Cell line and protein engineering tools for production and characterization of biologics

Anna-Luisa Volk

Kungliga Tekniska Högskolan, KTH
Royal Institute of Technology
School of Biotechnology
Stockholm 2017
“Learning is the only way to turn failure into success.”

Museum of failure, Helsingborg, Sweden
Public defense of dissertation

This thesis will be defended on September 29, 2017 at 10:00 in E3, Osquars backe 14, KTH Campus, Stockholm, for the degree of “Teknologie Doktor” (Doctor of Philosophy, PhD) in Biotechnology.

Respondent:
Anna-Luisa Volk, MSc in Molecular Biotechnology
Division of Proteomics and Nanobiotechnology, School of Biotechnology, KTH – Royal Institute of Technology, Stockholm, Sweden

Faculty opponent:
Prof. Alan Dickson Manchester
Institute of Biotechnology, Faculty of Science and Engineering, University of Manchester, United Kingdom

Evaluation committee:
Dr. Sara Mangsbo
Department of Immunology, Genetics and Pathology, Uppsala University, Sweden

Dr. Morten Nørholm
DTU BIOSUSTAIN, Novo Nordisk Foundation Center for Biosustainability, Danmarks Tekniske Universitet, Kongens Lyngby, Denmark

Dr. Yun Jiang Senior
Scientist Project Team Leader Upstream, Drug Design & Development, Swedish Orphan Biovitrum AB, Solna, Sweden

Chairman:
Prof. Per-Åke Nygren
Division of Proteomics and Nanobiotechnology, School of Biotechnology, KTH – Royal Institute of Technology, Stockholm, Sweden

Main supervisor:
Dr. Johan Rockberg
Division of Proteomics and Nanobiotechnology, School of Biotechnology, KTH – Royal Institute of Technology, Stockholm, Sweden

Co-supervisor:
Prof. Mathias Uhlén
Division of Proteomics and Nanobiotechnology, School of Biotechnology, KTH – Royal Institute of Technology, Stockholm, Sweden
Abstract

Our increasing understanding of disease mechanisms coupled with technological advances has facilitated the generation of pharmaceutical proteins, which are able to address yet unmet medical needs. Diseases that were fatal in the past can now successfully be treated with novel biological medications improving and prolonging life for many patients. Pharmaceutical protein production is, however, a complex undertaking, which is by no means problem-free. The demand for more complex proteins and the realization of the importance of post-translational modifications have led to an increasing use of mammalian cells for protein expression. Although improvements in development and production have allowed for unprecedentedly high product titers, the costs required for the development of pharmaceutical proteins still are far greater than those for conventional, small molecule drugs. To render such treatments affordable for healthcare suppliers and assist in the implementation of precision medicine, further progress is needed.

In five papers this thesis describes strategies and methods that can help to advance the development and manufacturing of pharmaceutical proteins. Two platforms for antibody engineering have been developed and evaluated, one of which allows for efficient screening of antibody libraries whilst the second enables the straightforward generation of bispecific antibodies. Moreover, a method for epitope mapping has been devised and applied to map the therapeutic antibody eculizumab’s epitope on its target protein, complement component 5. In a second step it was shown how this epitope information can be used to stratify patients and, thus, contribute to the realization of precision medicine. The fourth project focuses on the cell line development process during pharmaceutical protein production. A platform is described combining split-GFP and fluorescence-activated droplet sorting, which allows for the efficient selection of highly secreting cells from a heterogeneous cell pool. In an accompanying study, the split-GFP probe was improved to enable shorter assay times and increased sensitivity, desirable characteristics for high-throughput screening of cell pools.

In summary, this thesis provides tools to improve design, development and production of future pharmaceutical proteins and as a result, it makes a contribution to the goal of implementing precision medicine through the generation of more cost-effective biopharmaceuticals for well characterized patient groups.

Keywords: Pharmaceutical proteins, precision medicine, antibody engineering, epitope mapping, cell line development, split-GFP
Popular science summary

Proteins are important building blocks of our body. Each cell in our body consists of a myriad of proteins. There, they fulfill a variety of different functions. Keratin, for example, is a protein that gives our hair and nails structure. Insulin signals cells in our body to absorb more sugar and, thus, keeps our blood sugar level constant. Antibodies are proteins with an important function in our immune system. They recognize pathogens and other foreign particles and help to render them harmless.

The body is able to produce its own proteins. The instructions for this are stored in the genetic information, the DNA, in each cell of the body. In some cases, however, these instructions, the information stored in the DNA, is defective, which can lead to a protein being produced incorrectly or not at all. Considering the important tasks of proteins, this can have disastrous consequences for our health. A well-known example is diabetes, which can be attributed to the absence of insulin. In such cases, one can help the body and administer the missing or defective protein from outside. In the case of diabetes, the missing insulin is injected.

One approach to replacing a protein is to produce it outside the body. To this end, genetic engineering is often used today. This means that the protein’s building instructions, stored on a piece of DNA, the gene, are copied and inserted into a host cell, which now produces the protein instead of the body’s cells. The so produced protein can then be recovered and administered to a patient as a medicine. A drug manufactured in this way is also referred to as biopharmaceutical or biologic.

Advances in genetic engineering have enabled not just copying the protein’s building instructions, but also modifying this information in a very targeted manner so that the host cell produces a protein with new properties. At first glance, this may be a scary thought, and you may ask yourself what use this might have. Let me give you an example!

Cancer is a very malicious disease. It arises when cells in the body suddenly begin to multiply and spread uninhibitedly. Because the cancer is made up of our body’s own cells, our immune system does not recognize cancer cells as foreign particles and does not fight them. However, it would be very convenient if one could get the immune system to fight and kill cancer cells in the same way as it destroys pathogens. This is where genetic engineering comes in. As mentioned in the beginning, antibodies are important proteins of our immune system. Genetic engineering can be used to alter the building instructions for antibodies in such
way that they recognize cancer cells instead of pathogens. Thereby, the immune system is made aware of the cancer and begins to fight it.

This is only one example of how antibodies can be used as drugs. The last three decades have shown that antibodies are a very versatile biopharmaceutical. In my doctoral thesis, I have devised strategies for the development of even more efficacious antibody drugs, for the improved production of such biopharmaceuticals, and for evaluating whether a patient is likely to benefit from a certain antibody medication or should better be given an alternative drug.

**Populärwissenschaftliche Zusammenfassung**


Der menschliche Körper kann seine Proteine selbst herstellen. Die Anleitung dafür ist in der Erbmasse, der DNA, in jeder Körperzelle gespeichert. In manchen Fällen ist diese Anleitung, die in der DNA gespeicherte Information, allerdings fehlerhaft, was dazu führt, dass ein Protein falsch oder gar nicht produziert wird. Das kann verheerende Folgen für die Gesundheit haben in Hinsicht auf die wichtigen Aufgaben, die Proteine erfüllen. Ein bekanntes Beispiel ist die Zuckerkrankheit, Diabetes, die auf das Fehlen von Insulin zurückgeführt werden kann. In solchen Fällen kann man dem Körper helfen und das fehlende oder fehlerhafte Protein von außen verabreichen. Im Fall von Diabetes wird das fehlende Insulin gespritzt.

Um ein Protein zu ersetzen, kann es außerhalb des Körpers produziert werden. Dazu bedient man sich heutzutage oftmals der Gentechnik. Das bedeutet, dass man die Protein-Bauanleitung, geschrieben auf einem Stück DNA, dem Gen, kopiert und in eine Wirtszelle einfügt, die stellvertretend für die Körperzelle nun das Protein produziert. Das so hergestellte Protein kann dann gewonnen und einem Patienten als Medizin verabreicht werden. Ein so produziertes Medikament wird auch Biopharmazeutikum genannt.
Fortschritte in der Gentechnik haben dazu geführt, dass man heute die Protein-Bauanleitung nicht bloß kopieren kann, sondern diese Information auch ganz gezielt verändern kann, sodass die Wirtszelle ein Protein mit neuen Eigenschaften produziert. Das mag auf den ersten Blick ein unheimlicher Gedanke sein und du magst dich fragen, wozu das gut sein soll. Lass mich dir ein Beispiel geben!


Populärvetenskaplig sammanfattning


Kroppen kan tillverka sina egna proteiner. Instruktioner för detta lagras i varje cell i det genetiska materialet, DNA. I vissa fall är dessa instruktioner, den information

En strategi för att ersätta ett protein är att tillverka det utanför kroppen. För detta ändamål använder man sig numera ofta av genteknik. Detta innebär att man kopierar proteinbygginstruktionerna, skrivna i en DNA-bit som kallas gen, och infogar de i en värdcell som då istället för kroppscellen tillverkar proteinet. Protein som producerats på det sättet kan sedan utvinnas och administreras till en patient som ett läkemedel som då också kallas proteinläkemedel.

Framsteg inom gentekniken har lett till att vi idag inte bara kan kopiera proteinets bygginstruktioner, utan även målmedvetet ändra denna information så att värdcellen producerar ett protein med nya egenskaper. Detta kan vara en skrämmande tanke vid första anblick och du kanske undrar vad det kan vara bra för. Låt mig ge dig ett exempel!


Detta är bara ett exempel på hur antikroppar kan användas som läkemedel. Under de senaste tre decennierna har forskare visat att antikroppar är mycket användbara som proteinläkemedel. I min avhandling har jag studerat hur man kan utveckla ännu effektivare proteinläkemedel, hur man kan förbättra tillverkningen av sådana proteinläkemedel och hur man tar reda på om en patient kommer att ha nytta av ett visst proteinläkemedel eller hellre skulle få ett annat läkemedel.
List of appended papers

This thesis is based on the five articles and manuscripts listed below, which are referred to in the text by their roman numerals. Full versions of these papers can be found at the end of this thesis.


* These authors contributed equally to the work.

All articles are reprinted with permission of the copyright holder.
Respondent’s contributions to appended papers

I  Contributed to production, purification and affinity determination of proteins. Analyzed and interpreted data together with co-authors. Contributed to writing and revising the manuscript.

II  Designed experiments and performed experimental work with help of co-authors. Wrote the manuscript with contributions of co-authors.

III  Designed and planned the experimental procedures together with co-authors. Performed experiments with help of co-authors. Analyzed the data and wrote the manuscript with contribution of co-authors.

IV  Designed and planned the experimental procedures together with co-authors. Performed experiments with help of co-authors. Analyzed the data and wrote the manuscript together with co-authors.

V  Performed all cell culture related experiments. Analyzed and interpreted data together with co-authors. Contributed to writing the manuscript.
Published work not included in the thesis

Abbreviations

3D Three-dimensional aa Amino acid
ADCC Antibody-dependent cellular cytotoxicity
aHUS Atypical hemolytic uremic syndrome
BITE Bispecific T-cell engager
bsAb Bispecific antibody
C3 Complement component 3
C5 Complement component 5
CDC Complement-dependent cytotoxicity
CDR Complementary determining regions
CHO Chinese hamster ovary
CTLA-4 Cytotoxic T-lymphocyte-associated protein 4
DAF Decay accelerating factor
DHFR Dihydrofolate reductase
DNA Deoxyribonucleic acid
E. coli Escherichia coli
EGF Epidermal growth factor
ELISA Enzyme-linked immunosorbent assay
Fab Fragment antigen binding
FACS Fluorescence-activated cell sorting
Fc Fragment crystallizable
FDA Food and Drug Administration
Fv Fragment variable
GFP Green fluorescent protein

x
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>GS</td>
<td>glutamine synthetase</td>
</tr>
<tr>
<td>HC</td>
<td>Heavy chain</td>
</tr>
<tr>
<td>LC</td>
<td>Light chain</td>
</tr>
<tr>
<td>LDC</td>
<td>Limiting dilution cloning</td>
</tr>
<tr>
<td>mAb</td>
<td>Monoclonal antibody</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>NK cells</td>
<td>Natural killer cells</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>PNH</td>
<td>Paroxysmal nocturnal hemoglobinuria</td>
</tr>
<tr>
<td><em>P. pastoris</em></td>
<td><em>Pichia pastoris</em></td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td><em>S. carnosus</em></td>
<td><em>Staphylococcus carnosus</em></td>
</tr>
<tr>
<td><em>S. cerevisiae</em></td>
<td><em>Saccharomyces cerevisiae</em></td>
</tr>
<tr>
<td>scFv</td>
<td>Single chain fragment variable</td>
</tr>
<tr>
<td>TAA</td>
<td>Tumor-associated antigen</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-type</td>
</tr>
</tbody>
</table>
Contents

Public defense of dissertation ........................................ i
Abstract ................................................................. ii
Popular science summary ............................................... iii
Populärwissenschaftliche Zusammenfassung ....................... iv
Populärvetenskaplig sammanfattning ............................... v
List of appended papers ............................................... vii
Respondent’s contributions to appended papers ................ viii
Published work not included in the thesis ......................... ix
Abbreviations .......................................................... x
Preface ........................................................................ xv

1 Proteins ................................................................. 1

1.1 Proteins – the building blocks of life .......................... 1
1.2 Proteins in disease .................................................. 3
  1.2.1 The ErbB receptor family in cancer ....................... 4
  1.2.2 Complement component 5 in PNH and aHUS ............ 6
1.3 Proteins as therapeutic tools .................................... 7
  1.3.1 Antibody structure and function .......................... 8
  Antibodies as therapeutic tools .................................. 10
  1.3.2 Precision medicine ......................................... 12
  1.3.3 Immuno-oncology, a novel way of cancer therapy ....... 13
  1.3.4 Challenges of therapeutic proteins ....................... 14
1.4 Proteins as analytical tools ..................................... 15
  1.4.1 Antibodies as research tools .............................. 16
  1.4.2 GFP and split-GFP ........................................ 16
## CONTENTS

2 Protein engineering and characterization 19  
  2.1 Protein engineering strategies ................................. 20  
    2.1.1 Display technologies .................................. 22  
  2.2 Antibody engineering ........................................... 23  
  2.3 Bispecific antibodies ........................................... 25  
  2.4 Engineered protein scaffolds for molecular recognition .......... 27  
  2.5 Epitope mapping ............................................... 28  

3 Protein production 31  
  3.1 Choosing a production host ...................................... 32  
    3.1.1 Protein expression in *E. coli* ............................... 32  
    3.1.2 Protein expression in yeast ................................. 33  
    3.1.3 Protein expression in mammalian cells ....................... 35  
  3.2 Cell line development and engineering .......................... 36  
  3.3 Advances in cell culture technology ............................ 38  

Present investigation 41  
  Paper I: Combining phage and staphylococcal display for improved antibody discovery .................................................. 42  
  Paper II: Generation and mechanism of action of bispecific AffiMab molecules .................................................. 45  
  Paper III: Epitope mapping of eculizumab enables precision medicine 49  
  Paper IV: Combining split-GFP and droplet microfluidics for improved cell line development .......................... 53  
  Paper V: Improving speed and sensitivity of split-GFP by chromophore pre-maturation .................................. 57  
  Concluding remarks and future perspective .......................... 61  

Acknowledgements 63  

Bibliography 66
Preface

The terms “biologic”, “biopharmaceutical”, “biological” and “biotechnology medical product” are often used synonymously in scientific literature, but can have slightly different meaning to different authors [1, 2]. The American Food and Drug Administration (FDA) defines biologics as medical products that are used for treating, preventing or diagnosing diseases and that are derived from various natural sources [3, 4]. It is the latter that differentiates them from conventional, chemically synthesized drugs. Examples of biologics are vaccines, blood, gene therapy, and proteins produced by genetic engineering.

The following chapters will mainly discuss biopharmaceutical proteins, with a strong focus on their use in therapy but will also give a short digression on the non-pharmaceutical use of proteins. To grasp the full scope of biopharmaceutical proteins, we first have to understand what proteins are and how they can be used before we can take a closer look in how to design, develop and produce novel (biopharmaceutical) proteins.
Chapter 1

Proteins

Most people know of proteins due to the nutritional information present on food packages. Proteins are often regarded as a homogenous substance that should be part of a healthy diet. When taking a closer look, however, it becomes clear that proteins are far from homogenous. They come in all kinds of shapes and sizes and carry out a vast number of different tasks in the human body.

This first chapter will give a short introduction about proteins, which are one of the building blocks of cells, and discuss their involvement in disease. As well as being a potential cause of disease, proteins can also be exploited for treatment of different conditions and moreover, they present a versatile tool for use in molecular biology.

1.1 Proteins – the building blocks of life

The name protein, from the Greek proteios, meaning primary, was coined in 1838 by Jacob Berzelius in a correspondence with Gerardus Johannes Mulder [5, 6]. Mulder found in his studies that all of the proteins he analyzed had the same elemental composition but he wrongly came to the conclusion that they all consisted of a common, very large molecule and only differed in the amount of sulfur and phosphorus they contained [6, 7]. He also hypothesized that “this is the food-stuff of the whole animal kingdom and is probably formed only by plants” [6]. Thanks to Mulder’s and others’ continued research we know today that proteins consist of building blocks called amino acids and that they fulfill an enormous
Figure 1: A protein’s structure can be described on four levels. a) The primary structure refers to the linear amino acid sequence. b) Folding of the linear sequence into distinct structural elements, as illustrated here by a beta sheet (red) and an alpha helix (blue), is termed the secondary structure. c) The spatial arrangement of these structural elements is described by the tertiary structure, as exemplified here with the structure of an antibody’s light chain, characterized by numerous beta-sheets. d) A protein’s quaternary structure is defined as the interaction of several polypeptide chains that form the functional protein, here illustrated by the structure of an antibody consisting of two heavy chains (blue) and two light chains (red).

number of different functions in all organisms, not just plants. For example, enzymes catalyze different reactions, hormones enable signaling between different cells, proteins such as hemoglobin, fulfill roles as transporters, whereas proteins like tubulin and keratin form stable structures that form the cellular cytoskeleton and hair, respectively. Antibodies, which will be discussed in more detail in chapter 1.3.1, are proteins with a central role in the human immune system and, thus, help the body defend itself against pathogens.

To carry out these various functions, proteins must have the ability to adopt different shapes and structures. A protein’s structure is generally described on up to four different levels, namely, primary, secondary, tertiary and quaternary protein structures (Fig. 1). The primary structure of a protein refers to a protein’s amino acid (aa) sequence, i.e. the order of the protein’s building blocks. From a repertoire of 20 different, so-called natural amino acids, these building blocks are linked together to form a linear polypeptide chain and the information encoding the precise amino acid sequence is found in the DNA. A gene coding for a particular protein is transcribed into messenger RNA (mRNA) by the cellular machinery,
which then is translated into a polypeptide chain forming the primary protein structure. This transfer of information from DNA to protein is referred to as the central dogma of molecular biology [8]. All of the different amino acids share a common framework consisting of a central C-atom linked to an amino-group, a carboxylic acid group, a hydrogen atom and a differing side chain. It is this side chain that varies between the different amino acids and confers specific physical and chemical properties. As a result of these properties, the amino acids cause the polypeptide chains to fold into distinct structural elements [9]. These elements, including alpha helices, beta sheets, beta turns, and omega loops, are termed the secondary structures of proteins (1b) [10]. The tertiary protein structure describes the interaction of these secondary structural elements within a protein (1c). Proteins that contain several polypeptide chains also have a quaternary structure, which describes the relationship between the different protein subunits (1d).

As mentioned at the beginning, proteins are involved in innumerable biological processes. Many of these rely on interactions between different proteins, like the binding of a ligand to its receptor or an antibody to its antigen. To gain a better understanding of these interactions and their potential implications in disease, knowledge of a protein’s three-dimensional (3D) structure is desirable. Although it is known that the protein’s amino acid sequence determines its tertiary structure [9], in times where whole-genome sequencing can readily provide the primary structure of a protein from a DNA sequence, the de novo prediction of a protein’s structure from its amino acid sequence still remains a challenge [11, 12]. As such, scientists revert to experimental ways to try and determine protein structures and interaction sites, e.g. by x-ray crystallography, nuclear magnetic resonance (NMR) or mutational analysis (see chapter 2.5 for more details exemplified by the antibody-antigen interaction). This structural data is collected and made publicly available in the protein data bank (PDB) [13], where it can be viewed and analyzed to help unravel a protein’s role in disease and to guide the design of novel therapeutic and biotechnological tools, which are described in more detail in the following chapters.

1.2 Proteins in disease

As mentioned in the previous chapter, proteins are involved in virtually all biological processes in the human body. Via their roles in transporting, signaling,
providing structural support and immune defenses, and through regulating gene expression amongst other functions, proteins form a unique machinery ensuring that cells and organisms correctly function, even to the extent that they control when it is time for a cell to replicate or to die [10]. The lack, excess or malfunction of a protein can disturb this carefully controlled interplay and be detrimental to a cell and theoretically the whole organism. Consequently, proteins have a central role in many diseases. An exhaustive discussion of proteins in disease is practically impossible and doubtlessly outside the scope of this thesis. Therefore, the following sections will focus on proteins with relevance to the appended articles; the function of the ErbB-family proteins in cancer and the role of complement component 5 (C5) in paroxysmal nocturnal hemoglobinuria (PNH) and atypical hemolytic uremic syndrome (aHUS).

1.2.1 The ErbB receptor family in cancer

The ErbB receptor family, also referred to as the Her-family or epidermal growth factor receptor family, comprises of four members; EGFR/Her1/ErbB1, Her2/ErbB2, Her3/ErbB3 and Her4/ErbB4. All of these are protein tyrosine kinase receptors with very similar structures consisting of an extracellular region with four subdomains, a single pass trans-membrane region and an intracellular tyrosine kinase domain [14–16]. Except for Her2, ligands are known for all ErbB receptors [15, 17]. To date, eleven different EGF-like ligands have been identified, some of which have the ability to bind to different receptors, while others are specific for one receptor alone [15]. Most of the time, the receptors are in a non-activated, so-called tethered conformation (Fig. 2). In a few cases, the tether is broken, letting the receptor adopt an activated conformation preferred for ligand binding. Ligand binding to domains 1 and 3 stabilizes the extracellular domain of the receptor in the activated conformation, which enables homo- or heterodimerization with other family members via interaction with a exposed, so-called dimerization arm on domain 2 (2) [18–21]. Her2, lacking a ligand, is thought to be permanently active and, thus, regarded as the primary heterodimerization partner [22, 23]. Dimerization also brings the internal kinase domains together leading to reciprocal, allosteric activation and phosphorylation of tyrosine residues on the C-terminal tail of the kinase domain [15, 24]. These tyrosine residues serve as docking sites for downstream signaling molecules, which induce signaling cascades predominantly via the PI3K-AKT and MAPK pathways, thereby stimulating processes such as cell proliferation, survival, differentiation and migration [25, 26]. These vital func-
Figure 2: A conformational change allows for activation of ligand-binding ErbB family members and subsequent intracellular signalling. a) The ligand-binding family members mostly adopt a tethered, inactive conformation. b) In some receptors the intramolecular constraint is broken letting it adopt a more open conformation, which is preferred for ligand binding. c) Ligand binding to domains I and III stabilizes this conformational change exposing the dimerization arm on domain II and pushes the equilibrium towards the untethered conformation. d) Receptor homo- or heterodimerization leads to phosphorylation of the intracellular kinase domain and subsequent signaling cascades, mainly through the PI3K-AKT and MAPK pathways driving cell survival, proliferation, migration and differentiation.

As early as 1984, the first indications appeared suggesting that this receptor family might play a role in tumorigenesis [27, 28]. Overexpression of EGFR, Her2 or Her3 and mutation of EGFR, as well as autocrine ligand expression have subsequently been observed in various solid tumors and correlate with poor prognosis [25, 26]. Autocrine ligand secretion leads to a higher activation of receptors and consequently stronger signaling and faster tumor cell growth [29] and a similar effect is seen with receptor overexpression. A greater number of receptors on the cell surface increases the chance of dimerization and downstream activation. In particular, the overexpression of Her2 is important in this context due to the constantly active state of the receptor which can homodimerize without the need for ligand binding [26, 29]. A deletion mutant of EGFR, EGFRvIII, has also been reported rendering the receptor constantly active [30, 31], whereas a role for Her4 in cancer has still not been conclusively established. However, different studies suggest that expression of Her4, in contrast to the other EGF-receptors, has a
positive effect on cancer patients’ survival [32–35].

Due to their proven role in tumorigenesis and particularly HER2’s selective over-expression on cancer cells, the ErbB receptor family presents an interesting therapeutic target. One therapeutic strategy is to inhibit signaling through these receptors. Current approaches include the inhibition of the internal kinase domains, the prevention of ligand binding and receptor dimerization, or the fixation of the receptor in the inactive conformation. The methods through which these are achieved are the topic of chapter 1.3.

1.2.2 Complement component 5 in PNH and aHUS

PNH and aHUS are principally rather different disease. However, they both have in common a misguided complement system. The complement system is part of the immune system and provides a link between the adaptive and innate immune response [36]. It comprises of a group of approximately 50 serum proteins that act in cascades to defend the body against pathogens [37, 38]. Three different ways of complement activation have been described; the classical pathway is induced by the recognition of antibody-antigen-complexes, whereas pathogen-specific carbohydrates trigger the lectin pathway and the alternative pathway is initiated by spontaneous complement component 3 (C3) cleavage [36, 37]. All three pathways converge at C3 convertase, a protein complex that catalyzes the cleavage of C3 protein into C3a and C3b. Addition of C3b to C3 convertase generates complement component 5 (C5) convertase, which cleaves C5 into C5a and C5b. Both convertases lead to a rapid amplification of the immune response [38] and intensify the final outcome. C5b initiates the formation of the membrane attack complex (MAC, C5b-9), a lytic pore, leading to lysis of the pathogen. Additional outcomes from the complement cascade include the deposition of the opsonin C3b on the pathogen surface marking it for phagocytosis and recruitment of effector cells to the site via the proinflammatory anaphylatoxins C3a and C5a [37, 39]. In order to protect the body from damage by any of these processes, the complement system is strongly regulated [37, 38]. Consequently, the results are quite detrimental when any of these regulators fails.

One example, in which the regulation of the complement system fails, is the autoimmune disease PNH. PNH patients lack the proteins DAF (decay accelerating factor; CD55) and protectin (CD59) on the cell surface. DAF promotes the dissociation of C3 convertases and protectin prevents MAC formation, both of which pro-
Proteins as therapeutic tools

Protect the body’s cells from complement attack [36, 37]. However, the cause of PNH is not a mutation in the proteins DAF and protectin, but rather an acquired somatic mutation in the enzyme phosphatidylinositol N-acetylg glucosaminyltransferase subunit A (PIGA), which is needed for the synthesis of glycosylphosphatidylinositol (GPI) anchors that attach DAF and protectin to the cell surface [40, 41]. As a result of this, DAF- and protectin-deficient cells, especially erythrocytes, are sensitive to complement attack leading to hemolysis, which ultimately results in anemia, renal failure and thrombosis [42, 43]. The latter is the most common cause of mortality in PNH [40, 42, 44].

Similar to PNH, aHUS, another complement related disease, manifests in hemolytic anemia, thrombocyte reduction (thrombocytopenia) and renal failure [45]. aHUS is a genetic, chronic disease characterized by a dysregulated alternative complement pathway [45, 46]. However, the molecular causes of aHUS are far more diverse than those of PNH. Genetic mutations in alternative pathway proteins have been found in 60% of aHUS patients [47, 48]. Also, the presence of autoantibodies that neutralize complement regulatory proteins have been associated with aHUS [47, 49]. Many studies indicate that an external trigger is required for aHUS to manifest, although the significance in many cases remains unknown, which complicates the demarcation of aHUS from other disorders with similar disease patterns [45, 47].

Historically the treatment of both PNH and aHUS has relied on immunosuppressive and anticoagulation therapies [43, 44, 49], plasma infusion or exchange, or repeated dialysis for aHUS [47, 49] and blood transfusion in PNH [43, 44]. With the better understanding of the pathogenesis and the arrival of new recombinant protein therapeutics, more targeted treatment options have arisen and in particular, antibody-based therapeutics. The promise and challenges of this new era will be covered in the next section.

1.3 Proteins as therapeutic tools

With the growing understanding of disease mechanisms and of protein involvement in disease came the natural aim to replace faulty proteins or render them innocuous. One significant milestone was the isolation of insulin from bovine and porcine pancreas by Frederick Banting and coworkers in 1922 to treat diabetic patients, which earned him the Nobel Prize in 1923 [50, 51]. However, proteins isolated from natural sources are a limited resource that are quite expensive and
potentially immunogenic [52], therefore, the advent of recombinant DNA technology marked a turning point in the history of biotherapeutics and is sometimes referred to as the beginning of the “biologics” era [2, 53]. Recombinant DNA technology allows the insertion of an isolated gene into an organism to let this organism produce the so-called recombinant protein. Again it was insulin that was pioneering in this new era. By making the bacterium *Escherichia coli* (*E. coli*) express the human insulin gene, insulin was produced as the first recombinant protein in 1979, which gained approval as therapeutic by the American Food and Drug Administration (FDA) in 1982 [52, 54]. Since the production of insulin in *E. coli* many different production hosts have been employed, which will be reviewed in more detail in chapter 3. Today, there are over 400 approved recombinant proteins [55] and this number is constantly growing with biopharmaceuticals expected to account for 25% of the global pharmaceutical sales in 2017 [56]. One protein class in particular has gained special attention: antibodies. Of 62 FDA-approved protein therapeutics introduced between 2011 and 2016, almost 50% were antibodies [57]. Antibodies such as Humira (adalimumab) and Herceptin (trastuzumab) also accounted for 50% of the ten best-selling drugs in 2016 [58], which is reason enough to take a closer look at this particular protein class.

### 1.3.1 Antibody structure and function

Immunoglobulins, commonly known as antibodies, are part of the human immune system. They can specifically bind to structures that are recognized as foreign, and thereby marking them for elimination from the body.

The first indication of the existence of proteins that confer immunity was provided by Behring and Kitasato in 1890 when they published their results on “[t]he mechanism of immunity in animals to diphtheria and tetanus,” showing that the cell-free serum of an immunized rabbit is sufficient to protect a non-immunized mouse against tetanus toxin [59, 60]. These findings let Behring and Kitasato arrive at the conclusion that “Blut ist ein ganz besonderer Saft” (“Blood is a very unusual fluid”) [59]. It was in his side-chain theory in 1891, that Paul Ehrlich coined the name antibody to describe a molecule secreted by cells in response to encountering a toxin. The work of Astrid Fagraeus, Frank Macfarlane Burnet and many others paved the road for today’s understanding of antibodies as proteins secreted from specialized plasma cells, known as B-cells, after encountering an antigen [60]. Antibodies’ exceptional ability to bind almost any target structure stimulated research on how this function relates to an antibody’s structure. Even
Proteins as therapeutic tools

Figure 3: Schematic Y-structure of an IgG antibody. An antibody consists of two identical heavy (blue) and two identical light chains (red) with C-terminal constant regions (CH1-3 and CL) and N-terminal variable regions (VH and VL), also referred to as Fv. The arms of the Y are also known as the Fab region and the stem is termed the Fc region.

before the first crystal structure of an antibody was solved by Silverton et al. [61], the function of different antibody fragments was elucidated by Edelman and Porter among others [60, 62].

The characteristic Y-shaped structure of an antibody (Fig. 3) is probably one of the most well-known protein structures. It consists of two identical heavy chains (HC) and two identical light chains (LC) held together by disulfide bonds. Both chains have two distinct regions, namely an N-terminal variable region and a constant region. While the amino acid sequence of the constant region is conserved within an antibody subclass, it is the primary sequence of the variable region that is target specific and only common for antibodies from the same B-cell clone. Target-specificity and affinity is conferred by three highly variable stretches in the amino acid sequence of the variable region, referred to as complementary determining regions (CDRs). The heavy and the light chain’s variable regions together form the so-called variable fragment (Fv), which defines the binding site of the antibody, the so-called paratope. Each “arm” of the immunoglobulin is also referred to as fragment antigen binding (Fab), because of it’s ability to bind the antigen, in contrast to the fragment crystallizable (Fc), the stem of the Y, which does not bind the antigen but readily crystalizes at low temperatures [62]. However, the Fc also has a biological purpose; it confers so-called effector functions. By binding to different Fc receptors on human cells, immunoglobulins can attract different
effectors cells, e.g. opsonization of a pathogen by antibodies recruits phagocytotic cells and Fc binding to natural killer (NK) cells induces antibody-dependent cell-mediated cytotoxicity (ADCC). Antibodies can also initiate the complement cascade as described in chapter 1.2.2, which is referred to as complement dependent cytotoxicity (CDC) [37]. Furthermore, binding to the neonatal Fc receptor contributes to the long in vivo half-life of immunoglobulins. The Fc functions of an antibody depend on its isotype. There are five different immunoglobulin isotypes in human: IgG, IgA, IgM, IgD and IgE. The most frequent isotype in human blood is IgG, which itself can be grouped into four subclasses, IgG1 to 4, with differing propensities to bind various antigens and Fc receptors [10, 63]. The frequency is probably one reason why IgGs especially have been exploited as therapeutic, diagnostics and research tools [64]. Historically, IgG1 in particular, due to its ability to elicit both CDC and ADCC, was used as a therapeutic [50, 65]. These effector functions might, however, not always be desirable, which has lead to the use of IgG2 and IgG4 and the specific modification of Fc regions, known as Fc engineering, as illustrated by eculizumab in the next chapter [50, 66].

**Antibodies as therapeutic tools**

Antibodies’ long half-life in circulation and their ability to specifically bind virtually any antigen with high affinity and neutralize it via the different mechanisms explained above, qualifies them for therapeutic use. The most straight-forward way to generate antibodies is probably as Behring did at the end of the 19th century; immunize an animal and treat patients with serum from the immunized animal [60]. However, this procedure has some drawbacks. Besides antibodies against the desired antigen, serum also contains many other proteins and antibodies against additional antigens. These proteins could potentially trigger an immune reaction, which could be harmful to the patient. Even the purification of antibodies against the desired target from serum cannot eliminate the risk of immunogenicity, as they can still be recognized by the immune system as foreign. Furthermore, these antibodies are a limited resource, because only a finite amount of serum can be obtained from each animal. The purified, antigen-specific antibody preparation still contains a mixture of antibodies, each having different paratopes, as they were produced by different B-cells. For this reason, these antibodies are referred to as polyclonal. The desire to obtain an unlimited supply of identical antibodies derived from the same B-cell incited Köhler and Milstein in 1975 to develop a technique that fuses an antibody-secreting but
short-lived B-cell with an immortal myeloma cell [67]. This hybridoma technology marked the birth of monoclonal antibodies (mAbs). Since 1975, many monoclonal antibodies have been produced and the advent of recombinant DNA technology gave the antibody generation an additional boost. Not only could genes encoding human antibodies be transferred and produced in different organisms, but the technology also opened up for the targeted design of novel antibodies with new specificities and lower immunogenicity [68], and the generation of completely new antibody formats, like antibody-drug conjugates, antibody fragments and Fc fusion proteins [50]. A more detailed discussion about antibody engineering will be given in chapter 2.2.

As of January 2017, 68 monoclonal antibody products have been FDA-approved [69] against a variety of disease states including immunological and inflammatory disorders, with oncology applications taking the biggest share [52, 66, 70]. Some antibodies function mainly by occupying functionally important regions on a protein, while for others the Fc-mediated effector functions seem to be of bigger importance [52]. Here, two mAbs with relevance to the subsequently presented work will be described in more detail: trastuzumab, with the brand name Herceptin, and eculizumab, also known as Soliris.

Trastuzumab is a humanized monoclonal antibody targeting the Her2 receptor. It has been approved for the treatment of Her2-positive, metastatic breast cancer and gastric cancer [71]. Trastuzumab binds to domain IV of Her2’s extracellular region resulting in reduced tumor cell growth [65, 72, 73]. This effect is obtained through multiple mechanisms of action. Firstly, reduction in Her2 phosphorylation impairs downstream signaling and leads to subsequent cell cycle arrest [71, 74]. Secondly, inhibition of Her2 cleavage prevents the generation of a constitutively active Her2 variant termed p95 [71, 75]. Thirdly, being an IgG1 trastuzumab binding to Her2 recruits NK-cells and elicits ADCC [71, 76]. Although trastuzumab treatment leads to a prolonged survival of breast [77–79] and gastric cancer patients [80] recent reports about trastuzumab resistance [81, 82] underline the need for novel Her-targeted therapies.

Eculizumab is an example for a non-oncology-related mAb. By binding to complement component 5 (C5), it inhibits the cleavage of C5 into C5a and C5b and the subsequent formation of the MAC (see chapter 1.2.2). Eculizumab has been approved for the treatment of PNH and aHUS and prevents hemolysis and reduces transfusion requirements in these patients leading to an improved quality of life [83–86]. Eculizumab has been deliberately designed to trigger neither CDC
nor ADCC by containing a hybrid constant region with components from IgG2 and IgG4, which cannot bind Fc receptors or activate the complement system, respectively [42]. By targeting C5, MAC formation can be inhibited independently of the complement activation pathway. At the same time, the proinflammatory and opsonizing functions of C3a and C3b, respectively, are not compromised. Despite the clinical success, the high costs of eculizumab (approx. €460,000 for an adult per year [46]) hinder the use of eculizumab as first-line treatment [48, 87]. This highlights the need for additional treatment options and a careful analysis of which patients are going to profit from a particular treatment as described in the next paragraph.

1.3.2 Precision medicine

Precision medicine describes a concept where the appropriate disease treatment for a patient is chosen taking genetic, environmental and life-style factors into account [88]. In contrast to “personalized medicine”, the term precision medicine does not imply the development of unique, patient-tailored treatments. The idea of precision medicine is not new; blood typing as pre-requisite to transfusion has been performed for over a century [89]. However novel technologies that allow for the monitoring of genetic, proteomic or metabolomics markers, and databases which collate this information open up new possibilities for a more detailed classification of patients and diseases [88, 89]. The approach to map a patient’s molecular make-up and on that basis decide on a beneficial treatment holds the promise of both sparing unnecessary discomfort for the patient and reducing treatment costs [90, 91]. Oncology is at the forefront of this transition from a “one drug fits all” outlook, which is more or less the rationale behind chemotherapy, to “the right drug at the right time to the right patient” ideal, which is the key factor for targeted cancer therapy [92]. Once again, trastuzumab serves as an illustrative example. By assessing the Her2 levels in cancer tissues sampled from breast or gastric carcinoma patients, it is decided if the patient is likely to benefit from a trastuzumab therapy or should receive an alternative treatment [90]. A couple of targeted cancer therapies have been developed in this manner with companion diagnostics [92, 93]. However, they all struggle with tumor heterogeneity, i.e. the chosen medication may only be effective against a fraction of tumor cells and the subsequent growth of non-affected cells [92]. A more lasting effect might be achieved with combinations of different treatments [92] or by exploiting the natural cytotoxic properties of the immune system and directing them to the tumor [93].
The latter is the basic principle of cancer immunotherapy, which will be the focus of the following section.

1.3.3 Immuno-oncology, a novel way of cancer therapy

Immuno-oncology, also referred to as cancer immunotherapy, is a rather new approach in cancer treatment that holds a lot of potential for precision medicine. In contrast to classical, often mAb-based, targeted therapies, which directly inhibit proteins on the tumor cell surface, immunotherapy relies on the immune system to attack and eliminate tumor cells [91, 92]. Immuno-oncology comprises a variety of different strategies to harness the immune system for combating cancer mainly focused around T-cells, a type of lymphocytes involved in cell-mediated immunity. Generally, a T-cell mediated immune response against cancer cells starts with the release of tumor associated antigens (TAA), for instance from dying cancer cells, which are consequently taken up and presented on the surface of antigen-presenting cells such as dendritic cells. If these antigens are recognized by a T-cell through its T-cell receptor, the T-cell is activated and can subsequently infiltrate the tumor, bind to tumor cells via the T-cell receptor and kill them, resulting in the release of more TAAs to start a new cycle [94]. In cancer patients this sequence of steps, referred to as the cancer immunity cycle [94], is often impaired, e.g. due to limited release or presentation of TAAs and reduced T-cell activity within the tumor environment. The various immuno-oncology strategies aim to take corrective actions to restore a functioning cycle. These strategies can generally be divided in passive and active approaches [91].

Passive treatments facilitate the engagement of naturally occurring T-cells in preventing tumor growth. Maybe the most prominent representatives of this group are check-point inhibitors; mAbs that bind surface proteins on T-cells, like cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and programmed cell death protein 1 (PD-1), and in this way take the brakes off T-cells [91, 95]. Also, some small molecules have been reported to yield the same effect [91]. On the other hand, active immune-oncology treatment often involves a manipulation of T-cells in vitro and re-injection of selected cells into the patient. Adoptive cell transfer (ACT), the oldest method, involves isolation, expansion and re-injection of lymphocytes found in the patient’s tumor microenvironment. T-cell receptor therapy, an advancement of ACT omitting the need for surgery, takes T-cells from a patient’s blood and engineers them to express a T-cell receptor that recognizes a peptide specific for the patient’s tumor cell [96]. Because this method relies on
the presentation of these peptides, which is often downregulated in tumor cells, chimeric antigen receptor (CAR) T-cells were developed [97, 98]. CAR-T cells are engineered to express an antibody’s Fv on its surface that recognizes a TAA and is linked to intracellular signaling domains. Binding of the Fv to a cancer cell leads to T-cell activation and proliferation [98]. Besides the aforementioned T-cell strategies, there are also a variety of other active immuno-oncology treatments counted in the group, including cancer vaccines such as TAA-presenting dendritic cells, attenuated cancer cells or cancer proteins; oncolytic, i.e. tumor cell lysing viruses; and bispecific T-cell engagers (BITEs), engineered proteins that generate a physical link between a cancer and a T-cell [91].

Although most of the presented strategies are still in their infancy, with the first approved treatments and many ongoing clinical trials showing promising results (for a comprehensive review see [91]), immuno-oncology is referred to by some as a revolution in oncology [99] and presents a constantly growing branch within tumor therapy [91]. Besides the clinical success, these strategies with their inter-play of designed proteins and immune cells, as well as the engineering of cells to express certain proteins, give an excellent example of the huge potential of protein therapeutics and biologics. However, the use of proteins as therapeutic treatments still has some limitations and challenges to overcome, which will be covered in the following paragraph.

### 1.3.4 Challenges of therapeutic proteins

In the previous chapters we have established the enormous potential that protein therapeutics hold and in the last decades, many protein drugs have been successfully used in the treatment of a variety of diseases. However, the number of protein therapeutics that fail to reach the market far outweighs those that succeed [52]. This is partly due to challenges and limitations that are inherent in the development and administration of protein therapeutics. To date, protein therapeutics have to be administered in a manner that bypasses the digestive tract, with subcutaneous injection and IV infusion being the preferred routes [50]. Compared to taking a pill containing a small chemical drug, this involves much more discomfort for the patient, which might hamper patient compliance. Finding alternative routes of administration and reducing the number of injections by increasing the serum half-life of proteins is a field of active research [100, 101].

While the high specificity of protein therapeutics to their targets is one of their
biggest assets resulting in little off-target toxicity, the risk of the body mounting an immune response against a drug is higher for proteins than small chemical drugs [50]. Immunogenicity is generally regarded a bigger issue in protein replacement therapy because an immune response against the administered protein bears the risk that the natural protein will also be attacked and cleared, which can result in serious health problems. Whereas, in case of binding proteins like mAbs, an immune response merely leads to protein degradation and thus reduced efficacy [50, 52]. Generally immune reactions are often mounted against proteins of non-human origin. This has fueled efforts to engineer both proteins and their production hosts to become more human-like, and particularly with therapeutic antibodies this trend towards humanized and fully human mAbs is measurable [50]. Also the prediction of potential T-cell epitopes on a therapeutic protein using specialized algorithms, and the subsequent removal of such epitopes by protein engineering is becoming a common strategy to prevent immunogenicity (see chapter 2.2) [102, 103].

The quest for more human-like proteins, which are correctly processed, folded and glycosylated, has propelled the use of mammalian cells in this field, compared to the more traditional microorganisms being employed as production hosts (for a more detailed discussion of production hosts see chapter 3.1). This reorientation comes along with increased manufacturing costs as mammalian cells yield lower protein titers, have longer cultivation times and require more expensive growth medium than bacterial production organisms. Consequently, this leads to an increase in price for the final medicine and a higher financial burden on health care systems, which might culminate in an ethical dilemma where efficient treatment is available for a disease but the cost cannot be covered by the health-care system [52, 104]. This was the case for aHUS treatment with eculizumab in Sweden, where according to Swedish authorities the benefits did not justify the costs [87]. Pertinently, constant improvements in cultivation technology have contributed to higher protein yields, reducing the costs of goods [105–107] and ongoing efforts in cell line engineering hold much potential for the generation of novel, high-secreting production cell lines.

1.4 Proteins as analytical tools

Besides their use as therapeutics, as discussed in the previous chapter, proteins are also widely used and are indispensable as tools in life science laboratories. While
certain proteins might be needed for very specific problems, e.g. when aiming to elucidate the biological role of that particular protein, or for the determination of a protein’s 3D-structure [108], two proteins should be highlighted here that have gained more universal significance in the life sciences field, namely antibodies and the green-fluorescent protein.

1.4.1 Antibodies as research tools

Due to their outstanding ability to bind proteins with high specificity, antibodies have become an invaluable tool in a multitude of research applications. Both monoclonal and polyclonal antibodies are routinely used to identify, isolate, and quantify different proteins. For example, by immunoprecipitation, Western blotting or enzyme-linked immunosorbent assay (ELISA), the presence and quantity of an analyte in a sample can be measured and conclusions about protein-protein interactions can be drawn [109, 110]. Whole cells can also be stained with antibodies in immunocytochemistry, to demonstrate the localization of certain proteins on the surface of cells or even inside cells when the cells are permeabilized [109–111]. Combined with fluorescence-activated cell sorting (FACS), these methods can be used to separate different cell types, e.g. in a blood sample. When immobilized on a matrix, antibodies are additionally utilized to purify specific proteins from a mixture of different molecules, like cell lysates [109]. Many of these applications require the labeling of antibodies with enzymes, or radioactive or fluorescent probes to facilitate detection of the bound antibody. For some cases, especially in in vivo applications, the use of fluorescently labeled antibodies is not feasible, and here GFP, the main character of the next chapter, might come in handy.

1.4.2 GFP and split-GFP

The green fluorescent protein (GFP) from the jellyfish Aequorea victoria, a protein that emits green light upon excitation with blue light, was discovered in 1962 [112], but it was not until 1994, when GFP was recombinantly produced in E. coli and Caenorhabditis elegans (C. elegans), that its true potential was realized. GFP exhibits spontaneous chromophore formation without the need for any cofactors, enzymes or substrates besides oxygen, which enables the in vivo use of this protein in virtually any organism [113–115]. The fluorophore is formed by the cyclization and oxidation of the three amino acids Ser65, Tyr66, and Gly67, which are protected from external effects inside the characteristic β-barrel structure of
Proteins as analytical tools

Figure 4: The characteristic β-barrel structure of GFP. The GFP β-barrel consists of eleven β-strands and an axial helix that harbors the three amino acids forming the fluorophore (in green). The eleventh β-strand, which is taken out of the barrel in the split-GFP variant, is highlighted in blue.

GFP (Fig. 4) [116, 117]. Since the 1990’s, GFP has been extensively engineered and mutations have been introduced to improve the fluorescence and folding characteristics of this protein [115, 117]. Acknowledging the great impact that GFP had on molecular biology, Osamu Shimomura, Martin Chalfie and Roger Y. Tsien were jointly awarded the Nobel prize in chemistry in 2008 for the “discovery and development of GFP” [118].

Thanks to its stability and the possibility to non-invasively visualize cellular processes GFP has nowadays become a routine instrument in the biologist’s toolbox with very diverse applications as illustrated by the following examples. One of the first uses of GFP was to fuse it with other proteins to monitor their expression and localization in cells or to put the GFP-gene under control of a promoter of interest to visualize the promoter’s activity [113, 114, 117]. GFP is often included as part of expression plasmids and used to both calculate the transformation/transfection efficiency and to select for cells that have taken up a plasmid, e.g. by FACS [113].

pH-sensitive GFP variants have been generated that can serve as intracellular pH indicators [115, 117]. The use of GFP has also promoted our understanding of tumor growth and metastasis [119]. Together with other fluorescent proteins or GFP-derivatives with other fluorescent properties, protein interactions can be studied via Förster resonance energy transfer (FRET), i.e. when in close proximity, excitation of the donor fluorophore leads to activation of the second, acceptor fluorophore due to the excitation spectrum of the acceptor overlapping with the emission spectrum of the donor.
Proteins

However, for some proteins fusion of the full-length, approximately 27 kDa GFP can considerably affect the protein’s function or solubility. To circumvent this, the group around Geoffrey S. Waldo has developed a split-variant of GFP by separating the 11th β-sheet of GFP (GFP-11) from the rest of the β-barrel, termed GFP1-10 (Fig. 4) [120]. The short, 17-aa GFP-11 fragment is fused to the protein of interest and GFP1-10 is provided in solution or can be expressed intracellularly from a separate plasmid. These fragments, each non-fluorescent on its own, self-assemble to form a fluorescent protein complex [121]. Since it’s development, this technology has been utilized in many areas of application like monitoring protein solubility [121, 122], demonstrating protein-protein interactions [123, 124], measuring protein secretion [125], monitoring endosomal escape of biotherapeutics [126], and in various imaging applications [127, 128]. The generation of soluble, highly expressed split-GFP required significant protein engineering and combined, the GFP1-10 and GFP11 contain 19 aa mutations compared to wild-type (WT) GFP. How proteins are engineered and typical methods involved in this process are the subject of the next chapter.
Chapter 2

Protein engineering and characterization

Native proteins, also referred to as wild-type proteins, are adapted to fulfill their particular functions inside an organism but this often results in a rather poor suitability for industrial and pharmaceutical manufacturing, which often involves harsh conditions such as low pH or high salinity [129, 130]. The demand for better-suited proteins and biopharmaceuticals with new attributes has spurred protein engineering, defined as the purposeful modification of a protein to yield a protein with enhanced or completely new features or to simply further the understanding of a protein’s biology. Motivations for protein engineering differ and depend on the intended use of the protein, e.g. proteases are engineered to be functional in the harsh environment of a laundry detergent [131]. Although there are infinite applications for engineered proteins, this chapter will focus on the engineering of proteins for therapeutic use.

The history of therapeutic proteins is intrinsically linked to protein engineering. A constant striving for enhanced properties driven by clinical (e.g. improved efficacy and safety), scientific (e.g. better understanding of the mechanism of action), as well as economic rationale (e.g. reduction of production costs) has resulted in the vast number of different protein therapeutics that have been approved today [132]. Both minor changes in the protein sequence, i.e. optimizing single amino acids, and major modifications of the protein such as the fusion of two proteins, can confer improvements of protein properties. The different strategies for protein engineering are covered in the first section of this chapter. As a result of the domi-
nant role of antibodies amongst therapeutic proteins, antibody engineering is also a big area within the field of protein engineering and will therefore be discussed in a separate section. Protein engineering has further enabled the generation of completely new protein formats, some of which derived from antibodies, others based on so-called alternative scaffolds and yet others designed to engage two targets simultaneously, referred to as bispecifics. This chapter aims to give a short introduction to all of these different protein formats of affinity proteins.

Ultimately, the characterization of the newly engineered protein is an important step in the design process. Especially when developing antibodies or other affinity proteins, information regarding where these proteins bind to their target protein is desired. The process of identifying this binding site, i.e. the epitope, is known as epitope mapping, which is the topic of the last section in this chapter.

2.1 Protein engineering strategies

As mentioned in the introductory paragraph, the term protein engineering comprises many kinds of protein modifications ranging from single point mutations in the protein’s amino acid sequence to major truncations or extensions. But classification of different engineering techniques is rather diffuse. Generally, strategies are divided into rational design approaches and more random approaches, often referred to as directed evolution. In fact, these two approaches are not mutually exclusive and rather represent two different ends of a spectrum, with many commonly used approaches locating somewhere in the middle, therefore termed semi-rational approaches [133].

Historically, the beginning of the protein-engineering era is marked by the invention of site-directed mutagenesis in 1978 and it’s first use to study the mechanism of the enzyme tyrosyl–transfer RNA synthetase in 1982 [134, 135]. Until the arrival of high-throughput selection methods in the late 1980’s, rational design was the protein engineering approach of choice. Rational design relies on the understanding of structure-function relationships to guide the choice of modifications that are likely to result in a protein with desired properties. Potentially important amino acid residues are deduced from the protein structure, mutated, and the effect of the mutation is evaluated experimentally. This approach has greatly benefitted from the continued efforts to solve protein structures. While in the past this was mainly hypothesis driven, lately computational tools have facilitated the rational design of proteins. Using molecular modeling algorithms, novel protein sequences
Protein engineering strategies

have been devised that successfully fold into a desired structure [136, 137]. These computational tools allow for the evaluation of amino acid changes in silico on a scale not feasible experimentally [138] and enable the de novo design of proteins [137].

Strictly speaking, modifications like codon optimization to increase protein expression and protein fusions, with for instance polyethylene glycol (PEG) or with an antibody’s Fc region to increase the in vivo half-life of a protein, should also be regarded as forms of rational design but are rarely mentioned in this context [129].

However, the general applicability of rational design approaches is limited by the lack of three-dimensional structure for many proteins and our still incomplete understanding of the factors that govern protein folding and dynamics [138, 139]. An engineering strategy that does not rely on prior knowledge of a protein’s structure or its mechanism of action, is a process referred to as directed evolution. In directed evolution experiments, a protein’s gene is randomly mutated to generate a pool of protein variants, a so-called library. Protein mutants with desired properties are isolated from this library via screening or selection [136]. Commonly employed gene diversification methods include chemical mutagenesis, error-prone polymerase chain reaction (PCR) or different genetic recombination approaches, like DNA shuffling [140]. A limitation of the pure directed evolution approach is that even large libraries with millions of members often sample only a fraction of the available sequence space, e.g. a protein of 100 amino acid length would require the generation of a library with \(20^{100} \approx 10^{130}\) members, which is practically impossible [140, 141]. Therefore more directed, also called semi-rational approaches are commonly employed to generate smaller, high-quality libraries [133]. To achieve this, knowledge of protein structure and computational tools are employed to only mutate the parts of the protein that promise to yield the intended properties [133]. The synthesis of protein genes with degenerate codons and the use of mutagenic primers are commonly employed strategies to yield a more focused diversification of the amino acid sequence [140].

Regardless of whether a completely random or more directed diversification approach is chosen, a large number of mutants have to be screened or selected to identify and isolate variants with desired properties necessitating the use of high-throughput technologies [138]. The choice of an appropriate screening and selection method is highly dependent on the nature of the engineered protein. Generally, screening techniques monitoring substrate or product concentrations
Protein engineering and characterization are the most versatile but suffer from low throughput because they often rely on spatial separation of the different variants on a multi-well plate to maintain the necessary connection of the desired phenotype with the underlying genotype [140, 142]. Techniques that rely on compartmentalization for genotype-phenotype linkage, e.g. by using cells or droplets, and can be coupled to a fluorescent output provide a much higher throughput option thanks to the possibility to use FACS and fluorescence-activated droplet sorting [140]. For the engineering of protein-protein interactions, especially for the design of novel affinity proteins, like antibodies or alternative scaffolds, surface display of the engineered protein is a popular method for simultaneous screening and selection of desired variants [140]. Some commonly used display technologies are discussed in the following section.

2.1.1 Display technologies

Since its first description in 1985 [143] and its first use to display antibody Fv fragments [144], phage display is still frequently used today due to its easy implementation and ability to harbor large libraries [145]. In phage display, the protein to be engineered is expressed on the surface of phage particles, which are viruses that selectively infect bacteria, and then incubated with target protein. By immobilizing bound phages, non-binding phages can be washed away and phages displaying a protein that binds to the target are recovered for another round of enrichment or final evaluation [145]. However, limited control during the selection process and shortcomings in protein folding and post-translational modifications have lead researchers to devise other display platforms [145, 146]. An important group are cell surface display platforms, like yeast [147, 148], E. coli [149], Staphylococcus carnosus (S. carnosus) [150, 151] and mammalian display [152]. As in phage display, all these cell surface display platforms link the protein of interest to a membrane protein to anchor it on the cell surface. The different cell surface display systems come with advantages and limitations inherent to their respective cell type, which have been extensively reviewed elsewhere [146, 153]. One advantage that unites them is the possibility to monitor successful protein surface expression and target binding in real-time by flow cytometry, giving increased control over the selection process. They also have one common limitation; they depend on transformation or transfection, i.e. the process of introducing foreign DNA into a cell, for library generation, which limits the library size that can be practically achieved [140]. To circumvent this “transformation bottleneck”, cell-
free display systems like ribosome [154] and cis display [155] have been developed. They covalently link the protein of interest to its mRNA or DNA, respectively, to establish the link between gene and protein. Besides the possibility to generate large libraries, they also allow for the screening of proteins that are potentially toxic to cells and facilitate repeated PCR-based mutagenesis between selection cycles [156, 157].

One engineering discipline that has greatly benefited from surface display technology is antibody engineering. Different strategies utilized in engineering an antibody are presented in the following paragraph.

### 2.2 Antibody engineering

As outlined in the previous chapters, antibodies have multiple interaction partners: They recognize their antigen via the Fv region and bind to Fc receptors and complement via their Fc portion. Each of these interactions can be engineered to match the intended use of the antibody.

Engineering of the antigen-binding site can be achieved by the directed evolution strategies described above. Early antibody libraries were almost exclusively derived from natural repertoires, i.e. B-cells from an animal are isolated and the antibody genes are cloned into the chosen display system. Depending on whether the animal was previously immunized with the chosen antigen or not, the library is referred to as immune or naïve. Thanks to our increased knowledge of antibody structure and function, libraries can today be created from scratch, so-called synthetic libraries [158]. For this, an antibody framework is chosen and diversity within the antigen binding site is introduced via the above-mentioned methods. Synthetic libraries offer the advantage of selection against antigens that naturally do not elicit an antibody response, e.g. self-antigens [158]. Many display hosts, however, require the use of small antibody fragments, like single chain Fv (scFv) or Fab fragments (Fig. 5), instead of the full IgG molecule for functional expression necessitating re-cloning into a full-length format after successful selection [159]. The created library is then incubated with the selected antigen and potent binders are isolated. This scheme can also be applied to antibodies originally derived from hybridoma technology (see chapter 1.3.1) to increase binding strength to their antigen in a process called affinity maturation [160].

Due to the mouse origin of many hybridoma-derived antibodies, these molecules are often recognized as foreign by the human immune system provoking an immune
Figure 5: Structure of antibody derivatives and an alternative scaffold protein. a) A Fab fragment consists of an antibody’s light chain (red) and the corresponding heavy chain domains, CH1 and VH (blue), here depicted in cartoon form (top) and as ribbon model of the protein structure (bottom). b) In an scFv, the variable regions of both antibody chains, VH (blue) and VL (red), are linked by a protein linker (gray) to form a single protein. As for the Fab fragment, the structure of a scFv is here depicted in both cartoon form (top) and as ribbon model (bottom). c) Structure of the Affibody scaffold. The ribbon model illustrates the characteristic three-helix bundle structure of the Affibody protein.

response that can be harmful to the patient and renders the antibody ineffective. To reduce immunogenicity and increase the antibodies therapeutic potential, another engineering strategy is the humanization of an antibody, which describes the replacement of murine Fc and Fv framework regions with human sequences [160]. But even humanized or fully human antibodies, as they are when derived from surface display or transgenic mice carrying human antibody genes, still have a low risk of eliciting an immune response. This can be reduced by identifying known T-cell epitopes in the antibody’s primary structure with the help of computational tools. In a process known as de-immunization, single amino acids on the antibody are then selectively changed to disrupt these T-cell epitopes and, thus, preclude an immune response [161].

As mentioned at the beginning of this section, the Fc region of an antibody can also be subjected to engineering, which is of interest as it can confer effector functions like ADCC and CDC and contribute to the long in vivo half-life of antibodies via recycling through the FcRn receptor. The generation of a more effective and long-lived antibody, by improved effector functions and recycling properties, holds the promise of lower or less-frequent dosing and thus a reduction in treatment costs.
Both changes at selected positions in the Fc amino acid sequence as well as modifications of the sugar structures on the antibody, known as glycosylation, have been reported to enhance effector functions. However, for some applications effector functions are not desired. Then an antibody, like in the case of eculizumab (see chapter 1.3.1), can be engineered to contain Fc regions of isotypes that are known to have reduced effector functions like IgG2 and IgG4.

Examples of more drastic antibody engineering strategies that confer completely new functions to an antibody are the generation of antibody drug conjugates and bispecific antibodies. By linking an antibody to a cytotoxic drug, this drug can kill cancer cells with greater specificity, and it also obtains a longer half-life in circulation than the unconjugated drug. Bispecific antibodies are capable of binding two targets at the same time, which can impart a therapeutic benefit. Several different bispecific formats have been developed, some of which will be presented in the following section.

### 2.3 Bispecific antibodies

The concept of bispecific antibodies (bsAbs) is not new. As early as 1961, Nisonoff and Rivers combined antibody fragments of two different antibodies to yield bispecific molecules. But only recently have improvements in engineering and production techniques made it feasible to develop bsAbs for clinical use resulting in the approval of the first bsAb, catumaxomab, for cancer immunotherapy in 2009. One additional bsAb has been approved since then and more than 50 are currently in clinical development.

There are many reasons that motivate the design of bsAbs. For example, the understanding that cancer cells can compensate for reduced signaling from a receptor when it is blocked by a monospecific antibody through upregulation of another receptor has led researchers to hypothesize that inhibiting several signaling pathways simultaneously by blocking two receptors or their corresponding ligands should yield better therapeutic effects and prevent drug resistance. Several bsAbs, targeting for example different HER family members or various pro-inflammatory ligands, have confirmed this hypothesis and proved effective for the treatment of cancer and auto-immune diseases. Also the targeting of two epitopes on the same receptor has been shown to yield improved antiproliferative effects. While these effects could in part also be achieved by a combination
of monospecific antibodies [169, 170], bsAbs have both regulatory and economic advantages compared to the development of two monoclonal antibodies to be used in combination [166]. Moreover, bsAb can be designed to recruit effector cells, like T-cells. An example of this is the BiTe molecule mentioned in chapter 1.3.3, that binds simultaneously to CD3-receptor on T-cells and an antigen on the tumor cell surface, and in this way redirects the T-cell to the tumor [165]. Another mode of action for bsAbs is the mimicry of natural ligands like in the case of emicizumab (ACE910). This antibody forces the interaction of factors XI and X in hemophilia A patients which lack factor VIII that normally fulfills this task and who also are intolerant to replacement therapy, and in this way contributes to functional blood coagulation [165, 166, 171].

A plethora of different bsAb formats have been reported. The most obvious approach is to exploit the bivalent nature of an antibody and exchange one of the heavy-light-chain pairs for one arm from a different antibody to generate so-called asymmetric antibodies. However, this approach is likely to generate many mispaired antibodies and thus requires an elaborate purification scheme. Various strategies to prevent the mispairing of heavy and light chains have been developed based on rational design (see Spiess et al. [167] for a comprehensive review). For instance, introducing mutations in the Fc part of the antibody can disfavor HC homodimerization via steric hindrance, like with “knobs-into-holes” technology, or electrostatic repulsion consequently promoting HC-heterodimerization [172]. Methods to ensure the correct pairing of the LC with its corresponding HC include swapping the constant region of the light chain, CL, with its counterpart on the heavy chain, the CH1 domain, known as CrossMab technology. Electrostatic steering mutations have also been used to guarantee correct LC pairing [172, 173]. To avoid the pairing problem altogether, other, symmetric bsAb formats have been devised that rely on the fusion of other binding moieties to a monospecific antibody. These moieties include antibody fragments like scFvs or alternative scaffold molecules like Affibodies (see chapter 2.4) (Fig. 5). This format also allows for the generation of molecules with more than two specificities, which leads to further functional improvement [174, 175]. Additional formats that omit the Fc region are based on various antibody fragment fusion technologies. The above-mentioned BiTe, which consists of two genetically fused scFv molecules, is a representative of this group. This way of creating bispecifics is of course not limited to antibody fragments but can easily be adapted to non-IgG based, alternative scaffolds. What alternative scaffolds are and which advantages and limitations they have is topic of the following section.
2.4 Engineered protein scaffolds for molecular recognition

As the body’s natural way to generate highly specific binders to a variety of different targets, antibodies were the logical choice when seeking antigen-specific therapeutics. However, some characteristics inherent to the antibody molecule are unfavorable for a therapeutic drug. Antibodies are rather big (approx. 150 kDa for an IgG) proteins, consisting of four polypeptide chains and their functionality relies on the correct assembly of these chains via disulfide-bridges and the proper posttranslational glycosylation. These characteristics exclude the successful production in the cytosol of bacteria and necessitate the use of more expensive, eukaryotic production hosts (see chapter 3.1). Moreover, antibodies are comparably unstable and prone to aggregate, thereby restricting their formulation and administration [176, 177]. These drawbacks have spurred the search for alternative protein formats. The advances in protein engineering described above have facilitated the design of not only small antibody fragments, like scFv or Fab (Fig. 5), which overcome some of the limitations of a full-length IgG, but also completely new, non-IgG related antigen-binding proteins. These proteins, collectively referred to as alternative scaffolds, share the characteristic of a stable protein framework that tolerates both amino acid mutations and insertions to generate new binding specificities [178]. Alternative scaffolds are based on very structurally diverse proteins from both human and non-human sources. Generally, the frameworks are classified based on whether binding is mediated by residues in exposed loops similar to the CDRs of an antibody or by residues in more rigid secondary structural elements [179, 180]. Examples of the first group are Adnectins, Anticalins and Kuntiz domains. The second group includes DARPin molecules and Affibody molecules [181]. A comprehensive discussion and comparison of the different scaffolds goes beyond the scope of this thesis and the interested reader is referred elsewhere [181, 182]. Given the appended articles, only the Affibody technology will be described in a bit more detail.

The Affibody scaffold is derived from domain B of *Staphylococcus aureus* protein A, which in its natural form imparts the ability to bind IgG to the bacterium. The small, 6.5 kDa domain adopts the characteristic Affibody scaffold structure of a three-helix bundle (Fig. 5c). Novel binding specificities have been generated by randomizing 13 surface-exposed amino acids, including those that confer IgG binding, on helices one and two and selecting novel binders from phage libraries [183, 184]. To date Affibody molecules against more than 40 different targets have been selected [185]. Favorable properties of the Affibody scaffold like small size,
high thermostability, fast folding and the absence of cysteine residues and glycosylations are also shared by other frameworks. It is these characteristics that distinguish alternative scaffolds from antibodies. The lack of both disulfide bridges and glycosylations enables the easy and cost-efficient production of alternative scaffolds in prokaryotic host organisms like *E. coli*; the small size allows better tumor penetration compared to an IgG [177] and the high stability holds the promise of success in alternative routes of administration [132]. The monomeric nature of these proteins facilitates the fusion of other entities rendering them well-suited for the generation of multivalent or -specific molecules [132]. However, the small size of alternative scaffolds also represents a drawback in terms of therapeutic application because it leads to rapid clearance from the blood. While this might be advantageous for imaging applications, it is generally not desired for a drug as it entails the need for more frequent dosing. Different engineering strategies based on fusion with, for example, an Fc domain or PEG, have been devised to increase the time in circulation. However this solution via the fusion of another entity potentially undoes one of the biggest assets of alternative scaffolds; the small size [177].

Nevertheless, several different scaffold proteins have reached clinical development [177, 180]. But so far only one protein, the Kunitz domain-based molecule ecallas tide (DX-88) has gained approval for the treatment of hereditary angioedema [177, 186]. In general, much remains to be learned about the therapeutic potential of alternative scaffolds, but ongoing clinical trials are likely to shed light on some important aspects like the immunogenicity, correlating to safety and efficacy of alternative scaffolds, where data today is too scarce to draw general conclusions [177].

### 2.5 Epitope mapping

Desirable information that can help on the design process of affinity-based biopharmaceuticals such as antibodies is knowledge about where the protein binds on its target. This binding site on the antigen is termed the epitope. While there potentially are many epitopes on a chosen target and a directed evolution approach is likely to give binders against several of these, it cannot be assumed that all epitopes are equally functional from a therapeutic perspective. Moreover, it is not guaranteed that the most frequent or strongest binders from a selection are also the most effective ones [187, 188]. Therefore epitope mapping, i.e. the
determination of antibodies’ binding sites, is needed to further our understanding of the mechanism of action of an antibody or alternative binding molecule, and for identification of functional epitopes. This knowledge can consequently be used for the rational design of future antibodies and vaccines [189, 190]. Even when pure rational design is not possible, by knowing which epitope to target, the selection process can be adapted accordingly or antibodies against the desired epitope can be identified at an early stage during drug development, thereby streamlining the development process and, thus, reducing costs [187]. In times of rapid therapeutic antibody development with several companies developing mAbs against the same target another important rational for epitope mapping is intellectual property rights. By demonstrating that ones mAb recognizes a different site on the antigen than the competitor’s antibody, epitope mapping plays a central role in patentability [191]. Additionally, epitope information can be applied in pharmacogenomics and precision medicine. By comparing the known epitope of a therapeutic antibody with a patient’s antigen gene, possible mismatches can be identified that are likely to render the patient non-responsive to a drug and, thus, advise the design and use of an alternative drug [191–193].

Several different methods have been developed to help identify the epitope of an antibody on its target antigen. Co-crystallization of antigen and antibody and subsequent analysis of the 3D-structure by X-ray diffraction yields comprehensive information about the amino acids involved in the antigen-antibody interaction and is regarded as the gold standard in epitope mapping [188, 191]. However this method is rather labor-intensive, requires sophisticated equipment and large amounts of purified proteins, provides low-throughput and still may not be successful for all proteins [188, 191]. NMR-based structure determination can be used to identify the antibody-antigen interaction site in solution but struggles partly with the same drawbacks as X-ray crystallography and is limited to rather small molecules [188]. A more high-throughput way of epitope mapping is offered by peptide arrays. Short, overlapping peptides are immobilized on a solid surface, probed with antibody and binding events are registered [194, 195]. In this way linear stretches of antibody-antigen interaction are identified but the method has a limited ability to identify conformational epitopes, i.e. such epitopes that are dependent on correct protein folding [196], and is unable to pinpoint which residues actually make contact with the antibody. Mutational scanning approaches allow for a more functional analysis of epitopes. By mutating every single amino acid conclusions about which residues are involved in binding can be drawn.
Mutational scans have benefited a lot from surface display technologies described in chapter 2.1.1, as they render protein purification unnecessary and allow for simultaneous analysis and selection by, for example, FACS [191]. Alanine scanning, (the mutation of each amino acid to alanine), is a commonly used approach but offers low throughput and the effect of an alanine mutation might be too small to be detected [191]. More comprehensive, random mutagenesis gives more detailed information about single amino acid substitutions but bears the risk of introducing disruptive mutations requiring follow-up experiments [191, 192, 197]. Recently, strategies combining a rationally designed library with surface display and deep sequencing have yielded a comprehensive and efficient analysis of antibody epitopes allowing mapping of several antibodies in parallel which is desirable in early stage drug discovery [191, 198].
Chapter 3

Protein production

The protein engineering tools described in the previous chapter have enabled the fast and efficient generation of novel proteins. This demands equally fast and efficient ways of protein production. While some, non-engineered biologics are still derived from natural sources, most therapeutic, and in particular, engineered proteins are produced recombinantly in heterologous organisms, i.e. the protein gene is introduced into and then expressed in an organism that normally would not produce this protein. Choosing an appropriate production organism, also referred to as host, is one of the first steps in the production process. In 1982, insulin as the first FDA approved, recombinant protein was produced in *E. coli*. Since then various production hosts, including bacteria, yeast, insect cells, transgenic animals, and mammalian and plant cells have been employed commercially and significantly more have been explored in research applications [199]. The most commonly used production platforms are *E. coli*, yeast and mammalian cells. Their advantages and limitations will be described in the first part of this chapter. In recent years, the approval of biotherapeutics produced in mammalian expression systems has steadily increased thanks to technical advances and the demand for more complex protein therapeutics, which has resulted in the production of around 60% of all therapeutic proteins in mammalian cells today [199, 200]. The development of a mammalian production cell line, cell line engineering strategies and advancements in cell culture technology are addressed in the second part of this chapter. The final part of the production process, summarized as downstream processes, as opposed to upstream processes (which describes cell line development and cultivation) comprises of the purification and formulation of a protein. This is a separate
Protein production

Figure 6: Microscopy images of E. coli (a), S. cerevisiae (b) and Chinese hamster ovary (CHO) cells (c). a) The rod-shaped, bacterial cells of E. coli are barely visible when imaged at 400-times magnification in phase contrast. b) Yeast cells can be clearly visualized, when examined in bright-field at 400-times magnification. c) CHO cells are even larger than yeast cells and clearly visible in bright-field at 200-times magnification.

subject and will not be discussed any further in this thesis.

3.1 Choosing a production host

As described above, protein scientists have a wide variety of established and less conventional host systems to choose from for protein production, but the choice is everything but trivial. Besides the implications of production costs, many other aspects like the protein’s complexity, mode of action and delivery method have to be taken into account [201]. A simplified rule is that smaller proteins can efficiently be produced in microbial systems, while more complex proteins with requirements for correct folding and post-translational modifications necessitate the use of mammalian hosts [201]. The following paragraphs illustrate the benefits, limitations, and advances of the three most commonly used expression platforms: E. coli a prokaryotic host, the yeasts Saccharomyces cerevisiae (S. cerevisiae) and Pichia pastoris (P. pastoris) both examples of eukaryotic microorganisms and different mammalian cell lines of both human and non-human origin (Fig. 6).

3.1.1 Protein expression in E. coli

The Gram-negative bacterium E. coli (Fig. 6a) is the oldest production host for therapeutic proteins and it was the most prevalent expression system used during the beginning of the biotherapeutic era in the 1980s [55]. It is still today the single most used, non-mammalian production organism [199]. E. coli exhibits some highly desirable characteristics for a production host: its genetics are well-understood, which combined with readily available genetic tools allows
Choosing a production host

for easy genetic manipulation, and it has the ability to grow to high cell densities in inexpensive growth media [53, 202]. However, the reducing environment in the cytoplasm and the absence of enzymes often required for proper protein folding and glycosylation precludes the production of more complex proteins in E. coli, and even smaller proteins tend to aggregate and form so-called inclusion bodies partly due to insufficient protein secretion [201, 202]. On one hand, inclusion bodies provide a convenient way of protein isolation and proteins can be refolded from the aggregates [203], whereas on the other hand, refolding might not work for every protein and is often accompanied by a decrease in protein yield [201]. To overcome these limitations several engineering approaches have been undertaken. Co- or overexpression of chaperones and addition of signaling peptides promoting transport into the oxidative environment of the periplasm increases the amount of correctly folded protein and, thus, reduces inclusion body formation [53, 201]. Also, engineered E. coli strains with an oxidative cytoplasm have been reported to have improved protein folding capabilities [201]. Besides the folding problem, wild-type E. coli is not capable of protein glycosylation, which is a post-translational modification required to render many proteins, including antibodies fully functional. E. coli strains have been generated that contain the N-glycosylation pathway of another bacterium, Campylobacter jejuni [53, 201, 204]. While this allows some form of glycosylation of the recombinant protein, the glycosylation pattern is still different from that in humans, which might prove problematic when administered to patients. Another drawback of E. coli is the risk of contaminating endotoxins in the end product. Traditionally, removal of endotoxins which can elicit a dangerous immune response in the patient, is achieved by elaborate purification steps. However, recently, endotoxin-free E. coli strains have been developed to overcome this [201].

Despite the different strategies, application of E. coli remains focused on the production of small, non-glycosylated proteins like antibody fragments and insulin. For this task E. coli is well suited as the success of Lantus, an engineered insulin produced in E. coli, underscores, due to its status as the ninth best-selling drug in 2016 worldwide [58].

3.1.2 Protein expression in yeast

As an eukaryotic microorganism, yeast combines the ability of eukaryotic organisms to post-translationally modify proteins with the microbial property of fast growth to high cell densities in inexpensive media [202]. In contrast to E. coli,
yeast is also able to secrete the recombinant protein directly into the growth medium, which facilitates downstream processing [202]. Two yeast strains in particular, *S. cerevisiae* (Fig. 6b) and *P. pastoris* have been employed for the production of therapeutic proteins and rank in popularity right behind *E. coli* accounting for around 17% of all drug approvals [199]. Traditionally, *S. cerevisiae* was the only yeast species sufficiently genetically characterized to be employed for protein expression, explaining its broad commercial use. Analysis of other species, like *P. pastoris*, revealed though, that they are better suited for recombinant protein production because of e.g. *P. pastoris’* ability to grow on different carbon sources and to yield higher production and secretion rates [205].

Approaches to further improve protein production from yeast have mainly focused on increasing protein secretion rather than aiming for higher cell densities [68]. Strategies involving engineering at different levels, like codon optimization, generation of synthetic promoters and metabolic engineering have all proved successful in increasing protein production [201, 206]. Compared to *E. coli*, one of the biggest advantages in using yeast as a production host is their ability to proteolytically process a protein, form disulfide bonds and thus promote the secretion of correctly folded proteins. However, the additional burden of secreting a heterologous protein can overstrain the folding and secretion machinery leading to intracellular accumulation and degradation of the recombinant protein [53, 201]. Overexpression of chaperones as well as other helper proteins involved in vesicular trafficking and the generation of protease-deficient strains have all contributed to improved secretory capacity and productivity [53, 206].

Another advantage that yeast has over *E. coli* is the ability to glycosylate proteins, however, the glycosylation pattern is different to that in humans and is often high in mannose residues. Hypermannosylation of a therapeutic protein has been shown to result in faster degradation and, thus, reduced efficacy [206]. Therefore, glycoengineering (the process of controlling protein glycosylation) of yeast is an active field of research. Many studies have been performed with a slightly stronger focus on *P. pastoris* resulting in a strain capable of humanized N-glycosylation [206–208]. However, despite the availability of commercial technologies, heterogeneity in the glycosylation profile and reduction in protein yield have hampered the commercial use of glycoengineered yeast and as of 2013, all approved, therapeutic proteins produced in *S. cerevisiae* were non-glycosylated [201, 202].
3.1.3 Protein expression in mammalian cells

Mammalian cells are the production host of choice for more complex therapeutic proteins, like antibodies, because of their ability to properly fold, assemble and glycosylate proteins [200]. Since the approval of human tissue plasminogen activator produced in Chinese hamster ovary (CHO) cells (Fig. 6c) in 1986, which marked the first manufacture of a therapeutic protein from recombinant mammalian cells, several other host cell lines have been utilized and the use of mammalian as opposed to non-mammalian expression systems has steadily increased [199]. Cell lines approved for biotherapeutic production (besides CHO cells) include mouse myeloma derived NS0 and Sp2/0 cells, baby hamster kidney (BHK) cells, human embryo kidney (HEK-293) cells and human retina-derived PER-C6 cells [68, 200]. Human cell lines provide the advantage of fully human glycosylation profiles but the risk of human viral contamination and limited experience with them constrains their use [209]. Rodent-derived cell lines show human-like glycosylation, which in most cases are not immunogenic. However mouse cell lines have been reported to potentially produce N-glycans that elicit an immune response necessitating the selection of clones with agreeable glycosylation profiles [68, 209, 210]. By far, the most prevalent expression system are CHO cells owing to their rather easy adaptation to new culture conditions, their safety, their high productivity and not least due to the big body of knowledge and expertise regarding them, which has accumulated over the years [199, 211–213].

Drawbacks often mentioned in the context of mammalian protein production include low protein yields, requirements for complex medium and a lengthy development and production process due to extensive screening and low growth rates, which combined render mammalian protein production rather expensive [53, 214]. However, an improved understanding of mammalian gene expression, metabolism and growth has led to recent advancements in cell line development and engineering as well as in media formulations and cultivation technology, which have resulted in increased productivities, making mammalian protein production more competitive. This is illustrated by a direct comparison of antibody fragment production in \textit{P. pastoris} and CHO cells, where CHO cells reached a higher space time yield despite a higher cell mass in yeast [215, 216]. The improvements in cell line development and cultivation that have contributed to increased productivity are described in more detail in the following sections.
3.2 Cell line development and engineering

The aim of a cell line development process is the isolation of a cell clone that produces high amounts of the recombinant protein and is suited for the manufacturing process. It starts with the introduction of the protein-encoding DNA into the mammalian cell in a process known as transfection. Subsequently, transfected cells are selected, screened and clones showing desired characteristics are isolated for scale-up and further analysis resulting in the final choice of a clone for production which meets the specified criteria (Fig. 7).

Many different transfection methods exist that can generally be grouped into viral, chemical or physical gene delivery. The most suitable approach depends on cell type and purpose and often has to be empirically determined [217]. Once the DNA has entered the cell nucleus it either remains extrachromosomal, known as transient transfection, or is incorporated into the chromosomal DNA, which is referred to as stable transfection. After transient transfection the recombinant protein is only expressed for a limited time due to the eventual loss of the transfected DNA as a result of cell division and environmental factors [217]. Nevertheless, transient expression offers a possibility to quickly produce small amounts of recombinant proteins in mammalian cells as it omits the time-consuming development of a clonal cell line. It is therefore frequently used for drug screening and
in pre-clinical development [218, 219].

For large-scale protein production, transient gene expression is less well suited because of the requirement for large amounts of cells, DNA and where applicable transfection reagents [220]. Moreover, documentation of a clonally derived cell line is required during the approval process of protein therapeutics [221]. The selection and expansion of single clones requires the stable integration of the transgene into the genome. To promote this, the protein gene of interest is transfected together with a selection marker and a selective pressure is applied. Selection markers include classical antibiotic-resistance genes and the commonly used dihydrofolate reductase (DHFR) and glutamine synthetase (GS) system [220]. In the latter cases, selection relies on the use of DHFR- or GS-deficient cell lines and the co-transfection of the gene of interest with the DHFR- or GS gene, respectively. When cultured in hypoxantine- and thymidine-free (DHFR system) or glutamine-free (GS system) medium, only transfected cells are able to survive. Additional selective pressure is applied through addition of DHFR or GS inhibitors at increasing concentrations, which favors the survival of cells with a high transgene copy number which often show higher productivities [108]. The downside with these approaches is the random integration of the transgene into the genome generating a heterogeneous pool of production cells with varying stability, which necessitates the extensive and time-consuming screening and selection of isolated clones.

Several different methods for the identification and selection of high-producing clones have been devised. Traditionally, cells are isolated by limiting dilution in multi-well plates and clones exhibiting high product titers are chosen for expansion and further analysis. Besides the risk of non-clonal cell lines, this method is time-consuming and laborious and is restricted by a limited throughput even when automated with robotics [200, 221]. A higher throughput is offered by FACS-based methods but they can only measure the secreted product when it is captured or stuck on the cell surface or via a co-expressed surrogate protein that is fluorescent or can be detected on the cell surface [221–224]. Cell growth in semi-solid medium combined with automated clone-picking [222, 225] and the use of droplet microfluidics [226–228] are other strategies to spatially separate single cells while retaining the ability to monitor secretion of the protein of interest (see [222] for a comprehensive review). One such selection step does not, however, ensure the long-term expression of the transgene, therefore sustained screening is necessary to identify stably expressing clones. This instability is due to gene silencing mediated by epigenetic events like histone hypoacetylation and increased promoter
methylations [200, 206]. The realization that the site of integration greatly affects the transgene transcription, referred to as position effect, has spurred the development of technologies that circumvent negative position effects inherent to random integration. Addition of the histone deacetylation inhibitor butyrate or flanking the gene of interest with protective elements like scaffold/matrix attachment regions (S/MARs) or ubiquitous chromatin opening elements (UCOE\textsuperscript{TM}) have shown a positive reduction in the occurrence of gene silencing [108, 206, 229].

An even more convenient method that would render tedious screening obsolete, would be the targeted integration of the transgene into sites that are known to promote stable and high transcription rates. Different methods for targeted gene integration exist [230] and especially with the advent of CRISPR/Cas9, site-specific integration has become rather straightforward [231, 232]. While this procedure holds much promise for the rapid development of stable and high-producing cell lines, more knowledge about suitable integration sites is needed for targeted integration to become a standard procedure [232].

Other strategies that have led to improved transcription, aside from focus on the integration site, are those connected to improved vector design. The use of stronger promoters, codon optimization of the transgene, the addition of intron sequences and optimization of signal peptides have contributed to higher transcription rates [200, 233, 234]. Yet another approach to increase protein production is the direct engineering of the host cell line. Aside from the generation of DHFR- and GS-deficient cell lines as described above, mammalian cells have been engineered to be more apoptosis resistant, have higher proliferation and to be better at protein secretion, to name just a few improvements [213, 235]. The deciphering of the CHO genome [236, 237] and the increasing application of “omics” approaches, i.e. results from transcriptomics, proteomics and metabolomics analysis are expected to facilitate the rational engineering of superior production cell lines [213, 238, 239].

### 3.3 Advances in cell culture technology

Aside from the advances in host cell engineering, vector design and selection technology described in the previous section, improvements in medium composition and cultivation technology have made a major contribution to the high productivities achievable with mammalian cells today. As illustrated by Jayapal et al. [212], a standard mammalian production process in 1987 was run in batch-mode
for one week and yielded approximately 100 mg/L recombinant protein and a maximum cell density of $3 \times 10^6 \text{cells/mL}$. In 2007, titers had increased to 1-5 g/L of recombinant protein thanks to the achievement of higher cell densities in ten to twelve-day long fed-batch processes [212]. Nowadays, final product concentrations above 10 g/L can be achieved and perfusion cultures reach cell densities beyond $2 \times 10^8 \text{cells/mL}$ [68, 240, 241].

This progress illustrated above is a result of improvements in different aspects of mammalian cell cultivation rendering it a dependable and robust technology [68]. Maybe, the most striking development is the trend in the cultivation process towards higher cell densities and longer cultivation time. Generally, the use of suspension cultures as opposed to adherent cell culture is preferred as it allows higher cell densities [200]. As described above, in the beginning of the mammalian cell culture era, batch cultures were the most prevalent process mode, where rich medium is inoculated with cells that are allowed to grow without the addition or removal of any components until harvest, which occurs after approximately one week [242]. Nowadays, many high-yielding process are based on fed-batch cultivation, where a batch culture is regularly diluted, i.e. fed, with small volumes of concentrated feed medium [242]. Fed-batch processes can be maintained for several weeks and the time point of harvest is often determined by plant-capacity, productivity and product quality [200]. A change in paradigm aiming for consistent protein quality rather than high product titers has propelled the use of continuous perfusion cultures. Here the culture is constantly fed with feed medium and simultaneously, the spent medium is removed while the cells are recycled back into the bioreactor allowing the cultivation to go on for months. This generates a more uniform environment for the cells and ensures a greater consistency in product quality [68, 242]. Moreover, perfusion technology omits the need for big bioreactors. By feeding and harvesting several reactor volumes a day for a long period of time, and due to the high achievable cell densities, space-time yields can be obtained from a small bioreactor volume higher than those attainable by a fed-batch process [68, 206, 242].

This development in process technology was made possible by improvements in other areas. One major achievement was the establishment of serum-free, chemically defined medium, which increased process robustness, and the development and constant optimization of concentrated feed-medium, which first enabled fed-batch and continuous cultivation [68]. Growth and feed media have to be customized for every production process and various optimization strategies exist.
Moreover, improved cultivation monitoring allows for better control of the process and adaptation of feed-strategies [68, 242, 244]. Also, the host cell engineering achievements described above have contributed to more stable and robust cell lines that tolerate longer cultivation [68]. Despite the indisputable benefits for product yield and quality, the long cultivation times increase the risk for contamination [68, 218].

Another trend in cultivation technology is the increasing use of disposable technologies like bioreactors. They offer the advantage of lower initial investment, greater flexibility and good reproducibility characteristics. Since they arrive clean and sterile, they present a reliable resource with no need for validation [68, 243].

Taken together, the described improvements in host cells, vector design, selection process and cultivation technology have all contributed to higher productivities in mammalian cell culture and technological advances indicate there is still room for more improvement.
Present investigation

The biopharmaceutical industry, namely the generation of protein drugs with biotechnological methods, is constantly growing [58, 199]. Today, eight of the ten top-selling drugs are biopharmaceuticals with monoclonal antibodies accounting for five of these [58]. As outlined in the previous chapters, the development of such a recombinant antibody is a time-consuming and cost-intensive endeavor involving several steps. Starting with the identification of a potential target protein, functional antibodies have to be identified from an antibody library, a process that often employs display technology. Once a candidate antibody has been identified, its binding properties have to be characterized and at the same time, a production host has to be decided on. The correct, functional glycosylation of an antibody requires the production in mammalian cells and in this instance a single, high-producing clone has to be isolated to meet the regulatory requirements for monoclonality. This often involves extensive amounts of screening to identify a suitable clone, which alongside high productivity, also shows stable growth characteristics when scaled up to larger culture volumes.

The articles presented in this thesis highlight different stages of this biopharmaceutical production process. The first article examines the combination of different display technologies for the isolation of high affinity antibodies. In the second paper, the rational generation of a bispecific antibody format engaging two targets at the same time is described in the second paper. Antibody characterization is the focus of paper three and a method for identifying the epitope of an antibody is presented as well as how this knowledge can be used in a precision medicine approach to guide treatment. The fourth paper leaves the protein engineering part behind and centers on protein production and particularly cell line development, with a new method for the high-throughput screening and isolation of high-producing clones combining a split-GFP probe with droplet microfluidics being described. An improved probe could further speed up the selection process.
and the development of such a probe is covered in paper five.

**Paper I: Combining phage and staphylococcal display for improved antibody discovery**

Antibodies are powerful pharmaceutical proteins due to their ability to specifically bind and potentially inhibit a protein and to, additionally, mediate effector functions. The generation of novel pharmaceutical antibodies often involves the selection of candidates from an antibody library. To provide a direct link between the desired phenotype and its underlying genotype, libraries based on surface display of the antibody are frequently used. The most prevalent system is phage display [145]. While expressing an antibody, represented by its smaller scFv, on the surface of phage allows for the generation of large libraries, the selection via binding to immobilized target protein, in a process referred to as panning only offers limited control. Moreover, following selection all candidates have to be expressed in *E. coli* and are commonly analyzed by tedious ELISA screening [245, 246]. Conversely, antibody display on the Gram-positive bacterium *S. carnosus* enables the use of FACS for selection and screening allowing better control over the selection process [245] and omitting the need for ELISA screening. However, the libraries that can be generated in *S. carnosus* are generally smaller than phage libraries. In order to harness the advantages of both systems, we combined phage and staphylococcal display for the selection of scFvs against HER2 in this project, which is the first report of Gram-positive display of a human antibody library.

As illustrated in figure 8a, scFvs towards HER2 were selected in a pre-enrichment step from a previously described, phage displayed scFv library [247]. Subsequently, the enriched library was cloned into *S. carnosus*, which due to its size and stable cell wall permits the use of FACS. For FACS sorting the staphylococcal library was stained with fluorescently labeled HER2 and human serum albumin (HSA). The latter binds to the albumin binding domain and is used as a surface expression normalization tag (Fig. 8b), and thereby allows for a better discrimination of binders [245]. Clones exhibiting high target binding (HER2 signal) and surface expression (HSA signal) were sorted by FACS and propagated by cultivation in two consecutive rounds of selection. A clear enrichment of the desired population within the sorting gate could be observed following two rounds of flow cytometric sorting (Fig. 9a).

After two rounds of selection with soluble HER2 ECD by FACS, 96 clones were
isolated and the three best performing clones, named SC1, SC39 and SC66, were chosen for more detailed analysis. First, the three scFvs in their displayed form on *S. carnosus* were ranked depending on their ability to bind to soluble HER2 ECD in flow cytometry. The results demonstrated that SC1 binds the strongest to HER2 followed by SC66 and lastly SC39 (Fig. 9b). To gain an additional measure of binding strength, affinity of the three scFvs was measured by biolayer interferometry (BLI). To this end, all three scFvs were produced and purified from the periplasm of *E. coli* and their binding to immobilized HER2 ECD was monitored. The affinity values calculated from the BLI measurement confirmed the ranking determined by flow cytometry. Interestingly, none of the three binders were identified in a first ELISA screen of the phage library comprising of 47 clones. Only when 300 clones were analyzed by ELISA, were SC1 and SC66 identified. SC39 was not detected in this screen, which was likely to be a result of a low abundance of the clone.

Since full-length antibodies with a functional Fc region are often preferred for therapeutic applications, the three isolated scFvs were subsequently reformatted into full-length human antibodies. To ascertain that the reformatting did not eliminate HER2 binding and to ensure that the molecules are also capable
Present investigation

Figure 9: Flow cytometric enrichment of the staphylococcal library and analysis of isolated clones. a) The enriched phage library was cloned into *S. carnosus* and further enriched for HER2 binders by two rounds of flow cytometric sorting. A clear enrichment of HER2-binding clones within the indicated sorting gate is visible after two rounds of sorting (right diagram) compared to the pre-enriched library cloned from phage (left diagram). b) Following the two rounds of FACS, three clones were isolated and characterized in more detail. Clone SC1 shows the strongest binding to HER2 followed by SC39 and SC66. A Herceptin (Trastuzumab) scFv and a CTLA-4-binding scFv were used as positive and negative control, respectively.

of recognizing their target in a natural context on cells, full-length antibodies were produced in HEK-293 cells. Thereupon, HER2-overexpressing SKOV3 ovarian cancer cells were incubated with antibody-containing HEK-293 supernatants, which were adjusted to have comparable antibody concentrations based on an ELISA measurement. The degree of antibody binding to SKOV3 cells was assessed by flow cytometry and confirmed the binding ranking earlier established for the scFv format.

As staphylococcal display has so far been used for the affinity maturation of Affibody molecules [248–250], we aimed to verify its applicability for the maturation of scFvs. The best performing molecule SC1 was randomly mutated by error-prone PCR, expressed on *S. carnosus* and improved variants were selected by flow cytometric sorting. Two of these clones, SC1M2 and SC1M4, underwent a more detailed analysis. A comparison by BLI with the original SC1 molecule both in the staphylococcal anchored form and as soluble molecules, showed a modest improvement in affinity with a reduction in $K_D$ by factor four to five, which can be attributed to the location of the mutations being outside the CDRs. We anticipate, that more directed mutation strategies could give rise to higher affinity improvements.

In summary, we have devised a new system for antibody selection via combination of phage and staphylococcal display that was successfully applied for the fast and efficient isolation of high-affinity scFv molecules against Her2 also highlighting the suitability of *S. carnosus* for display of human scFv libraries. Recloning and dis-
play on *S. carnosus* following a pre-enrichment in phage allowed for a good differentiation and selection of binders, omitting the need for tedious ELISA screening, and also enabled the identification of binders that potentially would not have been detected by phage display and subsequent ELISA screening. Also, staphylococcal display offers a suitable platform for affinity maturation, eliminating the need for another recloning step. The good expression of the selected binders in *E. coli* as well as HEK-293 cells suggest that the selection of binders through a series of orthogonal selection systems is likely to favor robust proteins with better general folding behavior, which is a desirable trait for protein production.

To demonstrate the general applicability of the described platform, selections against several other targets would be an interesting continuation of the present study. Since this study was aimed at establishing a selection platform, no analysis of the cell growth inhibiting properties of the identified binders was performed. Considering that the molecules exist in both scFv and full-length antibody format, such a proliferation experiment could be worthwhile for assessing the therapeutic potential of these molecules.

**Paper II: Generation and mechanism of action of bispecific AffiMab molecules**

As described in chapter 1.3.1, trastuzumab is an established therapeutic antibody, which since its first approval in 1998 has been widely used for the treatment of HER2-positive breast and gastric cancer [71, 79]. But despite trastuzumab’s success, which was documented in several clinical trials to increase disease-free and overall survival [71, 251], in many patients the time until disease progression is limited due to the development or pre-existence of trastuzumab resistance [71, 252]. Several different mechanisms of resistance have been suggested [81, 251, 252], including compensatory signaling via other receptors of the ErbB family [81, 82, 252]. A number of studies indicate that upregulation of EGFR signaling could be a cause for acquired trastuzumab resistance [253–256]. Consequently, strategies were developed that target both receptors simultaneously. Synergistic anti-proliferative effects were reported for various tumor types and trastuzumab-resistant cells when combining trastuzumab with the EGFR-binding monoclonal antibody cetuximab [169, 254, 257, 258]. However, a combination treatment has the disadvantage of increased costs due to parallel development, production and approval of two molecules [166]. This problem can be bypassed by bispecific
binders, which combine two binding entities in one molecule. Therefore, we set out to develop bispecific antibodies targeting both Her2 and EGFR.

To this end, an EGFR-binding Affibody molecule, zEGFR, was genetically fused to the C-terminus of trastuzumab’s heavy or light chain (Fig. 10a). This was achieved by employing a solid-phase cloning technique previously developed in our lab [259], which allows the fusion of genetic elements without the need for restriction enzymes. The resulting proteins, termed AffiMab molecules, maintained the binding specificities of both entities and recognized their targets both in soluble form, when binding was measured by surface plasmon resonance (SPR) (Fig. 10b), and when presented on the surface of cancer cell lines. Also, for the C-terminal fusion protein, preservation of FcRn-binding was confirmed by binding to FcRn-overexpressing cells.

Having established the proper function of the AffiMab molecules, their cytotoxic effect was analyzed by treating NCI-N87 gastric carcinoma cells with varying concentration of these bispecifics or control proteins. Although zEGFR did not show any anti-proliferative effect on its own and did not enhance the effect of trastuzumab when applied in combination, the genetic fusion of zEGFR to trastuzumab conferred an increased cytotoxic effect compared to the combination treatment with both entities, irrespective of the site of fusion (Fig. 11). Aiming to elucidate the molecular mechanism behind this increased cytotoxic effect of the AffiMab molecules, we subsequently analyzed the surface levels of EGFR and HER2, the phosphorylation and transcriptional state of selected receptors and downstream signaling molecules as well as the number of apoptotic cells following treatment.
Figure 11: AffiMab molecules are potent inhibitors of NCI-N87 cell proliferation. NCI-N87 cells were incubated with different concentrations of AffiMab and control molecules. Proliferation was measured after four days. While the IgG control and zEGFR did not affect the number of viable cells, addition of trastuzumab with or without zEGFR had a modest cytotoxic effect. Both AffiMabs exhibited an significantly stronger cytotoxicity indicating a synergistic effect of simultaneous HER2- and EGFR-binding (unpaired t-test: * p<0.01). The average and standard deviation of triplicates are shown. Cell proliferation after addition of IgG control was set to 100%.

To assess the surface levels of HER2 and EGFR on NCI-N87 cells after treatment with AffiMab molecules or controls, cells were stained with affinity molecules towards HER2 and EGFR, respectively, and their binding was measured by FACS. Analysis of HER2 expression with a HER2-binding Affibody dimer, which does not compete with trastuzumab for binding, showed that both AffiMab molecules reduced HER2 surface levels to the same degree as trastuzumab or the combination of trastuzumab and zEGFR (Fig. 12a). On the other hand, EGFR surface levels, measured with a second EGFR-binding mAb, were lower after treatment with AffiMab molecules than after treatment with any of the controls (Fig. 12b). Moreover, the apparent reduction in EGFR levels following exposure to zEGFR or zEGFR combined with trastuzumab could be attributed to a certain degree of competition between the EGFR-binding antibody used for detection and zEGFR. However, the reduction in EGFR levels following AffiMab treatment was distinct from that which could be achieved by competition alone. Other studies have identified receptor clustering, increased internalization and impaired recycling as the reason for reduced receptor surface levels [168]. Nonetheless, when analyzing NCI-N87 cells post treatment by confocal microscopy, neither clustering nor increased intracellular receptor levels could be detected.

Reduced EGFR or HER2 expression due to decreased transcript levels could be another factor explaining the observed reduction in receptor surface levels. However no significant differences in transcript numbers of HER2 and EGFR, nor of any other receptors or downstream signaling molecules, which were selected due
Present investigation

**Figure 12:** AffiMab molecules reduce surface expression levels of HER2 and EGFR.
a) Compared to the IgG control (set to 100%), HER2 surface levels in AffiMab-treated samples were reduced equally when compared with samples treated with trastuzumab or a combination of trastuzumab and zEGFR. Mean and standard deviation of three independent experiments are shown. b) EGFR surface levels are significantly reduced after treatment with AffiMabs compared to trastuzumab and the combination of trastuzumab and zEGFR (paired t-test; *** p<0.0001; ** p<0.01). A certain degree of binding competition to EGFR between the detection mAb and zEGFR accounts for the reduced levels measured for zEGFR and zEGFR with trastuzumab. Mean and standard deviation of six independent experiments are shown. HER2 surface levels in IgG control samples were set to 100%.

to their reported link to HER signaling and trastuzumab resistance, could be detected. Consequently, potential differences in expression that could motivate the observed reduction in receptor surface levels and the larger cytotoxic effect are likely caused by other mechanisms such as reduced translation or increased protein degradation. The latter was earlier reported to be one of trastuzumab’s modes of action [260]. An analysis by Western blot could not identify differences in expression and phosphorylation level of the four ErbB receptors or the central downstream regulators AKT and MAPK.

Another clue to elucidating the mechanism of action of the AffiMab molecules was offered by the analysis of apoptosis levels in NCI-N87 cells. Measuring the percentage of apoptotic cells by flow cytometry after staining with propidium iodide and Annexin V, revealed a slight increase in apoptotic cells following treatment with AffiMabs. To our surprise, an increase was also observed in cells treated with the combination of trastuzumab and zEGFR, which suggests that simultaneous targeting of HER2 and EGFR, either by a bispecific binder or by a combination of binders leads to induction of apoptosis.

In summary, this work describes the successful generation of bispecific AffiMab
molecules, which are superior to their monomeric counterparts in inhibiting cancer cell growth. Their ability to inhibit cell proliferation to a stronger degree than the combination of trastuzumab and zEGFR, highlights their therapeutic potential even though the underlying mechanism of action could not be fully clarified. A potential explanation for the observed reduction in HER2 and EGFR levels and the cytotoxicity is that it results from the combined effect of several mechanisms, which on their own give rise to such small changes that they are not detectable by the chosen methods.

Taken together, the results motivate the further evaluation and optimization of the AffiMab format with regards to other targets, linker length and Affibody orientation. Also the generation of multivalent or multispecific AffiMab molecules could be worthwhile. The construction of such molecules is simplified by the modularity inherent to the solid phase cloning technique.

Paper III: Epitope mapping of eculizumab enables precision medicine

A worthwhile step during the antibody discovery process is the identification of the epitope of a novel antibody on its target antigen. This can help to understand the mechanism of action of an antibody and enables the improvement of the given antibody or the rational design of novel antibodies [261, 262]. Moreover epitope information fulfills an important function in intellectual property rights to distinguish a molecule from those of competitors and, thus, prove patentability [191]. The following work demonstrates another application of epitope information, namely in precision medicine.

The therapeutic monoclonal antibody eculizumab (Soliris) is approved for the treatment of the rare diseases PNH and aHUS, which when untreated can lead to fatal thrombosis [42, 44, 45]. Although both conditions have a different underlying molecular cause (see chapter 1.2.2), it was found that in both cases the complement system plays a central role in the development of the clinical picture [38, 40, 47]. Consequently, the inhibition of C5 cleavage by C5-targeting eculizumab proved successful in the treatment of PNH and aHUS [42, 46]. By hindering C5 cleavage, eculizumab prevents MAC formation and protects patients from hemolysis [39, 46]. However, a report about patients still showing hemolysis whilst receiving eculizumab, and the identification of a point mutation in these patients, have raised questions about the epitope of eculizumab [263]. Earlier, eculizumab’s
epitope had been reported to consist of up to three linear amino acid stretches, where none of these stretches comprises the reported point mutation at amino acid 885 [46, 264]. A more recent study found, based on transmission electron microscopy, that eculizumab interacts with the MG7 domain of C5 with R885 and one of the earlier reported peptides being part of the binding site, but this study could not map eculizumab’s epitope either [265]. This highlighted the need for a more accurate determination of eculizumab’s binding site on C5.

We employed an epitope mapping strategy based on deep mutational scanning of surface expressed protein domains. To this end, the twelve domains of C5 were cloned and expressed on the surface of *S. carnosus*. By measuring eculizumab binding to the different domains via flow cytometry, the MG7 domain was confirmed as the binding domain of eculizumab. Using error prone PCR, the MG7 domain was randomly mutated. In two rounds of flow cytometric sorting clones showing a high surface expression of the MG7 construct but a lack of eculizumab binding were isolated. From the 90 clones sequenced after sorting, only those coding for a single amino acid mutation were considered for analysis with the exclusion of mutations to or from cysteine, glycine or proline as these are thought to disturb the protein structure [197]. The only exception was the clone coding for R885C, which corresponded to the mutation reported in PNH patients [263] and therefore was included in further analysis. This left a set of eleven different clones encoding mutations at eight different residues, which all showed reduced eculizumab binding.

To confirm these results, nine mutants, covering the reported position R885 twice with a mutation to both cysteine and histidine, were produced as full-length pro-
teins in CHO cells. Binding analysis to eculizumab by ELISA revealed that seven mutants showed reduced binding to eculizumab, even though some of these to a lower degree as was earlier measured on the surface of *S. carnosus*. Two mutants, I829K and V896E, were subsequently excluded from further analysis, as their contribution to eculizumab binding could not be confirmed. This leaves six residues on C5 that seemingly are important for interaction with eculizumab. Strikingly, these six residues are not contiguous but instead scatter over the primary sequence of the MG7 domain (Fig. 13).

To ensure the functionality of the produced mutants and thereby exclude misfolding as a cause for the lack of binding, a hemolysis assay was performed showing that all seven mutants can induce MAC formation in a concentration dependent manner. When eculizumab was added to this assay the hemolytic activity of WT C5 was completely suppressed, while most mutants showed almost unchanged activity (Fig. 14). Eculizumab also exhibited an effect on mutants F855I, K887N and F918S, but compared to WT C5 a higher amount of eculizumab is needed to achieve full suppression. Remarkably, addition of the complement inhibitor OmCI, a protein derived from the tick *Ornithodorus moubata* [266], which does not compete with eculizumab for binding to C5 [265], successfully inhibited hemolysis in all mutants (Fig. 14).

Summarizing the above results, six residues can be considered to be part of eculizumabs epitope on C5 with varying impact on the antibody binding.
Present investigation

Figure 15: The six identified residues (in violet) form a conformational epitope on C5’s MG7 domain (blue) and co-locate with the contact regions identified by Schatz-Jakobsen et al. (pink). The protein structure is based on PDB5I5K.

izing these six non-adjacent positions on the reported crystal structure of C5 [39] illustrates their close proximity in the folded protein forming a conformational epitope on C5’s MG7 domain. While this work was under review, a co-crystal structure of C5 and an eculizumab Fab fragment was published [39]. In that work Schatz-Jakobsen et al. identified three contact regions (residues 851–858, 882–888 and 915–920) based on a less than 5 Å-distance between C5 and the Fab fragment. However, the resolution of the crystal structure did not allow the unambiguous identification of individual residues involved in antibody binding. All six residues that we found to interact with eculizumab co-localize with the contact regions identified by Schatz-Jakobsen et al. (Fig. 15) confirming the findings. Moreover, the epitope information helps to explain the earlier reported lack of cross-reactivity of eculizumab to C5 of other species than human [42]. Human C5 exhibits a tryptophan at position 917 whereas all other tested species hold a serine residue at that position. Analysis of the W917S mutant confirmed the inability of eculizumab to bind and inhibit hemolysis for this mutant.

Having successfully identified the earlier reported mutation at position R885 as part of eculizumab’s epitope, we aimed to analyze if other known germline mutations might fall into the binding site. While no additional mutations have been reported for the six identified residues, several genetic variations were identified within the 5 Å regions described above and, thus, in close proximity to the epitope residues. Such a mutation could potentially affect the binding and, consequently,
the efficacy of eculizumab. This highlights the important function that epitope information can fulfill in precision medicine. By analyzing a patient’s C5 gene for mutations within the eculizumab binding region, non- or poor responders can be identified before the start of treatment and consequently be prescribed an alternative therapy. In case of eculizumab non-responsiveness, OmCI, or its recombinant version Coversin, which has proved successful in a phase I clinical trial, presents a promising alternative treatment [267]. Such epitope-based screening would both save distress for the patient and unnecessary expenses for health-care systems.

Taken together, the presented epitope mapping approach based on the expression of independently folded protein domains on the surface of *S. carnosus* has proven a successful platform for the determination of eculizumab’s conformational epitope on C5. The results obtained from staphylococcal display were shown to be transferable and relevant for full-length C5. Therefore, we expect this strategy to be applicable to other protein pairs. We consider the described workflow of detailed epitope mapping and genetic profiling to be transferable to several other diseases and therapies, which could help to advance the implementation of precision medicine.

**Paper IV: Combining split-GFP and droplet microfluidics for improved cell line development**

Once an interesting, new pharmaceutical protein has been designed, irrespective of whether it is an antibody or recombinant protein, the following step is the bioproduction of the protein. The fourth paper addresses particularly the cell line development step within the production process. As mentioned in chapter 3.2, the establishment of a high-producing, mammalian cell line is a tedious endeavor and often regarded as a bottleneck in the production process [200, 268]. With more biopharmaceuticals being approved, both the competition in the pharmaceutical sector and the costs for healthcare systems increase, highlighting the requirement for a reduction in manufacturing cost and improvements in process efficiency to reduce the time to market.

Since current transfection methods often rely on random integration of the transgene, generating heterogeneous pools of cells with differences in expression level, a sub-cloning step is required to meet the approval requirements for a clonal-derived production cell line [221, 222]. Various strategies have been devised to render cell line selection more efficient. Commonly employed limiting dilution cloning (LDC),
Present investigation

Figure 16: Successful droplet sorting of a mock library containing nine-times excess of cells secreting Herceptin without S11-tag over cells secreting tagged Herceptin was confirmed by the absence of the 150 bp band after PCR amplification, which originates from the plasmid encoding HerceptinNotag before sorting.

Based on the spatial separation of single cells in multi-well plates and subsequent analysis of the supernatants, suffers from high labor input and low throughput even though automation has eased the workload and has allowed for the analysis of more clones. In industrial settings limiting dilution has been replaced by higher throughput methods that are typically based on flow cytometry [222]. However, most of these methods rely on reporter proteins or the measurement of the protein of interest on the cell surface and, thus, only give an indirect assessment of the secretion of the protein of interest.

Paper IV describes a method combining droplet microfluidics and split-GFP complementation for the general measurement of protein secretion at the single cell level, and the selection of highly secreting cells. Droplet microfluidics allows a thorough yet high-throughput analysis of single cells, while split-GFP technology offers a generic way of measuring protein secretion. To this end, the protein of interest is genetically labeled with the 11th beta-sheet of GFP (GFP 11 or S11) and the remainder of the GFP barrel is provided in solution inside the droplet. Upon protein secretion the two GFP parts combine and form a fluorescent construct. In this way, fluorescence intensity is linked to the amount of secreted protein allowing sorting of high-producing cells based on fluorescence. To show the applicability of this method, it was used for the selection of both antibody- and erythropoietin (EPO)-secreting CHO cells.

First, Herceptin (trastuzumab) secreting CHO cells were encapsulated as single cells in droplets and antibody secretion was monitored over time. After having
established that cells secreting S11-tagged Herceptin can be easily distinguished from cells secreting non-tagged antibody in droplets, a mock library sorting was performed. Herceptin$_{S11}$-secreting cells were sorted from a mixture of cells with nine times excess of Herceptin$_{nottag}$-secreting cells. Analysis of the sorted cell pool by PCR confirmed successful sorting through the presence of only one DNA-band corresponding to the Herceptin$_{S11}$ plasmid (Fig. 16).

The next logical step was to demonstrate sorting from a heterogeneous pool of stable integrants secreting a biologic of interest. To this end, CHO cells were transfected with a plasmid encoding for EPO-S11 and selected for stable genomic integration via antibiotic resistance. A first analysis of the generated cell pool was done by randomly picking 160 clones and measuring their EPO secretion. This confirmed the generation of a highly heterogeneous cell pool in terms of EPO secretion that we regarded as well suited for testing the performance of our microfluidic droplet-sorting platform. Also, a high (H)-, a medium (M)- and a low (L)-producing clone were isolated at this stage to be used as future reference clones.

To selectively isolate highly EPO-secreting cells, we generated single-cell droplets from the polyclonal EPO-S11 pool. After a four-hour incubation at $37^\circ$C, droplets were re-injected into a sorting chip and droplets with highest fluorescence intensity, corresponding to highest EPO secretion, were isolated. Fluorescence measurement of the sorted droplets showed a clear increase in mean fluorescence intensity com-

---

**Figure 17:** Enrichment of high secreting cells from an EPO-secreting, polyclonal cell pool by droplet sorting. Single-cells from the EPO-S11 producing polyclonal cell line were encapsulated in droplets. After four hours of incubation, droplets with high fluorescence intensity were sorted (the sorting threshold is indicated as the red dotted line). Subsequently, (a) non-sorted and (b) sorted droplets were measured after equal incubation times.
Present investigation

Figure 18: Analysis of specific productivity of selected clones after droplet sorting and S11-tag removal. a) The specific fluorescence, i.e. GFP signal divided by the integral of viable cell density, was measured for six clones isolated after droplet sorting in comparison to the unsorted polyclonal pool. Five out of these six showed productivities equal to or higher than the H-clone earlier isolated by LDC. b) One clone was chosen for S11-tag removal by CRISPR. Having confirmed successful tag removal, no difference in concentration of EPO secreted by the detagged clone compared to the tagged, parental clone could be detected when measured by BLI.

pared to the unsorted droplets (Fig. 17). Cells were recovered from the sorted droplets and single clones were isolated by means of LDC. A set of 20 clones was screened for EPO productivity and the six best performing clones were chosen for scale up and more detailed analysis. All six clones showed higher specific productivity than the polyclonal cell pool they originated from. Five out of six clones exhibited productivities on a par with or better than the high producing reference clone (H-clone) (Fig. 18a). Consequently, five out of 20 clones (25%) can be regarded as high-producing clones after droplet sorting as opposed to eight to twelve out of 160 (5-7.5%) in the unsorted, polyclonal pool. This equates to an enrichment factor of three to five. These results demonstrate that the developed platform can be used for the enrichment and isolation of high-producing clones.

Because the protein tag is not needed for the protein’s pharmaceutical function, it is reasonable to remove it. To this end, we also devised a generic way of S11-tag removal based on a CRISPR/Cas9-mediated double strand break and homology directed repair, which inserts EPO’s natural stop codon between the EPO coding sequence and the S11-tag. The H-clone was chosen for detagging and transfected with three plasmids coding for Cas9, guide RNA and donor DNA, respectively. Analysis of transfected cells by MiSeq and split-GFP complementation showed a detagging frequency of 6%, which is in line with previous experiments [269]. One detagged clone was isolated and analyzed. A size-shift on Western blot confirmed
successful detagging. Most importantly, the silencing of the S11-tag was achieved without compromising EPO productivity as illustrated in Figure 18b.

Taken together, we have developed a platform for the successful enrichment and recovery of high-secreting clones employing split-GFP complementation based droplet sorting and CRISPR/Cas9-mediated de-tagging of the isolated clone. Compared to many FACS-based selection methods this platform allows the sorting of cells based on protein secretion. Another advantage is that, in contrast to previous applications of cell screening and sorting in droplets [227, 270], the use of a split-GFP complementation assay in droplets does not depend on product-specific antibodies or enzymatic substrates and, thus, the platform will most likely be applicable to a wide range of proteins. Apart from screening and sorting pools of transfectants, we envision this technique to be useful for cell line engineering to evaluate the effect of, for instance, different chaperones, secretion enhancers or other components promoting protein production and secretion.

**Paper V: Improving speed and sensitivity of split-GFP by chromophore pre-maturation**

In the previous chapter, an incubation time of at least four hours between commencing cell encapsulation and performing droplet sorting was chosen. This is partly to allow for sufficient protein secretion but it is also owed to the rather slow kinetics of fluorescence development from split-GFP. Improving these kinetics would further increase the throughput of the presented cell selection platform. Likewise, other reported applications using split-GFP complementation, such as solubility assays [122], protein localization analysis [271] and phage display [272], are likely to benefit from faster kinetics and shorter incubation times. For example, split-GFP complementation has been used for protein quantification in cell and bacterial supernatants requiring an overnight or six hour incubation at room temperature, respectively, which might not be feasible for all proteins [120, 125].

While the association of the two GFP parts is a fast process, the slow kinetics are due to fluorophore formation as it requires the autocatalytic cyclization of the three residues forming the fluorophore [121, 273]. As it has been shown earlier that split-GFP can be disassembled after complementation and fluorophore formation, with the fluorophore in GFP 1-10 remaining in a mature state [274], we assumed that production of GFP 1-10 with a pre-matured chromophore should be possible and allow for faster fluorescent read-out.
Figure 19: Confirmation of chromophore cyclization by mass spectrometry. A comparison of GFP 1-10 before (a) and after (b) maturation on beads shows a 20.77 Da loss in mass. This corresponds to the cyclization and covalent rearrangement during the chromophore maturation process resulting in a loss of water and three hydrogen atoms.

For the production of GFP 1-10 with a mature fluorophore, GFP 11 was immobilized on sepharose beads and complemented with soluble GFP 1-10. After an overnight incubation at room temperature which allowed the fluorophore maturation to take place, mature GFP 1-10 was eluted by low pH and refolded. Successful maturation manifests in a molecular weight loss of 20 to 21 Da owing to the loss of water and two or three additional hydrogen atoms during cyclization depending on different maturation theories [273]. The detection of a mass reduction of 20.778 Da by mass spectrometry consequently confirmed the successful maturation of the eluted GFP 1-10 (Fig. 19).

The performance of the mature GFP 1-10 was subsequently evaluated with non-mature GFP 1-10 as benchmark. An excess of GFP 11 was added to both GFP 1-10 variants and development of fluorescence was monitored over time (Fig. 20a). Mature GFP 1-10 reaches its fluorescence half maximum after ten minutes while
non-mature GFP 1-10 exhibits a half-time of 100 minutes, which, thus, confirms our initial assumption of shorter incubation times and faster read-outs with pre-matured GFP 1-10. Remarkably, a concentration of mature GFP 1-10 equal to that of its non-mature counterpart gives rise to a much higher fluorescence. This can probably be attributed to the fact that the maturation procedure presents an additional purification step and hence, the mature GFP 1-10 preparation contains a higher degree of functional proteins. The mature GFP 1-10 showed some background fluorescence even in the absence of GFP 11, which may be ascribed to slight changes in spectral properties compared to non-mature GFP 1-10. Co-elution of GFP 11 was excluded as a reason for background fluorescence based on Western blot analysis. However, the higher background did not affect mature GFP 1-10’s assay performance as the signal to noise ratio was also increased by
factor four compared to non-mature GFP 1-10.

Next, the ability of mature GFP 1-10 to discriminate between different protein-secreting cell lines as initially envisioned was assessed. For this, both GFP 1-10 variants were added to the three EPO-S11-producing reference cell lines described in paper IV, which secrete high, medium or low amounts of EPO-S11 respectively. When monitoring protein secretion over time in a 96-well plate, the high-producer could be distinguished from medium and low-producing cells right from the beginning using mature GFP 1-10 (Fig. 20b). A differentiation between moderately and low-secreting cells was possible after five hours. With non-mature GFP 1-10 a statistically firm distinction of the three cell lines grown in a 96-well plate was not possible at any time point. Consequently, mature GFP 1-10 presents a promising reagent for the measurement of protein secretion and detection of high-producing cells.

In summary, the described procedure enables the production of pure and highly active, mature GFP 1-10. The matured form of GFP 1-10 is superior to its non-mature equivalent both in speed and sensitivity and its use is thus advised for all in vitro applications that require fast read-out and high sensitivity.
Concluding remarks and future perspective

The very diverse nature of biologics lets them fulfill functions and meet medical needs that cannot be satisfactorily answered by small chemical drugs [52, 129]. In this way, biologics have contributed to improve and prolong the live of millions of patients. One of the biggest assets of especially affinity proteins such as antibodies is their higher target specificity compared to small molecule drugs, leading to reduced adverse effects [52]. Consequently, the production of biologics with antibodies at the forefront is a multi-billion dollar industry with constantly increasing market shares. While in 2014 biologic-based products were forecast to account for 20% of the total pharmaceutical market in 2017 [199], this projection was recently raised to 25% [56] with monoclonal antibodies and antibody-based products accounting for the biggest share [56, 57].

However, as outlined in chapter 1.3.4, biotherapeutics are not without flaws. Risk for immunogenicity, need for parenteral administration and higher treatment costs can limit their application. Consequently, the last decades have seen many developments in the pursuit of generating more efficacious drugs at reduced costs.

The five papers that form the basis of this thesis present different tools with the potential to further advance biologics manufacturing. By covering protein engineering, precision medicine and cell line development, some of the key aspects that are thought to form and advance the biotherapeutics sector have been addressed. The realization of precision medicine requires the development of novel therapeutics to allow every patient to be offered a suitable treatment. The antibody engineering strategies described in papers I and II, focusing on directed evolution of mAbs and rational design of bispecific antibodies, respectively, demonstrate ways for the design of novel antibody therapeutics. The phage-staphylococcal display platform was shown to be a promising tool for the efficient selection of scFv molecules. It should be interesting to apply this platform also to other targets. AffiMabs show therapeutic potential, but their clinical applicability also in terms of alternative targets has to be evaluated.

The striving for precision medicine is accompanied by a demand for improved stratification methods to assign patients to a treatment regime that fits their needs, which is highlighted by the FDA’s encouragement of pharmacogenomic evaluation during drug development [275]. The third article highlights the importance of epitope information for the implementation of precision medicine. The epitope mapping approach described should be applicable to other binding pairs

61
but this has still to be verified in practice.

Ultimately, the transition towards precision medicine entails the need for a larger variety of pharmaceuticals that are produced in smaller batches. To render the production of smaller amounts economically viable, the production process needs to become more efficient and yields have to be improved. One strategy is the development of better production clones. Generic selection methods, like the one devised in paper IV, which employs split-GFP complementation-based droplet sorting, can help to select the highest secreting clones from a heterogeneous cell pool. Together with the improved, pre-matured split-GFP probe developed in paper V, the throughput can be further increased allowing for a more thorough screening of the cell pool. In the end, however, the production host that is utilized limits achievable production rates. The availability of genome information and editing techniques has spurred efforts to generate novel production strains that offer desirable characteristics not just in terms of productivity but also, for example, in terms of glycosylation pattern. Particularly, the CRISPR/Cas9 technology enables efficient genome editing, although it is mainly used in research and development as its industrial application is impeded by unresolved intellectual property issues [211]. Strain libraries generated by any genome editing approach could be screened with the split-GFP droplet-sorting platform to identify genomic modifications that increase the secretional capacity of a cell. Though an interesting utilization, the suitability of the platform for this application is yet to be proven.

In summary, the multitude of strategies devised to tackle the problems of biotherapeutics, of which some have been described in this thesis, will ensure a dominant role of biologics in the pharmaceutical sector in future. A combination of strategies involving both protein and cell line engineering, vector design and improved cultivation technologies, to name just a few, will probably be needed to achieve the desired outcome. Only time will tell which strategies will prove most applicable and successful in the future.
Acknowledgements

I could not have performed the work in this thesis without the support of many people who I was fortunate enough to meet and who I want to acknowledge at this point.

First, I want to express my gratitude to the ProNova Centre for Protein Technology, the NNF Center for Biosustainability, the WCPR -Wallenberg Center for Protein Research, the Knut and Alice Wallenberg foundation and Gålöstiftelsen for funding my research and enabling me to attend international research conferences.

Johan, my main supervisor, thanks for taking me in for a short research project, when I was an Erasmus student, which turned out to be the beginning of a so much bigger project. Your office door has always been open for me and your unwaning optimism and enthusiasm is inspiring. I could not have wished for a better “Doktorvater” (which would be your title in German) but, honestly, your e-mail address already predestined you for that role ;)

Mathias, my co-supervisor, you have always been involved and interested in the progress of my projects. Thanks, also, for taking the time to review and give me valuable feedback on my manuscripts. Your enthusiasm for science is truly inspirational.

Sophia, tack för att du granskade den här avhandlingen i en tid då du också hade så många andra saker på ditt bord. Dina kommentarer var till stor hjälp.

Per-Åke, vad skulle plan 3 vara utan dig?! Tack för att man alltid får störa dig med en fråga och när man kommer ut ur ditt kontor har man fått med sig så mycket mer än bara svaret på frågan.

Thanks to all my collaboration partners in the different projects. In particular, I want to thank Patrik, Magnus and Erik at SOBI for fruitful meetings and the
opportunity to work in your lab. Thanks to Petter and Håkan for teaching me how to sort droplets. I also want to thank Henning and Helene for inviting me to DTU and teaching me how to “CRISPR”.

Tack till alla ni som jobbar och har jobbat i HPA. Jag har tappat räkningen på hur många gånger ni har hjälpit mig med allt från att låna en sekundär antikropp till proteinrening. Ett särskilt tack till Henrik, Klas och Annelie C. för hjälp med just proteinrening.

Det var så roligt att få uppleva hur vår lilla grupp växte från tre doktorandar till de elva medlemmar som vi är idag. Tack till Johan, Magnus, Francis, Niklas, Magda, Mona, Ronia, Num, Aman och Max för att ni gör “the rockgroup” en så trevlig grupp att jobba i.

Jag hade tur och fick resa en del i vetenskapens namn. Tack till Sara, Johan S., Micke, Ken, Sarah, Anna B., Magda, Mona och Magnus för trevligt sällskap!

During my journey to a window seat, I have seen many people join and leave our fantastic office at the far end of the corridor. Thanks to Jonas, Mona, Jakob, Hao, Shengze, Magda, Johan S., Björn, Malin, Joel and Håkan for being such good company and for all the scientific and not so scientific conversations.

I also want to thank all current and former colleagues at “Plan 3” for making it such a great working environment. A special thanks goes to Lan Lan, Inger, Emma, Kristina B., Kristina J. and Kicki. Our everyday work would not be so smooth without your efforts.

Tack till mina fantasktiska kollegor som också har blivit riktigt goda vänner med åren. Josefine, Ken, Magnus, Linnea, Sara, Sarah, Micke och Tarek ni har hållit mitt humör uppe med bräddspelkvällar, middagsklubbar, sångkvällar, stickjuntor, segling och klättring. Jag hoppas det blir mer av det i framtiden!


What would my life be without rowing?! The ARF boathouse has become a second home to me. Thanks to all lovely people at ARF for making that club
such a great place to hang out at. Thanks to Marianne, Christina, Mikaela, Fredrica and Natalia for letting me be part of your crew and to Connla for coaching us. Thanks to the ladies’ quad with Marianne, Cat, Sara and Helena for my first SM-competition, for teaching me that you never regret a swim, for extensive Dolly Parton nights at the boathouse, and for ice skating on Norrviken. Let us continue like this! And Henley next summer is already in my calendar, girls.

Marianne, thanks to ARF you have become not only my double partner, but also a really good friend, and we have apparently become so inseparable that not even our coach can tell us apart. Thanks for teaching me to appreciate morning rows! I’m positive that we will manage to arrange more double races, cross-country skiing and Tatort evenings even in future.

Cat, Ginny, Tina and Alva, I still haven’t got a clue how netball works, but I will come and watch you one day. Until then, I’m looking forward to more, nice girls dinners, concerts at KMH and a weekend in Vålådalen next winter so we lure Marianne here from Bonn.

Ginny, thank you so much for proofreading my thesis. Your input was really valuable.


Thanks to all my friends back in Germany. Anna G., Lotti, Marta, Suse, Anna K., Marion, Robin und Thiemo, danke, dass ihr mich noch nicht aufgegeben habt und wir es schaffen den Kontakt zu halten trotz oft monatelanger Funkstille.


Last and least I want to the thank my tea pot. You had to stand a lot of comments, but kept me company and kept me hydrated throughout the whole writing process. More so, you helped me disprove an often-heard hypothesis: it is actually possible to write a PhD thesis without coffee :)
Bibliography


[23] Graus-Porta D, Beerli RR, Daly JM, Hynes NE (1997) ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. The EMBO journal 16(7):1647–55.


