Table 16: List of culture media

Culture media	Composition
H15 autoinduction medium	22 g/L D-glucose monohydrate
	8.9 g/L Glycerol
	7.6 g/L Lactose
	50 g/L Yeast extract
	52.3 g/L MOPS
	2 g/L $MgCl_2$
	0.2 mL/L Trace elements
LB agar	10 g/L Peptone
	5 g/L Yeast extract
	5 g/L NaCl
	15 g/L Agar agar
LBB autoinduction medium	10 g/L Tryptone
	5 g/L Yeast extract
	3.3 g/L $(NH_4)_2SO_4$
	6.8 g/L KH_2PO_4
	7.1 g/L Na_2HPO_4
	0.5 g/L D-glucose monohydrate
	2.0 g/L Lactose
	0.15 g/L $MgSO_4$
	0.03 g/L Trace elements from Formedium Ltd (UK)
M9 medium	12.8 g/L Na_2HPO_4
	3 g/L KH_2PO_4
	0.5 g/L NaCl
	0.25 g/L $MgSO_4$
	0.14 g/L $CaCl_2$
	1 g/L $^{15}NH_4Cl$
	2 g/L D-glucose-$^{13}C6$
	0.1 g/L Thiamine hydrochloride
	2 mL/L Trace elements

Culture media	Composition
SOC medium	20 g/L Peptone
	5 g/L Yeast extract
	0.6 g/L NaCl
	0.2 g/L KCl
TBY medium	20 g/L Peptone
	10 g/L Yeast extract
	10 g/L NaCl
	10 g/L $MgSO_4$
Trace elements	20 g/L Na_2EDTA
	17 g/L $FeCl_3$
	0.5 g/L $CaCl_2$
	0.2 g/L $CuSO_4$
	0.2 g/L $ZnSO_4$
	0.2 g/L $CoCl_2$
	0.1 g/L $MnSO_4$
2x YT medium	17 g/L Peptone
	10 g/L Yeast extract
	5 g/L NaCl
	0.41 g/L $MgCl_2$

2.1.10 Technical equipment

Table 17: List of technical equipment

Equipment	Type	Manufacturer
6-axis robot arm	LR Mate 200iC	FANUC (Neuhausen a.d.Fildern, GER)
Autoclave	Laboklav 55 MV	SHP Steriltechnik AG (Detzel, GER)
Barcode printer	CAB A4+/600	cab Produkttechnik (Sömmerda, GER)
CD spectrophotometer	J-810	Jasco (Pfungstadt, GER)
Centrifuge	Avanti J-30I	Beckman Coulter (Krefeld, GER)
	Biofuge fresco	Heraeus (Hanau, GER)
	5810R	Eppendorf (Hamburg, GER)
Centrifuge rotor	JA-25.50	Beckman Coulter (Krefeld, GER)
	75006445	Heraeus (Hanau, GER)
Chromatography system	ÄKTAxpress	GE Healthcare (Munich, GER)
Colony picker	QPix2	Molecular Devices (Biberach, GER)
Electrophoresis chamber	XCell SureLock Mini-Cel	Invitrogen (Carlsbad, CA, USA)
Electroporator	Genepulser II	Bio-Rad (Hercules, CA, USA)
Gel imaging system	G:BOX F3	Syngene (Frederick, MD, USA)
Heating block	SBH 130D/3	Stuart (Staffordshire, UK)

Equipment	Type	Manufacturer
HPLC	Ultimate 3000	Dionex (Sunnyvale, CA, USA)
Incubator	Kelvitron	Heraeus (Hanau, GER)
Liquid handling system	Freedom EVO	Tecan (Crailsheim, GER)
Peltier element	LFI-3751	Wavelength Electronics (Bozeman, MT, USA)
pH electrode	SenTix 81	WTW (Weilheim, GER)
pH meter	InoLab Level 1	WTW (Weilheim, GER)
Pipette	Eppendorf Research plus 10-1,000 µL	Eppendorf (Hamburg, GER)
Pipette controller	Accu-jet pro	Brand (Wertheim, GER)
Precision scale	BP 121 S	Satorius (Göttingen, GER)
Pure water system	Milli-Q Advantage A10	Millipore (Burlington, VT, USA)
Magnetic stirrer	MR 3000	Heidolph (Schwabach, GER)
Microplate reader	Infinite	Tecan (Crailsheim, GER)
Microplate washer	405	Biotek (Winooski, VT, USA)
Multi-channel pipette	Multipette plus Viaflow 300	Eppendorf (Hamburg, GER) INTEGRA Bioscience (Zizers, CH)
NMR spectrometer	Avance III 600 MHz, Avance III 800 MHz	Bruker BioSpin (Ettlingen, GER)
Octet	BLUE384	ForteBio (Fremont, CA, USA)
Safety workbench	HERAsafe	Heraeus (Hanau, GER)
Scale	572 TE601	Kern (Balingen-Frommern, GER) Satorius (Göttingen, GER)
Selection robot	KingFisher DUO	Thermo Scientific (Waltham, MA, USA)
Shaker	HT Ecotron Multitron RamBio Innova 43R	Infors (Bottmingen, CH) Infors (Bottmingen, CH) Applikon Inc. (Schiedam, NL) Eppendorf (Hamburg, GER)
Spectrophotometer	Nanodrop 2000c	Thermo Scientific (Waltham, MA, USA)
SPR device	SPR-32 Pro	Bruker Daltonik (Bremen, GER)
Thermocycler	Biometra Trio	Analytik Jena (Jena, GER)
Thermomixer	Comfort	Eppendorf (Hamburg, GER)
Ultrasonic homogenizer	Sonopuls HD 2200	Bandelin electronic (Berlin, GER)
Ultraturrax	T25 basic	IKA Labortechnik (Staufen, GER)
Vacuum pump	VP820	VWR (Darmstadt, GER)
Voltage source	PowerPac 1000	Bio-Rad (Hercules, CA, USA)

2.1.11 Software

Table 18: List of software

Software	Company
Chromeleon 6.80	Dionex (Sunnyvale, CA, USA)
Endnote X4	Clarivate (Philadelphia, PA, USA)
EVOware 2 Standard	Tecan (Crailsheim, GER)
Microsoft 365	Microsoft (Albuquerque, NM, USA)
Momentum 3.2.7	Fisher Scientific (Schwerte, GER)
NMRPipe	NIST, Frank Delaglio (Gaithersburg, MD, USA)
NMRViewJ 9.2	One Moon Scientific, Inc. (Newark, NJ, USA)
Octet Data Acquisition 7.1	ForteBio (Fremont, CA, USA)
Octet Data Analysis 7.1	ForteBio (Fremont, CA, USA)
PyMOL 0.99	DeLano Scientific (South San Francisco, CA, USA)
Sierra Analyzer 3.4.5	Bruker Daltonik (Bremen, GER)
Sierra SPR Control Software	Bruker Daltonik (Bremen, GER)
Sigma Plot 12.5	Systat Software (San Jose, CA, USA)
Spectra Manager I	Jasco (Pfungstadt, GER)
TALOS+	NIH, Yang Shen and Ad Bax (Bethesda, MD, USA)
Tecan Reader Software	Tecan (Crailsheim, GER)
TopSpin 3.6.4	Bruker BioSpin (Ettlingen, GER)
Unicorn 6.1	GE Healthcare (Munich, GER)

2.2 Methods

2.2.1 Molecular biological methods

2.2.1.1 DNA isolation from bacterial cells and agarose gels

Kits from Qiagen and Zymo Research were used for the isolation of DNA from bacterial cells according to the manufacturer's instructions. Depending on the scale, either the QIAprep Spin Miniprep Kit for 5 mL or the ZymoPURE Plasmid Maxiprep Kit for 150 mL preparations was used. XL2Blue cells were grown overnight at 37 °C and 220 rpm in an appropriate volume of 2x YT medium with antibiotics (Table 15). The Zymoclean Gel DNA Recovery Kit (Zymo Research) was used to isolate DNA from agarose gels after separation by agarose gel electrophoresis (2.2.1.4). DNA bands of expected size were excised from the agarose gel using a sterile scalpel. After gel extraction, DNA was eluted from the silica columns using RNase/DNase-free water (VWR). The concentration of the eluate was determined by UV/Vis spectroscopy as explained in the following section. The obtained DNA was stored at -20 °C for further use.

2.2.1.2 Photometric determination of nucleic acid concentration

The DNA concentration of solutions was determined by UV/Vis spectroscopy. A buffer-corrected spectrum of the DNA sample was measured at 260 nm using a spectrophotometer. The DNA concentration was calculated based on the Lambert-Beer's law (Equation 1). An extinction coefficient of 0.020 $(\mu g/mL)^{-1}$ cm^{-1} was used for double-stranded DNA, 0.027 $(\mu g/mL)^{-1}$ cm^{-1} for single-stranded DNA and 0.025 $(\mu g/mL)^{-1}$ cm^{-1} for single-stranded RNA (Lottspeich, 2006).

$$A_{260} = \varepsilon_{260} \cdot c \cdot d \qquad \text{(Equation 1)}$$

A_{260} = absorbance at 260 nm corrected by the extinction at 320 nm
ε_{260} = molar extinction coefficient at 260 nm $[\mu g/mL]^{-1}$ cm^{-1}
c = concentration $[ng/\mu L]$
d = path length of the beam path $[cm]$

2.2.1.3 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was used to identify and amplify DNA fragments (Mullis & Faloona, 1987). Phusion HF and Q5 Polymerases (NEB) were used for preparative and analytical approaches, while OneTaq Polymerase (NEB) was used for single colony PCR (scPCR). The reaction mixture was prepared on ice according to the manufacturer's instructions. The exact reaction conditions varied depending on the combination of template DNA, primers, and polymerase. The hybridization temperature was calculated based on the primer sequence.

The number of amplification cycles was adjusted depending on the polymerase and the size of DNA construct. Initial denaturation was followed by 20-25 cycles of denaturation, hybridization, DNA synthesis, and final amplification (Table 20). The elongation time was adjusted to template length and used DNA polymerase (Phusion HF and Q5: 30 s/kb, OneTaq: 60 s/kb). In addition, positive controls and empty vectors as negative controls were amplified. Table 19 and Table 20 give an overview of components and reaction conditions used for PCR. The obtained PCR products were separated by agarose gel electrophoresis (2.2.1.4).

Table 19: Components for 50 µL preparations

Reagent	Phusion HF	Q5	OneTaq
Polymerase buffer	1x	1x	1x
10 mM dNTPs	0.2 mM	0.2 mM	0.2 mM
100 µM Primer 1	0.5 µM	0.5 µM	0.2 µM
100 µM Primer 2	0.5 µM	0.5 µM	0.2 µM
Template DNA	1-5 ng	1-5 ng	single colony
2 U/µL Polymerase	2 U	2 U	2.5 U

Table 20: Overview of PCR reaction conditions

PCR step	Time [s]	Temperature [°C]	
Initial denaturation	30	95-98	
Denaturation	10	95-98	
Hybridization	20	55-72	20-25 cycles
Amplification	15-60	72	
Final amplification	120	72	

Overlap extension PCR (oePCR)

Overlap extension PCR (oePCR) was used to generate complex gene fragments (Hilgarth & Lanigan, 2019). In an initial PCR, fragments with complementary sequence overlaps at the 3' and 5' ends were amplified. Using overlap PCR, fragments were annealed at equimolar amounts of 100 ng/kb, and later extended and amplified using Phusion HF Polymerase (NEB) according to the reaction conditions listed in Table 19 and Table 20. Due to different annealing temperatures of DNA fragments and primers, both overlap and extension PCR were performed stepwise with ten amplification cycles each. The full-length gene fragments were isolated (2.2.1.1) and used for ribosome display (2.2.2).

Single colony PCR (scPCR)

ScPCR was performed to determine the presence of insert DNA in plasmid constructs. For this, a 20 µL ready-to-use reaction mixture with OneTaq Polymerase (NEB) was used per clone (Table 19). After transformation (2.2.3.2), individual colonies were picked with a toothpick, inoculated on a new agar plate with appropriate antibiotics (Table 15) and added directly to the PCR mixture. PCR was performed (Table 20) and the DNA of clones that showed a band at the expected level in the agarose gel (2.2.1.4) was sequenced (2.2.1.7).

2.2.1.4 Agarose gel electrophoresis

Separation of DNA fragments for analytical and preparative purposes was performed using agarose gel electrophoresis (Meyers *et al.*, 1976). DNA samples were mixed with 6x Gel Loading Dye Purple (NEB) prior to electrophoretic separation. For fragments $\leq 1,000$ bp, gels containing 1.6 % (w/v) Plaque Agarose (Biozym) were used. Fragments $> 1,000$ bp were separated using 1 % (w/v) UltraPure Agarose (Thermo) gels. All gels were prepared with 1x TAE running buffer. Electrophoretic separation was performed in horizontal gels for 1 h at a constant voltage of 5 V/cm. Quick-Load 1 kb DNA Ladder (NEB) was used as standard. To visualize DNA fragments after electrophoresis, agarose gels were panned in a freshly prepared GelRed solution for 30 min (VWR, 0.2 µL/mL in ddH$_2$0) and analyzed using a gel imaging system (Syngene). For preparative approaches, the gel was incubated in a freshly prepared SYBR Safe DNA Gel Stain solution (Fisher Scientific, 0.4 µL/mL in ddH$_2$0) for 20 min. Subsequently, DNA fragments were extracted from the gel (2.2.1.1).

2.2.1.5 Enzymatic cleavage and dephosphorylation of DNA

DNA was enzymatically cleaved to linearize circular plasmid DNA and obtain fragments for insertion into expression vectors (Nathans & Smith, 1975). Enzymatic cleavage was performed according to the manufacturer's instructions (NEB). In a 100 µL reaction, 2 µg DNA and 20 U/µg BsaI-HFv2 were mixed and incubated for 1 h. Subsequently, restriction enzymes were inactivated at 65 °C for 20 min and 1 U/µg Alkaline Phosphatase (NEB) was added to dephosphorylate the 5' ends in order to reduce religation of insert-free plasmids. Dephosphorylation was performed at 37 °C for at least 30 min. Finally, DNA fragments were separated by agarose gel electrophoresis (2.2.1.4) and isolated by gel extraction (2.2.1.1).

2.2.1.6 Ligation of DNA

Using ATP, DNA ligases catalyze the formation of phosphodiester bonds between fragments with complementary bases (Martin & MacNeill, 2002). DNA fragments with complementary overhangs obtained from enzymatic cleavage (2.2.1.5) were ligated with T4 DNA Ligase (Promega) at equimolar concentrations. A ratio of 1:3 (vector:insert) was used for the ligation of insert DNA with linearized plasmids. In addition, 5 mM ATP was added to the reaction mixtures before incubation at 16 °C for 16 h. After ligation, samples were purified using the DNA Clean & Concentrator-5 Kit (Zymo Research) according to the manufacturer's instructions. The purified ligations were diluted (1:10 in ddH$_2$O) and 1 µL was used to transform *E. coli* cells (2.2.3.2).

2.2.1.7 Sequencing of DNA

Sequencing of plasmid DNA and PCR products was performed by Microsynth Seqlab GmbH (Göttingen) based on the Sanger method (Sanger *et al.*, 1977).

Next-generation sequencing (NGS)

DNA libraries obtained after biopanning (2.2.2.1) were analyzed using NGS (GATC Biotech AG). In this so-called NGSelect Amplicon Adaptor Ligation, DNA fragments were amplified using the KAPA HiFi HotStart Polymerase (KapaBiosystems), which is included in a ready-to-use mixture. An overview of components and reaction conditions typical for this PCR amplification is given in Table 21 and Table 22. After PCR, samples were purified using the DNA Clean & Concentrator-5 Kit (Zymo Research) according to the manufacturer's instructions. By the so-called Illumina method, fragmented template DNA was covalently bound to a glass slide where the sequencing reaction took place. Starting from a bound DNA template, clusters of identical molecules were formed by bridge amplification. DNA sequencing was visualized by real-time detection of specific incorporated fluorescently labeled nucleotides (Canard & Sarfati, 1994).

Table 21: Components for PCR reactions (NGS)

Reagent	Final concentration
2x KAPA HiFi HotStart ReadyMix	1x
100 µM Primer 1	0.3 µM
100 µM Primer 2	0.3 µM
Template DNA	1-5 ng

Table 22: Overview of reaction conditions for initial PCR amplification (NGS)

PCR step	Time [s]	Temperature [°C]	
Initial denaturation	180	95	
Denaturation	20	98	
Hybridization	15	55-72	15-35 cycles
Amplification	15-60/kb	72	
Final amplification	60/kb	72	

2.2.2 Evolutionary *in vitro* selection of proteins by ribosome display (RD)

Evolutionary *in vitro* selection systems such as ribosome display (RD) can be used (Hanes & Plückthun, 1997) to enrich proteins with desired properties out of large libraries. This process is based on the coupling of phenotypes of library variants with their corresponding genotypes by formation of ternary complexes. First, stable ternary complexes are formed from nascent peptide chains (phenotype), ribosomes, and mRNA (genotype). Subsequently, these complexes are selected with regard to binding to an immobilized target. To obtain the genetic information after selection, ternary complexes are destroyed and mRNA is isolated. Finally, protein variants are obtained as DNA library by reverse transcription followed by cDNA amplification. The DNA library is used in a further selection round or is subcloned in an expression vector (2.2.2.5). The selection procedure is explained in more detail in the following sections and Table 24 gives an overview of the selection scheme.

2.2.2.1 Construction of domain B libraries

A gene fragment that encodes wild type domain B was used as template for the construction of nine libraries. Modular segments were defined at DNA level, named constant intermediate fragment (IF), randomized fragment (RF), start fragment (SF), and end fragment (EF). Compared to wild type domain B, IF1 and IF3 were swapped and both RF1 and RF2 were inserted to generate diversity (Figure 6, Table A 1). Depending on the length of RF1 and RF2, the libraries contained 9-13 randomized positions with 16 allowed amino acids per position (Δcysteine, Δphenylalanine, Δtryptophan, and Δtyrosine), resulting in a theoretical diversity of 1.8×10^{10}-4.5×10^{15} domain B variants at protein level (Table 23). First, IF2 was amplified by PCR (2.2.1.3). VNN oligonucleotides (wobbles) from Microsynth Seqlab GmbH were used in a second PCR to insert and assembly the gene fragments SF, IF1, IF3, EF, RF1 and RF2. For primer synthesis, prefabricated trinucleotides were used to avoid certain amino acids (Virnekas *et al.*, 1994). Codons of cysteine, phenylalanine, tryptophan, and tyrosine were excluded from the synthesis. After assembly of the complete gene fragment, libraries with comparable theoretical diversity at protein level were mixed in equal proportions (Table 23).
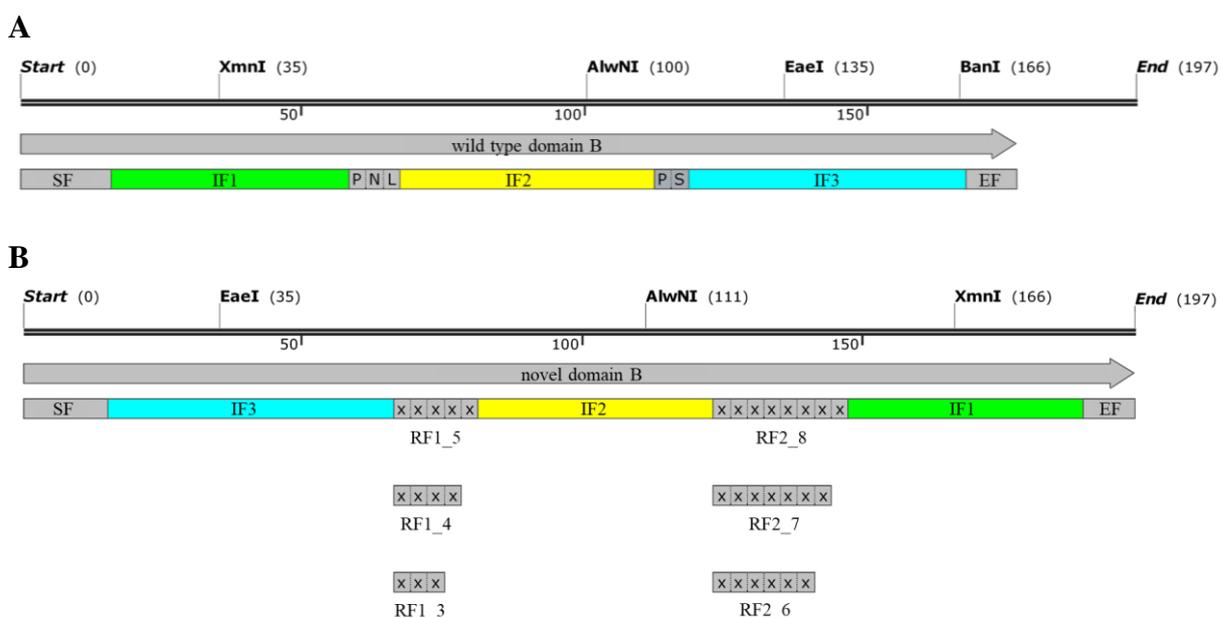
Figure 6: Modular design of generated domain B libraries at DNA level.
(**A**) Schematic representation of the gene fragment that encodes wild type domain B used as template for library generation. Modular segments are defined as constant intermediate fragment (IF), start fragment (SF), and end fragment (EF). Gene fragments encoding amino acids in the natural loops are separately labeled (PNL and PS). (**B**) Gene fragment of domain B libraries. Compared to wild type domain B, the order of IF1 and IF3 is swapped and the randomized regions RF1 (three, four, or five degenerate trinucleotides) and RF2 (six, seven, or eight degenerate trinucleotides) are inserted to generate diversity.

Table 23: Theoretical diversity of generated domain B libraries at protein level

Library	# Randomized positions in RF1	# Randomized positions in RF2	Theoretical diversity at protein level
PACB1	3	6	1.8×10^{10}
PACB2	3	7	1.1×10^{12}
	4	6	1.1×10^{12}

Library	# Randomized positions in RF1	# Randomized positions in RF2	Theoretical diversity at protein level
PACB3	3	8	1.7×10^{13}
	4	7	1.7×10^{13}
	5	6	1.7×10^{13}
PACB4	4	8	2.8×10^{14}
	5	7	2.8×10^{14}
	5	8	4.5×10^{15}

In order to use the generated DNA libraries in ribosome display, the 5' regulatory sequences such as the T7 promoter and the ribosome binding site (RBS) as well as the tolA spacer fragment were assembled, amplified by oePCR (2.2.1.3), and attached (Figure 7). The plasmid pPR-IBA1b-tolA1 was used as template to amplify the tolA spacer region. The obtained DNA was separated by agarose gel electrophoresis (2.2.1.4) and isolated by gel extraction (2.2.1.1). At RNA level, RNase digestion of the construct was blocked by 5' and 3' hairpin structures (stem loops) (Zahnd *et al.*, 2007).
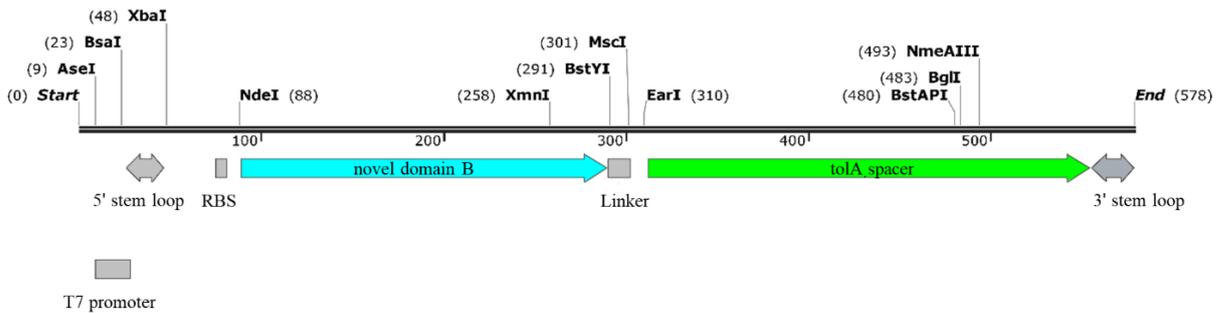
Figure 7: Modular design of domain B libraries with regulatory elements for ribosome display.
The DNA fragment generated by oePCR contains genes encoding the T7 promoter and the ribosome binding site (RBS) as well as inert and structure-breaking sequences. The stop codon of domain B was removed from the sequence to avoid a dissociation of the construct. At RNA level, RNase digestion of the construct was blocked by 5' and 3' hairpin structures.

The achieved diversity of the generated domain B libraries at DNA level was calculated based on the amount of purified DNA (Equation 2).

$$DNA_{molecules} = \frac{DNA_{amount} \cdot N_A}{DNA_{length} \cdot MW_{bp} \cdot CF} \qquad \text{(Equation 2)}$$

DNA_{amount} = amount of DNA amplicon [ng]
DNA_{length} = length of DNA amplicon [bp]
N_A = Avogadro's constant [molecules/mol]
MW_{bp} = average molecular weight of one base pair [g/mol]
CF = conversion factor

The following steps including *in vitro* transcription and translation (2.2.2.2), selection (2.2.2.3) and reverse transcription (2.2.2.4) were performed according to RNase-free guidelines.

2.2.2.2 *In vitro* transcription and translation

In this work, *in vitro* transcription and translation of the generated domain B libraries were performed separately and sequentially. This uncoupled approach allowed the accurate determination of the ratio of mRNA molecules and ribosomes and thus the efficient formation of ternary complexes. For mRNA synthesis, 0.5-1 µg DNA per 20 µL reaction mixture was transcribed using the HiScribe T7 High Yield RNA Synthesis Kit (NEB) according to the manufacturer's instructions. The samples were incubated for 2.5 h at 37 °C and 500 rpm in a Thermomixer (Eppendorf) followed by addition of 2 U DNase I (Thermo) for 15 min to remove DNA contaminations. The RNA was purified using the RNA Clean and Concentrator-25 Kit (Zymo Research) and 25 µL RNase/DNase-free water (VWR) was used to elute the RNA. The concentration was determined by UV/Vis spectroscopy (2.2.1.2).

The synthesis of ternary complexes by *in vitro* translation was performed using the *E. coli* S30 Extract System for Linear Templates from Promega. According to the manufacturer's instructions, a defined amount of 10 µg RNA ($\sim 1 \times 10^{13}$ molecules) per 50 µL reaction volume was mixed with *E. coli* S30 extract ($\sim 3.6 \times 10^{14}$ ribosomes) including NTPs, tRNAs, an ATP-regenerating system, IPTG, appropriate salts and an amino acid mix without cysteine and methionine. Anti-small stable RNA (ssrA, c = 2µM) was added to the reaction mixture. Thus, proteins translated from mRNAs lacking stop codons are modified by a ssrA-encoded peptide tag at the carboxyl terminus that prevents their degradation (Keiler *et al*., 1996). The samples were incubated at 37 °C for 15 min. The translation was stopped by addition of 500 µL ice-cold stop buffer including freshly added heparin [c = 2.5 mg/mL].

2.2.2.3 Selection of target-specific ternary complexes

The enrichment of IgG-specific variants from the generated domain B libraries (2.2.2.1) was performed by a modified biopanning, which included the presentation of nascent domain B variants via ribosomal ternary complexes. To prevent dissociation of the non-covalently linked ternary complexes, the selection was performed at a low temperature (15 °C) and in the presence of Mg^{2+} (Amstutz *et al*., 2006).

Preparation of magnetic beads

Both streptavidin (SA) and neutravidin (NA) beads from Invitrogen were used for the immobilization of biotinylated hIgG1-Fc. The two bead types were used alternately in each round of selection to present only hIgG1-Fc as a constant factor. The amount of beads (Table 24) was calculated based on the binding capacity for biotinylated hIgG1-Fc (2.2.4.4). The magnetic beads were washed three times with 1 mL TBST buffer each. The beads were focused with a magnet for 1 min to exchange the buffer between the wash steps. To reduce non-specific interactions of library variants with the carboxylated surface, the beads were blocked overnight at 4 °C with 1 mL TBST 0.5 % BSA (w/v). After the beads were washed three times with 1 mL TBST buffer each, they were mixed with biotinylated hIgG1-Fc and incubated for 2 h at 4 °C. Finally, the beads were washed three times with 1 mL TBST buffer each and followed by resuspension in 1 mL WBT buffer containing Mg^{2+}.

Selection of IgG-specific ternary complexes

Four rounds of biopanning were performed in 96-well format using the KingFisher Duo sample preparation device (Fisher Scientific). The ternary complexes formed by *in vitro* translation (2.2.2.2) were added to the prepared magnetic beads and incubated for 1 h. For each selection round, several wash steps each with 500 µL WBT 0.1 % BSA were performed. To generate selection pressure, the amount of biotinylated hIgG1-Fc used in the individual selection rounds was reduced and the wash steps were successively extended (Table 24). In round four, an off-rate selection was performed to obtain variants with high affinity. After binding of ternary complexes to 10 nM biotinylated hIgG1-Fc for 1 h, a 1,000-fold excess of non-biotinylated hIgG1-Fc was added and the mixture was incubated for another 30 min.

Table 24: Overview of selection parameters

Round	$c_{hIgG1-Fc}$ [nM]	Incubation time [min]	Magnetic beads	Wash steps
1	200	60	25 µL SA M-270	5×30 sec, 1×5 min
2	100	60	12 µL NA M-270	5×30 sec, 2×5 min
3	50	60	12 µL SA M-270	5×30 sec, 2×5 min, 1×10 min
4	10	60 + 30 (off-rate)	12 µL NA M-270	5×30 sec, 2×5 min, 2×10 min

2.2.2.4 mRNA elution, isolation and reverse transcription

IgG-specific ternary complexes were eluted from the magnetic beads by adding 100 µL 25 mM EDTA-containing EB buffer. The ion chelator EDTA complexes Mg^{2+} ions leading to the destabilization and dissociation of the ternary complexes and finally in the release of mRNA. The eluted mRNA was isolated using the RNA Clean & Concentrator5-Kit from Zymo Research according to the manufacturer's instructions. In addition, 2 U DNase I (Thermo) was added to the samples to remove DNA contaminations. The mixtures were incubated for 15 min at room temperature, and the purified mRNA was eluted from the column by adding 2x 25 µL RNase/DNase-free water (VWR).

Next, the isolated mRNA was transcribed into cDNA to generate a template for amplification. 4 µL reverse primer tolA-rev (100 pmol/µL) and 8 µL dNTPs (each 10 mM) were added to 50 µL purified RNA. The reaction mixture was divided into equal parts of 31 µL each and incubated for 10 min at 65 °C in a thermocycler. 1 µL RNasin Ribonuclease Inhibitor (Promega) and 1 µL Reverse Transcriptase (Roche) prepared in 8 µL of appropriate buffer (final buffer concentration 1x) were added. The samples were mixed and incubated for 30 min at 65 °C followed by heat inactivation for 5 min at 85 °C. In addition, a negative control containing water instead of RNA was transcribed. By a final PCR, the generated cDNA was transcribed in a RD compatible format and used in a new selection round (2.2.2.2). For this, Q5 Polymerase (NEB) and the primers F1B and tolA-stem-rev were added and PCR was performed as described (2.2.1.3). After agarose gel electrophoresis (2.2.1.4), the DNA fragments were isolated from the gel (2.2.1.1) using the Zymoclean Gel DNA Recovery Kit (Zymo Research).

2.2.2.5 Cloning of DNA libraries

After biopanning, the obtained DNA libraries of each selection round (2.2.2.4) were subcloned in the expression vector pNP-013 via BsaI restriction sites. Electrocompetent BL21(DE3) cells were transformed with the expression plasmids (2.2.3.2) and used to produce individual protein variants (2.2.3.3).

2.2.3 Microbiological methods

2.2.3.1 Production of competent cells

Chemically competent cells
Chemically competent BL21(DE3) cells were prepared using a modified protocol according to Cohen (Cohen *et al.*, 1972). The cells were plated on an agar plate and grown overnight at 37 °C. A single colony was transferred in 5 mL 2x YT medium and the cells were grown at 37 °C and 220 rpm. A fresh culture (2x YT medium) was inoculated with 1 mL preculture and the cells were grown as described until the optical density (OD_{600}) reached 0.5. The following procedure was performed under constant cooling in an ice bath or at 4 °C. The cell suspension was chilled on ice for 30 min followed by centrifugation at 3.000×g for 10 min. The obtained sediment was resuspended in 75 mL ice-cold 100 mM $CaCl_2$ and incubated for 1 h. After a second centrifugation step, the cells were resuspended in 10 mL $CaCl_2$, mixed with 300 µL DMSO and incubated for 10 min. The cells were aliquoted into 1.5 mL reaction tubes, frozen with liquid nitrogen for 5 min and stored at -80 °C.

Electrocompetent cells
An overnight culture with 50 mL 2x YT medium was inoculated with a single colony to produce electrocompetent BL21(DE3) cells. The cells were grown at 37 °C and 220 rpm. A fresh culture (1 L of 2x TBY medium) was inoculated with 5 mL preculture. The cells were grown as described until OD_{600} reached 0.5. The cell suspension was chilled on ice for 1 h followed by centrifugation at 3.000×g for 10 min at 4 °C. The obtained cell pellet was washed three times with 300 mL ice-cold 10 % glycerol (v/v). Finally, the cell pellet was resuspended in 1.3 mL ice-cold 10 % glycerol (v/v), aliquoted into 1.5 mL reaction tubes and stored at -80 °C.

2.2.3.2 Transformation of bacteria cells

Transformation of chemically competent cells
An aliquot of 40 µL chemically competent BL21(DE3) cells was mixed with 10-40 ng plasmid DNA (2.2.1.6) and incubated for 10 min at 4 °C followed by a heat shock for 30 sec at 42 °C. After a 2-min incubation on ice, 1 mL pre-warmed SOC medium was added and the cells were grown for 1 h at 37 °C and 220 rpm (Pope & Kent, 1996). Individual colonies were produced by plating 100 µL cell suspension onto an agar plate with appropriate antibiotics (Table 15). The bacterial colonies were grown at 37 °C for 18 h.

Transformation of electrocompetent cells

The uptake of plasmid DNA by BL21(DE3) cells was induced by electroporation (Dower *et al.*, 1988). On ice, 40 µL electrocompetent cells were mixed with 1-10 ng plasmid DNA (2.2.1.6) and transferred into an ice-cold transformation cuvette with an electrode spacing of 1 mm. The cells were exposed to a pulse of 1.8 kV, 50 µF, 100 Ω and 1 mL pre-warmed SOC medium was added. The cells were grown at 37 °C and 220 rpm for 1 h. Individual colonies were produced as described above.

2.2.3.3 Heterologous expression of recombinant genes

Protein production in shake flasks

BL21(DE3) cells were grown in shake flasks with a maximal culture volume of 20 % of the flask volume. Precultures in 2x YT medium containing 1 % (w/v) glucose, 2 mM $MgCl_2$ and appropriate antibiotics were used as inoculum (Table 15). First, the precultures were grown in 15 mL tubes or in 250 mL flasks for 8 h at 37 °C and 220 rpm. The main cultures were inoculated in 300 mL H15 autoinduction medium ($OD_{600} = 0.3$) that contains 0.2 mL/L trace elements, 0.3 % antifoam and appropriate antibiotics (Danielson *et al.*, 2004). The autoinduction system is based on a diauxic effect of the metabolic switch from glucose to lactose. Glucose functions as a carbon source, whereas lactose is not completely converted and acts as an inducer. Allolactose is produced as a by-product of the lactose metabolism and induces the lac operator of the expression plasmids by binding to the lac repressor (Studier, 2005). The main cultures were grown overnight in a resonance acoustic mixer (RAMbio) at 37 °C and $20 \times g$ acceleration. Subsequently, the cells were harvested by centrifugation at $4.000 \times g$ for 20 min and the obtained cell pellet was stored at -20 °C until further use.

For NMR spectroscopy, CID 25 was produced in M9 minimal medium. A preculture in M9 minimal medium with appropriate antibiotics (Table 15) supplemented with $^{15}NH_4Cl$ and $^{13}C6$-glucose, as the sole nitrogen and carbon sources, was used as inoculum. For cell adaptation, two precultures were grown in M9 minimal medium each for 24 h at 37 °C and 220 rpm. The main culture was inoculated in 500 mL M9 minimal medium ($OD_{600} = 0.2$) with appropriate antibiotics (Table 15). The cells were grown in 2.5 L Ultra Yield flasks for 2-3 h at 37 °C and 180 rpm until OD_{600} reached 0.6-0.8. Expression of recombinant genes was induced by addition of IPTG (final concentration 1 mM). After growing for 4 h at 30 °C and 180 rpm ($OD_{600} = 3$), the cells were harvested and stored as described above.

Protein production in microtiter plates

For the parallel characterization of recombinant proteins within the screening process, *E. coli* cells were grown in 384-well microtiter plates covered with breathable tape sheets. Individual colonies were picked using the automated colony picking system Qpix2 (program picking; pin type X4370). The cells were inoculated in 50 µL/well 2x YT medium containing 1 % (w/v) glucose, 2 mM $MgCl_2$ and appropriate antibiotics (Table 15). The cells were grown overnight in a plate shaker at 37 °C and 850 rpm. To prepare the main cultures, 60 µL/well autoinduction LBB medium with appropriate antibiotics (Table 15) was inoculated by automated pin dipping (program replicating; pin type X4376). After inoculation, 10 µL/well 50 % glycerol (v/v) was added to the precultures and the cells were stored at -20 °C. The main cultures were grown at

37 °C and 850 rpm for 5 h followed by 16 h at 31 °C. For µ-scale purification in 96-deepwell plates (2.2.4.2), 1 mL/well LBB autoinduction medium with appropriate antibiotics (Table 15) was inoculated with 5 µL/well preculture and the cells were grown as described above. The cell suspensions were centrifuged at 4.000×g for 10 min and the cell pellets were either stored at -20 °C or used directly for cell lysis (2.2.4.1).

2.2.4 Methods in protein chemistry

2.2.4.1 Cell lysis

After gene expression (2.2.3.3), the cells were lysed by a combination of enzymatic and physical methods. First, the cell pellets were resuspended in lysis buffer containing 15 µM lysozyme, 1 mM AEBSF, 2.5 mM $MgCl_2$ and 1 U/mL Benzonase Nuclease (VWR). Fixed volumes of lysis buffer were used for cells grown in microtiter plates (50 µL/well for 384-well plates, 70 µL/well for 96-deepwell plates). After addition of lysis buffer, the microtiter plates were sealed to avoid contaminations. Cell pellets obtained from cultivations in shake flasks were resuspended in 10 mL lysis buffer per gram of wet biomass. The enzymatic cell lysis was performed for 30 min at room temperature under constant stirring or plate shaking. Subsequently, the cells were mechanically disrupted either by ultrasonification or by freeze/thaw cycles.

Cell lysis by ultrasonification
The cell suspension was sonicated under constant cooling. Depending on the volume, either 6 mm (15-45 mL) or 10 mm (46-70 mL) sonotrodes were used. Three pulses with an amplitude of 60 % were performed for 20 sec each. The cell suspension was incubated for 30 min under constant stirring followed by centrifugation at 10.000×g and 4 °C for 20 min. The obtained supernatant was used for lab scale purification (2.2.4.2).

Cell lysis by freeze/thaw cycles
Cell pellets obtained from the cultivation in 384-well and 96-deepwell plates were disrupted by two freeze/thaw cycles. For this, the sealed microtiter plates were frozen at -80 °C for 1 h and thawed at room temperature. After a centrifugation step at 4.000×g and 4 °C for 30 min, the supernatants were used for ELISA (2.2.4.9) or µ-scale purification (2.2.4.2).

2.2.4.2 Automated purification of His10-tagged proteins

An automated two-step purification procedure was performed to purify domain B variants containing a His10-tag. First, immobilized metal affinity chromatography (IMAC) was used to isolate the proteins from supernatants (2.2.4.1). Second, size exclusion chromatography (SEC) was performed as a polishing step.

Lab scale purification

Proteins obtained from gene expression (2.2.3.3) and cell lysis (2.2.4.1) were purified in parallel using an ÄKTAxpress chromatography system (GE Healthcare). The supernatants were filtered and loaded on equilibrated 5 mL HisTrap HP columns (GE Healthcare) at a flow rate of 5 mL/min. To remove non-specifically bound proteins, the columns were washed with 15 CV IMAC buffer A. Bound proteins were eluted by imidazole competition (500 mM imidazole) using 5 CV IMAC buffer B (100 %) at a flow rate of 5 mL/min. Fractions, which contained the eluted proteins were automatically collected in in-system sample loops. The following parameters were used for the automatic detection of the eluted proteins.

Number of loops = 4
Maximum peak volume = 2 mL
Peak UV start value = 10 mAU
Peak UV end value = 10 mAU
Peak minimum width = 0.5 min
Peak UV initial slope = 10 mAU/min
Peak UV final slope = 10 mAU/min

A maximum sample volume of 5 mL was loaded on an equilibrated HiLoad 16/60 Superdex 200 prep grade column (GE Healthcare) at a flow rate of 1 mL/min. As running buffer, 1x PBS pH 7.4 was used. The purified proteins were automatically collected into 2 mL fractions. A calibration curve with Gel Filtration Calibration Standard (GE Healthcare) was used to estimate the apparent molecular mass of the proteins. To produce tagfree proteins, a proteolytic cleavage of the GFP-His10-tag was performed using histidine-tagged TVMV proteases (Navigo Proteins). The purified proteins were mixed with TVMV proteases at a ratio of 1:100 (TVMV proteases:POI) followed by overnight incubation at room temperature. The tagfree proteins were purified by reverse IMAC as described above and collected in the flowthrough fractions. The proteins were concentrated (2.2.4.7), manually loaded into the sample loops of the ÄKTAxpress chromatography system (GE Healthcare) and purified by SEC using a HiLoad 16/60 Superdex 75 prep grade column (GE Healthcare) as described above. The obtained elution fractions were analyzed by SDS-PAGE (2.2.4.5).

µ-scale purification

A Freedom Evo liquid handling robot (Tecan) was used for the parallel purification of proteins from *E. coli* cultures in 96-deepwell plates. After gene expression (2.2.3.3) and cell lysis (2.2.4.1), 280 µL/well supernatant was transferred into the wells of a 96-well F-bottom microplate (Nunc) and an IMAC was performed. For the purification, 200 µL tips (PhyNexus) were filled with 20 µL Ni Sepharose 6 Fast Flow (Cytiva), respectively. The tips were equilibrated by five up and down pipetting cycles with 150 µL IMAC buffer A each for 1 min at a speed of 5 mm/s. The supernatants were passed over the matrix by continuous up-and-down pipetting for 1 h at a speed of 2.5 mm/s. Subsequently, non-specifically bound proteins were removed by five up-and-down pipetting cycles with 170 µL IMAC buffer A each for 1 min at a speed of 10 mm/s. Bound proteins were eluted by seven up-and-down pipetting cycles with 100 µL IMAC buffer B (500 mM imidazole) each for 2 min at a speed of 2.5 mm/s. Finally, the

buffer of the samples was exchanged using a 96-well fritted plate filled with 800 µL/well Sephadex G-25 Fine (Cytiva) previously equilibrated with 5 CV 1x PBS pH 7.4. The protein solutions were passed over the matrix by centrifugation at 500×g for 2 min. The obtained proteins were used either for a biochemical characterization or for TVMV cleavage followed by reverse IMAC as described above.

2.2.4.3 Automated purification of proteins without an affinity tag

Tagfree IgG-binding domain B variants were purified by affinity chromatography (AIC) using IgG Sepharose 6 Fast Flow XK26/20 matrix (Cytiva). After gene expression (2.2.3.3) and cell lysis (2.2.4.1), the supernatants were loaded on equilibrated 50 mL IgG Sepharose 6 Fast Flow XK26/20 columns (Cytiva) at a flow rate of 15 mL/min. To remove impurities and non-specifically bound proteins, the columns were washed with 20 CV AIC buffer A. A pH shift (pH 3.0) with 10 CV AIC buffer B (100 %) eluted the bound proteins. The proteins were automatically collected into 2 mL fractions. The protein solutions were neutralized by titration with AIC buffer C until pH 7.4 was reached. Finally, the protein solutions were loaded into the sample loops of the chromatography system and SEC was performed as described (2.2.4.2). The obtained elution fractions were analyzed by SDS-PAGE (2.2.4.5). The purified proteins were stored at 4 °C and used for analytical approaches.

2.2.4.4 Biotinylation of hIgG1-Fc

In order to specifically detect and immobilize hIgG1-Fc, the protein was chemically modified by biotinylation. hIgG1-Fc was biotinylated using the EZ-Link Sulfo-NHS-LC-Biotinylation Kit from Thermo Scientific according to the manufacturer's instructions. The N-hydroxysuccinimide (NHS) esters of biotin react with the primary amine groups of hIgG1-Fc. Amine groups are located in the side chains of lysines and in the N-terminus of the protein. Sulfo-NHS-LC-biotin was dissolved, added in a 30-fold molar excess and the reaction mixture was incubated for 2 h at room temperature. Non-coupled biotin (Thermo Scientific) was removed by SEC using an equilibrated Superdex 200 prep grade column (GE Healthcare) (2.2.4.2). A pull-down assay was performed to determine the efficiency of biotinylation. For this, 40 µL biotinylated hIgG1-Fc [c = 100 µg/mL] was added to 20 µL prepared (2.2.2.3) streptavidin beads (Dynabeads M-270, Invitrogen) and the mixture was incubated for 1 h at 4 °C. The beads were focused with a magnet for 1 min and the supernatant was transferred into a fresh reaction tube. The beads were washed three times with 200 µL 1x PBS pH 7.4 each followed by a heat step for 5 min at 90 °C. All samples were analyzed by SDS-PAGE (2.2.4.5).

2.2.4.5 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

The molecular weight of proteins was analyzed by SDS polyacrylamide gel electrophoresis (SDS-PAGE) as described in the work of Laemmli (Laemmli, 1970). This method is based on the differential migration rate of negatively charged proteins in an electric field as a function of their size (Weber & Osborn, 1969). The electrophoresis was performed in a NuPAGE XCell

SureLock Mini-Cell electrophoresis chamber using NuPAGE Novex 4-12 % Bis-Tris gels (Invitrogen) and 1x MES running buffer. After expression of recombinant genes (2.2.3.3) and cell lysis (2.2.4.1), the soluble and insoluble protein fractions were separated by centrifugation for 30 min at 3.000×g and 4 °C. 20 µL supernatant was transferred into a new reaction tube. The obtained cell pellet was washed two times with 100 µL 1x PBS pH 7.4 each followed by resuspension in 20 µL 8 M urea. 5x SDS sample buffer (final concentration 1x) was added and the samples were incubated at 95 °C for 5 min. A sample volume of 10 µL was loaded on the gel. To determine the molecular weight of the proteins, 5 µL PageRuler Unstained Broad Range Protein Ladder (Thermo Scientific) was pipetted into one lane of each gel. The proteins were separated at a voltage of 200 V for 35 min. The gel was covered with staining solution, heated in the microwave for 30 sec and stained for at least 30 min. Destaining solution was used to decolorize the gel until the band intensity reached the desired level. Finally, the gel was digitized by scanning.

2.2.4.6 Photometric determination of protein concentration

The concentration of aqueous protein solutions was determined according to the Lambert-Beer law by measuring the absorbance at 280 nm (2.2.1.2). The molar extinction coefficient was calculated based on the primary protein sequence (Gill & von Hippel, 1989). A buffer-corrected spectrum was measured in a range of 250-320 nm using a spectrophotometer (Thermo Scientific). The 280 nm/260 nm ratio was used to estimate DNA contaminations. The concentration of µ-scale purified proteins was determined in a 96-well flat-bottom plate (Fisher Scientific) using the Bradford method (Bradford, 1976). A calibration curve prepared with an albumin standard was used to calculate the protein concentration. 5 µL/well sample and 250 µL/well Bradford reagent from the Coomassie Plus Protein Assay Kit (Pierce) were mixed and incubated for 10 min at room temperature. The absorption spectra of the samples were determined at a wavelength of 595 nm and the protein concentrations were calculated based on a second degree polynomial function used to fit the calibration curve.

2.2.4.7 Concentration and buffer exchange of protein solutions

For concentration and buffer exchange of protein solutions, Amicon Ultra-15 centrifugal filters (Merck Millipore, MWCO 3.000 Da) or disposable PD-10 columns filled with 1 mL Sephadex G-25 resin (Cytiva) were used according to the manufacturer's instructions. Buffer exchange was performed by applying a theoretical dilution of at least 1:100.

2.2.4.8 High-performance liquid chromatography (HPLC)

Size exclusion high-performance liquid chromatography (SE-HPLC) was used to determine the apparent molecular weight of the purified proteins and the homogeneity of the samples. SE-HPLC was performed using a modular Ultimate 3000 HPLC system (Dionex). Depending on the theoretical molecular weight of the proteins, Tricorn 5/150 GL columns filled with Superdex Increase 75 prep grade or Superdex Increase 200 prep grade (GE Healthcare) were used. The

column was equilibrated with 2 CV 1x PBS pH 7.4 at a flow rate of 0.3 mL/min. 50 µL protein solution [c = 0.2 mg/mL] in appropriate running buffer was loaded on the column at a flow rate of 0.3 mL/min. A Gel Filtration Standard (Bio-Rad) was included in each measurement for the calculation of the apparent molecular weight. Proteins were photometrically detected at 280 nm and a wavelength of 550 nm was used as reference. After the measurement, column cleaning was performed with 2 CV 6 M GdnHCl at a flow rate of 0.5 mL/min. The obtained chromatogram was evaluated using the Chromeleon 6.80 software.

2.2.4.9 Enzyme-linked immunosorbent assay (ELISA)

In this work, enzyme-linked immunosorbent assay (ELISA) was used to analyze protein interactions and to detect proteins in different fractions. The ELISA was performed in 384-well high binding microtiter plates (Greiner) within a working volume of 20 µL/well unless otherwise stated. Depending on the ELISA setup, hIgG1-Fc (on-target) and BSA (off-target) [c = 40 nM] or purified domain B variants [c = 120 nM] in 1x PBS pH 7.4 were immobilized overnight at 4 °C. In each assay, wild type domain B was used as positive control, whereas the corresponding empty expression vector was used as negative control. After immobilization, the microtiter plates were washed three times with 110 µL/well ELISA buffer A each using a semiautomated plate washer (Biotek). Non-specific binding sites were blocked by addition of 110 µL/well Blocking buffer followed by incubation for 2 h at room temperature. After three wash cycles with 110 µL/well ELISA buffer A each either cell lysates (2.2.4.1) or purified proteins (2.2.4.2) were added. The detailed assay procedure for the different ELISA types is described in the following section.

GFP-based pool ELISA
A GFP-based pool ELISA was performed to investigate the binding of a protein pool to immobilized hIgG1-Fc and BSA. After ribosome display, the obtained domain B libraries were subcloned in a GFP-containing expression vector (pNP-013) and the recombinant genes were expressed (2.2.3.3). The *E. coli* cells were lysed (2.2.4.1) and 30 µL/well supernatant was transferred into the cavities of the 384-well microtiter plate previously coated with hIgG1-Fc and BSA [c = 40 nM]. After incubation for 1 h at room temperature, the microtiter plate was washed three times with ELISA buffer A as described above followed by three wash cycles with ELISA buffer B. Finally, 30 µL/well 1x PBS pH 7.4 was added and the fluorescence intensity of fused GFP at 483 nm excitation and 535 nm emission was measured using a microplate reader (Tecan Infinite). Domain B libraries, which showed specific binding to hIgG1-Fc were selected for high-throughput screening (HTS).

GFP-based hit ELISA
A GFP-based hit ELISA was used for the high-throughput screening (HTS) of IgG-specific pools obtained from ribosome display. The screening process was performed using a multi-axis robot arm integrated in a momentum automation system (Thermo Scientific). Due to the 384-well format, this assay allowed the screening of 3,240 domain B variants per library. Both hIgG1-Fc and BSA [c = 40 nM] were immobilized on the surface of 384-well high binding plates (Greiner). After gene expression (2.2.3.3) and cell lysis (2.2.4.1), 30 µL/well supernatant

was transferred into the cavities of the microtiter plates followed by incubation for 1 h at room temperature. The microtiter plates were washed three times with ELISA buffer A as described followed by three wash cycles with ELISA buffer B. 30 µL/well 1x PBS pH 7.4 was added and the fluorescence intensity of fused GFP was measured as described above. Variants that showed specific binding to hIgG1-Fc were selected as "hits" and produced on larger scale for further characterization. Hits were selected based on their absolute signal amplitude and according to statistical parameters.

Concentration-dependent ELISA

A concentration-dependent ELISA was performed to determine binding affinities of domain B variants for hIgG1-Fc (Friguet *et al.*, 1985). For this, tagfree domain B variants were produced (2.2.3.3) and purified by a two-step procedure (2.2.4.2 and 2.2.4.3). Serial dilutions of biotinylated hIgG1-Fc in 1x PBS pH 7.4 were transferred into the cavities of a 384-well high binding plate previously coated with tagfree domain B variants [c = 120 nM]. After incubation for 1 h, the microtiter plate was washed three times with ELISA buffer A as described above. Bound biotinylated hIgG1-Fc was detected using Strep-Tactin-HRP (IBA) diluted 1:10,000 in ELISA buffer A. The Strep-Tactin-HRP solution was added to each cavity and incubated for 1 h at room temperature. The microtiter plate was washed three times with ELISA buffer A as described followed by three wash cycles with ELISA buffer B. Subsequently, the peroxidase substrate TMB PLUS 2 (Biotrend) was added. After 10-30 min, 0.2 M H_2SO_4 was added to stop the substrate conversion. The absorbance at 450 nm was measured and corrected by the reference wavelength (620 nm). The data were evaluated using the SigmaPlot software and the integrated regression model, assuming a 1:1 binding stoichiometry (Equation 3).

$$A_{450} = \frac{B_{max} \cdot c}{K_D + c}$$
(Equation 3)

A_{450} = specific binding signal at 450 nm
B_{max} = maximum available binding sites
c = concentration of free ligand [M]
K_D = concentration of free ligand at half-maximum saturation [M]

Competitive ELISA

A competitive ELISA was performed to investigate whether novel domain B variants compete with wild type domain B for Fc binding. For this, hIgG1-Fc [c = 40 nM] was immobilized on the surface of a 384-well high binding microtiter (Greiner) as described above. Serial dilutions of purified tagfree domain B variants [c = 0.01-3 µM] were produced in ELISA buffer A. The protein solutions were transferred into the cavities of the 384-well microtiter plate and Strep-tagged wild type domain B [c = 100 nM] was added as additional antigen. All further assay steps were performed according to the concentration-dependent ELISA described above. The inhibitory constant (K_i) of the domain B variants was determined based on to the graphical method of Dixon (Dixon, 1953).

2.2.4.10 CID 25 as ligand for the affinity chromatographic purification of hIgG1

To examine the suitability of CID 25 as ligand in affinity chromatography, the variant (5.1 mg) was coupled to the surface of a 1 mL HiTrap NHS-Activated HP column (Cytiva) according to the manufacturer's instructions. Wild type domain B (5.5 mg) was coupled as comparison. The coupling efficiency was determined by comparing the protein amount of both the coupling load and the flowthrough fraction (2.2.4.6). Unless otherwise stated, all column experiments were performed at 10 °C at a flow rate of 1 mL/min using an ÄKTAxpress chromatography system (GE Healthcare). The columns were equilibrated with 10 CV 1x PBS pH 7.4. After loading 2 mg hIgG1 diluted in a volume of 6.5 mL 1x PBS pH 7.4 [c = 0.3 mg/mL], non-specifically bound proteins were removed by a wash step with 20 CV 1x PBS pH 7.4. Finally, bound hIgG1 was eluted by a pH shift (pH 3.0) using 10 CV AIC buffer B (100 %).

Determination of dynamic binding capacity

To determine the dynamic binding capacity of both the CID 25 column and the matrix functionalized with wild type domain B, hIgG1 [c = 1.5 mg/mL] in 1x PBS pH 7.4 was passed through the columns at a flow rate of 0.2 mL/min with a contact time of 5 min. hIgG1 was detected in the flowthrough by measuring the absorbance at 280 nm. In addition, individual flowthrough factions were analyzed by SDS-PAGE (2.2.4.5). The dynamic binding capacity of the columns was calculated based on the total volume that was required until a significant amount (10 % of the maximum signal) of hIgG1 was detected in the flowthrough. The concentration of the saturation curve was normalized to the maximum concentration of used binding partner c/c_0 (Horstmann *et al.*, 1986) (Hage & Cazes, 2005).

Determination of elution pH

To determine the elution pH, 2 mg hIgG1 diluted in a volume of 6.5 mL 1x PBS pH 7.4 [c = 0.3 mg/mL] was loaded on the equilibrated CID 25 column at a flow rate of 1 mL/min. The wild type domain B column was used as comparison. After a wash step with 5 CV 1x PBS pH 7.4, the system and the columns were flushed with 10 CV 0.1 M citrate, 0.15 M NaCl pH 6.5. A gradient elution of hIgG1 in a range of pH 6.5-2.0 was performed within a volume of 20 CV 0.1 M citrate pH 2.0. Eluted hIgG1 was detected by measuring the absorbance at 280 nm. The obtained elution fractions were pooled and the pH value was determined using a pH meter (WTW).

Isolation of hIgG1 from E. coli and Expi293 cell lysates

As a proof of concept for the use of CID 25 as affinity ligand, hIgG1 was purified from lysates of different cell types. For this, *E. coli* and Expi293 cells were grown, harvested, and lysed (Dr. Manja Gloser-Bräunig, Navigo Proteins). 1.95 mg hIgG1 was added to each cell lysate and the sample volume was adjusted to 10 mL [c = 0.2 mg/mL]. A final optical density of $OD_{600} = 42$ was determined for the *E. coli* lysate and approximately 25×10^6 cells were used to prepare the Expi293 lysate. The cell lysates were injected in the super loops attached to the ÄKTAxpress chromatography system (GE Healthcare). Next, the samples were successively loaded on the equilibrated column at a flow rate of 0.5 mL/min. All further purification steps were performed as described above.

2.2.5 Biophysical methods

2.2.5.1 Surface plasmon resonance (SPR)

The affinity of domain B variants for the targets hIgG1, hIgG1-Fc and Avelumab-Fab was investigated by surface plasmon resonance (SPR) spectroscopy (Turbadar, 1959) (Liedberg *et al.*, 1983). For this, a multichannel SPR-32 system (Bruker Daltonik) with corresponding CM5 sensor chips was used. Carboxymethylated dextran is covalently attached to the gold surface of the CM5 chip, allowing the covalent coupling of molecules via amine functional groups. An NHS/EDC Coupling Kit (Bruker Daltonik) was used to couple hIgG1-Fc, Avelumab-Fab (~500 RU) and hIgG1 (~1,000 RU) on the surface of the CM5 chip according to the manufacturer's instructions. In further binding analyses, CID 25 was immobilized on the surface of the CM5 chip (~180 RU). CID 25 was immobilized by direct amine coupling using NHS/EDC chemistry (Johnsson *et al.*, 1991) similar to the coupling to HiTrap NHS-Activated HP resin (2.2.4.10). A non-coupled flow cell of the CM5 chip was used as reference. Serial dilutions [c = 0.02-2 µM] in SPR buffer were prepared for each domain B variant and the binding to hIgG1-Fc, hIgG1 and Avelumab-Fab was analyzed. The samples were passed over the CM5 chip at a temperature of 20 °C with a flow rate of 30 µL/min. After each cycle, the proteins on the surface of the chip were regenerated with 10 mM glycine pH 2.0 at a flow rate of 30 µL/min and a contact time of 30 sec. After regeneration, the baseline was stabilized for 2 min by equilibrating the chip with SPR buffer at a flow rate of 30 µL/min. To ensure the functionality of the coupled targets, binding of wild type domain B [c = 50 nM] to hIgG1-Fc, IgG1 and Avelumab-Fab was measured at the beginning and end of each curve set.

To investigate the suitability of CID 25 as ligand in affinity chromatography (2.2.4.10), serial dilutions of hIgG1-Fc [c = 0.01-1 µM] in SPR buffer were passed over the surface of a CM5 chip coupled with the variant. All further assay steps were performed as described above. The obtained data were evaluated using the Sierra Analyzer software (Bruker Daltonik). Numerical approximation and the kinetic parameters were calculated using the binding models integrated in the software, assuming a 1:1 binding stoichiometry. K_D values were calculated based on kinetic parameters k_{on} and k_{off}, assuming a 1:1 binding stoichiometry.

2.2.5.2 Biolayer interferometry (BLI)

As part of the screening process (2.2.4.9), biolayer interferometry (BLI) was used to characterize the binding of domain B variants to hIgG1-Fc. BLI allows the label-free detection of biomolecular interactions in real time based on a binding-induced change in refractive index (Kamat & Rafique, 2017). The BLI measurements were performed using an Octet BLUE384 device (ForteBio) equipped with Ni-NTA sensors and black 384-well microtiter plates (Greiner). Unless otherwise stated, all assay steps were performed in a volume of 80 µL/well. After expression of recombinant genes (2.2.3.3) and cell lysis (2.2.4.1), the concentration of soluble domain B variants was determined using a fluorescence standard curve prepared with GFP. Ni-NTA sensors were equilibrated with BLI buffer for 5 min. Normalized domain B variants fused to GFP-His10-tag [c ≈ 2 µM] were captured from the supernatants obtained from

cell lysis (2.2.4.1). Subsequently, a wash step with BLI buffer was performed for 1 min to remove non-specifically bound proteins. Binding of hIgG1-Fc and Avelumab-Fab [c = 0.5 µM] to immobilized domain B variants was measured in an association step for 2 min followed by a dissociation step for 1 min. The surface of the sensors was regenerated with 10 mM glycine pH 2.0 for 1 min and a recharge cycle with 10 mM NiCl$_2$ pH 7.4 was performed for 2 min. The Octet Data Analysis 7.1 software (ForteBio) was used to evaluate the obtained data. K$_D$ values were calculated based on kinetic parameters k$_{on}$ and k$_{off}$, assuming a 1:1 binding stoichiometry.

2.2.5.3 Circular dichroism spectroscopy (CD)

Circular dichroism spectroscopy (CD) can be used to measure the difference in the absorption of left and right circularly polarized light in optically active (chiral) molecules. This allows the analysis of protein structural elements and the investigation of conformational changes during temperature-induced unfolding and refolding of proteins (Holtzhauer, 1996). Peptide bonds can be detected in the far-UV spectral region (170-250 nm), allowing conclusions about the secondary structure of proteins. Moreover, the environment of aromatic side chains can be investigated in the near-UV spectral region (250-300 nm) to determine the tertiary structure of proteins.

In this work, CD measurements were performed in the far-UV spectral region (200-250 nm) using the CD spectrophotometer J-810 (Jasco). After buffer exchange (2.2.4.7), domain B variants were diluted in CD buffer [c = 0.2 mg/mL] and the samples were transferred into a 1 mm quartz cuvette (Hellma Optik). The spectral measurements were performed at 20 °C with a speed of 50 nm/min and a bandwidth of 1 nm. 16 spectra per sample were measured and baseline corrected by subtracting a buffer spectrum. Both the mean residual ellipticity and the proportion of α-helical secondary structure elements were calculated using the Spectra Manager I software according to the reference by Yang (Yang *et al*., 1986) (Equation 4).

$$[\Theta]_{MRE} = \frac{[\Theta] \cdot 100 \cdot MW}{c \cdot I \cdot n_A} \qquad \text{(Equation 4)}$$

$[\Theta]_{MRE}$ = mean residue ellipticity [degree·M^{-1}·cm^{-1}]
$[\Theta]$ = ellipticity [degree]
MW = molecular weight [g·mol^{-1}]
c = protein concentration [g·mol^{-1}]
I = layer thickness [cm]
n_A = amino acid residues

Measurement of temperature-induced structural transitions
To investigate temperature-induced unfolding and refolding of selected domain B variants, the minima of a single spectrum in the far-UV region (190-250 nm) were determined as described above. Temperature-induced structural transitions were measured using a 1 mm quartz cuvette (Hellma Optik) at a wavelength of 222 nm and a band width of 1 nm. With a slope of 1 °C/min, the initial temperature (20 °C) was increased up to 90 °C. To examine protein refolding, single CD spectra were measured after returning to 20 °C. The residual ellipticity (Equation 4) was

plotted as a function of temperature. The transition temperature (T_m) was calculated using the software Sigma Plot (Systat, category sigmoid four parameter).

2.2.5.4 Differential scanning calorimetry (DSC)

Dynamic scanning calorimetry (DSC) was performed to further investigate thermal transitions and folding mechanisms of CID 25 and wild type domain B. By DSC, the heat flow differences between a sample and a reference are measured directly as a function of temperature, which allows the calculation of thermodynamic parameters (Chiu & Prenner, 2011). Changes in heat capacity (ΔCp), melting temperature (T_m) and enthalpy (ΔH) during protein unfolding and refolding were measured using a PEAQ-DSC device (Malvern Panalytical). The proteins were diluted in 1x PBS pH 7.4 to a final concentration of 1.3-1.8 µM. Both the protein sample and the reference buffer (130 µL each) were transferred into the measuring cells and the temperature was adjusted (10-90 °C) with a heating rate of 60 °C/h. The thermograms were buffer-subtracted and evaluated using the Origin analysis software from MicroCal. The obtained transition peaks were fitted and integrated to determine the thermodynamic parameters.

2.2.5.5 Nuclear magnetic resonance (NMR) spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy represents an established method for the characterization of structural and dynamic properties of macromolecules in solution. NMR spectroscopy detects the resonance signals of magnetically active atomic nuclei that show a non-zero nuclear spin. In the context of protein structure determination and investigation of protein dynamics, the nuclei with spin such 1/2 as ^1H, ^{13}C and ^{15}N are of particular interest. In this work, NMR spectroscopy was used to investigate the secondary structure of the novel domain B variant CID 25. In addition, an NMR titration experiment with CID 25 and hIgG1-Fc was performed to localize Fc binding sites. All NMR measurements were performed by Dr. Ulrich Weininger (Institute of Physics, Faculty of Natural Sciences II, MLU Halle-Wittenberg). The NMR spectra were measured at 15 °C with Avance III 600 MHz and 800 MHz spectrometers (Bruker BioSpin) equipped with room temperature and cryoprobe heads, respectively. NMR experiments were performed on 1 mM ^{13}C and ^{15}N labeled CID 25 in 1x PBS, 10 % D_2O pH 6.5. The NMR spectra were recorded with 16 scans unless otherwise stated. The obtained NMR data were processed using the software NMRPipe (Delaglio *et al.*, 1995) applying zero filling and $\cos^2$ functions. NMR spectra were evaluated using NMRView (Johnson & Blevins, 1994). TALOS+ was used for deriving backbone φ/ψ dihedral angles and secondary structure (Shen *et al.*, 2009).

Two-dimensional NMR spectroscopy
To characterize the structure of $^{13}C^{15}N$ CID 25, several two-dimensional HSQC (heteronuclear single quantum coherence) experiments were performed. $^1H^{15}N$-HSQC spectra were measured to investigate all H-N correlations within the protein. Thus, the detected cross signals mainly result from the nitrogen located in the peptide bond and from nitrogenous side chains. The $^1H^{15}N$-HSQC spectra were recorded with 1024(^1H)×256(^{15}N) data points. For data evaluation,

only the low-field region of the proton dimension between 4.7 ppm and 12 ppm was used in which all amide protons were found. In addition, ^{1}H^{13}C-HSQC spectra were measured to detect all correlations between protons directly bound to carbons of the backbone or the side chains of the protein. ^{1}H^{13}C-HSQC spectra were recorded either with 8 scans and a data matrix of 1024(^{1}H)×512(^{13}C) points for aliphatic ^{13}C nuclei or with 16 scans and 1024(^{1}H)×256(^{13}C) data points for aromatic ^{13}C nuclei. A proton dimension between -3 ppm and 12 ppm was used for the data evaluation since all carbon protons were localized in the low-field region.

Three-dimensional NMR spectroscopy

Three-dimensional NMR experiments were performed to assign the cross peaks detected in the ^{1}H^{15}N-HSQC and ^{1}H^{13}C-HSQC spectra of CID 25. For all pulse sequences stated below, the name indicates the path of magnetization. In the HNCACB experiment, the resonances of ^{1}H and ^{15}N were correlated with the intra- and interresidue resonances of ^{13}C$_\alpha$ and ^{13}C$_\beta$. Using the scalar couplings of ^{1}J(N,H) ~90 Hz, ^{1}J(N,C) ~15 Hz and ^{2}J(N,C-1) ~7 Hz the amide proton can be linked via the amide nitrogen to both the carbon atom of its own amino acid ^{1}J(N,C) and to its predecessor amino acid ^{2}J(N,C-1).

In addition, a HN(CO)CACB experiment was performed to clearly distinguish between ^{1}J(N,C) and ^{2}J(N,C-1) resonances. Compared to HNCACB the magnetization was additionally transferred via the carbonyl group and only the resonances of the predecessor amino acid ^{2}J(N,C-1) were detected. The HNCACB and the HN(CO)CACB spectra were recorded with 1024(^{1}H)×64(^{15}N)×128(^{13}C) data points. In addition, ^{15}N-TOCSY-HSQC (total correlation spectroscopy) and HCCH-TOCSY experiments were performed to simplify the assignment of the side chains and to facilitate the identification of amino acid types. The ^{15}N-TOCSY-HSQC spectra were recorded with 1024(^{1}H)×64(^{15}N)×170(^{13}C) data points, whereas a matrix of 1024(^{1}H)×64(^{15}N)×256(^{13}C) data points was measured for the HCCH-TOCSY spectra.

NMR titration experiment for the localization of binding sites

An NMR titration experiment with hIgG1-Fc was performed to identify amino acid residues of CID 25 involved in Fc binding. Four titration steps were prepared and the detailed experimental conditions can be found in Table 25. The samples were incubated directly in the NMR tube for 5 min. ^{1}H^{15}N-TROSY-HSQC (transverse relaxation-optimized spectroscopy) spectra were recorded for each titration step. The TROSY experiment is beneficial for larger proteins and complexes ($\geq$ 20 kDa). The uncoupling of the dipole-dipole mechanism from the chemical shift anisotropy (CSA) generates a line narrowing of the resonance signals. By increasing the spectral resolution, sharp peaks can be detected in the spectrum. The ^{1}H^{15}N-TROSY-HSQC spectra were measured with 8 (1st titration step) or 64 scans each with 1024(^{1}H)×256(^{15}N) complex points. The obtained spectra were processed as described above. The peak volumes were determined by integration of the picked area. The experimental error was calculated based on the experimental noise adjusted for differences in concentration and number of scans in the four sub experiments.

Table 25: Overview of the experimental conditions of the NMR titration

Step	$V_{CID\ 25}$ [µL]	$V_{hIgG1\text{-}Fc}$ [µL]	$c_{CID\ 25}$ [µM]	$c_{hIgG1\text{-}Fc}$ [µM]	Molar ratio [CID 25:hIgG1-Fc]
1	550	0	91	0	1:0.0
2	550	38	85	30	1:0.35
3	588	55	78	70	1:0.89
4	643	89	68	130	1:1.91

3 Results

Hey *et al.* already described the suitability of helix bundle proteins as scaffolds for the generation of engineered binding proteins (Hey *et al.*, 2005). Within the scope of the present work, IgG-binding proteins based on three-helix bundle domain B of Protein A were generated and characterized. First, the influence of loops on folding and function of domain B was investigated. Second, helix shuffling was used to achieve a new arrangement of the α-helical structures. Two alternative loops with several randomization positions were inserted to generate diversity and connect the rearranged helices of the bundle (Figure 5). IgG-binding proteins were obtained by directed evolution and selected variants were characterized in detail.

3.1 Production and characterization of domain B variants

Molecular dynamics (MD) simulations of domain Z, an engineered B domain with A1V and G29A mutations were performed by PD Dr. Iris Thondorf (MLU Halle, WG Molecular Modelling, PDB: 1H0T) and revealed a variety of non-covalent interactions within the protein structure (Figure 8). Both amino acids in α-helical regions and residues in loops contribute to the formation of van-der-Waals interactions and hydrogen bonds. Interestingly, 50 % of all hydrogen bonds within the protein are formed by amino acids in the loop between helix 1 and helix 2. In particular, N21 and L22 form 2/3 of the entire non-covalent interactions between the α-helical structures and the loops.
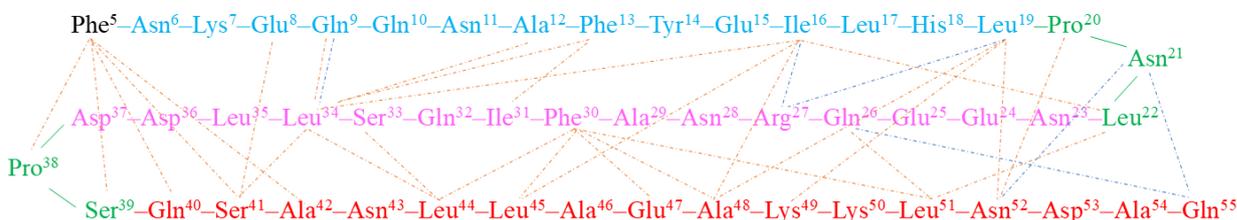
Figure 8: Interactions between the amino acids of a minimized average structure of domain Z (PDB: 1H0T). MD simulations of domain Z based on the PDB structure 1H0T. Non-covalent interactions were generated using Molecular Operating Environment (MOE) with filtering energy < -0.2 kcal/mol and > 4 residuals as distance. Starting at the N-terminus, a sequential numbering and an individual coloring of structural elements and amino acids are used: helix 1 (blue), helix 2 (magenta), helix 3 (red) and loops (green). Non-covalent interactions are marked as dashed lines: van-der-Waals interactions (orange) and hydrogen bonds (blue).

To experimentally investigate the influence of domain B loops on secondary structure and IgG-binding properties, five variants were produced in which the amino acids of the loops were partially or completely substituted for $(GS)_n$ linkers (Table 26). A consecutive identification number (clone ID: CID) was assigned for the variants and their affinity for hIgG1-Fc was investigated by SPR (2.2.5.1). In addition, the proportion of α-helical structures of the variants was determined by CD (2.2.5.3). For this, the gene sequences (Twist Bioscience) optimized for the expression in *E. coli* cells were subcloned in the expression vector pNP-013 using the restriction enzyme BsaI-HFv2 (2.2.1.5 and 2.2.1.6). After verification of correct sequences (2.2.1.7), the plasmids were used to transform *E. coli* cells (2.2.3.2). Five variants (CID 2-6) and wild type (wt) domain B (CID 1) were produced on shake flask scale (2.2.3.3). After inoculation of H15 autoinduction medium, recombinant *E.coli* cells were grown overnight at

37 °C and 20×g acceleration. As a result of this cultivation strategy, optical densities of $OD_{600} = 44\text{-}50$ were achieved. All variants showed exclusively soluble gene expression comparable to wild type domain B (data not shown). To obtain a pure tagfree product and monomeric protein species, the variants were purified using an ÄKTAxpress chromatography system (2.2.4.2). All variants were successfully obtained as monomeric tagfree proteins with yields ranging from 8.7-26.6 mg/L culture volume and 0.2-0.5 mg/g wet biomass (Table 26). For CID 3 and CID 5, additional peaks corresponding to the dimeric form of the proteins were observed in the chromatogram (Figure 9 A). However, these peaks were completely separated from the monomeric protein species. To monitor the progress of protein purification, fractions of the load, wash, and elution steps were analyzed by SDS-PAGE followed by Coomassie visualization (2.2.4.5). In the fractions of monomeric protein species, only bands that corresponded to the target structure were detected (exemplary CID 6, Figure 9 B).
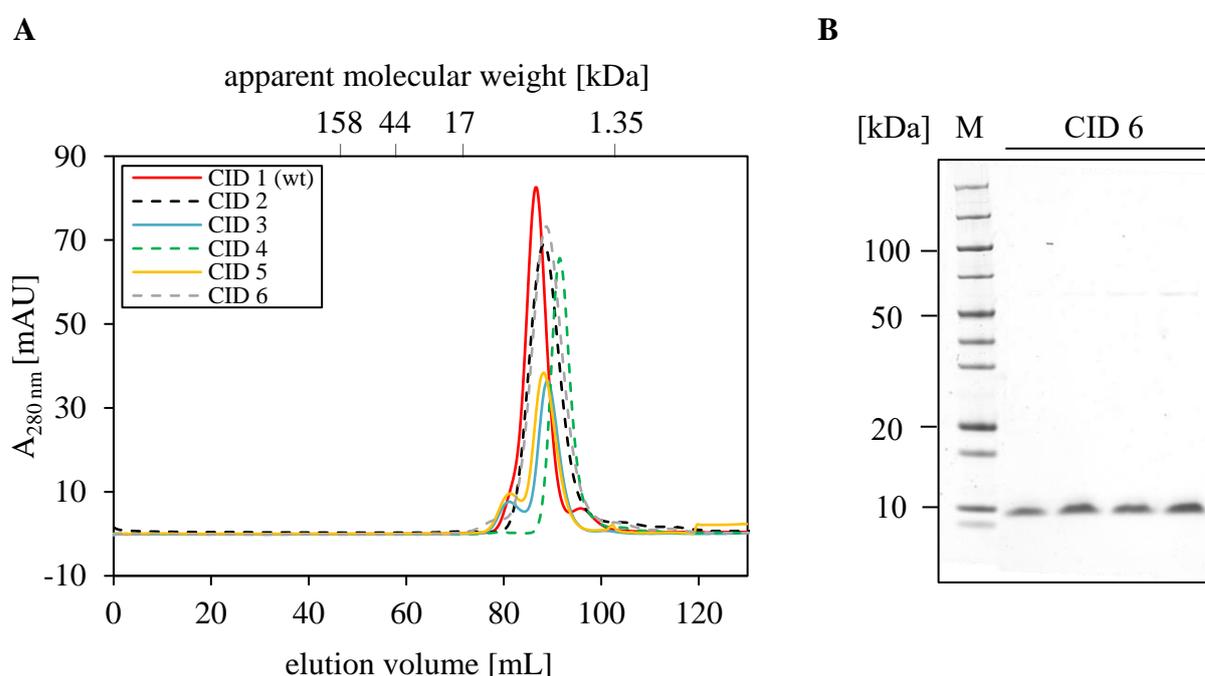
Figure 9: Purification of five domain B variants (CID 2-6) and wild type control (CID 1) by SEC.
(A) SEC of CID 2-6 purified by IMAC. Wild type domain B (CID 1) is shown as comparison (red curve). The samples were loaded on an equilibrated HiLoad XK16/60 Superdex 75 prep grade column. 1x PBS pH 7.4 was used as running buffer **(B)** SDS-PAGE analysis of the obtained monomeric species of CID 6 after SEC. 2 µg protein per lane was transferred into the slots of a NuPAGE Novex 4-12 % Bis-Tris gradient gel. Bands were visualized by Coomassie staining. M: PageRuler Unstained Broad Range Protein Ladder (Thermo Scientific).

Table 26: Yields of produced domain B variants

CID	Residues L19[loop 1]N23	Residues D37[loop 2]Q40	Yield per culture volume [mg/mL]	Yield per wet biomass [mg/g]
1 (wt)	PNL	PS	26.6	0.5
2	GSGS	PS	22.3	0.4
3	PNL	GSGS	8.7	0.2
4	GSGS	GSGS	18.7	0.4
5	PSGS	PS	9.6	0.2
6	GNL	GSGS	21.0	0.4

* The numbering of the residues was inspired by PDB: 1H0T and UniProt: 02976.

SPR analysis was performed to determine the affinity of the produced domain B variants for hIgG1-Fc (2.2.5.1). For this, hIgG1-Fc was immobilized on the surface of a CM5 chip using NHS chemistry and binding of serial diluted variants to the target was measured. The obtained curve progressions of each variant were approximated with a Langmuir fitting model assuming a 1:1 stoichiometry and resulted in affinities in a range of $K_D = 10^{-7}\text{-}10^{-8}$ M (Figure 10, Table 27). All variants showed specific binding to hIgG1-Fc. However, compared to wild type domain B, a decrease in affinity was observed when amino acids of the natural loops were partially substituted for $(GS)_n$ linkers (Table 27). Interestingly, CID 2, CID 4, and CID 5 in which the interactive amino acids asparagine (N21) and leucine (L22) were substituted, showed a significantly decreased affinity by a factor of 16-25 (Table 27). Moreover, an increase in the dissociation rate k_{off} was determined for these variants compared to wild type domain B (Figure 10). An overview of the complete data set can be found in the appendix (Figure A 2).
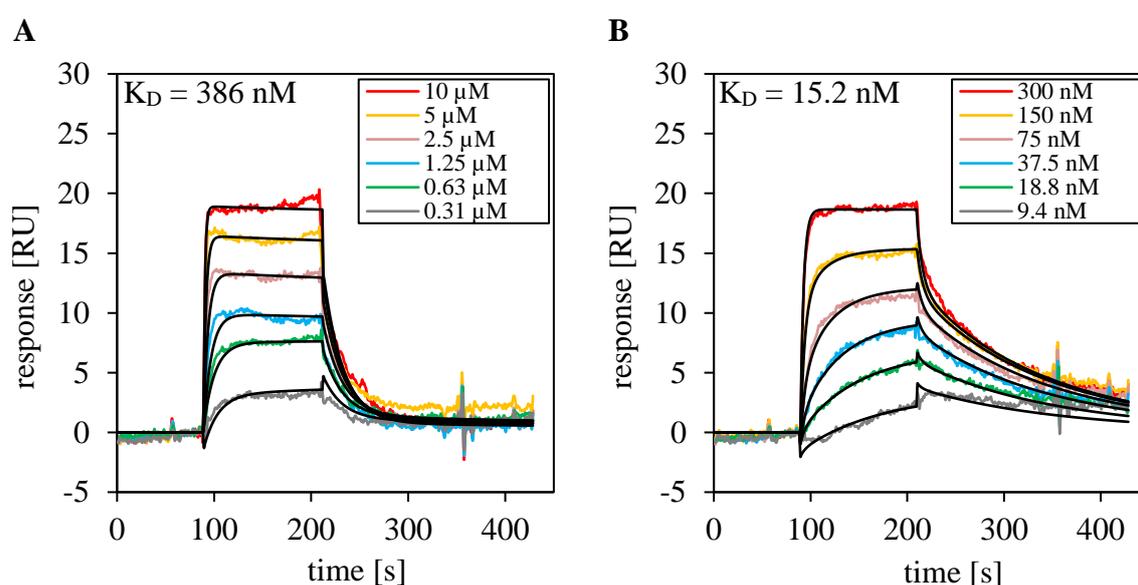
Figure 10: SPR sensorgrams of CID 4 and wild type control CID 1.
SPR sensorgrams of serial diluted (**A**) CID 4 and (**B**) wild type domain B (CID 1) are shown as examples. hIgG1-Fc (~850 RU) was immobilized on the surface of a CM5 sensor chip using NHS chemistry. 1x PBST 0.05 % (v/v) pH 7.4 was used as running buffer. The association and dissociation steps were performed for 120 sec each at a flow rate of 30 μL/min. After each cycle, proteins on the surface were regenerated with 10 mM glycine pH 2.0 for 30 sec. Kinetic parameters k_{off} and k_{on} were determined with a Langmuir fitting model assuming a 1:1 binding stoichiometry (black lines).

Table 27: Kinetic parameters of IgG-binding domain B variants determined by SPR

CID	Residues L19[loop 1]N23[*]	Residues D37[loop 2]Q40[*]	K_D [nM]	k_{on} [1/Ms]	k_{off} [1/s]
1 (wt)	PNL	PS	15.2	3.1×10^5	4.7×10^{-3}
2	GSGS	PS	252	8.5×10^4	2.1×10^{-2}
3	PNL	GSGS	46.8	1.3×10^5	3.4×10^{-3}
4	GSGS	GSGS	386	7.3×10^4	2.8×10^{-2}
5	PSGS	PS	311	7.7×10^4	1.1×10^{-2}
6	GNL	GSGS	111	4.1×10^4	3.3×10^{-3}

[*] The numbering of the residues was inspired by PDB: 1H0T and UniProt: 02976.

CD measurements (2.2.5.3) were performed to determine the secondary structure of the five domain B variants (CID 2-6) and the wild type control (CID 1). After buffer exchange (2.2.4.7), CID 5 showed precipitation and was excluded for further analysis. The obtained CD spectra were baseline corrected by subtracting a buffer spectrum. The mean residue ellipticity was calculated using Equation 4 and plotted as a function of wavelength (Figure 11). Based on the measured CD spectra, the proportion of α-helical structures was calculated using the reference by Yang included in the device software (Yang *et al.*, 1986). CID 3 and CID 6 showed a spectrum and a calculated α-helical proportion comparable to wild type domain B (Figure 11, Table 28). This result is consistent with the theoretical α-helical proportion of domain B (PDB: 1BDD, Gouda *et al.*, 1992), although the obtained spectral amplitudes were slightly decreased compared to CD spectra of different Protein A domains found in the literature (Myrhammar *et al.*, 2020). In contrast to wild type domain B, a significant increase in molar ellipticity at 210 nm and a decrease in α-helical structures were observed for CID 2 and CID 4, respectively (Figure 11, Table 28). However, substitution of natural loops for (GS)$_n$ linkers results in functional proteins (Figure 10, Table 27) with α-helical secondary structure (Figure 11, Table 28). Since both loops exhibit some degree of tolerance towards engineering, libraries based on domain B with alternative loops and shuffled helices were constructed (3.2).
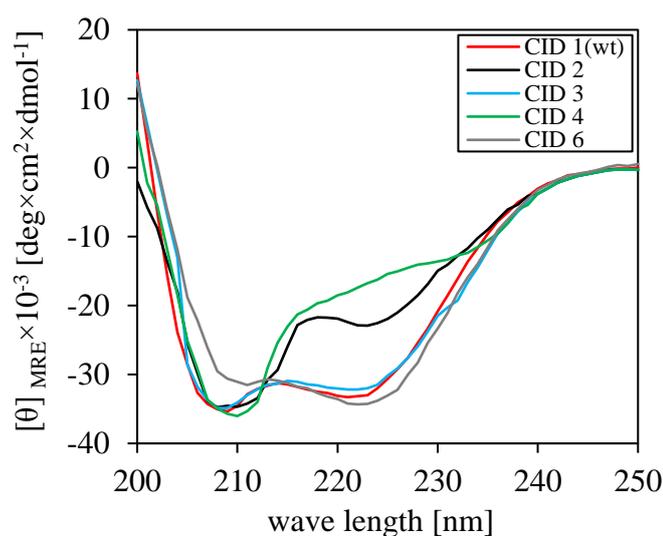
Figure 11: CD spectra of four domain B variants and wild type control CID 1.
CD analysis of CID 2-4, CID 6, and wild type domain B (CID 1) purified by IMAC and SEC. All variants [c = 0.2 mg/mL] were transferred into a 1 mm quartz cuvette. CD measurements were performed at 20 °C in the far-UV spectral region (200-250 nm) at a speed of 50 nm/min and a bandwidth of 1 nm. 16 spectra per sample were measured and baseline corrected by subtracting a buffer spectrum. The mean residue ellipticity is plotted as a function of wavelength.

Table 28: Proportion of calculated α-helical structures of domain B variants

CID	Residues L19[loop 1]N23[*]	Residues D37[loop 2]Q40[*]	Helix [%]
1 (wt)	PNL	PS	79.4
2	GSGS	PS	59.6
3	PNL	GSGS	81.4
4	GSGS	GSGS	53.1
6	GNL	GSGS	79.5

* The numbering of the residues was inspired by PDB: 1H0T and UniProt: 02976.

3.2 Design and generation of novel domain B variants

In this work, the B domain of Protein A was used as scaffold for the generation of novel binding proteins. Compared to wild type domain B, helix 1 and helix 3 of the three-helix bundle were swapped (1.4, Figure 5). Two alternative loops with three to five (loop 1) and six to eight (loop 2) randomized positions were inserted to connect the α-helical structures and generate diversity (Figure 6). To favor the generation of functional proteins with domain B-like fold, the length of the alternative loops was chosen according to the determined direct distance between the individual helices (PDB: 1BDD). In the following sections, the randomized loops are named RF1 (loop 1 between helix 3 and helix 2) and RF2 (loop 2 between helix 2 and helix 1).

3.2.1 Generation of domain B libraries

Domain B libraries were generated at DNA level and the gene fragments were divided into modular sections (2.2.2.1, Figure 6). Degenerate oligonucleotides from Microsynth Seqlab were used to generate diversity. Prefabricated trinucleotides encoding the amino acids cysteine, phenylalanine, tryptophan, and tyrosine were excluded from the synthesis. First, double-stranded DNA was generated based on diversified oligonucleotides. Using complementary sequences at the 5' and 3' termini, gene fragments were sequentially assembled to an overall construct suitable for ribosome display (2.2.2.1). In this way, 9 domain B libraries were generated, which had 9-13 randomized positions each, depending on the length of the randomized loops RF1 and RF2 (Table 29). DNA libraries with comparable theoretical diversities at DNA level were mixed in equal proportions, resulting in four libraries named PACB1-4. The diversity of PACB1-4 in RD-format was calculated based on the amount of purified DNA using Equation 2. As a result, 1.8×10^{10}-1.7×10^{12} domain B variants of each library were generated at protein level (Table 29).

Table 29: Overview of generated domain B libraries and their diversities at protein level

Library	# Randomized positions in RF1	# Randomized positions in RF1	Theoretical diversity	Diversity in RD format
PACB1	3	6	1.8×10^{10}	1.8×10^{10}
PACB2	3	7	1.1×10^{12}	1.1×10^{12}
	4	6	1.1×10^{12}	
PACB3	3	8	1.7×10^{13}	1.7×10^{12}
	4	7	1.7×10^{13}	
	5	6	1.7×10^{13}	
PACB4	4	8	2.8×10^{14}	1.7×10^{12}
	5	7	2.8×10^{14}	
	5	8	4.5×10^{15}	

Sequence analysis of 48 representative clones from each domain B library revealed a moderate approximation of amino acids in the randomized loops compared to the theoretical equal distribution of 6.25 % (Figure 12). Both codons of serine (S) and leucine (L) were slightly underrepresented compared to theory. Only single codons of cysteines (C) and aromatic amino acids (Y, F, and W) were detected and could be attributed to single point mutations. In addition, only a slight proportion of sequences with stop codons (~3 %) and frame shift mutations (~2 %) were observed.
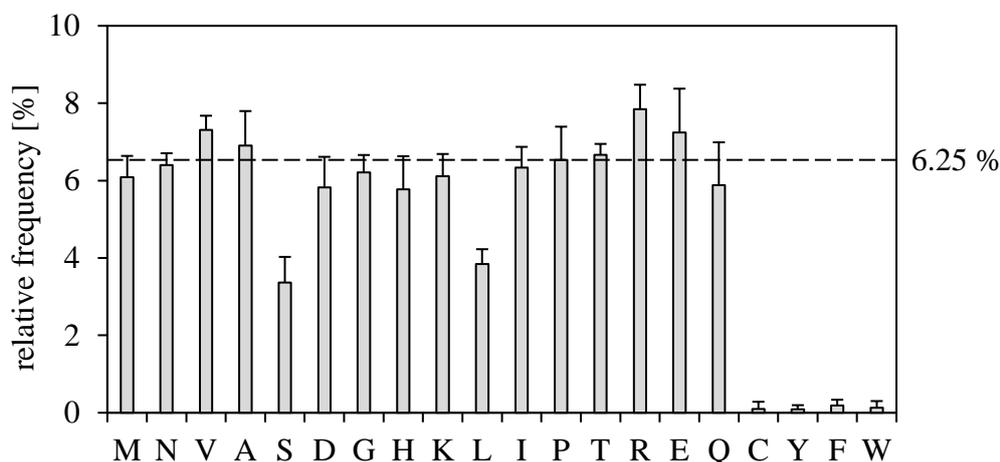
Figure 12: Amino acid distribution in the randomized loops of the domain B libraries PACB1-4.
The relative codon frequency of individual amino acids over all randomized positions of the four libraries PACB1-4 is averaged. For the calculation, 48 functional sequences of each library were used. The theoretical equal distribution of 6.25 % for codons of 16 amino acids (the library design excluded cysteines, phenylalanine, tryptophan, and tyrosine) is shown as black dashed horizontal line. Error bars represent the standard deviations over all randomized positions in all four libraries.

3.2.2 Selection of domain B libraries versus hIgG1-Fc by ribosome display

To enrich domain B variants with IgG-binding properties from complex libraries, ribosome display was used as selection method since a complete screening of all variants was technically not feasible. Ribosome display as *in vitro* selection technique offers the advantage of using larger library sizes compared to cell-based *in vivo* selection systems (He & Taussig, 2002). After construction of domain B libraries, four rounds of biopanning were performed and different selection conditions were applied (2.2.2, Table 24). Biotinylated hIgG1-Fc was immobilized on the surface of magnetic beads coated with streptavidin or neutravidin and selection pressure was increased between the selection rounds. During selection, the concentration of biotinylated hIgG1-Fc was reduced and an off-rate selection with non-biotinylated hIgG1-Fc was performed in round four to obtain domain B variants with high affinity (Table 24). The binding capacity of streptavidin and neutravidin beads for biotinylated hIgG1-Fc was determined in preliminary experiments (data not shown). Magnetic beads without immobilized hIgG1-Fc were used as negative control, whereas one selection strain that contained wild type domain B served as positive control. The formed ternary complexes were dissociated with EDTA-containing buffer and the eluted mRNA was transcribed into cDNA (2.2.2.4). The obtained DNA of the selection pools was amplified and subcloned in the expression vector pNP-013, which additionally encoded GFP (2.2.2.5). After expression of recombinant genes (2.2.3.3), the success of selection was determined by GFP-based pool

ELISA (2.2.4.9). The protein pools were transferred into the cavities of a 384-well high binding microtiter plate previously coated with hIgG1-Fc (on-target) and BSA (off-target). Proteins bound to the targets were detected via the fluorescence intensity of fused GFP at 483 nm excitation and 535 nm emission. From the third round of selection, a significant enrichment of domain B variants that specifically bound to hIgG1-Fc was detected for all four libraries (Figure 13). The protein pool of PACB1 obtained after the fourth round of selection showed a slight decrease in binding to hIgG1-Fc compared to the previous round.
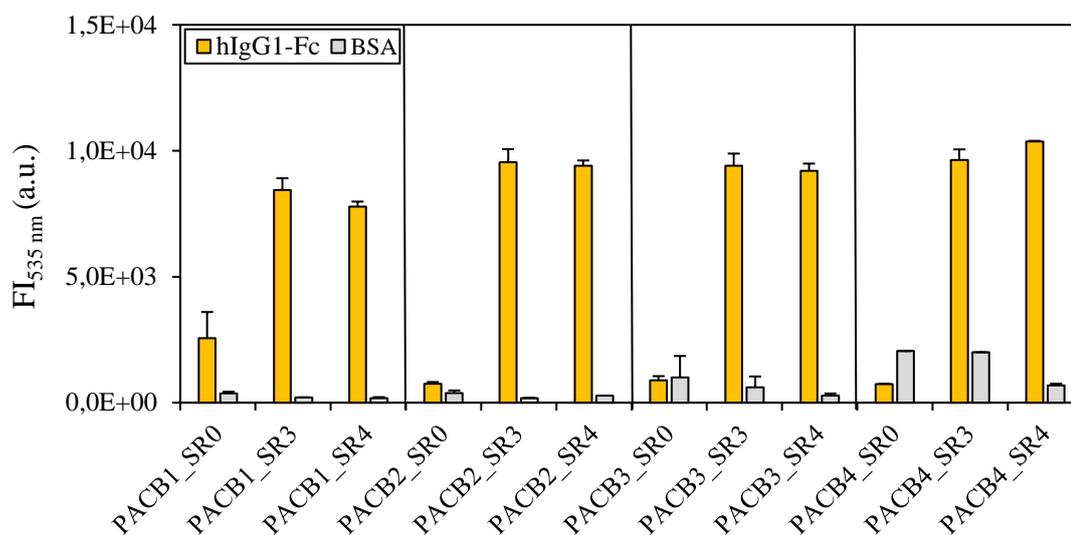
Figure 13: GFP-based pool ELISA of the libraries PACB1-4 selected versus hIgG1-Fc via ribosome display. After each selection round (SR), the obtained DNA was subcloned in the expression vector pNP-013 that additionally encoded GFP. For the ELISA, hIgG1-Fc (orange) and BSA (grey) were immobilized on the surface of a 384-well high binding microtiter plate [c = 40 nM]. After gene expression and cell lysis, the protein pools were transferred into the cavities of the microtiter plate. Bound domain B variants were detected via the fluorescence intensity of fused GFP at 483 nm excitation and 535 nm emission. All samples were analyzed as triplicates to determine measurement uncertainties. Standard deviations are shown as black error bars.

3.2.3 NGS of domain B libraries PACB1-4 after selection versus hIgG1-Fc

To fundamental characterize the domain B libraries PACB1-4, pools obtained after the fourth round of selection were selected for NGS (2.2.1.7, Figure 13, SR4). A total of 1,284,664 sequences were obtained from NGS (Table 30). The received sequences were evaluated using a company-owned software (Heiko Richter, Navigo Proteins GmbH). The number of sequences from each library was unevenly distributed due to slight pipetting inaccuracies during photometric determination of DNA concentration (2.2.1.2). The libraries PACB1-4 contained between 57.3 % and 83.4 % functional sequences, of which 1/3 were variants with additional point mutations (Table 30). Sequences grouped under "Others" contained either stop codons, insertions, or codons of cysteines (≤ 5 %) and, like sequences with a shifted reading frame, were excluded from further analyses. Compared to PACB1-3, a significantly increased proportion of sequences with frame shift mutations was detected for PACB4 (Table 30). All functional sequences of each library were used for the statistical evaluation.

Table 30: NGS of the libraries PACB1-4 after the fourth selection round versus hIgG1-Fc

Library	# Sequences	Functional sequences [%]	Frame shift mutations [%]	Others [%]
PACB1	427,440	83.4	4.3	12.3
PACB2	218,201	81.9	6.0	12.1
PACB3	349,339	82.4	6.0	11.6
PACB4	289,684	57.3	33.5	9.2

In RF1, codons of the hydrophobic amino acid leucine (L) and highly flexible glycine (G) as well as hydrophilic threonine (T) were enriched compared to the initial libraries (Figure 14, Figure 12). Codons of several amino acids (M, N, V, D, H, K, R, E, D, and Q) were underrepresented in RF1 compared to the theoretical equal distribution of 6.25 %. A high diversity was observed for RF2. Only codons of glycine (G), threonine (T), arginine (R), valine (V) and alanine (A) were slightly enriched compared to the initial libraries (Figure 14, Figure 12). Even after the fourth round of selection, codons of amino acids excluded from the library construction (C, F, Y, and W) were not significantly represented ($< 0.5 \%$).
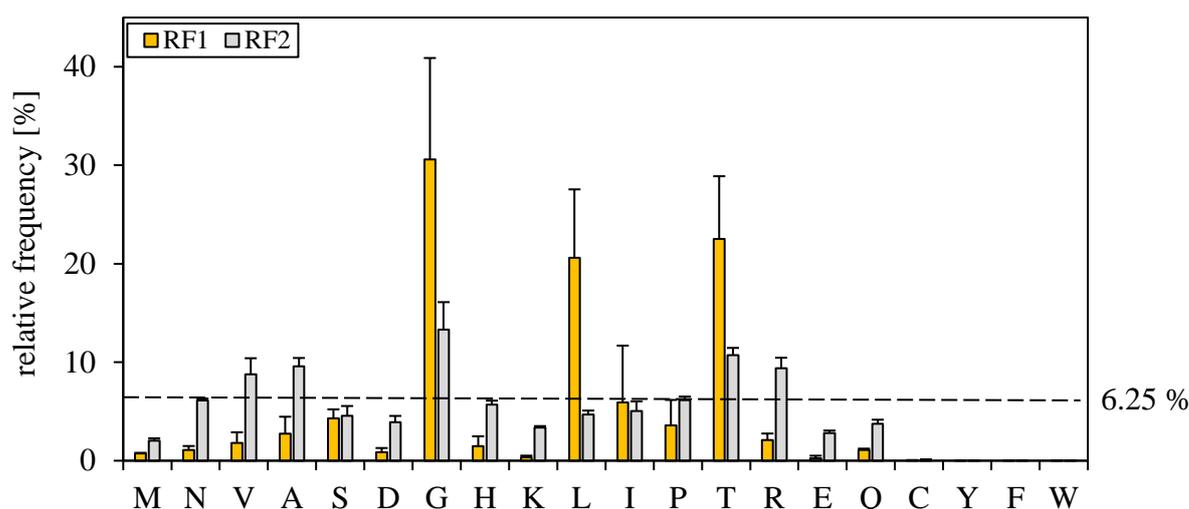
Figure 14: Amino acid distribution in the randomized loops of PACB1-4 after selection versus hIgG1-Fc. The relative codon frequency of the individual amino acids in the randomized loops RF1 and RF2 is averaged over all randomized positions of the domain B libraries PACB1-4. All functional sequences obtained from NGS after the fourth round of selection were used for the calculation (Table 30). The theoretical equal distribution of 6.25 % for codons of 16 amino acids (the library design excluded cysteines, phenylalanine, tryptophan, and tyrosine) is shown as black dashed horizontal line. The error bars represent the standard deviations over all randomized positions in the four libraries.

Sequences from four of the nine sub-libraries were obtained (Figure 15, Table 29). For PACB1, two specific sequence motifs (G/PLT) were identified in RF1 encoding either proline (P) or glycine (G) followed by leucine (L) and threonine (T). Sequences obtained from PACB2-4 showed also present sequence motifs ((P)GGI/LT) in RF1 encoding proline (P), only in PACB4, two glycines (G) followed by either leucine (L) or isoleucine (I) and threonine (T). In addition, accumulations of certain codons at individual positions were detected in RF2. For PACB1-4, codons of hydrophobic valine (V) were increasingly observed at the first position in RF2, whereas codons of small glycine (G) were mainly detected at the fifth position (only PACB1).
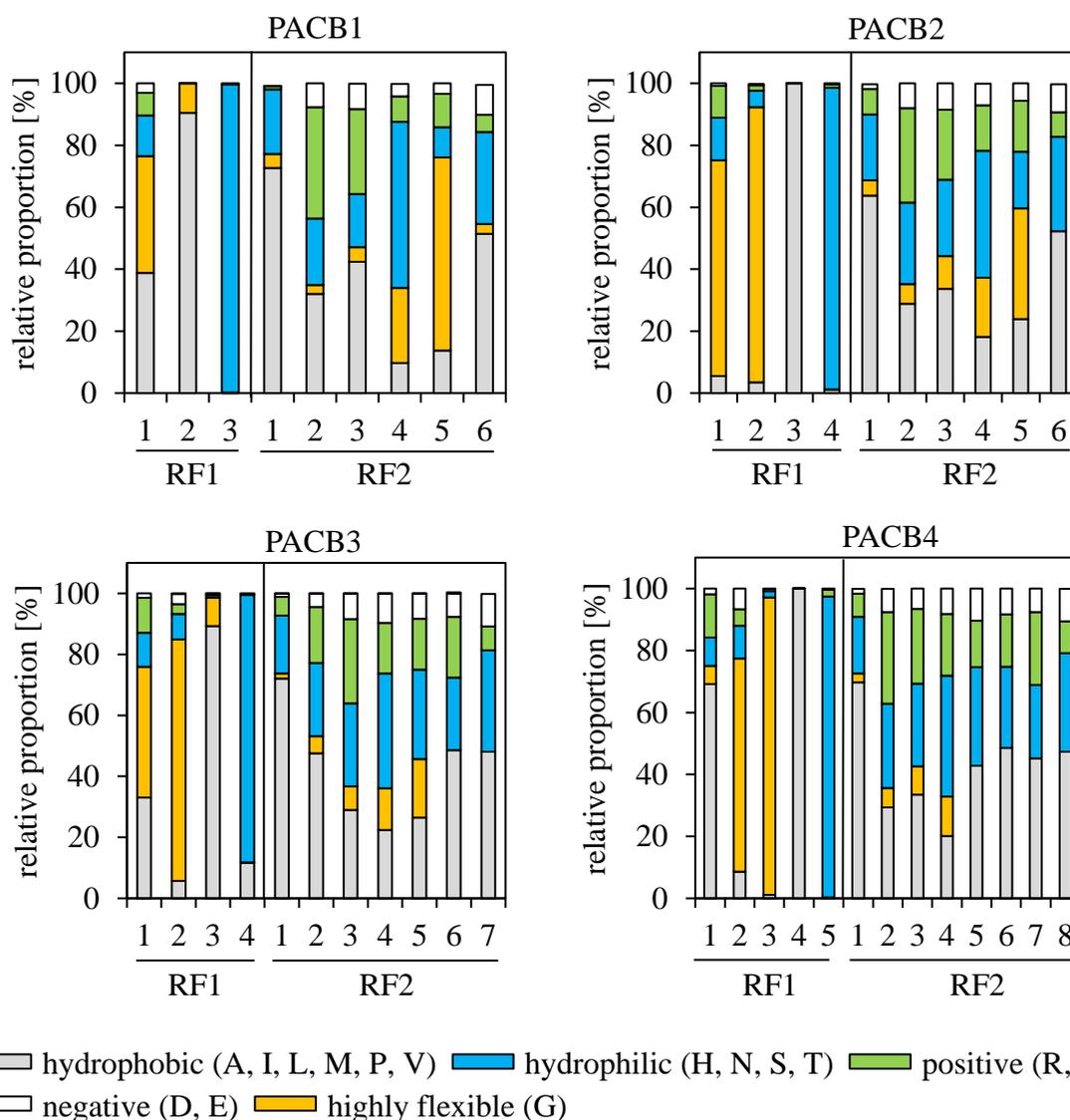
Figure 15: Relative proportion of amino acids in the randomized loops of the libraries PACB1-4 after selection versus hIgG1-Fc.
All functional sequences of the libraries PACB1-4 obtained from NGS after the fourth round of selection were used for the calculation (Table 30). The codon frequency of the individual amino acids was determined separately for each position in the randomized loops RF1 and RF2. The amino acids are classified according to their biochemical properties and assigned to the following groups: hydrophobic (A, I, L, M, P, and V), hydrophilic (H, N, S, and T), positive (K, R, and Q), negative (D and E), and highly flexible (G).

3.3 Identification and characterization of novel domain B variants

To identify individual domain B variants with IgG-binding property, proteins were produced and characterized on microgram scale. First, primary analysis was performed automatically using high-throughput (HTS) ELISA-based screening to obtain the broadest possible base of variants to choose for detailed analyses (2.2.4.9). Next, binding of selected variants to Fc was analyzed by BLI and SPR. Finally, a set of 24 IgG-binding variants was selected for sequence analysis.

3.3.1 Automated high-throughput screening (HTS)

For the primary analysis of individual domain B variants, an ELISA-based screening was performed using a multi-axis robot arm integrated in an automation system (2.2.4.9). The libraries PACB1-4 obtained after the fourth selection round (3.2.2) were subcloned in the expression vector pNP-013, which additionally encoded GFP. *E. coli* cells were transformed with the plasmids (2.2.3.2) and bacterial colonies were isolated on QTrays. A Qpix2 colony picker was used for the transfer of individual colonies into the cavities of 384-well microtiter plates (2.2.3.3). After subtracting the control wells (8× empty expression vector pNP-013 and 16× wild type domain B), 360 variants were produced per microtiter plate. 12,960 variants were screened with a comparable scope of 3,240 variants per library. The automated screening on 384-well scale included 36×384-well microtiter plates. After the variants were produced (2.2.3.3), *E. coli* cells were disrupted by freeze/thaw cycles (2.2.4.1) and the cell lysates were used for the automated screening (2.2.4.9). In parallel to protein production (2.2.3.3), samples of all investigated variants were stored as glycerol cultures (2.2.3.3). Both soluble proteins in the supernatant and proteins bound to hIgG1-Fc were detected by the fluorescence intensity of fused GFP measured at 483 nm excitation and 535 nm emission. 2,666 variants, so-called hits, showed an IgG-binding signal $\geq 3\times$SD of the mean of the negative control (empty expression vector pNP-013) and were selected for secondary screening (Table 31). Hence, hit rates between 10 % (PACB1) and 22 % (PACB2-4) were achieved.
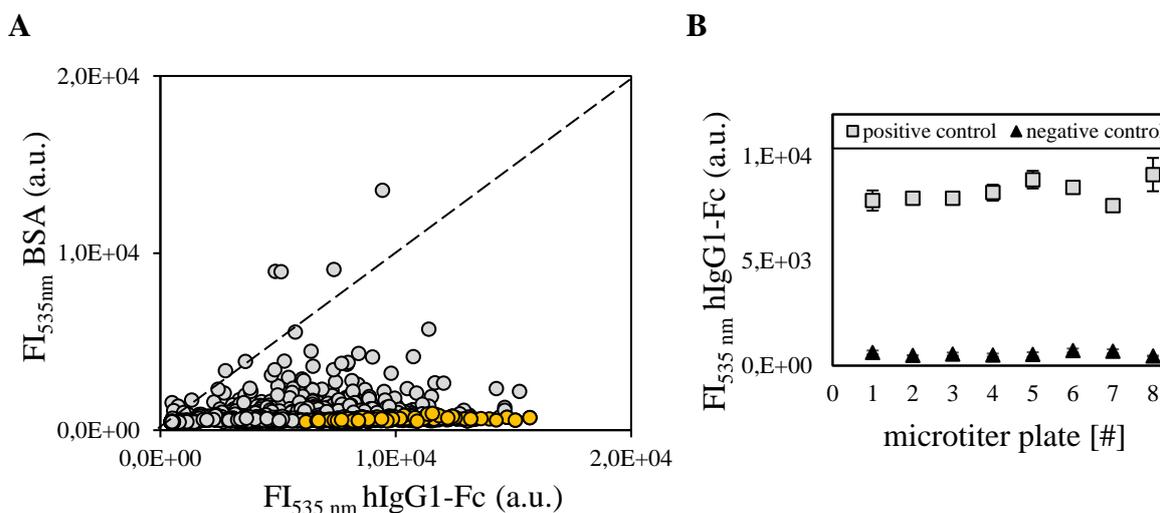
Figure 16: Automated GFP-based hit ELISA for the identification of IgG-binding domain B variants.
(A) 2,666 hits obtained from the primary screening of the libraries PACB1-4 after the fourth selection round were produced in 384-well microtiter plates. hIgG1-Fc (on-target) and BSA (off-target) were immobilized [c = 40 nM] on the surface of 384-well high binding microtiter plates. The bound proteins were detected by the fluorescence intensity of fused GFP at 483 nm excitation and 535 nm emission. The fluorescence intensities (FI) of domain B variants versus hIgG1-Fc (on-target) and BSA (off-target) are plotted. 461 hits (orange) were selected for further analyses by BLI and SPR. **(B)** Mean and standard deviation of 16× positive and 8× negative control per microtiter plate. Wild type domain B was used as positive control and *E.coli* cell lysates containing empty expression vector pNP-013 without subcloned domain B variant were used as negative control.

To investigate the specificity of the 2,666 IgG-binding hits obtained from primary screening, variants were repeatedly produced as described (2.2.3.3 and 2.2.4.1). Subsequently, hIgG1-Fc and BSA were immobilized on the surface of 384-well high binding microtiter plates and a GFP-based hit ELISA was performed as described (2.2.4.9). Except a marginal number (7.1 %), the variants showed IgG-specific binding signals (Figure 16 A). Based on protein expression, specificity, and absolute amplitude of the IgG binding signal, 461 hits were selected for µ-scale characterization by BLI and SPR (2.2.5.2) (Table 31). The signal variation of the assay controls over all microtiter plates was ≤ 5 % and corresponded well to values obtained from other in-house performed screenings (Figure 16 B).

Table 31: Hits obtained from screening of domain B libraries PACB1-4 versus hIgG1-Fc

Library	# Hits from primary screening	# Hits from secondary screening
PACB1	335	2
PACB2	827	65
PACB3	780	263
PACB4	724	131

3.3.2 Secondary analyses of novel domain B variants

To limit the number of domain B variants for a detailed characterization, the 461 hits obtained from the screening process (3.3.1) were produced on µ-scale (2.2.4.2) and analyzed by BLI (2.2.5.2) and SPR (2.2.5.1) with regard to their IgG-binding properties. For each variant, approximately 0.2 ± 0.01 mg/mL culture medium was purified, which was comparable to the yield of wild type domain B (0.25 mg/mL). After purification, binding of variants [c ≈ 2 µM] to hIgG1-Fc and Avelumab-Fab [c = 0.5 µM] was determined by BLI (2.2.5.2). As a result, 424 variants showed specific binding to hIgG1-Fc, whereas binding to Avelumab-Fab was not observed (data not shown). Interestingly, 37 variants showed no binding to hIgG1-Fc and Avelumab-Fab. 147 variants that showed the highest affinity for hIgG1-Fc compared to the others were selected for tagfree production and determination of IgG-binding by SPR (2.2.5.1). For this, proteolytic cleavage using TVMV protease with subsequent purification on µ-scale was performed (2.2.4.2). Approximately 0.12 ± 0.01 mg tagfree protein per mg of GFP-fused variant was obtained, corresponding to a theoretical yield of approximately 75 %. After protein concentration was determined (2.2.4.6), binding of purified variants [c = 1 µM] to hIgG1-Fc was analyzed by SPR (2.2.5.1). Except five, all variants showed specific binding to hIgG1-Fc with different kinetic parameters k_{on} and k_{off} (Figure 17). The dissociation rate of the variants was significantly increased by at least factor 15 compared to wild type domain B (CID 1), whereas k_{on} was only slightly affected by a factor of 2. A set of 24 variants that showed the highest affinity for hIgG1-Fc compared to the others was selected for sequence analysis (2.2.1.7) and a consecutive identification number (clone ID: CID) was assigned (Table 32). An overview of the K_D values of the 24 variants can be found in the appendix (Figure A 3).
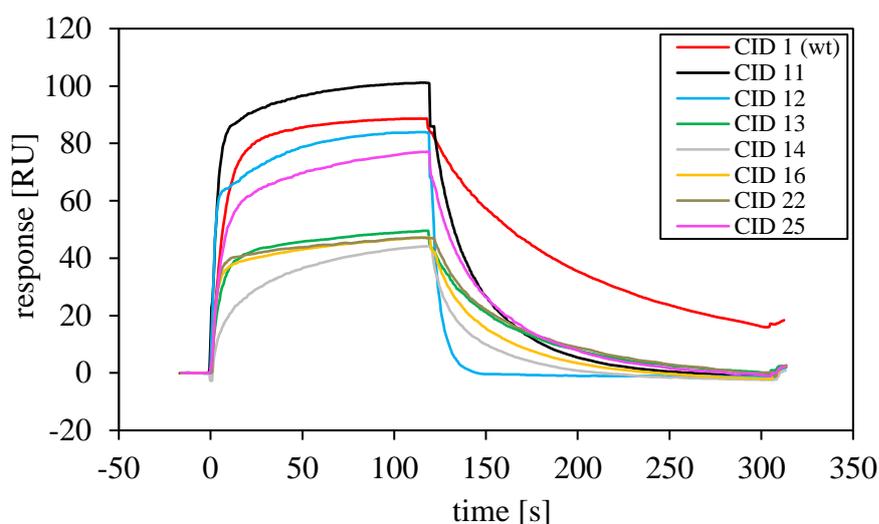
Figure 17: SPR sensorgrams of selected novel IgG-binding domain B variants.
SPR sensorgrams of seven novel domain B variants (CID 11-14, CID 16, CID 22, and CID 25) and wild type domain B (CID 1) are shown as example. hIgG1-Fc (~800 RU) was immobilized on the surface of a CM5 chip using NHS chemistry. The binding of domain B variants [c = 1 µM] to hIgG1-Fc was determined at a flow rate of 30 µL/min. The association and dissociation steps were performed for 120 sec each. 1x PBST 0.05 % (v/v) was used as running buffer and after each cycle proteins on the surface were regenerated with 10 mM glycine pH 2.0 for 30 sec.

3.3.3 Sequence analysis of novel domain B variants

Sequences of 24 variants detected in the automated screening process (3.3.1) with subsequent secondary analyses by BLI and SPR (3.3.2) were analyzed to fundamentally characterize the proteins (2.2.1.7). As a result, variants from the libraries PACB2 (12×), PACB3 (7×) and PACB4 (1×) each with four or five randomized positions in RF1 and six or seven randomized positions in RF2 were obtained (Table 32). Four of the examined variants contained a shifted reading frame at DNA level (4×) and were not used for further experiments. In addition, three variants showed signal overlaps due to mixed sequences. These variants were streaked out, the plasmid DNA of individual colonies was sequenced (2.2.1.7) and SPR analysis (2.2.5.1) was performed again.

Table 32: Sequence analysis of novel domain B variants obtained from screening

CID	RF1	RF2	Additional exchanges
1 (wt)	PNL	PS	/
7	GGLT	ASEKGA	/
8	GGLT	PVLRGKP	/
9	GGIT	ALRGHD	/
10	GGIT	AKRNPMT	/
11	GGLT	GTEGVH	/
12	GGLT	TAANPTA	/
13	GGLT	TKANPAR	/
14	GGLT	HTPTRG	/
15	QGLS	AVTHGA	N19S
16	GGLT	VRKLGP	/

CID	RF1	RF2	Additional exchanges
17	NGLT	VRTNGA	/
18	GGIT	VARAGP	E27D
19	RGLT	TKRHPKM	N19S
20	GGLT	VRRVGS	/
21	HGIS	VKRHPHQ	/
22	QGGLT	ALNSPTS	D35M, D36I, N42T
23	GGIT	AKANPEA	/
24	GGLT	MNVGKH	/
25	GGIT	AKTNPGN	/
26	GGLT	NRLTKD	/

In general, the analyzed variants showed high diversity in the randomized loops RF1 and RF2. In RF1, the variants showed two present motifs (GGI/LT) that consist of two glycines (G) followed by either leucine (L) or isoleucine (I) and threonine (T) (Table 32, Figure 18). Interestingly, in two variants the first glycine (G) of the motif was replaced by arginine (R) or glutamine (Q). A higher diversity was observed in RF2 compared to RF1. However, accumulations of certain amino acids at individual positions were detected (Figure 18). Small glycine (G) was cumulatively found at positions four and five in RF2 of PACB2 variants. Moreover, an enrichment of lysine (K) at position two, arginine (R) at position four and proline (P) at position five in RF2 was observed for PACB3 variants (Table 32, Figure 18). Four variants showed additional exchanges in the α-helical regions. However, these mutations were random in their position and type of inserted amino acid (Table 32).
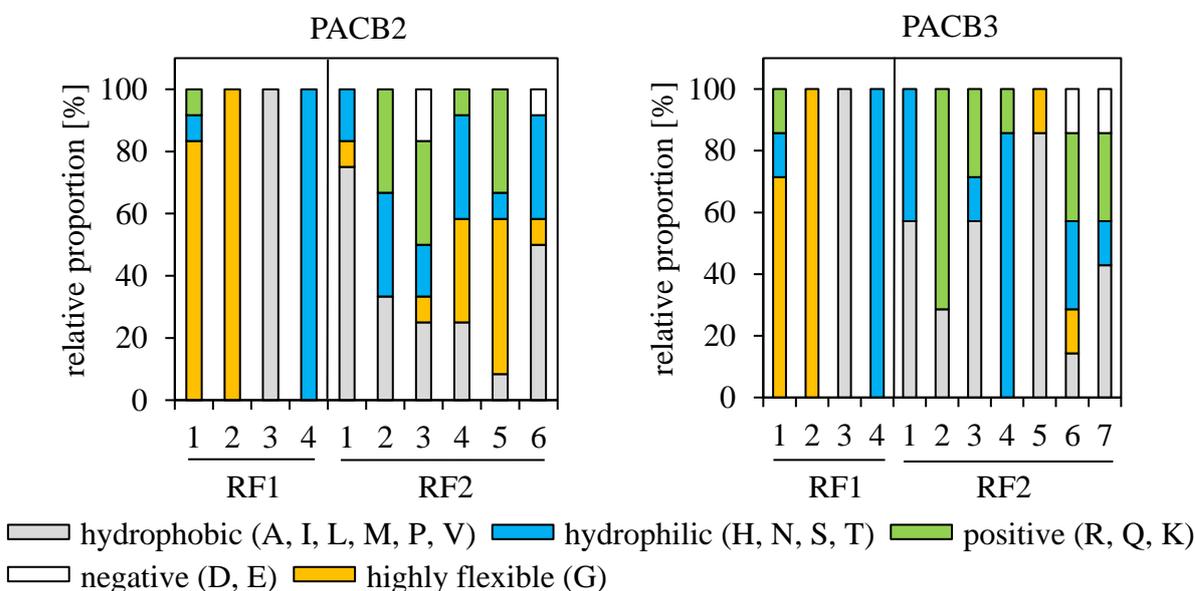
Figure 18: **Relative proportion of amino acids in the randomized loops of domain B variants obtained after selection and screening of the libraries PACB1-4.**
The sequences of 20 functional domain B variants obtained from the screening of the libraries PACB1-4 after selection round four were used for the calculation (Table 32). The number of occurring amino acids was counted separately for each position of the randomized loops RF1 and RF2 and is shown as relative proportion. The amino acids are classified according to their biochemical properties and assigned to the following groups: hydrophobic (A, I, L, M, P, and V), hydrophilic (H, N, S, and T), positive (K, R, and Q), negative (D and E), and highly flexible (G).

3.4 Production and characterization of novel domain B variants

A set of 20 domain B variants and wild type control (CID 1) were selected for protein production in shake flasks (Table 32). The variants were selected based on the obtained data with regard to solubility and IgG-binding properties (3.3.2). The sequence information was used as additional selection criterion to ensure diversity (3.3.3). After gene expression (2.2.3.3) and purification (2.2.4.2), binding of variants to hIgG1-Fc and Avelumab-Fab was analyzed by SPR (2.2.5.1) and ELISA (2.2.4.9). In addition, a competition experiment was performed by ELISA to investigate whether the novel domain B variants compete with wild type domain B for Fc binding. Furthermore, the secondary structure and the thermal stability of the variants were investigated by CD (2.2.5.3).

3.4.1 Production and purification of novel domain B variants

First, 20 domain B variants were produced in shake flasks for a detailed characterization by SPR, ELISA and CD. Recombinant *E.coli* cells were grown in H15 autoinduction medium. After inoculation of the medium, main cultures were grown overnight at 37 °C. As a result of this cultivation strategy, optical densities of $OD_{600} = 38\text{-}49$ and yields of 49-56 g wet biomass per liter culture medium were achieved. Exclusively soluble expression was observed for 16 variants but compared to wild type domain B these variants showed a slightly reduced expression level. CID 13, CID 15, CID 18, and CID 21 showed mainly insoluble expression and were excluded from the purification (Table 33).
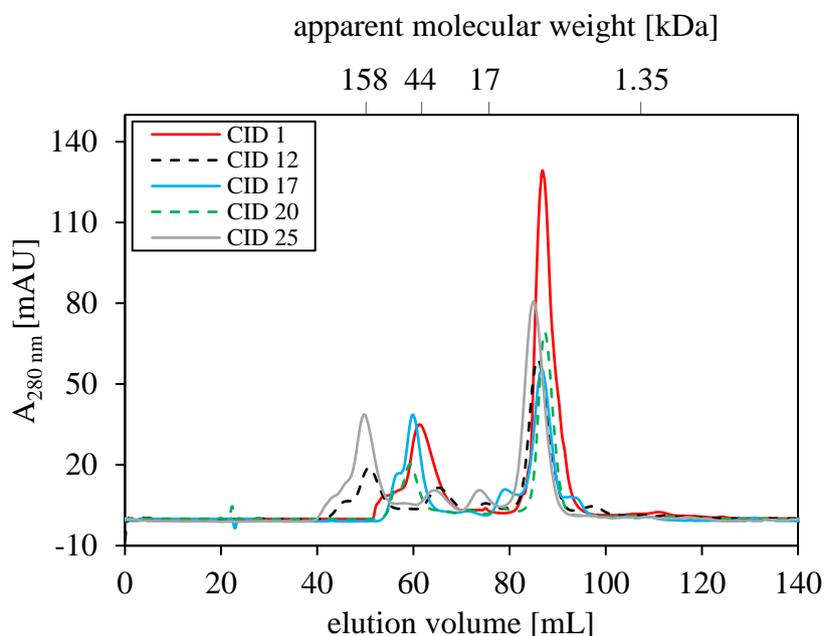
Figure 19: Purification of four novel domain B variants and wild type control (CID 1) by SEC.
SEC of the novel domain B variants CID 12, CID 17, CID 20, and CID 25 purified by IMAC. Wild type domain B (CID 1) is shown as comparison (red curve). Samples were loaded on an equilibrated HiLoad XK16/60 Superdex 75 prep grade column. 1x PBS pH 7.4 was used as running buffer. Monomeric variants were eluted at approximately 88 mL.

To obtain a pure tagfree product and monomeric protein species, variants were purified by IMAC and SEC using an ÄKTAxpress system (2.2.4.2). After IMAC, CID 19 and CID 22 showed precipitation and were not used further. The remaining variants showed additional peaks in the SEC chromatogram, corresponding to either multimers, aggregates or GFP-His as side product of the TVMV cleavage (Figure 19). However, these peaks were completely separated from the monomeric protein species (Figure 20 A). The yields of tagfree domain B variants were between 2.5-24.6 mg/L culture volume, which corresponded to 0.1-0.5 mg/g wet biomass (Table 33). To monitor the progress of protein purification, SEC elution fractions and soluble and insoluble fractions after cell lysis were analyzed by SDS-PAGE (2.2.4.5). Only the expected bands of variants were detected in the fractions of monomeric protein species (Figure 20 A). Bands indicating protein-based contamination were not observed. The purity of the pooled monomeric protein fractions was determined by SE-HPLC (2.2.4.8). As a result, 13 variants showed a monomeric peak (≥ 95 %) and a calculated recovery of 75.5-98 % (Figure 20 B, Table 33). Interestingly, one variant (CID 26) showed a significant proportion of aggregate (25.6 %) and a recovery of only 61.9 % (Table 33). However, no tendency towards aggregation was detected for the remaining variants. The variants showed a calculated molecular weight of 8.3-10.7 kDa, which was slightly increased compared to theory (data not shown). The SE-HPLC chromatograms of the 13 variants can be found in the appendix (Figure A 6).

A   **B**
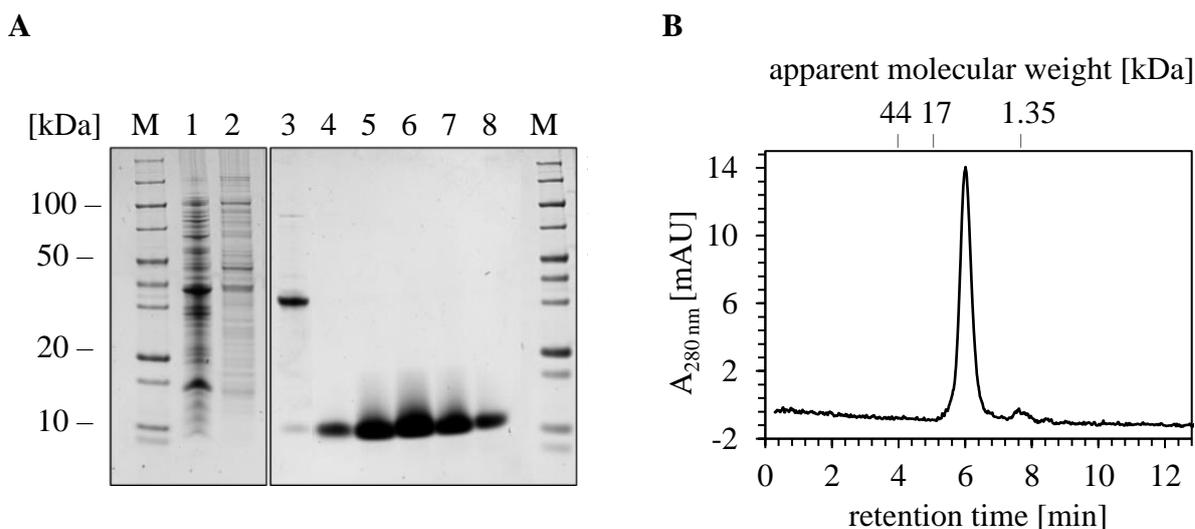
Figure 20: SDS-PAGE and SE-HPLC of the novel domain B variant CID 25 after SEC.
(**A**) SDS-PAGE of eluted tagfree monomeric species of CID 25 (lanes four, five, six, seven, and eight) after SEC. In addition, soluble (lane one) and insoluble (lane two) fractions after cell lysis of CID 25 fused to GFP-His10-tag were analyzed. A proteolytic cleavage was performed (lane three) to obtain tagfree CID 25 using histidine-tagged TVMV protease. A NuPAGE Novex 4-12 % Bis-Tris gradient gel was loaded with 10 µL sample and 5 µL PageRuler Unstained Broad Range Protein Ladder (M, Fisher Scientific). The bands were visualized by Coomassie staining. (**B**) SE-HPLC of the pooled elution fraction of tagfree, monomeric CID 25. The sample was loaded on a Tricorn 5/150 GL Superdex Increase 75 prep grade column (GE Healthcare) and 1x PBS pH 7.4 was used as running buffer.

Table 33: Yield and purity of produced novel domain B variants

CID	RF1	RF2	Yield [mg/L]*	Yield [mg/g]+	Purity [%]	Recovery [%]
1 (wt)	PNL	PS	40.1	0.8	99.8	88.9
7	GGLT	ASEKGA	2.5	0.1	98	75.5
8	GGLT	PVLRGKP	6.0	0.4	98	98
9	GGIT	ALRGHD	3.6	0.2	98.7	96.7
10	GGIT	AKRNPMT	2.3	0.1	96.8	98
11	GGLT	GTEGVH	21.3	0.5	98	84.4
12	GGLT	TAANPTA	20.3	0.4	97.1	70.7
13	GGLT	TKANPAR	n.p.	n.p.	n.m.	n.m.
14	GGLT	HTPTRG	14.4	0.3	95.7	75.7
15	QGLS	AVTHGA	n.p.	n.p.	n.m.	n.m.
16	GGLT	VRKLGP	24.4	0.4	95.8	75.8
17	NGLT	VRTNGA	22.0	0.5	98.7	82.3
18	GGIT	VARAGP	n.p.	n.p.	n.m.	n.m.
19	RGLT	TKRHPKM	3.7	0.1	n.m.	n.m.
20	GGLT	VRRVGS	23.4	0.5	96.4	71.3
21	HGIS	VKRHPHQ	n.p.	n.p.	n.m.	n.m.
22	QGGLT	ALNSPTS	4.1	0.1	n.m.	n.m.
23	GGIT	AKANPEA	3.7	0.2	98.5	98
24	GGLT	MNVGKH	22.6	0.5	98.9	75.9
25	GGIT	AKTNPGN	24.6	0.4	98	76.1
26	GGLT	NRLTKD	10.8	0.2	85.6	61.9

* = mg protein per L culture volume; + = mg protein per g wet biomass; n.m. = not measured; n.p. = not producible

3.4.2 Characterization of novel domain B variants

Binding analyses of selected domain B variants

ELISA and SPR analyses were performed to determine the affinity of novel domain B variants (Table 33) to hIgG1, hIgG1-Fc, and Avelumab-Fab. For SPR, targets were immobilized on the surface of a CM5 chip (2.2.5.1) and serial dilutions of recombinantly produced (2.2.3.3) and purified (2.2.4.2) variants were passed over the surface of the CM5 chip. Kinetic parameters were calculated using a Langmuir fitting model under assumption of a 1:1 binding stoichiometry. All variants showed specific binding to hIgG1-Fc and hIgG1 with affinities ranging from $K_D = 10^{-6}-10^{-8}$ M (Figure 21 A, Table 34). However, a significant increase in dissociation rate constant k_{off} was observed for the variants compared to wild type domain B (CID 1), whereas k_{on} was rather less affected. The sensorgrams of four domain B variants and the wild type control (CID 1) are shown as example in Figure 21. Interestingly, no binding of variants to Avelumab-Fab was observed (Figure 21 B). An overview of the complete data set can be found in the appendix (Figure A 4 and Figure A 5).
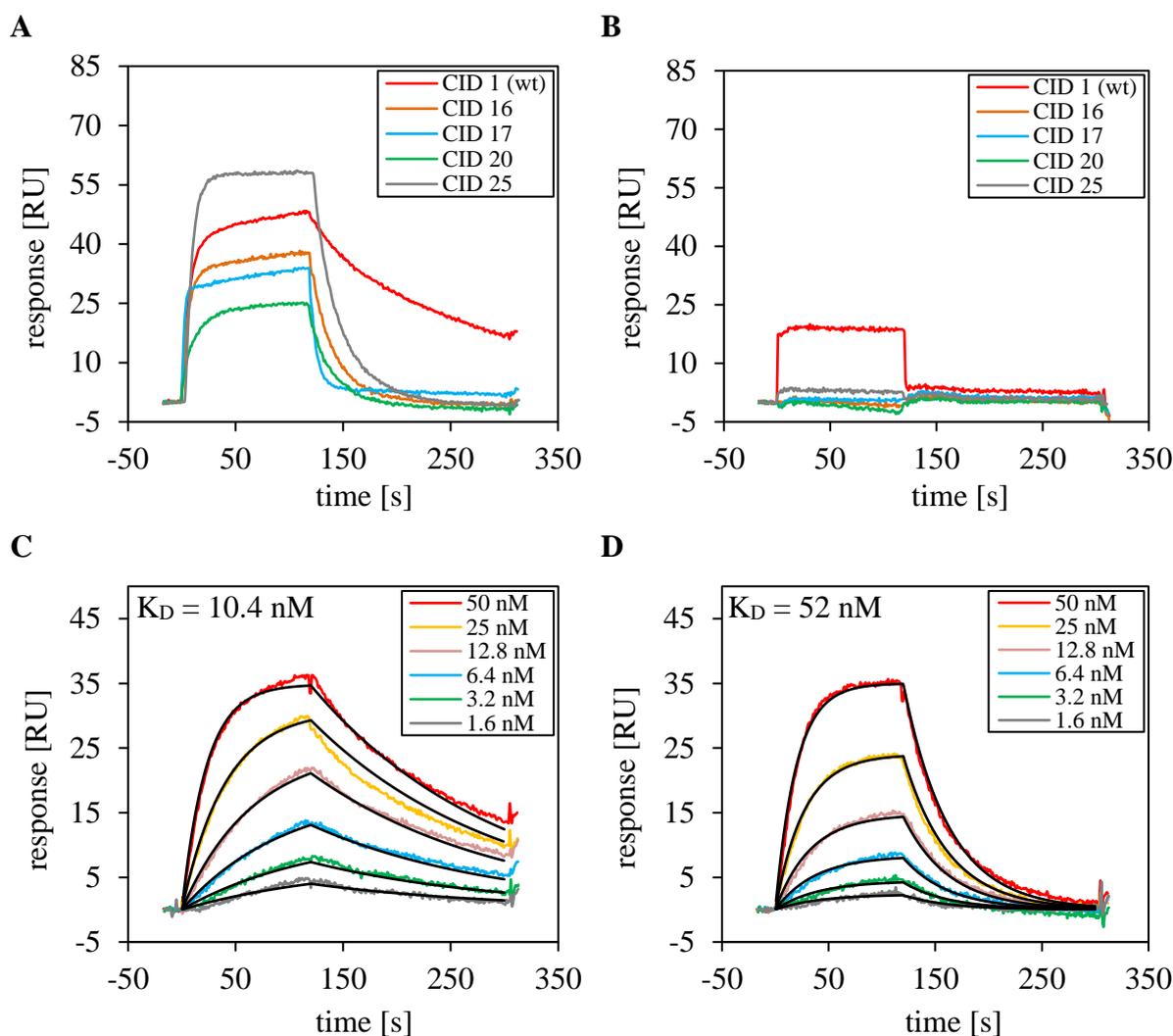
Figure 21: SPR sensorgrams of selected IgG-binding domain B variants.
(**A**) SPR sensorgrams of the novel domain B variants CID 16, CID 17, CID 20, and CID 25 are shown as example. Wild type domain B (CID 1) was used as comparison. hIgG1-Fc (~550 RU) was immobilized on the surface of a CM5 chip using NHS chemistry. The binding of variants [c = 100 nM] to hIgG1-Fc was determined at a flow rate of 30 µL/min. The association and dissociation steps were performed for 120 sec each. 1x PBST 0.05 % (v/v) pH 7.4 was used as running buffer. After each cycle, proteins on the surface were regenerated with 10 mM glycine pH 2.0 for 30 sec. (**B**) The binding of CID 16, CID 17, CID 20, and CID 25 [c = 2 µM] to immobilized Avelumab-Fab (~500 RU) is shown as example. Wild type domain B (CID 1) was used as comparison. Dilution series of (**C**) wild type domain B (CID 1) and (**D**) CID 25 versus immobilized hIgG1-Fc. The kinetic parameters k_{on} and k_{off} were determined using a Langmuir fitting model, assuming a 1:1 binding stoichiometry (black lines).

Based on the obtained SPR data, four variants (CID 16, CID 17, CID 20, CID 25) with the highest affinity for hIgG1-Fc compared to the others were selected for further characterization by concentration-dependent ELISA (2.2.4.9). For this, tagfree variants [c = 120 nM] were immobilized on the surface of a 384-well high binding microtiter plate. Binding of serially diluted biotinylated hIgG1-Fc was detected via Strep-Tactin-HRP (IBA). The obtained curve progressions for the individual variants were approximated using a hyperbolic function and resulted in affinities of $K_D = 10^{-7}$-10^{-8} M (Figure 22, Table 34). All variants showed reduced amplitudes of the plateau phases and a decrease in binding strength for hIgG1-Fc compared to wild type control (CID 1) (Figure 22, Table 34).
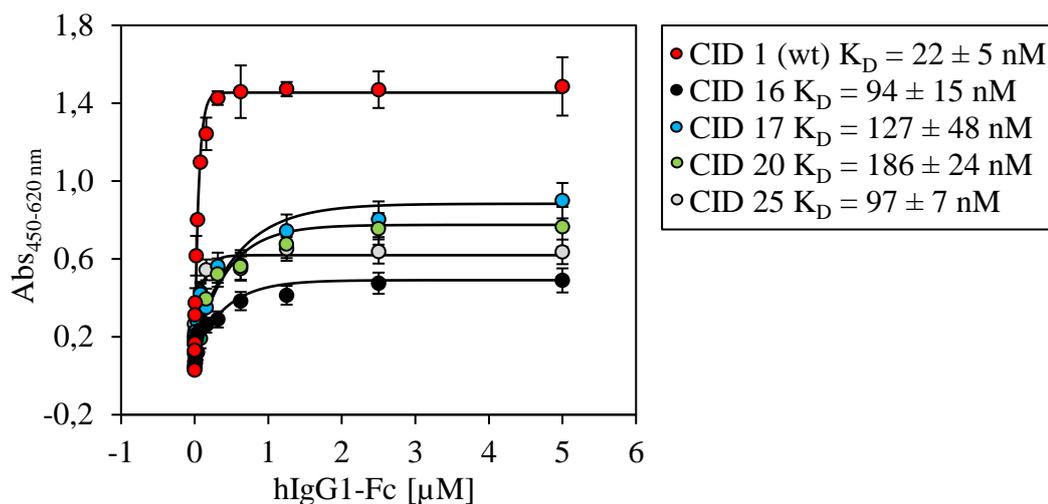
Figure 22: Binding of novel domain B variants and wild type control (CID 1) to hIgG1-Fc analyzed by concentration-dependent ELISA.
Concentration-dependent ELISA of the novel domain B variants CID 16, CID 17, CID 20, and CID 25. Wild type domain B (CID 1) was used as comparison (red dots). The tagfree variants [c = 120 nM] were immobilized on the surface of a 384-well high binding microtiter plate and incubated with a dilution series of biotinylated hIgG1-Fc [c = 0-5 µM]. Bound antibody fragments were detected via Strep-Tactin-HRP (IBA). Each series of measurements was performed in triplicate. Calculated standard deviations are shown as black error bars. A hyperbolic regression (solid black curve) was used to calculate the K_D values, assuming a 1:1 stoichiometry.

Table 34: Kinetic parameters of novel IgG-binding domain B variants determined by SPR and ELISA

CID	RF1	RF2	SPR K_D [nM]	SPR k_{on} [1/Ms]	SPR k_{off} [1/s]	ELISA K_D [nM]
1 (wt)	PNL	PS	10	7.6×10^5	5.7×10^{-3}	22 ± 5
7	GGLT	ASEKGA	793	1.1×10^5	7.9×10^{-2}	n.m.
8	GGLT	PVLRGKP	1,280	1.9×10^5	2.5×10^{-1}	n.m.
9	GGIT	ALRGHD	1,562	1.5×10^5	2.3×10^{-1}	n.m.
10	GGIT	AKRNPMT	507	2.2×10^5	1.1×10^{-1}	n.m.
11	GGLT	GTEGVH	357	1.7×10^5	6.1×10^{-2}	n.m.
12	GGLT	TAANPTA	815	2.1×10^5	1.7×10^{-1}	n.m.
14	GGLT	HTPTRG	6,423	1.9×10^4	1.3×10^{-1}	n.m.
16	GGLT	VRKLGP	190	4.7×10^5	2.2×10^{-2}	94 ± 15
17	NGLT	VRTNGA	220	6.1×10^5	2.3×10^{-2}	127 ± 48
20	GGLT	VRRVGS	272	2.2×10^5	6.0×10^{-2}	186 ± 24
23	GGIT	AKANPEA	1,180	3.8×10^4	4.4×10^{-2}	n.m.
24	GGLT	MNVGKH	783	4.2×10^5	4.1×10^{-2}	n.m.
25	GGIT	AKTNPGN	52	6.5×10^5	2.4×10^{-2}	97 ± 7
26	GGLT	NRLTKD	416	5.3×10^4	2.2×10^{-2}	n.m.

n.m. = not measured

A competition experiment measured via ELISA was performed (2.2.4.10) to investigate whether novel domain B variants compete with wild type domain B (CID 1) for Fc binding. For this, hIgG1-Fc [c = 40 nM] was immobilized on the surface of a 384-well high binding microtiter plate. The binding of Strep-tagged wild type domain B [c = 100 nM] in presence of serially diluted competing tagfree novel domain B variant [c = 0.01-3 µM] was determined. Bound wild type domain B was detected with Strep-Tactin-HRP (IBA). By increasing the concentration of tagfree domain B variant (competitor I), a significant decrease in binding signal of bound wild type domain B (ligand L) was observed (Figure 23 A). The inhibitory constants (K_i) of the novel domain B variants were determined inspired by the graphical method of Dixon (Dixon, 1953) and revealed values of $K_i = 10^{-7}-10^{-8}$ M (Figure 23 B, Table 35).

A **B**
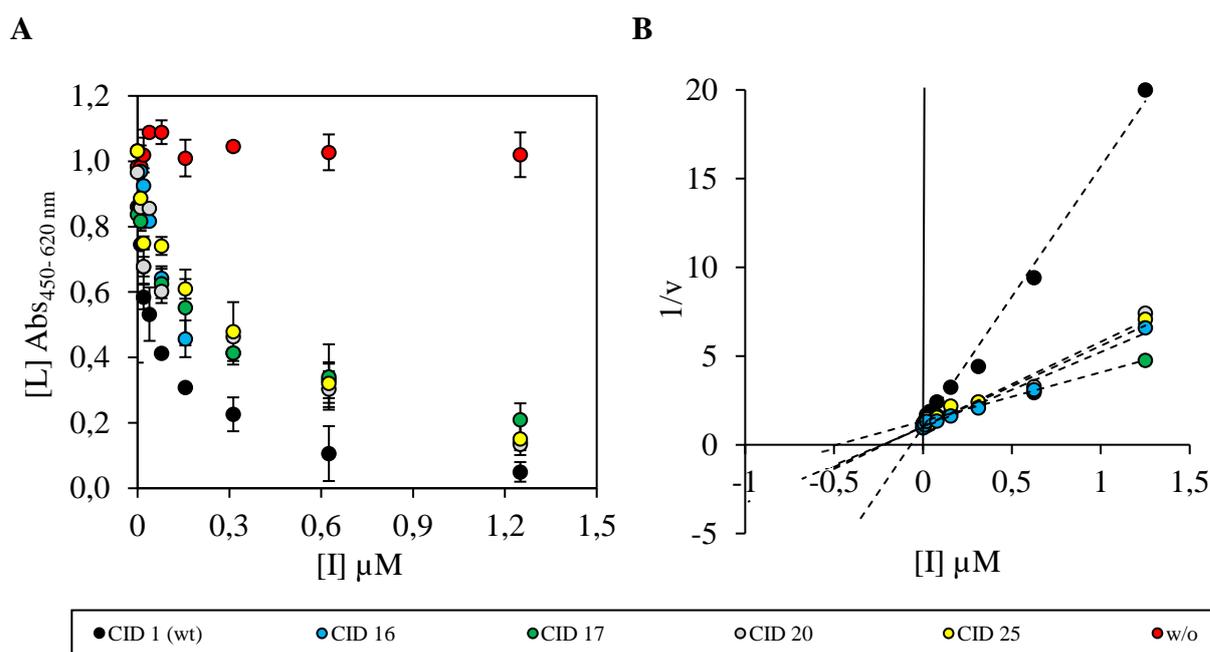
Figure 23: Competitive ELISA to determine the inhibitory constant (K_i) of four novel domain B variants. **(A)** Competitive ELISA with the novel domain B variants CID 16, CID 17, CID 20, and CID 25. Wild type domain B (CID 1) was used as comparison (black dots). hIgG1-Fc [c = 40 nM] was immobilized on the surface of a 384-well high binding microtiter plate and incubated with 100 nM Strep-tagged wild type domain B (ligand L) in presence of serially diluted competitive tagfree novel domain B variant (competitor I). Bound ligand was detected with Strep-Tactin-HRP (IBA). Binding signal of the ligand [L] is plotted versus inhibitor concentration [I]. Each series of measurements was performed in triplicate and calculated standard deviations are shown as black error bars. **(B)** Dixon plot used to calculate the inhibitory constant (K_i) of domain B variants. The reciprocal of binding rate (1/v) is plotted versus inhibitor concentration [I]. Graph lines are linearly fitted (black dashed lines) and the perpendicular of the fit to the x-axis is used as K_i value.

Table 35: K_i of novel domain B variants determined by competitive ELISA

CID	RF1	RF2	ELISA K_i [nM]
1 (wt)	PNL	PS	68 ± 12
16	GGLT	VRKLGP	228 ± 27
17	NGLT	VRTNGA	427 ± 61
20	GGLT	VRRVGS	249 ± 40
25	GGIT	AKTNPGN	215 ± 29

Structural characterization of selected domain B variants

Due to the shuffling of helices 1 and 3 and the insertion of randomized loops, it was possible that the structure of the novel domain B variants had changed compared to wild type domain B. To investigate the secondary structure of the novel domain B variants, CD measurements (2.2.5.3) were performed with CID 11, CID 16, CID 17, CID 20, and CID 25. The variants were selected based on the obtained data with regard to manufacturability (Table 33) and affinity for hIgG1-Fc (Table 34). Wild type domain B was used as comparison (CID 1). After buffer exchange (2.2.4.7), CID 17 showed precipitation and was excluded from the CD measurements. The measured CD spectra were baseline corrected by subtracting a buffer spectrum. The mean residue ellipticity was calculated using Equation 4 and plotted as a function of wavelength (Figure 24). Based on the measured CD spectra, the proportion of α-helical structures was calculated using the reference by Yang included in the device software (Yang *et al*., 1986). As a result, CID 20 and CID 25 showed a CD spectrum and a calculated α-helical proportion (≥ 70 %) comparable to wild type domain B (Figure 24, Table 36), although the obtained spectral amplitudes were slightly decreased compared to CD spectra of different Protein A domains found in literature (Myrhammar *et al.*, 2020). In contrast, a significant increase in molar ellipticity and a reduction of α-helical structures was observed for CID 11 and CID 16 compared to wild type domain B (Figure 24, Table 36).
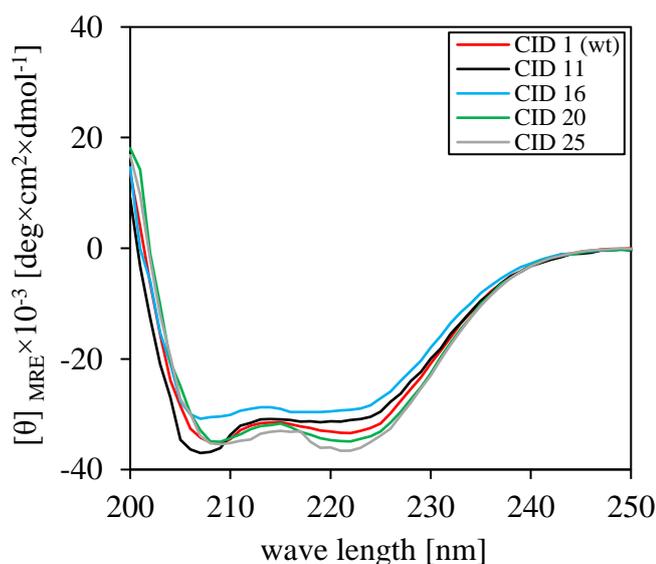
Figure 24: CD spectra of four novel domain B variants and wild type control CID 1.
CD analysis of novel domain B variants CID 11, CID 16, CID 20, and CID 25 purified by IMAC and SEC. The CD spectrum of wild type domain B (CID 1) is shown as comparison. Variants [c = 0.2 mg/mL] were transferred into a 1 mm quartz cuvette. CD measurements were performed at 20 °C in the far-UV spectral region (200-250 nm) at a speed of 50 nm/min and a bandwidth of 1 nm. 16 spectra per sample were measured and baseline corrected by subtracting a buffer spectrum. Mean residue ellipticity is plotted as a function of wavelength.

Table 36: Proportion of calculated α-helical structures of selected novel domain B variants

CID	RF1	RF2	Helix [%]
1 (wt)	PNL	PS	72.1
11	GGLT	GTEGVH	63.7
16	GGLT	VRKLGP	61.2
17	NGLT	VRTNGA	n.m.
20	GGLT	VRRVGS	70.1
25	GGIT	AKTNPGN	72.0

n.m. = not measurable due to precipitation

Next, temperature-induced unfolding and refolding of CID 11, CID 16, CID 17, CID 20, and CID 25 was measured to investigate the thermal stability of the proteins (2.2.5.3). Wild type domain B (CID 1) was used as comparison. The calculated residual ellipticity (Equation 4) was plotted as a function of temperature and the turning point of the curve represented the melting temperature T_m of the domain B variants (Figure 25 A). A significant decrease in T_m was detected for the novel domain B variants (T_m ~31 °C) compared to wild type domain B (T_m = 72.1 °C) (Figure 25 A, Table 37). Single spectra were recorded at 20 °C before unfolding and after refolding (90 °C) to further investigate the thermal transition of the variants. As a result, the novel domain B variants showed an irreversible refolding after thermal transition (Figure 25 B).
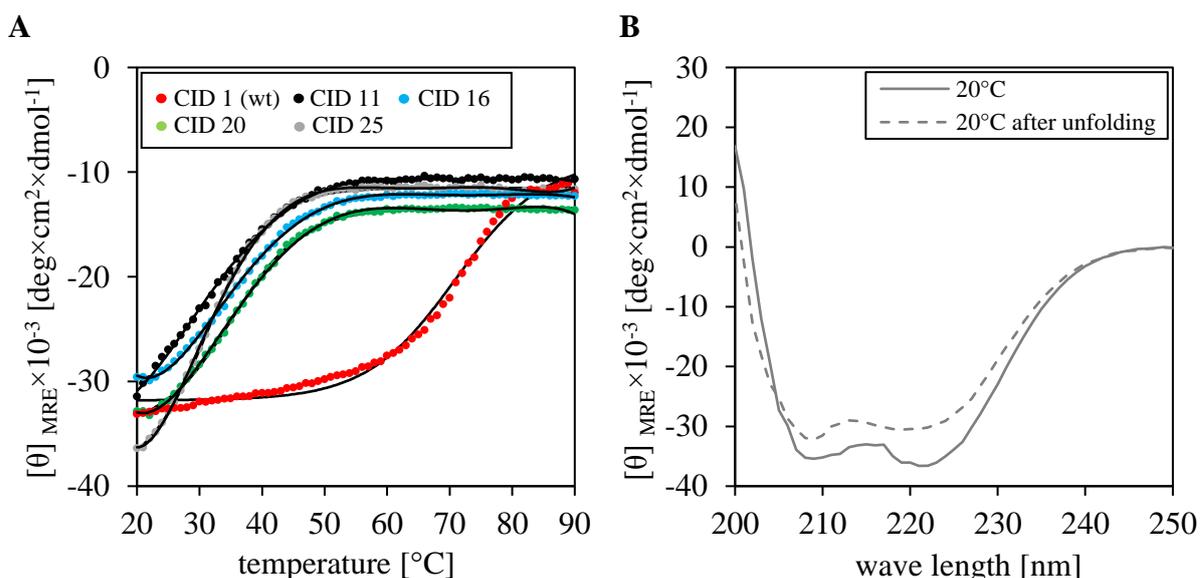
Figure 25: Thermal-induced structural transition of four novel domain B variants measured by CD.
(**A**) Temperature-induced structural transition of novel domain B variants CID 11, CID 16, CID 20, CID 25. Wild type domain B (CID 1) was used as comparison. Variants [c = 0.2 mg/mL] were transferred into a 1 mm quartz cuvette and CD measurement was performed at a wavelength of 222 nm. The initial temperature of 20 °C was increased up to 90 °C with a slope of 1 °C/min. The residual ellipticity is plotted as a function of temperature. The melting temperature (T_m) was calculated using a sigmoidal 4 parameter fit. (**B**) CD spectra of CID 25 recorded at 20 °C before unfolding and after refolding (90 °C) in the far-UV spectral region (200-250 nm) at a speed of 50 nm/min and a bandwidth of 1 nm. 16 spectra per sample were measured and baseline corrected by subtracting a buffer spectrum. The mean residue ellipticity is plotted as a function of wavelength.

Table 37: Calculated T_m values of novel domain B variants

CID	RF1	RF2	T_m [°C]
1 (wt)	PNL	PS	72.1 ± 0.6
11	GGLT	GTEGVH	28.1 ± 0.3
16	GGLT	VRKLGP	32.4 ± 0.1
17	NGLT	VRTNGA	n.m.
20	GGLT	VRRVGS	32.8 ± 0.2
25	GGIT	AKTNPGN	31.2 ± 0.1

n.m. = not measurable due to precipitation

Table 38: Summary of produced and characterized novel IgG-binding domain B variants

CID	RF1	RF2	Yield [mg/L]	K_D [nM]*	Helix [%]+	T_m [°C]+
1 (wt)	PNL	PS	40.1	10	72.1	72.1 ± 0.6
7	GGLT	ASEKGA	2.5	793	n.m.	n.m.
8	GGLT	PVLRGKP	6.0	1,280	n.m.	n.m.
9	GGIT	ALRGHD	3.6	1,562	n.m.	n.m.
10	GGIT	AKRNPMT	2.3	507	n.m.	n.m.
11	GGLT	GTEGVH	21.3	357	63.7	28.1 ± 0.3
12	GGLT	TAANPTA	2.3	815	n.m.	n.m.
14	GGLT	HTPTRG	14.4	6,423	n.m.	n.m.
16	GGLT	VRKLGP	24.4	190	71.2	32.4 ± 0.1
17	NGLT	VRTNGA	22.0	220	n.m.#	n.m.#
20	GGLT	VRRVGS	23.4	272	70.1	32.8 ± 0.2
23	GGIT	AKANPEA	3.7	1,180	n.m.	n.m.
24	GGLT	MNVGKH	22.6	783	n.m.	n.m.
25	GGIT	AKTNPGN	24.6	52	72.0	31.2 ± 0.1
26	GGLT	NRLTKD	10.8	416	n.m.	n.m.

n.m. = not selected for measurement; n.m.# = not measurable due to precipitation; * = determined by SPR; + = determined by CD

3.5 Detailed analysis of the novel domain B variant CID 25

The novel domain B variant CID 25 was selected for a detailed characterization since the protein showed the best manufacturability, the highest affinity for hIgG1-Fc and a domain B-like proportion of α-helical structures compared to the other variants (Table 38). First, thermal-induced structural transitions and folding mechanisms of CID 25 were investigated by DSC (2.2.5.4). Next, NMR experiments (2.2.5.5) were performed to further examine the secondary structure of CID 25 and to localize Fc binding sites on the surface of the protein.

3.5.1 Investigations on structural transitions and folding mechanisms

To investigate structural transitions and folding mechanisms of CID 25 and wild type domain B (CID 1), thermal-induced unfolding and refolding were analyzed by DSC (2.2.5.4). Thermodynamic parameters were determined in a temperature range of 10-60 °C (CID 25) or 10-90 °C (CID 1) with a heating rate of 60 °C/h. The thermograms were evaluated using the implemented Origin software from MicroCal. As a result, CID 25 showed a calculated T_m of 32.1 °C, whereas a T_m of 74.4 °C was observed for wild type domain B (Figure 26, Table 39). Thus, the T_m values obtained from CD analyses (3.4.2, Table 37) could be confirmed by DSC. However, CID 25 showed a significant decrease in heat capacity and a partially incomplete refolding compared to wild type domain B. Moreover, the calculated free enthalpy ΔH_{cal} of CID 25 unfolding was reduced by 25.7 % compared to wild type control. Furthermore, different enthalpies ΔH_{cal} and ΔH_{vH} were observed for CID 25 unfolding and refolding (Table 39). Interestingly, CID 25 showed a vertically shifted refolding curve, which was also detectable for wild type domain B (Figure 26).
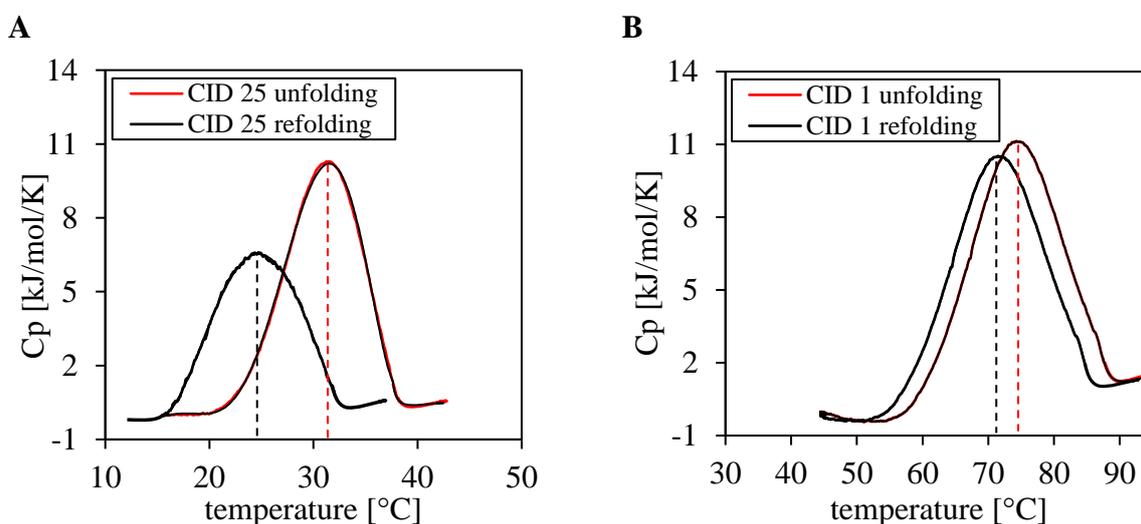
Figure 26: Thermal-induced structural transition of CID 25 and wild type control CID 1 measured by DSC. Thermal-induced unfolding and refolding of (**A**) CID 25 and (**B**) wild type domain B (CID 1) as comparison. Both variants were diluted in 1x PBS pH 7.4 to a final concentration of (**A**) 1.3 µM and (**B**) 1.8 µM. DSC measurements were performed in a temperature range of 10-60 °C (CID 25) or 10-90 °C (CID 1) with a heating rate of 60 °C/h. The thermograms are shown as red (unfolding) and black (refolding) colored lines. The calculated T_m values are marked as dashed lines.

Table 39: Thermodynamic parameters of unfolding and refolding of CID 25 and CID 1 (wt) determined by DSC

	CID 25		CID 1	
Parameter	**unfolding**	**refolding**	**unfolding**	**refolding**
T_{onset} [°C]	20.8 ± 0.6	18.2 ± 1.0	57.7 ± 1.1	52.7 ± 2.1
T_m [°C]	32.1 ± 0.2	26.2 ± 0.5	74.4 ± 0.4	70.5 ± 0.2
ΔH_{cal} [kJ/mol]	147 ± 0.4	93 ± 1.3	198 ± 0.3	196 ± 0.1
ΔH_{vH} [kJ/mol]	267 ± 0.5	150 ± 0.8	181 ± 0.8	186 ± 0.3

3.5.2 Investigations on secondary structure and Fc binding sites

In this work, NMR spectroscopy was used to study the structure of CID 25 and to localize Fc binding sites on the surface of the protein (2.2.5.5). About 80 % of the amid signals (proline excluded) in the $^1H^{15}N$-HSQC could be assigned by HNCA(CO)CB/HNCACB experiments. This is somewhat less than typical assignment levels of 95 % and above, caused by the limited spectral dispersion, reduced thermodynamic stability, and hence increased amide exchange in the artificially designed protein. Resonances from side chains in the $^1H^{13}C$-HSCQs have been assigned by ^{15}N-TOCSY-HSQC and HCCH-TOCSY experiments. Here, 96 % of the resonances could be assigned. This is a typical assignment level for well-behaving proteins and confirms the suspected problems by amide exchange for the NH assignment (data not shown).

Structural characterization of CID 25
Based on the chemical shifts of C_α and C_β atoms, backbone φ/ψ dihedral angles were determined using the program TALOS+. The combination of the phi (φ) and psi (ψ) angles of CID 25 illustrated in the Ramachandran plot revealed a mainly right-handed α-helical secondary structure (Figure 27). In total, 85.7 % of the amino acids determined for the expected α-helices were found in the allowed or in the partially allowed region. The sections A9-Q22 (helix 3), E27-N31, I34-K38 (helix 2) and E50-H60 (helix 1) were determined as α-helical structures of CID 25. Hence, an α-helical proportion of 55.5 % was calculated, which was less compared to the obtained CD data (72 %, Table 36). However, all three theoretically expected helices of CID 25 inspired by wild type domain B were experimentally verified but slight differences can be observed (Figure 28).
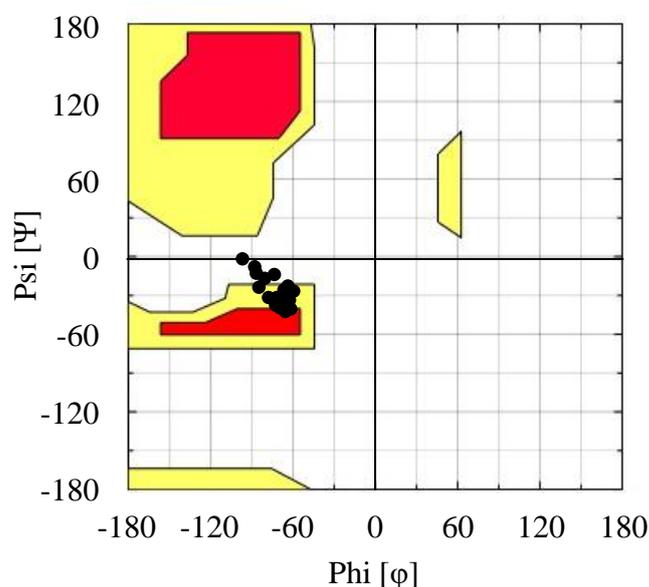
Figure 27: Ramachandran plot of determined backbone φ/ψ dihedral angles of CID 25.
The phi (φ) and psi (ψ) values of the amino acids were obtained by TALOS+ and typical values of α-helices visualized in a Ramachandran plot (black dots). The red areas of the plot represent allowed pairs of angles, whereas partially allowed regions are marked in yellow (Ramachandran *et al.*, 1963).

Since experimental 3D structure determination of CID 25 by NMR was not completed at the end of this work, the AI-based network AlphaFold (AG Schröder, Medical Faculty, University Leipzig) was used to predict the tertiary structure of the variant. The structure of CID 25 is characterized by three right-handed α-helices arranged in a compact bundle (Figure 28 A). Hence, predictions by TALOS+ and AlphaFold revealed comparable α-helical core stretches. Since NMR structures of wild type domain B with high resolution and complete α-helical structures were not available, the NMR structure of domain Z (PDB: 1Q2N), an engineered B domain with only two mutations, was used for the structural alignment. As a result, CID 25 showed a slightly different helix orientation compared to domain Z, particularly applicable for helix 2. Further structural differences were observed in comparison to this engineered domain B. Compared to theory, helix 3 was predicted to be N-terminal shortened by the residues S6 and Q7 (Figure 28 B). However, the helix 3/2 segment looked structurally as expected and these findings were in good accordance with the data obtained from NMR. For the short helix segment formed by the residues G32 and F33 (helix 2), a discrepancy was observed between AlphaFold and TALOS+. An interruption of the helix was predicted by TALOS+, whereas AlphaFold prediction revealed an α-helical structure. The structural prediction of the helix 2/1 segment by AlphaFold revealed further differences compared to theory and TALOS+. An extension of helix 2 by the theoretically loop-forming residues A41, K42, and T43 was predicted by AlphaFold, whereas TALOS+ prediction revealed a shortening of the helices 2 and 1 by the residues D39, D40 (helix 2), N48, and K49 (helix 1) (Figure 28 B).
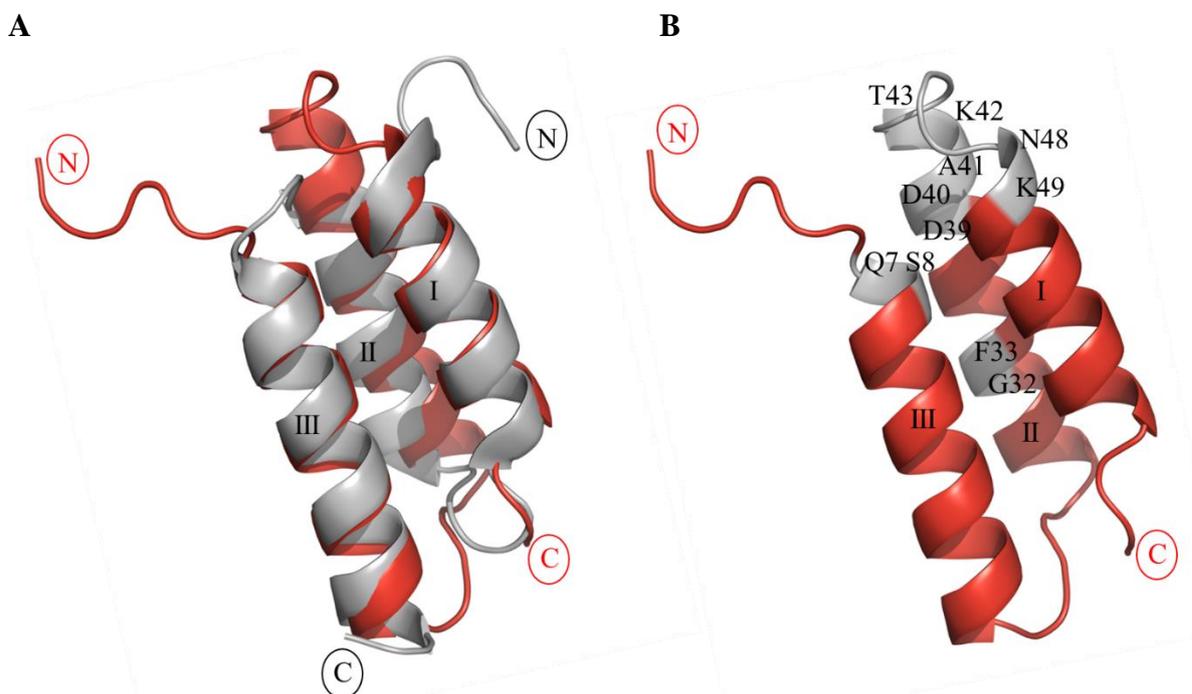
Figure 28: Structure prediction of the novel domain B variant CID 25 using AlphaFold.
(**A**) Structural alignment of CID 25 (red) predicted by AlphaFold and domain Z (gray, PDB: 1Q2N). The numbering of the helices is inspired by wild type domain B and is as follows: helix 1 (I), helix 2 (II), and helix 3 (III). The termini are named C (C-terminus) and N (N-terminus). (**B**) AlphaFold and TALOS+ predicted α-helical sections are colored red. Structural regions for which the prediction by TALOS+ and AlphaFold disagreed, are marked in gray and amino acids are shown as single letter codes. The figure was created as cartoon representation using the software PyMOL.

NMR titration for the localization of Fc binding sites

An NMR titration experiment was performed to identify amino acids of CID 25 involved in Fc binding (2.2.5.5). In four steps, unlabeled hIgG1-Fc [c = 0, 30, 70, 130 µM] was titrated to $^{13}C^{15}N$ CID 25 [c = 91 µM]. $^1H^{15}N$-TROSY-HSQC spectra showed changes in intensity but not in chemical shift, indicative for a slow exchange regime and a long lifetime of the formed complex. The binding of CID 25 to hIgG1-Fc induced a line-broadening and a signal loss of almost all amide resonances, which completely disappeared at the end of the titration (data not shown). In the end, almost all signals sense the binding to hIgG1-Fc either directly, in which the signal disappears to zero at one position in the spectrum and reappears at another position (if detectable), or indirectly by increased relaxation in the larger complex. Based on theoretical titration curves calculated by a basic Boltzmann sigmoid function, a 1:1 binding stoichiometry of the interaction between CID 25 and hIgG1-Fc could be verified (Figure 29).
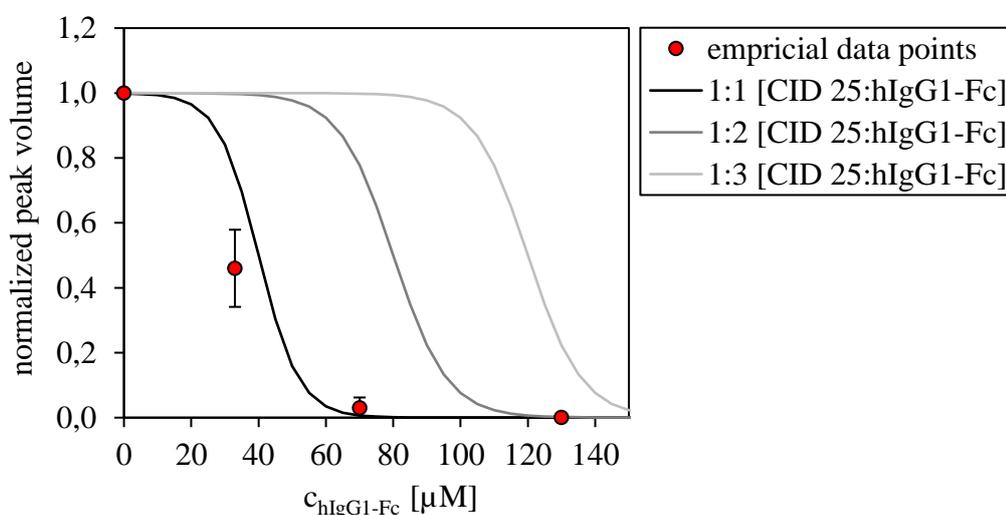
Figure 29: Normalized NMR peak volumes of the backbone amide resonances of $^{13}C^{15}N$ CID 25 in complex with hIgG1-Fc compared to theoretical NMR titration curves.
Integrated peak areas obtained from the $^1H^{15}N$-TROSY-HSQC spectra of the 1:0.35 complex (CID 25:hIgG1-Fc) were normalized to the spectrum of free $^{13}C^{15}N$ CID 25. The obtained values of amino acids that showed a significant decrease in NMR peak volume upon Fc binding (I25, T26, N31, I34, Q35, K38, Q51, N53, F55, Y56, and L59) were averaged and are shown as red dots. Standard deviations are represented as black error bars. Theoretical NMR titration curves assuming a 1:1 (black curve), a 1:2 (gray curve), or a 1:3 (light gray curve) binding stoichiometry of the interaction between CID 25 [c = 85 µM] and hIgG1-Fc [c = 0-150 µM] were calculated using a basic Boltzmann sigmoid function.

In order to distinguish direct and indirect effects, signal strengths were compared under sub-stoichiometric conditions. The obtained volumes of the integrated peak areas were normalized to the values of free $^{13}C^{15}N$ CID 25 and used to localize amino acids involved in Fc binding. The NMR signals of the amino acids I25, I34, Q35, K38, Q51, F55, Y56, and L59 changed significantly upon binding of hIgG1-Fc due to a considerable decrease in peak volume, making an involvement in the binding highly probable (plot signal ≤ 0.5, Figure 30). A loss of peak volume between 35 % and 50 % was detected for the amino acids T26, N31, and N53, which could function as further binding sites. Interestingly, the amino acids N3, D40, and K49 showed no changes in the backbone NMR signal during titration with hIgG1-Fc.
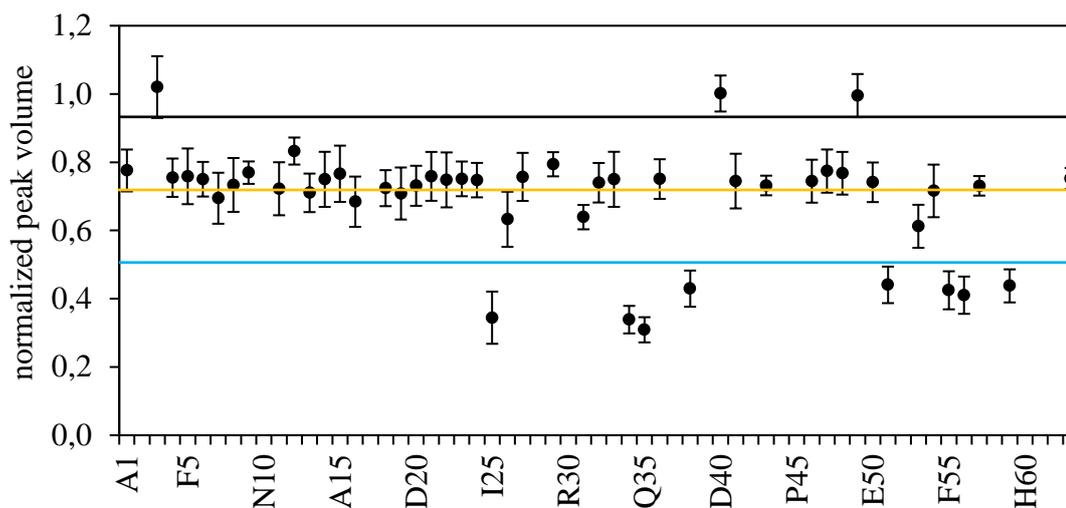
Figure 30: Normalized NMR peak volumes of the backbone amide resonances of $^{13}C^{15}N$ CID 25 in complex with hIgG1-Fc.
Integrated peak areas obtained from the $^{1}H^{15}N$-TROSY-HSQC spectra of the 1:0.35 complex (CID 25:hIgG1-Fc) were normalized to the spectrum of free $^{13}C^{15}N$ CID 25. The obtained values of the individual amino acids of CID 25 are shown as black dots. The experimental errors were calculated for each amino acid and are shown as black error bars. Limits are set for ≥ 0.9 (black line), ≤ 0.75 (orange line), and ≤ 0.5 (blue line).

A   **B**
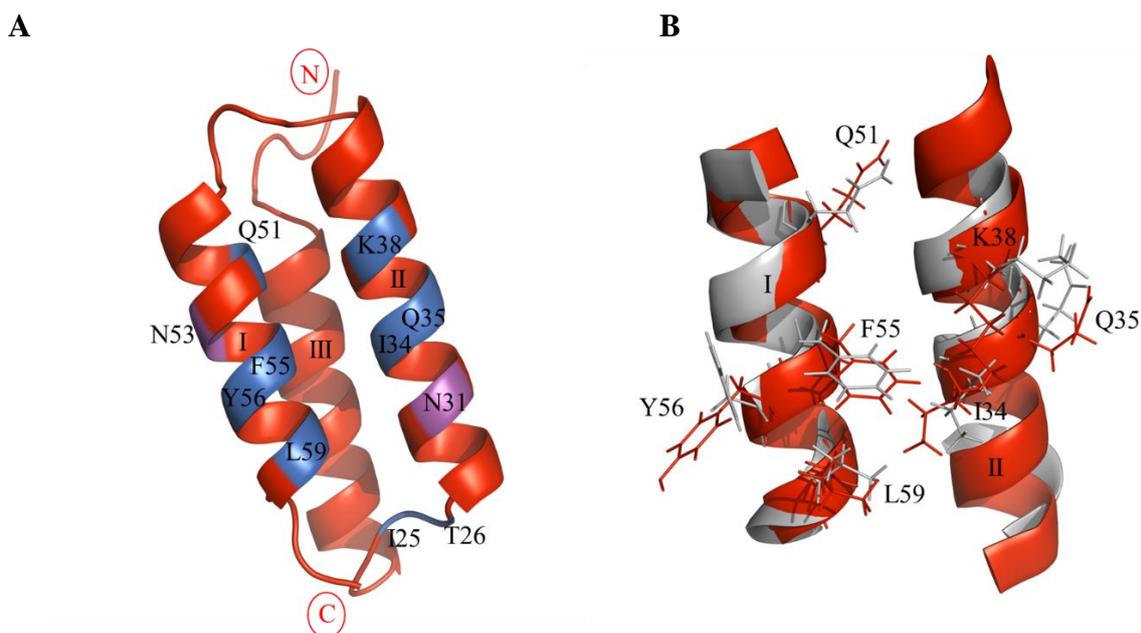
Figure 31: Amino acid residues of $^{13}C^{15}N$ CID 25 probably involved in Fc binding.
(**A**) The 3D structure of CID 25 was predicted by AlphaFold. All structural elements, including the three helices (I-III), the termini (N and C) and the alternative loops are colored red. Blue colored residues (I25, I34, Q35, K38, Q51, F55, Y56, and L59) showed a significant decrease in NMR peak volume (≤ 0.5) upon binding of hIgG1-Fc. Purple colored residues (T26, N31, and N53) lost between 35 % and 50 % (≤ 0.75) of their NMR peak volume. Amino acids probably involved in Fc binding are shown as single letter codes. (**B**) Structural alignment of CID 25 (red) predicted by AlphaFold and domain Z (gray, PDB: 1Q2N). The side chains of amino acids probably involved in Fc binding are shown as sticks. The figure was created as cartoon representation using the software PyMOL.

Residues that showed a significant change in NMR signal and a decrease in peak volume upon binding to hIgG1-Fc are located in helix 1 (Q51, N53, F55, Y56, and L59), helix 2 (N31, I34, Q35, and K38) and in the alternative loop 1 (I25 and T26) (Figure 31 A). Comparing these residues (N31, I34, Q35, K38, Q51, N53, F55, Y56, and L59) with the Fc binding sites of wild type domain B (Deis *et al*., 2015), a homology of 81.8 % can be determined. These findings are in accordance with data obtained from the competition experiment measured by ELISA (Figure 23 A). The area of residues that changed significantly in NMR signal upon hIgG1-Fc binding was calculated to be 616 Å, which is comparable to the binding area of wild type domain B (~650 Å).

3.6 CID 25 as ligand for the affinity chromatographic purification of hIgG1

In previous experiments, the novel domain B variant CID 25 showed a property profile that corresponded very well to the requirements for the use of the protein as ligand in affinity chromatography. In addition to an excellent manufacturability (3.4.1), CID 25 showed specific binding to hIgG1-Fc with an affinity in the nM range ($K_D = 52$ nM, 3.4.2). However, except for one experiment (Figure 22), CID 25 has been used exclusively in solution so far (Figure 21). First, the binding behavior of immobilized CID 25 to hIgG1-Fc was analyzed with an SPR experiment by coupling the variant to a stationary phase. Second, CID 25 was coupled to a column matrix and the characteristics of the functionalized chromatography column were investigated.

3.6.1 Coupling of CID 25 and purification of hIgG1

A basic requirement for the use of CID 25 as ligand for affinity chromatography is the preservation of functionality when the protein is coupled on a solid carrier matrix. First, the binding behavior of coupled CID 25 was analyzed by SPR. For this, approximately 180 RU of the variant was coupled on the surface of a CM5 chip using NHS chemistry (2.2.5.1). Wild type domain B was coupled as comparison (175 RU) and a set of curves with different concentrations of hIgG1-Fc was measured. As a result, CID 25 showed a 4.6-fold decrease in binding strength compared to the affinity determined with immobilized hIgG1-Fc ($K_D = 52$ nM) (Table 40). A lower affinity for hIgG1-Fc was also observed for wild type domain B (CID 1) compared to the reverse binding assay (Figure 32, Figure 21). Interestingly, both variants showed a significant increase in k_{off}, whereas k_{on} was rather less affected (Table 40, Table 34).

Table 40: Kinetic parameters of CID 25 and wild type domain B (CID 1) determined by SPR

CID	K_D [nM]	k_{on} [1/Ms]	k_{off} [1/s]
1 (wt)	31.4	5.9×10^5	1.1×10^{-2}
25	237	4.1×10^5	6.0×10^{-2}

A

B
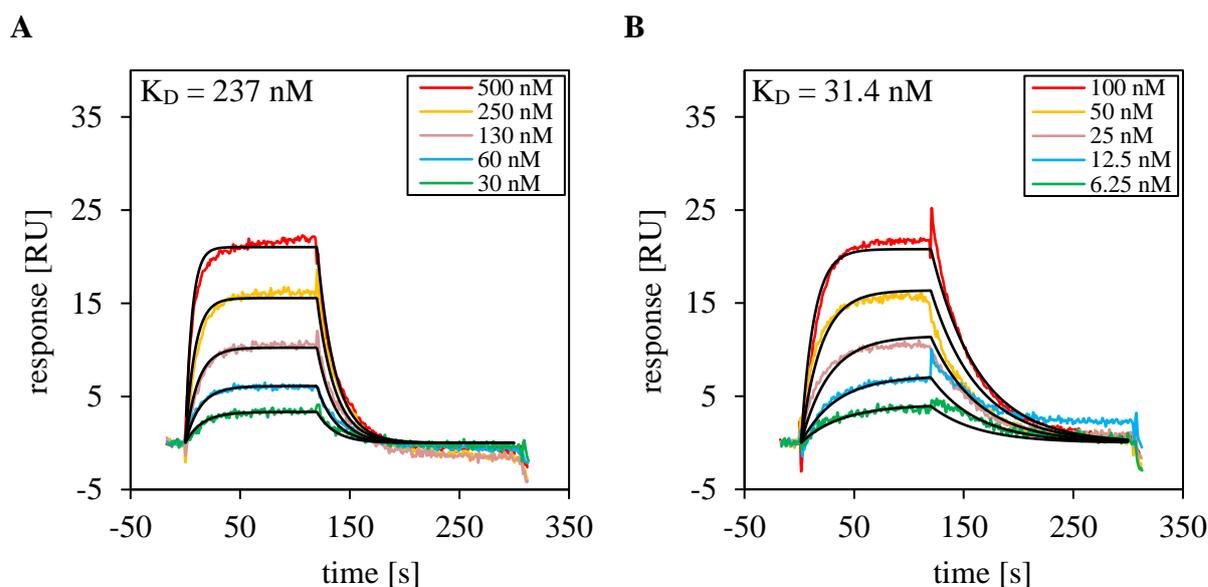
Figure 32: SPR sensorgrams of coupled CID 25 and wild type domain B (CID 1).
Approximately 180 RU of (**A**) the novel domain B variant CID 25 and (**B**) wild type domain B (CID 1) were coupled on the surface of a CM5 chip using NHS chemistry. Samples with different concentrations of hIgG1-Fc in 1x PBST 0.05 % (v/v) pH 7.4 were measured (colored lines) at a flow rate of 30 µL/min. The association and dissociation steps were performed for 120 sec each followed by regeneration of the coupled variants using 10 mM glycine pH 2.0. The kinetic parameters k_{on} and k_{off} were determined using a Langmuir fitting model integrated in the evaluation software Sierra Analyzer, assuming a 1:1 binding stoichiometry (black lines).

Since the obtained SPR data showed that CID 25 could be coupled on the surface of a stationary phase while retaining the IgG-binding properties, the variant was subsequently coupled to a column matrix (2.2.4.10). 5.1 mg CID 25 and 5.5 mg wild type domain B (CID 1), used as comparison, were coupled on the surface of 1 mL HiTrap NHS-Activated HP columns (Cytiva). The column experiments were performed at 10 °C at a flow rate of 1 mL/min using an ÄKTAxpress chromatography system (GE Healthcare). The functionalized columns were equilibrated with 10 CV 1x PBS pH 7.4 and 2 mg hIgG1 diluted in a volume of 6.5 mL 1x PBS pH 7.4 [c = 0.3 mg/mL] was loaded on each column (Figure 33, I). During the loading and washing steps, no hIgG1 was detected in the flowthrough fractions determined by measurement of the absorption at 280 nm (Figure 33, II) and by SDS-PAGE (data not shown). hIgG1 was eluted within a single peak by reducing the pH from 7.4 to 3.0. (Figure 33, III). The evaluation of the peak geometry according to IUPAC specifications revealed a tailing coefficient of T = 2.01 for both CID 25 and wild type domain B (Nic *et al*., 2006). After elution of hIgG1, the columns were re-equilibrated with 10 CV 1x PBS pH 7.4 (Figure 33, IV) and used again for the purification of antibodies. A total of ten cycles of hIgG1 loading and elution were performed with 2 mg hIgG1 each to investigate the column lifetimes. After each elution cycle, the columns were regenerated using 10 CV 100 mM citrate pH 2.0. As a result, a decrease in column performance of ~10 % was observed for CID 25, whereas no reduction was observed for the wild type domain B column (data not shown).
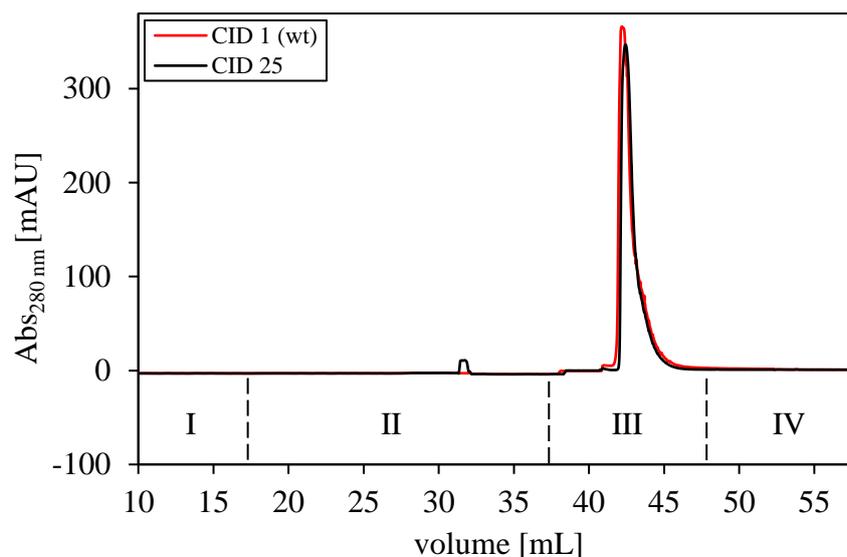
Figure 33: CID 25 as ligand for the affinity chromatographic purification of hIgG1.
CID 25 and wild type domain B (CID 1) were coupled on the surface of 1 mL HiTrap NHS-Activated HP columns (Cytiva) via NHS chemistry. CID 1 was used as comparison. The column experiments were performed at 10 °C using an ÄKTAxpress chromatography system (GE Healthcare) at a flow rate of 1 mL/min. After equilibration, 2 mg hIgG1 diluted in a volume of 6.5 mL 1x PBS pH 7.4 [c = 0.3 mg/mL] was loaded on each column (I). The columns were washed with 20 CV 1x PBS pH 7.4 (II), and hIgG1 was eluted by reducing the pH from 7.4 to 3.0 (III). The coupled proteins were regeneration with 10 CV 100 mM citrate pH 2.0 and re-equilibration with 15 CV 1x PBS pH 7.4 (IV) allowed to use the columns again for the purification of hIgG1.

3.6.2 Determination of dynamic binding capacity

The dynamic binding capacity of the CID 25 column was determined (2.2.4.10) to evaluate the coupling efficiency of functionally immobilized variant. For this, hIgG1 [c = 1.5 mg/mL] in 1x PBS pH 7.4 was passed over the column at a flow rate of 0.2 mL/min with a contact time of 5 min until saturation was reached (Figure 34 A).

A

B
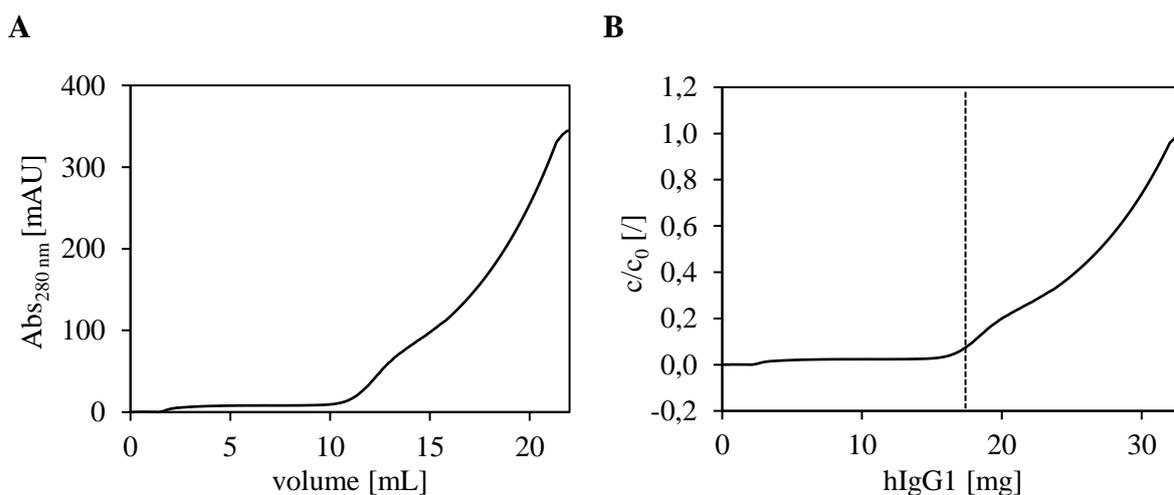
Figure 34: Determination of the dynamic binding capacity of the CID 25 column.
(**A**) CID 25 was coupled on the surface of a 1 mL HiTrap NHS-Activated HP column (Cytiva) using NHS chemistry. hIgG1 [c = 1.5 mg/mL] in 1x PBS pH 7.4 was passed over the column at a flow rate of 0.2 mL/min with a contact time of 5 min until saturation was reached. (**B**) Calculation of the dynamic binding capacity of the CID 25 column using the dynamic breakthrough method (Horstmann *et al.*, 1986) (Hage & Cazes, 2005*)*. The determined dynamic capacity for 10 % of the breakthrough is 18.2 mg/mL (vertical dashed line) compared to the theoretical value of 100 mg/mL.

The dynamic binding capacity of the CID 25 column was calculated based on the total volume that was needed until a significant amount (10 % of the maximum signal) of hIgG1 was detected in the flowthrough at an absorbance of 280 nm. The concentration of the saturation curve was normalized to the maximum concentration of the binding partner c/c_0 (Horstmann *et al.*, 1986) (Hage & Cazes, 2005). Based on the 10 % breakthrough (DBC10), the dynamic binding capacity of the CID 25 column was determined with 18.2 mg/mL (Figure 34 B). This is about 35 % less than the DBC10 determined for the column matrix with coupled wild type domain B (28.1 mg/mL).

3.6.3 Determination of elution pH

A pH gradient elution of hIgG1 was performed to determine the elution pH of CID 25 and compare it to that of wild type domain B (CID 1) (2.2.4.10). For this, 2 mg hIgG1 diluted in a volume of 6.5 mL 1x PBS pH 7.4 [c = 0.3 mg/mL] was loaded on each column. A gradient elution of hIgG1 ranging from pH 6.5 to pH 2.0 was performed within a volume of 20 CV using 0.1 mM citrate pH 2.0. The elution fractions were pooled and the pH was determined (2.2.4.10). As a result, a pH of 4.8 was sufficient to elute hIgG1 from the CID 25 column, which was significantly milder compared to wild type domain B (pH 3.7) (Figure 35). A second elution peak was observed for the CID 25 column. SDS-PAGE and SE-HPLC analysis of this elution fractions revealed slight impurities (~7 %) in addition to the target protein (data not shown).
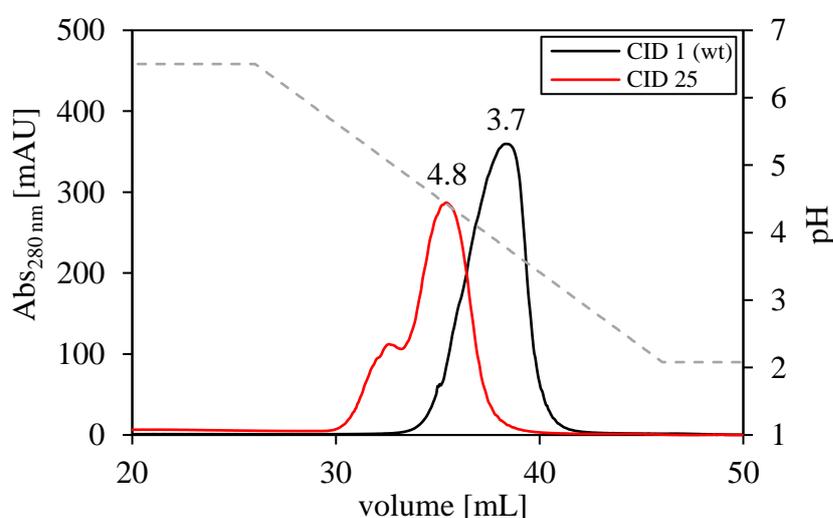
Figure 35: Gradient elution of hIgG1 purified by affinity chromatography using CID 25 as novel ligand. CID 25 and wild type domain B (CID 1) were coupled on the surface of 1 mL HiTrap NHS-Activated HP columns (Cytiva) via NHS chemistry. The column experiments were performed at 10 °C using an ÄKTAxpress chromatography system (GE Healthcare) at a flow rate of 1 mL/min. To determine the elution pH, 2 mg hIgG1 in 6.5 ml 1x PBS pH 7.4 [c = 0.3 mg/mL] was loaded on each column. The columns were washed with 5 CV 1x PBS pH 7.4 and 10 CV 0.1 M citrate, 0.15 M NaCl pH 6.5. A gradient elution of hIgG1 ranging from pH 6.5 to pH 2.0 was performed within a volume of 20 CV using 0.1 mM citrate pH 2.0. The theoretical pH gradient is shown as grey dashed line.

3.6.4 Purification of hIgG1 from different cell lysates

E.coli and Expi293 cells were prepared to demonstrate the feasibility of isolating hIgG1 from lysates of typical production strains using the CID 25 column (2.2.4.10). An optical density of $OD_{600} = 42$ was determined for the *E. coli* lysate and approximately 25×10^6 cells were used to produce the Expi293 lysate. 1.95 mg hIgG1 [c = 0.2 mg/mL] was added to each lysate and the samples were loaded on the equilibrated CID 25 column. Non-specifically bound proteins were removed by a wash step with 20 CV 1x PBS pH 7.4. Finally, bound hIgG1 was eluted by a pH shift (pH 3.0) using 10 CV AIC buffer B (100 %) (Figure 36). As a result, hIgG1 was successfully isolated from complex mixtures of endogenous proteins by using the CID 25 column.
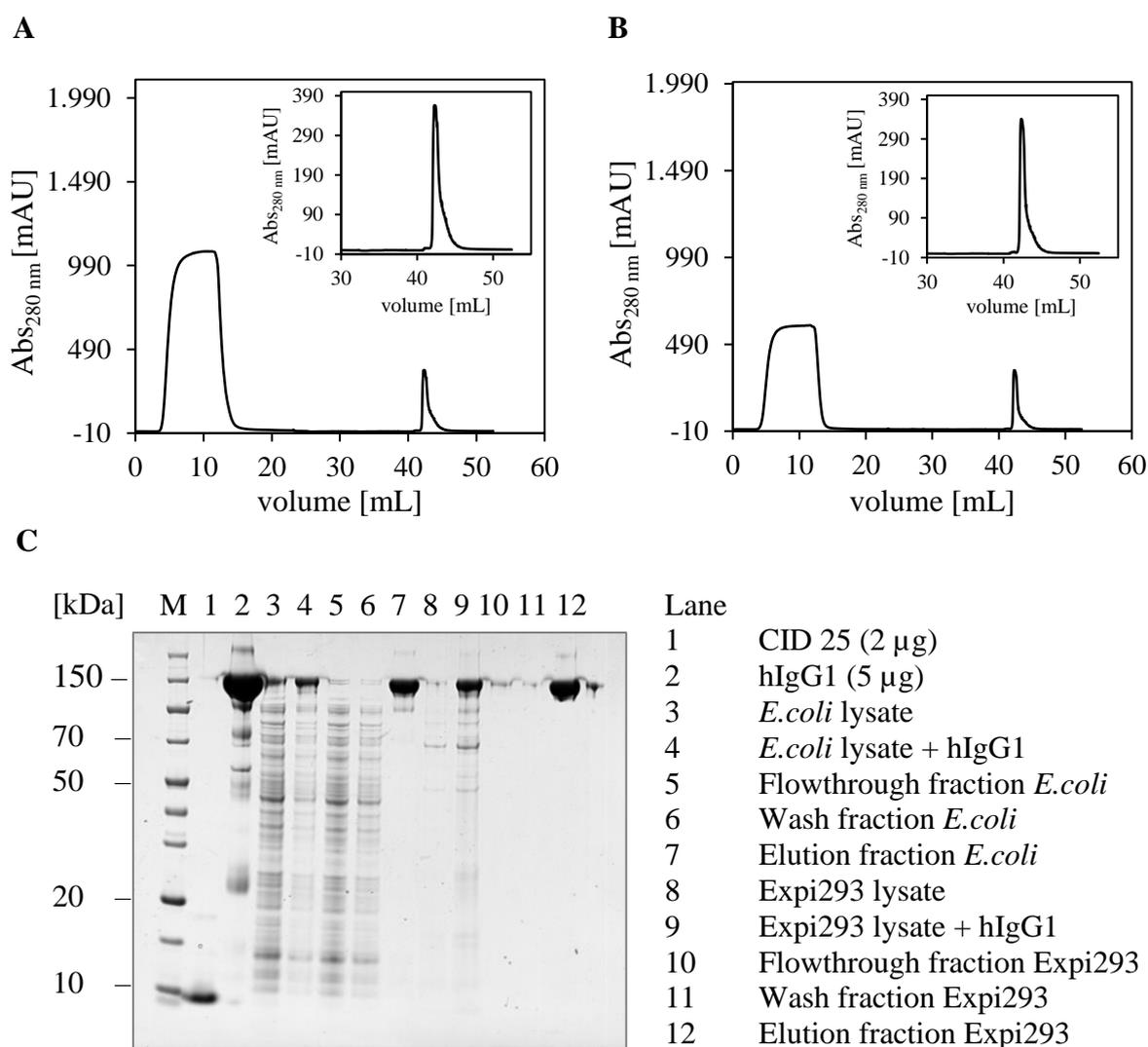
Figure 36: CID 25 as ligand for the affinity chromatographic purification of hIgG1.
CID 25 was coupled on the surface of a 1 mL HiTrap NHS-Activated HP column (Cytiva) via NHS chemistry. 1.95 mg hIgG1 [c = 0.3 mg/mL] was added to the prepared *E. coli* (**A**) and Expi293 (**B**) cell lysates, respectively. After sample loading and column washing, bound hIgG1 was eluted by a pH shift (pH 3.0) using 10 CV AIC buffer B (100 %) (smaller pictures). (**C**) SDS-PAGE analysis of selected purification fractions. 10 µL of each sample and 5 µL PageRuler Unstained Broad Range Protein Ladder (M, Fisher Scientific) were loaded on a NuPAGE Novex 4-12 % Bis-Tris gradient gel. The bands were visualized by Coomassie staining.

The integration of the peak areas revealed a yield of 91.8 % hIgG1 for the *E.coli* lysate (Figure 36 A) and 92.3 % for lysates from the Expi293 culture (Figure 36 B). The elution fractions obtained from the chromatographic purification of hIgG1 were analyzed by SDS-PAGE (2.2.4.5). As a result, only slight (< 5 %) protein-based impurities were detected in these fractions (Figure 36 C). The superimposition of the elution profiles of hIgG1 in 1x PBS pH 7.4 or in *E. coli* and Expi293 cell lysates revealed comparable curve progressions (Figure 37). The integration of the peak areas resulted in a yield of hIgG1 $\geq$ 90 %, respectively.
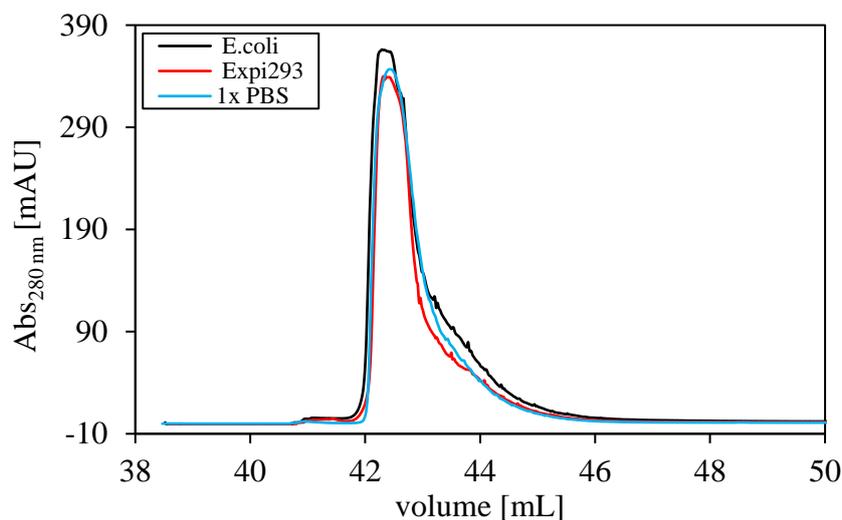
Figure 37: Purification of hIgG1 from different cell lysates using CID 25 as affinity ligand.
In each purification cycle, 1.95 mg hIgG1 [c = 0.3 mg/mL] was purified either from 1x PBS pH 7.4 or from *E. coli* (OD_{600} = 32) and Expi293 cell lysates (~17×10^6 cells). After sample loading and column washing (2.2.4.10), bound hIgG1 was eluted by a pH shift (pH 3.0) using 10 CV AIC buffer B (100 %).

4. Discussion

The high selectivity and affinity of antibodies to their corresponding antigens enables their use in therapy, diagnostics, and research (Lu *et al.* 2020). Although the complex structure of antibodies plays an important role in mediating effector functions, their complexity often limits their application. This has led to the development of antibody fragments (Borras *et al.*, 2010) that exhibit lower stability *in vivo* (Holliger & Hudson, 2005). Especially as ligands in affinity chromatography, antibody-based systems are often limited in their use despite their great advantages over conventional purification methods. High production costs of the comparatively complex antibodies are disadvantageous for a technical use since cost pressure plays an overriding role (Kaplon & Reichert, 2021). The use of polyclonal antibodies can reduce production costs but leads to a heterogeneous mixture and higher batch-dependent variances. Hence, the low reproducibility of these purification strategies can require an adaptation of process parameters.

As alternative approach, a variety of natural and engineered scaffolds have been investigated for the generation of novel binding proteins. Based on simply structured scaffolds, proteins with antibody-like properties can be obtained by diversification and selection. The feasibility has been demonstrated for various applications in therapy, diagnostics, and industry (Gebauer & Skerra, 2020). Due to their simple structure, independence of post-translational modifications and formation of disulfide bonds, these engineered binding proteins can be produced inexpensively in the cytosol of bacteria such as *E coli*. Small, naturally occurring helix bundle proteins are particularly well suited as scaffolds for the generation of engineered binding proteins. Based on individual domains of Protein A, the feasibility of generating inexpensive affinity ligands has already been demonstrated (Lu *et al.*, 2020).

In this work, domain B of staphylococcal Protein A was used as scaffold for the generation of IgG-binding proteins. First, the influence of natural loops on folding and function of domain B was investigated. For this, loops were partially or completely substituted for $(GS)_n$ linkers. Second, the order of helices in the three-helix bundle was changed by helix shuffling and two alternative loops with several randomized positions were inserted to connect the rearranged structural elements and generate diversity. It was investigated whether it is possible to change the helix 1/3 topology while maintaining the natural binding property, structure, and thermal stability of the three-helix bundle protein. Finally, the suitability of CID 25 as ligand for affinity chromatography was investigated. For this, CID 25 was coupled on the surface of a column matrix and purification of IgG from cell extracts was demonstrated. An affinity ligand generated in this way would allow inexpensive production in *E. coli* and specific, high-affinity IgG-binding. Antibodies and antibody fragments were selected as robustly detectable and scientifically relevant targets.

4.1 Library construction for the generation of novel domain B variants

The 3D structures of individual Protein A domains have been extensively studied in the past decades (Deis *et al.*, 2015). In this work, MD simulations with domain Z, an engineered domain B with A1V and G29A mutations, were performed to better understand the fold and the interhelical pattern of non-covalent interactions of the three-helix bundle protein. The Z domain

was used because high resolution NMR structures of wild type domain B with complete α-helical structures were not available. The results confirmed the close helix packing of the bundle inspired by Crick's "knobs-into-holes" (KIH) packing model, which is based on the idea that the hydrophobic amino acids "knobs" of one helix bury themselves into the "holes" between the hydrophobic knobs of their partner helix (Crick, 1953). The amphipathic helices of domain Z formed a hydrophobic core that is essential for the assembly of the polypeptide chain of the protein into its final folded form. Similar results can be obtained for a variety of three-helix bundles such as ABD (Nilvebrant & Hober, 2013), CAP (Schneider *et al.*, 1998), the villin headpiece (Lei *et al.*, 2007), and for most four-helix bundles such as cytochrome c and Rop (Bowles *et al.*, 2018). In addition to geometric aspects, orientation and topology of the helices and their connectivity can also determine the specific fold of helix bundle proteins. The results showed that domain Z is characterized by an antiparallel orientation of the helices (13-15 residues), which is typical for most three-helix bundle proteins. The helices are connected by two highly conserved loops. When comparing the topology of three-helix bundles with that of four-helix bundles, much higher topological diversity can be observed in arrangements of four helices, ranging from up-down-down-up (ferritin) to up-up-down-down (cytokine) conformations with long crossover connections between helices (Kamtekar & Hecht, 1995).

The obtained data revealed a variety of non-covalent interactions between the individual helices and, moreover, between the helices and the loops of domain Z. As the results showed, amino acids of both loops were highly involved in formation of van-der-Waals interactions and hydrogen bonds, suggesting that these regions could be involved in protein folding and stabilization. Rhys *et al.* reached a similar conclusion when they examined the energetic landscape of helix bundle proteins using computational energy-component analysis (Rhys *et al.*, 2019). In a study by Chou *et al.*, loop-helix interactions of three analyzed four-helix bundle proteins (cytochrome c/b-562, and methemerythrin) were even stronger than helix-helix interactions (Chou *et al.*, 1992). This suggests that loops can play a major role in stabilization and folding of helix bundle proteins such as domain B. However, further studies are required to gain a more detailed understanding of the energetic landscape and the contribution of loops to stabilization of domain B.

To investigate whether partial or complete substitution of the highly conserved loop residues affects the structural and functional behavior of domain B, five variants with $(GS)_n$ linkers were generated. Glycine-serine linkers were selected because they are rich in polar serine to ensure solubility and small size glycine that allows high conformational freedom (Tyshchuk *et al.*, 2017). Residues in loop 1, loop 2, or in both loops were substituted. In addition, only single amino acids in loop 1 within the helix 1/2 segment were substituted since this loop in particular shows only limited tolerance towards engineering (in-house company data). First, the secondary structure of the five variants was investigated by CD. Compared to wild type domain B, the proportion of α-helical structures was retained for variants possessing $(GS)_n$ substitutions only in loop 2 within the helix 2/3 segment. This result is in accordance with data from Gülich *et al.* where loop 2 of domain Z was replaced or extended by glycine linkers resulting in a retained secondary structure (Gülich *et al.*, 2000). A reduction in α-helical structures was observed for variants lacking the natural residues in loop 1 between the helices 1 and 2. In particular, domain B variants that had a N21G substitution showed significantly decreased α-helical content (CID 2 and CID 4), whereas a P20G substitution seemed to have only a marginal effect

on the secondary structure of domain B (CID 6). Substituting arginine for a glycine probably gives loop 1 and helix 2 a high flexibility. Arginine, which is known as a powerful turn-inducer residue tends to occur at the second position in β-turns. Moreover, arginine has been reported to form a variety of non-covalent interactions within domain B, indicating that this residue could be important for stabilization and folding of the protein.

To investigate functional differences of the domain B variants, binding to hIgG1-Fc was analyzed by SPR. In general, the binding behavior of the variants was different compared to wild type domain B. The observed decreases in binding affinities were primarily based on differences in k_{off} values. Variants with substitutions in loop 1 (except P20G) additionally differed in their k_{on} values, which could be caused by a structural rearrangement affecting the rate of target recognition. These results indicate that engineering of the loops, particularly loop 1, can have a major effect on structure and function of domain B. Gülich *et al.* also observed functional differences when comparing domain Z with loop 2 mutants mainly due to different k_{off} values (Gülich *et al.*, 2000). However, both loops to allow some degree of tolerance towards protein engineering attempts. Nevertheless, some questions remain unanswered and the influence of individual loop residues on protein folding and stabilization can only be assumed. In further experiments, the importance of individual loop residues could be determined by successively substituting single loop positions. Furthermore, variants with different loop lengths could be generated since it is known that besides composition also the length of connecting structures can influence folding and thermal stability of proteins (Linse *et al.*, 2020).

Based on the finding that the loops of domain B tolerate engineering attempts to a certain extent, two alternative loops were inserted to change the helix 1/3 topology of the three-helix bundle. Since there was no precise information on beneficial loop composition and length, loops with multiple randomized positions and a variable length were inserted to generate diversity in the protein libraries. A loop length of three to five (loop 1) and six to eight (loop 2) residues was selected to cover the calculated interhelical distances. Hence, the length of the novel loops was comparable to loops found in other helix bundle proteins, mostly consisting of up to 15 residues (Choi *et al.*, 2013). At DNA level, the gene fragments were divided into modular sections and VNN oligonucleotides (V = A/C/G, N = A/C/G/T) were used to generate diversity. The reduced set of nucleotides at the first position of the base triplet avoided internal stop codons and cysteines. Cysteine-coding sequences were excluded to prevent the formation of inter- and intramolecular disulfide bonds. If needed for later applications such as a directed coupling of proteins on matrices, single cysteines could be genetically added at the C- or N-terminus with minimal efforts. As an undesired side effect, the aromatic amino acids phenylalanine, tryptophan, and tyrosine were also excluded, which was considered as acceptable since these residues do not occur in the natural loops of domain B and generally show a weak tendency to appear in loops (Minuchehr & Goliaei, 2005). By combining the novel loop fragments, four domain B libraries (PACB1-4) that consist of nine sub-libraries were generated in parallel. The theoretical diversity of the libraries PACB1 and PACB2 was completely covered by the applied cloning strategy, while only larger sections of the highly diverse libraries PACB3 and PACB4 could be displayed. With a functionality of 65.4 % (PACB1), 69.3 % (PACB2), 66.9 % (PACB3), and 64.8 % (PACB4) at DNA level, the generated libraries showed a moderate quality. By including sequences that contain mutations, which did not lead to a reading frame

shift (nucleotide exchanges outside of the randomized regions or insertions and deletions), even a functionality of $> 85\,\%$ was achieved for all libraries. Other Protein A-based libraries generated by analogous methods show a comparable functionality (~89 %, in-house company data). The occurrence of sequences with stop codons (~3 %), reading frame shifts (~2 %) and mutations outside the randomized regions (~22 %) in all libraries is directly in line with results obtained for other libraries generated with degenerate oligonucleotides (Sayous *et al.*, 2020). However, codons of leucine ($< 20\,\%$) and serine ($< 36\,\%$) were underrepresented, which could be attributed to a lower proportion of corresponding base triplets due to the use of VNN oligonucleotides. Since codons of both amino acids were detected at different positions in the randomized loops, their reduced occurrence was considered as tolerable. Characterization of individual variants of the libraries PACB1-4 revealed comparable results with respect to thermal stability and soluble expression in *E. coli*. Although the examined domain B variants with (GS)$_n$ loops showed only marginal changes in their intrinsic properties, the sum of mutations within the library variants resulted in reduced solubility and stability (data not shown). The relatively large number of randomized positions (PACB4 = 13/65 $\triangleq$ 20 %) is probably the major reason for this finding (Soskine & Tawfik, 2010). A reduced solubility of library variants can also be observed for other in-house produced Protein A-based libraries (personal communication Dr. Lotze, Navigo Proteins).

Taken together, the combination of unevenly distributed codons and truncated or insoluble proteins that occur after gene expression reduces the effective library size and quality, a finding also observed for libraries generated with analogous methods (Nov, 2014). Since these circumstances can reduce the probability of success, it would be useful to optimize the domain B libraries for future applications. To improve the functionality and diversity of the libraries, pre-synthesized trinucleotides could be used for oligonucleotide synthesis that allow a defined composition of codons (Krumpe *et al.*, 2007). A more stringent cloning strategy with a reduced number of PCR cycles could significantly reduce the number of point mutations (Asif *et al.*, 2021). Alternative high-fidelity polymerases (PfuUltra) could be used to avoid point mutations by extending DNA fragments with high precision. To completely cover the theoretical diversity of the libraries PACB3 and PACB4, the amount of DNA templates used for PCR could be increased. To improve the stability and solubility of the library variants, the number of amino acid positions used for randomization could be reduced. Without knowing the exact structure and fold of the novel domain B variants, a re-analysis could be performed with different smaller libraries containing only two or three randomized positions in either loop 1 or loop 2. For this, the obtained data of domain B variants with (GS)$_n$ loops could serve as starting point (3.1). Thus, individual loop residues could be directly related to thermal stability and functionality of the proteins. Extension of natural loops by additionally inserted glycines could be another strategy to optimize the intrinsic properties of the library variants. In addition to loop optimization, engineering of α-helical regions could also be applied to improve the thermal stability. Based on high-throughput methods, it would be possible to randomize each amino acid position of the scaffold individually to identify stabilizing exchanges. Insertions of linking elements such as an intramolecular disulfide bond formed between two cysteines located on facing helices, could also contribute to protein stabilization (Wiedemann *et al.*, 2020). However, since data on folding and structure of the library variants and on favorable length and composition of alternative loops are currently insufficient, experiments on 3D structure

determination (X-ray or NMR) would be of great interest. The mentioned strategies for increasing the stability of the library variants could also be useful for later use of these proteins as affinity ligands. Typical strategies for regeneration of chromatography media mostly contain stringent cleaning steps at acidic and alkaline pH values. Established ligands such as domain Z are characterized by high chemical stability and allow multiple regeneration cycles under stringent conditions (Ramos-de-la-Peña *et al.*, 2019). Although no information on chemical stability of the novel domain B variants is available so far, the determined thermal stabilities (3.4.2, Table 37) suggest that such stringent conditions cannot be applied without a significant reduction of ligand activity.

4.2 Generation of novel domain B variants

Due to the high theoretical diversity of the generated libraries PACB1-4 (1.8×10^{10}-4.5×10^{15}), ribosome display was used for the selection and enrichment of IgG-binding domain B variants. Approximately 1.7×10^{12} DNA molecules per library served as input for the first round of biopanning. This value is comparable to diversities of 1×10^{12} commonly used for this selection system, although a manageable scope of 1×10^{14} molecules can be found in literature (Li *et al.*, 2019). The theoretical diversities of PACB3 and PACB4 could not be completely covered due to multiple cloning steps required to convert the protein variants into a RD-compatible format. If required for future selections, larger DNA quantities and ITT volumes could be used to display broader library frames. However, compared to other selection methods such as phage display, which are limited by the transformation efficiency of *E. coli* (10^9-10^{10}), high diversity was displayed for all libraries (Jaroszewicz *et al.*, 2022). Since only limited information on correct function and folding of variants was available, the goal of selection was a general enrichment of IgG-binding proteins by a moderate strategy. The first three selection rounds versus biotinylated hIgG1-Fc were performed under less stringent conditions including moderate wash steps and high concentrations of the target (Table 24). To enrich high-affinity variants with promising off-rates, selection pressure was significantly increased in the fourth selection round by addition of non-biotinylated hIgG1-Fc in solution. The RNA of target-bound ternary complexes was obtained by complexation of Mg^{2+} in the presence of the chelator EDTA instead of using pH-sensitive, enzymatic, or competitive elution methods (Dreier & Plückthun, 2018). The obtained RNA was transcribed into cDNA by reverse transcription, amplified by PCR, and subcloned in an expression vector that additionally encoded GFP. To evaluate the success of selection, binding of expressed protein pools to hIgG1-Fc was analyzed by GFP-based pool ELISA. The results showed that specific IgG-binding domain B variants were successfully enriched within four rounds of biopanning using ribosome display (Figure 13). The initial libraries also showed slight binding to hIgG1-Fc because intrinsic Fc binding of domain B was not eliminated during library generation. Interestingly, slight binding to BSA was observed for variants from PACB3 and PACB4, suggesting that the wash steps performed in the first three rounds were not sufficient to completely remove non-specifically bound proteins. However, after the fourth selection round, a reduction in binding signals versus BSA was observed, which can be probably attributed to the more stringent washing strategy. The addition of non-biotinylated hIgG1-Fc as competitor (off-rate selection) in round four resulted in a slight reduction in binding signals, which can be explained by the removal of variants with

weak affinities. Pools obtained after the fourth selection round were analyzed by NGS since the largest proportion of specific and highly affine IgG-binding variants was expected in these fractions. As a result, more than 1,200,000 sequences were obtained. The distribution of sequences differed considerably between the individual libraries probably caused by limitations associated with the determination of DNA concentrations by UV/Vis spectroscopy. All pools showed high diversity, although sequence motifs encoding G/PLT or (P)GGI/LT were enriched in loop 1 (~20 % each). As characteristic feature of a successful selection strategy, enrichments of sequence motifs or even entire protein variants (15-40 %) can also be observed for comparable in-house performed ribosome displays (personal communication Dr. Meysing, Navigo Proteins). Glycines and prolines are frequently found in loops of other proteins and are postulated to play an important role in chain compaction during the early stages of folding (Krieger *et al.*, 2005). Small and highly flexible glycine accelerates loop formation by decreasing the activation energy, whereas cis/trans proline allows significant changes in conformal directions. Hence, proline can be frequently found in β-turns of other helix bundle proteins due to its restricted φ angle, which is entropically favorable (Trevino *et al.*, 2007). The occurrence of threonine, which frequently forms hydrogen bonds in loops (Hu *et al.*, 2007) suggests that this amino acid could be involved in stabilizing the helix 3/2 interface. Hydrophobic leucine, an original loop residue of domain B involved in the formation of non-covalent interactions could also contribute to stabilization of the alternative scaffold. Loop 2 showed high diversity even after the fourth selection round, although enrichments of single amino acids at specific positions were observed. It can be assumed that loop 2 has a high degree of flexibility, which could affect the formation of the helix 2/1 interface and thus thermal stability and folding of the protein. Hence, loop 2 would be a good starting point for scaffold optimization to increase the probability of obtaining variants that exhibit improved thermal stability. To identify beneficial loops, smaller libraries in which only single loop 2 residues are randomized could be generated and used for ribosome display.

In addition to enrichment of sequence motifs and amino acids at specific positions, favorable combinations (3+6, 4+6, 4+7, 5+8) of the two loop fragments (RF1+RF2) were observed, indicating an appropriate length of the connective structures (Figure 15). A considerable proportion of functional variants with single, arbitrary point mutations in the α-helical regions (~30 % each) was detected, which could be attributed to non-specific binding of oligonucleotides used in PCR. In contrast to data found in literature, sequences with frame shift mutations and stop codons were not significantly reduced during selection (Plückthun, 2012). For PACB4, even an increase of variants with frame shift mutation at position E57 at the end of helix 1 was observed. This finding suggests that these variants shortened by six amino acids at the C-terminus still possess a certain functionality. Furthermore, it can be assumed that the combination of an increased amplification rate of the truncated DNA fragments and the moderate selection strategy has led to an enrichment of shortened but functional variants.

Per library, 3,240 variants were produced in microtiter scale, and binding to hIgG1-Fc and BSA was analyzed by automated GFP-based hit ELISA. An equal scope of PACB1-4 variants were analyzed, since NGS revealed a similar distribution of bases within the randomized loops and thus a remaining high diversity of all pools (Figure 15, Table 30). Although a hit rate of 1 % was technically covered by the applied screening scope, only a frame of the pools was represented due to the high diversity. Wild type domain B analyzed in parallel was used as

process control for cultivation, cell lysis, and ELISA. Lysates of empty vector pNP-013 were used as negative control and consistently low signals were observed, indicating that no significant cross contamination occurred during the screening process. A calculated Z-factor of 0.7, a statistical data quality indicator for HTS assays, represents an excellent screening assay (Iversen *et al.*, 2006). As a result of the primary screening, 2,666 IgG-binding variants that showed significantly increased signal amplitudes compared to empty vector pNP-013 were detected and defined as hits. Hit rates of ~10 % for PACB1 and ~22 % for PACB2-4 are in good accordance with successfully performed screening processes described in literature that revealed proteins with highly specific binding behavior against diverse targets (Goodwin *et al.*, 2020). The reduced hit rate observed for PACB1 suggests that the applied scaffold design could be unfavorable for generating functional variants. The loops of these library variants consist of three (loop 1) and six (loop 2) residues, which is shorter compared to the other libraries. Hence, significantly longer loops, as found in the libraries PACB2-4 with four or five (loop 1) and seven or eight residues (loop 2), are probably necessary to form functional helix bundles in a more promising manner. In secondary screening, BSA was used as off-target and specific binding to hIgG1-Fc was observed for most PACB2-4 variants. In contrast, almost all PACB1 variants showed non-specific binding to the targets, suggesting that the applied library design might not be suitable for generating functional IgG-specific binding proteins. A few variants (~7 %) obtained from PACB1-4 showed no binding to hIgG1-Fc at all. The exclusion of such false positive hits at the early stages of the screening process remains a key issue in the field of HTS, and many studies have been performed to eliminate such molecules by applying filters and thresholds based on statistical calculations (Sink *et al.*, 2010).

In total, 461 variants with soluble expression and specificity for hIgG1-Fc comparable to wild type domain B were selected as hits for µ-scale purification and BLI analysis (Figure 16, Table 31). The absolute signal amplitude (hIgG1-Fc) was used as further hit criterion, although this value does not necessarily correlate with high affinity because *E. coli* lysates were not normalized. As a result of BLI analysis, 424 variants (~92 %) were detected that showed specific binding to hIgG1-Fc (data not shown). In contrast, no binding of the variants to Avelumab Fab [c = 0.5 µM], another natural target of wild type domain B, was observed. The insufficient target concentration could be an explanation for this result, since it is known that wild type domain B binds to Fab fragments only with weak affinity of K_D ~12 µM (Jansson *et al.*, 1998). Much higher concentrations of Fab fragments should be used in future experiments. A few variants (~8 %), all obtained from PACB4, showed no binding to hIgG1-Fc and Fab at all. It remains unclear whether binding of these variants to the targets cannot be measured in the applied concentration range due to weak affinities or whether they show no binding due to weak thermal stability and structural unfolding caused by an unfavorable scaffold design. Nevertheless, 147 variants that showed soluble protein expression and the highest affinity for hIgG1-Fc compared to the others were selected for tagfree µ-scale production followed by SPR analysis. Sufficient quantities of tagfree protein were produced for all variants. The cleavage efficiency of TVMV protease (~75 %) was slightly decreased compared to data found in literature (Nallamsetty *et al.*, 2004). To obtain a higher yield of tagfree proteins, proteolytic cleavage could be optimized with regard to temperature, duration, and ratio of TVMV protease to POI. As a result of the SPR analysis, specific binding of almost all variants to hIgG1-Fc was confirmed (Figure 17). In general, the binding behavior of tagfree variants differed from that of

the corresponding GFP-format and wild type domain B. A decrease in affinity was observed mainly due to differences in k_{off} values. The k_{on} values of the variants also differed compared to wild type domain B, which could be explained by a structural rearrangement affecting the rate of target recognition. These results are in direct agreement with previously performed experiments that focused on substitution of domain B loops for $(GS)_n$ linkers (3.1). No binding to hIgG1-Fc was observed for five variants (2x PACB1 and 3x PACB4) that showed aggregation in tagfree form, suggesting that N-terminal GFP contributes to stabilization and folding of these proteins (Janczak *et al.*, 2015). Since there is no precise information on the sequences of these variants, no correlation between loop residues and protein stability can be made. However, 24 IgG-binding variants that showed the highest affinity for hIgG1-Fc compared to the others were selected for sequence analysis. The obtained sequences revealed 20 unique variants suitable for further analyses. Four variants showed a frame shift mutation at position E57 at the end of helix 1. This finding suggests that these variants shortened by six amino acids at the C-terminus still possess a certain functionality towards IgG-binding. In addition, four of the unique sequences showed point mutations outside the randomized loops. However, the distribution of these mutations and the type of inserted amino acid were random. Since none of the mutations prevailed, it is probable that they have only a marginal effect on the behavior of the variants. In all cases, point mutations of genes can be explained by process-related multiple PCR cycles and UV irradiation of DNA during library generation and selection. Moreover, undirected point mutations are frequently observed in ribosome display procedures and can be reinforced by performing error-prone PCR (Unkauf *et al.*, 2018).

A number of four or five (loop 1) and six or seven (loop 2) residues per randomized region was observed, which is in line with typical loop length of three-helix bundle proteins found in LIP database (Michalsky *et al.*, 2003). Interestingly, most variants were obtained from the libraries PACB2 (12×) and PACB3 (7×), underlining the favorable loop design of these libraries compared to PACB4 (1×) and PACB1, for which no IgG-binding variants were obtained. Significant differences in the amino acid distribution within the randomized loops were observed compared to the initial libraries (Figure 18, Table 32), indicating a successfully applied selection pressure during ribosome display. In particular, loop 1 within the helix 3/2 segment showed a present sequence motif (GGI/LT), whereas a higher diversity was observed in loop 2. However, codons of glycine, proline, arginine, and lysine were enriched at specific positions in loop 2. The beneficial properties of these amino acids in loops have already been discussed in this section. In addition, these amino acids can also be found in loops of other three- and four-helix bundle proteins such as ABD (PDB: 1PRB), cytochrome (PDB: 1CGO), DARPins (PDB: 2XEE) and the villin headpiece (PDB: 2RJY).

All variants except those with frame shift mutations were selected for the production in shake flasks. The expression rate of all variants was slightly reduced compared to wild type domain B. An impaired translation efficiency of the genetically modified proteins due to increased formation of mRNA secondary structure elements could be a possible explanation for this result (Mauger *et al.*, 2019). In addition to optimization of culture conditions, the use of alternative promotors, improved codon usage and alternative bacterial expression systems could be strategies to improve the manufacturability of the novel domain B variants (Francis & Page, 2010). Four variants (CID 13, CID 15, CID 18, and CID 21) showed mainly insoluble expression and were excluded from the purification. Compared to well-behaving domain B

variants, no pattern related to loop residues was identified that negatively affect the manufacturability of these proteins. Nevertheless, 14 variants were successfully purified. After purification by affinity chromatography, most variants showed proportions of multimers, aggregates or GFP-His cleavage products in the SEC (Figure 19). However, the different protein species were completely separated from the monomeric proteins by SEC. During the period of analysis, the purified variants showed no further tendency to aggregate, which was analyzed by SE-HPLC. The calculated molecular weight of the variants was slightly increased compared to theory, which can be explained by the non-globular shape of domain B (Figure 20). To overcome these limitations, it would be useful to perform alternative analytical methods such as MALS. Two variants (CID 19 and CID 22) showed precipitation after IMAC and could not be used for further purification by SEC. Interestingly, the amino acids at the first position in loop 1 were significantly different from those of variants that showed good manufacturability. Instead of highly flexible glycine (G), these two variants have either an arginine (R) or a histidine (H), suggesting that glycine at the first loop position might be important for the formation of correctly folded proteins.

To increase the probability of obtaining variants with good manufacturability and high affinity for IgG, the applied selection and screening conditions should be optimized in the future. Since all pools showed high diversity even after the fourth round of biopanning, future selections should be performed in a more stringent way with increased selection pressure. A pre-incubation of the library variants with empty, blocked beads could be performed at the beginning of the selection to remove non-specifically bound molecules. To more effectively isolate variants that exhibit a high affinity for hIgG1-Fc, a significant reduction of target amount, an increase of competitor concentration, and an extension of wash steps could be considered. To obtain variants that possess a domain-B like fold, yeast surface display could be used since a quality control mechanism via the endoplasmic reticulum ensures the efficient export of properly folded proteins (Cherf & Cochran, 2015). To enrich variants with high thermal stability, a heat step (≥ 50 °C) could be included in one of the selection rounds. Since ribosome display can only be applied to a limited extent due to the formation of heat-sensitive ternary complexes, more robust display methods such as mRNA display or phage display could be used (Wang & Lui, 2011). To achieve a comprehensive selection of heat-stable variants during screening, incubation of the *E. coli* lysates at ≥ 50 °C could be performed prior to ELISA. In addition, a reverse screening at pH 4-6 could be performed to identify proteins that show a mild elution profile suitable for chromatographic purification of pH-sensitive antibodies.

4.3 Functional and structural characterization of novel domain B variants

For each of the 14 successfully purified domain B variants (4.2), highly specific binding to hIgG1-Fc could be confirmed by two different binding assays, SPR and ELISA (Figure 21, Figure 22). SPR analyses of the variants revealed binding curves with different characteristics compared to wild type domain B. All variants showed a significant decrease in the dissociation rate constant k_{off}, whereas only slight differences in their association behavior were observed. These results are in accordance with previously performed experiments that focused on the substitution of loops for $(GS)_n$ linkers (3.1). Possible reasons and underlying causes for these

results have already been discussed in the sections 4.1 and 4.2. The affinities of the generated variants for hIgG1-Fc were in the range of $K_D = 10^{-6}$-10^{-8} M, which is less compared to wild type domain B that showed an affinity of $K_D = 10$ nM (Table 34). Slightly improved binding strengths of $K_D = 10^{-8}$-10^{-10} M are reported for other IgG-binding proteins such as bacterial Protein A/G/L/M or cyclic peptides (Choe *et al.*, 2016). However, the obtained K_D values of the novel domain B variants are in line with typical affinities of other engineered binding proteins (Mouratou *et al.*, 2015), antibodies or antibody fragments (Kikuchi *et al.*, 2005). Based on the biochemical properties of the amino acids in the novel loops, some speculations on binding affinity can be made. Loop 1 in the helix 3/2 segment consists of two highly flexible glycines, followed by hydrophobic leucin or isoleucine and hydrophilic threonine (GGI/LT). All analyzed variants showed these characteristic loop motifs, except CID 17, that had a hydrophilic asparagine instead of a highly flexible glycine at the first loop position (NGLT). However, the identified amino acids can be frequently found in loops and their beneficial properties in loops have already been discussed in sections 4.1 and 4.2. For loop 2 in the helix 2/1 segment, some patterns were observed that might be related to binding affinity of the novel domain B variants. Variants with hydrophobic residues such as valine, alanine, or proline at the first position in loop 2 showed an increased affinity for hIgG1-Fc compared to those having asparagine (CID 26), threonine (CID 12), or histidine (CID 14) at this or the first two positions. In particular, valine at the first position in loop 2 followed by positively charged arginine as observed for CID 16, CID 17, and CID 20 seems to have a beneficial effect on obtaining proteins with an affinity for Fc of at least $K_D = 10^{-7}$ M. Interestingly, variants in which charged amino acids occurred in the front region of loop 2 showed an increased affinity compared to variants, in which these amino acids were found at the end of the loop (CID 8, CID 9, CID 14, CID 23, and CID 24). However, apart from the patterns mentioned above, the loops of the novel domain B variants differed only slightly from each other and no direct correlation between sequence, folding and affinity for hIgG1-Fc could be drawn. Another finding of the SPR analyses was that the variants showed no binding to Avelumab-Fab. Since it is known that wild type B binds to Fab fragments with only weak affinity ($K_D \sim 12$ µM) (Jansson *et al.*, 1998), the insufficient concentration of provided target might be the main reason for this result. Therefore, higher concentrations of Fab fragments should be used for future analyses.

Four variants (CID 16, CID 17, CID 20, and CID 25) that showed the highest affinity for hIgG1-Fc ($K_D = 10^{-7}$-10^{-8} M) compared to the others were selected for further analyses by ELISA. As a result, specific binding of the variants to hIgG1-Fc was verified, but a decreased affinity was observed compared to wild type domain B (Figure 22, Table 34). The obtained K_D values were comparable to affinities determined by SPR, suggesting an excellent agreement of both methods. Compared to wild type domain B, the variants showed significantly decreased plateau phases, which did not reach a common level. Conformational changes caused by weak thermal stability of the variants ($T_m \sim 31$ °C) could affect the binding strength for Fc and the absolute signal amplitude. If necessary, the affinity of the novel domain B variants for hIgG1-Fc can be improved by various strategies previously discussed in this work. Error-prone PCR as a directed evolution approach can be used for affinity maturation of already obtained variants (Unkauf *et al.*, 2018). In addition, extended wash steps in combination with a reduction of target concentration during selection, a competitive screening process and successive engineering of

single loop positions can be mentioned as useful tools. To investigate whether the novel domain B variants compete with wild type domain B for Fc binding, a competitive inhibition ELISA was performed. As a result, all variants (CID 16, CID 17, CID 20, and CID 25) seemed to compete with wild type control for Fc binding. This finding suggests that the binding epitopes of the novel domain B variants could overlap with or even be identical to those of wild type domain B (Figure 23). The obtained K_i values were reduced by a factor of 3.1-6.2 compared to wild type domain B, indicating a weaker affinity of the variants for hIgG1-Fc (Table 35). For the variants, K_D calculation revealed affinities in the range of $K_D = 10^{-6}$-10^{-8} M, which were comparable to binding strengths determined by SPR and ELISA (Table 34). With this assay, it was not possible to identify specific amino acids involved in Fc binding because no structural information was considered. For this, an NMR titration experiment was performed, which will be discussed later in this section.

Due to previously obtained structural data of domain B variants containing $(GS)_n$ loops (3.1), it was assumed that structure and thermal stability of the empirically generated variants might also be affected. Based on manufacturability and affinity for hIgG1-Fc, five variants (CID 11, CID 16, CID 17, CID 20, and CID 25) were selected for characterization of secondary structure and thermal stability by CD (Table 36). After buffer exchange, CID 17 showed precipitation in low-salt buffer required for CD, suggesting that salts might stabilize the protein. This finding would be in line with numerous studies, examining the ability of salts to stabilize proteins (Hani & Cole, 2019). Interestingly, compared to the well-behaving variants, CID 17 exhibits a hydrophilic asparagine (N) at the first position in loop 1 instead of a highly flexible glycine. This suggests that glycine at the first loop position might be essential for the formation of proteins with domain B-like fold. As a result of the CD analysis, all domain B variants showed a primary α-helical structure (Figure 24). However, the variants showed slight differences in their mean residue ellipticity, which was expected because the number of loop-forming amino acids has been increased while the number of α-helical residues remained unchanged. CID 20 and CID 25 showed a spectrum and a calculated α-helical proportion almost comparable to wild type domain B, whereas for two variants (CID 11 and CID 16) a decrease of α-helical structures was observed. Since this method measures the mean residue ellipticity and not tertiary structure, the lower α-helical content could be related to an increased three-dimensional structural flexibility of these variants. Furthermore, partial denaturation of CID 11 could be assumed since thermal stability analysis revealed a T_m of 28.1 °C, which is decreased compared to the other variants ($T_m \sim 32$ °C) and wild type domain B ($T_m = 72.1$ °C). Thermal denaturation of variants was measured by heat-induced unfolding and refolding experiments in a temperature range of 10-90 °C using CD. The thermal stability analysis revealed a significant destabilization of all variants with T_m values between 28.1 °C and 32.8 °C (Figure 25). These findings are in line with an engineering approach by Gülich *et al.* that demonstrated a significant loss of thermal stability (~40 %) of domain Z when introducing glycine linkers (Gülich *et al.*, 2000). Reduced thermal stabilities were also observed for other helix bundle proteins with engineered loops such as ROP (Nagi & Regan, 1997). Since CD provides no information on tertiary structure and intramolecular interaction pattern of the novel domain B variants proteins, we can only speculate on the underlying causes of the decreased thermal stability. The reduced T_m of the variants suggest that structural factors such as helix orientation, interactivity of the novel loops,

packing, and hydrophobic core formation are attributed to destabilization of these proteins. Hence, experiments for 3D structure determination (X-Ray or NMR) would be of strong interest. Without knowing the tertiary structure of the variants, successive randomization of individual amino acids in the loops could reveal critical positions for maintaining thermal stability. It would be useful to insert cysteines in order to stabilize the entire protein since classical theories suggest that disulfide bond formation reduces the entropy of the denatured state (Betz, 1993). *In silico* MD simulations could also be useful to optimize the protein towards higher thermal stability by identifying residues, which do not tolerate substitutions.

To further characterize thermal-induced unfolding and refolding process, one variant (CID 25) that showed a nanomolar affinity for Fc (K_D = 52 nM) and a proportion of α-helical structures comparable to wild type domain B (CID 1) was selected for DSC. DSC measurements were performed in a temperature range of 10-60 °C (CID 25) or 10-90 °C (CID 1). As a result of the DSC analysis, thermal unfolding of CID 25 was observed at T_m = 32.1 °C, which is in good accordance with the T_m determined by CD (T_m = 31.2 °C, Table 37). For CID 25, a decreased enthalpy ΔH_{cal} of the unfolding reaction was observed compared to wild type domain B. Based on the calorimetric enthalpy ΔH_{cal}, conclusions regarding the compactness of the tertiary structure can be drawn (Dehghan-Nayeri & Rezaei-Tavirani, 2015), suggesting that the structure of CID 25 differs from that of wild type domain B. Since native loops are known to contribute significantly to the stabilization of domain B, we speculate that changing the helix topology and inserting novel loops results in a loose packing and a different helix orientation of the engineered three-helix bundle. In contrast to wild type domain B, irreversible refolding was observed for CID 25 using two methods DSC and CD. For protein species that show no or only partial refolding, a loss of structure and function can be assumed. In addition, refolding to non-native or misfolded states leads to accumulation of aggregates, which could be also observed for other proteins (Lyubarev & Kurganov, 2000). Interestingly, CID 25 showed a vertically shifted refolding curve, which was also detectable for wild type domain B. Since the temperature was successively increased up to 60 °C, protein modifications especially for heat-sensitive CID 25 can be assumed. For future DSC measurements, heating up to 45 °C should be sufficient to completely unfold CID 25 and decrease heat-induced protein modifications.

As a result of comparing the calorimetric enthalpy ΔH_{cal} with the van't Hoff enthalpy ΔH_{vH}, a non-two state folding of CID 25 can be assumed since enthalpies were different within the unfolding and refolding processes. This finding is in contrast to wild type domain B, where reversible two-state folding was confirmed by comparable enthalpies ΔH_{cal} and ΔH_{vH}. The two-state folding mechanism of domain B was already observed in the mid-90s by Bai *et al.* (Bai *et al.*, 1997). They found that domain B exhibits a very fast unfolding and refolding kinetic within six milliseconds even in the dead time of the stopped-flow apparatus. Currently, different folding theories exist, which differ in the chronological order of helix formation. Yang *et al.* performed a series of *ab initio* Monte Carlo simulations to determine the folding kinetics of domain B (Yang *et al.*, 2008). They concluded that the folding of the three-helix bundle proceeds in a fast burst phase as a relatively fast initial chain collapse. First, helix 3 (H3) is formed during initial collapse followed by partial folding of helix 1 (H1) and helix 2 (H2). In the slower phase, H1 and H2 are completely formed as relatively ordered α-helical hairpins. In contrast, other theories assume that starting from a H1-H2 nucleus, H3 is formed at the end (Nelson & Grishin, 2008). Sequential folding can also be observed in other helix bundle

proteins such as the headpiece of villin, where the temporal sequence of helix formation is H3 followed by H2 and H1 at a slower rate (Yang *et al.*, 2008). To analyze unfolding pathways of domain B, Alonso and Daggett performed native and denaturation MD simulations (Alonso & Daggett, 2000). Their results demonstrated that H3, located at the C-terminus of the protein, consistently denatures later than the other two helices, H1 and H2. These results are consistent with other experiments, suggesting that H3, which forms early in folding is more stable than H1 and H2 (Yang *et al.*, 2008). In contrast to wild type domain B, more folding intermediates (molten globules) can be assumed for CID 25. Molten globules can be found in the folding process of several small proteins such as albumin and cytochrome in early states of folding (Kuwajima, 2020). These folding intermediates are characterized by a substantial moiety of native-like secondary structure, resulting in a loosely organized hydrophobic core. However, at this point we can only speculate on the detailed mechanisms of folding and unfolding of CID 25 since DSC provides only limited kinetic information. For a temporal high resolution of protein folding and unfolding, powerful techniques such as stopped-flow, NMR, and more recently in-gel folding should be considered.

Next, CID 25 was selected for a detailed structural analysis by NMR. As a result, 80 % of the amid signals (prolines excluded) of $^{13}C^{15}N$ CID 25 detected in the $^{1}H^{15}N$-HSQC could be assigned, which is somewhat less than typical assignment levels of 95 % and above (personal communication Dr. Weininger, MLU Halle). The limited spectral dispersion, reduced thermodynamic stability, and thus increased amide exchange in the artificially designed protein can be named as obvious explanations of these findings. However, 96 % of the side chain resonances in the $^{1}H^{13}C$-HSQC could be assigned by ^{15}N-TOCSY-HSQC and HCCH-TOCSY, which is a typical assignment level for well-behaving proteins. The backbone φ/ψ dihedral angles and the secondary structure of CID 25 were predicted by TALOS+. The combination of φ and ψ angles illustrated in a Ramachandran plot revealed a mainly right-handed α-helical secondary structure of the variant comparable to wild type domain B (PDB: 1BDD). These findings are in excellent accordance with the obtained CD data, which revealed a mainly α-helical proportion of CID 25 and a domain B-like spectrum (Figure 24). The main moiety (85.7 %) of residues, which form α-helices, were detected in the allowed or in the partially allowed regions of the Ramachandran plot since the plot areas were applied with very high stringency. Some residues at the loop-helix transitions were categorized in the unallowed region, indicating that the secondary structure of these sections cannot be unambiguously determined. Since experimental 3D structure determination of CID 25 by NMR was not completed at the end of this work, tertiary structure prediction using the AI-based network AlphaFold was performed. As a result, three right-handed α-helices arranged in a compact bundle with a domain B-like structure were predicted for CID 25 (Figure 28 A). The three α-helices could also be verified experimentally by NMR. Hence, predictions by both TALOS+ and AlphaFold revealed comparable α-helical core sections. Because no NMR structures of wild type domain B with high resolution and complete helix sections were available, the NMR structure of domain Z was used for the structural alignment. Compared to that engineered domain B (PDB: 1Q2N), slight differences in helix orientation were observed, which could explain the reduced affinity of CID 25 for hIgG1-Fc. Marginal changes in the solvent accessible surface area, packing, non-covalent interactions, and salt bridge patterns can have a significant influence on the binding affinity of proteins. Using a large dataset of protein-protein interaction

structures with documented affinities, Erijman *et al.* examined structure-based features that could potentially define binding energetics. They found that affinity is highly correlated with the total number of non-covalent interactions such as hydrogen bonds, geometric complementarity, and conformational changes of side chains upon binding (Erijman *et al.*, 2014). Comparing the length of the α-helical structures of CID 25 with those of the theoretical expected helices inspired by wild type domain B, further differences can be observed (Figure 28 B). Compared to theory, helix 3 was predicted to be N-terminal shortened by the residues S6 and Q7. However, starting from Q7 the structure of the helix 3/2 segment predicted by AlphaFold looked structurally as expected, which is in good accordance with the TALOS+ prediction based on experimental NMR data. This finding suggests that the chosen length for loop 1 was appropriate to obtain proteins with a helix 3/2 segment comparable to wild type domain B. The structural prediction of the helix 2/1 segment by AlphaFold revealed slight differences compared to both TALOS+ and theory. An extension of helix 2 by the loop-forming residues A41, K42, and T43 was predicted by AlphaFold, whereas TALOS+ prediction revealed a shortening by the residues D39, D40, N48, and K49. It is well conceivable that the helices of the NMR structure are slightly expanded compared to the length predicted by TALOS+. Based on these results, we speculate that loop 2 possesses a certain degree of flexibility, which could cause these structural differences. In addition, the heterogeneity within the helix 2/1 segment could explain the decreased affinity of CID 25 for hIgG1Fc since this structural region is essential for the functionality of the protein. Considering these structural differences and the high diversity of loop 2 within the helix 2/1 segment even after the selection and screening process, it can be speculated that this region could be accessible for optimization. An NMR titration experiment using unlabeled hIgG1-Fc was performed to identify individual amino acids of ^{13}C^{15}N CID 25 involved in Fc binding. A ^{1}H^{15}N-TROSY-HSQC spectra was recorded for each of the four titration points. As a result, a line-broadening and a signal loss of almost all amide resonances, which completely disappeared at the end of the titration, were observed. In the end, almost all signals sensed the binding to hIgG1-Fc either directly, in which the signal disappeared to zero at one position in the spectrum and reappeared at another position (if detectable), or indirectly by increased relaxation in the larger complex. Differences in intensity and not in chemical shift are indicative for a slow exchange regime and a long lifetime of the formed complex (Günther, 2013). A similar conclusion was drawn for other engineered binding proteins such as Nanofitins, DARPins, and Affilin variants, which bind their respective targets with affinities in the low nM range (Gebauer & Skerra, 2020). In order to distinguish direct and indirect effects, signal strengths were compared under sub-stoichiometric conditions and the ratio of 1:0.35 (^{13}C^{15}N CID 25:hIgG1-Fc) was used to identify potential binding sites. By comparing the obtained titration points with theoretical titration curves, a 1:1 stoichiometry of the CID 25/hIgG1-Fc binding could be verified (Figure 29). NMR signals of the residues I25 (loop 1), I34, Q35, K38 (helix 2), Q51, F55, Y56, and L59 (helix 1) changed significantly upon binding of hIgG1-Fc due to a considerable decrease in NMR peak volume, suggesting that the environment of these amino acids was changed upon complex formation (Figure 30, Figure 31). Hence, the involvement of these amino acids in binding to hIgG1-Fc is highly probable. In addition, the residues T26 (loop 1), N31 (helix 2), and N53 (helix 1) showed a slight loss of peak volume, suggesting that these amino acids could function as further binding sites. We assume that the identified residues in helix 1 (Q51, N53, F55, Y56, L59) and helix 2 (N31, I34,

Q35, K38) are directly involved in Fc binding due to their high similarity (81.8 %) compared to the binding sites of wild type domain B (Figure 31). These findings are in good accordance with the results obtained from the competitive inhibition ELISA. In contrast, the identified loop residues I25 and T26 differ from the natural binding epitope of domain B, which is formed exclusively by the helices 1 and 2. In this context, different effects can be discussed. On the one hand, these loop-forming residues could be involved in Fc binding since they are surface exposed and potentially accessible to the target. On the other hand, the decrease in intensity could be caused by conformational changes of $^{13}C^{15}N$ CID 25 during complex formation. Binding between proteins is usually associated with structural rearrangements, ranging from small local changes in the side chains (lock-and-key) to global changes in the entire tertiary structure (induced-fit) (Tobi & Bahar, 2005). Such conformational changes upon Fc binding have also been described for domain B and its engineered descendant domain Z. Deis *et al.* observed a decrease in conformational heterogeneity upon Fc binding and a side chain substitution of domain B for an Fc side chain adjected to the binding surface (Deis *et al.*, 2015). Wahlberg's group discovered conformational changes in the domain Z residue Y14 upon binding to Fc (Wahlberg *et al.*, 2003). These findings could explain the slightly different side chain orientations of the identified residues of CID 25 probably involved in Fc binding. Due to the conformationally heterogenous nature of individual Protein A domains, exact side chain orientations are only transferrable to a very limited extent to those of CID 25. Experimental 3D structure determination of CID 25 by NMR would be useful to solve this issue. When comparing the obtained affinity of CID 25 (K_D = 52 nM) with regard to the presumed binding area of wild type domain B (~650 Å), similar magnitudes can be assumed. Gilbrecht and Koide compared affinities of binding proteins obtained from different scaffolds with the area hidden in the interaction surface (Gilbreth & Koide, 2012). In this publication, variants that had a binding area of 700-1,000 Å were summarized with affinities in the nanomolar range. For CID 25, the area of solvent exposed amino acids probably involved in Fc binding was calculated *in silico* to be almost ~616 Å, resulting in an affinity of K_D = 62 nM for hIgG1-Fc in solution (Figure 31). However, without a resolved 3D structure of the CID 25/hIgG1-Fc complex, only limited conclusion can be drawn and this value can only be used for an initial assumption.

4.4 CID 25 as ligand for the affinity chromatographic purification of hIgG1

Various preparative chromatography techniques can be used for the purification of antibodies and antibody fragments. More than 86 % of all approved monoclonal antibodies are purified using a Protein A chromatography step (Bolton & Mehta, 2016). In this process, antibodies are separated from impurities due to the reversible and highly specific binding of Protein A and its individual domains to Fc fragments. Engineered Protein A domains with improved chemical stability and high affinity for antibodies such as domain Z (MabSelect PrismA, Cytiva) or the mosaic domain (NGL-Impact A, Repligen) are industrially used as ligands. Bound antibodies can be eluted with remarkable purity and yield by drastically lowering the pH value (pH 3.0). However, one major limitation of this approach is the harsh elution of antibodies, which can cause denaturation of pH-sensitive targets multidomain proteins and virus particles (Mazzer *et al.*, 2015). In addition, lowering the pH can lead to the elution of non-specifically bound impurities since their interactions with the stationary phase can also be impaired by the applied

buffer condition. The preservation of functionality after immobilization on a solid carrier surface is a substantial requirement for the use of Protein A as affinity ligand. Initially, CID 25 was immobilized on the surface of a CM5-chip via NHS chemistry and the influence of undirected coupling on the binding behavior was analyzed by SPR. As a result, CID 25 showed specific binding to hIgG1-Fc with a nanomolar affinity of $K_D = 237$ nM (Figure 32, Table 40). Compared to data obtained from SPR analyses of CID 25 in solution (Figure 21), a significant increase in the dissociation rate constant k_{off} was observed, resulting in a 4.6-fold decrease in binding strength. Reduced affinities due to immobilization were also observed for wild type domain B (factor 3.1) and have been described in the literature for other binding proteins (Twair *et al.*, 2014). Compared to theory, the SPR signal amplitudes obtained for both CID 25 and wild type domain B were significantly lower. This suggests that the undirected coupling of the variants could partially result in non-functional proteins, possibly due to the blocking of binding sites. To increase the proportion of functional and accessible proteins, CID 25 could be immobilized site-specifically via C-terminal cysteine or biotin. The variant could be fused to a tag (His10-tag, Strep-tag II), although this strategy, based on direct modification of the molecule, could cause disadvantages depending on the application. Nevertheless, CID 25 fulfilled many criteria required for the use as affinity ligand. CID 25 would allow the purification of antibodies and antibody fragments with high affinity in nanomolar range. Compared to purification strategies based on fusion tags, only natively folded proteins are recognized due to the specific recognition of the target epitope.

Immobilization of CID 25 on the surface of the chromatography column was achieved by NHS chemistry, targeting the primary amino groups in the side chains of lysines and in the amino terminus (Fischer, 2010). CID 25 possesses seven solvent-exposed lysines (K4, K16, K17, K38, K42, K49, K63), six of which can be also found in wild type domain B. Lysines are located either in the α-helical structures or in the termini, K42 even in the randomized loop of the variant. The broad distribution of lysines indicates that CID 25 was coupled to the stationary phase in a rather undirected orientation since the applied immobilization method via NHS chemistry is characterized by low specificity. The proportion of functional immobilized CID 25 was estimated based on the ratio of coupled protein (5.1 mg) and dynamic binding capacity for hIgG1 (18.2 mg/mL) of the prepared column. For both CID 25 and wild type domain B, the measurements revealed only a small proportion (17.2 % and 24.7 %, respectively) of functional immobilized proteins. Similar results were observed for the undirected coupling of other engineered Protein A domains via primary amine groups (personal communication Dr. Fiedler, Navigo Proteins). It can be assumed that coupling of CID 25 via the side chains of the residues K38, K42, and K49 could cause steric hindrance and inaccessibility of the helix 2/1 interface required for Fc binding. In particular, proteins coupled via K38 are most likely non-functional, as this residue was identified as an Fc binding site (Figure 31). Nevertheless, an effective capturing of hIgG1 by immobilized CID 25 was demonstrated. A hIgG1 solution [c = 2 μM] (10× K_D) was loaded on the column, which corresponded well to typical antibody loads used in industrial purifications (Liu *et al.*, 2010). Even after a single purification step, the eluted target protein exhibited excellent purity ($\geq$ 95 %) and yield ($\geq$ 91 %). Similar results were observed for the purification of hIgG1 from *E. coli* and Expi293 cell extracts (Figure 37). In this work, lysates from Expi293 cultures were used to determine the specificity of CID 25 since, especially for eucaryotic cell lines, impurities could bind non-specifically to the column. The results

demonstrate that CID 25 can be used for the chromatographic purification of antibodies from pro- and eucaryotic cell extracts. In addition to specificity of the interaction between ligand and target, capacity of the chromatographic column is also an important criterion for the use in technical applications. A dynamic binding capacity of 18.2 mg/mL was determined for CID 25 (5.1 mg coupled), whereas 28 .1 mg/mL could be purified using a column functionalized with wild type domain B (5.5 mg coupled, Figure 34). Compared to commercial Protein A resins, the capacity of the CID 25 column was significantly lower due to the low proportion of functional coupled proteins. For alternative IgG-specific chromatography columns based on engineered Protein A domains (NGL-Impact A, MabSelect PrismA), binding capacities of ~80 mg/mL at a contact time of 5 min are described (manufacturer's specifications). However, the binding capacities of these ligands are only partially comparable to CID 25 since both MabSelect PrismA and NGL-Impact A are next-generation tandem repeat proteins. Such tandem repeats in combination with directed coupling of CID 25 could be useful strategies to significantly increase the binding capacity of the chromatographic column. In addition, flexible placeholder sequences such as PAS or G4S linkers could be genetically inserted between the tandem repeats to avoid steric hindrances (Schlapschy *et al.*, 2013). A lifetime study with ten cycles of hIgG1 capturing and elution each followed by a cleaning step at pH 2.0 revealed a 10 % decrease in binding capacity for CID 25, whereas the performance of the wild type domain B column was not affected (data not shown). We speculate that CID 25 could be pH-sensitive, but precise acid stability data were not available. In this context, further experiments on acid and alkaline stability of CID 25 would be of great interest.

Antibodies and antibody fragments bound to a Protein A are usually eluted from the column by a pH shift from neutral to low pH (3.0-3.7) (Ramos-de-la-Peña *et al.*, 2019). The results showed that a pH of 4.8 was required to elute hIgG1 from the CID 25 column, whereas hIgG1 purified via the wild type domain B column eluted at pH 3.7 (Figure 35). These results suggest that CID 25 could be used for mild elution of antibodies and Fc-fusion proteins, which could be particularly useful for pH-sensitive proteins (Mazzer *et al.*, 2015). For CID 25, a pre-peak was observed at pH 5.0 containing minor impurities (~7 %) and the target. The number of wash cycles should be increased in future purifications to reduce such non-specific binding of proteins to the column. Currently, there is only a limited selection of mild eluting Protein A ligands available in academia and industry. For example, NGL-Impact A HipH (Repligen) shows elution of pH-sensitive antibodies at pH 4.5, which prevents the formation of aggregates of the target. In other approaches, elution of antibodies at higher pH was enabled by inserting loop extensions (Gülich *et al.*, 2000) and binding site mutations in Protein A (Pabst *et al.*, 2014). Scheffel *et al.* demonstrated mild elution of antibodies using an engineered domain Z, so-called Z_{Ca} (Scheffel *et al.*, 2021). The original loop of domain Z between helix 2 and helix 3 was substituted for a calcium-binding motif inspired by the calmodulin loop. Thus, antibodies could be captured in the presence of calcium and eluted by removing the calcium ions. As another engineering approach, thermo-sensitive Protein A domains such as Bryzen Pro can be mentioned (Koguma *et al.*, 2013). Temperature-dependent elution of antibodies is enabled by introducing mutations in Protein A that reduce the hydrophobicity of the side chains buried inside the ligand. CID 25 with its T_m of 31.2 °C could also be used for a gentle elution of antibodies by raising the temperature shift (40 °C). However, such a column could probably be used only once because CID 25 shows incomplete refolding.

In conclusion, CID 25 offers high potential particularly for technical applications. It can be assumed that the structural and functional properties of CID 25 allow a mild elution of antibodies, which could be an advantage over established Protein A-based purification methods. By using this customized ligand, antibodies could be gently eluted by a moderate pH shift (pH 4.8). However, it should be noted that a column functionalized CID 25 exhibits a significantly reduced binding capacity and lifetime compared to commercial next-generation Protein A-based ligands. Hence, further engineering is necessary to improve the stability and affinity of CID 25.

References

Alfaleh, M.A., Alsaab, H.O., Mahmoud, A.B., Alkayyal, A.A., Jones. M.L., Mahler, S.M. and Hashem, A.M. (2020). Phage Display Derived Monoclonal Antibodies: From Bench to Bedside. Front Immunol *11*, 1986.

Alhuseinalkhudhur, A., Lubberink, M., Lindman, H., Tolmachev, V., Frejd, F.Y., Feldwisch, J., Velikyan, I. and Sörensen, J. (2020). Kinetic analysis of HER2-binding ABY-025 Affibody molecule using dynamic PET in patients with metastatic breast cancer. EJNMMI Res *10*(1), 21.

Alonso, D.O. and Daggett, V. (2000). Staphylococcal Protein A: unfolding pathways, unfolded states, and differences between the B and E domains. Proc Natl Acad Sci U S A *97*(1), 133-8.

Amstutz, P., Binz, H.K., Zahnd, C. and Plückthun, A. (2006). Ribosome Display: In Vitro Selection of Protein-Protein Interactions. Cell Biology *1*(3), 497-509.

Amstutz, P., Forrer, P., Zahnd, C. and Plückthun, A. (2001). In vitro display technologies: novel developments and applications. Curr Opin Biotechnol *12*(4), 400-5.

Asif, S., Khan, M., Arshad, M.W. and Shabbir, M.I. (2021). PCR Optimization for Beginners: A Step by Step Guide. Res Mol Med *9*(2), 81-102.

Bai, Y., Karimi, A., Dyson, H.J. and Wright, P.E. (1997). Absence of a stable intermediate on the folding pathway of Protein A. Protein Sci *6*(7), 1449-57.

Barendt, P.A., Ng, D.T., McQuade, C.N. and Sarkar, C.A. (2013). Streamlined protocol for mRNA display. ACS Comb Sci *15*(2), 77-81.

Bates, A. and Power, C.A. (2019). David vs. Goliath: The Structure, Function, and Clinical Prospects of Antibody Fragments. Antibodies *8*(2), 28.

Betz, S.F. (1993). Disulfide bonds and the stability of globular proteins. Protein Sci *2*(10), 1551-8.

Boder, E.T. and Wittrup, K.D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. Nat Biotechnol *15*(6), 553-7.

Bolton, G.R. and Mehta, K.K. (2016). The role of more than 40 years of improvement in Protein A chromatography in the growth of the therapeutic antibody industry. Biotechnol Prog *32*(5), 1193-1202.

Borras, L., Gunde, T., Tietz, J., Bauer, U., Hulmann-Cottier, V., Grimshaw, J.P. and Urech, D.M. (2010). Generic approach for the generation of stable humanized single-chain Fv fragments from rabbit monoclonal antibodies. J Biol Chem *285*(12), 9054-66.

Bowers, P.M., Horlick, R.A., Kehry, M.R., Neben, T.Y., Tomlinson, G.L., Altobell, L., Zhang, X., Macomber, J.L., Krapf, I.P., Wu, B.F., McConnell, A.D., Chau, B., Berkebile, A.D., Hare, E., Verdino, P. and King, D.J. (2014). Mammalian cell display for the discovery and optimization of antibody therapeutics. Methods *65*(1), 44-56.

Bowles, D.P., Yuan, C., Stephany, K.R., Lavinder, J.J., Hansen, A.L. and Magliery, T.J. (2018). Resonance assignments of wild-type and two cysteine-free variants of the four-helix bundle protein, Rop. Biomol NMR Assign *12*(2), 345-350.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem *72*, 248-54.

Buchanan, A. (2012). Evolution of protein stability using ribosome display. Methods Mol Biol *805*, 191-212.

Canard, B. and Sarfati, R.S. (1994). DNA polymerase fluorescent substrates with reversible 3'-tags. Gene *148*(1), 1-6.

Cherf, G.M. and Cochran, J.R. (2015). Applications of Yeast Surface Display for Protein Engineering. Methods Mol Biol *1319*, 155-75.

Chiu, M.H. and Prenner E.J. (2011). Differential scanning calorimetry: An invaluable tool for a detailed thermodynamic characterization of macromolecules and their interactions. J Pharm Bioallied Sci *3*(1), 39-59.

Chiu, M.L., Goulet D.R., Teplyakov, A. and Gilliland, G.L. (2019). Antibody Structure and Function: The Basis for Engineering Therapeutics. Antibodies *8*(4), 55.

Choe, W., Durgannavar, T.A. and Chung, S.J. (2016). Fc-Binding Ligands of Immunoglobulin G: An Overview of High Affinity Proteins and Peptides. Materials *9*(12), 994.

Choi, Y., Agarwal, S. and Deane, C.M. (2013). How long is a piece of loop? PeerJ *1*, e1.

Chou, K.C., Maggiora, G.M. and Scheraga, H.A. (1992). Role of loop-helix interactions in stabilizing four-helix bundle proteins. Proc Natl Acad Sci U S A *89*(16), 7315-9.

Cohen, S.N., Chang, A.C. and Hsu, L. (1972). Nonchromosomal antibiotic resistance in bacteria: genetic transformation of Escherichia coli by R-factor DNA. Proc Natl Acad Sci U S A *69*(8), 2110-4.

Crick, F.H.C. (1953). The packing of α-helices: simple coiled-coils. Acta Crystallographica *6*(8-9), 689-697.

Danielson, P.B., Büchs, J., Stöckmann, C. and Fogleman, J.C. (2004). Maximizing cell densities in miniprep-scale cultures with H15 medium and improved oxygen transfer. Biochem Eng J *17*(3), 175-180.

Dehghan-Nayeri, N. and Rezaei-Tavirani, M. (2015). The Interpretation of Protein Structure Through Relationship of Melting Point (Tm) and Enthalpy of Unfolding (ΔHU). Int J Anal Pharma Biomed Sci *4*(1), 47-50.

Deisenhofer, J. (1981). Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of Protein A from Staphylococcus aureus at 2.9- and 2.8-A resolution. Biochemistry *20*(9), 2361-70.

Deis, L.N., Pemble, C.W.4ᵗʰ, Qi, Y., Hagarman, A., Richardson, D.C., Richardson, J.S. and Oas, T.G. (2014). Multiscale conformational heterogeneity in staphylococcal Protein A: possible determinant of functional plasticity. Structure *22*(10), 1467-77.

Deis, L.N., Wu, Q., Wang, Y., Qi, Y., Daniels, K.G., Zhou, P. and Oas, T.G. (2015). Suppression of conformational heterogeneity at a protein-protein interface. Proc Natl Acad Sci U S A *112*(29), 9028-33.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*(3), 277-93.

Deveuve, Q., Lajoie, L., Barrault, B. and Thibault, G. (2020). The Proteolytic Cleavage of Therapeutic Monoclonal Antibody Hinge Region: More Than a Matter of Subclass. Front Immunol *11*, 168.

Ditzel, H.J. (2002). Rescue of a broader range of antibody specificities using an epitope-masking strategy. Methods Mol Biol *178*, 179-86.

Dixon, M. (1953). The determination of enzyme inhibitor constants. Biochem J *55*(1), 170-1.

Dower, W.J., Miller, J.F. and Ragsdale, C.W. (1988). High efficiency transformation of E. coli by high voltage electroporation. Nucleic Acids Res *16*(13), 6127-45.

Dreier, B. and Plückthun, A. (2018). Rapid Selection of High-Affinity Antibody scFv Fragments Using Ribosome Display. Methods Mol Biol 1827, 235-268.

Erijman, A., Rosenthal, E. and Shifman, J.M. (2014). How structure defines affinity in protein-protein interactions. PLoS One *9*(10), e110085.

Ey, P.L., Prowse, S.J. and Jenkin, C.R. (1978). Isolation of pure IgG1, IgG2a and IgG2b immunoglobulins from mouse serum using Protein A-sepharose. Immunochemistry *15*(7), 429-36.

Fischer, M.J. (2010). Amine coupling through EDC/NHS: a practical approach. Methods Mol Biol *627*, 55-73.

Francisco, J.A., Campbell, R., Iverson, B.L. and Georgiou, G. (1993). Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. Proc Natl Acad Sci U S A *90*(22), 10444-8.

Francis, D.M. and Page, R. (2010). Strategies to optimize protein expression in E. coli. Curr Protoc Protein Sci Chapter *5*(1), Unit 5.24.1-29.

Friedman, M., Nordberg, E., Höidén-Guthenberg, I., Brismar, H., Adams, G.P., Nilsson, F.Y., Carlsson, J. and Ståhl, S. (2007). Phage display selection of Affibody molecules with specific binding to the extracellular domain of the epidermal growth factor receptor. Protein Eng Des Sel *20*(4), 189-99.

Friguet, B., Chaffotte, A.F., Djavadi-Ohaniance, L. and Goldberg, M.E. (1985). Measurements of the true affinity constant in solution of antigen-antibody complexes by enzyme-linked immunosorbent assay. J Immunol Methods *77*(2), 305-19.

Garg, P. (2021). Selective Preference of Antibody Mimetics over Antibody, as Binding Molecules, for Diagnostic and Therapeutic Applications in Cancer Therapy. *11*(3), 10679-10689.

Gebauer, M. and Skerra, A. (2020). Engineered Protein Scaffolds as Next-Generation Therapeutics. Annu Rev Pharmacol Toxicol *60*, 391-415.

Gilbreth, R.N. and Koide, S. (2012). Structural insights for engineering binding proteins based on non-antibody scaffolds. Curr Opin Struct Biol *22*(4), 413-20.

Gill, S.C. and von Hippel, P.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem *182*(2), 319-26.

Goodwin, S., Shahtahmassebi, G. and Hanley, Q.S. (2020). Statistical models for identifying frequent hitters in high throughput screening. Sci Rep *10*(1), 17200.

Gouda, H., Torigoe, H., Saito, A., Sato, M., Arata, Y., and Shimada, I. (1992). Three-dimensional solution structure of the B domain of staphylococcal Protein A: comparisons of the solution and crystal structures. Biochemistry *31*(40), 9665-72.

Graille, M., Stura, E.A., Corper, A.L., Sutton, B.J., Taussig, M.J., Charbonnier, J.B. and Silverman, G.J. (2000). Crystal structure of a Staphylococcus aureus Protein A domain complexed with the Fab fragment of a human IgM antibody: structural basis for recognition of B-cell receptors and superantigen activity. Proc Natl Acad Sci U S A *97*(10), 5399-404.

Gülich, S., Uhlén, M. and Hober, S. (2000). Protein engineering of an IgG-binding domain allows milder elution conditions during affinity chromatography. J Biotechnol *76*(2-3), 233-44.

Gumbart, J., Schreiner, E., Wilson, D.N., Beckmann, R. and Schulten, K. (2012). Mechanisms of SecM-mediated stalling in the ribosome. Biophys J *103*(2), 331-41.

Günther, H. (2013). NMR Spectroscopy Basic Principles, Concepts, and Applications in Chemistry. Wiley-VCH 3.Auflage.

Guo, Z., Gibson, M., Sitha, S., Chu, S. and Mohanty, U. (2011). Role of large thermal fluctuations and magnesium ions in t-RNA selectivity of the ribosome. Proc Natl Acad Sci U S A *108*(10), 3947-51.

Hage, D.S. and Cazes, J. (2005). Handbook of Affinity Chromatography. CRC Press

Hanes, J. and Plückthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. Proc Natl Acad Sci U S A *94*(10), 4937-42.

Hanes, J., Schaffitzel, C., Knappik, A. and Plückthun, A. (2000). Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. Nat Biotechnol *18*(12), 1287-92.

Hani, F.M., Cole, A.E. and Altman, E. (2019). The ability of salts to stabilize proteins *in vivo* or intracellularly correlates with the Hofmeister series of ions. Int J Biochem Mol Biol *10*(3), 23-31.

Hankin, E. (1890). A Cure for Tetanus and Diphtheria. Nature *43*, 121-123.

Heinrich, P.C., Müller, M., Graeve, L., and Koch, H. G. (2021). Löffler/Petrides Biochemie und Pathobiochemie. Springer 10.Auflage.

He, M. and Taussig, M.J. (2002). Ribosome display: cell-free protein display technology. Brief Funct Genomic Proteomic *1*(2), 204-12.

Hey, T., Fiedler, E., Rudolph, R. and Fiedler, M. (2005). Artificial, non-antibody binding proteins for pharmaceutical and industrial applications. Trends Biotechnol *23*(10), 514-22.

Hilgarth, R.S. and Lanigan, T.M. (2019). Optimization of overlap extension PCR for efficient transgene construction. MethodsX *7*, 100759.

Holliger, P. and Hudson, P.J. (2005). Engineered antibody fragments and the rise of single domains. Nat Biotechnol *23*(9), 1126-36.

Holtzhauer, M. (1996). Methoden in der Proteinanalytik. Springer

Horstmann, B.J., Kenney, C.N. and Chase, H.A. (1986). Adsorption of proteins on Sepharose affinity adsorbents of varying particle size. J Chromatogr *361*, 179-90.

Hu, X., Wang, H., Ke, H. and Kuhlman, B. (2007). High-resolution design of a protein loop. Proc Natl Acad Sci U S A *104*(45), 17668-73.

Iversen, P.W., Eastwood, B.J., Sittampalam, G.S. and Cox, K.L. (2006). A comparison of assay performance measures in screening assays: signal window, Z' factor, and assay variability ratio. J Biomol Screen *11*(3), 247-52.

Janczak, M., Bukowski, M., Górecki, A., Dubin, G., Dubin, A. and Wladyka, B. (2015). A systematic investigation of the stability of green fluorescent protein fusion proteins. Acta Biochim Pol *62*(3), 407-11.

Jansson, B., Uhlén, M. and Nygren, P.A. (1998). All individual domains of staphylococcal Protein A show Fab binding. FEMS Immunol Med Microbiol *20*(1), 69-78.

Jaroszewicz, W., Morcinek-Orłowska, J., Pierzynowska, K., Gaffke, L. and Węgrzyn, G. (2022). Phage display and other peptide display technologies. FEMS Microbiol Rev *46*(2), fuab052.

Jensen, K. (1958). A normally occurring Staphylococcus antibody in human serum. APMIS *44*, 421-428.

Johnson, B.A. and Blevins, R.A. (1994). NMR View: A computer program for the visualization and analysis of NMR data. J Biomol NMR *4*(5), 603-14.

Johnsson, B., Löfås, S. and Lindquist, G. (1991). Immobilization of proteins to a carboxymethyldextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. Anal Biochem *198*(2), 268-77.

Kahl, M., Fiedler, E., Bosse-Doenecke, E., Katzschmann, A., Bobolowski, H., Lotze, J. and Meysing, M. (2021). Binding protein specific for the spike protein of severe acute respiratory syndrome corona virus 2 (SARS-COV-2). International patent publication WO2021/260075A1.

Kamat, V. and Rafique, A. (2017). Designing binding kinetic assay on the bio-layer interferometry (BLI) biosensor to characterize antibody-antigen interactions. Anal Biochem *536*, 16-31.

Kamtekar, S. and Hecht, M.H. (1995). The four-helix bundle: what determines a fold?. FASEB *9*(11), 1013-1022.

Kanje, S., Scheffel, J., Nilvebrant, J. and Hober, S. (2020). Approaches to the Purification, Analysis and Characterization of Antibody-Based Therapeutics. Elsevier Chapter 2, 35-54.

Kanje, S., Venskutonytė, R., Scheffel, J., Nilvebrant, J., Lindkvist-Petersson, K. and Hober, S. (2018). Protein Engineering Allows for Mild Affinity-based Elution of Therapeutic Antibodies. J Mol Biol *430*(18 Pt B), 3427-3438.

Kaplon, H., Muralidharan, M., Schneider, Z. and Reichert, J.M. (2020). Antibodies to watch in 2020. MAbs *12*(1), 1703531.

Kaplon, H. and Reichert, J.M. (2021). Antibodies to watch in 2021. MAbs *13*(1), 1860476.

Keiler, K.C., Waller, P.R. and Sauer, R.T. (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. Science *271*(5251), 990-3.

Kikuchi, Y., Uno, S., Nanami, M., Yoshimura, Y., Iida, S., Fukushima, N. and Tsuchiya, M. (2005). Determination of concentration and binding affinity of antibody fragments by use of surface plasmon resonance. J Biosci Bioeng *100*(3), 311-7.

Kim, H.K., Cheng, A.G., Kim, H.Y., Missiakas, D.M. and Schneewind, O. (2010). Nontoxigenic Protein A vaccine for methicillin-resistant Staphylococcus aureus infections in mice. J Exp Med *207*(9), 1863-70.

Knick, P., Bobolowski, H., Kahl, M., Fiedler, E. and Haupts, U. (2022). Immunoglobulin binding proteins for affinity purification. International patent publication WO2022/013272A1

Koguma, I., Yamashita, S., Sato, S., Okuyama, K. and Katakura, Y. (2013). Novel purification method of human immunoglobulin by using a thermo-responsive Protein A. J Chromatogr A *1305*, 149-53.

Köhler, G. and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. Nature *256*(5517), 495-7.

Korendovych, I.V. and DeGrado, W.F. (2020). *De novo* protein design, a retrospective. Q Rev Biophys *53*, e3.

Krieger, F., Möglich, A. and Kiefhaber, T. (2005). Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. J Am Chem Soc *127*(10), 3346-52.

Krumpe, L.R., Schumacher, K.M., McMahon, J.B., Makowski, L. and Mori, T. (2007). Trinucleotide cassettes increase diversity of T7 phage-displayed peptide library. BMC Biotechnol *7*, 65.

Ku, J. and Schultz, P.G. (1995). Alternate protein frameworks for molecular recognition. Proc Natl Acad Sci U S A *92*(14), 6552-6.

Kunamneni, A., Ogaugwu, C., Bradfute, S. and Durvasula, R. (2020). Ribosome Display Technology: Applications in Disease Diagnosis and Control. Antibodies *9*(3), 28.

Kuwajima, K. (2020). The Molten Globule, and Two-State vs. Non-Two-State Folding of Globular Proteins. Biomolecules *10*(3), 407.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*(5259), 680-5.

Lei, H., Wu, C., Liu, H. and Duan, Y. (2007). Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. Proc Natl Acad Sci U S A *104*(12), 4925-30.

Liedberg, B., Nylander, C. and Lunström, I. (1983). Surface plasmon resonance for gas detection and biosensing. Sensors and Actuators *4*, 299-304.

Linhult, M., Gülich, S., Gräslund, T., Simon, A., Karlsson, M., Sjöberg, A., Nord, K. and Hober, S. (2004). Improving the tolerance of a Protein A analogue to repeated alkaline exposures using a bypass mutagenesis approach. Proteins *55*(2), 407-16.

Linse, S., Thulin, E., Nilsson, H. and Stigler, J. (2020). Benefits and constrains of covalency: the role of loop length in protein stability and ligand binding. Sci Rep *10*(1), 20108.

Li, R., Kang, G., Hu, M. and Huang, H. (2019). Ribosome Display: A Potent Display Technology used for Selecting and Evolving Specific Binders with Desired Properties. Mol Biotechnol *61*(1), 60-71.

Litwack, G. (2021). Human Biochemistry. Elsevier 2.Auflage.

Liu, H.F., Ma, J., Winter, C. and Bayer, R. (2010). Recovery and purification process development for monoclonal antibody production. MAbs *2*(5), 480-99.

Löfkvist, T. and Sjöquist, J. (1962). Chemical and serological analysis of antigen preparations from staphylococcus aureus. Acta Pathol Microbiol Scand 56(3), 295-304.

Lottspeich, F. (2006). Bioanalytik. Spektrum Akademischer Verlag 2.Auflage.

Lu, R.M., Hwang, Y.C., Liu, I.J., Lee, C.C., Tsai, H.Z., Li, H.J. and Wu, H.C. (2020). Development of therapeutic antibodies for the treatment of diseases. J Biomed Sci *27*(1), 1.

Lyubarev, A.E. and Kurganov, B.I. (2000). Analysis of DSC data relating to proteins undergoing irreversible thermal denaturation. Journal of Thermal Analysis and Calorimetry *62*, 49-60.

Martin, I.V. and MacNeill, S.A. (2002). ATP-dependent DNA ligases. Genome Biol *3*(4), REVIEWS3005.

Mattheakis, L.C., Bhatt, R.R. and Dower, W.J. (1994). An in vitro polysome display system for identifying ligands from very large peptide libraries. Proc Natl Acad Sci U S A *91*(19), 9022-6.

Mauger, D.M., Cabral, B.J., Presnyak, V., Su, S.V., Reid, D.W., Goodman, B., Link, K., Khatwani, N., Reynders, J., Moore, M.J. and McFadyen, I.J. (2019). mRNA structure regulates protein expression through changes in functional half-life. Proc Natl Acad Sci U S A *116*(48), 24075-24083.

Mazzer, A.R., Perraud, X., Halley, J., O'Hara, J. and Bracewell, D.G. (2015). Protein A chromatography increases monoclonal antibody aggregation rate during subsequent low pH virus inactivation hold. J Chromatogr A *1415*, 83-90.

Meyers, J.A., Sanchez, D., Elwell, L.P. and Falkow, S. (1976). Simple agarose gel electrophoretic method for the identification and characterization of plasmid deoxyribonucleic acid. J Bacteriol *127*(3), 1529-37.

Michalsky, E., Goede, A. and Preissner, R. (2003). Loops In Proteins (LIP)--a comprehensive loop database for homology modelling. Protein Eng *16*(12), 979-85.

Minakuchi, K., Murata, D., Okubo, Y., Nakano, Y. and Yoshida, S. (2013). Remarkable alkaline stability of an engineered Protein A as immunoglobulin affinity ligand: C domain having only one amino acid substitution. Protein Sci *22*(9), 1230-8.

Minuchehr, Z. and Goliaei, B. (2005). Propensity of amino acids in loop regions connecting beta-strands. Protein Pept Lett *12*(4), 379-82.

Mouratou, B., Béhar, G. and Pecorari, F. (2015). Artificial affinity proteins as ligands of immunoglobulins. Biomolecules *5*(1), 60-75.

Müller-Esterl, W. (2018). Biochemie. Springer 3.Auflage.

Mullis, K.B. and Faloona, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol *155*, 335-50.

Myers, J.K. and Oas, T.G. (2001). Preorganized secondary structure as an important determinant of fast protein folding. Nat Struct Biol *8*(6), 552-8.

Myrhammar, A., Rosik, D. and Karlström, A.E. (2020). Photocontrolled Reversible Binding between the Protein A-Derived Z Domain and Immunoglobulin G. Bioconjug Chem *31*(3), 622-630.

Nagi, A.D. and Regan, L. (1997). An inverse correlation between loop length and stability in a four-helix-bundle protein. Fold Des *2*(1), 67-75.

Nallamsetty, S., Kapust, R.B., Tözsér, J., Cherry, S., Tropea, J.E., Copeland, T.D. and Waugh, D.S. (2004). Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. Protein Expr Purif *38*(1), 108-15.

Nathans, D. and Smith, H.O. (1975). Restriction endonucleases in the analysis and restructuring of dna molecules. Annu Rev Biochem *44*, 273-93.

Nelson, E.D. and Grishin, N.V. (2008). Folding domain B of Protein A on a dynamically partitioned free energy landscape. Proc Natl Acad Sci U S A *105*(5), 1489-93.

Nic, M., Jirat, J. and Kosata, B. (2006). The XML Gold Book Online

Nilsson, B., Moks, T., Jansson, B., Abrahmsén, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T.A. and Uhlén, M. (1987). A synthetic IgG-binding domain based on staphylococcal Protein A. Protein Eng *1*(2), 107-13.

Nilvebrant, J. and Hober, S. (2013). The albumin-binding domain as a scaffold for protein engineering. Comput Struct Biotechnol J *6*, e201303009.

Nov, Y. (2014). Probabilistic methods in directed evolution: library size, mutation rate, and diversity. Methods Mol Biol *1179*, 261-78.

Odegrip, R., Coomber, D., Eldridge, B., Hederer, R., Kuhlman, P.A., Ullman, C., FitzGerald, K. and McGregor, D. (2004). CIS display: In vitro selection of peptides from libraries of protein-DNA complexes. Proc Natl Acad Sci U S A *101*(9), 2806-10.

Pabst, T.M., Palmgren, R., Forss, A., Vasic, J., Fonseca, M., Thompson, C., Wang, W.K., Wang, X. and Hunter, A.K. (2014). Engineering of novel Staphylococcal Protein A ligands to enable milder elution pH and high dynamic binding capacity. J Chromatogr A *1362*, 180-5.

Plückthun, A. (2012). Ribosome display: a perspective. Methods Mol Biol *805*, 3-28.

Pope, B. and Kent, H.M. (1996). High efficiency 5 min transformation of Escherichia coli. Nucleic Acids Res *24*(3), 536-7.

Popova, B., Schubert, S., Bulla, I., Buchwald, D. and Kramer, W. (2015). A Robust and Versatile Method of Combinatorial Chemical Synthesis of Gene Libraries via Hierarchical Assembly of Partially Randomized Modules. PLoS One *10*(9), e0136778.

Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. J Mol Biol *7*, 95-9.

Ramos-de-la-Peña, A.M., González-Valdez, J. and Aguilar, O. (2019). Protein A chromatography: Challenges and progress in the purification of monoclonal antibodies. J Sep Sci *42*(9), 1816-1827.

Reiersen, H. and Rees, A.R. (1999). An engineered minidomain containing an elastin turn exhibits a reversible temperature-induced IgG binding. Biochemistry *38*(45), 14897-905.

Rhys, G.G., Wood, C.W., Beesley, J.L., Zaccai, N.R., Burton, A.J., Brady, R.L., Thomson, A.R. and Woolfson, D.N. (2019). Navigating the Structural Landscape of De Novo α-Helical Bundles. J Am Chem Soc *141*(22), 8787-8797.

Rigi, G., Ghaedmohammadi, S. and Ahmadian, G. (2019). A comprehensive review on staphylococcal Protein A (SpA): Its production and applications. Biotechnol Appl Biochem *66*(3), 454-464.

Roberts, R.W. and Szostak, J.W. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc Natl Acad Sci U S A *94*(23), 12297-302.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A *74*(12), 5463-7.

Sayous, V., Lubrano, P., Li, Y. and Acevedo-Rocha, C.G. (2020). Unbiased libraries in protein directed evolution. Biochim Biophys Acta Proteins Proteom *1868*(2), 140321.

Scheffel, J., Kanje, S. and Hober, S. (2021). Z_{Ca}: A Protein A-Derived Domain with Calcium-Dependent Affinity for Mild Antibody Purification. Methods Mol Biol *2178*, 245-249.

Schilling, J., Schöppe, J. and Plückthun, A. (2014). From DARPins to LoopDARPins: novel LoopDARPin design allows the selection of low picomolar binders in a single round of ribosome display. J Mol Biol *426*(3), 691-721.

Schlapschy, M., Binder, U., Börger, C., Theobald, I., Wachinger, K., Kisling, S., Haller, D. and Skerra, A. (2013). PASylation: a biological alternative to PEGylation for extending the plasma half-life of pharmaceutically active proteins. Protein Eng Des Sel *26*(8), 489-501.

Schneider, J.P., Lombardi, A. and DeGrado, W.F. (1998). Analysis and design of three-stranded coiled coils and three-helix bundles. Fold Des *3*(2), R29-40.

Shao, Q. and Gao, Y.Q. (2011). The relative helix and hydrogen bond stability in the B domain of Protein A as revealed by integrated tempering sampling molecular dynamics simulation. J Chem Phys *135*(13), 135102.

Shen, Y., Delaglio, F., Cornilescu, G. and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR *44*(4), 213-23.

Sink, R., Gobec, S., Pečar, S. and Zega, A. (2010). False positives in the early stages of drug discovery. Curr Med Chem *17*(34), 4231-55.

Smith, G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science *228*(4705), 1315-7.

Soskine, M. and Tawfik, D.S. (2010). Mutational effects and the evolution of new protein functions. Nat Rev Genet *11*(8), 572-82.

Ståhl, S., Gräslund, T., Eriksson Karlström, A., Frejd, F.Y., Nygren, P.Å. and Löfblom, J. (2017). Affibody Molecules in Biotechnological and Medical Applications. Trends Biotechnol *35*(8), 691-712.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr Purif *41*(1), 207-34.

Tobi, D. and Bahar, I. (2005). Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. Proc Natl Acad Sci U S A *102*(52), 18908-13.

Trevino, S.R., Schaefer, S., Scholtz, J.M. and Pace, C.N. (2007). Increasing protein conformational stability by optimizing beta-turn sequence. J Mol Biol *373*(1), 211-8.

Turbadar, T. (1959). Complete absorption of light by thin metal films. Proc Phys Soc *73*(1), 40.

Twair, A., Al-Okla, S., Zarkawi, M. and Abbady, A.Q. (2014). Characterization of camel nanobodies specific for superfolder GFP fusion proteins. Mol Biol Rep *41*(10), 6887-98.

Tyshchuk, O., Völger, H.R., Ferrara, C., Bulau, P., Koll, H. and Mølhøj, M. (2017). Detection of a phosphorylated glycine-serine linker in an IgG-based fusion protein. MAbs *9*(1), 94-103.

Unkauf, T., Hust, M. and Frenzel, A. (2018). Antibody Affinity and Stability Maturation by Error-Prone PCR. Methods Mol Biol *1701*, 393-407.

Virnekäs, B., Ge, L., Plückthun, A., Schneider, K.C., Wellnhofer, G. and Moroney, S.E. (1994). Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. Nucleic Acids Res *22*(25), 5600-7.

Wahlberg, E., Lendel, C., Helgstrand, M., Allard, P., Dincbas-Renqvist, V., Hedqvist, A., Berglund, H., Nygren, P.A. and Härd, T. (2003). An affibody in complex with a target protein: structure and coupled folding. Proc Natl Acad Sci U S A *100*(6), 3185-90.

Wang, H. and Liu, R. (2011). Advantages of mRNA display selections over other selection techniques for investigation of protein-protein interactions. Expert Rev Proteomics *8*(3), 335-46.

Weber, K. and Osborn, M. (1969). The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. J Biol Chem *244*(16), 4406-12.

Wiedemann, C., Kumar, A., Lang, A. and Ohlenschläger, O. (2020). Cysteines and Disulfide Bonds as Structure-Forming Units: Insights From Different Domains of Life and the Potential for Characterization by NMR. Front Chem *8*, 280.

Yan, Q., Huang, M., Lewis, M.J. and Hu, P. (2018). Structure Based Prediction of Asparagine Deamidation Propensity in Monoclonal Antibodies. MAbs *10*(6), 901-912.

Yang, J.S., Wallin, S., Shakhnovich, E.I. (2008). Universality and diversity of folding mechanics for three-helix bundle proteins. Proc Natl Acad Sci U S A *105*(3), 895-900.

Yang, J.T., Wu, C.S. and Martinez, H.M. (1986). Calculation of protein conformation from circular dichroism. Methods Enzymol *130*, 208-69.

Zahnd, C., Amstutz, P. and Plückthun, A. (2007). Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. Nat Methods *4*(3), 269-79.

Zajc, C.U., Salzer, B., Taft, J.M., Reddy, S.T., Lehner, M. and Traxlmayr, M.W. (2021). Driving CARs with alternative navigation tools - the potential of engineered binding scaffolds. FEBS J *288*(7), 2103-2118.

Appendix
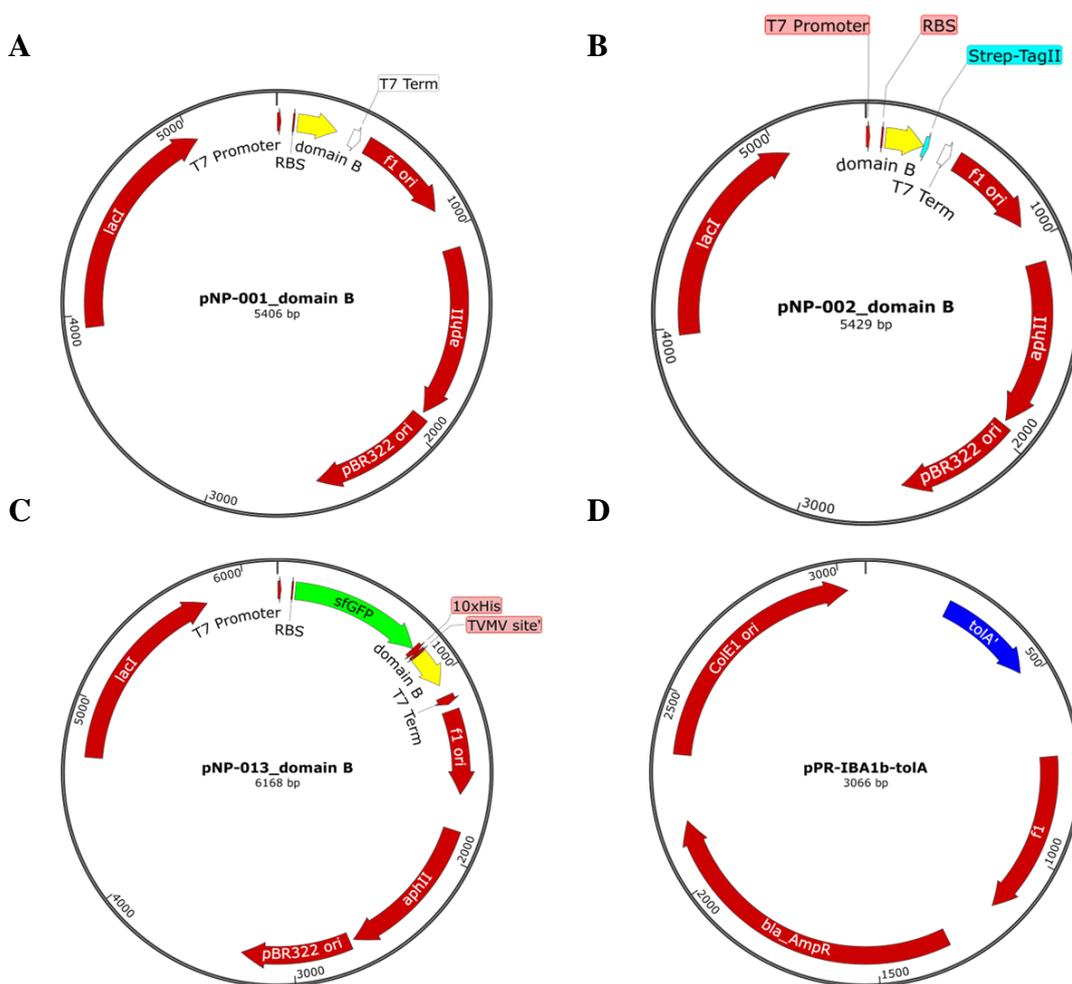
Figure A1: Expression plasmids for the production of recombinant proteins in *E. coli*.
Genes encoding domain B variants were subcloned in the expression plasmids pNP-001 (**A**), pNP-002 (**B**), pNP-013 (**C**) using the restriction enzyme BsaI-HFv2. All plasmids contain a T7*lac* promoter to control gene expression, a pBR322 ori as origin of replication and a kanamycin resistance gene (aphII_KanR). pNP-001 was used to produce tagfree domain B variants and pNP-002 for variants with C-terminal Strep-tag II. Domain B variants with an N-terminal sfGFP followed by a His10-tag and a TVMV cleavage site were produced in pNP-013. The plasmid pPR-IBA1b-tolA (**D**) was used for the amplification of tolA spacers. The plasmid contains a ColE1 ori as origin of replication and an ampicillin resistance gene (bla_AmpR).

Table A 1: Amino acid sequence of modular regions SF, EF and IF1-3 used for library design

Modular section	Wild type domain B	Novel domain B
SF	ADNKF	ADNKF
IF1	NKEQQNAFYEILHL	NKEQQNAFYEILHL
IF2	NEEQRNGFIQSLKDD	EEQRNGFIQSLKDD
IF3	QSANLLAEAKKLNDAQA	SQSANLLAEAKKLNDAQ
EF	PK	PK
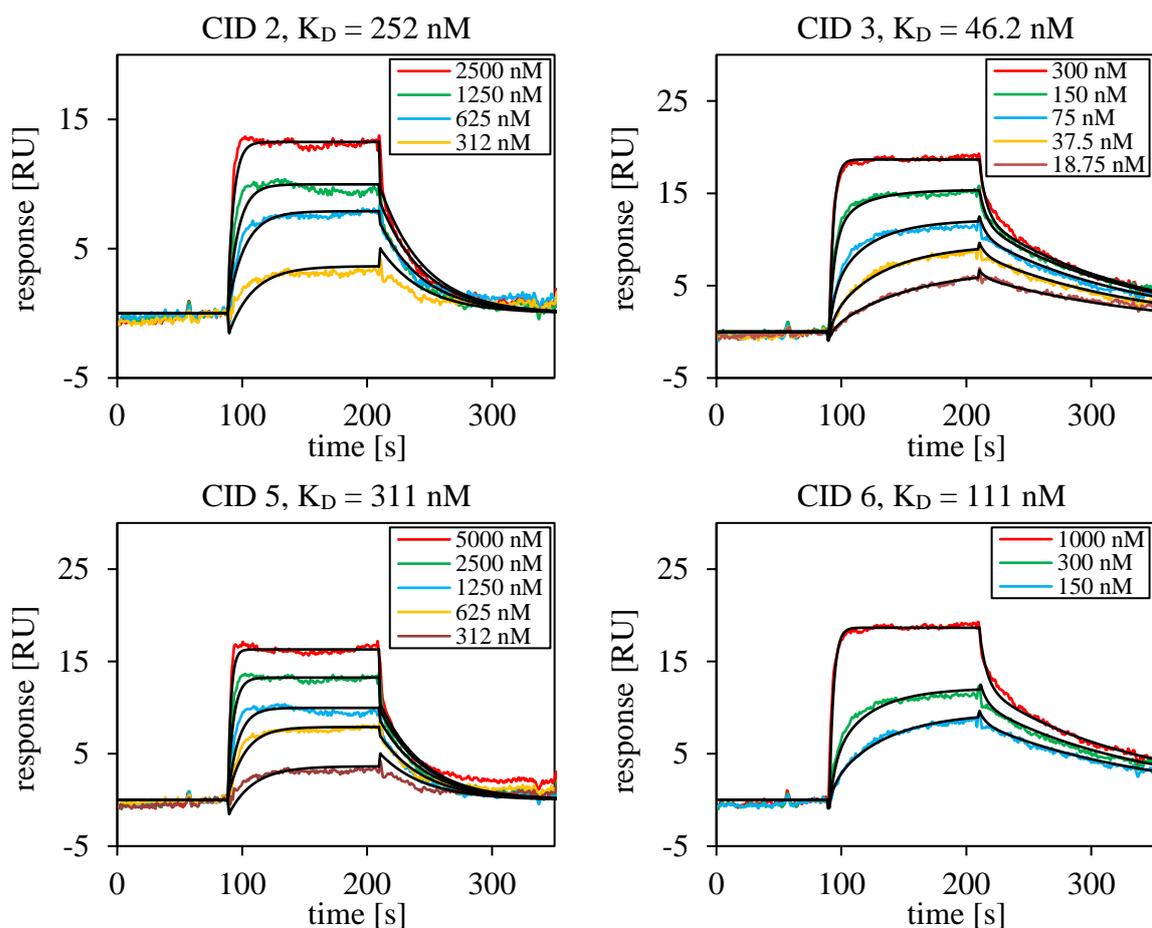
Figure A 2: SPR sensorgrams of CID 2, CID 3, CID 5, and CID 6.
hIgG1-Fc (~850 RU) was immobilized on the surface of a CM5 chip using NHS chemistry. The binding of serial diluted domain B variants to hIgG1-Fc was determined at a flow rate of 30 µL/min. The association and dissociation steps were performed for 120 sec each. 1x PBST 0.05 % (v/v) was used as running buffer. After each cycle, the proteins on the surface were regenerated with 10 mM glycine pH 2.0 for 30 sec. The kinetic parameters k_{on} and k_{off} were determined using a Langmuir fitting model, assuming a 1:1 binding stoichiometry (black lines).
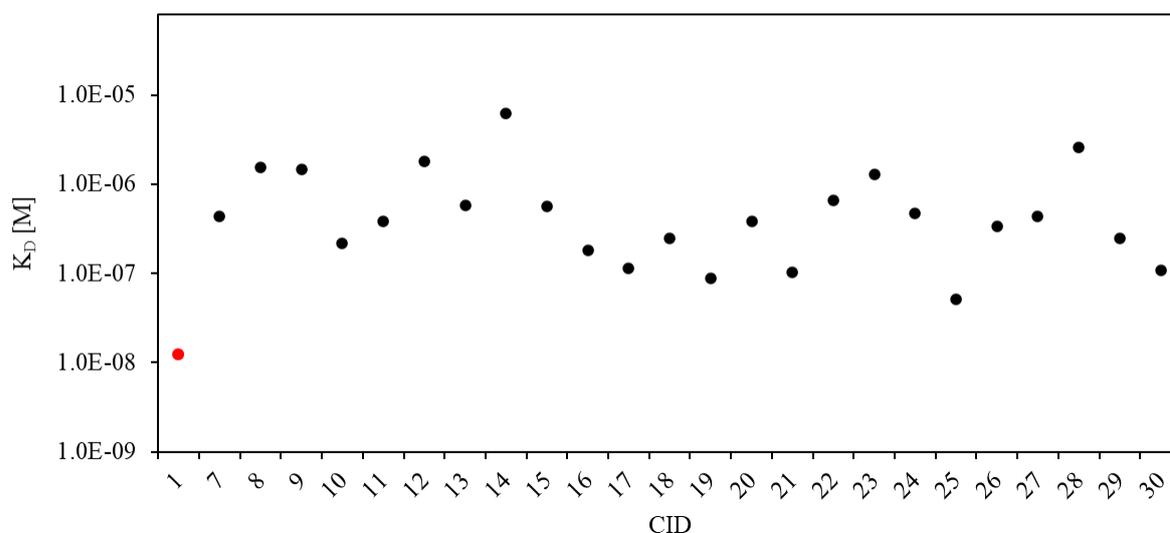
Figure A 3: K_D values of novel IgG-binding domain B variants determined by SPR.
24 novel domain B variants (CID 7-30) and wild type control (CID 1) were produced on µ-scale. hIgG1-Fc (~800 RU) was immobilized on the surface of a CM5 chip using NHS chemistry. The binding of variants [c = 1 µM] to hIgG1-Fc was determined at a flow rate of 30 µL/min. 1x PBST 0.05 % (v/v) was used as running buffer. After each cycle, the proteins on the surface were regenerated with 10 mM glycine pH 2.0 for 30 sec. The K_D values were determined using a Langmuir fitting model, assuming a 1:1 binding stoichiometry.

XXIII
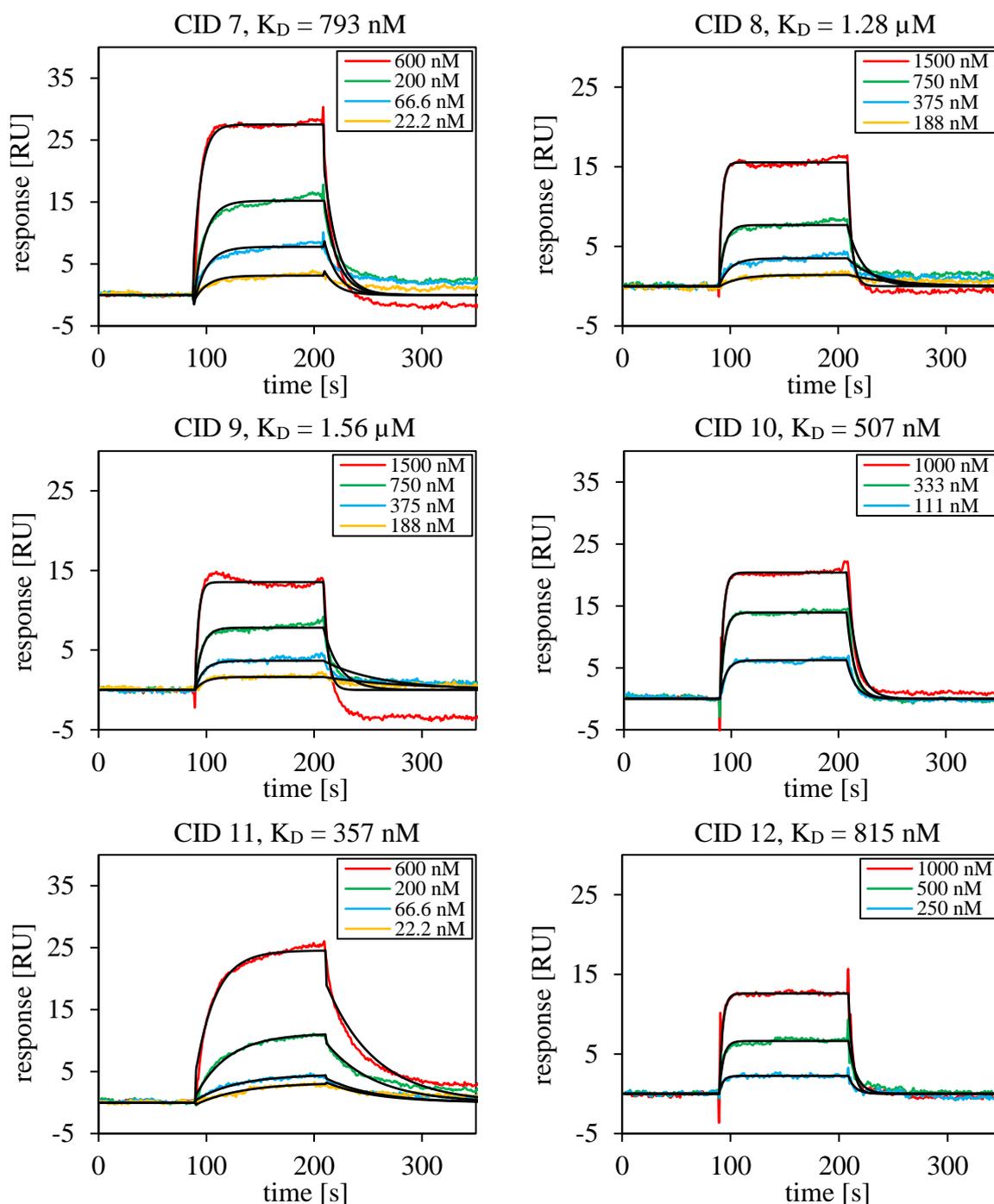
Figure A 4: SPR sensorgrams of novel IgG-binding domain B variants purified by SEC (I).
SPR sensorgrams of six novel domain B variants (CID 7-12) named with individual identification numbers (CID). hIgG1-Fc (~550 RU) was immobilized on the surface of a CM5 chip via NHS chemistry. The binding of serial diluted variants to hIgG1-Fc was determined at a flow rate of 30 μL/min. The association and dissociation steps were performed for 120 sec each. 1x PBST 0.05 % (v/v) was used as running buffer. After each cycle, the proteins on the surface were regenerated with 10 mM glycine pH 2.0 for 30 sec. The kinetic parameters k_{on} and k_{off} were determined using a Langmuir fitting model, assuming a 1:1 binding stoichiometry (black lines).
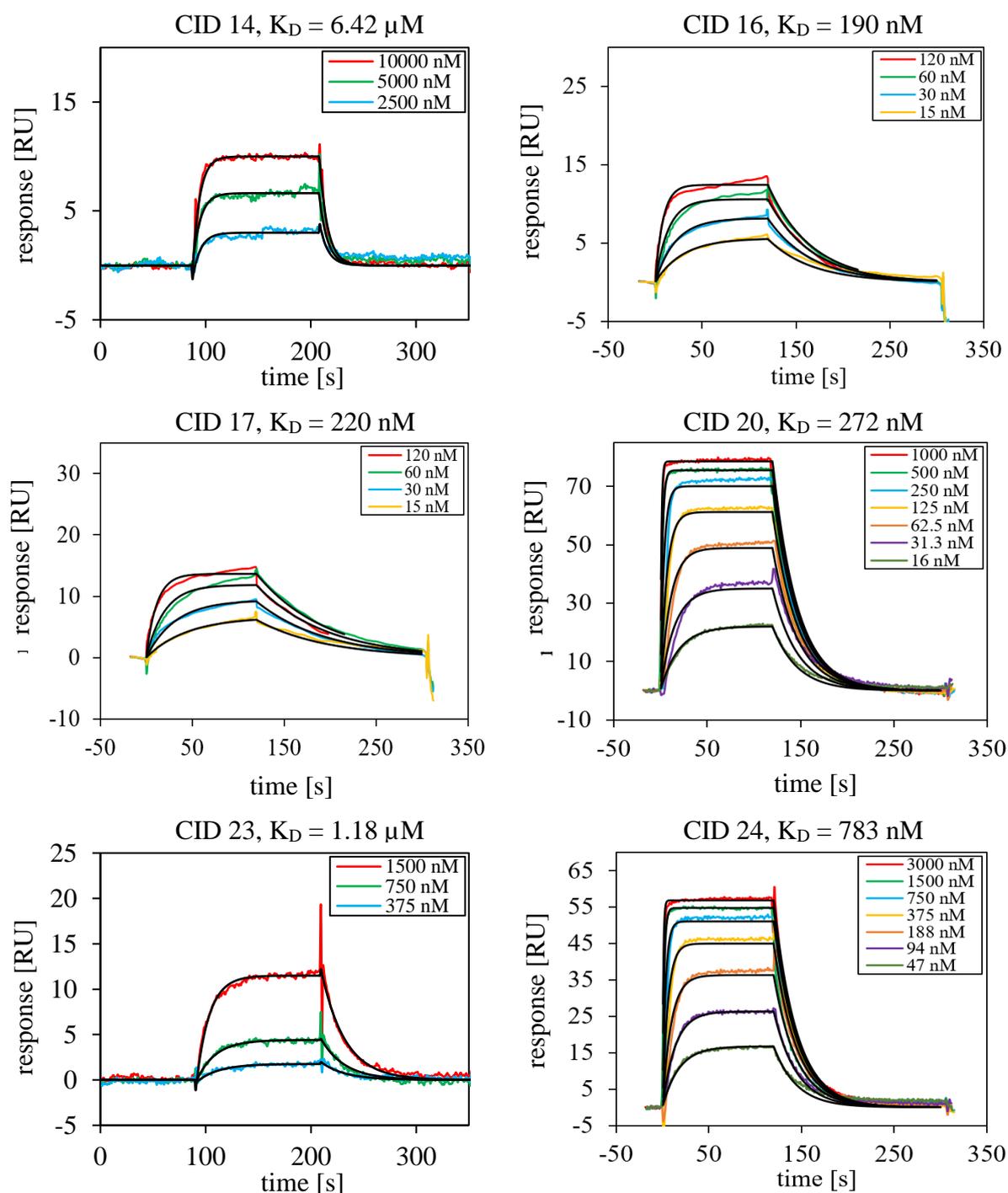
Figure A 5: SPR sensorgrams of novel IgG-binding domain B variants purified by SEC (II).
SPR sensorgrams of six novel domain B variants (CID 14, CID 16-17, CID 20, CID 23, and CID 24) named with individual identification numbers (CID). hIgG1-Fc (~550 RU) was immobilized on the surface of a CM5 chip via NHS chemistry. The binding of serial diluted variants to hIgG1-Fc was determined at a flow rate of 30 µL/min. The association and dissociation steps were performed for 120 sec each. 1x PBST 0.05 % (v/v) was used as running buffer. After each cycle, the proteins on the surface were regenerated with 10 mM glycine pH 2.0 for 30 sec. The kinetic parameters k_{on} and k_{off} were determined using a Langmuir fitting model, assuming a 1:1 binding stoichiometry (black lines).

A

B
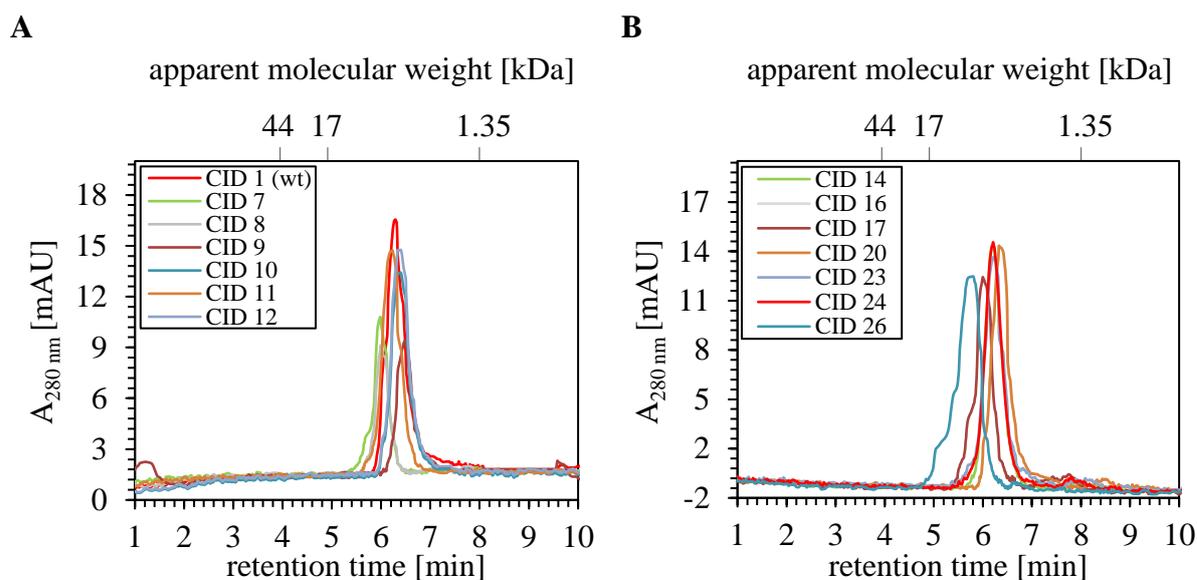
Figure A 6: SE-HPLC of novel IgG-binding domain B variants purified by SEC.
SE-HPLC of 13 novel domain B variants named with individual identification numbers (CID). Wild type domain B (CID 1) was used as comparison. The tagfree variants [c = 0.2 mg/mL] obtained from SEC were analyzed via a Tricorn 5/150 GL Superdex Increase 75 prep grade column (GE Healthcare). The absorbance at 280 nm was detected and visualized in a chromatogram. 1x PBS pH 7.4 was used as running buffer. The column was calibrated with the Bio-Rad Gel Filtration Standard to determine the apparent molecular mass.

Curriculum vitae

Name: Hanna Bobolowski
Date of birth: August 2nd, 1990

Professional career

03/2020 – present	**Navigo Proteins GmbH** Team leader and scientist at the Department of Screening and Protein Engineering
06/2015 – 06/2017	**Navigo Proteins GmbH** Laboratory technician at the Department of Screening and Protein Engineering

Academic career

04/2018 – present	**Martin-Luther-Universität Halle-Wittenberg** Industrial PhD at Navigo Proteins GmbH Dissertation: "Helix Shuffling in Staphylococcal Protein A – Design, Selection and Characterization of Novel Domain B Variants"
10/2012 – 06/2015	**Friedrich-Schiller-Universität Jena** Biochemistry, Master of Science Master thesis: "Regulation of endothelial glucose metabolism by inflammatory cytokines"
10/2009 – 09/2012	**Friedrich-Schiller-Universität Jena** Biology, Bachelor of Science Bachelor thesis: "Klonierung und Expression von Dimer-Konstrukten des *death receptor six* (DR6)"

Patent list

Bobolowski, H., Fiedler, E. and Haupts, U. (2021). Novel immunoglobulin binding polypeptides. International patent publication WO2021/176075A1.

Bosse-Doenecke, E., Gloser-Braeunig, M., Kahl, M., **Bobolowski, H.** and Lotze, J. (2022). Novel human Programmed Death Ligand 1 (PD-L1) specific binding molecules. International patent publication WO2022/223790.

Bosse-Doenecke, E., Gloser-Braeunig, M., Lotze, J., **Bobolowski, H.** and Fiedler, E. (2022). Novel type II collagen binding proteins. International patent publication WO2022096738A1.

Kahl, M., Fiedler, E., Bosse-Doenecke, E., Katzschmann, A., **Bobolowski, H.**, Lotze, J. and Meysing, M. (2021). Binding protein specific for the spike protein of severe acute respiratory syndrome corona virus 2 (SARS-COV-2). International patent publication WO2021/260075A1.

Kahl, M., Fiedler, E., Bosse-Doenecke, E., Katzschmann, A., **Bobolowski, H.**, Lotze, J. and Meysing, M. (2021). Binding protein specific for the spike protein of severe acute respiratory syndrome corona virus 2 (SARS-COV-2). International patent publication WO2021/260052A1.

Knick, P., **Bobolowski, H.**, Kahl, M., Fiedler, E. and Haupts, U. (2022). Immunoglobulin binding proteins for affinity purification. International patent publication WO2022/013272A1.

Statement of independent work

I hereby declare that I have written the present thesis independently, without enlisting any external assistance, and only using the specified aids.

....
Date   Original Signature

Acknowledgement

An dieser Stelle möchte ich mich bei den vielen Menschen bedanken, die diese Arbeit ermöglicht haben.

Mein tiefer Dank gilt zuallererst PD Dr. Hauke Lilie für die universitäre Betreuung dieses sehr spannenden Themas. Mit seiner konstruktiven Kritik und seinem kritischen Blick hat er entscheidenden zum Gelingen dieser Arbeit beigetragen.

Mein besonderer Dank gilt Dr. Erik Fiedler für die exzellente fachliche Betreuung während der gesamten Promotionszeit. Sein Ideenreichtum hat mich stets beeindruckt. Ebenso bedanke ich mich bei der Geschäftsführung der Navigo Proteins GmbH, Dr. Ulrich Haupts und Dr. Henning Afflerbach für die Ermöglichung dieser Arbeit im Unternehmen.

Für den Einblick in die sehr interessante Welt der NMR-Spektroskopie danke ich Dr. Ulrich Weininger. Der Arbeitsgruppe von Dr. Clara Schröder danke ich für die 3D Strukturvorhersage mittels AlphaFold. Außerdem danke ich PD Dr. Iris Thondorf für die Durchführung der Molekulardynamik-Simulationen.

Ein großer Dank geht an alle ehemaligen und derzeitigen Mitarbeiter der Navigo Proteins GmbH, die mich während der letzten Jahre unterstützt haben. Dr. Madlen Zwarg, Dr. Maren Meysing, Dr. Jonathan Lotze und Antje Bodemann danke ich für die Unterstützung bei den Bibliotheksklonierungen und für die Einführung in molekularen Displaymethoden. Bei Dr. Anja Katzschmann und Heike Böcker bedanke ich mich für die Unterstützung in der Proteinreinigung und Analytik. Von Herzen danke ich meinem ehemaligen „Sitznachbarn" Paul Knick für die vielen wissenschaftlichen Diskussionen, seine kritischen Einwände und guten Ideen. Dr. Kerstin Haucke, Dr. David Havlik, Dr. Manja Gloser-Bräunig, Dr. Florian Settele, Vitali Jakobi, Tobias Radtke, Anett Swoboda, Anja Wiech und Marion Herrmann haben mir in unterschiedlichster Form geholfen. Ich danke euch dafür.

Abschließend danke ich meinen guten Freunden und meiner Familie. Stets an meiner Seite war dabei Dominique. Rückblickend erscheint mir ihr Maß an Verständnis doch erstaunlich groß.