Engineering strategies for ABD-derived affinity proteins for therapeutic and diagnostic applications

Mikael Åstrand

Kungliga Tekniska Högskolan, KTH
Royal Institute of Technology
School of Biotechnology
Stockholm 2016
Abstract

Small stable protein domains are attractive scaffolds for engineering affinity proteins due to their high tolerance to mutagenesis without losing structural integrity. The albumin-binding domain is a 5 kDa three-helix bundle derived from the bacterial receptor Protein G with low-nanomolar affinity to albumin. In this thesis, the albumin-binding domain is explored as a scaffold for engineering novel affinity proteins with the possible benefit of combining a prolonged serum half-life with specific targeting in a single small scaffold protein. Previously, a library was created by randomizing surface-exposed residues in order to engineer affinity to a new target antigen in addition to the inherent albumin affinity. Here, phage display selections were separately performed against the tumor antigens ERBB2 and ERBB3. The ERBB3 selection resulted in a panel of candidates that were found to have varying affinities to ERBB3 in the nanomolar range, while still retaining a high affinity to albumin. Further characterization concluded that the clones also competed for binding to ERBB3 with the natural activating ligand Heregulin. The selections against ERBB2 resulted in sub-nanomolar affinities to ERBB2 where the binding site was found to overlap with the antibody Trastuzumab. The binding sites on ABD to albumin and either target were found in both selections to be mutually exclusive, as increased concentrations of albumin reduced the level of binding to ERBB2 or ERBB3. An affinity-matured ERBB2 binder, denoted ADAPT6, which lacked affinity to albumin was evaluated as a radionuclide-labeled imaging tracer for diagnosing ERBB2-positive tumors. Biodistribution studies in mice showed a high renal uptake consistent with affinity proteins in the same size range and the injected ADAPT quickly localized to the implanted tumor. High contrast images could be generated and ERBB2-expressing tissue could be distinguished from normal tissue with high contrast, demonstrating the feasibility of the scaffold for use as diagnostic tool. In a fourth study, affinity maturation strategies using staphylococcal cell-surface display were evaluated by comparing two replicate selections and varying the stringency. A sub-nanomolar target concentration was concluded to be inappropriate for equilibrium selection as the resulting output was highly variable between replicates. In contrast, equilibrium sorting at higher concentrations followed by kinetic-focused off-rate selection resulted in high output overlap between attempts and a clear correlation between affinity and enrichment.

Keywords: ABD, ADAPT, Protein G, ERBB2, ERBB3, Directed evolution, phage display, staphylococcal display, bispecific
Commonly used abbreviations

ABD - Albumin-binding domain
ADAPT - ABD-derived affinity protein
HSA - Human serum albumin
ERBB2/3 - Epidermal growth factor receptor 2/3
DNA - Deoxyribonucleic acid
RNA - Ribonucleic acid
ELISA - Enzyme-linked immunosorbent assay
FACS - Fluorescence-activated cell sorting
PCR - Polymerase chain reaction
SPR - Surface plasmon resonance
CDR - Complementarity-determining region
IgG - Immunoglobulin G
scFv - Single-chain fragment variable
\( F_c \) - Fragment crystallizable
\( F_v \) - Fragment variable
\( F_{ab} \) - Fragment antigen binding
List of publications


* These authors contributed equally

All papers are reproduced with permission from the copyright holders
Respondents contributions to appended papers

I. Designed and planned the experimental procedures together with co-authors. Performed affinity selections and characterization of protein variants together with co-authors. Contributed to interpretation and analysis of data and writing of the manuscript.

II. Designed and planned the experimental procedures together with co-authors. Performed initial phage display selection and protein characterization. Performed competition experiments and FACS analysis of mammalian cells. Contributed to interpretation and analysis of data and writing of the manuscript.

III. Contributed to design and preparation of protein imaging probes. Analyzed and interpreted data together with co-authors. Contributed to writing the manuscript.

IV. Designed and planned the experimental procedures together with co-authors. Performed FACS selections, screening and sequencing together with co-authors. Supervised and performed data analysis and presentation. Wrote the manuscript together with co-authors.
Published work not included in the thesis


Populärvetenskaplig sammanfattning

Proteiner är fascinerande biologiska molekyler som har en avgörande roll för att upp- rätthålla liv i alla dess former. I den genetiska arvsmassan finns ritningar för att skapa alla proteiner som finns i människan där dessa uppfyller ett mångfald av funktioner som krävs för att våra kroppar ska fungera. De sköter transport av viktiga molekyler som t.ex syret i vårt blod, de utgör det strukturella skelettet i våra celler och de är även vårt främsta försvar mot smittsamma sjukdomar, för att bara nämna några funktioner som proteiner utför. Det sistnämnda exemplet involverar antikroppar; stora målsökande proteiner som immunförsvarvet producerar som svar på en attack från en smittsam bakterie eller ett virus.

Man kan likna proteiner vid stora kedjor där varje länk i kedjan utgörs av mindre byggestenar som kallas aminosyror. Dessa förekommer i olika varianter med bred variation av kemiska egenskaper. Eftersom ett protein kan bestå av flera hundra aminosyror som sitter ihop i rad är möjligheterna för att skapa olika proteiner obegränsade. De olika kemiska egenskaperna i aminosyrorerna gör även att proteinet antar olika tre-dimensionella strukturer på grund av att vissa aminosyror inte trivs med att vara i kontakt med vatten; så kallade hydrofoba aminosyror. Proteinkedjorna brukar därför oftastilda stora trassliga nystan där de hydrofoba aminosyrorna hamnar i mitten.


Arbetet i denna avhandling behandlar ett bakteriellt affinitetsprotein som har visat sig binda starkt till ett protein i blodet som kallas albumin. Affinitetsproteinet har fått namn
av denna egenskap och kallas därför den albuminbindande domänen (ABD). Bakterier kan använda ABD för att förklä sig i mänskliga proteiner och därmed undgå detektion av immunförsvarset. Detta bakteriella protein har väckt stort intresse inom medicinska tillämpningar på grund av dess förmåga att stanna under en längre tid i blodet utan att brytas ner eller filtreras ut i njurarna, vilket annars är ett öde som väntar andra proteiner i blodet med liknande storlek som ABD. Det är tack vare bindningen till albumin som detta är möjligt eftersom det finns mekanismer i kroppen som upprätthåller albuminkoncentrationen i blodet. ABD åker därför snålskjuts på albumin och tar del av dess fördelar. ABD har därför använts som en adapter där den har kopplats ihop med andra medicinska proteiner, som då får avnjuta samma fördelar som ABD.

Projekten i denna avhandling handlar om att modifiera ABD på olika sätt. Målet med projekt 1 och 2 var att få ABD i sig själv att binda till målproteiner som finns i stora mängder på ytan av cancerceller men samtidigt ha kvar sin ursprungliga funktion, något som är utmanande på grund av dess lilla storlek. I projekt 3 används en av dessa modifierade ABD-varianter som ett diagnostiskt verktyg i medicinsk avbildning av tumörer inplanterade i möss. För att detta skulle vara möjligt togs den ursprungliga albuminbindningen först bort. I det fjärde och sista projektet utvärderas en metod för att ta fram proteiner med nya egenskaper från så kallade bibliotek; en enorm grupp av olika proteinindivider där varje molekyl har olika modifieringar. Dessa metoder används flitigt inom framtagning av nya proteinläkemedel och därför behövs pålitliga verktyg för att man ska försäkra sig om att man utvecklar proteiner med rätt egenskaper.
# Contents

Abstract i

List of publications iii

Respondents contributions to appended papers iv

Published work not included in the thesis v

Populärvetenskaplig sammanfattning vi

1 Affinity proteins 1

1.1 Introduction ........................................... 1

1.2 Molecular recognition .................................. 2

1.3 Antibody technology ................................... 5

1.4 Alternatives to antibodies ............................. 7

   Antibody fragments ...................................... 7
   Affibody molecules ...................................... 8
   Designed ankyrin repeat proteins ....................... 10
   Aptamers ................................................... 11
   Knottins .................................................... 12
## 2 Protein engineering

2.1 Engineering approaches ........................................... 14
   Rational design .................................................. 14
   Random mutagenesis ............................................. 15
   Combinatorial protein engineering ............................. 16

2.2 Directed evolution ................................................ 18
   Phage display ..................................................... 18
   Cell-surface display ............................................. 21
     Yeast display .................................................... 22
     Staphylococcal display ...................................... 24
   Cell-independent systems ................................... 25
     Ribosome display .............................................. 25
     CIS display .................................................... 26

## 3 Engineered proteins in clinical applications  

3.1 Affinity proteins as therapeutic and diagnostic agents .......... 29
   Therapeutic modalities for affinity proteins .................... 31

3.2 *In vivo* half-life extension strategies ....................... 33
   Increasing the hydrodynamic radius ............................ 33
   Exploiting the FcRn rescue pathway ........................... 34

3.3 The albumin-binding domain ................................ 36

## 4 Present investigation  

4.1 Investigation aim ............................................... 41

4.2 Development and characterization of small bispecific albumin-binding domains with high affinity for ERBB3 (Project I) ........ 42

4.3 Engineering of bispecific affinity proteins with high affinity for ERBB2 and adaptable binding to albumin (Paper II) ............ 45
4.4 ADAPT, a novel scaffold protein-based probe for radionuclide imaging of molecular targets that are expressed in disseminated cancers (Project III) 49

4.5 Investigating affinity-maturation strategies and reproducibility of fluorescence-activated cell sorting using a recombinant ADAPT library displayed on staphylococci (Paper IV) .................................................. 52

4.6 Concluding remarks and future outlook ........................................... 55

Bibliography ..................................................... 59
Chapter 1

Affinity proteins

1.1 Introduction

Proteins are macro-molecular polymers built up by amino acids. The most basic forms of proteins are short linear arrangements of amino acids that are more commonly referred to as peptides. The amino acid chain can be further arranged in a specific three-dimensional orientation, which is generally required for the protein to exert its function. There are 20 standard amino acids, appearing in proteins observed in all living organisms on earth, with vastly different chemical properties. Common among all amino acids is an amine head group (-NH$_2$) and a carboxyl end group (-COOH), allowing for the formation of peptide bonds between the amino acids connecting the carboxyl end of one amino acid to the amine end of another. This linear arrangement is known as the primary structure (Figure 1.1A). Theoretically, any number of any amino acid can be arranged in any sequence, which leads to near infinite possible protein arrangements.

The primary amino acid chain is then further ordered in local arrangements forming what is known as secondary structure, e.g. $\alpha$-helices and $\beta$-sheets, which can both be seen in the small 56-amino acid protein domain C2 (Figure 1.1B). Those secondary arrangements are generally also structured in higher order configurations known as tertiary structure (Figure 1.1C), giving rise to a stable protein domain. Further higher order arrangements among individually folded domains are also formed in very large proteins. The proteins quaternary structure describes the relationship and ordering among these subunit domains and how they interact with each other. The natural state of a protein,
1.2 Molecular recognition

Affinity proteins is a class of proteins that are able to reversibly bind other molecules. The most typical type of affinity proteins are antibodies that play a key role in the immune system where they specifically recognize and bind foreign antigens. This ability to bind other molecules is known as affinity. Any compound $A$ binding reversibly to compound $B$ forming the complex $AB$ (Equation 1.1) is governed by the law of mass action. The association equilibrium constant $K_a$ determines the ratio between the con-
centration of the formed complex and the free binding partners when equilibrium has been established (Equation 1.2).

\[ A + B \rightleftharpoons AB \tag{1.1} \]

\[ K_a = \frac{[AB]}{[A][B]} \tag{1.2} \]

The equilibrium constant is also defined as the ratio between the rate of formation \( k_a \left( M^{-1}s^{-1} \right) \) and the rate of dissociation \( k_d \left( s^{-1} \right) \). Typically, when affinity proteins are discussed, the affinity of an interaction is defined as the dissociation equilibrium constant \( K_d \) (Equation 1.3), which is measured in molar (M).

\[ K_d = \frac{k_d}{k_a} = \frac{1}{K_a} \tag{1.3} \]

This means that the relationship between the dissociation rate constant (\( k_d \), also known as “off rate”) and the association rate constant (\( k_a \), also known as “on rate”) determines the affinity. However, the kinetics can be dramatically different for two interactions measuring the same affinity. Figure 1.2 illustrates this by showing the theoretical interactions between two binding pairs having the same affinity but different kinetics. It is therefore important to distinguish between these terms when discussing the behavior of affinity proteins. First-order kinetics gives us that the formation of \( AB \) follows equation 1.4.

\[ \frac{d[AB]}{dt} = [A][B]k_a - [AB]k_d = [A]k_a([B]_{max} - [AB]) - [AB]k_d \tag{1.4} \]

At equilibrium \( \frac{d[AB]}{dt} = 0 \), and we can acquire an expression describing the amount of complex formed at equilibrium where \([B]_{max}\) is constant and represents the maximum number of binding sites for \( A \) (Equation 1.5).

\[ [AB] = \frac{[A][B]_{max}}{[A] + K_d} \tag{1.5} \]

Equation 1.4 also show that both association and dissociation govern the rate of complex formation on the way to equilibrium.
As such, affinity is clearly defined and it can be measured and even monitored in real time using several assays such as surface plasmon resonance (SPR) and quartz crystal microbalance [2,3]. Ultimately, asking why and how a protein interaction occurs rather than if it occurs can be a more fruitful endeavor as it can offer insights into general aspects in protein behavior and determinants of high-affinity interactions, which in turn helps guide the design and engineering of affinity proteins. The amino acids that directly take part in the binding interface of a protein-protein interaction are the most likely culprits to examine but there might also be secondary effects in play such as changes in conformation upon binding.

The 20 standard amino acids have varying properties regarding hydrophobicity, charge and what types of bonds they can form with each other. Hydrogen bonds, occurring between electronegative atoms (oxygen and nitrogen) and hydrogen atoms in close proximity, are strong determinants of binding strength between proteins. Negatively and positively charged amino acids also act as an attractive force, as do hydrophobic interactions between surfaces containing non-polar amino acids. The binding equilibrium between two proteins is ultimately affected by the total energetic contributions involved in binding, which can be expressed as the change in Gibb’s free energy upon binding (Eq. 1.6).

\[ \Delta G = \Delta H - T \Delta S \]  

The energetic contributions are thus categorized in an enthalpic term (\( \Delta H \)) and a
temperature-dependent entropic term \((T\Delta S)\). In simplistic terms, the enthalpy \((H)\) describes the total amount of energy contained in a system, whereas the entropy \((S)\) governs spacial freedoms of movement and disorder. The formation of a protein complex thus result in a reduction in entropy due to the loss of conformational freedom and increased structural rigidity. This unfavorable entropy loss is overcome by the enthalpy loss during the formation of bonds in the interface between the interacting proteins. The change in Gibb’s free energy upon binding is directly related to the equilibrium constant, and thus affinity, by the relation described in equation 1.7 where \(R\) is the ideal gas constant, \(T\) is temperature and \(\ln K\) is the natural logarithm of the equilibrium constant.

\[
\Delta G = -RT\ln K \tag{1.7}
\]

Equation 1.4 above describes a rather simple but general case where only two molecules interact with one another, which is also known as a monovalent interaction. There are many examples of multivalent affinity proteins with two or more binding sites that may or may not be identical. Models exist that describe these interactions [4] but they are significantly more complicated than the widely applicable monovalent model described above. The contribution of several binding sites on the overall binding strength is known as the avidity effect.

### 1.3 Antibody technology

The most well-known example of affinity proteins are antibodies that play a prominent role in the immune system. There they are responsible for binding to foreign molecules and cells in order to activate other key members of the immune system, such as T cells and NK cells [5, 6]. Among our genes there is a large immune repertoire of variable genes encoding the binding domains of antibodies. Antibodies, or immunoglobulins, is the collective term for a number of isotypes that are produced by different cells and ultimately play different roles in the defense of the body. Common among antibodies is a Y-like shape (Figure 1.3) where the binding of antigens take place at the ends of the arms. These regions are therefore variable and selected based on their ability to bind a target antigen. The rest of the molecule is constant and play various roles in signaling by binding to immune cells and taking part in the activation of the complement cascade [7]. The IgG isotype is the most common antibody during the later stages of the immune
response after affinity maturation has taken place. As a result, they are capable of the highest affinities to the antigen, typically in the low nanomolar range. Similar to the other isotypes, they consist of two types of polypeptide chains, called the heavy and light chain that, after expression, associate into homodimers that are held together by disulphide bridges [8]. An antibody has six unstructured loops on each antigen-binding arm that interact with target proteins in a highly specific manner, making multiple contacts with amino acids located in the target binding sites. These loops are called complementarity-determining regions (CDR). For IgG, the total molecular weight is in the range of 150 kDa (Figure 1.3). The remarkable ability of antibodies to be capable of binding to virtually any target has raised the interest of researchers around the world. Because of this property, antibodies are commonly used as tools in all fields of biological research. However, the field where the most interest can be found is within medical applications as diagnostic tools and therapeutics. Novel antibodies are continuously being developed and improved to combat a multitude of widespread diseases where more conventional approaches fail, such as cancer or inflammatory disorders.

The genetic repertoire within immunoglobulin genes is formed through recombination of exons that encode for the variable domains at the ends of the antibody arms. These are termed the variable light ($v_L$) and the variable heavy ($v_H$) for the light and heavy
chains, respectively. Antibodies can be raised against novel targets in many ways, the earliest of which was through immunization [9]. By introducing an antigen into an animal, typically by injection, antibodies will naturally be produced by the immune system and a polyclonal pool of antibodies can then be purified from the blood. However, using such an antibody preparation for disease treatment in humans is problematic due to several reasons. First, the antibody framework of the animal in question will not be identical to human, which will result in an immune response to the introduced antibodies in the patient. Second, purifying antibodies from animal sera is challenging since many contaminants such as other host protein or viruses can be difficult to remove. Furthermore, polyclonal antibodies acquired from animals stem from a heterogeneous mixture of antibody-producing cells with variations in the immunoglobulin genes. This means that they will often bind several epitopes and have varying affinity to the antigen. Polyclonal antibody preparations may also suffer from batch-to-batch variations due in part to the immunization of different individual animals [10]. In addition, it is possible that only a small fraction of all isolated antibodies from an immunized animal shows activity toward the antigen in question. This lowers the reliability of these antibodies and therefore limits their potential usefulness in therapy. The use of monoclonal antibodies instead, i.e. a single clone with a specific amino acid sequence, circumvents these problems. However, producing monoclonal antibodies is a much more expensive and challenging endeavor since single clones need to be identified, analyzed and isolated. Antibodies can then be produced using for example hybridoma technology or recombinant production in mammalian cell lines [11,12].

### 1.4 Alternatives to antibodies

While antibodies are highly useful as affinity reagents, there are many examples of proteins in nature with the capability of taking part in high affinity interactions with other proteins. The antibody structure is therefore not a requirement in order to facilitate a strong binding interface. Indeed, many different proteins with varying sizes and structure are being explored for engineering new affinity proteins.

**Antibody fragments**

A full-length antibody has a considerable size (150 kDa for IgG), but it is only the variable fragments of the $F_{ab}$ arms that interact with the target antigen. If binding is
the only desirable feature of the molecule, then smaller versions of antibodies should be feasible. Indeed, several different antibody formats have been developed that retain the strong and specific interaction of its full-length counterpart, albeit with loss of the avidity effect due to their monovalency, but with a significant reduction in size [13–15]. The \( F_{ab} \) arm can be isolated alone, consisting of the full-length light chain and the corresponding antigen-binding portion of the heavy chain, together weighing in at roughly 50 kDa in size. Even smaller fragments are also in use including the single-chain variable fragment (scFv, \( \sim 28 \) kDa) and the heavy-chain variable fragment (\( v_H \), \( \sim 15 \) kDa), also known as a single-domain antibody [16]. The former is a tethering of the \( v_H \) and \( v_L \) by a peptide linker that ensures the formation of the natural heterodimer between the \( v_H \) and \( v_L \) by hydrophobic contacts. Notably, the linker length between the two fragments can be varied, which results in the formation of different topological structures with increasing valency the shorter the linker is. These have been termed diabodies (dimeric), triabodies (trimeric) and tetrabodies (tetravalent), respectively [15,17,18] (Figure 1.3).

**Affibody molecules**

Protein A from *Staphylococcus aureus* has been extensively characterized and has raised interest due to its function as an immune escape mediator for increased virulence [19], but also for its now widespread use as a purification handle in affinity chromatography for antibodies. This 42 kDa protein consists of a C-terminal membrane anchoring region and an N-terminal portion with five domains of roughly 6 kDa each with high sequence homology that all show remarkable affinity for immunoglobulins [20–23]. Together, they provide recognition of antibodies of a wide range of classes from multiple species [24]. It was early concluded that Protein A binds exclusively to the constant domain of the heavy chain of antibodies [25,26]. However, further experimentation showed that all individual domains of Protein A also have affinity for the \( F_{ab} \) fragment [27], which could further explain the versatility of the molecule as a virulence factor and its involvement in immunomodulatory functions [28].

The B domain of Protein A has attracted the most attention. The crystal structure of the B domain in complex with the \( F_c \) portions of IgG has been determined, revealing that helix 1 and 2 of domain B interacts with \( F_c \) at the hinge region between \( CH_2 \) and \( CH_3 \) [26]. The B domain was chosen for mutagenesis into a synthetic domain, denoted the Z domain, lacking the methionine present in domains E, D and A as well as the conserved aspargine-glycine peptide in helix 2 [29].
The Z domain has since been used not only as means of purifying antibodies, but has also been engineered to create novel affinity reagents, denoted Affibody molecules, by replacing the affinity to IgG with binding affinity to a plethora of protein targets including ERBB2, ERBB3, IL-6, TNF-α and the amyloid-β peptide to name only a few [30–32]. This has generally been accomplished by thorough mutation of 13 solvent-exposed residues in helix 1 and 2 [33] (Figure 1.4A). The mutated Z domains are able to retain the three-helical fold of the parent protein. A notable exception is the group of Affibody variants binding the amyloid-β peptide that sequesters the target peptide in a 2:1 stoichiometry. This is achieved by the formation of a β-sheet involving amino acids in what in the parent Z domain is helix 1 [34,35].

Affibody molecules are currently being pursued as both therapeutic and diagnostic agents with several promising candidates in development for a number of diseases and conditions. An ERBB2-binding Affibody molecule, named ABY-025 with picomolar affinity, has recently undergone clinical trials as a radionuclide imaging tracer [36]. The study showed that the Affibody tracers were well tolerated in humans and accurately detected the presence of ERBB2-expressing metastases.

![Figure 1.4: Structural models of (A) an Affibody, (B) a DARPin, (C) an Aptamer in a complex with thrombin (pink) and (D) a Knottin. Disulphide bridges are indicated in yellow. The models were generated from PDB files 1fcc, 4ydw, 3dd2 and 4gux, respectively.](image-url)
Designed ankyrin repeat proteins

Another example of an engineered non-immunoglobulin protein are binders based on the ankyrin repeat motif, a 33 amino acid peptide present in thousands of proteins of eukaryotic and prokaryotic origin [37]. The ankyrin repeat consists of a β-turn followed by two anti-parallel α-helices. Much like other repeat motifs found in nature, e.g the leucine-rich repeat or the armadillo repeat, the proteins in which the repeats are found do not necessarily share the same functional properties but they do share striking structural similarities. This implies that the repeating units form a structurally conserved scaffold from which protein-protein interactions can be evolved. By comparing an abundant set of ankyrin repeat sequences from the PFAM database ¹, Mosavi et al. created consensus designs with up to four repeats in an attempt to optimally represent the structural unit [38]. The stability and fold of the resulting proteins were found to be highly dependent on the number of repeats since one or two repeats failed to produce a stable protein, whereas three or four repeats yielded circular dichroism spectra consistent with high α-helical content. It was also shown that four repeats resulted in the highest thermal stability.

Binz and colleagues used a similar idea where they also designed a consensus sequence of a repeating unit that could be expressed as a tandem protein of variable size. Surface-exposed residues on the α-helices could then be randomized in defined positions in order to engineer new protein-protein interactions from the resulting libraries [39]. Flanking N- and C-terminal repeats were included in this study, in contrast to the former, which contributed to the increased stability and proper folding of the proteins. These repeats, referred to as capping repeats, were left constant and corresponded to N- and C-terminal repeats found in natural proteins. Since these repeats are more solvent-exposed than the internal ones, the distribution of charged and hydrophobic residues in the capping repeats is drastically different in order to facilitate proper packing of hydrophobic patches. These resulting designed ankyrin repeat proteins (abbreviated DARPINs, figure 1.4B) have since been used as a general scaffold for affinity protein engineering, yielding binders with very high affinities, in some cases reaching the picomolar range [40–42]. Much like in the study by Mosavi et al. [38], adding more internal repeats have been shown to increase the stability of the molecule, almost rendering it resistant to thermal and chemical denaturing up to 100°C in guanidine-hydrochloride [43].

Initially, six surface-exposed residues were mutated in each repeat in order to introduce diversity [42]. More variability can be added by adding repeats, which not only increases

¹http://pfam.xfam.org/
the size of the protein, but the size of the potential binding interface as well. Novel improvements of the DARPin scaffold has been made as of late, including a redesign of the scaffold to reduce the overall hydrophobicity of the scaffold [44] in the hopes of reducing the occurrence of non-specific interactions during the selection of binders from the DARPin libraries. More recently, changes have been made to the DARPin scaffold where the loops connecting the repeats were elongated and diversified, yielding a new scaffold denoted loopDARPin [45]. This intriguing design approach more resembles the CDR loops of an antibody variable domain and possibly enables the isolation of binders to different epitopes than conventional DARPin.

Aptamers

Not only protein-based scaffolds have been developed to compete with antibody-derived affinity proteins. While DARPin and affibodies are good examples of well-established alternative protein scaffolds, there are more possibilities that need mentioning in the context of developing molecules with the ability to specifically recognize a target. A notable example are Aptamers, which are DNA or RNA-based single-stranded oligonucleotides. As opposed to double-stranded DNA, single-stranded DNA or RNA can form various three-dimensional structures depending on the nucleotide sequence [46], quite reminiscent of how proteins fold. The work on Aptamers started in the early nineties when it was shown that RNA sequences with the right three-dimensional fold to bind to a protein target can be isolated from large pools of RNA with varying sequences [47]. A concern early on was the high susceptibility of DNA and RNA to endonucleases (DNAse and RNAse, respectively). This issue was however partially resolved by the addition of a reversed 3’ thymidine that binds with high affinity to the 3’ end of the aptamer. This modification was shown to strongly inhibit the activity of 3’-5’ exonucleases [48].

Crystal structures of several types of aptamers in complex with their binding partner have been solved, revealing that the aptamer indeed folds into a specific conformation [49,50] (Figure 1.4C). In the study by Long et al. from 2008 [49], the structure shows that amino acid side chains of the target thrombin interacts with the annealed bases in the aptamer hairpin. Since RNA and DNA naturally only contains four possible nucleotides; adenosine, thymine/uracil, cytosine and guanine, the total chemical diversity that can be achieved in a random pool of oligonucleotides is somewhat limited, at least in comparison to proteins where 20 amino acids can be encoded by the standard genetic code. This limitation has been thwarted to an extent through DNA synthesis in which modified bases can be included in order to expand the range of chemical properties, enabling new
possibilities for high-affinity interactions [51].

Several aptamers are currently being investigated in clinical trials for the treatment of several disorders affecting the eye. Since the presence of exonucleases is significantly lower in the eye vitreous than in blood serum, this is a suitable area in which treatment with aptamers is highly suitable. Indeed, one aptamer (commercial name “Macugen”) targeting the vascular endothelial growth factor (VEGF) has already been approved for the treatment of ocular vascular disease and other promising leads are being investigated. [52,53].

**Knottins**

The scaffolds described above, including aptamers, serve as a few examples of successful engineering of biomacromolecules for use in *in vivo* applications. At one end of the spectrum we have affibodies and DARPins representing proteins completely lacking disulphides and post-translational modifications, enabling production in bacteria or through solid-phase peptide synthesis in the case of affibodies. At the other end we have non protein-based affinity reagents such as aptamers that are reliant on *in vitro* DNA and RNA synthesis. The middle-ground between those two extremes can be exemplified by Knottins; a polypeptide scaffold held together by internal disulphide bridges, which makes them unsuitable for stable expression in the reducing environment of the *E. coli* cytoplasm, instead relying on transport to the periplasmic space before folding can take place [54]. However, they make up for this limitation by being the smallest alternative scaffold protein to date with only 30 amino acids for the smallest variants [55], and showing impressive structural stability owing to the disulphide bridges [56]. The structural motif of the knottin fold is characterized by the presence of a $3_{10}$ helix as well as a an anti-parallel $\beta$-sheet composed of three strands (Figure 1.4D). The tight conformation of the molecule is held together by the constraints imposed by the disulphide bonds where the back-bone chain of the peptide is oriented as a knot around itself, hence the name of the scaffold. Knottins are derived from a family of proteins known as squash inhibitors [57]. The 29 amino acid trypsin inhibitor II from *Echallium elaterium* was first explored for inserting additional amino acids in the loops of the scaffold including a 17 residue epitope for Sendai virus L protein. Knottins have since been engineered to bind various targets and show promise in several clinical applications [55].
CHAPTER 2. PROTEIN ENGINEERING

Chapter 2

Protein engineering

The engineering of proteins, i.e. artificial protein modification, is a large field encompassing multiple areas of industry and research. It has developed during the course of the past decades as recombinant DNA technologies emerged, including the groundbreaking discovery of the polymerase chain-reaction (PCR) [58], and the understanding that the evolutionary change of genes is guided by the positive selection of favorable phenotypes over generations [59]. Since genes encode proteins, recombinant DNA technology also enabled facile expression in multiple host systems, e.g. *E. coli*; the most frequently used host system in biotechnology [60]. By mutating the genes encoding the protein primary structure we can directly modify the function and/or behavior of the protein. Due to the complexity of proteins as described above, deducing the relationship between amino acid composition, i.e. the only information that can easily be extracted from the genetic sequence, and function is very difficult. Therefore, more information is needed about what properties of the protein are essential for function and how it can be improved. For example, knowledge about the three-dimensional structure of the protein is highly useful in designing engineering strategies. Such an approach is known as rational protein design. For example, detailed knowledge about the catalytic mechanism of an enzyme’s active site and the structure of the substrate transition state allows for tailoring of the amino acids interacting with the substrate, resulting in altered or even novel specificities [61]. The amount of knowledge about the general dynamics of protein folding and structure has increased greatly over the years, allowing researchers to even design proteins *de novo* [62].

Rational approaches are often not practical or even possible when large complex proteins
are involved whose functions and interactions are not fully understood. Instead, large pools of protein mutants, so called protein libraries, can be screened for activity using some method that reflects the desired function appropriately. This requires that each protein mutant can be physically separated and identifiable in order to attribute the amino acid sequence to the observed trait. To overcome this, many techniques have been developed where the gene encoding a protein, the genotype, is physically linked to the expressed protein, the phenotype. This effectively labels each protein with its unique DNA sequence and removes the necessity to individually screen proteins, provided that we can separate the proteins based on a function, e.g. affinity. Many techniques are available to create libraries of varying sizes where the introduced mutations range from completely random to more focused on certain residues or regions of the protein. This approach also allows for combinatorial protein engineering where several mutations are combined in order to find variants where unexpected interactions and synergies occur.

This chapter will highlight the key differences between rational protein engineering and the creation and screening of protein diversity for the discovery of improved proteins; a principle known generally as directed evolution.

2.1 Engineering approaches

Rational design

The key to engineering a protein using either a limited number of modifications to the amino acid sequence, or indeed developing a protein de novo, is being able to predict the outcome of a given modification. The field of rational protein engineering has been greatly accelerated by the introduction of X-ray crystallography and nuclear magnetic resonance (NMR), which has resulted in accurate, high-resolution structural models of various proteins. From this came also the ability to solve the structure of a wide range of proteins in complex with other affinity proteins, such as antibodies, or substrates in the case of enzymes. The effect of mutations can then be studied in detail, not only in the protein itself, but the effect it has on binding activity as well.

In enzyme engineering, information about the catalyzed reaction, and in particular information about the transition state of the substrate, is needed in order to design active sites that accurately accommodates the intended chemical reaction. As an example, being able to selectively catalyze a reaction into a specific enantiomer of a chiral product is invaluable in industrial production of chemical substances [61]. More often than not,
however, the structure of a protein is not available and can be laborious to solve due to difficulties with recombinant expression, stability or solubility if the protein contains membrane-bound components [63]. Efforts to circumvent the limitation of structure determination is dominated by computational prediction of protein structures. The ability to accurately predict the secondary and tertiary structure from the amino-acid sequence of a protein still remains one of the holy grails in protein science where software such as the Rosetta structural prediction algorithm is making large contributions [64]. Recent efforts in computational structure prediction has lead to the design of proteins completely de novo by tailoring a peptide to fit a binding pocket or making a peptide adopt a specified fold [62,65,66]. This approach has also successfully been combined with directed evolution methods to create powerful platforms for the discovery of novel affinity reagents [67,68].

Rational protein engineering can be the favoured approach in certain applications. It can for example be difficult to discover antibodies targeting certain epitopes that are disordered or exhibiting low immunogenicity, thus favoring the generation of antibodies to other epitopes. In such cases, rational approaches can be employed where peptides with known binding or aggregation activity are grafted in place of the CDR regions on an antibody [69,70]. Similarly, this approach can also be employed for vaccine design to allow for the generation of more immunogenic antigens in cases such as influenza where the immune system fail to sufficiently generate antibodies [71–73].

Random mutagenesis

The other end of the design approach spectrum is screening mutants of a protein that have been randomly generated. This is a straight-forward method of creating protein diversity as it does not require any detailed knowledge about the function or structure of the protein, in stark contrast to rational protein design. The drawback of such an approach is that many, if not a majority, of the random mutations will influence the activity of the protein negatively. The choice of mutagenesis method influences what type and amount of mutations are generated, which can be crucial depending on what downstream application is used to screen the prospective mutants since the size of the mutant library strongly affects what methods can be used to screen it [74]. A random approach is appealing as it more closely resembles the way proteins have naturally evolved and it can be applied not only to single genes, but to whole genomes as well when more broadly acting mutational methods are used such as UV-irradiation and mutator strains [75,76]. Mutations can also be induced chemically in order to mimic the
action of somatic hypermutation of B-cells when antibodies undergo affinity maturation in the later stages of antigen response [77, 78].

One of the more common methods of introducing random mutations, error-prone polymerase chain reaction (epPCR), relies on increasing the error rate of DNA polymerases by the addition of magnesium and manganese ions to the PCR mixture [79–81]. The use of modified dNTPs is also a popular choice as this can lead to even higher mutational rates more frequent than $10^{-2}$ mutations per base. [82]. EpPCR libraries are also inherently biased because some types of mutations will be more frequent than others. Due to the way the genetic code functions, more than one mutation will sometimes be required to change a certain codon to another, resulting in a bias towards amino acids only requiring a single point mutation. Nonetheless, epPCR is an attractive approach as it offers a simple way to generate diversity, either focused on a single gene or on a whole plasmid [83], that can be rapidly screened by any of the directed evolution strategies outlined below.

**Combinatorial protein engineering**

If very little information is available about the target molecule, or indeed cell type, it is not feasible to rationally design a molecule with which the target should interact in a specific way. It is therefore generally more practical to screen large amounts of candidate libraries where each member might have similar structural skeletons with a number of possible active groups as substitutions. All possible permutations of the molecule must then be tested for function, which requires that (i) the chosen assay is relevant for the intended application and that (ii) the data acquired from such an assay is reliable and accurate. Proteins, on the other hand, can be expressed in cells and require only the genetic sequence encoding that protein. Chemical diversity can be created by varying the amino acid sequence of the protein and the 20 amino acids available in the genetic code yields a vast array of possible variants even for short peptides. Indeed, even a 5-mer peptide where every position is mutated to every possible amino acid results in a total variant space of $20^5 = 3.2 \cdot 10^6$.

Similarly, vast libraries of protein mutants can be created by varying the gene sequences encoding the protein. 64 different codons are possible encoding 20 amino acids and 3 stop codons. A common way of introducing diversity into a protein is therefore to synthesize the encoding DNA with a mix of nucleotides for specific positions. NNN denotes a variable codon where every position can be any nucleotide. This yields complete
coverage of the entire genetic code but will naturally contain biases since the genetic code is redundant, with multiple codons encoding the same amino acid. There are many reasons why this is not always a good way to create a library. A high degree of degeneracy in the library means that the size of the library is vast on the genetic level, but much smaller on the protein level. In addition, codons are used with varying frequency in different organisms, which means that many codons will not be optimized for the intended expression host leading to decreased levels of expression and a biased diversity [60,84]. Other ways to randomize codons have therefore become more common to use, such as NNB, NNS or NNK randomization (where N=G,T,C,G; B=C,G,T; S=C,G and K=G,T), which includes 48, 32 and 32 codons, respectively [85]. Not only do these randomization strategies lower the bias to certain amino acids, but they also lower the probability of forming any of the three stop codons Ochre (TAA), Amber (TAG) or Opal (TGA), the prevalence of which would be detrimental to any protein engineering experiment. Another alternative is to remove all redundancy and bias from the library diversity altogether. This can be done by using one single codon for each amino acid and incorporating the codon in a single reaction step. This was first described by Virnекас et al. in 1994 [86] with the development of trinucleotide phosphoramidite synthesis where one codon for each of the 20 amino acids are synthesized as trimers and then coupled in a single reaction step to an oligonucleotide. Other variants of this approach have also been attempted, including MAX randomization [87] where a conventional NNK-randomized gene is modified by primers containing optimized single codons for each amino acid. The prevalence of unwanted codons could thus be depleted.

Using trinucleotide phosphoramidites in the creation of libraries allows for further optimization of the diversity by exclusion of such amino acids that are rare or not observed in certain protein environments, e.g. structural motifs. For instance, the presence of glycines and prolines in regions known to be involved in α-helices or β-sheets should be avoided since the backbone orientation of these amino acids do not support those secondary structure elements. In addition, some amino acids are more likely to take place in antigen-binding interactions of antibodies due to their ability to engage in a wide variety of molecular contacts. By analyzing the frequency of certain amino acids in antibody CDRs it was concluded that certain amino acids are more favored, e.g. tyrosine, and that the distribution is different whether or not the antibody had undergone affinity maturation [88].
2.2 Directed evolution

Directed evolution methods are widely employed in protein engineering to isolate mutants with specific functions or properties from large varied libraries. The following sections will describe the most frequently used methods in the selection of affinity proteins and they all share the common principle of physically linking a gene to the polypeptide it encodes. The gene thus functions as an ID-tag for the protein as the amino acid sequence can be directly derived by DNA sequencing. In the case of affinity selections, specificity and high affinity are attractive properties that usually go hand in hand. This is because protein-protein interactions are reliant on multiple direct contacts between amino acids, which makes it unlikely, although not impossible, that a given protein has high affinity for multiple varying proteins. Affinity proteins can thus be isolated by exposing them to a target antigen, to which a binder is desirable, during successive rounds of selection where non-binding clones are washed away and binding clones are retained. The recovered library members can then be amplified by for example bacterial growth or PCR and subjected to another round of selection. A simple way to steer the selection process towards isolating only clones over a certain affinity threshold is to lower the target antigen concentration in each round, thus enriching the output for clones with the desired affinity. Since affinity is defined as an equilibrium constant, it makes sense to let the mutant library pool reach equilibrium with the target antigen before removing negative clones. Protein variants with poor affinity will have a lower chance of binding the target during the specified equilibration period compared to a higher affinity competitor. Alternatively, a bound low-affinity variant is removed from the selection during washing if the lower affinity is due to faster dissociation kinetics. This can also be exploited to increase the selection stringency even further by letting bound proteins dissociate from the target over a specified time-frame before rescuing the remaining candidates. This is a principle more commonly referred to as off-rate selection.

Phage display

Bacteriophages are virus particles that only infect bacteria. Filamentous phages belonging to the Ff class bacteriophages are the most utilized for display of heterologous proteins. These phages only infect gram-negative bacteria through interaction with the F pilus on the cell surface [89]. The use of phage display for directed evolution purposes started with the discovery by George P. Smith in 1985 that peptides can be genetically fused to the minor coat protein PIII of filamentous bacteriophage M13 without
disrupting the normal function of the virion [90]. Later, other phage types have been investigated for the display of proteins, including lambda \( \lambda \) phage [91–94] but the M13 phage remains the most prominent tool for directed evolution purposes. Typically, the gene encoding the protein of interest is sub-cloned into a vector containing phage genes where it is fused to the gene encoding minor coat protein PIII (Figure 2.1A). This provides a physical linkage between the genotype and phenotype and the genetic material stored inside the phage can be propagated and amplified by infection of a bacterial host (Figure 2.1B). It was later proposed that these properties could easily be exploited for screening peptide libraries to isolate sequences reactive to antibodies [95]. This idea has since been further developed to screen large protein libraries to find novel affinity reagents binding to almost any target imaginable including proteins, peptides, carbohydrates [96] and DNA [97].

Figure 2.1: Overview of a phage display selection protocol for the discovery of affinity reagents. (A) Schematic of an Ff filamentous phage particle. Major and minor coat proteins PVIII and PIII, respectively, are indicated. A displayed protein is shown in blue in fusion to PIII. (B) A panning round is carried out by first producing a phage library in \( E. coli \) carrying a phagemid encoding the library insert in fusion to PIII. The production of library-displaying phage is initiated by infection with a helper phage carrying a plasmid encoding a complete set of phage components. Due to the disabled packaging signal on the helper phage plasmid, the phagemid will be preferentially packaged into the nascent phage. The library phage pool can then be incubated with immobilized target antigen and binding phage can be isolated by washing away non-desirable clones followed by elution. The eluted phage in the eluate is then used to infect \( E. coli \), which can be amplified by shake-flask cultivation in preparation for the next round. Individual clones in the output can be sequenced and/or ranked by phage-ELISA.

The phage particles mostly consist of the major coat protein PVIII, making up most of the phage shaft along the length of the virion (Figure 2.1A). Interestingly, the number of
CHAPTER 2. PROTEIN ENGINEERING

copies of PVIII that are produced per particle is adapted to match the size of the genetic material being encapsulated [98], i.e. the phage plasmid containing the M13 packaging signal [99]. This effectively means that the total size of the virion is proportional to the size of the viral DNA contained therein. Minor coat protein PIII and PIX are present at opposite ends of the particle with 5 copies each where PIII is responsible for interacting with the pilus on the bacteria, which confers infectivity to the phage. PIII has been shown to be the most suitable fusion partner for display proteins due to the large range in protein sizes that are possible to display without hampering infectivity or lowering phage production titers [100]. This is in contrast to fusion to PVIII where large proteins have been proven difficult to display since the number of copies of successfully displayed proteins per capsid decreases dramatically with the size of the fusion partner [101,102]. The key issue to consider when choosing which coat protein to display on is valency. Displaying peptides in fusion to PVIII, while effective up to 6-8 amino acids, will potentially display the peptide in hundreds of copies on each phage. Increasing the length of the displayed peptide above 10, however, reduces the potential to generate infective particles significantly [101,103]. The high valency this option offers limits the affinity of the peptides that can be discovered using this form of display [104]. This is because even low affinity clones will have a high apparent affinity due to the added avidity effect, owing to the high valency imposed by the high copy number of Protein VIII. Thus, it would be difficult to isolate high affinity clones from less desirable low-affinity variants.

The term phage display actually refers to the system where all the genes encoding for the phage virion components are present on a single vector along with an M13 origin of replication and one for *E. coli*. The library to be displayed can then be fused to e.g. gene gIII encoding PIII, which results in the production of a homogeneous group of phages where every single copy of PIII will be displaying a member of the library. The components are expressed in the cytoplasm and secreted to the periplasmic space where assembly of the particles subsequently takes place. Again, this means that multiple copies of the displayed peptide will be present on each phage, resulting in multivalent binding of positive clones. In addition, in some instances it may also lead to an uneven growth rate of clones since the production of each phage is also dependent on the expression of the foreign display insert [105]. In the interest of overcoming this, a system called the phagemid system has been developed [106,107]. A phagemid is a plasmid containing the gene for the peptide to be displayed in fusion to a gene encoding a coat protein. A functioning packaging signal together with an M13 and *E. coli* origin of replication and an antibiotic resistance marker is also present to allow for
simple propagation in *E. coli* and packaging of the phagemid into the phage particles. However, no other phage components are present. The rest of the mandatory phage components are instead supplied through infection of a helper phage carrying a plasmid containing the genes of all phage components, including a wild-type version of the coat protein used for display, together with an additional antibiotic resistance marker. The packaging signal on the additional helper phage plasmid has been mutated to prevent it from being assembled into the phage particles. The result is a lower degree of display where a majority of the coat proteins will be wild-type and only a small number will be displaying the library [108].

Many versions of phagemid display exist with many improvements such as inclusion of protease sites in the PIII fusion gene to allow for mild and specific elution conditions to decrease background enrichment [100]. Other notable improvements include the insertion of an amber stop codon between the gene of the displayed protein and the gene encoding PIII to allow for soluble expression of the display protein in non-amber suppressive strains, and functional phage display in amber-suppressive strains [109]. The enrichment of background phage can be decreased even further with the use of a non-infective truncated PIII in the helper phage plasmid. This means that only the phage carrying full-length PIII in fusion to the displayed protein will be infective [110].

Cell-surface display

While phage display is the most frequently used method for directed evolution in the field, it does feature drawbacks that have lead to the pursuit of alternative strategies. Phage display generally requires multiple rounds of panning followed by sequencing of output clones. This also needs to be supplemented with functional screening of individual clones, e.g. using ELISA, if multiple sequences are isolated with minimal convergence in similarity [111]. This means that several days of laborious work needs to be performed in order to deduce if a selection procedure was successful. Prokaryotic systems also does not allow for post-translational modifications, which can be crucial in certain applications where the protein in question is heavily reliant on for example glycosylations [112].

Multiple technologies now exist to express recombinant proteins anchored to the outside of the expressing cell on the cell surface. Since the protein is presented by the same cell carrying the gene encoding the protein, a linkage between phenotype and genotype is naturally established. This also means that several copies of the protein will be present on each cell at any given time. Having the protein presented on the surface of a cell
CHAPTER 2. PROTEIN ENGINEERING

also allows for the use of fluorescence-activated cell sorting (FACS) in order to analyze a pool of cells carrying a library of protein mutants. Potential binders can be discovered by incubating the cells with a fluorescently-labeled target antigen. These properties are universal among cell-surface display platforms but they differ widely in terms of the library size tolerance and suitability for different displayed proteins. Using FACS also offers additional benefits such as being able to monitor the selection progression over the course of several rounds. The amount of bound antigen to the entire cell pool is clearly visualized when they are interrogated in the flow cytometer. The principle is illustrated in figure 2.2. Many flow cytometers also has the ability to measure the fluorescence in several spectra at once by the use of multiple excitation lasers and emission filters. This can be exploited to monitor binding to additional antigens as well as measuring the amount of displayed protein on each cell surface. This is typically implemented by the addition of an affinity tag in the display gene construct.

Clones showing a high ratio between antigen binding fluorescence and display level fluorescence have bound the most target molecules per displayed protein, which translates into a high affinity of that clone (figure 2.2B). Desirable clones can therefore be sorted by employing a selection gate reflecting this principle. The FACS selection methodology therefore functions as an additional stringency mechanism in addition to varying the target concentration.

Yeast display

Directed evolution of proteins using yeast was developed by Boder and Wittrup [113, 114] as an alternative strategy to phage display primarily for the display of antibodies. Since yeast strains are eukaryotic they support the incorporation of post-translational modifications as discussed above and are thus suitable for the production of mammalian proteins that more closely resemble their naturally expressed counterparts [115,116]. Although several methods to express proteins on the outside of yeast exist, the method developed by Boder et al. seems to be the most popular. The cell-adhesion molecule α-agglutinin is used as the fusion partner where a protein of interest is fused to the N- or C-terminus of the aga2p subunit. The aga2p subunit is linked to the aga1p subunit by two disulphide bonds, which in turn is covalently attached to the cell wall glucan [117]. The expression of the protein library fused to agglutinin can be regulated by inducible expression in order to minimize the impact of expression on growth rate [118]. Typically, a GalI promoter is used that can be induced by the addition of galactose [119].
Figure 2.2: Principles of fluorescence-activated cell sorting (FACS). (A) Cartoon depicting the basic setup of a flow-cytometric cell sorter. Cells are suspended in droplets formed in the interface between a sample flow and surrounding sheath fluid. The droplets can either contain none, one or several cells depending on the amount of cells in the sample and the flow-rate. The droplets are interrogated by an excitation laser and the fluorescence is determined by a detector (not shown). A decision is then made whether or not to keep or discard the droplet depending on the chosen gate parameters. A charge is applied to the droplet, which allows for directing the flow by electromagnets into collection tubes. (B) A typical read-out of cell interrogation in a flow-cytometer. Two fluorescence parameters are often monitored at the same time and a sorting area is chosen by a gate (blue polygon).

A potential hindrance when screening cell display libraries is generally that cells may express the displayed protein to varying extents. This could be due to differences inherent to the specific clone but also to morphological differences among the cells, resulting in a difference in the total amount of available surface area. Therefore, a good way to eliminate this problem is to introduce an additional affinity tag or handle that can be used to counter-stain the cells in order to monitor the expression level of each individual cell. The fluorescent signals measured in FACS; one from the bound labeled antigen and one from the expression label, can be monitored simultaneously and the binding signal can be normalized against the expression level. A common expression normalization tag in yeast display is a c-myc epitope tag or hemagglutinin either N- or C-terminally of the displayed protein of interest [120].

Yeast display has been successfully used to enrich antibody fragments and other non-immunoglobulin scaffolds with high affinity to various targets [121]. Notably, it was used to isolate a scFv binder with femtomolar monomeric affinity to fluorescein from a library obtained by a combination of random mutagenesis using a mutator strain and
DNA shuffling [122]. Functional scFv libraries with a size in the order of $10^9$ have been created. For example a scFv library by Feldhaus and colleagues, using homologous recombination to insert the diversity into the display vector followed by electroporation into *S. cerevisiae* [123]. The electroporation transformation efficiencies have since been greatly improved over the years with efficiencies now reaching over $10^8$ transformants per µg DNA, yielding libraries in the range of $10^{10}$ [124].

Displaying full-length antibodies on any display platform has been a challenge and a desire for many researchers as it would eliminate the need for reformatting the genes encoding the $v_H$ and $v_L$ into a full-length framework. Full-length antibodies are significantly more complex molecules compared to scFv or Fab fragment since four chains have to assemble into one structure. While this has not yet been accomplished through traditional means by fusing the antibodies to aga2p, it has been accomplished by excreting the antibodies into the extracellular space and anchoring them to two Z domains, derived from Protein A, expressed in tandem. The dimeric Z domain was first attached to the cell-surface by conjugation to streptavidin and binding to the cell-surface, which had been biotinylated non-specifically [125].

**Staphylococcal display**

Bacterial cell display systems have also been developed in parallel to yeast display. Bacterial hosts remain the most attractive for expression of proteins lacking post-translational modifications or disulphide bridges due to their easy manipulation, fast growth rate and high resilience to physical stress compared to mammalian or yeast cells. For the same reason, bacterial hosts are also attractive for surface expression in directed evolution applications. Gram positive bacteria, especially *Staphylococcus carnosus*, have been investigated for use as a display platform [126–129]. Their thick peptidoglycan layer serves as a suitable anchoring point for secreted proteins that are covalently attached to the cell wall by endogenous sortase. In contrast to gram negative bacteria, the presence of a single cell membrane simplifies the transportation of heterologously expressed proteins to the surface. The fact that the displayed protein is not fused to any endogenous protein spanning the membrane allows for flexibility in terms of the types of proteins that can be displayed [130–133]. However, the electroporation transformation efficiency is a limiting factor with this system and has restricted the maximum library size that can be achieved, necessitating the need for pre-enrichment of larger libraries in other display platforms such as phage display before further selection by display on *S. carnosus* [134,135]. Efforts to increase the transformation efficiency has been done, which has resulted in $10^6$
transformants from a single electroporation [136].

Much like in yeast display, monitoring of the expression level of individual cells is an important component when using FACS to screen protein libraries for affinity protein candidates. This has been achieved by the C-terminal fusion of ABP from Protein G binding fluorescent HSA [135]. Alternatively, the Z domain has been used as an expression normalization tag with fluorescent IgG [137].

**Cell-independent systems**

In all display systems mentioned above, genetic material has been housed in a particle or container of some sort; be it a phage virion or cell, while the encoded protein is displayed on the outside. Other ways of physically linking the genotype and phenotype have been explored that do not rely on a living host to function. Bacterial or fungi host cells have the benefit of being easy to clonally expand thanks to their high growth rate, but PCR methods can also be employed as a reliable way to amplify a pool of DNA molecules. All cell-independent systems have in common that expression of the encoded genes occur completely *in vitro*. This means that cellular extracts containing all the required components; including ribosomes, amino acids and RNA polymerases, are used instead of whole cells. The benefit that cell-independent system offers is complete circumvention of cellular transformation with library plasmids. This enables the screening of much larger libraries as the upper limit of practical library size is set by how many genetic variants can be expressed and subsequently screened for binding to the antigen.

**Ribosome display**

One such system utilizes the mechanism of ribosome stalling upon translation of an RNA strand into its polypeptide product. It was discovered that such a ternary complex between mRNA, ribosome(s) and nascent peptides is stable enough to allow for co-purification of the mRNA when the protein was pulled out by binding to immobilized antigens [138]. The idea to use this property for directed evolution purposes came later as the interest for fast and reliable methods to screen protein diversity grew [139]. Another useful component was the development of cell extracts for cell-free protein synthesis, which has enabled efficient production of many proteins of interest, including antibodies, to be expressed entirely *in vitro* [140]. This system is now known as ribosome display and it has been established as a reliable way to screen for novel binding
proteins [40,141–143]. The library is sub-cloned into a genetic construct containing a promoter sequence but lacking a stop codon. From this construct, mRNA can be transcribed by RNA polymerases and subsequently purified. The mRNA can then be used as template for in vitro translation, which results in the formation of a ternary complex between mRNA, ribosome and the nascent peptide. The mixture must then be kept at low temperatures in order for the complex to stay associated throughout the selection procedure. RNA molecules encoding binding peptides or proteins can then be isolated by using immobilized target antigens and eluting the mRNA by heat or chemical denaturation of the RNA-protein complex. Additional rounds can then be carried out by converting the RNA into cDNA by using reverse transcriptase.

Isolated clones after several rounds of ribosome display can be screened for activity by either cloning the isolated genes into another display format. Phage display is a suitable format for screening since ELISA can be performed using antibodies specific for M13 coat proteins [139]. Alternatively, appropriate tags can be included in the expression construct, which means that crude expression lysates can be assessed directly for binding activity in ELISA [141].

CIS display

A method to anchor a protein-encoding gene to its peptide product has been developed that does not involve a physical linkage between the mRNA and the protein. This system uses the gene encoding the replication initiator protein RepA, which, when located upstream of the ori sequence, acts in cis. In other words, expression of RepA on a plasmid promotes the replication of that plasmid but not other plasmids harboring the ori but lacking the RepA gene [144]. The mechanism behind this property is not yet fully understood but an additional sequence bridging the RepA gene and the ori, a sequence simply denoted CIS, is believed to be responsible. The CIS region has been shown to serve as a strong transcription terminator sequence, causing the RNA polymerase to stall [144]. In addition, it might also serve as an initial binding site for the nascent RepA, which would explain the cis activity [145]. These observations made RepA interesting for fashioning a novel technique for directed evolution. Odegrip and colleagues fused the RepA gene downstream of two different antibody epitope tags followed by a glycine- and serine-rich linker. From this they constructed a linear DNA construct with a promoter and ribosome binding site and expressed the epitope-RepA fusion using a cell-free protein synthesis kit. The resulting product could be recovered by binding to the corresponding antibody, and the bound DNA was subsequently eluted and amplified with PCR [146].
selection was later carried out on a peptide library, from which sequences were isolated with structures capable of sterically inhibiting proteases [147]. This study served as a proof of concept and showed that large protein libraries can be screened using this technique similarly to ribosomal and mRNA display. However, a benefit of this technique over ribosomal display is the option to perform the selection at room temperature [146].
Chapter 3

Engineered proteins in clinical applications

3.1 Affinity proteins as therapeutic and diagnostic agents

Proteins have become more and more interesting to pursue as useful tools in treating wide-spread diseases where other more conventional treatment methods fail or are lacking altogether. The advent of recombinant protein technology gave us the possibility for large-scale production of human proteins in bacteria, which started the era of proteins as therapeutics with insulin being the first approved protein for use in humans almost 30 years ago [148,149]. The number of possible diseases where recombinant proteins can be useful for treatment is too vast to accurately describe here. Therefore, the following discussion will focus primarily on cancer. A common way to treat cancerous tissues is to utilize broadly acting cytostatic drugs that act through various mechanisms with the end result usually being induction of apoptosis or arrest of DNA replication [150]. Cisplatin is an example of one of the more potent drugs, which acts through the former mechanism during treatment of solid tumors [151].

As outlined in the previous chapter, antibodies are highly versatile tools that can be adapted to bind virtually any protein target. Antibodies belonging to the IgG isotype typically have the highest affinities to their respective antigens due to the extensive
affinity maturation they undergo through somatic hypermutation during B-cell clonal selection in germinal centers [77]. The large diversity present in the immunoglobulin genes and the potential to raise antibodies binding specifically and with high affinity to a wide array of different targets has inspired researchers to harness this power and apply it to other proteins as well. Full-length antibodies are capable of much more than just molecular recognition including immunological functions such as recruitment of natural killer (NK) cells and T-cells to initiate activation of apoptosis of target cells but also label a target for opsonization and degradation by monocyte/macrophages [6]. This makes them useful in therapeutic settings where these functions might be helpful or even essential to reach the required efficacy [5].

The size of affinity proteins is a source of debate in the field of biological therapeutics. It has been clearly shown that there is a correlation between molecular weight and distribution in different tissues. Recent evidence suggests that a size increase of around 35 kDa for different antibody fragments leads to a 50% decrease in the amount of absorbed protein into tissues [152]. Thus, it seems there is an inverse relation between molecular weight and tissue absorption. By this comparison alone, a smaller protein such as an Affibody molecule with only a fraction of the molecular weight of an antibody would be the obvious choice when choosing a protein to engineer for binding to a target on solid tumors. However, another important determinant of biodistribution is the in vivo half-life of the protein therapeutic, which is primarily limited by the renal elimination of proteins. This mechanism is highly complex and involves multiple cellular mechanisms, which have been extensively reviewed [153]. A plethora of evidence exist that support the general notion that smaller molecular weight compounds and proteins are rapidly renally filtered, degraded and ultimately passed through the urine. Proteins of larger molecular weight, such as antibodies, remain in blood circulation for a longer period of time, which has also been attributed to a higher activity in vivo during treatment of EGFR (ERBB1)-expressing solid tumors [154]. The question is if a longer in vivo residence time of protein therapeutics is always desirable since the mode of action and physical location of the malady being treated varies widely. Binding to a target molecule expressed in a solid tumor, for instance, requires that the targeting moiety be able to extravasate from the blood stream into the surrounding tissue and then infiltrate the tumor microenvironment [155]. This is in stark contrast to an antigen readily available for binding in the peripheral blood, which does not impose any such constraints.

The intention behind the administration of a protein moiety in a patient needs to be clearly outlined before any discussions are held regarding optimal pharmacokinetics. To
put it simply, we can divide the usage scenarios into two groups; therapy, i.e. treatment of a disease; or diagnosis where the affinity protein can be used as a radionuclide tracer for positron emission tomography (PET) or single-photon emission computed tomography (SPECT). In the former case, a biological effect is required such as inhibiting the proliferation of cancer cells or the dampening of an auto-immune response by sequestering cytokines. This entails that the therapeutic first reach the affected tissue or site and then exert its function. Conversely, in diagnostic imaging of cancerous tissue, it is desirable to reach the diseased site rapidly while simultaneously clearing the blood circulation to achieve a sufficient contrast between the tumor and surrounding tissue and blood vessels. Having radioactive compounds active in circulation for a long time is not desirable either, which is why the imaging tracer should preferably be cleared quickly.

When comparing these two, albeit broad, usage scenarios it becomes reasonable that the same protein compound cannot be used in both cases. Indeed, full-length antibodies have evolved to remain active in the blood circulation for long periods of time making them ideal candidates for therapy. This is partly due to their large size, as already discussed, but also through a mechanism involving the neonatal $F_c$ receptor (FcRn), which will be discussed below. Smaller affinity proteins are thus more readily advantageous for diagnostic applications [156,157].

Therapeutic modalities for affinity proteins

One of the first antibodies approved for use in humans was Herceptin (also known as Trastuzumab) binding the epidermal growth factor receptor 2 (ERBB2). This monoclonal antibody was developed by Genentech starting with a mouse antibody. It was later humanized to avoid eliciting an immune response to the therapeutic itself [158] and it was approved for use in humans in 1998 for the treatment of metastatic breast cancer. ERBB2 was chosen as a therapeutic target owing to its role in the progression of the disease and its correlation with poor survival and has since been a popular choice for targeted therapy [159–162], much due to the success of Trastuzumab. The mechanism of action of Herceptin has been, and still is, a topic of discussion. Being an antibody, it was believed that Trastuzumab acted through its immunomodulatory functions, i.e. antibody-dependent cellular cytotoxicity (ADCC), which entails the recruitment and cytotoxic action of NK cells. While this has been shown to be the case [163,164], it cannot account for the whole therapeutic effect as the interaction with ERBB2 has also been shown to interfere with the downstream phosphoinositide 3-kinase (PI3K) pathway, thus preventing the activation of survival and proliferation mechanisms [165,166]. Moreover,
treatment with other human antibodies specific for ERBB2, such as Pertuzumab, did not yield similar results as trastuzumab when used alone. However, these antibodies have been shown to be highly effective in combination, which has been attributed to pertuzumab binding to a different epitope on ERBB2. This has been shown to block the dimerization domain of Her2, perturbing its ability to bind with the other epidermal growth factor receptors ERBB1 and ERBB3 [167,168]. Dimerization is crucial in order for ERBB2 to transduce signaling, due to its inactive tyrosine kinase domain, which explains why this mode of action is effective [160].

Trastuzumab and Pertuzumab serve as excellent examples of cancer therapy using targeted affinity proteins. They demonstrate that eliciting a therapeutic effect on a cell-bound receptor can happen through multiple mechanisms and it also shows that predicting that effect a priori is very challenging. The epidermal growth factor family of receptors belongs to a signaling network where almost every constituent, both extracellular and intracellular, have been targeted for therapy in some form. Simply targeting one member of these pathways has generally proven to be a strategy with limited efficacy since not all patients involved in therapeutic trials respond to the treatment. For example, mono-specific ERBB2 treatment may eventually fail in inhibiting tumor growth due to development of resistance mechanisms such as upregulation of other oncogenic markers or simply survival of tumor cells with low ERBB2 expression [169–172]. The prevalence of these resistance mechanisms observed during clinical studies has motivated the investigation of alternate or multi-pronged methods of engaging the tumor tissue.

A relatively simple way to increase the efficacy of a treatment is to use two or more affinity proteins at the same time, as exemplified again by Trastuzumab and Pertuzumab, which have seen the most success when combined in one injection [168]. However, yet another avenue in exploring treatment options is to combine two different affinities in a single molecule, thereby generating multispecific affinity proteins. Antibodies offer a way to engineer bispecificity into a molecule due to its two binding arms. The challenge has thus been to combine two heavy- and light-chain pairs from two different antibody molecules; a problem that has been overcome through a number of accomplishments, arguably the most famous of which is the “knobs-into-holes” strategy presented by Genentech in 1996 [173]. By introducing mutations into each respective heavy chain CH3 domain, the authors made a favorable heterodimerization interface that resulted in the recovery of bispecific antibodies to 92 % when the two heavy chains were co-expressed. However, this strategy does not solve the problem of exclusive heavy- and light-chain pairing. That was solved later by engineering the Fab regions in a similar fashion as the
“knobs-into-holes” method [174].

Recombinant protein technology facilitates other ways of producing multispecific proteins. Many scaffold proteins, of immunoglobulin origin or otherwise, can be expressed alone or in fusions with other proteins without compromising expression levels or stability. By introducing linker peptides between two recombinant genes, a stable fusion can be created, provided the linker region is appropriately designed. Properties such as length, solubility and flexibility need to be considered in order to facilitate optimal linkage between two therapeutic proteins. A popular option is a linker that is both conformationally flexible as well as water soluble as the linker amino acids will be constantly exposed to the surrounding aqueous environment. An example of such a linker is a varying number of repeats of four glycines interrupted by a serine [175].

Multispecific strategies offer near infinite possibilities in terms of target combinations. The ERBB receptors have been popular targets for multispecific proteins and many such attempts using antibodies have been shown to be more effective than simply targeting one of the receptors [176–178]. Another intriguing strategy is targeting the same receptor molecule with two domains binding to non-overlapping epitopes. This has been demonstrated using DARPins targeting ERBB2 [179]. Jost and colleagues showed that the linker length between the two DARPins was critical in eliciting a biological effect, where a short linker of only 5 residues showed the strongest inhibition of cell proliferation. The authors suggested that this effect is caused by the formation of receptor oligomers that lock the receptors in inactive states. The specific nature in which this effect occurred, including factors such as linker length and domain orientation, illustrates the complexity in targeting these signaling pathways when multispecific and multivalent therapeutics are used.

### 3.2 *In vivo* half-life extension strategies

#### Increasing the hydrodynamic radius

Protein of low molecular weight are known to be easily filtered through the glomerulus in the kidneys. The size limit that a protein has to surpass in order to avoid this is not precisely determined but is roughly estimated to be around 50 kDa [153]. Proteins that enter the kidneys are filtered through pores that have been determined to be 4-5 nm for the most numerous, and around 10 nm for less frequent pores [180]. This means that discussing the hydrodynamic radius of proteins is more meaningful since proteins of the
same size can have varying topology and structure, and therefore different hydrodynamic radii. However, the molecular weight is an easier metric to measure and compare than radius, which is why it is used as an estimate. The charge of a protein is also a factor affecting the renal elimination to a certain extent but the major determinant is still the size when comparing proteins with vastly different molecular weight. IgG1 with a molecular weight of 150 kDa has an \textit{in vivo} half-life, i.e. the time it takes for half of the molecules to be degraded and eliminated, of several weeks. In contrast, a typical scaffold protein, such as an Affibody with a molecular weight of 6-7 kDa, has an \textit{in vivo} half-life in the order of minutes [181]. A popular method to increase the \textit{in vivo} half-life of scaffold proteins is therefore to increase the effective molecular weight over the glomerular filtration cut-off. This can be done by conjugating the protein to some form of inert polymer that does not modulate the affinity or function of protein, but merely increases the hydrodynamic radius. A common method is conjugation to polyethylene glycol (PEG), which can be varied in size to optimize the pharmacokinetics of the therapeutic in question. The conjugation can be done either at multiple sites or on a single specific amino acid, e.g. a single reduced cysteine residue. This is a strategy that has seen frequent use over the years and its effectiveness has been proven [182–184]. Indeed, studies done by Tolcher et al. in 2011 [183] using a binder based on fibronectin and Mross et al. in 2013 [182] using the Adnectin scaffold both showed comparable pharmacokinetic profiles when using a 40 kDa PEG molecule to extend the half-life to about a week. The effective molecular weight of those examples were 50-60 kDa, thereby reaching and exceeding the kidney cut-off.

\textbf{Exploiting the FcRn rescue pathway}

The extraordinary half-life of IgG was further explained by the discovery of FcRn, which was shown to be ubiquitously expressed, primarily on endothelial cells and immune cells [185]. FcRn is structurally similar to MHC class I and is also bound to β-microglobulin, but instead of presenting foreign peptides to the immune system, it has pH-dependent affinity to the constant region of various immunoglobulins [185–190]. At pH 7.4, i.e. the normal hematopoietic pH level, FcRn does not bind IgG. The interaction only occurs at pH levels around 6, which is reached in acidified endosomes following the random engulfment of pinocytosis. FcRn binds the IgG in the endosome and traffics it back into circulation. Other proteins that do not bind FcRn are instead degraded in the endolysosome. The solved structure between FcRn and \( F_c \) shows that the interaction occurs at the \( C_H2-C_H3 \) interface on \( F_c \) [191]. Attempts have since been made to make
antibodies with engineered $F_c$ regions to increase the affinity to FcRn at pH 6, and this also lead to an increased serum half-life, as shown by Zalevsky et al. [154]. Since the $F_c$ domain is seemingly responsible for the prolonged half-life of IgG, many groups have genetically fused their proteins of choice to $F_c$ in order to gain the long \textit{in vivo} half-life of antibodies, while also forming homo-dimers through the natural pairing of the $C_H2$ and $C_H3$ domains. This increases the \textit{in vivo} half-life, not only through the interaction with FcRn, but by also dramatically increasing the molecular weight. The dimerization of the therapeutic is also expected to increase the effective affinity to the target antigen by the avidity contribution of having two binding moieties in the same molecule. Despite these improvements, $F_c$ fusion proteins generally do not reach the \textit{in vivo} half-life of natural IgG [192], but still has \textit{in vivo} half-lives in the range of a few days. The reasons for this are unknown but it indicates that the interaction to FcRn might be dependent on a structural conformation only found in full-length IgG.

Serum albumin, the most abundant protein in serum with a concentration around 35-50 mg/ml, also has a prolonged serum residence time compared to other proteins of its size (66 kDa). Albumin is remarkably persistent \textit{in vivo}, having a half-life of 19 days in humans. Similarly to antibodies, albumin is also rescued from degradation by interacting with FcRn, which has thus evolved to interact in a pH-dependent manner to two different serum proteins. IgG and albumin binds to two non-overlapping sites on FcRn and can thus be rescued simultaneously [185, 187, 193]. Albumin has thus also become a tool to rescue pharmaceutical proteins from degradation by either binding reversibly or covalently to albumin. The latter approach can be done by chemical conjugation or genetic fusion. Since this approach requires preparation or expression of albumin, modified versions of albumin can be used in this way, which, similarly to engineered $F_c$, can result in a higher affinity to FcRn and improved pharmacokinetics. Andersen and colleagues identified a single residue in domain III of human serum albumin (HSA) that, when mutated to any other amino acid from the wild-type lysine, increased the affinity to FcRn at pH 6 [194]. Albumin fusion proteins have been constructed in multiple studies and the half-life of many protein pharmaceuticals have been extended using this method, including single-chain antibodies [195] and recombinant factor IX, which has currently undergone phase III clinical trials [196].

Fusing HSA to pharmaceutical proteins complicates its production as HSA contains many paired cysteines forming disulphide bonds, although recombinant production of albumin in yeast has proven successful [195, 197, 198]. An alternative strategy is to enable the protein to associate to albumin reversibly using one of several available routes,
including molecules with natural affinity to albumin such as fatty acids and protein domains, but also engineered proteins and peptides [199,200]. The albumin-binding domain, described in detail in the next section, has gained attention for being particularly effective at mediating an increased in vivo half-life of a wide range of fusion partners.

### 3.3 The albumin-binding domain

Many streptococcal and staphylococcal strains exhibit a high virulence and pathogenicity in humans and animals. A large number of bacterial receptors have been identified, collectively termed Receptins, which bind several different serum proteins [201]. In the case of *Staphylococcus aureus*, as described above in section 1.4, the $F_\text{c}$-binding membrane-anchored protein A is expressed, which hinders immuno-activated opsonization [19]. Similar proteins are also expressed on several pathogenic streptococcal strains that bind serum proteins for the same purpose; avoiding immune-driven elimination and also scavenging nutrients [202]. Many such proteins have been identified including Protein G from members of group C and G streptococci [203–205] and Protein PAB from *Finegoldia magna* [206]. Protein G and PAB are examples of proteins showing a high affinity to albumin from several mammalian species [207]. The N-terminal half of the 63 kDa Protein G facilitates binding to albumin and the C-terminal end binds immunoglobulins [208], whereas protein PAB only binds albumin. A closer examination of these albumin-binding regions revealed repeating domains with a high sequence homology within the Protein G N-terminal region that are also homologous to regions in other albumin-binding receptors, indicating that these proteins have evolved by duplication events and module shuffling [206]. The structure for the third albumin domain (henceforth referred to as ABD) from Protein G originating from strain G148 has been determined by NMR, showing that the domain folds as a bundle of three right-handed anti-parallel $\alpha$-helices [209,210]. This fold is also shared among other albumin-binding domains such as the GA module from Protein PAB, as confirmed by both NMR [211,212] and crystallization studies [213,214].

G148-ABD3 (Figure 3.1A & B) has been extensively characterized and used for engineering purposes. When expressed alone, ABD3 is highly stable with a melting point around 80°C and contains no cysteines [210,215]. The affinity to albumin has been attributed mainly to residues in helix 2, as determined by an alanine scan involving residues in helix 2 and 3 and the connecting loops [216]. The study identified several residues that, when mutated to alanine, significantly decreased the affinity. The crystal structure between
the homologous domain ALB8-GA from protein PAB and HSA shows that residues in helix 2 and the flanking loops make direct contact with the albumin molecule [213,214]. Since this domain is structurally very similar to ABD3 from Protein G, and considering the fact that they share 55% pair-wise sequence identity (Figure 3.1A), it is reasonable to assume that the interaction to albumin is conserved. The residues that were identified by Linhult et al. [216] agree well with the findings regarding the corresponding amino acids in ALB8-GA by Lejon et al [213]. A central hydrophobic interaction between ABD and HSA is formed by residues Tyr-20 (Phe in ALB8-GA) and Tyr-21 on ABD. The interaction is further supplemented by two flanking hydrogen bonding networks by residues in the loops preceding and following helix 2, respectively (Figure 3.1C).

Due to the low nanomolar affinity even for the single-domain ABD, several derivatives of the albumin-binding portion of Protein G has seen frequent use in affinity chromatography in order to purify or deplete albumin from serum samples [217]. Improvements have been made to this end by engineering a variant of ABD with improved resistance to alkaline treatment, thereby making it compatible with cleaning-in-place procedures. This was done by substituting several asparagines with aspartates without negatively effecting the affinity to albumin [215]. In addition, these mutations also increased the melting temperature of ABD by around 10°C. Several efforts have been made to the deduce the determinants governing stability and species specificity in order to facilitate engineering of these domains. Rozak et al. managed to isolate a sequence from a phage display library constructed by shuffling the genes of 16 ABD homologs using offset recombinant PCR. The most prevalent clone, denoted PSD-1 (figure 3.1A) had increased stability and increased affinity to human albumin as well as guinea pig albumin compared to G148-ABD3 [218].

ABD forms a stable fold independently of intramolecular disulphide bonds and can be expressed either alone or in fusion with other heterologous domains without loosing its albumin-binding properties [221,222]. These features have made ABD an attractive domain for use in protein engineering. Ahmad and co-workers replaced the albumin-binding surface by mutating residues in helix 2 and 3 and managed to isolate binders to interferon gamma [223]. In a later publication, the same group presented binders to human PSP94 [224], confirming that the scaffold is indeed resilient to mutation of surface-exposed amino acids without compromising the structure.

The high affinity to HSA also makes ABD an interesting scaffold for use in therapeutic applications where it has been used as a general fusion partner in order to prolong the in vivo half-life of many different protein pharmaceuticals, including Affibody molecules.
CHAPTER 3. ENGINEERED PROTEINS IN CLINICAL APPLICATIONS

Figure 3.1: Sequences of ABD homologs and structures. (A) Amino-acid sequence alignment of the two natural albumin-binding domain homologues ALB8-GA and G148-ABD3 as well as four engineered variants. ABD* is the stabilized variant of ABD3 by Gülich et al. [215]. ABD035 is the femtomolar affinity ABD engineered by Jonsson et al. [219]. PSD-1 (phage-selected domain 1) was selected by Rozak et al. from a phage library constructed by shuffling the sequences of 16 ABD homologues [218]. ABDcon is a consensus sequence presented by Jacobs et al. [220], resulting from aligning 20 publicly available sequences that most closely resembles G148-ABD3. (B) The protein structure of G148-ABD3 as determined by NMR (PDB file: 1gjt) [210]. (C) The protein structure of ALB8-GA (blue) in complex with HSA as determined by Lejon et al. [213]. Domains 1, 2 and 3 of HSA are colored green, orange and red, respectively (PDB file: 1tf0).

and antibody fragments [30,225–227]. The ABD from Protein G has a broad species specificity, making it particularly useful in clinical settings due to its compatibility with mouse and primate albumin [210]. The ability to tune the pharmacokinetic properties of a protein therapeutic is a valuable prospect. It is highly desirable to minimize dosing frequency and toxicity but at the same time maximizing efficacy. Several studies show that the affinity of the albumin-binding domain used as a fusion partner affect the in vivo half-life. Hopp et al. [225] showed that an increase in mouse albumin affinity from 21.4 nM to 1.8 nM, i.e. a factor of almost 10, increased the elimination half-time in mice by about 30 % (36.2±4.8 h to 47.5±14.8 h). Jacobs et al. [220] studied a consensus sequence of ABD (ABDcon, figure 3.1A), which they designed by aligning the sequences of 20 ABD homologs. A panel of mutants was constructed from this consensus sequence, resulting in affinities to mouse albumin ranging from 1.2 nM to 1.6 μM. Both of these studies concluded that even a low affinity for albumin provides the fusion protein with a significantly prolonged half-life. Jonsson et al. applied directed evolution to a library based on ABD and managed to generate a variant (ABD035, figure 3.1A) with femtomolar affinity (50-500 fM) to HSA [219]. This is the strongest albumin-binding variant described to date and should thus enable the longest serum half-life possible.
using this strategy.

Previously, ABD was engineered in order to introduce a binding site into the domain in addition to the albumin-binding paratope, primarily localized to helix 2 [213, 216]. Alm et al. [228] created a phage display library based on a stabilized version of the third albumin-binding domain of Protein G [215]. They panned this library against the IgG-binding Z domain in an effort to create bispecific proteins for use as an affinity purification handle. A variant with affinity for both HSA and the Z domain was successfully isolated, but was found to be dependent on a cysteine residue introduced in helix 1. The new protein thus bound the Z domain exclusively as a dimer. The same library was later used to select for binders against TNF-\(\alpha\) [137], this time yielding two candidates with moderate affinity to the target. Affinity maturation was subsequently employed, which resulted in higher affinity to TNF-\(\alpha\) but it was also shown that the affinity to albumin was also negatively affected to varying extents. Interestingly, the isolated variants all had glycine residues, and in one case a proline residue, in helix 1, meaning that these binders most likely did not retain the original three-helical fold.
Chapter 4

Present investigation

4.1 Investigation aim

The work in this thesis focuses on evaluating different engineering strategies for the albumin-binding domain with a primary focus on creating bispecific proteins aimed at clinical applications. Papers I and II present selection and characterization of variants with affinity for tumor antigens ERBB3 and ERBB2, respectively. Both these targets have been identified as important markers for several cancer types and over-expression of either is associated with a poor prognosis [160]. Signaling is activated by dimerization of the extracellular domains, which allows the intracellular tyrosine kinase domains to autophosphorylate. This in turn activates several signaling pathways by binding and phosphorylating intermediary signaling transducers. ERBB2 is unique among the receptors as it currently does not have any identified ligands and its extracellular domain is constitutively arranged in a conformation that allows for dimerization with other receptors. In contrast, ERBB3 lacks a functioning tyrosine kinase domain and thus requires heterodimerization with another receptor in order to transduce signaling. This can occur when the receptor has been activated by binding of a ligand, which changes the conformation of the receptor and exposes a dimerization interface.

Affinity proteins targeting ERBB2 and ERBB3 receptors can potentially be useful in combating cancerous disease where the high proliferation rate of cells is contributed by abnormally high expression of these receptors. Binding of the extracellular portions may abrogate their function by sterically hindering binding of natural ligands or formation of
dimers. In addition, over-expressed tumor markers can be targeted to detect abnormal tissue in diagnostic imaging applications. This potential avenue is explored in paper III where an ERBB2-binding ABD variant (or ABD-Derived Affinity Protein, ADAPT) is used as a radionuclide imaging probe. For this application, it was important to remove the affinity to albumin, which had been found to be possible in paper II, without sacrificing affinity to ERBB2.

Finally, in paper IV the established staphylococcal display platform is evaluated as a method for directed evolution in terms of reproducibility. An affinity maturation was performed in two separate attempts to compare the resulting outputs in relation to chosen selection strategy.

### 4.2 Development and characterization of small bispecific albumin-binding domains with high affinity for ERBB3 (Project I)

The ABD library that was previously used by Alm et al. to select for binders to the Z domain [228], and later also by Nilvebrant et al. in selections against TNF-α [137], was again used here in order to isolate binders with affinity to ERBB3. The library was created by subjecting 11 surface-exposed residues to randomization using NNK codons (Figure 4.1A). The library was assembled and sub-cloned into a phagemid vector and the size of the library was estimated to $10^7$ based on the number of achieved transformants when electroporating the phagemid into *E. coli*. The selection was carried out in four rounds where the ERBB3 concentration was lowered in each successive round. Sequencing of random clones after the fourth round resulted in a panel of variants with similarities in several randomized positions, indicating that convergence had occurred to a common binding motif. In addition, several candidates were found to be more frequent than others, which showed that some sequences were enriched and thus potentially favourable. No sequences were found that did not follow the design of the library. Seven clones were chosen for further characterization based on their representation of different similarity clusters when aligning all output sequences and the number of times they were observed during sequencing. This panel of variants thus had high sequence variability relative to each other but at the same included more frequent clones that were expected to be of most interest in terms of affinity. Figure 4.1B shows the amino acid sequences of the seven chosen clones compared to the scaffold ABD. A few striking similarities
could be seen including positions 6 and 7 where a positive amino acid is always paired with a hydrophobic amino acid. The more frequent clones had an arginine in position 6 and a leucine in position 7 whereas $ABD_{ErbB3−20}$ has an alanine in position 6 and an arginine in position 7. Arginine also seemed to be a frequent amino acid in positions closer to helix 2. In helix 3, similarities were less apparent but mostly charged and polar amino acids were observed with a few exceptions, such as a proline in position 43 for $ABD_{ErbB3−18}$.

The panel of seven binders were successfully expressed in E. coli incorporating an N-terminal histidine tag for use as a purification handle in case the intended method of affinity purification on an albumin-functionalized sepharose resin failed. However, all seven variants were successfully expressed and purified to high purity using the HSA affinity as determined by gel electrophoresis and mass spectrometry. This confirmed that the affinity to HSA had been retained, consistent with the results from the previous selections against the Z domain and TNF-α [137, 228]. The affinities to albumin and ERBB3 were determined separately using SPR and varied between 10 and 94 nM to ERBB3 for $ABD_{ErbB3−3}$ and $ABD_{ErbB3−18}$, respectively. Unexpectedly, the affinities for the two most enriched clones did not stand out as particularly high compared to the other clones. However, the lowest target concentration during selection was 10 nM and only in one of four separate tracks. The other tracks used 80 or 50 nM as the lowest concentration, which would not favor a 10 nM binder over $ABD_{ErbB3−1}$ with a 75 nM $K_D$. The selection stringency was thus not high enough to enrich the most high-affinity clones $ABD_{ErbB3−3}$ and $ABD_{ErbB3−27}$. However, it does not explain the high enrichment observed for $ABD_{ErbB3−1}$ and $ABD_{ErbB3−2}$. 

Figure 4.1: Library design and sequences for isolated clones. (A) A model of ABD is shown with the randomized positions indicated in red. (B) The panel of seven binders that were investigated in paper I are shown with single-letter amino acid codes at each position. Dashes are used in positions that were not changed relative to the scaffold ABD shown at the top. Approximate positions of the three helices are shown, corresponding to the scaffold ABD.
The affinities to albumin were generally very high with $K_D$ values below or around 1 nM for all clones except $ABD_{Er\beta B3-18}$, which had an affinity of 81 nM to HSA. The high number of positive amino acids in the sequences could potentially explain the higher affinity to HSA compared to the scaffold ABD. The theoretical pI of HSA at pH 7.4 is 5.7, which results in a negative net charge. The presence of positive amino acids in ABD could thus contribute to higher association kinetics, despite not directly participating in the interaction upon binding. This showed that the introduced binding affinity to ERBB3 also affected the affinity to HSA, which has been shown before [137], although this time the affinity to HSA was increased.

To get an idea of the potential usefulness of these binders in a therapeutic setting, a competition experiment was performed where the natural ligand Heregulin (NRG-$\beta$1) was immobilized on a SPR sensor chip. Injections of the extracellular domain of ERBB3, pre-incubated with varying amounts of different competing ABD variants and control proteins, were then injected over the surface. The max responses from these injections were recorded and plotted as a function of competitor concentration. The results clearly showed that pre-incubation with ERBB3-binding ABD variants resulted in reduced binding levels to NRG-$\beta$1 (Figure 4.2A). The reduction in binding signal also seemed to follow in an affinity-dependent manner. These results do not answer conclusively if this observed binding inhibition is due to overlapping binding sites or due to allosteric interference. However, the binders, in particular $ABD_{Er\beta B3,3}$, show similar concentration-dependent binding inhibition as NRG-$\beta$1 itself, which indicates that the binding sites are indeed overlapping. This behavior was confirmed when a flow-cytometric binding analysis was performed on the ERBB3-expressing breast-cancer cell-line AU565. In this case, NRG-$\beta$1 acted as competitor, which confirmed the earlier results. NRG-$\beta$1 managed to completely block binding to the cells when they had been incubated with 100 nM NRG-$\beta$1 to equilibrium (Figure 4.2B).

In a final experiment, the two binding sites to ERBB3 and HSA on ABD were tested at the same time in order to determine if the two molecules could be bound simultaneously. A competition experiment was performed again using SPR where pre-incubations with HSA and albumin were injected over a surface of ERBB3. The maximum response from each injection was normalized to reference injections where HSA was omitted, and the data was plotted as a function of albumin concentration. The results, shown in figure 4.3, show that increasing amounts of albumin lead to decreased binding levels. This shows that the interactions cannot occur at the same time. In the case of $ABD_{Er\beta B3-27}$ the binding is completely abolished, whereas $ABD_{Er\beta B3-18}$ exhibited a lower degree
Figure 4.2: ERBB3-binding ABD variants compete with Heregulin for binding to ERBB3. (A) Heregulin (NRG-β1) was immobilized on a SPR sensor chip and ERBB3 pre-incubated with varying concentrations of competitor was injected over the surface. Bars represent mean values from identical injections where the maximum binding signal was recorded for each injection. Error bars show the standard deviation. (B) FACS analysis of the indicated cell lines incubated with labeled ABD binder. BG=Background fluorescence. Error bars show the standard deviation.

of signal reduction at increased concentrations of albumin, which might be explained by the reduced affinity to albumin for this binder. The behavior of \(ABD_{ErbB3-3}\) is more difficult to explain, however. It still retains roughly 50 % binding signal even at 2000 nM HSA. One explanation is non-specific binding caused by the ABD-HSA complex, meaning that the observed binding is an experimental artifact. Another remote possibility, albeit unlikely, is that the complex between ABD and HSA can still bind ERBB3, but with a significantly reduced affinity. The binding would thus be most likely mediated by helix 1 alone, considering the close proximity of HSA to helix 3. This would explain the data shown for \(ABD_{ErbB3-3}\) but it would require further exploration to determine conclusively.

4.3 Engineering of bispecific affinity proteins with high affinity for ERBB2 and adaptable binding to albumin (Paper II)

An additional selection with the ABD library aimed at generating bispecific proteins was attempted against ERBB2. The selection was carried out in a similar fashion as in paper I, but this selection resulted in the isolation of a single binder, termed \(ABD_{ErbB2-1}\) with vast majority among sequenced clones. This binder was found to have 75 nM affinity to
Figure 4.3: Data from a competition SPR experiment. ERBB3 was immobilized on a sensor chip surface and a constant concentration of ABD binder pre-incubated with increasing amounts of albumin was injected. The bars represent mean maximum binding signals obtained per injection type and the error bars indicate standard deviation. All values were normalized to reference injections without albumin.

ERBB2 and an HSA affinity around 1 nM, again very similar to the selection against ERBB3 in paper 1. The amino acids in the randomized positions were primarily of polar and charged varieties, except for two alanines in position 39 and 43. The binder was thus expected to have a retained three-helical structure, which was confirmed using circular dichroism spectroscopy.

Since only one binder was isolated, more information about the binding interface to ERBB2 was needed to draw conclusions about which positions are most important. To this end, an alanine scan was performed where each position was separately mutated to alanine. The positions that were already occupied by alanine were mutated to the larger hydrophobic residue valine in order to assess the importance of the alanine residue. Each mutant was expressed and the binding affinity to ERBB2 was evaluated. A39V was omitted after failing expression multiple times. Positions 7, 10, 11 and 43 had the most profound impact on binding affinity, since these mutants did not bind ERBB2 at all. The other mutants had little or no impact on overall binding affinity. With this information, two new libraries were constructed where these four important positions were either completely retained and all other positions randomized by NNK, or only retained to 50%. These libraries were denoted conservative and semi-conservative, respectively.

The libraries were panned against ERBB2 again but with a key change in selection methodology. HSA was included during the selection in order to favor clones that are
able to bind ERBB2 in presence of high concentrations of HSA. In previous selection attempts, HSA was not included as a selection parameter, which has consistently lead to an inability to support both binding affinities at the same time. This affinity maturation was performed in two stages: First, the phage display selection was performed either with or without albumin. Second, the entire phage-selected binder pools were sub-cloned into a staphylococcal cell-surface display vector and electroporated into *S. carnosus*. Further selections were then made using FACS by incubating the binders expressed on staphylococci with labeled ERBB2 with or without an excess of HSA.

Figure 4.4 summarizes the results from the selections. The plots show density outlines of FACS-interrogated cells with ERBB2-binding signal as a function of surface expression level as gauged by labeled IgG bound to the included Z domain expression tag. The cells were incubated with 50 nM ERBB2 with and without 1 μM HSA. Panel 4.4A shows the behavior of phage selected binders before being subjected to FACS selection. Binders isolated from tracks without HSA present show a decreased binding signal when HSA is included. In contrast, the populations originating from selection tracks where HSA was included show a partly retained binding signal. Panel 4.4B show the impact of including HSA in the selection on the resulting HSA affinity of the binding clones. Binders that were not challenged with HSA (tracks A, C, E and G) retain their binding affinity to albumin, whereas binders challenged with HSA (tracks B, D, F and H) divide into two populations; one with retained binding affinity to HSA and one where the affinity to HSA had been lost.

The explanation to this was found during sequencing of the respective clones from the selection tracks. Binders that had lost the HSA affinity had unintended mutations in scaffold residues, mostly in the linker between helix 2 and 3 and in helix 3. These mutations caused almost complete loss of HSA affinity. This shows that binders with the ability to bind both ERBB2 and HSA at the same time was not found in any track. However, many of the clones isolated from the FACS selection had high affinity to ERBB2, with some clones showing sub-nanomolar affinities.

The effects of the identified mutations were further examined by grafting them onto a binder with high affinity to ERBB2; \( ADAPT_{ERBB2-FACS-12} \). Figure 4.5A show the effect of increased concentrations of albumin on the ERBB2-binding capability with either of the three mutations A28V, G33D and L37R. The results clearly show that the binding signal to ERBB2 is retained even in the presence of 600 μM HSA. These scaffold mutations were thus responsible for conferring ERBB2 binding at high levels of HSA.

Finally, competition experiments were also done in order to gain information about the
Figure 4.4: Flow-cytometric analysis of (A) phage-display selected outputs after step in the affinity maturation. Staphylococcal cells expressing the polyclonal binder pools were incubated with ERBB2 with/without HSA present. (B) Flow-cytometric analysis of the HSA-binding capacity of polyclonal outputs from different FACS selection tracks.
CHAPTER 4. PRESENT INVESTIGATION

Figure 4.5: Two different competition experiments are shown. One was performed in order to evaluate the HSA and ERBB2 binding capability of three different scaffold mutants. The other was performed using soluble proteins on ERBB2-expressing SKOV-3 cells. (A) Staphylococcal cells expressing the indicated clones were incubated with labeled ERBB2 and varying amounts of HSA up to 600 μM. (B) SKOV-3 cells were incubated with labeled ADAPT binder with or without the indicated competitor. Bound ADAPT was detected with labeled streptavidin using a site-specific biotin. * indicates p-value=0.0067.

epitope that the ERBB2-binding ADAPT variants interact with on ERBB2. SKOV-3 cells over-expressing ERBB2 were incubated with a labeled representative ERBB2-binding ADAPT with or without competing binding molecules. Figure 4.5B shows fluorescence intensities recorded from flow-cytometric analysis of the treated cells. Incubating the cells with a Trastuzumab scFv or the ADAPT molecule itself lead to a complete loss in binding signal. This shows that the isolated ADAPT variants compete with Trastuzumab and thus bind to the membrane-proximal domain IV on ERBB2 [229]. This is interesting since Affibody molecules selected against ERBB2 bind to domain III and are thus non-overlapping with ADAPT [230]. Affibody molecules are highly structurally similar to albumin-binding domains and it would thus be likely that the two scaffolds target the same epitope.

4.4 ADAPT, a novel scaffold protein-based probe for radionuclide imaging of molecular targets that are expressed in disseminated cancers (Project III)

The ADAPT candidates identified in paper II with high affinity to ERBB2 were encouraging and prompted an investigation to apply them to a practical application. Several
candidates had already been characterized that had lost their inherent affinity to albumin through a few key mutations that did not lead to decreased ERBB2 affinity. One such candidate, referred to as \( ADAPT_{ERBB2-FACS-6} \) in paper 2 had the L37R mutation in helix 3 that led to a loss of albumin affinity as measured by SPR. However, to make sure that the affinity to albumin was completely removed, two more previously well-known mutations, S18A and Y20A were introduced [231]. The new protein was from then on denoted ADAPT6. The high level of ERBB2 expression in many tumors make affinity proteins targeting this receptor attractive for diagnostic applications. Many different Affibody molecules have already been investigated for use as imaging agents and an ERBB2-binding variant has been used for human trials [36]. The ADAPT scaffold is slightly smaller than Affibody molecules with a difference in molecular weight of approximately 1 kDa. This reduction in size make the ADAPT scaffold promising for use as an imaging agent since it might show differences in biodistribution and pharmacokinetics compared to Affibody molecules.

To this end, a cysteine was introduced as well as a histidine-containing tag \((HE)_3\). The cysteine was used for site-specific conjugation of the radionuclide chelator DOTA and the \((HE)_3\)-tag was used for purification and as a modifier of biodistribution. Previous studies have shown that \((HE)_3\)-tags, when used in place of \(H_6\)-tags, reduces the liver uptake of the affinity protein [232]. The two radionuclides \(^{111}\text{In}\) and \(^{68}\text{Ga}\) were used, the former for imaging with a gamma camera and the latter for PET imaging. Imaging with \(^{68}\text{Ga}\) imposes greater requirements on the affinity protein tissue penetration and tumor localization rate due to the short half-life of the radionuclide (68 min) compared with 2.8 days for \(^{111}\text{In}\).

Both radionuclides were successfully chelated by the DOTA conjugates with high yields.
CHAPTER 4. PRESENT INVESTIGATION

Figure 4.7: Imaging of SKOV-3 tumors in mice. (A) Imaging of ERBB2-expressing tumors using \(^{111}In\)-labeled ADAPT6. All three mice carried SKOV-3 xenografts and the right-most mouse was injected with a saturating amount of non-labeled ADAPT6. The imaging was performed using a gamma camera. (B) PET imaging of ERBB2-expressing SKOV-3 tumors with \(^{68}Ga\)-labeled ADAPT6.

\(^{111}In\) and \(^{68}Ga\)-labeled ADAPT6 was also used to generate images of implanted tumors. Figure 4.7A show images generated 1 h after injection of \(^{111}In\)-labeled ADAPT6 into three tumor-bearing mice, two of which carrying SKOV-3 cells expressing high levels of ERBB2 and the third mouse carrying a control LS174T tumor with low ERBB2 expression. The tumors are clearly visible and no background signal can be seen in normal tissue, except for the high expected signal in the kidneys whose defined shape

(>98 \%) resulting in stable radiolabeled protein conjugates that withstood treatment with EDTA without loosing radioactivity. The performance of ADAPT as an imaging tracer was then evaluated, first by assessing the overall biodistribution in normal mice, and then by performing imaging of implanted tumors in BALB/C nu/nu mice. The initial biodistribution experiments showed that the injected ADAPT6 was rapidly eliminated from circulation through the kidneys, consistent with other low molecular-weight proteins. This confirmed that the affinity to albumin had indeed been completely removed by the three introduced mutations (Figure 4.6A). Next, the biodistribution in tumor-bearing mice was evaluated by injecting \(^{111}In\)-labeled ADAPT6. The specific binding to the tumor was also assessed in a separate experiment by first injecting non-labeled ADAPT6 as a binding site blocking agent. Figure 4.6B shows the observed biodistribution in non-blocked and blocked mice, respectively, which shows high uptake of ADAPT6 in the tumor compared to other organs and tissues. The signal in the blocked tumor was significantly lower than in the comparable non-blocked sample. Again, the uptake in the kidneys was very high as expected. The black bars in figure 4.6B correspond to data retrieved from control mice with implanted tumors lacking ERBB2-expression. The uptake in these tumors were comparable to normal tissue, which strengthens the evidence that uptake in ERBB2-expressing tumors was specific.
and known location is easily distinguishable from the tumors. Similar images can be seen in figure 4.7B showing PET images generated with $^{68}$Ga-labeled ADAPT6 1 h after injection. An implanted SKOV3 tumor is clearly visualized together with the kidneys. The high contrast image of the tumor was achieved owing to the rapid elimination of ADAPT6 from circulation and normal tissue.

Taken together, the data show that ADAPT is a suitable scaffold for use in vivo as an imaging agent. The injected ADAPT6 accumulated in ERBB2-expressing tumors rapidly, possibly owing to the small size of the scaffold, allowing for high-contrast images already 1 h after injection.

4.5 Investigating affinity-maturation strategies and reproducibility of fluorescence-activated cell sorting using a recombinant ADAPT library displayed on staphylococci (Paper IV)

The expected outcome of any affinity selection is to only have binders left with affinities that are proportional to the chosen selection stringency. However, this is not always the case. Paper 1 in this thesis is a good example, where more highly enriched binders were found to be of lower affinity than other rarer clones. The problem can be overcome by carefully screening the selection output for desired properties. ELISA or FACS can be employed in this endeavor, depending on the selection method, in order to gauge the equilibrium binding level at a certain target concentration. This is highly laborious, however, and if several libraries are used to select binders to multiple targets at once thousands of clones will have to be individually tested for binding, which is not practically feasible.

The reasons why a selection method would enrich for undesirable clones or generate false positives may be several, but fine-tuning selection parameters and quality-assurance of the library are good first measures. The selection method itself might also be a source for error as certain library members might be toxic to express for the cells, affecting their growth-rates or viability. Measures to control the expression of library members have been implemented in for example phage display and yeast display by using tightly regulated inducible promoters. The staphylococcal display method was used in paper II for affinity maturation purposes. The expression of library members
in the staphylococcal system is not governed by a regulated inducible promoter but is instead constitutive.

In this study, an affinity maturation based on binders in paper I was performed and, in doing so, the staphylococcal display method was evaluated as a directed evolution platform by performing the selection in duplicate and comparing the outputs. Several rounds of selection was performed using a library of roughly $10^6$ protein variants. The first three rounds of FACS selection was performed at equilibrium and the fourth round was performed with an off-rate focused selection. In round 3, the target concentration was varied over three orders of magnitude from 10 nM down to 0.1 nM in order to find an optimal stringency resulting in the highest-affinity clones. FACS was used in all selection tracks and the ERBB3 used as target was fluorescently labeled. Cells displaying the highest fluorescent intensities in relation to their expression levels were sorted in each round, much like the FACS selections in paper II. Figure 4.8A outlines the whole selection scheme where the fluorescent intensities for the analyzed cells are shown in density dot plots. The blue polygons in each dot plot indicate the placement of the sort gates, i.e. only cells with fluorescent intensities within each gate were sorted and amplified for the next round. Figure 4.8B shows the outcome of choosing different selection stringencies in round 3. The lowest target concentration resulted in a lower overall affinity for the output clones. This was consistent for the two selection attempts.

The sequences of the clones in the outputs were further analyzed and compared between the replicate selections. Figure 4.9 illustrates the sequence overlap between the two attempts, i.e. the amount of sequences that were found both times. A clear difference can be seen here as well in regards to the selection stringency. The lower target concentrations in round 3 (1 nM and 0.1 nM) resulted in the lowest sequence overlap, and thus the poorest reproducibility. Round 4 was performed with clones originating from the 10-nM track and showed the highest reproducibility with a high degree of enrichment of a single clone represented by all theoretically possible genetic sequences allowed in the library design (not shown).

As it was not known if the most highly enriched binder in round 4 also had the highest affinity, attempts were made to screen the clones one-by-one. This was done by incubating cells expressing each clone separately with a fixed concentration of ERBB3 and measuring binding intensity at equilibrium. This allowed low affinity clones to be clearly distinguishable from the more higher affinity variants, which also happened to be generally more frequent. However, in order to rank the higher affinity clones more reliable affinity measurements were needed due to the high degree of experimental er-
Figure 4.8: Scheme depicting the selection methodology. (A) The binding profile observed in each selection round is shown and the sorting gates are shown as blue outlines. The arrows indicate selection progression. (B) The clones isolated from round 3 were flow-cytometrically analyzed after incubation with 1 nM labeled ERBB3. Below each plot is a normalized value calculated from the ratio of binding fluorescent intensity over expression fluorescent intensity.
ror observed. SPR was therefore employed for a select group of binders. The affinities ranged from 0.6 to 2.3 nM for clones with the highest affinity. The most highly enriched clone had an affinity constant similar to other binders but it had more favorable dissociation kinetics, reflecting the selection methodology.

The results thus show that the staphylococcal display system is reliable and robust in the isolation of affinity reagents, although the selection stringency should be carefully considered in each subsequent round as too low target concentrations resulted in poor reliability and isolation of lower affinity variants. These results and principles can reasonably be applied to other cell-surface display systems utilizing FACS as well.

![Diagram of affinity data](image)

Figure 4.9: Random clones originating from rounds 3 and 4 were sequenced and the corresponding amino acid sequences were compared between selection attempts. Venn diagrams show the sequence overlap between data sets.

### 4.6 Concluding remarks and future outlook

The albumin-binding domain has been extensively characterized and used in many varied applications. The inherent affinity has been mapped by structural studies [210, 212, 213] and it has also been significantly improved by directed evolution, reaching affinities in the femtomolar range [219]. Its usefulness has been explored and proven in therapeutic applications as a general fusion partner to facilitate a prolonged *in vivo* half-life of protein pharmaceuticals. In this thesis, the albumin-binding domain from streptococcal protein
G has been further explored as a protein engineering scaffold (denoted ADAPT) by introducing an additional affinity to either ERBB2 or ERBB3, while attempting to retain the affinity to albumin. The studies included herein show that this endeavor is possible, even in a small 46 amino acid scaffold consisting of only three helices and connecting loops. However, consistent for all selections is the lack of capability of binding both HSA and the target molecule at the same time. Variants with mutants in scaffold positions were enriched when attempts were made to perform a selection in the presence of HSA. This demonstrates that variants with capacity to support binding to both molecules at the same time were rare, if at all present. The close proximity of the bound HSA molecule to the randomized surface makes it reasonable to assume that any capacity of forming a ternary complex (target-ADAPT-HSA) would be restricted in terms of the number of interacting amino acids available for binding on ADAPT. This also means that mutants with substitutions in the albumin-binding surface, as those observed in paper II, would be favorable in a selection as they could potentially bind with higher affinities. Unintentional mutants are challenging to exclude from a library and would require high fidelity DNA synthesis coupled with a stringent purification scheme to avoid.

The possibility of modifying the affinity to albumin and ERBB2/ERBB3 independently is useful. The affinity to albumin was completely removed in paper III, which allowed us to use the resulting ADAPT molecule as an imaging tracer with suitable pharmacokinetics and tissue penetration. The ERBB2-binding ADAPT presented here demonstrate high tumor-to-healthy tissue contrasts with an ERBB2 affinity in the low nanomolar range. In addition, it was shown to be suitable for PET imaging using $^{68}$Ga. This is a positive indication for future studies considering the lower radioactive exposure to patients that this technique offers. The ADAPT scaffold will be further investigated and improved for use as a general imaging agent also to other targets beside ERBB2.

The question is if a bispecific ADAPT in the current form can be clinically useful in a therapeutic setting. It has been shown that even affinities in the micromolar range to albumin offers a significantly prolonged in vivo half-life [220,225]. If this would be coupled with a high affinity to a target antigen, the bispecific protein could potentially be accumulated in a tissue with high target expression. This alone would not make it favorable over a therapeutic protein combining targeting and albumin-binding without interference, unless penetration into the tumor tissue would be limited by constantly being associated to albumin. The most optimal pharmacokinetics for a given therapeutic presumably varies depending on the situation. Especially in cases where payloads such as toxins or cytostatic drugs are used where prolonged exposure to the toxic compound
could lead to adverse effects. The possibility of binding albumin and an additional target
simultaneously in a single domain remains an attractive prospect that would combine
the benefits of small scaffold proteins such as high stability and simple expression with
a long in vivo half-life comparable to antibodies. This would abolish the need for fusing
a therapeutic protein to ABD, or other half-life extending moieties such as PEG, thus
simplifying development and reducing the risk of negatively affecting the stability, affin-
ity or functionality of the therapeutic protein. The design of a library enabling such a
bispecific domain would have to be carefully considered and the selection methodology
optimized to generate the desired proteins. Prospective targets should also be selected
carefully for potential future studies, possibly favoring smaller proteins.
Acknowledgements

Först och främst skulle jag vilja tacka Vetenskapsrådet, Cancerfonden samt Knut och Alice Wallenbergs Stiftelse för finansiering av dessa projekt. Jag vill också tacka Gålöstiftelsen och Apotekarsocieteten för finansiering av konferensdeltagande.

Jag hade givetvis inte klarat detta själv. Jag skulle vilja tacka:


Alla nuvarande och gamla medlemmar i Hobergruppen; Sara, Sarah, Emma, Johan N, Tove B, Hanna, Jenny, Anna K, Tove A och Johanna S. Tack för alla frukostar och Måladagar!

Johan N, tack för grym handledning och allt bra samarbete i labbet och i skrivandet och att du alltid stöttar och hjälper till även när du är på andra sidan Atlanten! Lycka till med alla dina otroligt ambitiösa forskningsplaner. Det kommer gå hur bra som helst!


Thank you Vladimir and Anna and the rest of the Uppsala crew for a wonderful collaboration. Who knew we would get such nice results? You are a joy to work with and I wish you the best of luck in the future!

My, för kritiskt granskande av denna avhandling och för alla tips och råd.

Per-Åke. Tack för allt intressanta diskussioner genom åren och för att din dörr alltid står öppen. Du har motiverat mig otroligt mycket att ständigt lära mig mer om selektion och bindare. Tack också för all hjälp och feedback i mina projekt!
Alla samarbetspartners och hjälpsamma människor på Affibody: Elin, Caroline, Gunilla, Fredrik och Per. Tack för att jag fick selektera med er under några månader i CD44-projektet. Jag lärde mig mycket som jag kommer ha nytta av länge. Tack också för alla intressanta och roliga möten om nya ABD-bibliotek!


**Human Protein Atlas** på plan 3! Tack för all hjälp med sekvensering och BCA-analys genom åren. Tack också för att ni bidrog med Rosettaceller. Ni är alla fantastiskt hjälpsamma och duktiga!

**Johan S** och **Feifan.** Tack för sällskap hos Affibody och all hjälp med fag display genom åren.

Alla kontorskamrater, gamla som nya: Sara, Emma, Erik, Linnea, Mona, Jossan, Tove, Staffan, Prem och Greg.

Tack **Luciakören** för många roliga timmar! Det har lättat upp höstmörkret under åren.

Tack **Anna-Luisa, Tarek och Josefine** för klättersällskap och coachning. Jag ska ta tag i det igen, jag lovar ;).

Till sist skulle jag förstås även vilja tacka min fantastiska familj för att ni alltid finns där och alla vänner som undrar vad jag har hållit på med så länge. Tack för att ni stått ut med mig under alla stressade stunder.
Bibliography


BIBLIOGRAPHY


