Staphylococcal surface display in directed evolution

Nina Kronqvist

School of Biotechnology
Stockholm 2009
Nina Kronqvist (2009): **Staphylococcal surface display in directed evolution**, School of Biotechnology, Royal Institute of Technology (KTH), Stockholm, Sweden.

**Abstract**

Engineered affinity proteins have together with naturally derived antibodies become indispensable tools in many areas of life-science and with the increasing number of applications, the need for high-throughput methods for generation of such different affinity proteins is evident. Today, combinatorial protein engineering is the most successful strategy to isolate novel non-immunoglobulin affinity proteins. In this approach, generally termed *directed evolution*, high-complexity combinatorial libraries are created from which affinity proteins are isolated using an appropriate selection method, thus circumventing the need for detailed knowledge of the protein structure or the binding mechanism, often necessary in more rational approaches. Since the introduction of the phage display technology that pioneered the field of combinatorial engineering, several alternative selection systems have been developed for this purpose.

This thesis describes the development of a novel selection system based on staphylococcal surface display and its implementation in directed evolution approaches. In the first study, the transformation efficiency to the gram-positive bacteria *Staphylococcus carnosus* was successfully improved around 10,000-fold to a level that would allow cell surface display of complex combinatorial protein libraries. In two separate studies, the staphylococcal display system was investigated for the applicability in both *de novo* selection and affinity maturation of affibody molecules. First, using a pre-selection strategy with one round of phage display, a high-complexity affibody library was displayed on staphylococcal cells. Using fluorescence-activated cell sorting, binders with sub-nanomolar affinity to tumor necrosis factor-alpha (TNF-α) were isolated. Second, a combined approach using phage display for *de novo* selection of first-generation affibody binders and staphylococcal display in a subsequent affinity maturation selection was applied to generate binders with low nanomolar affinity to the human epidermal growth factor receptor-3 (ErbB3). Moreover, in an additional study, the staphylococcal surface display system was improved by the introduction of a protease 3C cleavage sequence in the displayed fusion products in order to facilitate straightforward production of soluble proteins for further downstream characterization.

Altogether, the presented studies demonstrate that the staphylococcal selection system indeed is a powerful tool for selection and characterization of novel affinity proteins and could become an attractive alternative to existing selection techniques.

**Keywords:** affibody, combinatorial library, directed evolution, Gram-positive bacteria, protein engineering, staphylococcal surface display

© Nina Kronqvist 2009
Main references

This thesis is based upon the following papers, which are referred to in the text by their Roman numerals (I-IV). The four papers are found in the appendix.


*These authors contributed equally to this work.

All papers are reproduced with permission from the copyright holders.
Contents

1 Protein engineering ................................................................. 1
2 Affinity proteins ......................................................................... 3
  2.1 Antibodies .............................................................................. 4
  2.2 Antibody fragments ................................................................. 7
  2.3 Alternative scaffolds ............................................................... 9
3 Generation of library diversity .................................................. 14
  3.1 Error-prone PCR ................................................................. 15
  3.2 Oligonucleotide-directed randomization ................................. 17
  3.3 DNA shuffling ...................................................................... 18
4 Selection systems ...................................................................... 21
  4.1 Phage display ................................................................. 22
  4.2 Cell surface display ............................................................. 24
  4.3 Other selection systems ......................................................... 30
5 Staphylococcal surface display ............................................... 32
  5.1 Cell-display principle ......................................................... 33
  5.2 Applications ........................................................................ 35
6 Present investigation ................................................................. 37
  6.1 Optimization of electroporation-mediated transformation: Staphylococcus carnosus as model organism (I) .................... 38
  6.2 A novel affinity protein selection system based on staphylococcal cell surface display and flow cytometry (II) .................... 42
  6.3 Simplified characterization through site-specific protease-mediated release of affinity proteins selected by staphylococcal display (III) .... 49
  6.4 Combining phage and staphylococcal surface display for generation of ErbB3-specific Affibody molecules (IV) ............ 55
  6.5 Future perspectives .......................................................... 62
Acknowledgements ................................................................. 65
References ............................................................................... 68
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABD</td>
<td>Albumin-binding domain</td>
</tr>
<tr>
<td>ABP</td>
<td>Albumin-binding protein</td>
</tr>
<tr>
<td>CDR</td>
<td>Complementarity determining region</td>
</tr>
<tr>
<td>CH</td>
<td>Constant domain of the antibody heavy chain</td>
</tr>
<tr>
<td>CL</td>
<td>Constant domain of the antibody light chain</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-activated cell sorting</td>
</tr>
<tr>
<td>EGFR</td>
<td>Epidermal growth factor receptor (ErbB1, HER1)</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>ErbB3</td>
<td>Epidermal growth factor receptor-3 (HER3)</td>
</tr>
<tr>
<td>Fab</td>
<td>Fragment, antigen binding (Antibody)</td>
</tr>
<tr>
<td>Fc</td>
<td>Fragment, crystallizable (Antibody)</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and drug administration</td>
</tr>
<tr>
<td>GRAS</td>
<td>Generally recognized as safe (FDA designation)</td>
</tr>
<tr>
<td>FcRn</td>
<td>Neonatal Fc receptor</td>
</tr>
<tr>
<td>HER2</td>
<td>Epidermal growth factor receptor-2 (HER2/neu, ErbB2)</td>
</tr>
<tr>
<td>His6</td>
<td>Hexahistidyl</td>
</tr>
<tr>
<td>HSA</td>
<td>Human serum albumin</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>IMAC</td>
<td>Immobilized metal ion affinity chromatography</td>
</tr>
<tr>
<td>K\text{ass}</td>
<td>Association rate constant</td>
</tr>
<tr>
<td>K\text{off}</td>
<td>Dissociation rate constant</td>
</tr>
<tr>
<td>K\text{d}</td>
<td>Dissociation equilibrium constant</td>
</tr>
<tr>
<td>kDa</td>
<td>Kilodalton</td>
</tr>
<tr>
<td>mAb</td>
<td>Monoclonal antibody</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>pAb</td>
<td>Polyclonal antibody</td>
</tr>
<tr>
<td>PCA</td>
<td>Protein complementation assay</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>scFv</td>
<td>Single chain variable fragment (Antibody)</td>
</tr>
<tr>
<td>TNF-alpha</td>
<td>Tumor necrosis factor-alpha</td>
</tr>
<tr>
<td>V\text{H}</td>
<td>Variable domain of the antibody heavy chain</td>
</tr>
<tr>
<td>V\text{L}</td>
<td>Variable domain of the antibody light chain</td>
</tr>
</tbody>
</table>
1 Protein engineering

The developments in protein science were dramatically impacted by the advent of recombinant DNA technology (Linn & Arber, 1968) and the introduction of PCR for amplification of DNA by polymerase chain reaction (Saiki, et al., 1985), making it possible for the first time to isolate the gene for any natural protein, and to produce pure soluble proteins by expression in various host organisms. As a result of the growing number of applications for which proteins could be used, ranging from reagents to therapeutics, a requirement for proteins with physical or chemical properties that are not found in nature became evident and hence the field of protein engineering emerged. It is however no simple matter to generate novel gene sequences with predetermined properties on the protein level due to the complex nature of protein composition and folding. The DNA sequence translates to a peptide sequence built of 20 different amino acids, which in turn is folded into a complex three-dimensional structure. It is not obvious from the primary structure how individual amino acids are arranged in space, neither how they contribute to particular interactions and protein properties. It is nevertheless quite feasible to modify an already known protein by introduction of point mutations at specific positions in the gene sequence through site-directed mutagenesis (Hutchison, et al., 1978, Winter, et al., 1982) and to subsequently characterize the slightly altered protein with partly new properties. Protein engineering is today applied to improve numerous protein properties, including molecular recognition, enzymatic activity, folding or increased stability and solubility. This thesis will however focus on the engineering of protein-protein interactions which include qualities such as affinity and specificity.

There are several approaches for protein engineering in this respect, and they can be divided into two major disciplines: rational or combinatorial engineering. In rational design, a detailed knowledge of the three-dimensional structure and physicochemical properties of proteins is very valuable for prediction of the outcome from changes in
the amino acid sequence. Modifications are typically made as one, or a few, point mutations by site-directed mutagenesis or as insertions, deletions and fusion to other protein genes. Rational engineering has been most extensively used in enzymatic engineering (Bloom, et al., 2005) and also as a method for manipulating the binding interface of affinity proteins. An increasing number of protein structures are becoming available through X-ray crystallography or NMR spectroscopy and computational algorithms are constantly being improved to facilitate predictions of protein-protein interactions (Lippow & Tidor, 2007). When structure information is not available, a possible, yet time consuming, approach to determine the importance of individual amino acids is referred to as alanine scanning mutagenesis, a method where amino acids are individually substituted with the rather small and uncharged alanine, and variants are compared to determine the altered properties (Cunningham & Wells, 1989). By any approach, rational methods have the disadvantage of being highly laborious, and are therefore limited to a small number of mutants which can be designed, produced and characterized.

Combinatorial engineering, has evolved as an alternative to rational engineering and rely on a process that mimics the natural evolutionary principle, where randomly occurring mutations eventually generate proteins with novel functions. Accordingly, these methods are commonly termed “directed evolution”. In principal, combinatorial methods are based on random mutation of a protein to generate a library consisting of a large repertoire of variants, from which proteins with new properties can be isolated using a suitable system for screening or selection (such selection systems are further described in chapter 4). In contrast to rational engineering, this approach circumvents the requirement for available three-dimensional structures, and a large number of protein mutants can be assayed simultaneously in a high throughput manner. In general, the probability of finding variants with improved functions (e.g. the highest affinity) in an assay depends on the complexity of the library. Though, no available selection technique is able to hold the immense amount of protein variants that are generated on the genetic level from a fully randomized protein, combinatorial libraries can often be created with elements of rational design, directing the mutagenesis to certain positions or a certain part of the protein in order to ensure a maximal functionality of the library.
2 Affinity proteins

Biomolecular recognition is the key aspect of most biological processes and relies on specific protein-protein interactions and the capability to discriminate between different molecules. Molecular recognition is the product of several non-covalent amino acid interactions (e.g. hydrogen bonds, electrostatic interactions and van der Waals interactions). These interactions contribute to what we commonly term affinity, more precisely denoted the dissociation equilibrium constant (K_d). When two proteins interact, exemplified by [A] and [B], an equilibrium state will arise between the concentration of free protein and the binding complex [AB]. The equilibrium state formula for this interaction provides the K_d and also the kinetic association and dissociation rate constants k_{on} and k_{off} (Eq 1).

\[
\frac{[A][B]}{[AB]} = \frac{k_{off}}{k_{on}} = K_d \quad (Eq 1)
\]

Affinity proteins have become indispensable tools in biological, biotechnological and medical science as specific reagents for a wide variety of applications, including bioseparation, detection, biopharmaceutics and diagnostics. Antibodies (further described in chapter 2.1) were early discovered as naturally evolved affinity proteins, and are at present the most widely used and examined affinity reagents. The development of protein engineering techniques, and combinatorial engineering in particular, led to the development of a new generation of designed affinity proteins. In resemblance to the antibody molecule, these proteins constitute a stable framework, termed scaffold, with flexible loops or solvent exposed amino acids which are targeted for randomization. One such group of scaffolds is the antibody derivatives, described in chapter 2.2, which are affinity proteins produced as smaller units of full sized antibodies, but with the same binding properties. In addition, several non-antibody
proteins have emerged as functional scaffolds for combinatorial engineering of affinity proteins, as will be described in chapter 2.3.

2.1 Antibodies

Immunoglobulins, more generally termed antibodies, are naturally evolved affinity proteins produced by B cells as part of the humoral response in the immune system of vertebrates, and function as a defense towards toxins and invading pathogens. Foreign molecules (antigens) are recognized by specific antibodies and targeted for elimination by effector cells of the immune system. There is a large repertoire of antibodies displayed on the surface of B cells and remarkably, a reasonably specific antibody can be found to target almost any given antigen to which an organism is subjected to.

Antibody molecules (Fig. 2.1) form a Y-shaped structure comprising two identical heavy chains (H) and two identical light chains (L) linked together by disulfide bonds. Both heavy and light chains consist of a constant (C) and a variable (V) part. The amino-terminal domain of each chain consists of entirely variable parts and comprises the antigen binding properties of the antibody molecule. Within each variable domain reside three small loop-shaped regions called hypervariable regions or complementarity determining regions (CDR1-CDR3). These regions are characterized by a high degree of sequence variation among different antibody molecules and are thus mainly responsible for the varied binding properties. The highly diverse repertoire of antibodies present in the immune system is generated by random arrangements of limited numbers of antibody genes, and further diversification by somatic hypermutation. In addition to specific antigen binding, antibodies are responsible for recruiting appropriate effector cells and proteins for elimination of the encountered antigen. The constant Fc part of antibodies is accountable for mediating such biological effector functions, including binding to gamma immunoglobulin receptors present on natural killer cells and macrophages or activation of the complement pathway (Ward & Ghetie, 1995, Gelderman, et al., 2004). Based on the structure of the constant domains, human antibodies are divided into five major isotypes: IgA, IgD, IgE, IgG and IgM. Among these, IgG is the most abundant isotype in sera (Goldsby, et al., 2003), as a first line of defense against foreign antigens.
With their discovery in the late nineteenth century, antibodies have become extensively used as affinity reagents in a wide range of applications. The ability to bind antigens with high affinity and specificity was early exploited for purification and detection in biological and biotechnological research, and more recently, antibodies have become important agents in medical diagnostics and therapeutics. In addition, awareness of the natural process in which antibodies are diversified and selected inspired the development of methods to generate novel affinity proteins \textit{in vitro} by combinatorial protein engineering.

Traditionally, antibodies have been produced through immunization of animals with an antigen in order to produce a pool of antigen-specific antibodies, which can be isolated from the sera. However, antibodies produced by the immune system are polyclonal as they are expressed by multiple B-cell lines with different amino acid sequences for antigen recognition, and consequently they recognize different epitopes on the antigen. Production of polyclonal antibodies as described is performed rapidly and cost efficiently, but lacks in reproducibility since the obtained pool of antibodies differ between each immunization. In addition, the precise antibody-target interaction is unspecified due to the heterogeneous mixture of antibodies targeting multiple epitopes. Although detection of multiple epitopes by polyclonal antibodies may be advantageous in some applications such as purification and analysis (e.g. immunoblotting, ELISA, immunohistochemistry), therapeutic applications often require the more defined specificity of monoclonal antibodies (mAbs). Reproducible production of natural mAbs became plausible in the mid-1970s when Köhler and Milstein developed the hybridoma technology (Köhler & Milstein, 1975), for which they later were awarded the Nobel Prize. Basically, an antibody-expressing B-cell - typically derived from the spleen of immunized mice - is fused to a myeloma cell to form an immortal hybridoma with the capacity to provide a continuous production of a unique mAb. The hybridoma technology has become widely used for production of mAbs, especially for analytical purposes, but suffers from several limitations for clinical use. Most importantly, the produced antibodies are of animal origin, hence likely to induce an immune response in the human recipient. In order to overcome this limitation, a common strategy is the generation of humanized antibodies by grafting of the CDRs of antibodies produced by hybridoma technology into a human antibody framework (Jones, \textit{et al.}, 1986). Another issue is the low level of control
during the selection process since it occurs in vivo as a response to immunization with an antigen and become dependent on the toxicity and immunogenicity of a chosen antigen and limitations in the natural selection pressure. In addition, the hybridoma technology is quite laborious and does not permit maturation of obtained antibodies to increase properties such as affinity, specificity and stability. Today, recombinant antibodies can be generated in a more controlled manner by generation of diverse combinatorial libraries comprising the variable domains V\textsubscript{L} and V\textsubscript{H} from which specific variants are selected in vitro using phage display or other selection methods (further described in chapter 4).

The dual property of antibodies facilitate antigen binding as well as effector functions and is valuable for certain therapeutic concepts in which they contribute to antigen recognition together with the recruitment of a proper immune response (Adams & Weiner, 2005). Nevertheless, antibodies suffer from several disadvantages that make them less appropriate for certain therapeutic applications. As mentioned, the Fc mediated effector functions are only beneficial in certain therapeutic applications and may as well lead to undesired side-effects (Wu & Senter, 2005). Moreover, the large size of antibody molecules (~150 kDa) lead to a decreased tissue penetration (Graff & Wittrup, 2003, Holliger & Hudson, 2005) as well as a long serum half-life, which is mainly prolonged through recycling by interaction to the neonatal Fc receptor (Ghetie & Ward, 1997), and thus antibodies are less suitable for in vivo diagnostics by medical imaging where extended half lives result in poorer contrast. In addition to the size, antibodies are glycoproteins with a complex multi-domain structure composed of several polypeptide chains, and hence production of correctly glycosylated antibodies is costly due to the need of a mammalian expression system (Jefferis, 2007). These aspects together with the intricate intellectual property rights obstructing commercialization of antibody-derived products have motivated the development of affinity proteins based on antibody fragments and alternative scaffolds.
2.2 Antibody fragments

Even though full-size antibodies have become valuable affinity reagents, there are as mentioned above several limitations associated with large complex structure and the effector functions mediated by the Fc region. Antibodies are comprised of distinct protein domains and it was early understood that enzymatic cleavage of antibodies could be applied to remove parts of the constant region to produce Fab fragments with retained antigen-binding activity (Porter, 1959). Advances in combinatorial protein engineering led to the development of even smaller single-chain antibody fragments (scFv) appropriate for in vitro selection using phage display (McCafferty, et al., 1990) and, as will be further described below, several other fragment formats have been reported (Holliger & Hudson, 2005). At present, most recombinant antibody fragments are produced by combinatorial means from diverse libraries of natural or synthetic antibody genes.

The Fab and scFv fragments (Fig. 2.1) constitute smaller units of the antibody structure with preserved pairing of the variable domains, and an antigen-binding surface equal to the original antibody. While the 55 kDa Fab fragments (55 kDa) are held together by the natural disulfide linkage, scFv fragments (28 kDa) are produced as the smallest entities of the heavy and light chains, held together with a flexible polypeptide linker to increase stability and allow production in E. coli (Bird, et al., 1988, Huston, et al., 1988). In order to further decrease the fragment size, single-domain antibody fragments (15 kDa) consisting of either VH or VL have been thoroughly investigated as an alternative to scFv fragments. Although the use of single-domain fragments initially was hampered by the tendency for aggregation (Holliger & Hudson, 2005), further engineering to improve their stability and solubility has led to successful selections of specific binders (Holt, et al., 2003, Colby, et al., 2004, Jespers, et al., 2004). Interestingly, single-domain antibodies that lack light chains (sdAb) have been found to be the naturally evolved antibody format in two entirely different organisms, camels and cartilaginous fish. Engineered sdAbs have been exploited as affinity scaffolds in several applications (Holliger & Hudson, 2005, Streltsov, et al., 2005, Harmsen & De Haard, 2007).
Fig. 2.1. Ribbon drawings illustrating an IgG antibody molecule 1IGT.pdb (upper), a Fab fragment 1AD9.pdb (lower left) and a scFv fragment 1X9Q.pdb (lower right). The shaded regions represent the binding surface consisting of the six hypervariable loops (CDRs).
2.3 Alternative scaffolds

Inspired by the progresses in combinatorial protein engineering, and the limitations experienced with full length antibodies, as mentioned above, researchers began to investigate other proteins that could serve as alternative scaffolds for randomization and generation of affinity proteins. Although antibodies have evolved as naturally diversified affinity proteins, one must recall that the function of almost any protein involves some specific interaction with other proteins or smaller molecules. For instance, major parts of the cell signaling, cell to cell communications and even interactions between higher organisms, depend on specific molecular recognitions. In order to be useful as an affinity scaffold, a protein must comprise a stable core, commonly termed as framework, which can be expressed at a high yield and preferably in a bacterial system. Some additional attractive features are: a highly stable architecture, recombinant expression as a single polypeptide chain, and a relatively small size (Nygren & Skerra, 2004). Most important, some part of the protein must have the ability to withstand amino acid alterations by diversification, equivalent to the antibody variable regions, without compromising the overall structure and stability. While the mentioned properties are desired for most protein scaffolds, it is equally important to consider the end-product applications. If the intended use of an affinity protein is a therapeutic drug, the protein origin might become a concern in order to avoid immunogenicity – an undesired side-effect due to the recognition of foreign antigens by the human immune system. Moreover, intracellular applications require the absence of disulfide bonds, which are commonly present in proteins to stabilize the tertiary or quaternary structure. If present, the reducing environment inside the cells will break disulfide bonds and potentially compromise the overall function and stability of the scaffold.

At present, there are over 50 proteins reported as potential affinity scaffolds (Nygren & Skerra, 2004, Binz, et al., 2005, Hosse, et al., 2006, Skerra, 2007, Gebauer & Skerra, 2009) and more will certainly be suggested in the future. Protein scaffolds are commonly classified based on the overall structure, the size, or the composition of the binding region. Here, scaffolds are categorized into two groups based on the
properties of the binding surface. The first group includes immunoglobulin-like scaffolds comprising a framework with one, or several, flexible loops which are targeted for randomization. The second group includes scaffolds proteins that are randomized directly on the more rigid, solvent exposed parts of the secondary structure. This section will describe a few examples in each category, and a brief summary of some commonly used protein scaffolds is presented in table 2.1. However, the section will mostly focus on the surface randomized affibody molecule which is the protein scaffold used for generation of affinity proteins in this thesis.

### Table 2.1 Examples of protein scaffolds for generation of affinity proteins

<table>
<thead>
<tr>
<th>Name</th>
<th>Scaffold</th>
<th>Species</th>
<th>Class</th>
<th>Residues</th>
<th>References</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aptamer</td>
<td>TrxA</td>
<td>Bacterial</td>
<td>Single loop</td>
<td>108</td>
<td>(Borghouts, et al., 2005)</td>
<td>Aptanomics</td>
</tr>
<tr>
<td>Adnectin</td>
<td>Fibronectin</td>
<td>Human</td>
<td>Several Loops</td>
<td>94</td>
<td>(Koide, et al., 1998) (Xu, et al., 2002)</td>
<td>Compound Therapeutics</td>
</tr>
<tr>
<td>Avimer</td>
<td>LDLR-A Repeat</td>
<td>Human</td>
<td>Oligomeric n x 40</td>
<td>(Silverman, et al., 2005)</td>
<td>Amgen</td>
<td></td>
</tr>
<tr>
<td>Darpin</td>
<td>Ankyrin Repeat</td>
<td>Human (designed)</td>
<td>Secondary structure</td>
<td>67 + n x 33</td>
<td>(Binz, et al., 2004)</td>
<td>Molecular Partners</td>
</tr>
</tbody>
</table>

#### 2.3.1 Immunoglobulin-like scaffolds

Antibodies and antibody fragments comprise a rigid β-sheet framework with diversified loops for antigen recognition. This architecture has proven to be successful for holding highly diversified libraries of natural antibodies as well as in vitro generated antibody fragments. Thus, several other proteins with similar structural features have been investigated and utilized as affinity scaffolds in combinatorial engineering. One example is the 10Fn3 scaffold that has been derived from the 10th domain of
fibronectin type III (Koide, *et al.*, 1998), which is a 15-domain repeat protein involved in a wide range of cellular processes through its binding to various macromolecules. 10Fn3 (also referred to as adnectin or monobody) is a 10 kDa monomeric domain consisting of seven strands and six loops assembled into a stable β-sandwich structure, which resembles the variable domain fold and CDR-loops of antibodies. However, this structure is independent of disulfide bonds for folding and stability. Initially, phage display was used to successfully select binders subjected to randomization on two of the flexible loops (Koide, *et al.*, 1998). Since then, combinatorial libraries are typically generated by randomization of three loops, and binders have been selected against several targets, typically using mRNA display (Xu, *et al.*, 2002) and yeast display (Lipovsek, *et al.*, 2007).

A second example belonging to the immunoglobulin-like group is the lipocalin scaffold (Beste, *et al.*, 1999, Skerra, 2000). Lipocalins are a family of proteins involved in transport or storage of mainly small and hydrophobic molecules such as steroids and lipids. These approximately 20 kDa proteins comprise an eight stranded anti-parallel β-barrel structure with four connecting loops forming the ligand binding site. In contrast to the adnectin scaffold, lipocalins usually have one to two disulfide bonds although these are not necessary for protein folding and/or stability. The lipocalin scaffold has been used for generation of specific binders, i.e. anticalins, against small molecules (haptens) but also for selection of binders against larger proteins (Schlehuber & Skerra, 2005).

### 2.3.2 Surface-randomized scaffolds

In contrast to the antibody mimicking scaffolds mentioned above, the second group of binding proteins is generated by randomization of surface exposed residues resting directly on secondary structure. Normally, this strategy targets a surface or secondary structure elements known for its interaction with other proteins. Although novel binding surfaces can be generated *de novo* by randomization of any solvent exposed area of a protein, the likelihood for the protein framework to withstand the amino acid modifications, and maintain stability, is assumed to be higher upon randomization of naturally occurring binding regions which have been subjected to several substitution events during natural evolution. Scaffolds belonging to this group
typically have rather flat binding-surfaces, which provide slightly different binding properties compared to immunoglobulin-like scaffolds. In one respect, the constrained binding surface is ideal due to low losses of entropy during the binding event. On the other hand, the low flexibility is less suited for interactions with small molecules, peptides and protein epitopes within poorly exposed binding pockets or grooves. This section will describe two surface-randomized scaffolds generated from similar structural elements: ankyrin repeat proteins and affibody molecules, with a certain focus on the latter.

The 33-residue ankyrin repeat protein is one of nature’s most common protein interacting motifs found in a number of eukaryotic proteins. Ankyrin repeat domains usually consist of four to six repeats of one β-turn and two anti-parallel α-helices and lack disulfide-bonding cysteines. Designed ankyrin repeat proteins (DARPins) have been generated by partial secondary structure randomization of the β-turn and the adjacent α-helix, while preserving the repeat properties (Forrer, et al., 2003). High-affinity DARPins have been selected using ribosome display (Binz, et al., 2004), and more recently by a modified phage display protocol that circumvents the low display levels due to the fast folding of ankyrin repeat proteins (Steiner, et al., 2006).

Affibody molecules
Affibody molecules originate from the staphylococcal protein A (SpA), which is displayed on the *S. aureus* cell surface and involved in host infection through its binding to the Fc and Fab region of IgG molecules. SpA contains five homologous domains (E, D, A, B, and C), each with Ig-binding properties directed to different species and subclasses (Moks, et al., 1986). Hence, the full-length SpA has been extensively used for purification and detection of antibodies (Ståhl & Nygren, 1997, Hober, et al., 2007). In order to produce a smaller affinity reagent with retained IgG binding properties, the domain B was early isolated and engineered at two amino acid positions in order to increase resistance to hydroxylamine cleavage (Nilsson, et al., 1987). The improved domain (denoted Z) demonstrated preserved binding to the Fc part human IgG1, with an affinity of 10-60 nM (Jendeberg, et al., 1995). However, the affinity to the Fab region was significantly lower compared to the non-engineered domain B (Jansson, et al., 1998).
The Z domain consists of a single polypeptide, folding into a compact structure of three anti-parallel α-helices (Fig. 2.2). This rather small domain (6.5 kDa) has a high stability and solubility without the presence of cysteine bonds, and is easily produced in prokaryotic hosts at high yield and low expense. Furthermore, the small size and rapid folding kinetics allow for solid-phase peptide synthesis in order to produce high purity proteins and to introduce functional groups at specific positions (Engfeldt, et al., 2005). The proven binding capacity of the Z domain and the traits mentioned above form an ideal basis for combinatorial engineering of the binding surface to generate mutants with affinity for other proteins. Initially, a library of $4 \times 10^7$ variants of Z domain mutants was generated by randomization of 13 solvent exposed residues of helices one and two (including nine residues which participate in the original Fc interaction) using synthetic oligonucleotides produced by NNG/T degenerate codons (described in chapter 3.2). From this library, binders (termed affibody molecules) with micromolar affinities against Taq DNA polymerase, insulin, and apolipoprotein A-1 were isolated by phage display (Nord, et al., 1997). More recently, a larger library of $3 \times 10^9$ variants displayed on phage (Grönwall, et al., 2007) has been used for selection of binders with nanomolar affinities against various targets, for example EGFR (Friedman, et al., 2007), HER2 (Wikman, et al., 2004) and amyloid beta peptide (Grönwall, et al., 2007). In addition, binders with even higher affinity (down to 20 pM) have been isolated by affinity maturation selections using second-generation affibody libraries (Gunneriusson, et al., 1999a, Nord, et al., 2001, Orlova, et al., 2006, Friedman, et al., 2008). Affibody molecules have been investigated for a wide range of applications including separation, diagnostics and in vivo tumor imaging/therapy, reviewed more extensively elsewhere (Nygren, 2008, Friedman & Ståhl, 2009, Grönwall & Ståhl, 2009).

Fig. 2.2 The α-helical affibody protein scaffold. Combinatorial affibody libraries have been constructed by randomization of 13 amino acid residues (red) on helix one and two.
3 Generation of library diversity

Combinatorial libraries constitute a repertoire of protein mutants which differ more or less on the genetic level and consequently in the amino acid composition. Library diversity may be obtained from a natural source such as antibody genes from B-cells (described in chapter 2.1). Antibody libraries from natural sources are either naïve and derived from the natural non-immunized repertoire, or immune, due to a previous immunization with an antigen. In addition to antibodies, smaller protein libraries can be obtained from sets of closely related homologues. However, synthetic libraries have become the most frequently used in combinatorial protein engineering due to the high flexibility and the growing number of alternative scaffolds lacking a natural process for diversification. A crucial first step to generate a synthetic diversified gene library is to select an appropriate mutagenesis strategy. There is a large variety of techniques to introduce alteration to a gene sequence in a random manner, which are essentially categorized into random mutagenesis or recombination methods. Random mutagenesis is applied to introduce point mutations into a gene, either in an entirely random mode (e.g. error-prone PCR) or by randomization of pre-determined positions in a semi-rational manner (oligonucleotide-directed randomization), and these methods will be further discussed in chapters 3.1 and 3.2, respectively. Recombination methods mimic the evolutionary process of exon shuffling and are often applied as a complement to random mutagenesis in order to further increase library diversity or on natural families of related (homologous) genes. DNA shuffling, the most commonly employed recombination method will be described in chapter 3.3. The methods discussed in this chapter are summarized in Fig. 3.1.
The available methods for random mutagenesis all have their pros and cons depending on the application. As mentioned in chapter 1, aiming for a very high diversity may only result in a loss of coverage during selection, and eventually the applicability of a mutagenesis method is highly dependent on the nature of the protein targeted for randomization (e.g. protein size and binding properties) as well as the selection technique.

3.1 Error-prone PCR

Since the introduction of the PCR technology for rapid amplification of DNA fragments, entirely random diversifications of gene libraries are almost solely generated by error-prone PCR (epPCR). Some alternative methods for introducing random point mutations to a given gene sequence are *in vitro* or *in vivo* exposure to chemical agents, such as ethyl methanesulfonate (EMS), nitrous acid or hydroxylamine, or the use of mutator strains with defects in the DNA repair machinery (Nguyen & Daugherty, 2003). However, *in vivo* mutagenesis (i.e. whole-cell mutagenesis) is indiscriminate because mutations cannot be directed to a specific gene of a vector construct, leading to adverse affects on protein expression, plasmid stability and cell survival. Today, the preferred choice for diversification of large libraries by the use of chemical agents would be *in vitro* mutagenesis of PCR amplified gene fragments for successive sub-cloning to an expression vector. Still, chemical mutagens must be handled with great care and a proper risk assessment is required. In addition to the obvious limitations to chemical and biological methods, the overall mutation rates generally are below those obtained by epPCR, which was shown in a recent study comparing random mutagenesis by hydroxylamine treatment, *Escherichia coli* mutator strain and epPCR (Rasila, et al., 2009).

In principle, epPCR is a polymerase chain reaction amplification of a gene fragment performed under mutagenic reaction conditions in order to increase the natural error-prone properties of DNA polymerases. In theory, low fidelity polymerases, such as Taq DNA polymerase, could be used in standard PCR to introduce a certain degree of mutations to a gene sequence due to the lack of proof-reading activity, which result in an error rate of about 1 in 9000 bases (Tindall & Kunkel, 1988). In practice, however,
the accuracy is still too high for construction of diverse gene libraries and several protocols have been developed in order to increase the rate in which bases are misincorporated by DNA polymerases. Mutagenic reaction conditions are typically achieved by the addition of small amounts of Mn$^{2+}$ or increased amounts of the Mg$^{2+}$ cofactor together with unequal amounts of dNTPs in the reaction mixture (Cadwell & Joyce, 1994, Cirino, et al., 2003). The mutation rate correlates to the reaction conditions as well as template concentration and the number of PCR cycles.

Random mutagenesis using epPCR is affected by several bias issues, and one must be aware of their impact on the final library diversity. First, for most DNA polymerases the base-pair substitutions are biased towards transitions, where structurally similar bases are exchanged (i.e. A-G, T-C) rather than transversions, where purines are exchanged for pyrimidines. However, a few available DNA polymerases are more biased towards transversion events, which provide the means to slightly control the transition to transversion ratio by using combinations of different polymerases during the epPCR reaction (Rowe, et al., 2003). Second, there is a codon bias due to the natural degeneracy of the genetic code. There are 61 existing codons translating to 20 amino acids and three stop codons. The amino acids serine, leucine and arginine are specified at a high redundancy by six different codons respectively, while methionine and tryptophan are only specified by one codon each. This results in possible stop codons and an uneven distribution of amino acids produced by a single point mutation in epPCR (Neylon, 2004). Additionally, a single point mutation in a gene allows only 5.7 different amino acids to be translated (Hermes, et al., 1990). Last, there is a natural bias attributed to mutagenesis during PCR amplification, since mutations introduced in an early PCR cycle will be over-represented in the end product. This problem is usually circumvented by producing a pool of multiple amplification reactions performed with a decreased number of cycles. Nevertheless, as for all bias issues, statistical analysis becomes increasingly complicated.
3.2 Oligonucleotide-directed randomization

In some library applications it may be advantageous to randomize only amino acids which are found to be solvent exposed and unimportant for protein stability and folding. Oligonucleotide-directed randomization by synthesis, allows for a semi-rational approach where only specific positions are subjected to random mutagenesis. Accordingly, the level of diversification is highly controlled, both regarding to the amino acid identities and randomized positions.

One of the most established techniques for oligonucleotide-directed randomization utilizes sets of mixed mononucleotides in the synthesis reaction to generate a pool of degenerate oligonucleotides. The codon in a specific position can either be fully randomized through synthetic oligonucleotides, randomly incorporating any of the four bases (NNN), or by restricting the mixture of bases in the last position (NNT/C, NNG/T, NNG/C and NNT/C/G). However, mononucleotide synthesis is susceptible to the same codon bias as epPCR, as mentioned in chapter 3.1, due to the degeneracy of the genetic code. Hence, at any randomized position in the pool of oligonucleotides, there is a bias towards certain amino acid residues and a potential for undesired stop codons. The frequency of introduced stop codons is slightly reduced by restricted mixtures of bases, for instance NNG/T randomization, which introduces one stop codon in every 32 sequences, compared to 3 stop codons in every 64 sequences introduced by NNN randomization (Neylon, 2004). Even so, the final complexity and functionality of the product will be much lower than anticipated.

In order to circumvent the bias issues of mononucleotide synthesis, several techniques have been developed for synthesis using defined sets of triplet nucleotides, coding for specified amino acids. One strategy is to prepare trinucleotide phosphoramidities, representing defined codons for each of the 20 amino acids (Virnekäs, et al., 1994). While randomizing one amino acid position, a mixture of 20 codons, or subsets of desired codons, are added in one step during synthesis. Furthermore, to obtain a higher control of the level of diversification, a combined synthesis approach using Fmoc-protected trinucleotides together with DMTr-protected mononucleotides has
been devised (Yanez, et al., 2004). However, trinucleotide phosphoramidities are
difficult and expensive to produce, although progress has been made for large-scale
solid phase synthesis (Kayushin, et al., 2000). Trinucleotide building blocks for
synthesis of randomized oligonucleotides eliminates the bias issues common for
mononucleotides, and offer a high control over the ratio of amino acids in any
position.

The Slonomics™ technique (Van den Brulle, et al., 2008) is another elegant approach
to produce non-biased randomized libraries, exploiting the advantage of triplet-based
building blocks, while avoiding any requirement of presynthesized trinucleotides. In
this technique, universal building blocks are synthesized as single stranded
oligonucleotides forming double-stranded hair-pin structures having trinucleotide
single-stranded overhangs. During production, defined sets of triplet building blocks
are assembled on a solid-phase by sequential ligation and restriction in a highly
automated and standardized procedure. This novel randomization approach was used
for the library generation in paper IV included in this thesis (described in chapter 6.4)

3.3 DNA shuffling

In eukaryotic organisms, recombination of genes is a natural evolutionary process
referred to as exon shuffling, in which domains or exons are recombined to generate
proteins with novel functions. A wide range of methods have been developed to
mimic the natural recombination process in vitro. This chapter will focus on the most
common method referred to as DNA shuffling or sexual PCR while other methods
are reviewed elsewhere (Lutz & Patrick, 2004, Neylon, 2004). Willem Stemmer first
described DNA shuffling as a method for directed evolution by recombination of
closely related gene fragments (Stemmer, 1994a) and it has become an important tool
for generation of library diversity. DNA shuffling is performed in a three-step
process in which a pool of related genes is randomly fragmented by DNase I,
reassembled by multiple cycles of PCR without added primer but in the presence of
DNA polymerase, and amplified with conventional PCR with primers to generate full-
length chimeric genes. A natural source of related sequences is the genes encoding
natural homologues, which have the advantage of being functional proteins and thus
provide a “functional diversity”. An alternative source is synthetic combinatorial gene libraries generated by random mutagenesis such as error-prone PCR or oligonucleotide-directed randomization (described in chapter 3.1-3.2). DNA shuffling of such libraries is either performed in order to increase the diversity generated by random mutagenesis or more commonly as a rapid tool for affinity maturation between selection cycles. *In vitro* recombination has the advantage of being the only method which allows beneficial mutations to be brought together in a single gene. In addition, most produced variants are functional since deleterious or neutral mutations can be eliminated by backcrossing with an excess of the wild-type gene (Stemmer, 1994b). The DNA shuffling procedure does not only have the power to recombine gene fragments, but is in addition usually performed with error-prone PCR during amplification of the recombined chimeric genes. Random point mutations are consequently introduced at a slow rate, which further adds to the final diversity. The mutation rate is mainly controlled by the fidelity of the DNA polymerase (Barnes, 1994).

Antibodies represent a diverse pool of natural functional homologues and hence, libraries of antibody fragments are particularly suited for diversification by DNA shuffling. It was early shown that a scFv library could be created by random combination of the separate V<sub>L</sub> and V<sub>H</sub> genes isolated from peripheral blood lymphocytes (Marks, *et al.*, 1991). Random recombination was later shown to be applicable also for synthetically randomized antibody genes when Cramèri and co-workers successfully applied DNA shuffling on synthetically mutated CDRs for generation of a large scFv library from which antibodies specific for the human granulocyte colony-stimulating factor receptor were selected (Crameri, *et al.*, 1996).
Fig. 3.1 Overview of common methods for generation of library diversity in combinatorial protein engineering. Starting with the protein encoding gene (top), randomization is performed to generate a library of different genes which translates into proteins with an altered amino acid composition (bottom). The figure illustrates one possible randomization event although a typical randomized library would comprise several billions of variants. Oligonucleotide-directed mutagenesis introduces changes at one or several pre-determined positions in the gene. Error-prone PCR, belonging to the completely random methods, introduces changes randomly throughout the gene sequence. DNA shuffling is performed on a set of already diversified genes in order to bring them together in novel combinations. The starting material can either be a library randomized by any of the mentioned methods or a set of naturally derived homologous genes.
4 Selection systems

The natural evolutionary process of diversification, selection and amplification of protein variants with favorable traits (e.g. high affinity), constitute the essence of combinatorial protein engineering. As described in chapter 3, there are several methods for generation of diversity that allow scientists to construct protein mutant libraries with higher complexities than ever seen in nature. However, there is a subsequent need for high-throughput assays in order to discriminate between different protein variants and isolate those with the highest affinity. Such technologies for affinity protein selection from combinatorial libraries are termed selection systems.

Since there is no available method for amplification of proteins themselves while both sequencing and amplification of DNA are straightforward techniques, all selection systems have the common feature of physically linking the phenotype of individual proteins to their corresponding genotype. Some properties which are desired in a selection system are: 1) functional display of the combinatorial protein library, 2) possibility to construct large libraries in order to increase the likelihood of finding high-affinity binders, 3) selection pressure focused on affinity between proteins and their target, 4) easy isolation and amplification of clones between selection rounds, 5) high-throughput. Furthermore, some features are advantageous although not essential and include the possibility to further diversify during selection, automation options and monitoring possibilities. Even though the main purpose of any selection system is the isolation of high-affinity binders, systems with means for rapid and easy characterization of selected candidates are beneficial since selections typically generate large amounts of binders with different affinities and protein traits.
4.1 Phage display

The most intuitive approach to obtain a physical linkage between the genotype and the phenotype for individual protein mutants would be by utilizing the natural machinery of microbial organisms for translation and display of solvent exposed proteins. This was indeed shown to be a realistic strategy over two decades ago, when George Smith demonstrated that peptides could be displayed on the surface of filamentous bacteriophage M13 through genetic fusion to the native exterior coat protein pIII and that peptide displaying phages could be enriched for using immobilized peptide-specific antibodies (Smith, 1985). This was a major leap towards what we today entitle phage display, a technology which has become one of the most established for high-throughput screening and selection from large combinatorial libraries. The first proof-of-concept experiments using phage display as a selection system were performed by screening and selection from libraries of peptide mutants displayed by phages to isolate variants with properties matching antibody epitopes (Cwirla, et al., 1990, Scott & Smith, 1990) or with streptavidin binding activity (Devlin, et al., 1990). Soon thereafter, Jim Wells and co-workers developed the phagemid system, which will be described below, and demonstrated that phage display was applicable also for larger polypeptides by the successful display of two variants of a 191 residue human growth hormone and subsequent selection of the variant with highest affinity for the hormone receptor (Bass, et al., 1990). Phage display was early recognized as a powerful tool for the generation of affinity proteins which could substitute the time consuming hybridoma technology, which was the standard method for monoclonal antibody production, and soon successful selections were performed using phage displayed libraries of antibody fragments, such as scFv (McCafferty, et al., 1990) and Fab (Garrard, et al., 1991, Hoogenboom, et al., 1991). Over the years, the phage display technology has been subjected to constant improvements and has become an important method for selection of antibodies as well as affinity proteins based on non-immunoglobulin scaffolds (Rader & Barbas, 1997, Bradbury & Marks, 2004, Sergeeva, et al., 2006).
The filamentous bacteriophages are a family of bacterial viruses that specifically infect gram-negative bacteria and the Ff group of this family (including M13, fd and f1) is known to exclusively infect *E. coli* cells bearing F-pili. Pioneered by Smith and co-workers, the filamentous phage M13 has become the most extensively used for phage display selections although systems using other viral species, such as T7 and lambda bacteriophages, have been described for this purpose (Mikawa, et al., 1996, Houshmand, et al., 1999). The rod-shaped bacteriophage M13 has a filamentous capsid enclosing a circular single-stranded DNA molecule. The viral capsid is formed primarily by ~2700 copies of a major coat protein pVIII, which assembles into a rod-shaped structure, and ~5 copies each of the four coat proteins pIII, pVI, pVIII and pIX, positioned at the ends of the phage particle (Webster, 2001). The minor coat protein pIII has an important function as it is responsible for cell entry by attaching to F-pili of the *E. coli* host. The principle of phage display is the recombinant fusion of a protein to a capsid protein, which in theory would be feasible for any of the capsid proteins present on the phage surface. Still, most protein libraries have been created by fusion to the coat proteins pIII and to a certain extent pVIII. Incorporation of a foreign protein directly into the phage genome facilitates the display of proteins in a multivalent manner, especially for the more abundant major coat protein pVIII. However, the capsid proteins have been demonstrated to only tolerate insertions of small peptides in order to maintain their natural functions (Kretzschmar & Geiser, 1995) and also, manipulation of all five pIII proteins on an individual phage may affect the ability of phages to infect the host, which is critical for the phage replication cycle. In addition, multivalent display itself might generate avidity effects during selection. Monovalent display on the phage surface was accomplished by the introduction and use of phagemid vectors (Bass, et al., 1990, Smith & Petrenko, 1997). The phagemid vector contains the gene sequence encoding the recombinant protein fused to the wild-type capsid protein but lacks the components for phage replication and assembly. Infection of *E. coli* cells harboring the phagemid using wild-type phages, i.e. helper phages, allows the production of phage particles consisting of both wild-type and mosaic proteins. Phage display using the phagemid system has become the strategy of choice for monovalent display of larger proteins. Additionally, the use of an *E. coli* host, to which phagemid vectors are typically transformed by electroporation, permit the creation of protein libraries with a relatively high complexity and phage displayed antibody libraries comprising over 10^10 individual
variants have been reported (Sblattero & Bradbury, 2000, Hoet, et al., 2005). In this thesis, phage display of affibody molecules have been used in two of the presented papers (II and IV).

The selection procedure using phage display (also referred to as biopanning) starts with transformation of a library of phagemid vectors to \( E. \text{coli} \) cells followed by infection with the helper phage to initiate production of phage particles. The phage library is harvested and incubated with the target molecule in solution or on solid-phase (e.g. paramagnetic beads) in order to allow specific interactions with phage displayed library variants. After a series of washing steps, bound phages are eluted and used to infect new \( E. \text{coli} \) cells for amplification of selected clones by growth. Biopanning is typically performed in three to five cycles to enrich clones having the highest affinity to the target. Normally, the selection stringency is increased in each selection round by an increased number of washes and/or a decreased target concentration. After the selection procedure, isolated clones are identified by sequencing and usually screened in an ELISA to verify antigen binding (Bradbury & Marks, 2004). In order to further characterize binding and biophysical properties, genes of candidate clones are typically sub-cloned into an expression vector for production of soluble protein.

4.2 Cell surface display

Although microbial display for selection from combinatorial libraries was first developed on filamentous phages, cell surface display systems employing yeast, bacteria and mammalian cells are becoming attractive alternatives for isolation of novel affinity proteins. Similar to phage display, cells provide a physical linkage between the phenotype and genotype by harboring plasmid DNA inside the cell, while displaying the corresponding proteins on the cell surface. Cell display systems are however advantageous by two key features: the larger size and the multivalent display of recombinant proteins. First, the large size of cells compared to phages allows cells to be detected with flow cytometry, and hence, fluorescence activated cell sorting (FACS) can be applied for high-throughput screening and sorting of cells binding to fluorescently labeled target protein. In contrast to the “blind” capture and elution
principle of phage display, FACS sorting provides a real-time monitoring and visualization of each individual cell together with its target binding capacity. Second, the multivalent display (i.e. display of multiple copies) allows for a quantitative measurement of the affinity. The target protein is usually labeled with a fluorophore and incubated with the cell-displayed library in solution. Since the recombinant proteins are typically expressed in fusion with a reporter protein, the uneven expression levels on individual cells can be normalized with respect to the target binding signal by simultaneous two-color detection. The obtained data for each cell displayed library member therefore represent an accurate measurement of the relative affinity. The possibility to quantitatively measure the relative affinity during selection, as well as to characterize selected binders on-cell, is an exclusive feature for cell display systems and offers an advantage compared to alternative selection systems.

Cell-based selection techniques follow some basic steps (Fig. 4.1). First, a plasmid library created in a host-specific expression vector is transformed to the cells, typically by electroporation. Next, the cell-displayed protein library is incubated with both target protein and reporter binding protein that are labeled with differently colored fluorophores. The stringency is controlled by target concentration, referred to as equilibrium screening, or by variations in the dissociation time, referred to as kinetic screening. The stringency can further be controlled with several built-in sorting parameters of the flow cytometer. During FACS sorting, the relative affinities are monitored in real-time and a sorting gate can be selected to isolate a chosen percentage of clones having the highest relative affinities. Between sorting rounds, isolated clones are easily amplified by cell growth. Last, when a set of candidate clones is obtained, the binding constants can be determined by incubation with target protein at a series of concentrations.

The development and implementation of a cell display selection system utilizing the gram-positive bacterium *Staphylococcus carnosus* was the objective for the research which this thesis is based on (extensively described in chapter 5 and 6). A number of other microbial cells have been used for surface display of recombinant proteins (Lee, *et al.*, 2003, Wernérs & Ståhl, 2004), but only a few have been successfully applied in combinatorial approaches. This chapter will focus on two cell display systems which have been extensively used in combinatorial protein engineering: the yeast strain.
Staphylococcal surface display in directed evolution

*Saccharomyces cerevisiae* (Gai & Wittrup, 2007) and the gram-negative bacterium *Escherichia coli* (Daugherty, 2007).

**Fig. 4.1** Schematic overview of the cell display selection procedure. Combinatorial libraries are typically subcloned to an expression vector and transformed to a particular species for subsequent display on the cell surface. Following incubation with fluorescently labeled target protein, library members showing the highest target-binding signals are isolated using FACS sorting. Retained variants are amplified by growth and subjected to several rounds of sorting until a required enrichment is obtained.
4.2.1 Yeast surface display

In the late 90’s, Wittrup and co-workers were the first to describe yeast surface display for applications in combinatorial protein engineering (Boder & Wittrup, 1997, Gai & Wittrup, 2007). This was one of the first microbial selection systems developed as an alternative to the phage display technology, and has since then become one of the most widely and successfully used selection techniques besides phage display and ribosome display (further described in chapter 4.1 and 4.3 respectively).

The yeast display system developed by Wittrup utilizes C-terminal fusion of library proteins to a native agglutinin receptor (aga2p) for recombinant display on the surface of the yeast strain *Saccharomyces cerevisiae*. As for all cell-based systems, yeast cells display multiple copies of the recombinant proteins (~10^5 copies per cell). However, the expression level differs between individual cells and hence, the signal obtained from labeled target protein is reflecting both affinity and level of expression. Therefore, the recombinant proteins are also fused to an hemagglutinin tag (N-terminal) or a c-myc tag (C-terminal) in order to normalize the antigen binding signal to the surface expression levels using tag-specific antibodies (either directly labeled or detected by labeled secondary antibodies). The C-terminal c-myc tag also provides means to determine the level of full-length recombinant proteins displayed by the cell population. The advantages of the yeast display is the possibility of using the powerful FACS instrumentation during the selection process together with the multivalent display for monitoring of relative affinities, as mentioned in the introduction to this chapter. Furthermore, yeast cells provide eukaryotic secretory processing during expression, which might be advantageous for expression of some mammalian proteins requiring disulfide bonds and/or glycosylations in order to be functional.

Since the transformation efficiency of yeast is much lower compared to *E. coli*, the final diversity of a yeast-displayed library typically is around 10^7 members. For this reason, yeast display has been most successfully applied in affinity maturation approaches where a large diversity is less important. In an early study, yeast surface display was used for affinity maturation of a high-affinity scFv antibody binding to a
fluorescein-biotin, and a 1000-fold improvement to low femtomolar affinity was reported (Boder, et al., 2000). Since then, yeast display has been applied for affinity maturation in a number of studies (van den Beucken, et al., 2003, Rajpal, et al., 2005, Razai, et al., 2005, Weber, et al., 2005, Wang, et al., 2007), and in two additional selections, the reported increase in affinity was as high as 200,000 (Jin, et al., 2006) and 3,000,000-fold (Buonpane, et al., 2007). Although a really large diversity is difficult to obtain, yeast display has also been successfully implemented for de novo selections. The largest reported library displayed on yeast (~10⁹ members) was constructed from non-immune human scFv antibody fragments by Wittrup and co-workers in 2003, and was used for FACS selection of scFv fragments with nanomolar affinity against several target proteins (Feldhaus, et al., 2003). In later years, other de novo selections have been described using both scFv (Lee, et al., 2006) and Fab fragments (Weaver-Feldhaus, et al., 2004).

Yeast cells exhibit some interesting features that might be advantageous in order to generate or increase library diversity. First, yeasts have the ability for in vivo homologous recombination of DNA with high sequence similarity. This property was utilized by co-transformation of two scFv genes with 89.8 % homology to produce a chimeric library of 10⁷ members (Swers, et al., 2004). However, recombination events might also be disadvantageous due to a lower control of the functional diversity of a yeast displayed library. Second, smaller sub-libraries can be generated in haploid yeast strains and combined by mating to produce diploid yeasts with significantly higher diversity, thus circumventing the low transformation efficiency. By this approach, the heavy and light chains from a repertoire of Fab fragments were separately transformed into opposite mating strains a and α respectively, and mated to produce a haploid yeast displaying randomly combined Fab fragments on the cell surface (Blaise, et al., 2004, Weaver-Feldhaus, et al., 2004).

### 4.2.2 E. coli surface display

Although yeast display was the first cell-based system to be truly implemented for selection of affinity proteins using combinatorial engineering, the prospect of using surface display on the gram-negative bacterium E. coli for the same purpose was investigated already in the early 90’s. After some early attempts using short peptide
libraries, Georgiou and co-workers demonstrated that a single scFv fragment, specific for a hapten molecule, could be functionally displayed on the *E. coli* cell surface and enriched by FACS sorting from an excess of control cells (Francisco, *et al.*, 1993). In 1998, an affinity maturation library (~10^6 members) of scFv fragments displayed on *E. coli* cells was used to isolate fragments with a three-fold higher affinity (Daugherty, *et al.*, 1998).

In *E. coli* surface display systems, the recombinant protein is fused to a natural surface protein capable of transporting the protein through the periplasmic space and across the outer membrane. However, the translocation of larger and more complex proteins have proven to be challenging. Although a number of natural surface proteins have been investigated for outer-membrane display, only a few scaffolds have been successfully displayed by this strategy (Daugherty, 2007). In order to improve the *E. coli* display system, Georgiou and co-workers developed the anchored periplasmic expression (APEx) technology which was used to isolate a scFv fragments with 200-fold increased affinity (Harvey, *et al.*, 2004). By this method, recombinant proteins are anchored into the periplasmic space of the cells and hence, translocation through the outer membrane is circumvented. Upon selection, interaction with target protein is facilitated by disruption of the outer membrane to produce spheroplasts with solvent exposed recombinant proteins. However, the spheroplasts have a rather low viability during FACS sorting. In contrast to yeast display or outer-membrane *E. coli* display, systems where selected clones are readily amplified by cell growth, the APEx method requires PCR rescue, sub-cloning to the display vector and transformation to new *E. coli* cells between each selection cycle. The APEx technology was further developed into a similar technique which allows display and selection of full-length antibodies, termed *E*-clonal antibodies. In principle, the system allows full-length heavy and light chains to be transported into the periplasm, followed by assembly and capturing by an Fc binding protein anchored to the inner membrane (Mazor, *et al.*, 2007).
Apart from the microbial display systems described in the previous sections of this chapter, several other selection systems have been developed and successfully implemented. In this section, one of the most common cell-free display systems (ribosome display), and one example of a non-display system (protein complementation assay, PCA) will be briefly discussed.

Ribosome display was first described in the mid-90’s (Mattheakis, et al., 1994) and has since then become one of the most extensively used cell-free selection systems. In this method, the DNA sequence coding for the protein of interest is fused to a ribosome binding site, and in addition, the stop codon which normally terminates translation of the gene is removed. Following *in vitro* transcription, the resulting mRNA is incubated with a ribosome extract for translation into protein. Since the mRNA lacks the stop codon for termination, the translation becomes stalled and hence, the ribosome provides a physical linkage between the protein (i.e. the phenotype) and the mRNA (i.e. the genotype). The selection procedure using ribosome displayed protein libraries resembles the biopanning in phage display and is typically performed in solution with target molecules bound to paramagnetic beads. The key feature of ribosome display is the possibility to readily produce very large libraries of up to $10^{13}$ variants without the limitations of cell transformation efficiencies (Hoogenboom, 2005). In addition, the amplification step after selection can be performed using error-prone PCR in order to introduce additional diversity between selection rounds (Roberts, 1999). A similar cell-free selection system, mRNA display (Roberts & Szostak, 1997, Lipovsek & Plückthun, 2004) share the same advantages as ribosome display but will not be further discussed.

In contrast to the previously described methods that rely on protein display, incubation with a target protein, and the subsequent capturing or sorting of high-affinity variants, the protein complementation assay (PCA) is typically based on growth selection and entirely performed inside a cell. In this method, a reporter protein with important functions for cell survival is divided into two inactive parts.
two separate vector systems, one part is in turn genetically fused to the target protein while the other part is fused to the recombinant protein library. Upon transformation of both vectors to a cellular host, only library variants with affinity to the target protein should be able to reconstitute the function of the reporter protein by creating proximity of the two parts, and subsequently survive by selective growth. PCAs generally refer to systems developed for *E. coli*, utilizing the reporter function of a genetically split enzyme (Michnick, *et al*., 2000). However, a number of related methods have been described, such as the yeast two-hybrid (Bruckner, *et al*., 2009) and bacterial n-hybrid systems (Hu, 2001) that function according to the same principle and typically utilizes the reconstitution of a genetically split transcription factor. PCA and hybrid systems have become valuable tools for discovery of protein-protein interactions, in particular for systems biology in order to identify biochemical networks (Michnick, *et al*., 2000, Morell, *et al*., 2009). For combinatorial protein engineering purposes, PCA based methods utilizing either a split dihydrofolate reductase (mDHFR) gene (Pelletier, *et al*., 1998, Arndt, *et al*., 2000) or a split β-lactamase gene (Galarneau, *et al*., 2002, Wehrman, *et al*., 2002) have been reported. It has been speculated that unspecific binding to intracellular proteins and a difficulty to distinguish between binders with different affinity to the target may complicate the use of PCA for this purpose (Koch, *et al*., 2006). However, the simple selection strategy (i.e. over-night growth), and the co-expression of the target protein, are attractive features that could provide a very rapid selection of affinity proteins. The fact that there is no prerequisite to have access to target proteins of good quality in terms of purity and stability is a key advantage for these systems. Indeed, the potential of the PCA selection platform was recently demonstrated in two separate studies where improved PCA systems using β-lactamase as reporter protein were employed in successful selections of low nanomolar scFv fragments (Secco, *et al*., 2009) and affibody molecules (Löfdahl, *et al*., 2009) from complex naïve libraries.
5 Staphylococcal surface display

The gram-positive bacterium *Staphylococcus carnosus* is a food-grade bacteria commonly found in fermented meat products. For several decades, *S. carnosus* has been exploited in the meat industry, and in particular as a component of starter cultures used for ripening of dried sausage, since the species contribute to several characteristics of the final product. Some chemical processes facilitated by *S. carnosus* and exploited in sausage ripening are: the production of methyl ketones for development of the distinctive taste and aroma (Fadda *et al.*, 2002), and the reduction of nitrate to nitrite, which prevents growth of food pathogens and further reacts with myoglobin to produce a red color (Götz, 1990). Initially, this small coccus (around 0.5-1.5 μm) was considered to belong to the micrococcus family of bacteria. However, the bacterium was later classified as a new staphylococcal species in 1982 by Schleifer and Fischer (Götz, 1990). Despite their close phylogenetic distance based on 16S rRNA (Saitou & Nei, 1987, Woese, 1987), *S. carnosus* share little DNA homology to the pathogenic *S. aureus* and does not produce virulence factors such as protein A, coagulase, haemolysins and enterotoxins (Götz, 1990). The genome sequence of *S. carnosus* was recently published (Rosenstein *et al.*, 2009). In contrast to *S. aureus* that appears in grape-resembling clusters *S. carnosus* grows predominantly in pairs or as single cells. The extensive use in the food industry together with the non-pathogenic behavior has led to the classification of *S. carnosus* as a GRAS-organism (generally regarded as safe) by the U.S. Food and Drug Administration (FDA).

Since *S. carnosus* cells neither secrete or display virulence factors on the cell surface, and also show a very low level of extracellular proteolytic activity, the species has been extensively used as bacterial host for production of recombinant proteins (Götz, 1990,
Pschorr, et al., 1994, Wieland, et al., 1995, Dilsen, et al., 2000, Hansson, et al., 2002). Furthermore, *S. carnosus* has been employed for cell-surface display applications, which will be further discussed in the next sections.

### 5.1 Cell-display principle

Over a decade ago, Ståhl and co-workers described a first plasmid vector system for display of heterologous proteins on *S. carnosus* (Samuelson, et al., 1995). The expression vector was successfully employed for display of recombinant proteins in several applications as will be further discussed below (Wernérus, et al., 2002). In order to improve the genetic stability, the vector was later modified by the exchange of elements from a related vector, previously used for recombinant display on *Staphylococcus xylosus* (Wernérus & Ståhl, 2002).

The staphylococcal display vector (Fig. 5.1) uses the promoter (P<sub>lip</sub>), signal sequence (S) and propeptide (PP) from a lipase gene found in the related *Staphylococcus hyicus* (Götz, et al., 1985, Liebl & Götz, 1986). The propeptide has an essential role in the efficient translocation of recombinant proteins to the cell surface (Demleitner & Götz, 1994, Samuelson, et al., 1999, Sturmfels, et al., 2001), and is subsequently removed by proteolytic cleavage in *S. hyicus*. This mechanism is however not present in *S. carnosus* and hence, the propeptide remains as part of the displayed recombinant protein (Samuelson, et al., 1999). Following the propeptide is a multiple cloning site for insertion of gene fragments encoding the protein of interest, and thereafter an albumin-binding protein (ABP) from streptococcal protein G (Sjölander, et al., 1997),

![Fig. 5.1. The staphylococcal display vector pSCXm (7655 bp). Bla, β-lactamase encoding gene; Cml, chloramphenicol acetyl transferase encoding gene; OriS, origin of replication from *S. aureus*; OriE, origin of replication from *E. coli*; ABP, albumin binding protein; P<sub>lip</sub>, promoter; S, signal sequence; PP, propeptide; XM, cell wall anchoring region.](image-url)
which will be further discussed below. The C-terminal element of the expression cassette is the cell-wall anchoring region (XM), derived from staphylococcal protein A (SpA). This region consists of a charged repetitive part (X), believed to interact with the cell wall (Guss, et al., 1984), and a cell wall sorting signal (M) that is highly conserved among gram-positive species and contains an LPXTG-motif, a hydrophobic region, and a charged tail. The charged C-terminal tail has an important function to trap the protein in the cell membrane in order to avoid secretion, while the LPXTG-motif is recognized by an enzyme (i.e. a sortase) that proteolytically cleaves between the threonine and glycine residues and covalently attaches the protein to cross-linking peptides in the cell wall (Ton-That, et al., 1997, Navarre & Schneewind, 1999). In addition to the expression cassette, the vector also contains an origin of replication derived from \textit{S. aureus} and a chloramphenicol acetyl transferase gene for replication and selection in staphylococci. Moreover, in order to facilitate subcloning in \textit{E. coli}, the so-called shuttle-vector also contains an origin of replication for \textit{E. coli} and a β-lactamase gene conferring ampicillin resistance.

\textbf{Fig. 5.2.} Schematic illustration of the processed fusion product displayed on the staphylococcal cell surface. PP, the propeptide from a \textit{S. hyicus} lipase gene that remains part of the fusion protein when expressed in \textit{S. carnosus}. X, part of the SpA-derived cell wall anchoring region. M', the second part of the anchoring region from SpA, now in its processed and covalently anchored form. ABP, the three-domain albumin-binding protein from streptococcal protein G. Z, the protein of interest, here represented by the SpA-derived protein Z.

The processed fusion protein displayed on the staphylococcal cell surface is illustrated in Fig. 5.2. As mentioned, the propeptide is not processed in \textit{S. carnosus} and hence, remains as the N-terminal part of the fusion protein. The albumin binding protein, consisting of three albumin-binding domains, is situated between the protein of interest and the cell-wall anchoring region and serves the purposes of: (i) introducing space between displayed proteins and the cell wall to ensure accessibility, (ii) functioning as an affinity handle for purification of cell-wall extracted fusion proteins.
on human serum albumin (HSA) columns (Samuelson, et al., 1995), and (iii) providing means for flow cytometric monitoring of the level of surface-accessible fusion proteins displayed on the cell surface. In fact, the latter method was used to estimate the number of displayed recombinant proteins to approximately 10,000 per staphylococcal cell (Andreoni, et al., 1997).

5.2 Applications

Since the system was first described in 1995, recombinant display on S. carnosus has become extensively employed in a number of applications. Considering the GRAS-designation, S. carnosus was early investigated as a potential vehicle for live delivery of subunit vaccines (Nguyen, et al., 1993, Liljeqvist, et al., 1997, Liljeqvist & Ståhl, 1999). Moreover, the system has been used for display of recombinant proteins including: (i) metal-binding peptides for environmental applications (Samuelson, et al., 2000, Wernérus, et al., 2001), (ii) scFv fragments and affibody molecules for potential use as whole-cell diagnostic devises (Gunneriusson, et al., 1996, Gunneriusson, et al., 1999b), and (iii) a cellulose-binding domain for surface immobilization purposes (Lehtiö, et al., 2001).

The successful applications of yeast and E. coli display in combinatorial protein engineering in the early 2000, and the potential of cell-based systems to become attractive alternatives to phage display, inspired the evaluation of staphylococcal cell-surface display also for this purpose. Employing a second generation display vector, the staphylococcal system was evaluated in a model library application using flow cytometry (Wernérus, et al., 2003). At a 1:100,000 ratio, staphylococcal cells displaying the IgG-binding affibody molecule Z_wt were spiked with cells displaying a non-related affibody molecule. Using fluorescently labeled IgG as a target molecule, two rounds of FACS sorting was sufficient to obtain a 25,000-fold enrichment of Z_wt-displaying cells. The system was further optimized by the introduction of a surface expression normalization strategy in order to improve the ability of discriminating between closely related binders and to account for the natural variations of surface expression levels for individual cells (Löfblom, et al., 2005). This strategy utilizes the albumin binding protein, present as part of the expressed fusion protein, as a reporter protein.
for monitoring of the expression level by the binding to fluorescently labeled HSA. Hence, by a two-color labeling, the target binding signal is normalized with respect to the surface expression on individual cells. As a consequence, the discrimination ability increases by an overall decrease in the binding signal distribution. The novel approach was evaluated by flow cytometric sorting of staphylococcal cells displaying Z\textsubscript{wt}, mixed at a 1:1000 ratio with a single amino-acid mutant demonstrating an eight-fold weaker affinity to the target IgG \cite{Cedergren1993}. After one round of FACS sorting of the top 0.1\% fraction (i.e. the fraction having the highest normalized target binding signal), Z\textsubscript{wt}-expressing cells were enriched 130-150-fold. Furthermore, the staphylococcal display system was evaluated for post-selectional characterization directly on the cell surface using flow cytometry. In a comparative study by Löfblom and co-workers, the equilibrium dissociation constants for Z\textsubscript{wt} and two single amino-acid mutants with different affinity for IgG were determined both on-cell using flow cytometry and in a biosensor assay \cite{Löfblom2007}. The results demonstrated that affinity determination using staphylococcal surface display showed a correct relative order of affinities for the three IgG-binding variants, in agreement with the results from biosensor analysis. However, the absolute affinities differed approximately 40-fold between the methods. The main difference was expected to be due to avidity effects on the staphylococcal surface when using a bivalent target protein, i.e. IgG. Indeed, a further study performed on a monovalent albumin-binding domain (ABD) with affinity for HSA resulted in only a six-fold difference in the absolute affinity between the methods. These successful results obtained from model systems indicated that staphylococcal display had the potential to become a powerful selection platform for combinatorial protein engineering, and encouraged the continued work to implement the system for real-library selections which has been the focus of the studies included in this thesis.
6 Present investigation

The objective of the underlying studies which this thesis is based on has been the development and implementation of a staphylococcal display system for directed evolution of affinity proteins (i.e. affibody ligands). The four papers included in this thesis describe studies that can be divided into two categories: technical improvements, performed in order to optimize and further develop the system (papers I and III), and the applications of staphylococcal display for isolation and characterization of affinity proteins (papers II and IV). In this section, I will summarize this work and briefly discuss some future perspectives.

In paper I, the transformation protocol for *S. carnosus* was improved by a series of optimization steps in order to improve the transformation efficiency to levels required for combinatorial library approaches. The successful results in paper I, lead to the first library application using staphylococcal display. Hence, paper II describes the selection and characterization of three specific affibody molecules selected against human tumor necrosis factor-alpha (TNF-α). In paper III, the system was further improved to allow simplified post-selectional characterization of a larger set of candidate binders by the introduction of a specific protease recognition site for on-cell cleavage and production of soluble proteins to omit tedious subcloning procedures. Finally, paper IV describes a selection of affibody molecules with specific affinity for the human epidermal growth factor receptor 3 (ErbB3), by a combined approach using phage display for *de novo* selection and staphylococcal display for affinity maturation.
In order to generate recombinant protein libraries displayed on the surface of cells, the randomized DNA must be efficiently delivered into the cells of choice. Typically, high-frequency transformation is performed by electroporation techniques where an electric field is applied for creation of temporary pores to increase the cell permeability and allow uptake of macromolecules (Shigekawa & Dower, 1988). Although cell-based selection systems have several advantages over alternative systems, most importantly the possibility of using powerful flow cytometric analysis in the selection process and a multivalent display of recombinant proteins, any system based on replication inside cells (e.g. cell display and phage display) is more or less restricted by the transformation efficiency. Since this efficiency determines the final size of the library, the transformation step becomes a crucial bottleneck which limits the potential diversity.

For bacterial systems, gram-negative organisms such as *E. coli* usually have the highest transformation frequencies and therefore might be considered more suitable for protein library applications. However, recombinant display on gram-positive species such as *S. carnosus* provides additional, and very attractive, features. First, gram-positive cells have a robust structure built from a single cell membrane surrounded by a thick peptidoglycan cell wall. Consequently, these cells show a high viability when subjected to physical stress, such as pressure-induced shear-forces, a property that also increases the ability to withstand the harsh conditions during high-speed FACS sorting compared to gram-negative cells (Pagan, *et al.*, 1999). In addition, the simple architecture allows proteins to be efficiently translocated and attached to the cell surface. Second, recombinant proteins are typically displayed by a C-terminal anchoring to the cell wall, a mechanism able to withstand relatively large fusion proteins (Navarre & Schneewind, 1999).
Recombinant display on *S. carnosus* was early suggested to be a suitable system for selection of staphylococcal protein A-derived affibody molecules due to the same origin of the species, the scaffold, and the cell wall anchoring cassette, which potentially increases the probability for functional display on the cell surface. However, the low-efficiency transformation hampered the generation of large protein libraries. In this study, the electroporation procedure was thoroughly examined and optimized to reach transformation frequencies necessary for construction of high-diversity combinatorial protein libraries. Several parameters throughout the electroporation process were investigated, including preparation of competent cells, the electroporation step, and the recovery of transformed cells. One of the main concerns regarding the *S. carnosus* transformation procedure is the origin of the plasmid DNA which is used for electroporation. Since staphylococcal cells give low yields upon plasmid preparation, the expression vector used for recombinant display on *S. carnosus* is a shuttle vector, containing origins of replication for both *E. coli* and staphylococci. This enables efficient library construction and preparation of large amounts of the plasmid library in *E. coli*, followed by transformation to *S. carnosus*. However, interspecies transfer of DNA is generally problematic since foreign DNA is recognized by differences in the methylation patterns, and subsequently degraded by enzymes involved in the cell restriction machinery. Hence, in an effort to increase the transformation frequency of *S. carnosus*, we investigated a previously proposed method (Edwards, *et al.*, 1999) to temporarily inactivate the host-restriction enzymes by heat treatment of the cells prior to electroporation. In a series of experiments, the optimal heat-treatment protocol was determined by evaluation of different time and temperature combinations and the protocol for subsequent washing of the cells. Thereafter, treated cells were electroporated with the wild-type variant of the expression vector (pSCX:Zwt) in order to determine the effects on the transformation frequency. In a comparative experiment, the effect on the frequency using both staphylococcal and *E. coli*-prepared DNA, with or without the optimized heat-treatment protocol was evaluated (Fig. 6.1). The heat-treatment procedure resulted in a 20-fold increase in the transformation frequency for *E. coli*-prepared DNA. Most importantly, after heat-treatment, the efficiency for DNA prepared from *E. coli* was similar to the efficiency observed with DNA prepared from *S. carnosus*, strongly indicating an inactivation of the host restriction machinery in the staphylococcal cell.
Although the heat-induced inactivation of host restriction resulted in the highest increase in the transformation frequency, the altering of several other parameters contributed to the final optimized protocol. Besides heat-treatment, the most important changes in order to improve the transformation frequency were: addition of an osmotic stabilizer (i.e. sucrose) to the electroporation buffer, electroporation in room temperature, new settings for the electrical pulse during electroporation and a recovery medium with elevated concentrations of salt and yeast extract. In table 6.1 the final improved protocol is compared to the previously used protocol for electroporation of *S. carnosus*, including the approximate contribution to the increase in transformation frequency observed for each optimized parameter. It should be noted that regarding the amount of DNA used in the transformation, it is not the transformation frequency (obtained number of transformants per µg DNA) that is improved, but instead the actual number of colonies in a transformation event, since this parameter is what is more relevant in order to obtain increased library sizes. Increasing the amount of DNA above 4 µg did not significantly contribute to an increased number of transformants.

To confirm that all transformants obtained by the novel transformation procedure expressed the recombinant Zwt-domain together with the albumin binding protein (ABP) on the cell surface, an assay was performed using flow cytometry. In this experiment, an over-night culture of cells transformed with pSCX:Zwt, and selectively grown on agar plates, was incubated with fluorescently labeled IgG and HSA followed by flow cytometric analysis of the fluorescence intensity. The transformed population was almost identical to control cells obtained from a single-clone culture of pSCX:Zwt-containing cells, indicating that

![Fig. 6.1. Heat-induced inactivation of host restriction in *S. carnosus*. The black bars represent the transformation frequency obtained for non-treated cells using plasmid DNA prepared from *E. coli* and *S. carnosus*, respectively. The white bars represent the transformation frequency obtained for heat-treated cells, again using plasmid DNA prepared from both species.](image-url)
recombinant proteins were functionally displayed on the cell surface using the new transformation protocol.

**Table 6.1 Initial and optimized procedures for transformation of *S. carnosus***

<table>
<thead>
<tr>
<th>Step</th>
<th>Initial procedure</th>
<th>Optimized procedure</th>
<th>Increase in efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation of cells</td>
<td>Growth from OD_{578} 0.1-0.6</td>
<td>Growth from OD_{578} 0.5-0.6</td>
<td>2-fold</td>
</tr>
<tr>
<td>Thawing of cells</td>
<td>10 min on ice</td>
<td>5 min on ice, 30 min at room temperature</td>
<td>Included below (heat treatment)</td>
</tr>
<tr>
<td>Heat treatment</td>
<td>-</td>
<td>56°C for 2 min, followed by a washing step</td>
<td>20-fold</td>
</tr>
<tr>
<td>Electroporation buffer</td>
<td>10% glycerol</td>
<td>0.5 M sucrose and 10% glycerol</td>
<td>5-fold</td>
</tr>
<tr>
<td>Amount of plasmid DNA</td>
<td>1 µg</td>
<td>4 µg</td>
<td>3-fold</td>
</tr>
<tr>
<td>Electroporation conditions</td>
<td>Cells, cuvettes and medium on ice</td>
<td>Cells cuvettes and medium at room temperature</td>
<td>5-fold</td>
</tr>
<tr>
<td>Electric pulse</td>
<td>20 kV for 2.5 ms (untruncated)</td>
<td>21 kV for 1.1 ms (truncated)</td>
<td>5-fold</td>
</tr>
<tr>
<td>Recovery</td>
<td>SMMP medium, 37°C, 48 h</td>
<td>B2 medium, 37°C, 48 h</td>
<td>5-fold</td>
</tr>
<tr>
<td>Freezing of cells</td>
<td>Liquid nitrogen, -80°C freezer</td>
<td>-80°C freezer</td>
<td>3-fold</td>
</tr>
<tr>
<td>Transformation frequency</td>
<td>cfu transformation^{-1}</td>
<td>10^{1}-10^{2}</td>
<td>10,000-fold</td>
</tr>
</tbody>
</table>

In conclusion, the transformation frequency of *S. carnosus* was optimized by the introduction of a heat-induced inactivation of host cell restriction, and by changing several parameters of the initial transformation protocol. By these means, the transformation frequency was increased around 10,000-fold to 10^{5}-10^{6} cfu per transformation, a satisfactory level for construction of high-complexity combinatorial libraries.
6.2 A novel affinity protein selection system based on staphylococcal cell surface display and flow cytometry (II)

In this study, surface display on the gram-positive bacterium \textit{S. carnosus} was for the first time employed for isolation of novel affinity proteins in a library application. Based on the verified functional display of affibody molecules on the staphylococcal cell surface (described in chapter 5), the staphylococcal display system was considered to be well suited for selection of affibody binders. However, despite the successful optimization of the transformation frequency described in paper I, generation of a high-complexity library within the same size range (~$10^9$) as the currently used naïve phage-displayed affibody library (Grönwall, et al., 2007) would require no less than a thousand transformations. Although this might be feasible with some effort, flow cytometric sorting of cell displayed libraries with a complexity over $10^8$ is however not easily performed in practice due to the increasing sampling time and consequently, large cell displayed libraries are commonly pre-enriched using for example magnetic bead-based panning (Feldhaus, et al., 2003, Bessette, et al., 2004). Here we chose a pre-enrichment strategy based on one round of phage display selection followed by transfer to staphylococci. By this approach, the complexity was reduced to levels more suitable for flow cytometric sorting and also, tedious transformation of a naïve affibody library to \textit{S. carnosus} was circumvented.

In order to avoid PCR-induced biases upon library transfer between the phage and staphylococcal systems, the staphylococcal display vector pSCXm (Wernérus & Ståhl, 2002) was redesigned to facilitate direct subcloning from the phagemid vector into the staphylococcal display vector by a direct digestion and ligation procedure. The novel staphylococcal vector (pSCZ1) was constructed through: (i) the insertion of a dummy fragment for more efficient subeloning, (ii) the insertion of restriction sites identical to those flanking the randomized gene in the phagemid vector, and (iii) modification to ensure in-frame insertion of fragments obtained by restriction digestion of the phagemid vector (Fig. 6.2.A). The phage-displayed affibody library described above
was pre-enriched in one cycle of phage display using human tumor necrosis factor alpha (TNF-α) as target protein. After infection to *E. coli* for amplification and subsequent isolation of phagemid library plasmids, the library fragments were subcloned to pSCZ1. Finally, after transformation to *S. carnosus* according to the improved protocol described in paper I, a pre-enriched staphylococcal-displayed affibody library with a final complexity of $1.1 \cdot 10^6$ was obtained.

Prior to employing the novel staphylococcal displayed library for the intended selection experiment, an investigation was conducted to evaluate the pre-selected library obtained from phage display and the proportion of functionally displayed variants in the staphylococcal system. Sequencing performed on 96 individual *E. coli* clones isolated from the phage display pre-selection showed only unique sequences and approximately 70% were of the correct length (i.e. without insertions, deletions, truncations, or multiple inserts). In order to determine the surface-expression level for the staphylococcal displayed library, a flow cytometric analysis of the unsorted library was performed. The histogram obtained from the analysis showed a high percentage of ABP-expressing cells within the gated region. In subsequent passages, the percentage of ABP-expressing cells was monitored to assess the stability and efficiency of the library.

**Fig. 6.2.** A: Schematic illustration of the novel staphylococcal display vector pSCZ1. B: Histogram obtained from flow cytometric analysis of the unsorted library, showing fluorescently labeled HSA binding to functionally displayed ABP within the gated region. C: Histogram showing the percentage of ABP-expressing cells in the unsorted library after each passage.
Staphylococcal surface display in directed evolution

44

Staphylococcal surface display in directed evolution
cytometric assay was performed using HSA for monitoring of the expressed fusion
candidate ABP. Previously conducted assays using single clones such as the IgG-binding
affibody molecule Z_wt have demonstrated that around 90% of the staphylococcal cell
population display full-length recombinant fusion proteins (Löfblom, et al., 2005). In
this library experiment however, only around 20% of the population displayed
functional ABP on the cell surface (Fig. 6.2.B). The rather low proportion of
functionally expressed variants observed for the pre-enriched affibody library was
considered to be mainly due to problems arising from the NNK randomization
strategy used for synthesis of the affibody library, by which a certain level of amber-
stop codons are introduced in randomized positions. Although phage display
protocols utilize amber-suppressor E. coli strains in order to partly avoid termination
by amber-stop codons, truncations will still occur at a certain frequency. Since
truncated variants typically are beneficial for cell survival due to a lower pressure on
the expression machinery, phage displayed libraries have an inherent bias towards
variants containing amber-stop codons. Following transfer of such variants to S.
carnosus that lacks amber-suppressing capacity, translation is prematurely terminated
and hence variants containing amber stops are not successfully displayed on the
staphylococcal cell surface.

Since this was a first attempt of employing staphylococcal display for an affibody
library, notice was taken of the possibility for growth bias between selection cycles
that could result in an uneven distribution of library members. Upon a significant
growth bias, the ratio of non-displaying to displaying cells (80% and 20% respectively)
would shift when monitoring the ABP-binding signal using flow cytometry. The
staphylococcal library was therefore amplified around 10^{12}-fold by four sequential
passages of over-night growth with a 1000-fold amplification respectively. After each
passage, flow cytometric analysis was conducted to determine the percentage of ABP-
expressing cells in the population. The results showed a stable 20% proportion of
expressing cells after each passage, indicating no detectable growth bias between
staphylococcal-displayed library members (Fig. 6.2.C).

Affibody molecules binding to TNF-α were isolated by flow cytometric sorting of the
pre-enriched staphylococcal-displayed affibody library. A portion of the library was
incubated with fluorescently labeled TNF-α and HSA and subjected to three rounds
of FACS sorting with an increased stringency between rounds. In each round, normalized gating was employed to sort out the part of the cell population having the highest ratio of target binding compared to the surface expression level (Fig. 6.3.A). Upon sequencing of isolated clones, 15 unique full-length sequences were found and among these, 3 variants appeared more than once (Fig. 6.3.B). In order to verify target

![Fig. 6.3. A:](image1) Results from the flow cytometric sorting of the staphylococcal-displayed pre-selected affibody library presented as dot plots. The x-axis show the FL-4 channel fluorescence intensity corresponding to the surface expression level monitored via HSA binding. The y-axis show the FL-2 channel fluorescence intensity corresponding to TNF-α binding. Dot plots are showing the staphylococcal library prior to FACS sorting in each round (1-3) and the regions used for subsequent gating. **B:** Amino acid sequences of the isolated TNF-α binding affibody molecules $Z_{\text{TNF-\alpha 1}}$, $Z_{\text{TNF-\alpha 2}}$ and $Z_{\text{TNF-\alpha 3}}$ as compared to the parental $Z_{\text{wt}}$. Randomized positions are indicated (*) and the representation (%) of each TNF-α binder is shown after the sequence.
binding activity, all isolated clones were analyzed using a whole cell ELISA. The three clones that appeared several times in the sequencing experiment also were found to give positive signals in ELISA. The most abundant variant (Z_{\text{TNF-alpha1}}) was found to be present at 95% in the isolated population, indicating either a higher affinity or potentially a higher growth-rate between selection cycles and hence, to evaluate the latter, a growth-rate experiment was conducted for this variant. Briefly, Z_{\text{TNF-alpha1}}-expressing staphylococcal cells were mixed at a 1:20 ratio with the unsorted cell library and subjected to two growth passages as described earlier. Flow cytometric analysis of the TNF-\( \alpha \) binding signal revealed a retained 5% ratio between rounds, thus excluding the possibility of growth bias compared to other members of the library pool.

The equilibrium dissociation constants (K_d) for the three ELISA-positive affibody molecules Z_{\text{TNF-alpha1}}, Z_{\text{TNF-alpha2}}, and Z_{\text{TNF-alpha3}} were determined directly on the staphylococcal cell surface using flow cytometry. The results demonstrated apparent affinities ranging from 95 pM to 2.2 nM and as expected due to the quantitative sorting provided by normalized gating, the most abundant variant Z_{\text{TNF-alpha1}} showed the highest apparent affinity (95 pM) for TNF-\( \alpha \). Moreover, the results obtained by on-cell affinity determination were verified using a biosensor assay (Biacore analysis). The affibody molecules were subcloned to E. coli for soluble protein production and subsequently individually immobilized on sensor-chip surfaces for monitoring of the response signals using different concentrations of TNF-\( \alpha \). The results demonstrated similar relative affinities as obtained on-cell, indicating that the variants were correctly ranked relative to each other. However, the absolute affinities were three to six-fold lower when using Biacore. As previously shown, the difference in absolute affinity observed for the two methods becomes more significant when using multivalent target proteins (Löfblom, et al., 2007) and is thought to mainly result from avidity effects occurring on the staphylococcal cell surface. Since most TNF-\( \alpha \) monomers associate into homotrimers at the concentrations used here, avidity effects are most likely contributing to the apparent affinity.

In order to investigate the interaction between the isolated affibody molecules and TNF-\( \alpha \), a competition assay was performed directly on the cell surface using flow cytometry. Prior to incubation with affibody-displaying staphylococcal cells, TNF-\( \alpha \)
was pre-incubated with a ten-fold excess of two monoclonal antibodies (infliximab and adalimumab) and a recombinant TNF-\(\alpha\) receptor fusion protein (etanercept) respectively. The target binding signals were analyzed with flow cytometry and compared to the signals obtained without competition (i.e. only TNF-\(\alpha\)). The results demonstrated that competition with adalimumab almost totally blocked the affibody-TNF-\(\alpha\) interaction (Fig. 6.4.A) and therefore, the epitope is probably shared or

![Fig. 6.4. Histograms obtained from on-cell competition analysis using flow cytometry. The x-axis shows the FL-2 channel fluorescence intensity corresponding to TNF-\(\alpha\) binding. Rightmost peaks are showing the TNF-\(\alpha\) binding signal without competition (MFI_{Control}). Leftmost peaks are showing the TNF-\(\alpha\) binding after competition (MFI_{Comp}). In each histogram the percentage (%) of remaining binding activity after competition is calculated as: MFI_{Comp} / MFI_{Control} \times 100). A: Competition with Adalimumab at a 10-fold molar excess over TNF-\(\alpha\). B: Competition with Etanercept at a 10-fold molar excess over TNF-\(\alpha\). C: Competition with Infliximab at a 10-fold molar excess over TNF-\(\alpha\).]
overlapping. However, incubation with infliximab only partially blocked the interaction (Fig. 6.4.C), indicating a partially overlapping epitope or an allosteric effect due to a close proximity of the anti-TNF-α proteins and the surface displayed affibody molecules. Although pre-incubation with etanercept also resulted in a partial blocking of the affibody-TNF-α interaction (Fig. 6.4.B), this is most likely due to the binding mechanism of the TNF-α receptor. Since the receptor only binds between two subunits of the trimeric form of TNF-α, the third subunit remains free to interact with the affibody molecule. Hence, even if the epitope was shared only a partial blocking would be observed in the experiment. Moreover, the results from the competition assay showed similar pattern for all three affibody variants, indicating that they share the same binding epitope on TNF-α.

To conclude this study, staphylococcal surface display was for the first time used as a selection system for generation of affinity proteins from a large combinatorial library. Using a pre-enrichment strategy in order to reduce the library complexity, a naïve affibody library was subjected to one round of phage display with TNF-α as target protein before expression in the staphylococcal system. Evaluation of the staphylococcal-displayed affibody library prior to selection showed that 20% of the library population functionally displayed recombinant proteins on the cell surface and no detectable growth bias was observed. Following flow cytometric sorting of the staphylococcal-displayed affibody library, three new TNF-α binding affibody molecules were isolated and further characterized. The strongest binder, and also the most abundant variant obtained from the selection, showed an apparent affinity of 95 pM for TNF-α when analyzed directly on-cell using flow cytometry.
6.3 Simplified characterization through site-specific protease-mediated release of affinity proteins selected by staphylococcal display (III)

For any selection system, of utmost importance is the ability to discriminate between different protein variants based on the affinity to the target molecule, and to provide a phenotype to genotype linkage for further production and characterization of selected variants. However, a significant number of variants with similar affinities but different protein properties are typically generated, and the downstream characterization of candidate clones becomes an equally important step in the selection procedure. Although cell display systems provide means for a rapid affinity characterization directly on the cell surface using FACS, the investigation of several other properties (e.g. binding specificity, structural stability or affinity to the target displayed in its natural context, such as binding to mammalian cell surface receptors or signal molecules) is typically performed using soluble protein. In most selection systems, downstream characterization of candidate clones is performed by sub-cloning of the selected variants to an expression vector for subsequent production of soluble protein in E. coli. In order to avoid the rather tedious gene transfer to a different expression system, and thereby increasing the number of candidate clones to be feasibly characterized after selection, a more rapid protocol for production of soluble proteins was developed and evaluated in this study. The approach was to provide means for protease mediated release of displayed affinity proteins directly from the staphylococcal cell surface, by the introduction of a unique protease cleavage site in the displayed fusion protein. In addition, fusion of the affinity protein to different affinity tags was investigated to allow for efficient purification of released affinity proteins (Fig. 6.5.A).

This paper describes the construction and subsequent evaluation of three novel vectors based on the previously described display vector pSCX:Zwt (Wernérus & Ståhl, 2002). In principle, all vectors were constructed to express a fusion protein consisting of the IgG-binding staphylococcal protein A-derived affinity protein Zwt (Nilsson, et
al., 1987) fused to a C-terminal affinity tag (specified below), followed by a protease cleavage site. The chosen recognition sequence EALFQ/GP is specifically cleaved by protease 3C of coxsackievirus B3 (Miyashita, et al., 1992) and the commercially available PreScission, a closely related rhinovirus protease (Walker, et al., 1994). The first vector (pHis3C) was constructed by insertion of a gene fragment encoding a polyhistidine (His6) tag, allowing IMAC-based purification, followed by the protease recognition sequence on the C-terminal side of Zwt (Fig. 6.5.B). In the second vector (pABP3C) the already existing albumin binding protein (ABP), used as reporter protein for normalization of the surface expression levels and C-terminally fused to Zwt, was here also exploited for affinity purification on HSA-Sepharose. Hence, only a gene fragment encoding the protease recognition sequence was inserted on the C-terminal side of ABP (Fig. 6.5.B). The third vector (pABD3C) was constructed by insertion of a gene fragment encoding 46 amino acid albumin binding domain (ABD) followed by the protease recognition sequence on the C-terminal side of Zwt (Fig. 6.5.B).

It has previously been shown that insertions or deletions of gene fragments in the staphylococcal display vector occasionally change important properties, for example plasmid stability and surface expression levels. In addition, previous experiments have indicated that the transformation frequency commonly is reduced by insertions of gene fragments into the rather large expression vector. Therefore, the novel vector constructs were evaluated with respect to the transformation frequency in S. carnosus and the level of surface expressed recombinant proteins compared to the parental vector pSCX:Zwt. The transformation frequency was determined by individual electroporation experiments, using equal amounts of the three vector constructs, and compared to pSCX:Zwt. As expected, the vectors pHis3C and pABP3C showed no significant decrease in transformation frequency, probably as they are similar in size to the parental vector. However, the larger vector pABD3C was transformed with an approximately four-fold lower frequency than pSCX:Zwt and is hence less suited for combinatorial library applications where the maximum achievable transformation frequency is desired. In order to determine the surface expression levels, staphylococcal cells expressing the different vector constructs were analyzed with flow cytometry. The obtained histograms showed similar IgG-binding intensities for all vector constructs, comparable with the intensities obtained from cells expressing the
Fig. 6.5. A: Schematic overview of the devised procedure for on-cell proteolytic release and purification of staphylococcal-displayed affinity proteins. Selected clones obtained from FACS sorting are individually cultivated and treated with a 3C protease to produce soluble affinity proteins in fusion with a protein purification tag (black square) for subsequent purification using affinity chromatography. B: Expression cassettes of the three vectors pHis3C, pABP3C and pABD3C constructed and evaluated for proteolytic release of staphylococcal displayed fusion proteins using a site-specific 3C protease. Abbreviations: 3C, recognition sequence for the protease 3C; His<sub>6</sub>, hexahistidine used for protein purification. Additional abbreviations are described in figure 5.1.
parental vector. The results indicate that the insertion of any of the affinity tags together with the protease cleavage site had no negative impact on the functional expression of fusion proteins.

In order to investigate the novel strategy for on-cell proteolytic release of displayed fusion proteins, staphylococcal cells harboring the three vector variants were incubated with protease 3C followed by a series of experiments to determine the cleavage efficiency and the possibility to produce pure and functional fusion proteins. The overall cleavage efficiency is largely dependent on the accessibility of the protease cleavage sites on the cell surface (apart from the protease activity). However, the recognition sequence was positioned differently in all the vector constructs, dependent on the chosen affinity tag. To determine the cleavage efficiency, protease treated cells expressing the different fusion proteins were analyzed with flow cytometry. The results showed an approximately 80\% decrease in the IgG-binding signal for all constructs after cleavage compared to non-treated cells (Fig. 6.6), indicating a quite efficient release of fusion proteins independent of the location of the cleavage site. Moreover, the supernatants obtained from the cleavage reactions were collected and purified from protease and other contaminants using affinity chromatography (IMAC for His-tagged proteins and HSA affinity chromatography for ABP or ABD-tagged proteins). The most concentrated fractions of purified proteins were subsequently analyzed with polyacrylamide gel electrophoresis to confirm the presence of pure protein in the expected size range and with no detectable degradation. In order to

![Fig. 6.6. Histograms obtained from flow cytometric analysis of the protease 3C cleavage efficiency using the vectors pHis3C, pABP3C and pABD3C, respectively, for surface display of Zwt fusion proteins. The x-axis show the fluorescence intensities corresponding to the IgG binding before (rightmost peaks) and after proteolytic cleavage with protease 3C (leftmost peaks).](image)
confirm retained binding activity (i.e. functionality) for the different fusion proteins after proteolytic release and affinity purification, a biosensor analysis was performed. A series of concentrations of the three produced variants were individually injected over a flow-cell surface containing immobilized IgG, and the results revealed similar binding activity for all three fusion proteins.

To verify that the binding capability of the affibody molecule was not negatively influenced when expressed and released using the presented strategy, we investigated the applicability of the devised protocol for downstream characterization of target-binding in a cell-based bioassay, using a previously selected affibody molecule as model. The high-affinity affibody molecule ZHER2 (Orlova, et al., 2006), binding with picomolar affinity to the human epidermal growth factor 2 (HER2), was subcloned to the novel expression vector pHis3C for subsequent proteolytic cleavage and purification using IMAC. Without further manipulation, the ZHER2-His<sub>6</sub> fusion protein was used in a flow cytometric assay to determine the cell surface expression of HER2 on human breast cancer cells (Fig. 6.7). A labeled anti-His<sub>6</sub> antibody was used for detection of bound fusion proteins, thus circumventing direct labeling of the affibody molecule, a tedious procedure that would result in decreased yields and possible interference with the binding surface. The flow cytometric assay was performed with both on-cell cleaved and purified ZHER2-His<sub>6</sub> and a commercially available HER2-binding affibody molecule with similar binding affinity as a positive control. The results were similar for

![Figure 6.7](image_url)

**Fig. 6.7.** Histograms from flow cytometric analysis of on-cell proteolytically released and purified ZHER2-His<sub>6</sub> binding to human breast cancer cells. The ZHER2-His<sub>6</sub> binding signal was monitored with a biotinylated anti-Penta-His and stained with fluorescently labeled streptavidin. For negative control, cells were incubated with biotinylated anti-Penta-His and fluorescently labeled streptavidin only. A commercially available fluorescein conjugated anti-Her2 affibody molecule served as positive control.
both affinity proteins, demonstrating a high level of HER2 expression on the cancer cells and hence, the fusion protein produced by the devised protocol show a retained functionality.

To conclude, this study demonstrated that a protease recognition sequence could be introduced to the surface-displayed fusion protein in order to release affinity proteins directly from the staphylococcal cell surface for affinity purification of small amounts of soluble proteins. Three different vector constructs were evaluated and most importantly, a biosensor analysis and an on-cell bioassay showed a retained functionality for the proteolytically released and purified fusion proteins (consisting of the affinity protein in fusion with a purification tag). The devised method enables rapid production of soluble protein and provides means to characterize a larger set of candidate clones in different downstream assays, thus increasing the probability of identifying affinity proteins with desired properties.
6.4 Combining phage and staphylococcal surface display for generation of ErbB3-specific Affibody molecules (IV)

In this study, affibody molecules specific for the human epidermal growth factor receptor 3 (ErbB3 or HER3) were selected and characterized by a combined approach, using both phage and staphylococcal display.

So far, affibody molecules with new specificities have typically been isolated from both naïve and second-generation affinity maturation libraries using the phage display technology. However, since many second-generation libraries generate only slightly improved mutants in terms of affinity, a selection system used for affinity maturation purposes should preferably have a high capacity for fine affinity discrimination. Although the capture and elution principle of phage display often is sufficient to separate matured binders, the method less successful if the library contains only mutants with a few-fold increase in the affinity. In contrast, cell-based selection systems such as the staphylococcal display system provide a truly quantitative measurement of the affinity for each library member due to the multivalent display on the cell surface combined with fluorescence-activated cell sorting (FACS) that facilitates real-time monitoring of the binding interactions with a single-cell resolution. These properties enable discrimination of mutants with small differences in their relative affinities and hence, cell-based systems are particularly valuable for affinity maturation applications. However, for true de novo selections, phage display still has an advantage due to the higher transformation frequency to *E. coli*, enabling libraries with higher complexities to be generated. In order to isolate ErbB3-binding affibody molecules in this study, we chose a combined approach using both phage and staphylococcal display in order to fully utilize the advantages of each system. First, since a higher complexity can be obtained in *E. coli*, phage display was employed for de novo selection of first-generation binders from a naïve affibody library. Second, due to the fine affinity discrimination capacity, staphylococcal display was employed for a second-generation library based on the variants obtained from phage display.
A portion of a previously described phagemid affibody library with a complexity of around $3 \cdot 10^9$ (Grönwall et al, 2007), constructed with an NN(G/T) randomization strategy, was used for phage display selection of first-generation ErbB3-binding affibody molecules using the extracellular domain of ErbB3 (ErbB3-ECD) as target protein. After four rounds of biopanning in solution with increasing stringency between rounds, randomly picked clones were subjected to an ELISA screen for detection of ErbB3-binding activity using biotinylated ErbB3-Fc fusion protein or the biotinylated extracellular domain of ErbB3 (ErbB3-ECD) in two different assays. Upon sequencing of variants with the highest absorbance values from ELISA, 23 unique sequences were identified of which two were considered background binders. Among the obtained variants, a certain degree of sequence homology was observed in several positions and most significantly an entirely conserved tryptophan in position 17. Moreover, a sequence cluster analysis showed three different clusters of binders.

The first-generation ErbB3-binding affibody molecules showing the highest signals in ELISA were further characterized for specificity and binding activity to native target. A total of 16 clones were subjected to a dot blot analysis using 16 high-abundant human serum proteins as well as ErbB3, ErbB3-Fc, ErbB2 and ErbB4 blotted onto nitrocellulose membranes and all affibody molecules except one showed specific binding to ErbB3. In addition, 11 clones were produced as dimers fused to an N-terminal His$_6$ tag and a C-terminal cysteine for characterization using immunofluorescence microscopy. After incubation with ErbB3 positive AU565 cells (human

**Fig. 6.8.** A: Amino acid sequences (one letter code) of the parental Z$_{wt}$ with randomized positions indicated with • and phage display-selected ErbB3-binding affibody molecules. Representation among sequenced clones is shown after each sequence. Below the sequences of phage display-selected affibody molecules is a schematic overview of the design of the affinity maturation affibody library (Z$_{ErbB3LIB}$) for staphylococcal display selection. Amino acid representation (one letter code) in 12 randomized positions within helix 1 and helix 2. Note that position 17 (W) is not randomized in the maturation library. Bottom images: Immunofluorescence microscopy image of **B**: ErbB3-positive AU565 cells stained with His6-(Z01814)2-cys. **C**: ErbB3-positive AU565 cells stained with His6-(Z01820)2-cys. **D**: ErbB3-positive AU565 cells stained with Anti-HER3 antibody (positive control). **E**: ErbB3-positive AU565 cells stained with (Z01154)2 (negative control). **F**: ErbB3-negative/ErbB2-positive SK-OV-3 cells stained with His6-(Z01814)2-cys. **G**: ErbB3-negative/ErbB2-positive SK-OV-3 cells stained with His6-(Z01820)2-cys.
Staphylococcal surface display in directed evolution

mammary gland cell line), 6 clones showed specific staining patterns. The two clones Z01814 and Z01820 were selected for an additional staining using also ErbB3 negative/ErbB2 positive SK-OV-3 cells (human ovary carcinoma cell line). The results demonstrated distinct staining only on AU565, indicating specific ErbB3 binding (Fig. 6.8.B-G). To roughly determine the relative affinities and to verify functional expression on staphylococcal cells, the four clones were subcloned to the staphylococcal display vector and subsequently transformed to S. carnosus for display on the cell surface. Flow cytometric analysis of affibody-displaying cells incubated with biotinylated ErbB3-Fc and fluorescently labeled streptavidin showed specific binding activity and functional display for all clones. In addition, the lack of cross-reactivity to ErbB2 also in this context was confirmed using biotinylated ErbB2 as a negative control. The relative affinities were estimated in an on-cell ranking experiment using four different concentrations of biotinylated ErbB3-Fc and the results demonstrated no significant difference in the relative affinities among the analyzed clones.

A second-generation affinity maturation library was created based on the 21 sequences obtained from phage display (Fig. 6.8.A). In order to gain control over the amino acid distribution and identities in each randomized position and thereby restrict the final diversity, the second-generation library was generated using the Slonomics® technique, a novel and attractive method for production of double-stranded DNA libraries by sequential incorporation (i.e. ligation and digestion) of randomized sets of trinucleotide elements, each coding for a specific amino acid. While traditional synthesis with mononucleotide elements and degenerate codon strategies (e.g. NN(G/T) randomization) has inherent limitations due to codon bias and introduction of undesired amino acids or stop codons, the Slonomics® technique provides means to fully control the randomization process. The affinity maturation library was designed mainly with respect to the four clones with verified ErbB3-binding activity, whereas the complete set of first-generation binders also were considered. The second-generation affibody library was randomized in 12 positions to generate an oligonucleotide library with a total diversity of $7.4 \cdot 10^8$. After cloning to the staphylococcal display vector and transformation to S. carnosus, a cell-displayed library containing $1.3 \cdot 10^7$ variants was generated. Sequencing results showed an expected codon distribution compared to the theoretical design and a very low proportion of
frame shifts, multiple inserts and undesired codons. To verify that the affibody library was functionally displayed on the cell surface, a portion of the library was incubated with fluorescently labeled HSA for flow cytometric analysis of the ABP binding. The results demonstrated that around 72% of the library expressed full-length recombinant proteins on the cell surface.

In order to select novel affibody molecules with improved affinity for ErbB3, the staphylococcal-displayed library was incubated with biotinylated ErbB3-Fc fusion protein, as well as fluorescently labeled streptavidin and HSA and subjected to four rounds of FACS sorting with increasing stringency between rounds (Fig. 6.9). After the last round, sequencing of randomly picked clones yielded 443 unique sequences.

**Fig. 6.9.** Density plots showing the results from flow-cytometric sorting of Sc:ZErB3LIB. FL-4 channel fluorescence intensity corresponding to surface expression level (monitored via HSA binding) on the x-axis and FL-2 channel fluorescence corresponding to ErbB3 binding on the y-axis. The density plots are showing the staphylococcal library before flow-cytometric sorting round 1, 2, 3 and 4, respectively, with regions used in gating.
and among these, 45 clones appeared more than once. Since the most abundant clones were expected to have the highest target affinity due to the quantitative sorting properties, the clones that appeared more than once were selected for further characterization. In a whole-cell ELISA using flow cytometry, the 45 clones as well as one first-generation binder (Z01820) were ranked with respect to their ErbB3-binding activity. All analyzed clones showed ErbB3-binding activity in the assay and furthermore, 43 demonstrated a higher ErbB3-binding signal in comparison to the first-generation binder. Moreover, the four most abundant clones after sequencing also showed the highest signals in the assay, again confirming that flow cytometric sorting is truly quantitative.

The apparent equilibrium dissociation constants (K_d) for the four most promising candidates Z05405, Z05413, Z05416 and Z05417, and one of the first generation binders (Z01820), were determined by on-cell analysis (Fig. 6.10). Staphylococcal cell
populations displaying the different affibody variants were individually incubated with varying concentrations of biotinylated ErbB3-Fc and stained with fluorescently labeled streptavidin for analysis using flow cytometry. Upon plotting the obtained fluorescence intensities against the target concentrations, the apparent $K_d$ for each variant was determined. The four second-generation binders showed affinities for ErbB3 ranging from 1.8-3.0 nM, an improvement of affinity of around 15-fold compared to the first generation binder.

In conclusion, this study presents a combined approach using both phage and staphylococcal display in order to successfully isolate affibody molecules with nanomolar affinity for ErbB3. A set of 21 first-generation binders was generated using phage display, hence utilizing the higher library complexity obtained in this system. Based on the obtained sequences, a second-generation oligonucleotide library was generated using the novel Slonomics® technique that allows for a highly controlled randomization by incorporation of trinucleotide elements during synthesis. Utilizing the fine affinity discrimination capacity of the staphylococcal display system, affinity maturation was performed in order to quantitatively isolate 45 second-generation binders with confirmed ErbB3 binding activity. Upon on-cell affinity determination, four affibody molecules demonstrated binding affinities for ErbB3 and, for the strongest binder, a more than 15-fold improvement as compared to the first-generation binders.
6.5 Future perspectives

Combinatorial protein engineering has become an important principle that allows for rapid generation of novel or improved affinity proteins. As described in chapter 4 of this thesis, selection from combinatorial libraries is today facilitated with a number of available systems, to which our contribution is the staphylococcal display system.

In general, a high complexity is considered to increase the probability of finding high-affinity binders in a randomized library. In this respect, systems that require an in vivo step are limited by the transformation frequency for that species. Phage displayed libraries are typically in the size range of $10^9$-$10^{10}$ while yeast and staphylococcal displayed libraries generally have at least a 10-fold to a 100-fold lower complexity. In sharp contrast, in vitro systems such as ribosome display or mRNA display have been used to generate libraries with a complexity of $10^{13}$. The library size is however not the only parameter to consider and although not competing in complexity, cell-display systems provide several attractive features that are not achievable in other systems. First, the multivalent display on the cell surface together with flow cytometric sorting provides a quantitative selection based on the relative affinities for each library member. Cell-based systems therefore show a high capacity to discriminate between binders with even minor differences in their relative affinities. Second, after selection, these systems provide means for straightforward characterization of candidate clones directly on the cell surface without any requirement for subcloning, expression and purification that might become a major hurdle if a large set of candidates are to be evaluated. The latter is becoming important as the increasing use of affinity proteins in therapeutic and diagnostic applications put high demand on additional properties, e.g. specificity, to be considered at an early stage.

In two projects included in this thesis, staphylococcal surface display was employed in two successful selections to isolate novel affibody binders, proving the potentials for the system. Nevertheless, although the transformation frequency was improved around 10,000-fold in paper I, the achievable staphylococcal-displayed library complexity is still not to be compared with the in vitro systems or even phage display.
It would nevertheless be possible to consider staphylococcal libraries in the range of $10^8$-$10^9$, which would be sufficient for naïve libraries, especially using trinucleotide approaches for library synthesis in order to avoid unwanted or biased amino-acids and hence, decrease the theoretical complexity. In addition, a possible strategy to increase the total diversity using staphylococcal display is by applying a combined approach using a different system in the first selection cycle. In fact, paper II describes a pre-selection strategy using one round of phage display prior to staphylococcal selection in order to sample a naïve affibody library with a diversity of around $10^9$. In general, combined approaches are becoming increasingly popular since the most attractive features from different systems can be utilized.
Det är med en viss overklighetskänsla jag inser att jag kommit ända hit. Avhandling och disputation var ju något som skulle hända långt in i framtiden! Men så blev det några pek, en galen slutspurt och slutligen en aldeles egen liten blå bok. Jag kan inte annat än dela känslorna med Charles P. Conrad när han besteg månen med Apollo 12: "Whoopie! Man, that may have been a short one for Neil, but it sure was a long one for me."

Det är givetvis så att bakom varje nybliven doktor finns ett antal personer som hjälpt till på vägen. Ett stort och innerligt tack till:


**John**, som med ett oändligt tålamod handledde mitt exjobb och därefter varit en outsäg ogiv inspirationsbank. Det har varit fantastiskt roligt att dela omsorgen över staffarna med dig! Dessutom har du fortsatt att vara superengagerad i mina projekt. Glöm aldrig att du är bäst! Stort tack för all hjälp under slutspurten!
Denise, min första exjobbare. Jag hann knappt blinka så blev det ett pek och jag är fortfarande lite chockad över hur fort det gick!

Magdalena, min andra exjobbare, du arbetade uttröttligt för att förverkliga drömmen om den optimala transformationsfrekvensen och ditt fantastiska arbete lärde oss alla att plasmider är riktigt knepiga typer. Det har varit jätteroligt att jobba tillsammans med dig och superkul att du ville stanna kvar och ta över vårdnaden av staffarna! Tack för allt arbete du lagt ner för att min avhandling skulle bli en realitet.

Alla PIs på avdelningen som visar så mycket engagemang och tillför en kompetens som vi andra får ta del av. Särskilt tack till Per-Åke för att du alltid tar dig tid med en extra funderare och Amelie för att jag fick en suverän granskning av avhandlingen.

Elin, Fredrik, Ingmarie och Lars på Affibody AB för ett givande samarbete i HER3-projektet.

Gamla och nya medlemmar och exjobbare i Stefans grupp. Jag kommer alltid att minnas våra roliga gruppträffar och resorna till Gotland och Skärgården! Stor kram till John, Henrik, Mikaela, Caroline, Andreas, Johan, Sara, Basia, Julia, Magdalena, Helena, Denise, Anna, Sverker, Alex, Josefine, Carmen och Filippa.

Emma L och Anna K, för att ni funnits som det bästa stöd jag kunnat ha under doktorandtiden och för allt roligt vi haft tillsammans! Tack för alla fester, luncher, fikastunder, befriande diskussioner (okej, kanske lite skvaller också) och den ovärderliga uppmuntran ni alltid gett mig. Som om det inte räckte så har ni också bjudit på benbrytande eskapader (på två, en, eller noll skidor) i Kittelfjäll och filmkvällar med Mr. Darcy et al. Ni är bäst! Kram!!

Julia, Sara och Cilla för filmkvällar (romance nights) och annat roligt som vi gjort.

Tack till alla nuvarande och föregående som gjort vårt skrivrum så trevligt: Hanna, Mikaela, Caroline, Johanna, Margareta, Henrik, Camilla, Peter, Erik m.fl. Tänk vad mycket hjälp jag fått från er, och så mycket roligt sällskap!
Nina Kronqvist

Sebastian, för alla våra diskussioner och för en superkul proteomics-kurs i Portugal. Tack vare dig vet jag nu att det är surfa jag ska göra på nästa semester. Per-Ake L, för djupgående analyser av det mesta, allt från bibliotekskonstruktion till livsåskådning, samt för trevligt sällskap på immunologikursen och i Prag. Feifan, för always being so nice and eager to help. Also, thanks for introducing me to genuine Chinese cooking! Joije, allas vår enzym-langare som alltid livar upp stämningen i labbet.

Bahram, för all hjälp med sekvensering även om det ibland (okej, oftast) råkat bli i sista sekunden.

Tack till alla roliga, snälla och hjälpsamma personer på labbet. Det hade varit betydligt svårare och tråkigare utan er! Jag vill också tacka alla på plan 3 för den underbara attityden att man alltid ställer upp för varandra.

Diana och Rosa, för alla mysiga pizzakvällar och fikastunder och som gett perspektiv till kth-livet.

Jag älskar er hela bunten!

Sist och minst, kissekatten Smilla som hållit mig sällskap under avhandlingsskrivandet, och hållit mina pek varma genom att sova på dem i tur och ordning.
References


Ghetie, V. & Ward, E. S. (1997) FeRn: the MHC class I-related receptor that is more than an IgG transporter. *Immunol Today*. **18**: 592-598.


Nina Kronqvist  77


