Utilizing Solid Phase Cloning, Surface Display
And Epitope Information for
Antibody Generation and Characterization

FRANCIS JINGXIN HU

Doctoral Thesis in Biotechnology
KTH – Kungliga Tekniska Högskolan
Royal Institute of Technology
School of Biotechnology
April 2017
Stockholm, Sweden
Abstract

Antibodies have become indispensable tools in diagnostics, research and as therapeutics. Although the use of polyclonal antibodies (pAbs) has seen considerable use in research and diagnostics, the generation of monoclonal antibodies (mAbs) is necessary to avoid the drawbacks of polyclonal antibodies, in particular when aimed at therapeutic use. There are several strategies to generate monoclonal antibodies, including the hybridoma technology and display technologies. For therapeutic uses, the growing interest in precision medicine requires a well-characterized target and antibody in order to predict the responsiveness of a treatment towards a patient group. This thesis describes the usage of epitope information and display technologies to generate and characterize antibodies. In Paper I, we evaluated whether the epitope information of a well-characterized pAb could be used to generate mAbs with retained binding characteristics towards the same epitope region. In Paper II, the epitope on the complement protein C5 towards Eculizumab was mapped in detail using surface display. The results from the epitope mapping explained the non-responsiveness towards Eculizumab treatment among a Japanese patient group due to a mutation in their C5 gene. With this in mind we showed efficacy in treatment of the Japanese C5 variants using a drug aimed at different site of C5, suggesting that our approach can be used to guide treatment in precision medicine. In Paper III, a Gram-positive bacterial display platform was evaluated with the aim to complement existing platforms for selection of human scFv libraries. When used in conjunction with phage display, a thorough screening was possible and low affinity binders were isolated. In Paper IV, a solid phase method for directed mutagenesis was developed to allow for generation of highly functional scFv affinity maturation libraries by simultaneous targeting of all six CDRs. The method was also capable of creating numerous individual mutants which were used to map the paratope of the parent scFv. The paratope information was used to create directed libraries and deep sequencing of the affinity maturation libraries confirmed the viability of the combination approach. Taken together, precise epitope/paratope information together with display technologies have the potential to generate attractive therapeutic antibodies and direct treatment in precision medicine.

Keywords: antibody engineering, antibody affinity maturation, combinatorial protein engineering, epitope mapping, FACS, HER2, complement C5, Staphylococcal surface display, surface display.
Popular Science Summary

Proteins are not only found in our diet, but also in our cells. There, proteins are involved in communication, cell growth, transportation and defense. The proteins in our immune system are extremely vital in keeping us healthy and safe from diseases. A key player in this formidable defense network is the antibody, which is produced in response to an invading pathogen. Because the antibody is able to bind to its target at a specific location (called the epitope), the antibody molecule has been subjected to many types of artificial modifications for various applications. One of these applications is treatment of diseases because proteins, ironically, are also involved in numerous diseases and are responsible for making us sick. Therefore, there has been considerable interest to generate and modify antibodies to combat such diseases by targeting these disease related proteins. The oldest method to generate antibodies is to immunize an animal with the desired molecule, but this approach produces a finite amount of antibodies (called polyclonal antibodies) and further immunizations unfortunately lead to batch-to-batch variations. A breakthrough technique pioneered in 1975, called the hybridoma technology, enabled the generation of an infinite source of identical antibodies (called monoclonal antibodies). In Paper I, we determined the binding site of a well-performing polyclonal antibody and investigated whether that information can be used to generate monoclonal antibodies which retained similar properties. However, simply possessing a therapeutic antibody against a disease related protein does not guarantee that treatment will be successful. In Paper II, we investigated the reason why a certain patient group was not responding to a specific therapeutic antibody. Our experiments revealed that a mutation in the DNA of that patient group prevented the therapeutic effect of that antibody from taking place. This means that such information can be used to tailor different treatments to different patients, the concept of which is known as precision medicine.

If you are feeling anxious over the idea of using animal antibodies in humans, your anxiety is fully justified. Simply injecting animal derived antibodies into humans cause unwanted effects and thus various efforts have been made to generate fully human antibodies. But since immunizing humans is usually not possible, several techniques have been developed to circumvent this issue. One of these technologies is called surface display technologies and can be described as “artificial immune systems in a test tube”. Here, a pool consisting of millions to billions of antibodies can be artificially created using either artificial human genetic material or genetic material isolated from human donors. This pool can then be used to find a suitable antibody towards a target molecule. First pioneered in 1985, surface display technologies currently consist of several different platforms. In Paper III, we presented and evaluated yet another surface display platform for generating human antibodies. However, simply finding a suitable antibody is only half the story. Often antibodies are subjected to modifications to improve the binding strength, called affinity maturation. Paper IV, we present a method that improves the ease of handling and allows the affinity maturation process to take place more conveniently compared to several currently existing methods.
**List of Abbreviations**

Ab – antibody  
ABP – albumin binding protein  
ABR – antigen binding region  
Ag – antigen  
BLI – biolayer interferometry  
CDR – complementarity determining region  
ELISA – enzyme linked immunosorbent assay  
FACS – fluorescence activated cytometric sorting  
Fab – fragment antigen binding  
Fc – antibody constant domain  
FDA – Food and Drug Administration  
FR – framework region  
Fv – antibody variable domain  
HSA – human serum albumin  
Ig – immunoglobulin  
IgG – immunoglobulin G  
mAb – monoclonal antibody  
pAb – polyclonal antibody  
scFv – single chain variable fragment  
SDR – specificity determining region  
SPR – surface plasmon resonance
Structure of the thesis

This thesis is based on the work presented in the four appended publications and manuscripts. The thesis provides a perspective on how properly characterized antibody information can aid in the generation of focused antibodies and guide therapeutic treatment. Moreover, this thesis will also present the methods utilized to achieve such purposes. This section will first present the objective of the thesis followed an outline of five chapters and a list of the papers that the work is based upon.

Objectives

The general aim of this thesis is to investigate how the epitope and paratope information of an antigen or antibody can be used to generate and characterize antibodies for therapeutic use. For such purposes, directed evolution and semi-rational based methods have been employed. Studies I and II aimed to investigate whether properly characterized epitope information could be used to generate renewable monoclonal antibodies and guide treatment in precision medicine. Study III presents and evaluates a directed evolution antibody engineering platform that can complement existing methods. Finally, study IV introduces a cloning method that can be conveniently used to improve antibodies and obtain paratope information. The specific details of each study is presented in greater detail in Chapter 5.

Thesis outline

The five chapters of this thesis offer a theoretical and general overview of the antibody molecule and how various information and methods can aid in antibody engineering.

In Chapter 1, proteins and a general description of protein engineering are introduced. Protein engineering refers to approaches that can be used to modify proteins and consists of rational design and directed evolution.

Chapter 2 proceeds to present how proteins can be both the scourge and the savior in various diseases. Specifically, proteins involved in cancer and a malfunctioning immune system are presented. The chapter then continues to introduce the antibody molecule, its structure and applications in therapeutic settings.

In Chapter 3, the molecular basis of antibody – antigen recognition is discussed. The definitions of epitope and paratope are provided in addition to various methods that can be used to determine the epitope information.

Chapter 4 introduces various approaches used to generate and isolate human antibodies. These can consist of in vivo based (transgenic mice), in vitro based (display technologies) or modification based (humanization). In addition, this chapter also presents a few strategies used to improve some commonly modified antibody properties (e.g. affinity & stability) and how display technologies can be used for such purposes.

Finally in Chapter 5, the work of the appended papers and manuscripts is reviewed.
List of appended papers

This thesis is based on the work presented in the four articles and manuscripts listed below. Full versions of all articles and manuscripts are appended at the end of the thesis.

**Paper I:** Hu FJ, Uhlen M, Rockberg J

Generation of HER2 monoclonal antibodies using epitopes of a rabbit polyclonal antibody.

*New Biotechnology 2014, 31:35-43*

**Paper II:** Volk AL, Hu FJ, Berglund MM, Nordling E, Strömberg P, Uhlen M, Rockberg J

Stratification of responders towards eculizumab using a structural epitope mapping strategy.

*Scientific Reports 2016, 6:31365*

**Paper III:** Hu FJ, Volk AL, Persson H, Säll A, Borrebaeck C, Uhlen M, Rockberg J

Phage and Gram-positive bacterial display of human antibody repertoires enables isolation of functional high affinity binders.

*Scientific Reports, submitted*

**Paper IV:** Hu FJ, Lundqvist M, Uhlen M, Rockberg J


*Manuscript*

All papers are reproduced with permission from the copyright holders
# Table of Contents

1. Proteins  
1.1 Antibodies  
1.2 Protein Engineering  
1.3 Rational Design  
1.4 Directed Evolution  

2. Antibodies and Proteins in Diseases  
2.1 Proteins in Cancer  
2.2 ErbB family of Tyrosine Kinase Receptors  
2.3 HER2/ErbB2  
2.4 The Complement System of the Immune System  
2.5 Antibodies  
2.6 The Structure of the Antibody Molecule  
2.7 Different Classes of Immunoglobulins  
2.8 How Antibody Diversity is Created in vivo  
2.9 Other Antibody-Based Affinity Scaffolds  
2.10 Applications of Therapeutic Antibodies  
2.11 Eculizumab for Treatment of PNH and aHUS  
2.12 Therapeutic Treatment of HER2 Positive Patients  
2.13 Cancer Immunotherapy  

3. The Antibody – Antigen Interface  
3.1 The Epitope  
3.2 The Paratope  
3.3 The Epitope – Paratope Interface  
3.4 Mapping of the Epitope  

4. Generation and Engineering of Human Antibodies  
4.1 Humanization  
4.2 Transgenic Mice  
4.3 Display Technologies and Library Construction  
4.4 Phage Display  
4.5 Flow Cytometry and Cell-Based Platforms  
4.6 Bacterial Display  
4.7 Yeast Display  
4.8 Mammalian Display  
4.9 Cell-Free Based Techniques and Combination of Methods  
4.10 Engineering of Antibodies  

5. Present Investigation  
5.1 Paper I – Generation of HER2 monoclonal antibodies using epitopes of a rabbit polyclonal antibody  
5.2 Paper II – Stratification of responders towards eculizumab using a structural epitope mapping strategy  
5.3 Paper III – Phage and Gram-positive bacterial display of human antibody repertoires enables isolation of functional high affinity binders  
5.4 Paper IV – SPUX – A solid phase uracil excision method for antibody affinity maturation and paratope mapping  

6. Concluding remarks and future outlook  

7. Acknowledgments
1. Proteins

Proteins are involved in almost every major process in the body and are essential for the existence of life. They play major roles in structural support of the cells, regulating biochemical processes, signaling between cells, enzymatic reactions, transporting molecules and protecting the host from diseases. However, at the same time proteins can be seen as a double-edged sword. While properly functioning proteins are essential and responsible for the well-being of the individual, the opposite can also be true. Malfunctioning, mutated or up-/down-regulated proteins are involved in the pathogenesis of many diseases, of which the most notorious is perhaps cancer. A more detailed description of cancer can be found in Chapter 2.

Although proteins are complex structures and can vary greatly in size and form, the basic building blocks of proteins consist typically of only 20 amino acids. All amino acids have a similar structure consisting of an amino group (-NH$_2$) and a carboxyl group (-COOH) bound to a central carbon (also called alpha carbon). The side chain, unique for each amino acid and responsible for the unique physical and chemical properties, is also bound to the alpha carbon. Two individual amino acids are joined by a peptide bond formed between the carboxyl group of one amino acid and the amino group of another, displacing a water molecule. Many amino acids joined by peptide bonds form a polypeptide chain. Once incorporated, the amino acids are usually referred to by their side chains and are called amino acid residues. The amino acid sequence of a protein is called the primary structure, and is created in nature in a process called the central dogma where genetic information is copied to an mRNA molecule from which the polypeptide chain is synthesized. The primary amino acid sequence can be organized to form localized alpha helices or beta sheets, known as secondary structure. Several alpha helices and/or beta sheets can be structured in higher order configurations to form the tertiary structure of a protein, giving a folded protein. Several tertiary structures or polypeptide chains can be organized into a quaternary structure, consisting of several subunits. The protein’s quaternary structure also describes the relationship between its subunits and how they interact.

1.1 Antibodies

Antibodies are a special class of proteins responsible for the defense of the host against pathogenic invaders. Highly specific, these molecules are produced by the B-lymphocyte cells (B-cells) of the host’s own immune system and are able to form strong binding interactions with a wide spectrum of target proteins (also known as antigens). In addition, antibodies are also able to recruit other cells within the immune system to eliminate pathogens. Therefore antibodies have become popular therapeutic agents, but they have also found immense use in other areas such as industrial processes, research, imaging and diagnostics. Chapter 2 provides more information on the antibody structure and its application in cancer treatment. In contrast to other protein-protein interactions that have evolved over the course of evolution, antibodies must be generated swiftly and be able to bind their respective antigens strongly. This means that the antibody-binding surface must be highly adaptable considering the large differences in shape, size and amino acid composition between various antigens. How this adaptability is possible is summarized in Chapter 3. Chapter 3 will also describe methods that can be used to determine the binding site of antibodies (called epitopes) on their respective antigens. These methods are collectively known as epitope mapping methods.

1.2 Protein Engineering

Protein engineering can be considered a sub-field of genetic engineering that involves modifying and improving the structure and properties of proteins in the laboratory for various purposes, including medicine (e.g. therapeutic antibodies), diagnostics (e.g. imaging),
manufacturing (e.g. industrial enzymes in pharmaceutical manufacturing) and household products (e.g. proteases in laundry detergents). Protein engineering can be used to study the relationship between the structure and the function of a protein of interest. The properties that are targeted for modification include, improved stability, improved solubility, improved pharmacokinetics, increased expression level, reduced immunogenicity and better molecular recognition.

The workflow of protein engineering often involves a few common steps. First, where the protein modification should occur is decided, followed by making those changes (using rational design or directed evolution) and finally assessing the modified proteins for desired properties (screening)(1). In detail, the three steps need to take the following into consideration: 1) location of the mutations (throughout the protein or at specific places), 2) types of mutations (single amino acid substitutions, deletions/insertions or homologous recombination), 3) mutagenesis strategies (single site or multi site), 4) choice of screening(1).

The protein of interest can either derive from a natural source (e.g. antibody genes isolated from mice) or have a synthetic origin. Synthetic proteins, also called recombinant proteins, refer to proteins that have been produced in the laboratory by recombinant DNA technology or molecular cloning. Here, various techniques have made it possible to modify and create novel genetic material absent in nature. One of the most important breakthroughs in recombinant technology is the polymerase chain reaction(2). This technology made it possible to generate sufficient amounts of genetic material for downstream processes. Other equally vital techniques include restriction enzymes and sequencing platforms to modify and assess the genetic blueprint. This genetic blueprint can then be introduced into various production hosts to produce (also called express) such recombinant proteins in large quantities. Nowadays, several hosts in addition to the popular E. coli are available that can be used as protein factories. These include, bacterial, fungal, plant and mammalian cells(3).

It has been suggested that the methods used for protein engineering can be visualized as a two-dimensional grid(1). On the X-axis is amount of information required, while the Y-axis shows the magnitude of modification. For instance, design of a new enzymatic reaction would require transition state information and multiple mutations in the protein(1). In contrast, other methods only require the amino acid sequence. Nonetheless, these different methods are often classified as either rational design or directed evolution.

1.3 Rational Design

In rational design, the amino acid positions in proteins that are targeted for mutations are chosen based on knowledge of the protein structure, function and/or computational modeling(4). These targeted amino acids are then usually subjected to site directed amino acid replacements, insertions or deletions at deliberately chosen positions within the protein. Rational design is commonly used in enzyme engineering(5) since it is a practical method when a straightforward relation exists between the structure and function of the protein(6). If there is a lack of structural data, alanine scanning(7) can be used to identify the crucial residues that contribute to the trait which is targeted for modification. In alanine scanning, selected residue positions are individually mutated to alanine and the effects of the mutations are assessed. Another similar but more laborious approach is saturation mutagenesis, where each residue position is mutated to all 20 amino acids(8). Computational modeling, a third alternative, is based on thermodynamically simulations to find energetically favorable folding and interaction states(9).

Despite many successful outcomes, rational design tends to be laborious where the main challenge is being able to unravel the complex relationship between the protein structure and its function and being able to correctly predict the effect that the mutation has on the
structure and function of the protein\(^4\)\(^,\)\(^10\). Prediction of the effect of a mutation can still be an issue even in the presence of structural data since small changes in the amino acid sequence can hugely impact the function of the protein\(^11\). Rational design methods are also considered time consuming and low-throughput since the mutants are typically produced and characterized one by one\(^9\), hence restricting each iterative cycle to a small number of variants.

1.4 Directed Evolution

Directed evolution, also called combinatorial protein engineering, is an alternative approach that attempts to mimic natural evolution via a specific artificially chosen evolutionary pressure. Generally, directed evolution involves iterative rounds of creating numerous protein variants (also called a library), typically in the range of \(10^5 - 10^{10}\), and applying a selection pressure to screen and isolate candidates with desired/improved properties. If we go back to the two-dimensional grid mentioned earlier\(^1\), directed evolution benefits from only requiring the amino acid sequence but is capable of both minor and major modifications to the protein\(^1\). The directed evolution methodology usually involves two separate steps. In the first step, genetic diversity is generated to create a library and in the second step the library is screened and the variants with the desired function are isolated by different selection methods. Not requiring structural data or computational simulations is often seen as the major advantage with directed evolution.

1.4.1 Creating the Library

The genes encoding a protein library can either be created by recombinant DNA technology or isolated from a natural source (e.g. antibody genes from a human B-cell repertoire\(^12\)). If the B-cell repertoire is derived from previously infected patients or immunized animals, the resulting library is called an “immune” library as it is enriched with target specific antibodies. If the library is not pre-enriched for any specific target protein it is called a “naïve” library\(^12\). Libraries created completely \textit{in vitro} using sequence information from various databases are called synthetic libraries\(^13\).

Two common methods used to construct recombinant protein libraries are random mutagenesis and site-directed randomization. Random mutagenesis refers to random diversification of the template gene whereas site directed mutagenesis involves incorporation of random codons at specific positions in the template gene. A common method used in random mutagenesis is the error-prone polymerase chain reaction\(^14\). Error prone PCR is a rapid and cost efficient approach that does not require any mutagenic oligonucleotides. Here, a polymerase with a high mutation rate is used to amplify the template gene of interest. The PCR reaction condition can also be modified to encourage further introduction of errors in the amplified gene. The resulting gene will have mutations randomly distributed along the sequence, which may be detrimental as stop codons and undesirable amino acids can be incorporated at unwanted positions. Another drawback with error prone PCR is unwanted biases in the library since certain amino acids are more likely to be incorporated than others. This is because some amino acids are encoded by several codons, whereas others are only encoded by one codon. Furthermore, error-prone PCR is likely to only substitute on average of six of the possible amino acids\(^15\) and struggles with altering two nucleotides in the same codon. In contrast, site directed mutagenesis utilizing mutagenic degenerate oligonucleotides is able to achieve a higher mutation rate at each codon position. A common degenerate codon is the NNN (\(N = \text{A, G, C, T}\)) combination where each residue position can be mutated to all twenty amino acids.

While complete random directed evolution sounds simple and straight-forward, such an approach can easily generate a library with a complexity (size) too large to be analyzed with most existing selection methods\(^15\). On the other hand, the likelihood of finding a protein
variant with a desired property increases with library size(12). To solve this dilemma, semi-rational approaches can be employed where the focus is to reduce the library size while maintaining the functionality relative to a complete random library(16). One approach is to reduce codon redundancy. For example, the NNN degenerate codon unnecessarily covers all 64 different codons (including 3 stop codons), resulting in over-representation of certain amino acids since some amino acids are encoded by multiple codons whereas others are encoded by a single codon. The NNK (K = G, T) codon is a better less redundant alternative as it still manages to cover all 20 amino acids (and one stop codon) via 32 different codons. The NDT (D = A, G, T) is another popular codon that encodes a balanced mix of 12 different amino acids(17).

While the NNK and NDT codons can reduce redundancy, the distribution and bias of certain encoded amino acids cannot be altered. The development of trinucleotide synthons, on the other hand, allows each amino acid to be represented by only one codon(18). Trinucleotide synthons also enable the possibility to introduce higher relative amounts of certain amino acids at desired positions. For example, unwanted codons such as stop codons, cysteines or prolines can be avoided completely, and the level of desirable codons such as tyrosine can be increased. Using trinucleotide randomization is useful for generation of antibody libraries, where certain amino acid residues are vital for binding to the target protein, and therefore needs to be randomized to a lower degree. A final interesting approach to reduce the resulting library size focuses on “hot spot” residue positions or “attractive residue codons”. Either, a smaller number of positions can be selected for randomization, or a larger number of positions can be randomized using a very limited number of amino acids, such as tyrosine/serine/alanine/aspartate or tyrosine/serine only as has been reported earlier(19–21).

1.4.2 Screening the Library

Being able to subsequently identify and isolate the interesting variants within the diverse library is just important as generating said library. To achieve this, a suitable selection / evolution pressure is applied to the entire library and the desired mutants are identified by screening and isolated.

Since the chances of finding a suitable candidate increases with library size(12), high-throughput selection- or screening methods are necessary to cover these large library sizes. In addition, these screening/selection methods also need to have good sensitivity (high signal to noise ratio), ability to isolate rare clones, ability to find small improvements (fine discrimination), robustness, reproducibility and sensitivity to the desired function of the protein(22). Therefore, different screening strategies / methods will be appropriate for different protein properties.

A common requirement for all screening / selection methods is the crucial genotype-phenotype link, meaning that each protein variant in a library (the phenotype) should be coupled to its corresponding genotype, i.e. the gene encoding that particular protein sequence(23). This link allows all variants to be analyzed simultaneously in a single batch experiment instead of analyzing them individually. Once an interesting candidate is found, the corresponding encoding gene can be used to identify the sequence and downstream individual characterization of selected variants is straightforward after amplification. Various selection technologies are available today and these will be described more in detail in Chapter 4. Generally these technologies are divided into cell based and cell free systems(24). In cell-based systems (also called surface display technologies) the protein of interest (phenotype) is expressed on the surface of a microorganism with its corresponding genetic sequence (genotype) located inside the cell. The displayed protein library is subjected to an appropriate selection strategy and the encoding genes from isolated desirable candidates can be recovered. The recovered genes can then be used for either sub-cloning and downstream
characterization, or as starting material for additional selection rounds. The microorganisms used in surface display technologies can be bacteriophages (phage display), bacteria (bacterial display), yeast (yeast display) and mammalian cells (mammalian display). Cell free systems (e.g. ribosome display) are similar to cell based systems but the proteins are expressed without the use of living cells.

I hope that this chapter has provided you with an adequate introduction to the topics and techniques that will be covered in this thesis. For now, let us first agree that proteins can be considered both the scourge and the savior when it comes to disease. The next few chapters will first take a closer look at the scourge: the proteins involved in cancer and a malfunctioning immune system. The savior, therapeutic antibodies, will be discussed further in Chapters 2 and 3 while Chapter 4 will look at how combinatorial protein engineering can be utilized to improve these antibodies.
2. Antibodies and Proteins in Diseases

This chapter will first take a general look at proteins involved in cancer and a malfunctioning immune system. Later, a more detailed description of antibodies is provided, which is followed by a brief introduction into how therapeutic antibodies can be used to combat the mentioned diseases.

2.1 Proteins in Cancer

Cancer is a generic term referring to over one hundred different types of diseases that can be found within different specific organs(25). According to the World Health Organization (www.who.int), cancer is one of the leading causes of mortality worldwide responsible for almost 9 million deaths each year, with lung, liver, colorectal stomach and breast cancer being the most common subtypes. A common characteristic for all cancer types is the uncontrolled growth of cells while also possessing the ability to invade nearby tissues. In 2000, Hanahan and Weinberg made an attempt to provide common traits, or hallmarks, that are present in all cancer subtypes. By 2011 the same authors provided several more additional hallmarks(26) based on more updated research. Both the original six hallmarks(25) and the newer hallmarks(26) described by Hanahan and Weinberg are briefly summarized below(25, 26) to provide a general introduction to understand why cancer remains an elusive disease to cure.

The first hallmark is termed “self sufficiency in growth signals”, meaning that tumor cells show a reduced dependency on growth signals produced by another cell type. While normal cells require such stimulus to proliferate, tumor cells are able to generate their own growth signals to which they are responsive, resulting in a positive feedback loop called autocrine stimulation(25). Another strategy involves overexpression of growth signal transducing surface receptors, leading to hypersensitivity to levels of growth signals that would normally not trigger proliferation. Yet a third strategy involves changes that occur in components involved in cytoplasmic downstream signaling(25, 26). For example, in the SOS-Ras-Raf-MAPK cascade, altered forms of the Ras protein are able to release growth signals to cells without upstream stimulation(27) and approximately 40% of human melanomas contain mutations affecting the structure of the B-raf protein(28).

The second hallmark is called “insensitivity to antigrowth signals” implying evading various anti-proliferative signals. Most programs involved in negative regulation depend on tumor suppressor genes and two important tumor suppressors are the RB (retinoblastoma associated) and TP53 proteins as they occupy central roles within two key cellular regulatory circuits that determine cell proliferation or cell apoptosis(25, 26). The role of RB is to prevent uncontrolled cell growth by inhibiting proliferation until a cell is ready to divide. In contrast, TP53 receives inputs from stress sensors from within the cell and can halt cell division if the stress levels reach a threshold. In case of overwhelming stress indicating severe or irreparable damage to cellular systems, TP53 can trigger cell apoptosis(25, 26).

Apart from evading anti-growth signals, cancer cells are also able to evade apoptosis, or programmed cell death (third hallmark). The apoptotic system consists of upstream regulators and downstream effectors(29). The regulators can be divided into two programs: the extrinsic (receiving extracellular apoptotic signals) and the intrinsic (sensing intracellular signals). Both circuits lead to activation of caspase 8 and 9, which initiate the execution phase of apoptosis(25, 26). The most common evasion strategy used by tumor cells is the inactivation of TP53 tumor suppressor function, a component of the apoptotic signaling circuitry. Other ways include increasing expression of anti-apoptotic regulators, by down regulating pro-apoptotic factors, or by disrupting the extrinsic ligand-induced apoptosis pathway(25, 26).
The fourth and fifth hallmarks suggested by Hanahan and Weinberg is called “inducing angiogenesis” and “replicative immortality” respectively(25). Angiogenesis, the process of new blood vessel growth, is usually only temporarily activated for example during wound healing. However, in tumor progression, an “angiogenetic switch” is constantly on, resulting in new vessels that sustain tumor growth(30). This angiogenetic switch(31) is a controlled by a balance between angiogenesis inducers (e.g. VEGF(32)) and angiogenesis opposers (e.g. TSP-1(33)). Replicative immortality refers to the acquisition of unlimited replicative potential by cancer cells, in contrast to normal cells that experience senescence after a certain number of cell division cycles. Telomerase plays a vital role in the capability for unlimited proliferation(34).

Genome instability and reprogramming energy metabolism are additional cancer traits that refer to a breakdown in the genomic maintenance machinery and the existence of a metabolic switch in cancer cells(26). The genomic maintenance machinery is responsible for detecting and repairing defects in DNA and keeping mutations low during each cell generation(35, 36) while cancer cells show increased rates of mutations(37). The metabolic switch refers to reprogramming of glycolysis metabolism in cancer cells, favoring glycolysis under aerobic conditions instead of oxidative phosphorylation(38). To compensate for lower ATP yields cancer cells upregulate glycolysis(39), thus allowing diversion of glycolytic intermediates into various biosynthetic pathways used to generate nucleosides and amino acids, which leads to synthesis of macromolecules and organelles used for assembly of new cells(40).

The final and perhaps the most important hallmark is "activation of metastasis"(25, 26). Being able to invade neighboring tissues is often seen as the difference between benign and malignant cancer. Metastasis begins with local invasion of host stroma often by intravasation into thin walled venules and lymphatic channels. The next stage is transit of cancer cells though the lymphatic and hematogenous systems, during which interactions with hematopoietic cells can lead to increase in cancer cell size. This is followed by escape of cancer cells from the lumina of vessels via extravasation, formation of nodules of cancer cells (micrometastasis) and finally growth into macroscopic tumors(41, 42). The newly formed tumors can themselves induce metastasis and the cycle repeats.

2.2 ErbB family of Tyrosine Kinase Receptors

As described earlier one of the cancer hallmarks is a reduced dependency on external growth signals and one strategy utilized by tumor cells is the overexpression of cell surface receptors that transduce growth(25, 26). One class of such receptors is the human epidermal growth factor (EGFR, HER or ErbB) family of tyrosine kinase receptors, which plays a central role and is involved in many types of tumors(43). The ErbB-family consists of four members: EGFR (ErbB1 or HER1), HER2 (ErbB2), HER3 (ErbB3) and HER4 (ErbB4), all of which possess a highly conserved extracellular ligand-binding domain (ca. 620 amino acids), a single trans-membrane α-helix domain and an intracellular domain (ca. 550 amino acids) with a tyrosine kinase domain (ca. 270 amino acids)(44) although the HER3 kinase domain has impaired activity(45).

Eleven different growth factors are known today to be able to bind to the four ErbB family members. Some such growth factors include EGF, TGF-α, Amphiregulin, Betacellulin, Epigen, Epirregulin and Neuregulin 1-4 (46). Binding of these agonist ligands to the extracellular domain of the receptor induces formation of homo- and heterodimers(47) followed by activation of the cytoplasmic kinase domain, although formation of ErbB1, ErbB2 and ErbB4 homodimers results in relatively weak signaling. HER2, being the exception, does not bind any of the ligands directly, but can be activated via homo-dimerization as well as hetero-dimerization with the other ligand-bound receptors(48). HER2 is also the preferred partner in receptor hetero-dimerization(49) and overexpression of HER2 not only favors hetero-
dimerization but also leads to more potent signaling(47). Several signaling pathways that members of the ErbB family are involved in include the Ras/MAPK, PI3K/Akt and Jak/STAT pathways, which are responsible for cellular processes such as apoptosis, migration, proliferation, adhesion and differentiation(45, 47). Nowadays, it is clear that the ErbB family members are activated and play major roles in the development of many types of tumors(48). This has made them attractive targets for cancer therapy which has led to the emergence and development of specific ErbB inhibitors. Since HER2 is the target antigen used for selection in several papers in this thesis, it will be described more in detail below.

2.3 HER2/ErbB2
First described in 1984 (50), HER2 is a 185 kDa glycoprotein with several alternative names such as ErbB2, CD340, or HER2/Neu. The HER2 gene is located on chromosome 17q21 (51) and a mutation at position 655 results in a structural change, promoting homodimerization(52) and correlation with a higher risk of breast cancer(53). Up-regulation or overexpression of HER2 is mostly associated with breast cancer with approximately a 20 – 30% occurrence rate(54, 55). In an overexpressing cancer tissue, the number of HER2 receptors per cell can increase from 20000-50000 in normal cells to 2 million in cancer cells(56). In addition, overexpression of HER2 in breast carcinomas is also often associated with poor prognosis, increased risk of developing metastasis and overall poor survival(57, 58). The underlying mechanism is that overexpression of HER2 induces anti-apoptotic signals (via PI3K/AKT pathway) while promoting proliferation (via activation of RAS-MAPK pathway) and inducing angiogenesis(52). Furthermore, unregulated HER2 expression has also been correlated with resistance to anti-HER2 therapy, usually through one of the following mechanisms: alteration to HER2 itself, utilizing other signaling pathways and defects in cell cycle and apoptosis(59). Apart from the most commonly associated breast cancer, overexpressed HER2 has also been found in ovarian cancer(60), stomach cancer(54, 61) and lung cancer, where mutation in the HER2 protein was present in 10% of patients(62).

2.4 The Complement System of the Immune System
The immune system is divided into two branches, the innate system and the adaptive system. However, the two branches of the immune system do not operate independently of each other. The complement system has a vital role in bridging the two systems together as described below.

The complement system’s main task is to enhance or complement (hence the name) the actions of antibodies and consists of a complex network of proteins while also playing a key role in both innate and adaptive immunity(63). Apart from affecting the innate immune responses, the complement also interacts with and influences T- and B-cell responses in the adaptive branch of immunity. The three biochemical pathways that activate the complement system include: the classical pathway, the alternative pathway and the lectin pathway. After activation, the various complement components interact to carry out a number of functions including: 1) lysis of cells and bacteria in an antibody dependent or antibody independent manner, 2) opsonization which promotes phagocytosis and 3) activating inflammatory responses(63).

The classical pathway is activated by the formation of soluble antibody-antigen complex or binding of antibody to a pathogenic surface. Both IgM and IgG can activate the classical pathway which first involves C1, C2, C3, C4 present in plasma in inactive forms. In short, the pathway is triggered by the activation of the C1 complex when it binds antigen bound antibody. C1 then cleaves C2 and C4 to form C2a, C2b, C4a and C4b. C4b and C2a bind to form the classical pathway C3-convertase (C4b2a complex), which promotes cleavage of C3 into C3a and C3b. C3b later joins with C3 convertase to form C5 convertase (C4b2a3b complex). The C3b component of C5 convertase binds C5, allowing C5 convertase to cleave C5
into C₅a (which diffuses away) and C₅b which forms the membrane attack complex (MAC, also called C₅b-9) together with C₆, C₇, C₈ and C₉.

The alternative pathway is a component of the innate immunity system since it is constantly activated at low levels in an antibody independent manner. The alternative pathway is initiated by the spontaneous hydrolysis of serum C₃ (which is abundant in plasma and contains an unstable thioester bond) to C₃a & C₃b. The pathway is triggered when C₃b binds cell surface constituents of foreign origin. The next few steps involves binding of factor B, factor D and properdin to create a complex with C₃ convertase activity. The alternative pathway C₃ convertase can then bind C₃b to form the alternative pathway C₅ convertase.

The lectin pathway also functions in an antibody-independent fashion and uses host proteins (e.g. mannose-binding lectin (MBL)), instead of C₁, to conduct nonself recognition. This pathway is initiated by binding of MBL to mannose residues on the pathogen surface, which activates the MBL-associated serine proteases (MASP-1 & MASP-2) which in turn cleave C₄ into C₄a & C₄b and C₂ into C₂a & C₂b. C₄b and C₂b then bind together to form the classical C₃-convertase, as in the classical pathway.

Nonetheless, complement activation regardless of pathway ultimately leads to the generation of C₃ and C₅ convertase, the crossroads of all three pathways. The C₅ convertase cleaves the important C₅ protein into anaphylatoxins C₅a and C₅b. C₅b combines with C₆, C₇, C₈ and C₉ to form the MAC (C₅b-9) complex, capable of lysing a broad spectrum of cells. In addition, the other fragments generated during the formation of C₅b-9 play a critical role in developing inflammatory responses (C₃a and C₅a) or facilitate opsonization (C₃b and C₄b).

As illustrated above, the complement system is a key player of the immune system and a defective complement is unfortunately involved in the pathogenesis of paroxysmal nocturnal hemoglobinuria (PNH) and atypical haemolytic uremic syndrome (aHUS). PNH is caused by mutations in the PIG-A gene in pluripotent hematopoietic stem cells. The mutation affects GPI-anchor biosynthesis which serves to anchor proteins to the surface of blood cells(64). Cells deficient in GPI-anchored proteins are in turn vulnerable for complement-mediated attack, leading to intravascular hemolysis, venous thrombosis, and hemoglobinuria(65, 66). Atypical hemolytic uremic syndrome (aHUS) is rare disease characterized by an unregulated activation of the alternative pathway of the complement(67), resulting in the formation of blood clots in capillaries and arterioles due to endothelial injuries. In most patients with aHUS, the underlying cause is production of anti-factor H autoantibodies or from genetic mutations in any of several complement regulatory proteins such as complement factor H, factor I and CD46(67).

So far, we have established that certain proteins play important roles in disease pathogenesis, as illustrated by cancer and a defective complement system. Next, let us take a look at how other proteins can be used to combat disease proteins using a “fight fire with fire” approach. One class of therapeutic proteins is (therapeutic) antibodies and the next section will expand the information on this important protein group.

### 2.5 Antibodies

Antibodies play a vital role in adaptive immunity and are produced by B cells which derive from pluripotent hematopoietic stem cells. Mature B cells leave the bone marrow and circulate in the blood and lymphatic systems. Antibodies exist in a membrane form, called B-cell receptor (BCR) on the surface of a naïve B-cell (one that has not previously encountered antigen) before interaction with a foreign antigen. Following an encounter with an antigen, the naïve B-cell rapidly divides and differentiates into memory B cells and effector B cells (also called plasma cells). Memory B cells have a longer life span than naïve B cells, and they
express the same membrane bound antibody as naïve B cells. Plasma cells, which have little or no membrane bound antibody, produce the secreted version of the BCR as free antibodies. In addition to antigen recognition, B-cells also require T-cell stimuli to differentiate into plasma cells. Although each B-cell produces one unique antibody, the vast repertoire of B-cells results in an enormous diversity. Apart from the immune system’s capability to create a vast diversity of antibodies, each antibody also possesses the ability to bind to a specific region, called epitope, on its respective antigen. When several different B-cells produce an antibody response towards the same antigen, such antibodies are called polyclonal antibodies (pAbs). The ability for several polyclonal antibodies to bind different epitopes can produce a high overall response although the contribution from each member can vary greatly. Therefore animal immunization is the most straightforward way to obtain antibodies for practical use due to the ease of manufacture. However, pAbs suffer from the disadvantage of batch-to-batch variation as the anti-serum is finite. Monoclonal antibodies (mAbs), in contrast, refer to identical copies of an antibody produced by the same B-cell. A huge breakthrough in 1975 allowed B-cells with a finite lifetime to be immortalized and thus produce an infinite amount of mAbs(68). This technology, called the hybridoma technology, has revolutionized the field of antibody generation for practical applications. However, using animal derived antibodies for treatment in humans leads to undesired immunogenic effects and considerable efforts have been made to “humanize” animal derived antibodies. Chapter 4 will provide more information on humanization as well as methods that can used to generate human monoclonal antibodies. But before that, the antibody structure needs to be introduced.

2.6 The Structure of the Antibody Molecule

The antibody molecule has two distinct tasks. Firstly, they need to bind to foreign antigens encountered by the host and secondly they need to mediate effector functions to neutralize foreign invaders. The Fc region is responsible for binding to effector cells and activating other parts of the immune system, such as ADCC, phagocytosis and the complement system(69, 70). In contrast the antigen-binding region has to bind to a wide array of antigens and to accomplish this, the antibody molecule has adopted an unique appearance in the form of a Y-shaped molecule (Figure 2.6.1).

The basic antibody molecule (of the immunoglobulin G subclass) consists of four polypeptide chains, two identical light chains and two identical heavy chains. The molecular weight of one light chain (L) and one heavy chain (H) is approximately 22 kDa and 55 kDa respectively. Each heavy chain is connected to a light chain by one disulphide to form a heterodimer (H+L). The two heterodimers are joined by similar disulphide bridges to form the basic four chain (H+L)₂ antibody structure with an approximate molecular weight of 150kDa. Depending on isotype, each antibody molecule can either adopt the basic four chain (H+L)₂ form or multiple units ((H+L)_n)ₙ of this basic structure.

The terminal regions of the two arms of the Y-molecule are called variable regions (V regions: VL for light chain and VH for heavy) and are responsible for the variety between different antibodies. Interestingly, most of the variety within the V regions is concentrated in distinct loop regions called complementary determining regions (CDRs)(71). The three CDRs in each VL and VH constitute the antigen binding site of the antibody molecule, whereas the segments between the CDR regions within VL and VH show much less variation and are thus called framework (FR) regions for their role in maintaining the structure of the antibody molecule. The regions of rather conserved sequences beyond the variable regions are called constant regions: CL in light chain and CH in heavy chain.

There are two isotypes of light chains in humans and mice, kappa (κ) and lambda (λ) and a normal antibody molecule contains only one light chain isotype, either kappa or lambda, but not both. As for the heavy chain, there are five major isotypes in humans, each corresponding
to a different heavy chain constant region: µ, δ, γ, ε, and α. The length of the constant regions is approximately 330 amino acids for δ, γ, and α and 440 amino acids for µ and ε. As a result, the five different isotypes give rise to five different antibody classes: IgM (µ), IgD (δ), IgG (γ), IgE (ε), and IgA (α). Furthermore, minor amino acid sequence differences within the human α and γ heavy chains have led to further subisotypes: α1 and α2 for α and γ1, γ2, γ3 and γ4 for γ. Since the isotype of the heavy chain does not affect its pairing ability to the light chain, heavy chains of any class may pair with either κ or λ light chains. The light chain consists of one variable (VL) and one constant (CL) domain. The heavy chain also contains one variable (VH) domain and either three (CH1, CH2, CH3) or four (CH1, CH2, CH3, CH4) constant domains depending on isotype. IgG, IgD and IgA have three heavy chain constant domains whereas IgM and IgE have four. Also depending on isotype, the different domains are connected by flexible hinge regions that allow movement and adjustments between the domains. While IgA, IgD and IgG possess a highly flexible hinge between CH1 and CH2, no such hinge exists in IgE and IgM.

Figure 2.6.1. Structure of the antibody molecule and antibody fragment based scaffolds
2.7 Different Classes of Immunoglobulins

Immunoglobulin G (IgG), the most abundant isotype in serum, is made up of two $\gamma$ heavy chains, two $\kappa$ or $\lambda$ light chains and consists of the basic ($H+L)_2$ monomeric structure. As described earlier, differences within the four $\gamma$ heavy chains have given IgG four subclasses: IgG1, IgG2, IgG3 and IgG4. IgM constitutes about 5-10% of the total serum immunoglobulin and monomeric IgM is expressed as membrane bound antibody on B-cells whereas plasma cells secrete IgM as a pentamer. The five units are arranged with their Fc regions in the center of the pentamer and the 10 antigen binding arms on the periphery of the molecule. IgM is the first immunoglobulin class produced in a primary response. Due to its pentameric structure with 10 antigen-binding sites, IgM has higher valency than the other isotypes. In addition, IgM is more efficient then IgG at activating complement since two or more Fc regions in close proximity are required. IgA exists primarily as a monomer in serum, but polymeric forms can exist, while IgE antibodies play a role in hypersensitivity reactions.

Immunoglobulin domains are folded into a structure called the immunoglobulin fold with an approximate size of 110 amino acids. This structure consists of a “sandwich” of two beta sheets, each containing antiparallel beta strands connected by loops. Antibodies have two types of immunoglobulin folds, consisting of either seven or nine beta strands (and an extra loop). The former is structural unit of the constant regions, whereas the latter is found in the variable region. Unsurprisingly, it is the loops connecting the antiparallel beta strands of the VL and VH domains that correspond to the CDRs.

2.8 How Antibody Diversity is Created in vivo

The kappa and lambda light chains and heavy chains are coded by separate multigene families situated on different chromosomes. In unarranged germ-line DNA, each of these multigene families contains several coding sequences, called gene segments, separated by noncoding regions. During B-cell maturation, these gene segments are brought together to form functional immunoglobulin genes. The kappa and lambda light chain families contain V, J and C gene segments and the rearranged VJ segments encode the variable domain (VL) of the light chains. The heavy chain family contains V, D, J and C segments and the rearranged VDJ gene segments encode the variable domain (VH) of the heavy chain. In each gene family, the C gene segments encode the constant domains.

During B-cell maturation, the heavy chain variable region genes rearrange first, followed by rearrangement of the light chain variable gene regions. During rearrangement in the kappa light chain, any of the functional Vk genes can combine with any one of the Jk genes to form the VL domain. The C gene is then joined to the VJ segments, resulting in a complete rearranged antibody light chain. The lambda light chain in humans undergoes a similar rearrangement. On the other hand, generation of a functional heavy chain requires two separate rearrangement events within the variable region. The first step involves joining any one of the DH genes with any JH gene to form a resulting DHJH unit. A VH segment then joins to form a VDJ unit that encodes the entire heavy variable region. In the final step, a C gene segments is attached to form the final immunoglobulin heavy chain construct. Joining of V-(D)-J segments is imprecise. Further junctional diversity at the VJ and VDJ is created by trimming of the coding joints as well as addition of extra nucleotides at the junctions.

In summary, antibody diversity in humans is created in several different ways in vivo. Firstly, multiple V, D and J gene segments can be randomly combined and imprecise joining of V-(D)-J gene segments leads to additional diversity. It is the diversity at the V-(D)-J interface that forms the third hypervariable region (CDR3) of the heavy and light chains. Furthermore, addition of extra nucleotides at the coding joints adds more diversity. Finally, after a mature B-cell encounters an antigen, an in vivo affinity maturation process called somatic hypermutation further introduces diversity and improves the affinity of the antibody.
2.9 Other Antibody-Based Affinity Scaffolds
While the full 150 kDa immunoglobulin G is an antibody in its natural form in humans, other smaller antibody based formats have also been created. These are known collectively as antibody fragments and some examples are illustrated in Figure 2.6.1 (74). Two popular formats are the Fab and single chain variable fragment (scFv) formats. In a Fab format, one VH-CH1 chain is paired with a VL-CL chain, essentially forming one binding arm of immunoglobulin G. The scFv is an even smaller format where only one VH domain is linked to one VL domain via a polypeptide (most commonly glycine-serine) linker. Both the Fab and scFv formats retain the specificity and monovalent antigen binding capability as the parent IgG. Alternative bispecific (or trispecific) versions of the Fab include bispecific Fab2 (or Fab3) where two (or three) different Fab domains, each recognizing a different antigen, have been fused together. This form allows a single Fab2 (or Fab3) molecule to bind two (or three) separate targets simultaneously. In a similar fashion, bispecific scFvs can either be two separate scFvs fused in a sequential manner (bis-scFv) or assembled with a very short linker forcing two scFvs to dimerize (diabody)(75). Attempts to make even smaller antibody based formats involve human single domain antibodies (sdAbs), which are essentially lone VH domains without the VL domain. Although the idea of human single domain antibodies may sound somewhat daunting, stable single-domain antibodies do naturally exist in camels and sharks(76), where the longer CDR3 loop of the VH protects the VH-VL interface.

2.10 Applications of Therapeutic Antibodies
The applications of therapeutic antibodies for treatment of various diseases such as autoimmune, inflammatory, cardiovascular, infectious and Alzheimer’s disease(77–80) are unfortunately too vast to be covered in this section. Instead, this section will first focus on providing some examples where therapeutic antibodies have been employed to treat patients with HER2 overexpressing breast cancer as well as patients with PNH and aHUS. The final section of this chapter will look at an emerging, and perhaps the latest trend in cancer treatment: cancer immunotherapy.

2.11 Eculizumab for Treatment of PNH and aHUS
The humanized monoclonal antibody eculizumab (Soliris) is a FDA approved therapeutic antibody for treatment of PNH(66, 81, 82) and aHUS(83, 84). Eculizumab targets the key C5 protein preventing the subsequent cleavage of C5 by C5 convertase(83, 85). This absence of C5 cleavage prevents the formation of the membrane attack complex (MAC) and generation of the anaphylatoxin C5a(81, 82). In contrast, the upstream generation of C3b is unaffected and C3b-mediated opsonization is not impaired.

Despite the approval of eculizumab, there have been reports on patients who either respond poorly or do not respond at all to the antibody(86–89). One of these reports concerns Japanese PNH patients with mutation in the C5 gene, making eculizumab unable to bind to the C5 protein(86). In Paper II we used an epitope mapping strategy based on scanning of surface-displayed protein domains in an attempt to understand the mechanism of this nonresponsiveness. The identified epitope residues of eculizumab on C5 were compared with the structural data reported by another group working in parallel(85).

2.12 Therapeutic Treatment of HER2 Positive Patients
There are several antibody cancer treatment strategies which focus on targeting the tumor cells themselves via receptor inhibition, induction of apoptosis signaling and antibody conjugated cytotoxic payload delivery (e.g. antibody drug conjugates and radioimmunotherapy)(70, 90). A similar strategy involves activation of effector functions once the antibody binds the tumor, resulting in antibody-dependent cell-mediated cytoxicity, complement-dependent cytoxicity, phagocytosis or CAR T-cell mediated therapy(70, 90,
The tumor associated antigens being targeted in these examples include haematopoietic differentiation antigens, glycoproteins, angiogenesis agents as well as growth and differentiation signaling agents(70, 90). Since HER2 is a well-studied receptor protein, several therapies are available for the treatment of HER2 positive metastatic breast cancer. Because antibody engineering is the focal point of this thesis, only antibody related therapies will be described more in detail below(92).

Herceptin (Trastuzumab) was the first anti-HER2 therapeutic agent approved in 1998 by the US Food and Drug Administration (FDA) for treatment of metastatic breast cancer patients who had HER2 overexpressing tumors. A few years later, in 2006, the FDA issued another approval, allowing Herceptin to be used in conjunction with chemotherapy for treatment of early stage disease. Herceptin is a humanized recombinant monoclonal antibody that binds to the extracellular domain IV of HER2(93), preventing the receptor from dimerization as a result. Other mechanisms that Herceptin uses to inhibit cell growth include: downregulation of the HER2 receptor by endocytic destruction, accumulation of the cyclin-dependent kinase inhibitor p27 and cell cycle arrest, induction of antibody-dependent cellular cytotoxicity in vivo and inhibition of constitutive metalloprotease mediated HER2 cleavage/shedding(94).

Although Herceptin was able to inhibit tumor growth when used alone(95), the addition of Herceptin to standard chemotherapy has significantly improved survival rates for patients compared to when either is used alone(96–99). Moreover, the combination of chemotherapy with Herceptin has also been shown to have a positive outcome leading to an increased overall survival in both early stage(100, 101) and advanced(96) breast cancer with HER2 overexpression. This does however not mean that single agent Herceptin treatment is obsolete. For example, single agent Herceptin can be used as maintenance therapy for patients who have had Herceptin / chemotherapy combination therapy(92). Due to the numerous clinical trials with positive outcome, many countries have therefore adopted Herceptin combination treatment as standard for patients with HER2 breast cancer regardless of the chemotherapy agent used and its efficacy(102).

Despite the success of Herceptin, there has also been an emergence of resistance to the antibody. Generally speaking, drug resistance can either be classified as de novo resistance or acquired resistance. De novo resistance refers to a lack of response to therapy in patients who have not received such previous therapy whereas acquired resistance refers to the development of drug resistance as the disease progresses after an initial clinical benefit. Approximately 30–50% of breast cancer patients who overexpress HER2 are either initially nonresponsive to Herceptin-based therapy even in conjunction with chemotherapy or eventually become resistant to Herceptin during treatment leading to short survival rates(96, 102–104). As a consequence, other therapeutic agents have emerged in an attempt to combat this issue.

One example is Laptatinib, an oral, small molecule tyrosine kinase inhibitor capable of both inhibiting HER2 and EGFR(105). Being a small synthetic molecule, Laptatinib inhibits HER2 by competing with ATP for the ATP binding domain in cytoplasmic region of the tyrosine kinase receptor. The anti-proliferative effects of the molecule in combination with Herceptin was demonstrated in a pre-clinical study(106). Response to laptatinib was also correlated with HER2 expression and its ability to inhibit HER2, Raf, AKT, and ERK phosphorylation(106). Another study demonstrated synergy and lack of cross-resistance between Laptatinib and Herceptin in human patients indicating that a combination of Laptatinib with Herceptin improved progression free survival versus Laptatinib alone while also offering a chemotherapy free treatment option(107).

Another strategy involves combining Herceptin with Pertuzumab (Perjeta), another humanized monoclonal antibody that binds to a different epitope (domain II) of the HER2
extracellular domain while also being able to stimulate ADCC(108). Pertuzumab is a dimer inhibitor and prevents HER2 from dimerizing with other members of the HER family (ErbB1, ErbB3, ErbB4), especially ErbB3(109, 110). The combination of pertuzumab and Herceptin provided a strongly enhanced anti-tumor effect and tumor regression in xenograft models, whereas the same effect could not be achieved when either agent was used alone. This enhanced efficacy could also be observed after tumor progression during Herceptin monotherapy(108). A phase II clinical trial in 2010 evaluated pertuzumab alone or pertuzumab in combination with trastuzumab in patients who had experienced progression during prior Herceptin / chemotherapy treatment. Although pertuzumab alone showed low antitumor activity, the combination therapy was active and well tolerated with mild/moderate adverse effects(111).

Two years later in June 2012, results from the Clinical Evaluation Of Pertuzumab And Trastuzumab (CLEOPATRA) phase III study led to approval by the FDA of combining pertuzumab with trastuzumab and docetaxel for first-line breast cancer treatment. In the CLEOPATRA trial, patients in the assessment group were given pertuzumab plus trastuzumab plus docetaxel whereas patients in the control group were administered placebo plus trastuzumab plus docetaxel. The results showed prolonged progression-free survival, and significant overall survival rate with no increase in cardiac toxic effects. In the addition, the pertuzumab/trastuzumab/docetaxel combination had no adverse impact on health-related quality-of-life and was able to prolong time to worsening of breast cancer-specific symptoms(104, 112–114).

A third strategy takes utilizes antibody-drug conjugate approach. Ado-trastuzumab emtansine (T-DM1) is an antibody drug conjugate composed of trastuzumab coupled to emtansine (also called DM1), a highly potent antimicrotubule cytotoxic agent. With an average of 3.5 DM1 molecules carried per molecule of trastuzumab(115), intracellular drug delivery specifically to HER2 overexpressing cells is possible. T-DM1 was approved by the FDA in 2013 for treatment of breast cancer that had progressed after trastuzumab/taxane or following early relapse on adjuvant trastuzumab based therapy. T-DM1 is the standard choice as second-line therapy for HER2 cancer following two phase III trials(92). In the first trial, patients given T-DM1 showed significantly prolonged progression-free and overall survival with less overall toxicity compared with patients given lapatinib plus capecitabine(116), whereas the second trial showed that T-DM1 provided a clinically significant benefit for patients who had previously received trastuzumab / lapatinib therapy and/or chemotherapy(117). Furthermore, attempts to make T-DM1 a first line treatment has also shown promising results. Using a combination of T-DM1 and pertuzumab in cultured tumor cells and mouse xenograft models led to synergistic inhibition of cell proliferation and induction of apoptotic cell death(118). Another phase II study showed that treatment with T-DM1 resulted in a significant improvement in progression free survival over trastuzumab plus docetaxel(119).

In summary, so far the FDA has approved four HER2-targeting agents (trastuzumab, pertuzumab, lapatinib and T-DM1) for patients with HER2 positive breast cancer(92). Combinations of these four agents together with chemotherapy are used as first, second and third line treatments. However, two other emerging anti-HER2 protein therapeutics are currently in clinical trials: MM302 and Margetuximab. MM-302 is an antibody–liposomal doxorubicin conjugate and is composed of doxorubicin encapsulated by a liposome of approximately 70-100 nm in diameter with anti-HER2 scFvs conjugated to its surface. This enables selective delivery of doxorubicin to tumor cells while minimizing exposure to healthy tissues(120). Due to the liposome, MM-302 is able to remain in the circulation for extended periods of time and can thus accumulate in tumors. Following attachment to HER2 overexpressing cells, receptor mediated endocytosis releases the doxorubicin inside the cell(120). A clinical phase II using MM-302 in conjunction with Herceptin is currently
ongoing(120). Margetuximab is a chimeric anti-HER2 monoclonal antibody that exhibits an elevated binding ability to both the lower and higher affinity forms of CD16A. CD16A is an important Fc-receptor involved in antibody dependent cell-mediated cytotoxicity (ADCC) against tumor cells. A recent phase I trial aimed to assess toxicity and anti-tumor properties in patients failed to respond to Herceptin/lapatinib showed that Margetuximab was well tolerated with promising single-agent activity(121).

2.13 Cancer Immunotherapy

While more conventional antibody treatment strategies target the cancer cells, as exemplified in Chapter 2.12, combating cancer with therapeutic antibodies can also focus on the immune system itself instead of the cancer cells. This concept is called cancer immunotherapy and involves blockade of immune checkpoints. Immune checkpoints refer to a collection of inhibitory pathways within the immune system which play a central role in the balance between activation and inhibition of antigen-specific and tumor-specific T cells(122). Such control is crucial for the immune system to maintain self-tolerance and avoid autoimmunity that could result from an uncontrolled T-cell response. However, it has become clear that tumors co-opt certain immune-checkpoint pathways as a major mechanism of immune resistance, especially against T-cells specific for tumor antigens(122).

One particularly important receptor is cytotoxic T lymphocyte-associated antigen 4 (CTLA4) expressed by activated T cells that results in the down regulation of the T cell response. CTLA4 and other negative regulators of T cell activation have a role in blunting the T cell response. In other words, CTLA4 acts like brakes on T-cells, preventing them from initiating immune attacks. Such blunting of the T cell response, as described previously, is important for avoiding autoimmunity but is very undesirable when the goal is to generate and maintain an effective antitumor T cell response(123). Therefore it was hypothesized that blocking the blocker (the CTLA4 molecule) would allow the immune system to attack tumors, and this has proven to be the case. One of the earlier milestones took place in 1996 when it was shown that anti-CTLA 4 antibodies erased tumors in mice(124). Since then, several mAbs that interfere with the inhibitory signals responsible for limiting T cell activation have been developed. These mAbs, called checkpoint blockade mAbs, can maintain T cell activation and enhance T cell-mediated lysis, and in the case of cancer, induce a more robust and sustained antitumor T cell response(122, 125).

One such mAb, Ipilimumab, that blocks CTLA4, has been approved by the FDA and has led to improvement in overall survival of patients with previously treated metastatic melanoma, earlier considered an elusive goal(126). Due to the positive outcome of Ipilimumab monotherapy, combination therapy using Ipilimumab with dacarbazine has also been conducted on patients with previously untreated metastatic melanoma with equally promising results(127).

Several mAbs that interfere with a different T cell regulatory pathway, the programmed cell death protein 1 (PD1) receptor–PD1 ligand 1 (PDL1) pathway, are also in the spotlight. Similar to CTLA4, a PD1 is also recognized as a brake on T-cells, but is instead expressed in dying T-cells. Monoclonal antibodies that target the PD1-PDL1 pathway include those that bind to the ligand and others that bind to the receptor(128, 129). Both approaches have been promising in early clinical trials and the FDA approved two anti-PD1 mAbs (pembrolizumab and nivolumab) for treatment of melanoma. Although, both anti CTLA4 and anti PD1 targeting the immune system, their pathways are different. Whereas CTLA4 regulates de novo immune responses, PD1 has a greater influence on ongoing T-cell immune responses(130, 131). Combination trials using co-inhibition of CTLA4 and PD1 sound attractive, considering, although blocking both pathways raises concerns about potential autoimmunity.
3. The Antibody – Antigen Interface

3.1 The Epitope
Antibodies recognize specific immunogenic markers, called epitopes, on their target antigen. The corresponding interacting surface on the antibody is called the paratope and consists of different amino acids in the CDRs. Epitopes are usually classified as continuous or discontinuous, depending on whether the amino acids that form the epitope are contiguous or not. It is generally acknowledged that most protein epitopes are discontinuous and consist of two to five short stretches of residues that are distant in the protein sequence but are brought together into a single unity by the folding of the protein chain(132, 133). Discontinuous epitopes are structurally defined by the amino acids in contact with the paratope residues in a crystallographic complex. In contrast, continuous epitopes are identified by the ability of short peptide fragments of the protein to bind to an antibody. The distinction between these two epitope classes is however not black and white. A discontinuous epitope can consist of several continuous stretches. In other cases, not all residues in a continuous epitope will interact with a paratope, with only a subset of residues present at the surface of a native protein where they are usually a part of a more complex discontinuous epitope(134). Additional types of protein epitopes have also been suggested. One example is neotopes which are formed when several subunits are brought together and recognized by the antibody as a single antigenic entity(135). Neotopes, therefore arise from the quaternary protein structure. Another example is mimotopes which are peptides that bind to antibodies but share little in common with the protein whose epitope the mimotope “minics”(136).

3.2 The Paratope
As of today, there are several methods and approaches available to identify and number the antigen binding regions (CDR regions)(137). The oldest scheme (Kabat numbering scheme) relied solely on alignment of antibody sequences and determining the positions with the highest sequence variability(71) which was the most viable approach before structural data was present. A few years later as more 3D structures became available, a second numbering scheme (Chothia numbering scheme) emerged which is based on location of the hypervariable loop and constant framework regions in the antibody structure. Interestingly, despite the high sequence diversity, the CDR loops could only adopt a restricted set of conformations termed canonical structures(138). The way different canonical structures combine alters the shape, or topography of the antigen binding site which in turn determines the size/type of the antigen with which the antibody is able to interact(139–142). Yet, a third way to define CDRs (IMGT) uses the database of germline variable immunoglobulin sequences(143, 144), and as expected, all three different approaches produce slightly different residue numbering systems.

A more recent fourth CDR identification/numbering scheme called the Paratome was introduced in 2012. This scheme took advantage of the large number of antibody-antigen complex structures which had been made available in recent years. The Paratome method is based on large-scale protein structure complex analysis by utilizing all antibody-protein complexes in the PDB database. A multiple structural alignment of all these antibody-antigen complexes was conducted to identify the structural regions in which antigen binding actually occurs(145, 146). The alignment revealed structural consensus regions where the pattern of structural positions that bind the antigen is highly similar among all antibodies. These regions were called antigen binding regions (ABRs) and corresponded roughly to CDRs. Compared with older CDR identification tools (Kabat, Chothia and IMGT) which may miss 20% of the Ag binding residues, ABRs managed to cover much as 96% of all the residues that actually bind the antigen(137, 145, 146). Paratome also found that several residues in the FR region and constant regions of IgGs also contributed significantly to antigen binding.
While CDRs are often used to identify paratopes, not all residues contained in the CDRs are found to bind the antigen. According earlier studies, only about 20–30% of the residues within the CDRs actually form contacts with the antigen(147) with the contacting residues more commonly located at the central regions of the binding site. The non-contacting residues within the CDRs, in contrast are important for maintaining the structural conformations of the hypervariable loops(139). Furthermore, Padlan and co-workers(148) found that the residues directly involved in the antigen interaction are usually also the most variable ones. They therefore suggested that the residues in direct contact with the antigen should be called Specificity Determining Residues (SDRs).

3.3 The Epitope – Paratope Interface

The interacting interface between the antigen binding site (epitope) and the antibody binding site (paratope) consists of several non-covalent chemical interactions including hydrogen bonds, van der Waals forces, electrostatic interactions and hydrophobic interactions. These non-covalent interactions determine the binding strength (also called affinity) between the antibody and antigen. The concept of affinity is not only applied to antibody-antigen interactions, but also to any two interacting proteins in general. The driving force behind complex formation between an antigen and antibody is the change in Gibbs free energy ($\Delta G=\Delta H-T\Delta S$). The changes in the enthalpy component ($\Delta H$) are the results of hydrogen bonds, electrostatic and van der Waals interactions whereas changes in the entropy component ($\Delta S$) are associated with changes in conformational freedom upon binding.

The affinity between an antibody and antigen can be expressed as an association equilibrium constant ($K_A$, unit $M^{-1}$) between the concentration of the antibody-antigen complex (Ab-Ag) and the concentrations of free antibody/antigen at equilibrium ($K_A = [\text{Ab-Ag}] / ([\text{Ab}][\text{Ag}])$). The affinity can also be defined as a change in Gibbs free energy ($\Delta G=RT\ln K_A$), where $R$ is the gas constant and $T$ is the temperature. However, a more commonly used term to describe affinity is dissociation constant (or $K_D$, unit M), which is essentially the inverse of the association constant ($K_D=1/ K_A$). Therefore, a low $K_D$ value is regarded as a stronger affinity, although no kinetic information is provided about the formation of the complex. Fortunately, the dissociation equilibrium constant can also be defined using rate constants ($K_D=k_{\text{off}}/k_{\text{on}}$), where $k_{\text{on}}$ ($M^{-1}S^{-1}$) is the rate of formation of the complex (also called on rate) and $k_{\text{off}}$ ($S^{-1}$) is the rate of dissociation of the complex (also called off rate). This means that two antibody-antigen pairs can have the same affinity (same $K_D$) but with different kinetics. For example, an antibody with low on-rate but also a low off-rate will have the same affinity towards its antigen as another antibody with a high on-rate but also a high off-rate. Several instrumental approaches can be used to measure the affinity and kinetics of antibody-antigen formation, including surface plasmon resonance (SPR) and biolayer interferometry (BLI).

Apart from affinity, another important concept is antibody specificity, which refers to the ability for the antibody to distinguish an antigen from another. However, specificity is not an absolute binary phenomenon but instead can be regarded as varying degrees of “goodness of fit” between the antibody and antigen(149). Antibody poly-specificity, on the other hand, is the ability for an antibody to recognize several different epitopes present on different antigens whereas antibody hetero-specificity refers to the phenomenon where an antibody has better recognition towards a different antigen other than the one used to obtain the antibody(149). When discussing the affinity and specificity between the antibody and its antigen, the geometric and chemical complementary between the epitope and paratope need to be considered(137, 149).
3.3.1 Geometric Complementarity

From a protein engineering point of view, it is important to be able to understand the length and amino acid composition preferences found in antibody CDR domains as well as any differences in these parameters between different targets. As most antibody display platforms have an upper limit in the library size that can be created, a well constructed library is beneficial. Several studies have analyzed the length and composition of matured antibodies as well as the specific CDR residues involved in binding (71, 140, 142, 150–153).

For example, in 2003, Collis et al correlated CDR length with different types of antigens bound and reported that there were some differences in the distribution of CDR lengths in antibodies binding to different classes antigens. The most common lengths in human antibodies were also reported (L1: 11; L2: 7; L3: 9; H1: 5; H2: 17; H3: 10-16) (152). Another example is a study conducted in 2011 by North et. al, who provided the following most common CDR lengths: L1: 11; L2: 8; L3: 9; H1: 13; H2: 10; H3: 7-16 (150). A third example of a more recent analysis examined the amino acid composition of the 6 ABRs in 200 antibody-antigen complexes using the Paratome method (153). The analysis reported the median lengths of the six ABRs (L1: 8, L2: 11, L3: 8, H1: 9, H2: 14, H3: 11) while also revealing that H1, L2, and L3 had limited length diversity. In addition, the number of SDRs in each ABR was also determined and what is noteworthy is that there is a large diversity in the number of SDRs in each ABR. For all ABRs, there are instances where they do not contain a single SDR, but at the same time there are instances where there are 8-10 or more SDRs in each ABR (153).

Unfortunately, the data obtained by the different studies provided some slightly different results with regards to length diversity. For instance, H3 was suggested as the longest CDR by North et. al whereas Collis et al claimed that H2 might be as long or even longer than H3 (150, 152). However, these minor differences can be due to the fact that different antibody sets as well as different numbering schemes were used to define the CDRs (137). Regardless of the inconsistencies between the different papers, some trends can be seen. One of the most obvious trend is that CDRH3 had the highest length diversity / distribution. Whereas the other CDR loops each had a one over-represented length, numerous antibodies had various lengths (7 to 18 residues long) in the CDRH3 loop. In addition, the largest spread in CDR length is also found in CDRH3. This means two things. Firstly, for instance, while CDRL3 rarely has a length below 6 residues and above 13 residues, CDRH3 can have lengths as low as 3 and as high as 25. Secondly, CDRH3 also shows a more considerable variation between different classes (e.g. peptides & proteins) of antigen binders compared to other CDR loops. This is not surprising in any way since the generated diversity in vivo is distributed differently in the antigen binding site due to V-D-J recombination, as described in Chapter 2. The diversity of the first and second CDR regions of the heavy and light chains are restricted by the number of V genes available. On the contrary, the CDR3 regions are the results of V-(D)-J recombination, making these loops much more diverse with regards to both sequence and length. Since H3 has an extra recombination (V-D-J) compared to L3 (V-J), H3 is the most variable loop in a B cell produced antibody and it occupies a central location in the binding site and provides many important antigen contact residues.

Apart from analyzing the different lengths of the CDR loop, it has also been demonstrated that the number of residues the antibody uses to interact with the antigen differs in antibodies recognizing antigens of different type and size (139, 140, 151). Generally speaking, antibodies use about 5 CDR residues fewer when interacting with haptens compared to peptides and proteins. Overall, all these studies have led to the consensus that three main topographies exist. These three main topographies are cavities, groove and planar, each correlating with a different type of antigen. Deep pockets are often associated with anti-hapten antibodies, grooves are associated with peptide binders, while protein binding antibodies have the largest
and relatively flat combining sites (139, 154). In addition, the six CDR loops play an important role in determining the shape of the binding site. For example, long L1, L2, H1 and H2 with short L3 and H3 would result in a groove. A long L1 and H1 with a medium-to-long H2 and short L3 form a pocket. A flat binding site can be obtained by filling the central part of the binding site using L3 and H3, while a very long H3 can lead to a protruding binding surface. As of CDR usage, anti-protein antibodies have high CDR usage located on the edge of the surface whereas anti-hapten antibodies have the lowest usage, with hot spots buried in the VL:VH interface (140). The CDR usage of anti-peptide antibodies is in between the anti-protein and anti-hapten antibodies, with hot spots located in the interior of the antigen-binding site but not as deeply buried as in anti-hapten antibodies (140).

3.3.2 Chemical Complementarity

Apart from geometric complementarity, the amino acids in the epitope-paratope interface need to provide the necessary chemical complementarity. Although the CDRs are also known as hypervariable regions, their sequences are not completely random. Several studies (in addition to / including those studies mentioned in the section above) have shown that the amino acid composition of CDR loops differs from antibody framework regions and from generic protein loops (140, 151–153, 155–158).

Tyrosine is the most abundant amino acid in antigen binding sites, accounting for about 10% of the total CDR composition and 25% of the antigen contacts in functional antibodies (21, 155). The tyrosine bias is even greater when naïve CDRH3 loops are examined. There, about 40% of the sequence is predicted to be tyrosine and 30% is predicted to be small amino acids (serine, glycine, alanine and threonine) (159). In addition to tyrosine, tryptophan, which has similar properties as tyrosine is also overrepresented in CDR loops compared to generic loops. Phenylalanine can be considered as a neutral amino acid, as it is claimed to be favored by some (157, 158) whereas others suggest that phenylalanine is underrepresented (151).

The reason behind aromatic side chain preference is simple: these side chains can interact with a diverse group of targets through a cumulative collection of relatively weak noncovalent interactions. For instance, it has been reported that aromatic side chains interact with other aromatic side chains via face-to-edge or parallel π-stacking. They can also interact with positively charged side chains through cation–π interaction and with backbone/sidechain hydrogen bond donors by having hydrogen bonding to aromatic π-systems. Finally, aromatic side chains can also interact with alkyl carbons through the C–H–π interaction, and with negative charged side chains through anion–π interactions (160). Although each individual interaction is weak, the combined effect is considerable. Tyrosine possesses hydrophilic character in addition to these aromatic and hydrophobic traits, and it becomes clear why tyrosine is able to form hydrogen bonds, hydrophobic and van der Waals interactions with a wide array of antigens.

As of the remaining seventeen amino acids, there are again some differences between analyses since many of these amino acids do not exhibit a strong bias. Some authors claim that hydrophilic residues predominate over hydrophobic ones (151) whereas others do not make this statement (152). However, cysteine, proline and glutamic acid are often all considered detrimental. On the other hand, asparagine, arginine and small (polar) amino acids (serine, threonine, glycine) that allow conformational freedom to accommodate the large side chains of tyrosine/tryptophan are usually favored. The rest of the amino acids (mostly hydrophobic) can be considered neutral, since their preference in the antibody CDR loop varies between slightly favored and slightly disfavored (152, 153, 158). The charged amino acid aspartate is considered detrimental in one study (152) which determined the frequency of SDRs averaged over an entire paratope (i.e. six CDRs). However in another study that determined the frequency of SDRs in each ABR, aspartate was one of the abundant amino acids together with
tyrosine, asparagine, arginine, and tryptophan, constituting between 48% and 63% of the residue composition(153). Again, different sets of analyzed antibodies binding to different targets can be said to account for these contradictions. There are also minor exceptions where certain amino acids are over- or underrepresented only in individual classes of antibodies. For instance, threonine is under-represented only in protein binders while the very strong over-representation of arginine in nucleotide binders is a reflection of interactions with the negative charge of the phosphate backbone(152).

Regarding energy contributions, no link was found between the abundance of an amino acid in SDRs and its energetic contribution to antigen binding. In the paper published by Kunik and Ofran, the most energetically important residues in the ABRs were Tyr, Asp, Asn, and Arg with H3 having the highest percentage (29%) of SDRs that are energetically important for binding followed by L1 (24%) and H2 (22%)(153).

3.3.3 Epitopes and Paratopes are not Intrinsic Properties

A short introduction to epitopes was given at the beginning of this chapter. Now that it has been established that CDR loops of antibodies differ significantly from generic protein loops, one would be tempted to believe that epitopes also have characteristics that distinguish them from the general protein surface. This is however, not the case. While some papers suggest that the amino acid composition of epitopes is not distinguishable from that of protein surfaces(157), others imply that epitopes may be enriched with certain of amino acids(137, 149, 156, 161, 162).

In a way, both sides are correct(137, 149) as a relatively recent study revealed that the amino acid composition in each of the six CDRs differed significantly from the other CDRs and that each CDR favors binding to a different set of protein surface amino acids. Because the six CDRs have different amino acids compositions, they also have different contact preferences regarding epitope residues. However, when the amino acid composition of all combined epitopes was compared with that of entire protein surfaces, no noticeable differences were observed(153). In other words, although each CDR has its own set of contact preferences, the total effect of all individual CDR preferences resulted in an overall epitope amino acid composition that is very similar to the composition of generic protein surfaces(137, 153, 157).

Because the entire accessible surface of a protein is a continuum of potential epitopes(163), this “distributed specialization” of CDR loops is what allows antibodies to bind any protein surface. This makes sense, as the immune system has to fulfill the “time is of the essence” requirement. Whereas other protein-protein interactions have the luxury of fine-tuning their interface throughout the course of evolution, the immune system must generate antibodies swiftly in response to an immunogen. Therefore, epitopes and paratopes are not intrinsic properties of antigens or antibodies but depend on each other to acquire a recognizable identity as they cannot exist in the absence of the other(137, 149). This means that the number of epitopes in a protein is equal to the number of different antibodies that can be raised against it assuming they bind different epitopes. But how does one know if these different antibodies actually bind different epitopes?

3.4 Mapping of the Epitope

Epitope mapping is a term used to describe techniques employed to identify and locate the epitope on the protein surface. Mapping the interactions between an antibody and antigen has several benefits, allowing us to understand immune responses and autoimmunity as well as aiding vaccine design and unravel the mechanism of action of an antibody(164). The experimental methods developed to map epitopes can roughly be divided into structural and functional. Structural methods interpret the protein structure residues in direct contact with the antibody but may not always reveal contribution of amino acids in binding strength.
Functional methods use assays to identify and characterize residues important for binding. As each method has its own advantage and disadvantage, a combination of different methods almost always is preferred.

3.4 Structural Mapping

The most accurate and reliable method, and hence for many the favorite weapon of choice, for identifying epitopes and paratopes is via X-ray crystallography and determining the interacting amino acids in the two partners(165). Often regarded as the only method to define a structural epitope(166), very accurate key contacts between the epitope and paratope side chains can be extracted from the crystal complex. As of interpretation, amino acids within 4 Å of each other are generally considered to be contacting residues(164). This technique not only provides precise identification of both continuous and discontinuous epitopes, but also information about binding strength(167). However this technique is also dependent on the availability of the antigen and antibody with a certain degree of purity. Obtaining the pure antibody is usually relatively straightforward, whereas the antigen can be more problematic or even very difficult in the case of membrane proteins(164). The crystallization process starts with pure antibody and antigen, followed by formation of the complex and frequent purification of this complex by size exclusion chromatography. The next few steps involve successive rounds of crystallization to obtain diffraction-quality crystals(164). Therefore, the main disadvantages with crystallography include high cost and requirement of diffraction quality crystals, in the absence of which, can make certain antibody-antigen complexes impossible.

Nuclear magnetic resonance (NMR) spectroscopy is another technology performed in solution which provides fine epitope data. In this technique, the protons of the complex are subjected to a magnetic field and pulsed electromagnetic radiation. Not only can linear and conformational epitopes be resolved, but NMR can also provide information about the dynamics, structure and nature of epitope recognition(168). However, NMR is limited in molecular size and only small proteins and peptides (<25-30 kDa) are suitable for analysis(169).

3.4.2 Functional Mapping

Functional methods include techniques based on detection of antibody binding to synthetic peptides (peptide scanning), antigen fragments or recombinant antigens (e.g. antigen mutants, or antigens expressed using selection systems such as phage display) and alternation of the target protein to attribute a change in binding to a specific residue.

In peptide scanning, synthetic overlapping peptides covering the targeted antigen sequence are immobilized on a solid support and binding of an antibody is detected using western blot, dot blot, or an ELISA format(170, 171). The peptides can be synthesized on pins (PEPSCAN), on a cellulose membrane support (SPOT method)(172), or on peptide microarrays(173, 174). The length of the synthesized peptides have been suggested to range between 6 and 18 amino acids(175) and the larger overlap between peptides, the higher resolution of the determined epitopes. This approach does not require expensive specialist equipment and is able to quantify the immune response towards a specific epitope. But this method is unlikely able to identify complex conformational epitopes involving tertiary and/or quaternary structure. Although it is possible to mimic discontinuous and conformational epitopes by constraining peptides via disulphide bridges(176), peptide scanning is nonetheless most ideal for mapping of linear and relatively simple conformational epitopes.

Mutagenesis is another relatively cheap approach to determine epitopes without the need for high level expertise or equipment. This method is based on the creation and screening of numerous mutant versions of the target protein. The main rationale behind mutagenesis
based mapping is the fact that substitution of individual residue(s) part of a functional epitope causes loss of antibody binding. Another advantage with mutagenesis is that it is possible to screen numerous (hundreds to thousands or even protein libraries) using alanine scanning or shotgun mutagenesis quickly and can be used to map the epitopes of difficult membrane proteins(177). Several approaches can either be used alone or in combination to create these mutants in order to find energetically favorable residues. Random or site-directed mutagenesis refer to either complete random mutation of the amino acid positions in the target region or specific positions are selected for substitution. Substitution can either be to alanine or another of the remaining 19 amino acids to investigate changes in binding affinity(178). It is important to point out that it can be difficult to know whether the mutation will cause a disruption in protein folding. With the availability of protein structure data, mutagenesis can be optimized to only allow mutations near the surface of the protein, in order not to disrupt the hydrophobic core(179–181).

Combinatorial libraries and display technologies are versatile platforms that are also popular for epitope mapping in addition to selecting proteins with novel properties due to their relative cheapness and quickness(182). Display technologies can be combined with mutagenesis or peptide scanning by expressing peptide or protein/protein fragment libraries on the surface of different hosts (e.g. phage, bacteria, yeast or mammalian) and testing the binding capacity to the antibody of interest through affinity selection. Therefore the costs of peptide synthesis and production/purification of protein mutants/libraries can be circumvented. Following selection and screening, the sequences of selected binders or non-binders can be analyzed and an epitope or relevant residues for binding can be identified. The peptide libraries used in display systems can either be random peptide libraries or antigen-derived libraries, fragmented or randomized by mutation(183). A comparison between random peptide libraries and antigen fragment libraries showed that latter had higher success rate of identifying epitopes(184).

3.4.3 Computational Epitope Prediction
Accurate computational tools for prediction of epitopes have long been highly desired in peptide vaccine design to overcome the experimental determination of epitopes. Despite the rapid current advancements in computation and artificial intelligence, as demonstrated by the recent defeat of multi world champion Lee Sedol to AlphaGo, many B-cell epitope prediction methods are still not accurate enough(185, 186). The explanation, is that current methods attempt to identify epitope residues based on the characteristics associated with residues that bind the antibody(187–193). As described earlier, the differences between epitopes and generic protein surface residues are not substantial and the amino-acid composition of epitopes is indistinguishable from that of surface-exposed non-epitope residues. Therefore it has been suggested that epitope prediction should be done for each antibody instead by taking both the epitope and paratope into account(137) as most of the antigen surface may become a part of an epitope under certain circumstances(194–196).

3.4.4 Applications of Epitope Information
Once the epitope / paratope interaction has been identified, such information can be used to characterize autoantibodies and provide insights into autoimmune diseases(197). Combination of display technologies and epitope mapping have also been utilized to study the natural humoral immune response and identify, as an example, crucial conserved epitopes among various HIV strains(198). Another area where epitope information has been extensively used is in the field of vaccine design. Instead of using conventional attenuated or inactivated pathogens, epitope mapping can be used to build directed immunogens(199). In one example, epitope mapping of MntC (MntC is considered an attractive candidate for S. aureus vaccine) revealed three linear stretches considered to be immunodominant epitopes when the crystal structure was studied. A designed vaccine based these linear B-cell epitopes
was successful in eliciting an effective immune response\(^{(200)}\). Epitope mapping has also been used to gain knowledge of an antibody’s mechanism of action. In one example, the 3D structure of human ephrin type A receptor 2 (EphA2) bound to an agonistic human antibody was solved. EphA2 is involved in formation and progression of several cancers and a structural and functional mapping of an EphA2 antibody showed that CDRH3 mimics the ligand, leading to tumor inhibition by inducing receptor internalization\(^{(179)}\). In another example the mechanism of Lebrikizumab towards IL13 was revealed. The cytokine IL13 is involved in the pathogenesis of allergic diseases by binding to a heterodimeric receptor consisting of IL-13Ra1 and IL-4Ra. X-ray crystallography showed that Lebrikizumab was able to inhibit IL13 signaling by binding with a very high affinity to IL-13 at an epitope that overlaps strongly with the binding site of IL-4Ra. This in turn prevented IL13 from interacting with IL-4Ra and an additional site directed mutagenesis allowed the identification of key residues in the CDRs of Lebrikizumab responsible for the strong affinity towards IL13\(^{(201)}\).

In addition to vaccine design, unraveling of mechanism of action and understanding autoimmune diseases, epitope information can also be used to aid both antibody (library) design and assessment (described more in detail in Chapter 4).
4. Generation and Engineering of Human Antibodies

The previous chapters have provided information about the structure, recognition and therapeutic uses of antibodies / antibody fragments. This chapter will describe different approaches to obtain such antibodies as well as how these antibodies can be improved even further.

As described earlier in Chapter 2, the hybridoma technology has been a revolutionary technique for generating renewable monoclonal antibodies. Nonetheless, using animal-derived antibodies in humans, whether it is pAbs or mAbs, ultimately has issues with immunogenicity. In addition, murine hybridomas can be labor intensive, low yielding or genetically unstable(202). Therefore, several techniques have emerged over the years to generate human-like antibodies derived from mouse (humanization) or fully human antibodies altogether. This is especially important for therapeutic use as mAbs have become one the most powerful therapeutic and diagnostic tools in modern medicine with some estimates suggesting that worldwide mAbs market can reach $125 billion in the next few years(203).

4.1 Humanization

Humanization began with chimerization where murine variable regions were combined with human constant regions to generate antibodies with 70% human content(204). This reduced the human-anti-mouse-antibody (HAMA) response but could still result in a human anti-chimeric antibody (HACA) response(205). In the years following chimerization, several other humanization approaches have emerged, falling into two main categories: rational and empirical. Rational methods rely on iterative cycles, while empirical methods rely on selection of large combinatorial libraries by surface display or high throughput screening.

4.1.1 Rational Methods for Humanization

A popular method to humanize murine antibodies is CDR-grafting, envisioned by Winter’s group in 1986(206). In short, CDR grafting involves replacing the CDR regions from a human antibody with murine CDRs(207, 208) and consists of three major steps: 1) defining target residues for grafting; 2) source of human framework regions; 3) back mutation of residues outside of the grafted region to restore affinity. This process (CDR grafting) often reduces the mAb affinity, and hence, it is usually followed by affinity maturation.

The most contributing target residues can be identified via the antibody-antigen complex structure complemented with alanine scanning mutagenesis(209) and/or combinatorial mutagenesis(210). In the absence of a structure, CDRs grafting(208, 211, 212) or SDR grafting may be employed(213). The source of human framework can either be chosen based on known structure (fixed FR)(206, 208), sequence homology (best fit)(214) , or a combination of both fixed FR and best fit(215, 216) as well as generating consensus sequences to select human FRs as template(93, 217). Back mutation of key residues on the human FR back to murine is necessary since a decrease in affinity after grafting is common, and which these key residues are depends on each case(211, 212).

In addition to CDR grafting, other rational humanization approaches include resurfacing and superhumanization. Resurfacing is similar to CDR-grafting, but only changes the surface exposed residues to human(218–220). Superhumanization relies on selecting human germline sequences with the highest CDR homology as the FR donor from a subset of human germline sequences with similar canonical structures as the murine antibody(221–223).

4.1.2 Empirical Methods for Humanization

One example of an empirical method is creating a library by introducing mutations at specific
positions in the human FR region. The library is then panned to select the FR construct that best supports the murine CDR(224). Another example is guided selection which consists of combining for example the VH of a non-human antibody specific for a particular antigen with the VL library of a human antibody. Following selection, the human VL is now specific for the target protein and is then combined with a human VH library and selected again. This strategy led to the first human antibody approved by the FDA: Adalimumab (Humira)(225, 226). However, a potential disadvantage of this approach is that shuffling of one antibody chain while keeping the other constant can result in epitope drift(227, 228). In order to maintain the epitope recognized by the non-human antibody, CDR retention can be applied(229), most commonly to the non-human CDR-H3, but also to both CDR-H3 and CDR-L3 (230) as well as CDR-H3, CDRL3 and CDR-L2 (231).

In summary, humanization has proven to be a successful method for antibody discovery but re-engineering and re-validation may be necessary since the humanization process can lead to detrimental effects, including loss of affinity(222), changes in specificity(232) and reductions in solubility and/or expression in production systems(233). Fortunately, other available techniques (presented below) can complement the shortcomings of humanization.

4.2 Transgenic Mice

In the transgenic mice approach, human immunoglobulin loci are introduced in mice where rearrangement and somatic hypermutation can occur within the animal's own immune system. While the first fully human FDA approved monoclonal antibody Adalimumab was selected using phage display, transgenic mice led to the approval of the second fully human monoclonal antibody Anitumumab in 2006. This was possible since human and mouse immunoglobulin loci work in very similar ways(234). The groundwork for commercial transgenic mice strains was laid in 1989 when Bruggemann et. al. generated mice with an unrearranged human minilocus consisting of 2 VH, 4 D, 6 JH, and CL genes. These mice were able to rearrange the antibody genes to produce IgM antibodies with human heavy chains, albeit at low levels(234).

A major challenge for transgenic technology is integrating the large size of the human Ig loci. Advances in genomic technology e.g. yeast artificial chromosome (YAC), bacterial artificial chromosome (BAC), oocytes microinjection and embryonic stem cells have enabled the cloning, modification and integration of large DNA regions into the mouse genome(235–238). Together with advances in knock-out mouse technology with where the endogenous mouse IgH and IgK loci have been inactivated(239–242), the stage was set for the entrance of human monoclonal antibodies from transgenic mice.

In 1994, two biotechnology companies, GenPharm International (later Medarex, now part of Bristol-Meyers Squibb) and Cell Genesys Corporation (later Abgenix, now part of Amgen) working in parallel, published separate papers where gene knockout technology was used to inactivate both the murine heavy and kappa light chain loci and introducing fragments of the human heavy and light chain loci. From Cell Genesys, a heavy chain minilocus was introduced which comprised of 4 VH, 25 D, 6 JH, Cµ, and Cδ genes together with a kappa light chain minilocus comprised of 2 Vκ, 5 Jκ, and Cκ genes(243). From GenPharm International, a heavy chain minilocus consisting of 4 VH, 15 D, 6 JH, Cµ, and Cγ1 was introduced alongside a kappa light chain minilocus consisting of 4 Vκ, 5 Jκ, and Cκ genes(244). Immunization of both mice strains gave similar results in that both the heavy and light chain loci underwent gene rearrangement yielding fully human antigen-specific antibodies. These two independently developed transgenic mice strains later became branded as HuMAb-Mouse (GenPharm Inc) and XenoMouse mice (Cell Genesys, Inc.) respectively.
Further improvements to XenoMouse included increased diversity of the variable repertoire (e.g. 42 VH genes and 34 VK genes together with the complete complement of D and J genes)(245), and addition of the complete human Igλ locus yielding picomolar / sub-picomolar affinities IgGλ mAbs(235). Similarly, improvements to the HuMAB strain included additional expanded diversity to the VK and VH gene repertoire(246) and development of new HuMAB strains with fully human VH transgenes and fully-mouse Cμ, Cdelta and Cgamma genes. Another transgenic mouse strain is the TC Mouse, engineered by Kirin Brewery Company, Ltd., (now Kyowa Hakko Kirin Co., Ltd. following the merger of Kyowa Hakko Kogyo Co., LTD. with Kirin Pharma Company, Limited) which used separate human transchromosomes carrying the complete human IgH and IgK loci bred into a mouse with inactivated endogenous IgH and IgK loci. While it sounds attractive to have a complete human IgH and IgK loci (and thus a full set of human IgHK antibodies), both human transchromosome fragments underwent mitotic and meiotic loss, diminishing the efficiency of antibody recovery(247).

Ironically, despite all the effort invested into transgenic mice, a paradigm shift was taking place. The XenoMouse and HuMAb-Mouse technologies were originally engineered with one huge benefit in mind: that hybridomas could be used directly from these mice to produce fully human mAbs before clinical development. However, transgenic mice do not produce human mAbs as efficiently as regular wild type mice(248, 249) due to the incompatibility between human Ig loci and other components of the immune system, which are essentially murine. Furthermore, studies have shown that chimeric constructs with human VDJ linked to rodent CH are most efficient(249, 250). The reason was that endogenous rodent CH genes enable optimal physiological interactions necessary in the regulation of immune responses by the various host Fc receptors(235, 251). Since human VH genes can be efficiently rearranged in transgenic mice, their linkage to endogenous CH genes provides a flawless interplay with the cellular signaling machinery(235). In other words, transgenic mice only needed to produce chimeric mAbs instead of fully human mAbs and production of recombinant antibodies in cell lines prior to clinical development would be more advantageous. Since only the DNA sequences encoding an antibody’s fully human variable regions was required, transgenic mice only needed to generate diverse high quality fully human variable regions instead of producing complete fully-human mAbs. The DNA encoding these fully human variable regions could then be attached to any given human constant region sequence of choice.

Based on this paradigm change, the company Regeneron created the VelocImmune mouse platform (using BAC-based Velocigene technology) for the generation of therapeutic mAbs(252). In contrast to HuMAB and XenoMouse, the VelocImmune mouse technology produces chimeric antibodies that have human variable regions attached to fully mouse constant regions. Further expansion of the VelocImmune Mouse platform added human Vλ repertoires, either as chimeric Igλ transgenes (human Vλ-mouse Cl) or hybrid Igλ/K transgenes (human Vλ-mouse CK)(253). Nonetheless, even these chimeric transgenic mice suffer problems, such as loss of mAb potency when converted to fully human composition by changing into human constant regions. This is because the constant domain of the heavy can influence the structure and activity of its respective variable region(254–256). Furthermore, there is also a risk that exchanging mouse and human constant regions alters binding affinity and thermodynamics(232). A solution is to engineer a mouse to make antibodies comprising fully human heavy chain variable regions-CH1-upper hinge combined with mouse middle hinge and Fc regions paired with fully human light chains(253).

In summary, obtaining fully human antibodies directly from transgenic mice immunizations is beautiful in concept but time consuming in practice. Since recloning of chimeric mAbs for production in mammalian cell lines is regarded ironically as the better option, the engineered mice strains have somehow lost their original intended purpose. Moreover, the fact that many
transgenic strains are properties of corporations also means that licensing is limited making these platforms unavailable to the academic world.

4.3 Display Technologies and Library Construction

When using the transgenic mice (or any immunization based) technology, there are two issues of concern that need to be mentioned: a hypothetical “affinity ceiling” and immune tolerance.

It has been hypothesized that there should be an “affinity ceiling” of approximately $10^{-10}$ M (100 pM) in mice, beyond which there can be no driving force due to the on rate being limited by the diffusion coefficient and the off rate being limited by the rate of internalization of surface immunoglobulin (257–259). However this does not seem to apply in a strict manner with transgenic mice since antibodies with affinities beyond this theoretical limit have been isolated against human antigens (253). As for immune tolerance, there have been arguments that immune tolerance might prevent the mouse from reacting to human proteins highly homologous to mouse. But proponents of transgenic mice claim that the mice are capable of breaking this immune tolerance since the amino acid sequence of most human antigens is sufficiently different from their mouse orthologues (253). Proponents also point out that reports in the patent literature show no correlation between sequence identity and recovery of high affinity antibodies from transgenic mice (253).

Nonetheless, it cannot be denied that deletion of “self antigen” recognizing B-cells in the immune system removes an important portion of the antibody repertoire when isolation of self-reactive antibodies is desired. With display technologies that circumvent immune tolerance, antibodies that recognize “self antigens” can be preserved. Also considering the fact that far more variants can be screened using less time, display technologies have become formidable tools in protein engineering (a few display platforms are shown in Figure 4.3.1). Since there is no reliance on the immune system, various sources of immunoglobulin genetic material are available when using display technologies in conjunction with de novo selection. The genetic material can derive form natural, synthetic or semi-synthetic sources. Natural sources include immunized animals (260) or infected humans (261) (immunized libraries). The immunoglobulin genes can also derive from non-immunized sources and these are used to construct natural naïve libraries. Naïve libraries can also be made by semi-synthetic or fully synthetic approaches and are therefore more appropriate for selection towards a wider array of antigens.

Natural libraries (naïve or immune) are constructed with VH and VL genes cloned from B-cells using PCR (262, 263) or other PCR based methods, such as oligonucleotide-assisted cleavage and ligation (OACL), anchor PCR, PCR adaptors and rapid amplification of cDNA ends (RACE) (264–267). The B-cells can originate from peripheral blood lymphocytes, bone marrow, lymph nodes or the spleen. The desired natural VH-VL pairing can also be maintained by parallel amplification of the VH and VL genes from single B-cells with the rationale that original VH-VL pairing can better reflect the natural immune composition (13). Semi synthetic libraries are also based on the natural repertoire but contains certain synthetic aspects such as PCR rearrangement of germline VH genes, shuffling of VH and VL domains or CDR oligonucleotide randomization on B-cell isolated framework regions (13). However, naïve libraries ultimately depend on the immune system for B-cell diversity and germline rearrangements can restrict the obtained diversity (268).

In contrast, synthetic libraries are generated using in vitro and in silico methods without the need to extract B-cells from donors. Various strategies are available to generate fully synthetic libraries and these strategies usually differ in 1) the choice of framework, 2) which CDRs / residues are mutated, 3) how diversity is introduced and 4) the choice of amino acid composition in the diversified positions (269).
The framework of the synthetic library can be chosen based on sequence analysis of the human antibody repertoire. This strategy was employed in the development of the HuCAL library where it was found that seven VH and seven VL germline families covered 95% of the antibody diversity. A consensus framework was created for each family and optimized for expression in E. coli(270). The HuCAL library has later been optimized for improved expression in mammalian systems(271). An alternative option is to use single framework based libraries, where the single framework is selected based on high occurrence in natural repertoire(272) or reliable expression and shown to be successful in therapeutics(273). Biophysical characteristics favorable to manufacturing and development can also be used as a key consideration when choosing a framework. This rationale was behind the development of the Ylanthia synthetic library consisting of fixed VH/VL pairs which were selected based on parameters such as thermal stability, expression and glycosylation(274).

Once the framework has been chosen, diversity can be introduced via annealing of degenerate oligonucleotides(19, 275). Alternatively trinucleotide phosphoramidite cassette technology can be used where CDRs are inserted as cassettes into framework regions flanked by restriction sites(270). A third option involves a ligation building block based technique called Slonomics(276). The CDRs which are targeted for diversification can either be CDRH3 / CDRL3 (270, 272), CDRH1-H3 & CDRL3 (19, 275, 277) or all six CDRs(271). More directed approaches can involve diversification at solvent exposed positions while omitting residues not believed to be involved in antigen contact(273, 278). While diversification only in one or two CDRs / one or two chains is adequate for isolation of antibodies(279), diversity in multiple CDRs are generally more successful(24, 280). Alanine scanning can be used to assess synthetic library designs which has showed that CDRH3 is not always required for generation of high affinity binders(275). Lastly, when choosing the CDR amino acid composition, the most popular approach is to mimic the natural amino acid distribution(270, 281), although using a very limited set of amino acids has also been reported(19, 20). Overall, despite the numerous ways to create natural non-immune and synthetic libraries, there does not seem to be a large difference in performance between them with respect to affinity(24). However, library design and selection need to take other considerations into account in addition to affinity, including expression levels, stability, compatibility with automation and ease of conversion to other formats. Once the source of immunoglobulin genetic starting material has been chosen, a suitable display platform can be selected, starting with phage display.
4.4 Phage Display

The most widespread and popular display platform within combinatorial protein engineering is phage display. As the name implies, proteins are displayed on the surface of filamentous phage as fusions to one of the phage coat proteins, while all genetic information encoding a phage particle is present inside the phage.

The phage display platform was first conceived in 1985 by George Smith when peptides were displayed on filamentous phages and enriched by affinity for an antibody(282). A few years later McCafferty et al. showed the possibility to display scFvs on fd bacteriophages and subsequent enrichment using antigen based affinity chromatography and affinity selection(283). These successful attempts naturally led to the application of phage display for selection of both peptide and protein libraries(260, 284, 285).

The selection process uses a “capture and elute” procedure (Figure 4.4.1) and involves panning against surface immobilized, biotinylated or antigens present on cell surfaces(286). Free antigens in solution can also be panned against using, for instance, magnetic beads to capture binding clones. Typically, after incubation of the phage library with the target protein, the phage particles expressing a library member with an affinity for the target are captured, while non-binding phage are washed away followed by elution and isolation of binding phage particles. Eluted phages are used to infect E. coli cells enabling library-gene amplification as well as the production and assembly of new phage particles without lysing the host cells(12, 13). The amplified phage particles can thus be recovered and used for subsequent panning rounds.

**Figure 4.3.1.** Examples of three (yeast, staphylococcal and phage) surface display platforms.

---

**Figure 4.4.1.** Selection process for phage display.
Figure 4.4.1. The selection process in phage display.

Filamentous bacteriophages, such as M13 and fd phages are the most commonly used\(^\text{287}\), although other phages (e.g. \(\lambda\), T4 and T7) have also been employed\(^\text{13}\). Both M13 and fd phages are very similar in genome\(^\text{288}\), where both have a rod shaped structure with a genome of 6000-8000 bases encased in a coat composed of five different proteins (pIII, pVI, pVII, pVIII, pIX)\(^\text{289, 290}\). The major coat protein (pVIII) is present in 2700 copies and covers the length of the phage and on one end of the phage particle are 5 copies of the proteins pIII and pVI, while the other end has 3-5 copies of pVII and pIX. The coat proteins pIII and pVIII remain the most commonly used coat proteins for phage display\(^\text{289}\) although foreign peptides have been displayed when fused to the surface exposed N-terminus of coat proteins pIII, pVII, pVIII, pIX and C-terminus of pVI\(^\text{289}\). Both pIII and pVIII coat proteins contain an N-terminal signal sequence that directs them to the inner bacterial membrane before phage assembly and fusion to these proteins allows multivalent display of the recombinant protein\(^\text{12, 13}\). However, the pIII coat protein is involved in phage-host interaction during infection by binding to the F-pili of \(E. coli\) and protein-fusions to all 5 copies of pIII can interfere with this process. Moreover, multivalent display is not always desired since avidity effects may limit the isolation of high affinity binders. The solution is the development of phagemid systems\(^\text{13, 291}\), which allows monovalent display on pIII. This technique is based on a phagemid vector that encodes the library protein fused to pIII along with a phage DNA packaging signal, while all other proteins essential for phage assembly are provided by a helper phage that also must infect the \(E. coli\) cell to facilitate phage production\(^\text{12, 13}\). In addition, the helper phage also encodes wild-type pIII, thus promoting monovalent display of...
fused pIII along with wild-type pIII. The phagemid display system has successfully been utilized for display of many different antibody fragments, including Fabs(292), scFvs(293) and even full-length antibodies(294). Due to the stability of the phage particle, harsh selection conditions can be used without jeopardizing infectivity(291).

In addition to in vitro display of peptides and antibody fragments, phage display has also been used to select tumor binding ligands by intravenous administration and in vivo panning of phage libraries(295). Other applications include using an expanded genetic code to select a more efficient binder containing un-natural amino acids(296) as well as selection of antibodies binding to specific conformations of caspase-1 (297). Phage display has also been successfully used as vaccine delivery systems(298) and selection of antibodies recognizing post translational modifications independently of sequence context(299).

In summary, phage display has proven to be a very useful, versatile and inexpensive technology with the possibility of generating large libraries in the range of $10^{10}$ - $10^{12}$ variants(13, 300). However, several panning rounds are usually needed and the selected clones can only be assessed after the selection process. Another challenge is that the display efficiency can be influenced by properties such as protein size and charge(301). These limitations with the phage display platform have led to the development of other display systems, which are described in the following sections.

4.5 Flow Cytometry and Cell-Based Platforms

In contrast to the “capture and elute” process used in phage display, cell based display platforms express the protein library on the surface of bacterial or eukaryotic cells. The main advantage with cell based display technologies is that cells are large enough to be analyzed in a flow cytometer using fluorescence activated cell sorting (FACS). In flow cytometry (Figure 4.5.1), each cell is analyzed individually with regards to size and granularity. Apart from analyzing the cell itself, multiple protein interactions on the surface of the cell can also be assessed with different fluorescently labeled proteins (Figure 4.5.1a). Whereas clones selected with phage display cannot be analyzed during the panning procedure, flow cytometry allows simultaneous real-time analysis and selection of the library. By using different protein labeling approaches (e.g. equilibrium binding, competition for limited target, and kinetic competition), quantitative and fine discrimination of protein affinities is possible(302, 303). Since it is also possible to normalize expression levels, bias towards high expression clones is minimized and the selected clones can be subsequently individually analyzed without sub-cloning. The selection process using flow cytometry is similar to phage panning, where the library is first incubated with the antigen followed by washing and real time sorting. The sorted cells can then be expanded for another repeated cycle of screening (Figure 4.5.1b). Over the years, bacteria, yeast and mammalian cells have all been used for cell based surface display, and the following section provides more insight into these cell-based techniques.
**A**

- Cells
- Sorted Cells
- ExpTg
- POI
- Target
- Non-binding

**B**

- Target
- Label
- Amplify
- Sort
- Antigen Binding
- Surface Expression
- Gate

Diagram B illustrates the process of Amplify and Sort, followed by antigen binding and surface expression analysis. Diagram A shows the sorting of cells based on antigen binding and surface expression.
Figure 4.5.1. The selection process using cell-based display and flow cytometric sorting. The protein displaying cells are individually analyzed and multiple protein interactions can be assessed using different fluorophores (a). The surface expression can be monitored via an expression tag (ExpTg, blue fluorophore) while binding to the target antigen is monitored via labeled target antigen (red fluorophore). In the example shown, a non-binding versus a binding displayed protein of interest (POI) will have different graphical profiles, and this profile can be used for screening and selection of cell display libraries (b).

4.6 Bacterial Display

4.6.1 E. coli Display

The Gram-negative E. coli is the most popular host strain utilized in bacterial display(304) due its well-characterized nature, rapid growth rate and ability to yield large libraries (up to $10^{11}$)(304). Being a Gram-negative organism means that the displayed protein needs to be transported through an inner membrane, a periplasmic space and an outer membrane.

In E. coli display, the protein of interest is fused to a carrier protein and transported to the cell surface where it is anchored. Various carrier proteins employed in E. coli display include outer membrane proteins (e.g. OmpX, OmpA, Omp1, OmpC), surface appendages (e.g. FltTrx), lipoproteins (e.g. INP, Lpp-OmpA) and autotransporters (e.g. AIDA)(304, 305). These different carrier proteins can also be broadly grouped into 1) insertional scaffolds (e.g. OmpA) where the protein of interest is inserted into the mid-section of the exposed scaffold; 2) N-terminal (e.g. AIDA) scaffolds where the target is fused to the N-terminal of the carrier and 3) C-terminal scaffolds (e.g. Lpp-OmpA) where the target is fused to the C-terminal of the carrier(304, 305). Despite the variety, most carrier proteins based on outer membrane proteins and surface appendage (i.e. flagella, pili and fimbriae) are mostly suitable for display of peptide libraries(304, 305). In contrast, systems based on lipoproteins and autotransporters have shown to be appropriate for display of larger and more complex proteins(305).

The first example of E. coli cell surface display in combination with flow-cytometric sorting was described in 1993 where scFvs were fused to the C-terminal of the fusion protein Lpp-OmpA and enriched(306). This technology was used a few years later for screening and selection from a scFv affinity maturation library(307). However, drawbacks with Lpp-OmpA include leakiness and permeability disruption of the outer membrane as well as cellular toxicity(308). Autotransporters are a class of naturally occurring proteins in Gram-negative bacteria responsible for secreting proteins to the outer cell surface(309). The general structure of an autotransporter based display system consists of a C-terminal β-barrel domain and a protein of interest fused to the N-terminal domain(309). A few studies have confirmed the possibility of using autotransporters for display of scFvs(309, 310).

A third strategy is to display the protein on the periplasmic side of the inner membrane of E. coli. The first example of this approach was Anchored periplasmic expression (APEx) technology described in 2004 (311) which enabled display and affinity maturation of scFvs. In the APEx system, the recombinant protein can be displayed either as an N-terminal fusion to the six first amino acids of the E. coli lipoprotein NlpA, or as a C-terminal fusion to the M13 phage pIII coat protein(311). The APEx technique was later modified to allow the successful display of full-length antibodies on the inner membrane, which can also be combined with pre-enrichment steps using phage(294, 312). The soluble full length IgG is produced in the periplasm, and is captured to the inner membrane by a dimer version of the Fc-binding Z
domain fused to the lipoprotein NlpA, displayed on the inner membrane. The very slow dissociation of Fc from the dimeric Z domain ensures that the IgG remains captured on the inner membrane even after disruption of the outer membrane(312). A drawback with periplasmic display is that it requires disruption of the outer membrane prior to incubation with labeled target protein and library screening by flow cytometry. This prevents further amplification by cell growth and the spheroplasts have reduced viability in the sorting process, meaning that the genes encoding the isolated clones must be recovered by PCR, subcloned and re-transformed for further sorting rounds, which is time-consuming and laborious.

4.6.2 Gram-Positive Staphylococcal Display
The Gram-positive bacterium *Staphylococcus carnosus* was originally isolated from sausages and is used in the production of meat products(313). *S. carnosus* is classified as a GRAS (generally regarded as safe) organism and shows low genetic homology to the pathogenic *S. aureus*, making *S. carnosus* a good alternative for surface display. Firstly, the displayed protein only needs to be translocated through a single membrane, as opposed to two membranes in Gram-negative bacteria. Secondly, the thick peptidoglycan cell wall acts as a physical barrier protecting the cells during the harsh conditions during flow cytometric sorting(303, 315, 316). Moreover, *S. carnosus* grows as single cells with low extracellular proteolytic activity. To allow successful display, the protein of interest is fused to a signal peptide and a propeptide (originally isolated from *S. hyicus*) that directs the protein of interest to the surface of *S. carnosus*. An albumin binding protein (ABP) is present at the C-terminus of the displayed protein followed by an anchoring mechanism which uses a cell wall anchor protein derived from *S. aureus* protein A(317). The strong interaction between ABP and human serum albumin (HSA) enables efficient monitor and normalization of surface expression levels(303, 315, 316).

Gram-positive bacteria have traditionally suffered from low transformation efficiency due to the thick peptidoglycan cell wall, but optimization has allowed up to $10^6$ cells to be transformed per electroporation(318). Moreover, incorporation of a protease recognition site at the C-terminal of the displayed protein together with a purification tag allows rapid characterization by cleaving the protein from the surface without the need for subcloning(319). The staphylococcal display platform has previously been extensively used for *de novo* selection of Affibody (a non-immunoglobulin based affinity protein scaffold) libraries and affibody affinity maturation(315, 316) as well as epitope mapping using peptide libraries(320). Regarding immunoglobulin based scaffolds, the platform has also been proven successful for the display and sorting of a single domain library(321) and display of a single mouse scFv(322). In Paper III, we expanded on the previous work and used phage display in conjunction with staphylococcal display to select, for the first time, binders from a fully human synthetic scFv library against HER2. Due to the resiliency of the Gram-positive *S. carnosus* and the quick growth rate, the staphylococcal display platform is an interesting complement to the other more mainstream display platforms described in this chapter.

4.7 Yeast Display
The yeast *Saccharomyces cerevisiae* is about 5-10µm in diameter and a household microorganism used by millions around the globe, many of whom may not even know its full name. Apart from baking, winemaking and brewing, *Saccharomyces cerevisiae* also possesses eukaryotic folding machinery and an oxidative endoplasmic reticulum. This makes it a useful host for producing disulphide bond-containing proteins since its eukaryotic secretion system also prevents misfolded proteins from export(323). In 1997 *Saccharomyces cerevisiae* was used by Wittrup’s group to display scFv antibody fragments in combination with flow cytometric screening(324), and thereon the yeast display technology was born.
A disadvantage with yeast, however, is the maximum size of the library that can be created. Yeast library size (~10^7–10^9) has traditionally been a few orders of magnitude lower than bacterial (~10^9–10^10), phage (~10^9–10^11), and mRNA (~10^10–10^11) display(325). Newer more optimized transformation methods have alleviated this issue and creation of yeast libraries exceeding 10^10 clones have been reported(326). But due to the large culture volumes (>1 L) required to generate more than 10^10 cells and the volume occupied by this amount of biomass (>50 mL), libraries larger than this magnitude would be difficult to manipulate(327). Another disadvantage is the difference in glycosylation patterns between yeast and mammalian cells. This can potentially cause issues when displaying human glycosylated protein scaffolds on yeast. While humanization of N-glycosylation pathway in yeast has solved some of this problem(328), differences in other pathways still exist and not all pathways have been humanized(329). Thirdly, the presence of multiple surface protein copies can lead to unwanted multivalent binding with oligomeric proteins, leading to isolation of high avidity binders with low affinity. To overcome this limitation, kinetic selections can be applied(330).

So, does that mean that the yeast display platform has gone the way of the dodo? The answer is no, of course not. Compared to phage display, yeast display was shown to be able to sample the library more comprehensively with a higher chance to isolate rare binders from large libraries(331). A study in 2007 made a direct comparison between yeast and phage display for de novo isolation of anti-HIV-1 scFvs from an immune library. Clone analysis after selection revealed that the yeast platform found a set of new clones not discovered from the phage library(331). A similar study using parallel phage and yeast selection of the same library revealed that each route generated its own distinct set of binders. But more importantly, yeast allowed isolation of binders with both high affinity and expression whereas phage had to sacrifice expression levels for affinity(332). This means that yeast libraries have a high functional diversity due to its eukaryotic post-translational machinery as mentioned earlier. Despite some limitations, the yeast display platform is still very useful and the following section will describe this technology and its applications in more detail.

The yeast display platform relies on naturally present proteins anchored to the cell wall of yeast. This allows foreign proteins to be targeted to the cell-surface via fusion proteins containing a C-terminal glycosylphosphatidylinositol attachment signal (GPI anchor)(333). Several cell wall proteins have been identified nowadays, all proven to be capable of displaying different fusion proteins on the cell surface. These include for instance A-agglutinin(324), Flo1p(334), and alpha-agglutinin(335). Among these, the A-agglutinin-system has shown to be the most successful(333).

A-agglutinin is composed of a 725-residue core-subunit (Aga1p) which includes the C-terminal truncated GPI anchor, and a 69-amino acid subunit (Aga2p). Aga2p in turn contains the binding site for the A-agglutinin ligand Alpha-agglutinin at its C-terminus. Aga1p is linked to Aga2p through two disulphide bridges and due to the GPI-attachment-signal of Aga1p, the core-subunit covalently anchors the whole complex to the cell wall. Each yeast cell is then able to display approximately 10^4–10^5 copies of the fusion-protein on its surface(324). The gene encoding the Aga1 protein is integrated into the yeast chromosome, while the gene encoding the protein of interest-Aga2 fusion is cloned into a yeast display plasmid vector that is maintained in the yeast using a nutritional marker for selective growth. Expression of both Aga1 and Aga2 is under the control of the galactose-inducible promoter. Furthermore, the modular property of A-agglutinin makes it possible to fuse the protein of interest to either the C- or N-terminus of Aga2p(336) whereas single-unit GPI-proteins only allow an N-terminal fusion of heterologous proteins because of the required C-terminal GPI-attachment signal. While proteins have been displayed on yeast fused to both the N- (337, 338) and C-termini of Aga2p, antibody display studies have almost universally used fusion to the C-terminus(339, 340). In this case, the native Aga2p signal sequence directs the fusion to the secretory...
pathway, where assembly of Aga2p and Aga1p occurs intracellularly prior to export to the cell surface.

Yeast surface display has been used to engineer numerous protein scaffolds for therapeutic, diagnostic, and biotechnological applications in addition to its most popular use in antibody engineering. Yeast display has been utilized for both affinity maturation as well as de novo selection of immunoglobulin and non-immunoglobulin scaffolds towards a wide spectrum of targets, including small organic molecules, peptides, and both soluble and membrane proteins(325, 327).

To date, the vast majority of immunoglobulin scaffolds that have been displayed and engineered are antibody single-chain variable fragments (scFvs)(324, 341) and the antibody fragment (Fab) domains(342). In addition to these two classic formats, more examples include the display of camelid VHH domains(343, 344) and the single chain T cell receptor (scTCR)(345). Others have also displayed a hybrid between a scFv and Fab, called single chain Fab fragments (scFab), and shown its feasibility to isolate high affinity binders(346). Single chain Fab fragments is a construct where a VH-CH1 domain is fused via a linker to a VL-CL domain and the rationale is that Fab requires assembly of two separate chains and certain scFvs can lose antigen binding and/or neutralization activity after conversion into Fabs or immunoglobulin Gs (IgGs)(347–349). Apart from antibody fragments based on the variable binding domain, the constant (Fc) domain of immunoglobulin G has also been targeted for engineering into an antigen binding Fc affinity protein (Fcab). By introducing diversity in the loop sequences at the C-terminal of the CH3 domain of the Fc, binding towards HER2 and integrin were made possible(350, 351).

Small antibody fragments such as scFv and Fab can already be reliably expressed on phage and bacteria, but these two organisms struggle with larger proteins. By taking advantage of yeast as an eukaryotic host, several successful attempts have managed display of full length IgGs. These display platforms use a “secrete and capture” approach where an antibody is produced intracellularly and captured non-covalently by a surface adaptor. In contrast to conventional methods where the protein of interest is expressed as a part of the cell wall, a major advantage with the “secrete and capture” method is the ability to easily switch between display for screening and production for functional characterization.

A challenge with early efforts of full IgG yeast display was the inefficient secretion of full length antibodies(352). Later progress succeeded in optimizing the secretory leader sequence which enabled higher secretion(353) and two later the SECANT platform was introduced by the same group. In the SECANT system, avidin is attached to the yeast cell wall and the antibody molecule is expressed as a fusion to a biotin acceptor peptide. A biotin ligase later attaches biotin to the acceptor peptide and allows the secreted biotinylated antibody to be captured by avidin(354). One newer example of full length IgG yeast display is the Fc-Sed1p display system employed in P. pastoris which enables simultaneous display and secretion. Here, the single hinge-CH2-CH3 domain is fused to the cell wall anchored protein Sed1p. Co-expression of a secretable full length IgG leads to dimerization between the single hinge-CH2-CH3 domain and a half IgG molecule. The result is monovalent display of full length IgGs. However, assembly of soluble full length IgGs can also occur with an equal probability and the end result is simultaneous display of monovalent and secretion of bivalent IgG molecules(355). Another more recent example uses the Fc binding ZZ-domain from S. aureus in conjunction with antibody secretion from from respective heavy-chain and light-chain plasmids. The ZZ-domain is displayed in a conventional Aga1/Aga2 manner and is employed to capture full length IgGs on the surface(356).

Moving away from antibody engineering and affinity maturation/selection, other applications
of yeast display include identification of tumor antigens(357) and identification of protein-protein interactions(358). Apart from identification of protein-protein interactions, protein-protein or protein-peptide interactions can also be quantitatively analyzed using flow cytometry. This yeast surface 2-hybrid system involves co-expression of a surface anchored “bait” and a soluble “prey”. If intercellular binding interaction occurs between the “bait” and “prey”, the complex is anchored to the surface and can be assessed without exogenous ligands(338, 359). Another area is increasing thermal stability and expression. A study conducted in 1999 showed that yeast display can be used as a screening variable for enhanced thermal stability and soluble expression of mutant proteins(360) as a more efficiently displayed protein also possessed a higher thermal stability. Now, improved soluble expression can also be achieved by engineering the expression host instead of the protein itself. For example, yeast surface display was used to screen a yeast cDNA library to reveal yeast genes that increased secretion of scFvs and other proteins(361). Another powerful application of yeast surface display is the identification of binding epitopes (both continuous and discontinuous) on a target protein (mentioned earlier in Chapter 3). The approach uses yeast libraries of mutagenesis modified full-length or single-domain protein targets and assaying the binding phenotypes by flow cytometry(362, 363). The majority of examples of yeast display protein / antibody engineering use soluble antigens or extracellular domains of membrane proteins, primarily with fluorescence activated cell sorting and commonly with a pre-selection on solid-phase antigen as a preliminary step(339, 364, 365). However, yeast display has also been adapted to allow selection of antibody libraries by against surface antigens by panning directly on cultured cells(366). Biotinylation of either whole cells or surface proteins and library “panning” against cell lysates has also been applied to recover anti-membrane protein targeting scFvs(367).

4.8 Mammalian Display

Microbial display platforms have generated numerous useful proteins and antibodies with desirable traits. However, large-scale production of therapeutic antibodies currently takes place in mammalian cell line systems and as a consequence, features such as high-level expression, reduced aggregation and correct glycosylation are necessary. These are unfortunately difficult issues for microbial based platforms and as a result there has been a considerable interest to develop mammalian based display systems(333).

In 2006, Ho and Pastan showed that single-chain antibodies (scFvs) could be displayed on the surface of human embryonic kidney 293T cells (HEK293T) by adapting Wittrup’s yeast display platform(368, 369). The scFv was fused to the transmembrane domain of human platelet-derived growth factor receptor (PDGFR) via a myc tag to monitor the expression level and the construct was used for affinity maturation against CD22. In another 2008 example, peripheral blood mononuclear cells from human donors were used to directly isolate B cells specific for an antigen of interest. RNA isolated from these specific B cells was used to generate an enriched random combinatorial scFv library. The library was then converted to a Sindbis virus expression library used to infect mammalian cells, each expressing at their surface one specific antibody. This method enabled isolation of antigen-specific antibodies fragments by a single round of FACS but had to be re-formatted to full length antibodies and expressed in different standard mammalian expression systems if full length IgG were desired(370).

More interestingly, full-length functional antibodies were able to be presented on the surface of mammalian CHO cells by applying the Flp-In technology (Invitrogen). The library diversity was generated by recombinase mediated chromosomal DNA integration at a specific location so that each host cell contained only one copy of the gene of interest in the genome and the expression levels were high and comparable between clones. FACS sorting of a randomly mutagenized library identified antibodies with improved neutralizing activity and affinities. A
500-fold enrichment (from a single round) as well as discrimination of affinities within picomolar range was also possible (371). Another approach utilizes vaccinia virus as a display technology. This method can be seen as a hybrid between phage and cell display and is based on co-infection of mammalian cells with separate libraries of human heavy chain and light chain immunoglobulin libraries that have been constructed in vaccinia virus. The heavy chain constant region contains a trans-membrane domain, so that the antibody can be expressed, assembled, and transported to the cell surface. In addition, the library of fully human antibodies is also displayed on the surface of vaccinia virus itself, so that panning and selection on antigen-coated magnetic beads can be made. Upon infection of mammalian cells, the antibodies are expressed on the mammalian host cell surface, so that infected cells can be stained with antigen and single positive binders can be separated with cell sorting and the virus encoding the specific antibody heavy and light chains readily recovered and analyzed (372).

In the examples mentioned above, the library sizes varied between $10^6$ and $10^7$. As the generation of diversity is crucial and important in any selection process, including mammalian cell display, methods have been developed to combat this issue. Several approached aimed to solve the issue with small library sizes, and hence small original diversity, by integrating in vitro somatic hypermutation with cell display. Somatic hypermutation is dependent on the action of the B cell enzyme activation-induced cytidine deaminase (AID), and can be replicated in non-B cells through expression of recombinant AID.

In one example from 2011, Bowers and co-workers isolated low affinity IgM from naïve B-cells and grafted the CDR3 regions of the heavy and light chains into an IgG framework with a C-terminal transmembrane domain to produce a library of full length germline antibodies on the surface of HEK293T cells. Antigen-coated beads were used to isolate low-affinity antibodies, followed by FACS sorting and AID-induced affinity maturation to produce high affinity antibodies, one of which with low pM $K_D$ (373). The same platform has also been used to generate functional high affinity humanized IgGs from mice against C5. Rearranged heavy chain CDR3 regions were amplified from spleen and lymph tissues of immunized mice and fused with a repertoire of human germline heavy chain genes to form intact humanized heavy chains. The heavy chain library was paired with a human light chain library with diversity in CDR3. FACS selection and subsequent affinity maturation by AID-induced mutagenesis resulted in a high-affinity functional antibody with 1000-fold increase in $K_D$ (374).

Also worth mentioning is the ADLib system (Autonomously Diversifying Library system) which has the advantage of allowing the acquisition of whole IgM molecules in a rapid and convenient manner as well as hypermutation based affinity maturation. ADLib technology relies on using enhancing gene conversion (the main diversification process of immunoglobulin (Ig) genes in chicken B cells) to obtain naturally expanding libraries of mAbs displayed at the cell surface as membrane-bound IgM. DT40 (a chicken B-cell derived cell line) clones expressing antigen-specific mAbs can then be isolated using magnetic beads conjugated to any target antigen of interest. An improvement to the platform allowed generation of chimeric human IgG in DT40 cell display libraries. By introducing the human IgG constant region into the chicken IgM heavy-chain locus, DT40 cells could express both membrane bound chicken IgM and secreted chimeric IgG sharing the same antigen-binding domain. Thus it was possible to simultaneously isolate and produce specific chimeric IgGs of interest (375).

### 4.9 Cell-Free Based Techniques and Combination of Methods

The library size based on living host organisms is limited by the DNA transformation or transfection efficiency. In phage and bacterial display, E. coli or Gram positive cells need to be
transformed whereas in yeast and mammalian display eukaryotic cells need to be transfected. When extremely large libraries are necessary, cell-free methods can be utilized. The cell free display platform is based on the transcription of library genes into mRNA and subsequent translation into proteins in a cell-free environment. By avoiding the transformation/transfection bottleneck, cell free systems are able to produce the largest library size ($10^{12}$ - $10^{13}$ library members) in the display realm (376, 377).

The most common cell-free technique is ribosome display, pioneered in 1994 (378). In ribosome display, the ribosome forms a complex with the corresponding mRNA sequence after translation of the protein due to the lack of a stop codon. This is what provides the essential linkage between the phenotype and genotype. After panning against an immobilized target, the mRNA is released, reverse transcribed into cDNA and amplified using PCR. A few years later, this technology was further developed from displaying peptides to enable display and selection from scFv libraries (379).

Apart from the large obtained library size, ribosome display also minimizes introduction of bias since neither amplification nor secretion in living organisms is present (380). A third benefit of ribosome display is that PCR can be used to introduce diversity between selection rounds. Since PCR is needed regardless to amplify clones between each round, incorporating mutations by randomization in this step can be considered “two birds with one stone” (378, 379). A similar cell-free display technology is mRNA display where puromycin coupled to a DNA primer leads to the formation of a covalent link between the displayed protein and the mRNA. The mRNA is stabilized by the synthesis of a complementary DNA strand before selection (381).

Not using a host organism for display does come with its disadvantages, the biggest of which is sensitivity to RNAse degradation as the crucial phenotype-genotype link relies on mRNA (380). CIS display and CAD display are two alternative approaches, where the peptide is cross-linked to double-stranded DNA, either by the bacterial replication initiator protein, RepA (CIS display) (382), or cis-acting DNA binding protein (CAD display) (383).

The emergence of massively parallel sequencing (MPS, also called deep sequencing) has enabled thorough high-throughput analysis of both the natural and synthetic antibody repertoire (384, 385). The information gained can be used to guide vaccine design, identify endogenous VH:VL pairs and understand the immune repertoire development (384). In addition, deep sequencing can be used for de novo discovery of antibodies by circumventing screening steps. For example, Reddy et al. analyzed the V-gene repertoire of immunized mice using deep sequencing and the most abundant VH and VL genes were paired (386). However, one is not restricted to a single antibody generation platform despite the existence of several different methods. A more recent trend involves using a combination of different display platforms as well as combining deep sequencing with surface display. For instance, phage display can be combined with yeast display in order to take advantage of the ability to create large libraries (phage) with the benefits of FACS (yeast) to screen the library more comprehensively (333, 387). Another alternative uses an antigen displaying phage library for selection against a scFv displaying yeast library (388). This library against library panning allows simultaneous screening against numerous antigens and isolation of cognate antigen-antibody pairs (388). Other combination methods can involve deep sequencing and display platforms where MPS can be used to assess library design, guide the in vitro selection process and identify low abundance clones (385, 389).

4.10 Engineering of Antibodies

Antibody engineering refers to modifying and/or improving the properties of antibodies or antibody fragments using directed evolution. Rational design, chemical and computational
methods are also employed for engineering, but these approaches are outside the scope of this thesis and will not be described in detail. The most popular subfields within antibody engineering are de novo selection and affinity maturation in addition to other subfields including improving stability, engineering for bi-specificity, reducing immunogenicity, modifying Fc effector function, improving pharmacokinetics, and manufacturability.

While engineering / characterization of the antigen binding domain is the main focus of this thesis, other aspects of antibody engineering are just as essential. It is an important reminder that a full antibody is a 150 kDa molecule and having tunnel vision by just focusing on affinity in the variable domain is suboptimal. For example, therapeutic or in vivo diagnostic applications require a long serum half-life and be resistant to aggregation, precipitation and protease degradation(390, 391). Industrial applications demand antibodies with a long half-life and/or activity in organic solvents at high temperatures(392–394). Furthermore, manufacturing of (therapeutic) antibodies needs high levels of expression in addition to chemical and physical stability to withstand degradation (physical and chemical degradation include: oxidation, aggregation, degradation, fragmentation, deamination(395)) during manufacturing, transport, storage and delivery(396).

4.10.1 Improving Affinity

Affinity maturation, or improving the affinity of an antibody / antibody fragment (or any affinity protein in general) is the probably the most popular antibody engineering subfield. An improved affinity is often correlated with enhanced efficacy, improved pharmacokinetics and lower dosing/toxicity(287).

There are numerous strategies that exist for in vitro affinity maturation(287), either by introducing mutations in a targeted or random manner. An easy way is to randomize the entire variable region using error prone PCR(397, 398) in conjunction with display technologies. But as described earlier in Chapter 1, error prone PCR will introduce mutations throughout the entire variable region and complete randomization can create libraries too large to be properly analyzed. Therefore a very popular approach is to either mutate the CDR3 regions only(248) as they form the majority of antigen binding site. Alternatively, mutating the CDR3 in combination with peripheral residues can also be a viable option(399, 400). Random mutations can also be introduced in a procedure called gene shuffling. Here, one chain is fixed (e.g. VH) and is combined with a repertoire of the other chain (VL) to yield a secondary library(401). Following selection of the two libraries (fixed VH-mutated VL and mutated VH-fixed VL), the two mutated chains can be combined to allow further selections(402). A variant of the DNA shuffling procedure can be implemented in yeast display where sub-cloning can be circumvented. In yeast chain shuffling affinity maturation, haploid yeast strains of opposite mating types are used where each strain contains a separate heavy and light chain library respectively. Following yeast mating, the heavy and light chain libraries are combined to form a Fab library(403).

On the opposite end of the spectrum, targeted mutagenesis strategies have also been applied to affinity maturation. One such strategy is to mutate one CDR residue per clone to create a small directed library for each CDR. The library members are assessed and the mutations that lead to improved affinity are then combined to form a smaller group of combinatorial variants in the next iterative cycle(404). One can also choose to mutate and analyze/optimize each CDR sequentially in a process called CDR walking. Here, the information or best candidate(s) after each round is used as the template for the next cycle of mutagenesis and subsequent library generation(405, 406). The arrival of massively parallel sequencing has further enabled more time efficient analysis of each CDR library. For instance, Hu et al generated 5 separate libraries where each library had one mutated CDR. Phage panning and deep sequencing of the selected clones was used to provide information on the enrichment landscape and the most
common abundant variants were combined to form a secondary library (407).

Yet other groups have attempted to mimic natural in vivo somatic hypermutation for in vitro affinity maturation. During in vivo somatic hypermutation, point mutations are introduced in the variable regions at a rate of $10^{-3}$ mutations per base pair per cell division (408). Somatic hypermutation also tends to spread diversity to more peripheral regions not in contact with the antigen, whereas germline recombination mostly introduces diversity in the central region of the antigen binding site (151). It is believed that these more peripheral positions are more tolerant to mutations without causing disruptions to antigen binding (409). Some examples were already given in the mammalian display section (Chapter 4.8) where expression of the recombinant AID enzyme in antibody displaying mammalian cells induced in vitro somatic hypermutation. Another way to mimic somatic hypermutation is called hot spot mutagenesis and involves targeting and introducing mutations around hot spot motifs (410, 411). These hot spot motifs are regions where a high mutation frequency is found during somatic hypermutation (approx 50-60%) and consists of DGYW (D=A, G, T; Y = C, T; W= A, T) (408).

There is a variety of selection strategies and protocols described that can be used to improve affinity and kinetics using phage- and ribosome display (412). A common approach is to use a low or decreasing amount of antigen after each successive round, which favors clones with lower $K_D$. If an improved off-rate is desired, a long washing step after incubation with target antigen can be used. To select for improved on-rate, a fast incubation with the antigen is employed (24). Affinity maturation using FACS based platforms are similar to that of phage display. Due to the possibility of real-time visualization of the selection process, FACS based platforms allow fine affinity discrimination by varying antigen concentration and dissociation of antibody-antigen complexes (302).

Apart from affinity, selecting for specificity is can be just as an important factor when performing selections. To achieve this, the phage library can be panned against immobilized target antigen in the presence of an excess of a related non-target antigen in solution. Since the majority of phages will bind to the free non-target antigen, the cross-reactive phages recognizing both targets can be washed away (413). This approach, called competitive deselection, can be used to avoid cross-reactivity towards conserved epitopes and direct the selection process towards certain epitopes (413).

### 4.10.2 Improving Stability

The stability of antibodies can depend on isotype (414) or the presence of N-glycans in the Fc domain. As N-glycans stabilize the CH2 domain of IgGs, deglycosylation makes mAbs less thermostable and more prone to unfolding and aggregation (415). Apart from the Fc domain, it has also been shown that different germline families of human variable domains are responsible for differences in biophysical properties, with VH families 1, 3 and 5 being more stable (394).

To increase the stability, one can engineer the constant domains by mutating the CH2/CH3 or introducing additional di-sulphide bridges in the respective domains (414). Alternatively, the human variable domains can also be chosen as the subject of engineering. One approach to engineer more stable human variable domains is to introduce elements of camelid single domain antibodies (VHH) (416). Camelid VHHs are generally well expressed, have good solubility (417) and stable at high (ca 90 °C) temperatures (418). To compensate for the lack of a VL chain, VHHs have longer CDRH3 loops to protect the VH-VL interface, but camelization attempts of human VH domains tends to lead to reduced expression levels (416). Surface display has been another valuable tool to enhance stability by finding mutations that can be used to make single VH human domains. For example, Sidhu’s group used a protein that was able to bind to intact (but not aggregated/unfolded) human VH3 domains in conjunction with
phage display to identify beneficial mutations. Several mutations in the VH3/VL interface were identified that improved solubility, thermal refolding and melting temperature. More importantly, some of these mutations were independent of CDR3 diversity (419). Another way to improve stability involves identification of variable germline sequences with favorable properties (394) and using alignment of the framework repertoire to identify amino acids occurring with high frequency. The consensus framework sequences can later be optimized for expression (270, 420). Apart from framework mutations, the CDRs have also been engineered to increase stability by displaying human VH single domain antibodies on phage. The phages were heated to 80 °C to induce unfolding and to promote aggregation of the single domain antibodies. After cooling, the single domain antibodies were selected for binding to protein A (a ligand common to these folded antibodies) and those antibodies that unfold reversibly were enriched (421).

4.10.3 Engineering of Antibody Fragments

While the full 150 kDa immunoglobulin G is an antibody in its natural form in humans, there are certain applications where the full-length antibody is disadvantageous. For instance, long serum half-life (due to binding of the Fc domain to the FcRn receptor) is undesired in imaging applications due to poor image contrast and Fc activated effector functions can lead to unwanted cytotoxicity. In addition, large antibodies penetrate solid tumors more slowly, and are non-uniform in their final distribution (74). Apart from size, tumor uptake also depends on affinity. In this regard avidity seems to be more important than affinity (422). This is due to the fact that tissues possess a non-equilibrium environment and even high affinity monovalent binders experience poor target retention because of fast dissociation rates. In other words, smaller antibody fragments are advantageous with regards to tumor penetration and blood clearance, but suffer from poor tumor retention due to their monovalent nature. It is therefore advantageous that designing multivalent antibody fragments would increase their functional affinity (i.e. avidity) while maintaining improved tissue penetration and rapid blood clearance relative full-length IgG (423–426).

4.10.4 Engineering of Bispecificity

Bispecific antibodies (bsAbs) possess specificities from two different antibodies and can simultaneously bind different antigens. There are approximately 50 different formats of bsMabs available (reviewed in (427) and (428)) and these can generally be divided into those that possess a Fc domain and those that lack a Fc domain. The Fc region facilitates purification, gives improved solubility/stability/long half-life, as well as mediates Fc effector functions (e.g. ADCC and complement activation). Those bsAbs that lack an Fc region rely entirely on their antigen-binding capacity to exert their therapeutic activities. These “scFv bsMabs” are relative easier to make by linking the two different scFvs with a peptide linker (429).

To make full length bsMabs, two separate mAbs can be expressed by separate cell lines and then combined with protein engineering or biochemical methods (430, 431). Alternatively, two transformed E.coli cell lines (each expressing a half antibody) can be cultured together (432). But being able to produce bsMabs from a single cell line is more attractive and desirable. However, a major challenge with this is concept is the random assembly of different chains from two separate co-expressed antibodies within the same host cell. Fortunately a major breakthrough called the knobs-into-holes technology (433) was presented in 1996. Here, different mutations were introduced into the two CH3 domains, thus forcing heterodimerization during co-expression. The knob-into-holes technology has been subsequently followed by several variations (reviewed in (434)). Nonetheless, none of these techniques could solve the "light chain problem" i.e. random pairing of the two different light chains with the two chains of the heavy chain heterodimer. An elegant solution to this "light chain problem" was provided by CrossMab technology (435). Here, correct pairing of the light
chains was accomplished by exchanging the CH1 domain of one heavy chain with its corresponding light chain CL domain. Another solution was presented in 2014 where it was found that VH-VL interactions can dominate the interaction specificities between the heavy and light chains. By introducing mutations into the CH1-CL and VH-VL interface of the Fab domains, forced correct paring of the light chains with their corresponding heavy chains was achieved(436).

Being able to bind two distinct epitopes as a single molecule has given bsMabs several useful applications. These include 1) recruitment and activation of immune cells (T cells and NK cells), 2) interference with signaling pathways, as multiple inhibitions of surface receptors is possible, 3) delivery of targeted payload and 4) forced formation of protein complexes by placing targets in close proximity(427).

4.10.5 Engineering for Improved Pharmacokinetics and Effector Functions

Engineering of the Fc portion of the antibody molecule aims to enhance effector functions such as antibody dependent cell mediated cytotoxicity (ADCC), complement dependent cytotoxicity (CDC) and pharmacokinetics. In the Fc domain, each heavy chain CH2 contains one N-glycosylation site at approximately Asn297 and about 20% of human IgGs carry a N-glycosylation site in the variable region(437, 438).

The long serum half-life of full length mAbs can be attributed to their binding to the neonatal receptor FcRn in a pH dependent manner (strongest at slightly acidic pH to marginal under neutral and basic conditions)(439). Increasing half-life, which results in lower dosage, is commonly achieved by engineering stronger binding to the FcRn(440, 441). Although the pharmacokinetic profile of immunoglobulin G (IgG) isotype is strongly dependent on interaction with the FcRn receptor, the Fab domain has also been suggested to have an impact on FcRn binding(442). Another later analysis revealed that the charge distribution in the Fv domain is involved in excessive binding to the FcRn. This excessive binding prevents efficient FcRn–IgG dissociation at physiological pH, thereby reducing FcRn-dependent terminal half-lives(443).

Apart from pharmacokinetics, the Fc region has also been the target for improved effector functions. The most popular approach is via glycosylation to increase the affinity towards Fc receptors on effector cells(444, 445). For example obinutuzumab (an ADCC enhanced version of rituximab) was given the breakthrough therapy designation and approved in 2013 by the FDA(446, 447). Obinutuzumab is an afucosylated (lacking fucose sugar units in the Fc oligosaccharides) antibody and was manufactured using cell lines that removed fucose from the N-glycans, which greatly enhanced affinity towards FcγRIII, and increased NK mediated ADCC(448, 449).

Two good reviews on antibody glycosylation are provided in (450) and (451). There, the authors describe different types of antibody glycosylation and the impact that glycosylation has on antibody structure, effector function (in addition to the example given above), pharmacokinetics (e.g. clearance from circulation) and immunogenicity.
5. Present Investigation

In this chapter, the work in the four appended papers and manuscripts is presented and reviewed.

In Paper I, we compared two immunogen design approaches and investigated whether the epitope information of a carefully characterized rabbit polyclonal antibody can be used to generate renewable murine monoclonal antibodies with retained binding properties.

In Paper II, we characterized the epitope of the therapeutic antibody Eculizumab and explained the poor response of a patient group towards Eculizumab treatment. We also present evidence suggesting that the genotype of patients can be used to guide treatment in a precision medicine setting.

In Paper III, a Gram-positive bacterial surface display was evaluated for selections from human scFv libraries in conjunction with phage display. This is to our knowledge the first reported use of Gram-positive display for selections of human scFv libraries using flow cytometry and the selected clones showed similar affinity ranking when expressed in different hosts.

In Paper IV, we present a solid phase cloning method for generation of affinity maturation libraries where multiple site-directed mutations can be incorporated simultaneously. This method produces libraries with a low wild-type content. The method was also used to map the paratope of the parent binder, the information of which was used to create directed maturation libraries.
Contributions to the appended papers

**Paper I:** Designed and planned the experimental procedures together with co-authors. Conducted experimental characterization and of mAbs (with the exception of IHC) and wrote and reviewed the manuscript with co-authors.

**Paper II:** Purified C5 mutants and conducted C5 functional assays together with co-authors and reviewed the manuscript with co-authors.

**Paper III:** Designed and planned the experimental procedures together with co-authors. Conducted experimental bacterial display selection, affinity maturation and characterization of selected clones (with the exception of BLI affinity measurement). Wrote and reviewed the manuscript with co-authors.

**Paper IV:** Designed and planned the experimental procedures together with co-authors. Conducted all experimental procedures (with the exception of cloning of individual mutants and MiSeq deep sequencing of libraries). Wrote and reviewed the manuscript with co-authors.
5.1 Paper I – Generation of HER2 monoclonal antibodies using epitopes of a rabbit polyclonal antibody

Polyclonal antibodies (pAbs) are widely used in research and diagnostics and part of their utility and popularity can be primarily traced to their inherent ability to recognize and target several epitopes of the antigen and ease of manufacture. A huge drawback of pAbs however, is the finite amount of reagent generated by each immunization and batch-to-batch variation(452). An attractive strategy would be to convert a polyclonal antibody with good binding performance into a renewable monoclonal with identical or similar binding specificity. In this paper, the hybridoma technology was used to generate monoclonal antibodies, as it is the most well-established method, mimicking the performance of a polyclonal rabbit antibody. However, simply using the same antigen for mouse immunizations will not necessarily lead to hybridomas targeting the desired region. Differences in immune response between individuals and species might generate a binder that not only possesses a new paratope surface, but also reactivity to a different part of its cognate antigen. This will in turn result in extensive supernatant screening of expanded hybridoma clones, which may or may not allow identification of binders towards the functional region.

In this paper, a 127 amino acid long protein fragment called PrEST (Protein Epitope Signature Tag) encompassing regions of domains 2-3 of the extracellular domain of HER2 (Green, Figure 5.1 a,b) was originally used to generate rabbit pAbs. An epitope mapping(320, 453, 454) and fractionation of the pAb serum(455, 456) together with reassessment of functionality of the pAb-fractions revealed the residues in the C-terminus of the original antigen as main contributors to functionality(457). This information, together with knowledge of the 3D structure of HER2, was used for design of a peptide antigen located in a loop region of domain 3 of HER2, depicted in red in Figure 5.1 a,c.
Figure 5.1. Schematic representation of the extracellular domain of HER2. The original PrEST antigen (green) and peptide antigen (red) used for immunization are shown together with the epitopes of the commercial antibodies trastuzumab and pertuzumab (purple).

In other words, two different approaches were undertaken to establish anti-HER2 mAb-secreting mouse hybridoma cell lines. The first approach utilized the original 127 amino acid PrEST antigen, whereas the second approach made use of the peptide antigen based on knowledge of which epitopes originally contributed to the characteristics of the rabbit pAb. In Paper I, we further evaluated whether having the knowledge of the pAb epitope would allow us to design a better immunogen to generate mAbs targeting the same epitope region as the pAb.

5.1.1 Epitope binning of anti-PrEST mAbs vs anti-peptide mAbs
The two different approaches were first assessed by epitope binning based on reactivity towards different protein fragments overlapping the PrEST antigen using Luminex bead array (Figure 5.1.1.1). The protein fragments consisted of the original PrEST immunogen (blue, HER2 position 274–400) and three overlapping fragments covering the N-terminus (red, HER2 position 236–363), Middle (green, HER2 position 347–492) and the C-terminus (purple, HER2 position 364–530) of the PrEST immunogen. The Luminex epitope binning experiment showed a clear binding to the PrEST antigen for all tested hybridoma supernatants. However, it also revealed focused reactivity towards the N-terminus of the PrEST with no clones binding C-terminal parts for the hybridoma generated by the complete...
PrEST immunogen, indicating a shift in epitope preference compared to the original pAb reactivity (Figure 5.1.1.1a). In contrast, the hybridoma supernatants from clones generated towards the peptide showed no reactivity with the N-terminal region but gave clear signals towards the original PrEST antigen as well as both PrESTs covering the C-terminus (green and purple) (Figure 5.1.1.1b).

**Figure 5.1.1.1** Epitope binning of anti-PrEST mAbs (a) versus anti-peptide mAbs (b) towards protein fragments overlapping the different regions of PrEST antigen.
5.1.2 Evaluation of HER2 binding capabilities of anti-peptide mAbs

The ability of the anti-peptide mAbs to bind to the native form of HER2 was evaluated by incubating HER2 overexpressing SKOV3 cells with purified mAbs. Following a second incubation with a fluorescently labeled goat anti-mouse antibody, the different anti-peptide mAbs were analyzed with flow cytometry (Figure 5.1.2.1). An anti-RBM3 mAb (Figure 5.1.2.1 black) was used as a negative control and Pertuzumab (Figure 5.1.2.1 red) was used as a positive control respectively. This is a necessary step, since neither the peptide nor the PrEST antigen necessarily presents the same folded three-dimensional structure as the native protein during immunization. Six of the eight anti-peptide mAbs (2F5, 3F6, 4E1, 7A7, 10G12 & 15A7) showed binding capability for live cells, albeit with a lower ability than Pertuzumab. When compared to the anti-PrEST mAbs (the best of which is clone 9F6), these anti-peptide mAbs exhibited equal or stronger binding to native HER2. When the median fluorescence intensity signals were normalized and compared, the four strongest cell binders originated from immunizations with the peptide approach.

![Flow cytometric binding analysis of the mAb 2F5 (anti-peptide with highest binding ability) and 9F6 (anti-PrEST with highest binding ability) towards HER2 overexpressing SKOV3 cells.](image)

Apart from binding to live cells, the ability of the anti-peptide mAbs to bind to HER2 using immunohistochemical staining was also assessed (Figure 5.1.2.2a) Five anti-peptide mAbs (2F5, 3F6, 10P3, 10G12 & 15A7) were analyzed by IHC on normal and carcinoma breast tissues, all of which showed similar staining profiles (only 2F5 is shown). Weak staining was present in normal tissues (Figure 5.1.2.2a left) in contrast to strong and specific staining of the cell membranes of the HER2 overexpressing carcinoma breast tissue (Figure 5.1.2.2a right). One anti-peptide antibody (2F5) was further investigated by immunohistochemical staining of HER2 in mice xenografted with SKOV3 tumors (naturally high HER2 expression) and A431 tumors (naturally limited HER2 expression) (Figure 5.1.2.2b). Images showed distinct staining of HER2 in the membranes of SKOV3 tumor cells (Figure 5.1.2.2b left), while HER2 staining in A431 tumors cells was considerably weaker or absent (Figure 5.1.2.2b right).
5.1.3 Epitope and paratope characterization of anti-peptide mAbs

A final epitope mapping and paratope characterization experiment was done to understand the differences in cell binding and specificity between different mAbs considering the peptide antigen used was only 27 amino acids long. All anti-peptide mAbs were subjected to detailed epitope mapping using tiling 15-mer peptides with 14 amino acid overlap. The 15-mers covered the 27 amino acid epitope region represented by the peptide antigen as well as 15 amino acids upstream and downstream of the epitope region respectively present in HER2 (Figure 5.1.3.1a,b). The epitope of the most promising binders (2F5, 3F6, 4E1, 10G12 and 15A7) was located to the exposed central loop region between two helices in the natural conformation of HER2 (Figure 5.1.3.1c). Almost all mAbs had identical or similar epitope profile as 2F5, the only exception being 16E11 (Figure 5.1.3.1b,d). Unable to bind SKOV3 cells, 16E11 only displayed binding to one 15-mer peptide containing a C-terminal phenylalanine. A free C-terminal carboxyl group was most likely required for 16E11 to bind to the peptide, which could explain its inability to bind the native protein in the live cell assay.

Figure 5.1.2.2. IHC (a) and xenograft staining (b) of the anti-peptide mAb 2F5.
Although the mAbs 2F5, 3F6, 10F3, 10G12 and 15A7 had the same epitope, there were clear differences in IHC and native protein binding. The same statement could be made for 4E1 and 7A7 and likely explanation would be differences in CDR composition. Full CDR information was obtained for three representative clones (2F5, 3F6 and 4E1) with varying binding strengths and unsurprisingly, the CDR sequences proved to be similar, but not identical. Differences in CDRL1 and heavy CDRH1 and CDRH2 seem to be the factors behind the weaker binding ability of 3F6 compared to 2F5. Similarly, mutations in CDRL3 and CDRH3 are probably the reason 4E1 is a considerably stronger binder than 3F6 despite identical CDRs elsewhere.

In summary, some of the potential issues when trying move from a pAb to a mAb could be circumvented by using a combination of structural information, knowledge of the desired epitope region of the mAb and epitope mapping of the pAb. Comparing the results from immunization of the original PrEST antigen and a new peptide antigen showed that the latter strategy was more successful in yielding several hybridoma mAbs with the ability to bind HER2 both natively on live cells and in paraffin embedded tissue sections. As these mAbs had a different epitope than pertuzumab and trastuzumab, one can speculate that this strategy could be used to generate therapeutic mAbs used for combination therapy with pertuzumab and trastuzumab.

**Figure 5.1.3.1.** Epitope mapping of the anti-peptide mAbs using tiling 15-mers (a, b) and location of the epitope of the anti-peptide mAbs on the native HER2 (c, d).
5.2 Paper II – Stratification of responders towards eculizumab using a structural epitope mapping strategy

Precision medicine refers to providing effective medical treatment based on an individual patient’s molecular characteristics. It does not mean the creation of pharmaceuticals specifically uniquely tailored to a patient, but rather the ability to classify individuals into subpopulations that differ in their susceptibility to a particular disease or in their response to a specific treatment. To allow for such classification, knowledge of the precise binding site of the drug on its target is of great importance. This paper describes how epitope information, using the therapeutic antibody eculizumab and its target complement component 5 (C5) protein, can be used to guide treatment. As described in Chapter 2, the complement protein C5 occupies a central role in the complement system, and upregulation of the complement system is the root cause of paroxysmal nocturnal hemoglobinuria (PNH) and atypical haemolytic uremic syndrome (aHUS) which leads to hemolysis. Although the humanized monoclonal antibody eculizumab (Soliris) is used for treatment of PNH(82, 458) and aHUS(66, 84) there have been reports on non- or poorly responding patients(86–88, 459). By binding to C5, Eculizumab should prevent the cleavage of C5 by C5 convertase, but a study on Japanese PNH patients carrying a mutation in the C5 gene showed that it no longer can be bound by eculizumab(86). Therefore, determining eculizumab’s epitope on its target protein could enable the development of genotyping kits for prediction of a patient’s response to eculizumab treatment.

5.2.1 Epitope characterization of Eculizumab

In this paper, the first step involved a crude mapping of eculizumab to the binding domain of C5. Each of the twelve domains of C5 was expressed on the surface of the Gram-positive Staphylococcus carnosus which allowed binding analysis of the eculizumab using a fluorescently labeled secondary antibody. The binding analysis showed that the eculizumab binds exclusively to the MG7 domain (residues 822–931), a result consistent with a structural study published after our experiments were conducted(460).

In the next step, error-prone PCR of the MG7 domain was used to construct a random mutagenesis library and expressed on the surface S. carnosus. The albumin-binding protein (ABP), fused to the c-terminal of the protein library, was used with fluorescently labeled human serum albumin (HSA) to detect surface expression of the library. Clones that retained a high HSA signal, but a reduction in eculizumab binding were sorted and subjected to sequence analysis and characterization. Clones containing multiple amino acid mutations were excluded from the analysis as well as clones containing mutations involving cysteine, glycine and proline due to the fact that these amino acid changes often result in disruption of the overall conformation(362). One exception was R885C which had earlier been reported as a mutation present in patients(86) and was therefore included. This clone together with ten others led to the identification of 11 single clones (I829K, Q854L, F855I, R885C, R885H, K887N, K887R, K887T, V896E, W917L, F918S) all with a single mutation in the MG7 domain. All eleven clones showed reduced or absent binding to eculizumab when analyzed individually by flow cytometry and since there three clones had mutations in the same position (position 887), the results suggested that eight different amino acids in the C5 domain are involved in the binding to the therapeutic antibody eculizumab. The epitope mapping experiments are summarized in Figure 5.2.1.1.
residues are brought together on the (PDB5I5K) comparison to the (binding to C5 inhibit all C5 mutants was due to the fact that OmCI does C5, whereas Eculizumab had no effect on R885H inhibition of OmCI against R885H had a similar profile as Eculizumab had on the wild (Figure 5.2.2 eculizumab was unable to inhibit the structure caused by the mutation (Figure 5.2.2.1). The epitope information together with a patient's genotype can help guide treatment in precision medicine (b).

5.2.2 Functional assays of full-length C5 mutants
Further analysis by ELISA using nine full-length C5 mutants (I829K, Q854L, F855I, R885C, R885H, K887N, V896E, W917L, F918S) produced in CHO cells showed that five mutants had reduced binding to eculizumab with two showing a non-significant loss of binding. These seven clones (Q854L, F855I, R885C, R885H, K887N, W917L, F918S) were further analyzed for hemolytic activities (i.e. formation of C5b-9) to confirm that these mutant clones were biologically active and that the drop in ELISA binding was not due to a major change in structure caused by the mutation. The hemolysis assay showed that all seven mutants were able to induce C5b-9 formation in a concentration-dependent manner when eculizumab was absent. Addition of 3.7 nM eculizumab led to the inhibition of C5b-9 formation in the wild-type C5 and lowered hemolytic activity in the mutants F855I, K887N, and F918S. In contrast, eculizumab was unable to inhibit the hemolytic activity in Q854L, R885C, R885H and W917L (Figure 5.2.2.1a). However, when 3 nM of the Ornithodorus moubata complement inhibitor (OmCI) was added instead of eculizumab, a clear anti-hemolysis effect was observed (Figure 5.2.2.1a). Moreover, a hemolysis titration assay showed that the hemolysis inhibition of OmCI against R885H had a similar profile as Eculizumab had on the wild-type C5, whereas Eculizumab had no effect on R885H (Figure 5.2.2.1b). The ability for OmCI to inhibit all C5 mutants was due to the fact that OmCI does not compete with eculizumab for binding to C5 (460). In other words, the ELISA and hemolytic assay suggest that six residues (Q854, F855, R885, K887, W917 and F918) in C5 are involved in binding to eculizumab. A comparison to the published co-crystal structure of C5 with an eculizumab Fab (PDB5J5K)(85), which was determined in parallel as our mapping effort, showed that all six residues were brought together on the -sheet in the folded C5 MG7 domain(85). Moreover, all six residues were found in three contact regions (positions 851–858, 882–888 and 915–920)
identified by the authors(85), who suggested from the binding pocket that residues W917, F918 and R885 are involved in binding.

In summary, this paper involves the successful identification of six important residues on the complement protein C5 which are responsible for binding to eculizumab. A coarse domain mapping was first conducted with bacterial display followed by a more detailed mutagenesis mapping. Several functional assays revealed the six epitope residues which comprise a structural epitope on C5 and which could explain the observed non-responsiveness in a patient group towards Eculizumab treatment. The successful inhibition of C5 hemolysis by an alternative C5 binder (OmCI) indicates a possibility for precision medicine, where epitope knowledge of the therapeutic antibody is crucial to guide treatment. We therefore believe that the epitope information together with the genotype information of patients can be used to determine an appropriate alternative treatment for those who fail to respond to the initial treatment.

**Figure 5.2.2.1. Hemolysis assay of C5 mutants.**
5.3 Paper III – Phage and Gram-positive bacterial display of human antibody repertoires enables isolation of functional high affinity binders

Display technologies have become extremely valuable tools for de novo antibody selection and affinity maturation. In addition, these technologies have also been successfully employed to engineer stable antibody framework sequences(270) and study protein interactions(330).

The main benefits of display technologies is the ability to bypass immunization, to spare animal life, avoid immune tolerance, and also to screen for functional binding in a more controlled setting using a desired diversity. Of the several display technologies that exist today, phage display(282, 284, 461) is the most widely used. The success of phage display is mainly attributed to its ability to sample large libraries and ease of handling. Other display formats include cell based display platforms such as E. coli(306), yeast(324) and mammalian display(462). These cell-based platforms offer several advantages compared to the capture and elution procedure of phage display. Firstly, cell based display platforms are compatible with flow cytometric sorting, where cells displaying an antibody library can be incubated with a fluorescently labeled antigen, allowing simultaneous real time screening and selection. The stringency of the sorting process can easily be manipulated by changing the gate from which the cells are sorted, enabling superior discriminatory superior(303). Lastly, surface expression can be normalized and sorted clones can be individually characterized without subcloning. The eukaryotic yeast cell has been a popular display platform mainly due to its post-translational processing machinery. Differences still exist however between yeast and human processing(463) and glycosylation is typically not necessary for selection of scFv clones, as subsequent reformatting and expression as full-length antibodies in mammalian cells is employed for production. The Gram-negative prokaryotic E. coli is a household name which has also been used for isolation of antibodies from libraries(312). Recognizing the capabilities of yeast and E. coli as display platforms, we believe display using the Gram-positive Staphylococcus carnosus would be a suitable complement to the existing cell based platforms. In Gram-positive bacteria, the displayed protein only needs to be translocated through one membrane and one layer of peptidoglycan. The thicker peptidoglycan layer also serves to protect the cells from the harsh conditions during flow cytometric sorting, leading to higher cell viability after sorting, which is an important requirement of cell surface display.

Previous work has been successful in increasing the transformation efficiency(318) of S. carnosus in addition to selection of affinity proteins(315, 316), camelid single domain antibodies(321) as well as for epitope mapping of linear(320) and structural epitopes(464). In this proof-of-principle Paper III, S. carnosus was utilized as a platform for selection of human scFvs, showing for the first time that Gram-positive bacterial display can be used for selection and isolation of binders from human scFv libraries.

5.3.1 Design and evaluation of the S. carnosus selection platform

The basic principle is illustrated in Figure 5.3.1.1. The scFv library is anchored to the cell wall of S. carnosus via an XM domain, originally isolated from protein A. At the N-terminus of the scFv a propeptide (originating from S. hyicus) helps to translocate the scFv to the cell surface. The albumin binding protein (ABP) at the C-terminus of the displayed scFv is used to monitor and normalize surface expression when bound to fluorescently labeled human serum albumin (HSA). After incubation with two fluorescently labeled proteins, one of which is the antigen while the other is HSA, the population exhibiting binding to the target antigen can be isolated in a flow cytometer. The sorted clones can be expanded for another round of sorting or characterized individually.
Before selection could take place, the platform was evaluated to confirm its discrimination ability. For this end, a mock library was made by spiking *S. carnosus* cells displaying Herceptin scFvs (positive control) into a background of negative control cells (*S. carnosus* cells displaying anti CTLA4 scFvs). Two rounds of sorting successfully enriched the positive clone from a ratio of 1:10000 to 1:30 (Figure 5.3.1.2).

5.3.2 Library selection and characterization

In the next step, HER2 was chosen as the target antigen for the selection of a human scFv library. The synthetic human scFv library was first subjected to three rounds of pre-selection using phage display to reduce the size of the library to $10^4$. The pre-enriched library was then...
sub-cloned into *S. carnosus* and subjected to flow cytometric sorting (Figure 5.3.2.1a). 200 nM of target antigen was used in the first sorting round whereas 40 nM was used in the second round. A clear enrichment could be seen after two rounds of sorting after which 96 clones were picked for subsequent sequencing.

Sequencing of the 96 clones revealed 18 unique clones, of which the three individual clones with the highest binding ability (Sc1, Sc39, Sc66) were subjected to a more detailed analysis. This involved flow cytometric analysis using scFv in surface displayed form (Figure 5.3.2.1b), as well as affinity determination by BLI using purified soluble scFv molecules. Apart from characterizing the selected clones as scFv, the clones were also reformatted as full length IgGs and assessed for binding to native HER2 on live cells (Figure 5.3.2.1c).

The affinity measured in the BLI indicated that the clone SC1 was the strongest binder, ahead of SC66 and SC39. This affinity data correlated very well with the data obtained from the on-cell flow cytometric analysis on *S. carnosus* and live cells, despite the scFvs having been expressed by three different host organisms (E. coli, *S. carnosus* and HEK293T).

![Figure 5.3.2.1](image)

*Figure 5.3.2.1. Selection (a) of a human scFv library and characterization using on-cell flow cytometry (b) and as re-cloned full length IgGs (c).*

5.3.3 Error prone affinity maturation

In a final experiment, the *S. carnosus* platform was used to assess whether it could be used for subsequent affinity maturation. One of the selected clones, SC1, was subjected to error prone PCR and the maturation library (size: 10⁴) was sorted in the same way as earlier. Two affinity matured clones were obtained from two rounds of sorting and a sequencing analysis revealed that the point mutations were located outside the CDR loops.

Overall, the experiments in this proof-of-principle paper illustrate the viability and to our knowledge the first example of selections from a human scFv library using Gram-positive bacterial surface display. Despite the lower transformation efficiency compared to *E. coli*, the
fast growth rate and resiliency of *S. carnosus* makes this platform a suitable complement to existing display methods for real time selection using flow cytometry, especially when combined with phage display.
5.4 Paper IV – SPUX – A solid phase uracil excision method for antibody affinity maturation and paratope mapping

The affinity maturation result from Paper III revealed that the mutations from an error prone PCR were, located outside the CDR loops. Due to the random nature of error prone PCR, various methods exist in an attempt to incorporate mutations specifically in the CDR loops.

This is important for several reasons. First, in addition to affinity, antibodies also need to possess adequate (biophysical) properties in order to allow successful downstream production. These (biophysical) properties include low susceptibility to aggregation, high expression levels in cells and low immunogenicity. Since different germline immunoglobulin genes have varying biophysical properties (394), framework genes used in synthetic antibody libraries have often been optimized with regards to these biophysical properties (270). Random mutations would thus require re-assessment.

There are numerous methods to generate affinity maturation libraries where the diversity is restricted to a region of choice (19, 20, 24, 405). Some examples include, CDR walking, Kunkel based approaches, Pfunkel, QuickChange and nick mutagenesis (405, 465–467). However, these methods have only reported the ability to mutate 1-4 CDRs simultaneously. Moreover, some of these methods require preparation of uracil containing dsDNA from E. coli. As a consequence, many affinity maturation libraries focus on the CDRH3 and CDRL3 loops as these are involved in most binding cases. However, in vivo somatic hypermutation tends to spread mutations to the more peripheral regions of the antigen binding site (409) and structural analysis of antibody-antigen complexes indicate that all six CDRs may contribute to antigen binding (146).

Therefore, in Paper IV, a method was developed to allow simultaneous mutation of six CDRs in the construction of a random scFv affinity maturation library. This method not only allowed for a very clean library to be generated, by eliminating the original wild type gene, but was also able to be used for swift generation of targeted single mutants in an automated and high-throughput manner. The single mutants were used to map the paratope of the wildtype scFv and the paratope information was used to construct two different directed maturation libraries. The two directed libraries were assessed and compared to the random library (with all positions randomly mutated in the six CDRs).

5.4.1 Design of the SPUX platform

The solid phase cloning approach (Figure 5.4.1.1) is based on enzymatic creation of the wildtype gene using an uracil PCR with the nucleotides A, G, C, U (no T) which allows 100% incorporation of U in the wildtype gene. Subsequent immobilization on magnetic beads enables easy handling during washing and elution of the antisense strand. Annealing of the mutagenic oligonucleotides and creation of the mutagenic antisense strand was done in a single step using DNA polymerase and a thermo-stable ligase. Addition of an uracil excision enzyme (USER) after strand synthesis completely degraded the wildtype strand and the mutated single strand could be amplified and cloned into a vector.

The abovementioned procedure can be used to create both libraries and individual single mutants alike. To make the latter, automation can be employed and the gene assembly can be done in a 96 well PCR plate. A Gram-positive Staphylococcal display platform (see Paper III) was used to both screen the library and single mutants.
5.4.2 Evaluation of the SPUX platform

Evaluation first consisted of confirming that the wildtype template gene could be created with 100% uracil content prior to immobilization on beads and mutagenic strand synthesis (Figure 5.4.2.1a). The ability of USER to completely degrade the uracil containing wildtype gene implies that carryover wildtype genetic material can be minimized after the mutagenesis process (Figure 5.4.2.1b). The annealing process of the mutagenic oligonucleotides was also evaluated. The oligonucleotides could either be annealed simultaneously or in a sequential manner. In this paper, the NNK degenerate oligonucleotides were designed to mutate one residue per CDR region. Initial evaluation revealed that sequential annealing was capable of mutating 4-6 CDRs, and the occurrence of the wildtype gene was below <0.5% (Figure 5.4.2.1c left). This implied that approx. 12% of the clones had 3 mutated CDRs, 35% had 4 mutated CDRs, 30% had 5 mutated CDRs and 23% had 6 mutated CDRs. Furthermore, sequential annealing revealed that all six CDRs were able to be mutated without a strong bias towards any specific CDR (Figure 5.4.2.1c right). Simultaneous annealing of all mutagenic oligonucleotides showed a lower mutation rate, with the majority of the clones possessing 3 mutated CDRs (data not shown) and a similar distribution mutation frequency across the six CDRs.
5.4.2.1. Evaluation of the SPUX platform. A) Creation of WT scFv gene with 100% uracil (right lane); b) degradation of uracil containing dsDNA with USER (lane 1) versus controls (lanes 2 – 4); c) mutations introduced by sequential annealing of mutagenic oligonucleotides.

5.4.3 Identification of critical binding paratope residues

Random mutation of all six CDRs is a straight-forward approach, but leads to large library sizes that need to be screened. Complete random mutation can also lead to a lower functional population in the library. Therefore, we also evaluated the possibility to create smaller directed libraries which could retain functionality. For this end, the SPUX method was used to create 50 individual alanine mutants to map the paratope of the wildtype scFv. Each mutant had one of the CDR residues mutated to alanine and assessment of these mutants revealed two groups of residues (Figure 5.4.3.1). One group consisted of residue positions which diminished binding when mutated to alanine (diminishing residues, red) whereas the other group consisted of residue positions which enhanced binding when mutated to alanine (enhancing residues, blue).
Figure 5.4.3.1. Paratope mapping of the parental scFv using alanine mutagenesis, revealing diminishing (red) and enhancing (blue) residues.

5.4.4 Creation and sorting of the random and directed libraries

Based on the knowledge from the paratope mapping experiment, three libraries were created: 1) random library where all CDRs were mutated to NNK; 2) diminishing library where six diminishing residues were mutated and 3) enhancing library where four enhancing residues were mutated.

Simultaneous annealing of mutagenic oligonucleotides was used as we believed sequential mutation would be a too aggressive affinity maturation approach. It has been suggested that amino acid substitutions in three to four CDR positions are usually sufficient to improve affinity(407). Mutagenic creation of the three libraries showed desired mutation frequencies (average of 3 CDR amino acid mutations for random library; 2 mutations for enhancing library; 3 mutations for diminishing library) and adequate mutation distribution across the different CDRs (Figure 5.4.4.1), as shown by deep sequencing of the raw libraries.
Figure 5.4.4.1. Amino acid mutation distribution in the raw random (a), enhancing (b) and diminishing libraries (c) following MiSeq deep sequencing.

Flow cytometric sorting of the three libraries was done by incubating library displaying S. carnosus with fluorescently labeled antigen and human serum albumin (to normalize expression levels). 500 nM, 200 nM and 50 nM of labeled antigen was used in the first, second and third round of sorting respectively. From the sorting data, the diminishing library seemed to contain the highest amount of non-functional binders when compared to the enhancing and random libraries respectively (Figure 5.4.4.2).
5.4.5 Assessment of directed library approach using deep sequencing

Thirty individual clones from each library (after sorting) were screened by flow-cytometry. As expected, no clone from the diminishing library was found capable of binding to the target antigen. The clones from the other two libraries were screened and ranked with regards to antigen binding while also taking surface expression into account. Several binders from each library showed an improvement in antigen binding when compared to the wildtype scFv. A subsequent BIACore measurement of the best binder from each library showed a small 5-fold increase in affinity, which could be explained by the relative high antigen concentration used in selection.

In addition to single clone analysis, cells from all raw libraries and cells after each round of sorting were subjected to MiSeq deep sequencing. The idea behind MiSeq deep sequencing was to assess the whether the rationale of creating directed libraries based on paratope alanine scanning was a viable approach. The sequencing data was used to generate a heat map (Figure 5.4.5.1) showing which amino acids at which CDR positions were enriched when compared to the raw unsorted library. This data had a good correlation with the observations from the other experiments. For instance, the wildtype amino acid was heavily enriched (approx. 10%) in the paratope residue positions of the random library. As for the enhancing library, three of four enhancing residue positions in the random library experienced...
enrichment to amino acids other than the wildtype. Furthermore, important supporter positions in CDRL3 were found which were not revealed during alanine scanning. Finally the best clone from the random library had two mutations in the enhancing positions in addition to a mutation in a CDRL3 supporter position.

**Figure 5.4.5.1.** Heatmap of amino acid enrichment after each round of library sorting for A) Random library and B) enhancing library.

In summary, we believe SPUX is a versatile method for creation of both random and site directed affinity maturation libraries in a single step while minimizing the presence of wildtype gene. We further showed the possibility to target all six CDRs simultaneously and selectively. In addition, the SPUX platform is also capable of creating multiple mutations in the same CDR as well as numerous single mutants in an automated and high throughput manner. The SPUX platform was applied for directed affinity maturation of a HER2 binding scFv by characterizing the paratope and creation of a directed library with mutations in non-contacting enhancing residues. Sequencing and sorting of the libraries showed that focused libraries is a viable approach. Finally, we believe the SPUX cloning method can be used for other purposes such as characterization of more therapeutic antibodies for precision medicine (see Paper II) and creation of other non-antibody single- or multi-site saturation mutagenesis libraries(467).
6. Concluding remarks and future outlook

Generation and engineering of (monoclonal) antibodies have come a long way since the dawn of the hybridoma technology. This thesis describes how epitope information and display technologies can be used to generate and characterize antibodies. In Paper I, the epitope information of a non-renewable pAb was used to generate a renewable source of mAb. The mAbs were also revealed to bind a different epitope than the two FDA approved Trastuzumab (Herceptin) and Pertuzumab (Perjeta). It is thus possible to imagine that this approach can be used to generate additional therapeutic antibodies that can be used in combination with existing therapeutic agents. In Paper II, surface display was used to map the epitope of the complement protein C5 towards Eculizumab. The results were able to explain the non-responsiveness of a specific patient group towards Eculizumab treatment due to a mutation in the C5 protein that affected binding to Eculizumab. This highlights the need to properly characterize a protein’s epitope in order to properly guide treatment and for precision medicine to become a reality.

The rise of display technologies has enabled swift and efficient screening of large combinatorial antibody libraries. The various display platforms have enabled the possibility to conduct affinity maturation, de novo selection and engineering of antibodies. This toolbox has been further expanded with the addition of a Gram-positive surface platform for isolation of human scFvs (Paper III). The resiliency of Gram-positive bacteria and fast growth rate allow them to become useful hosts for real time flow cytometric screening and sorting. By using Gram-positive bacterial display as a high-throughput ELISA equivalent in conjunction with phage display, we showed the first example of isolation of human scFv binders from a Gram-positive bacterial display library. As affinity is perhaps the most commonly engineered antibody property, the ability to introduce multiple simultaneous mutations is attractive. Furthermore, creating completely random libraries may result in ill-balanced mutational ratios, targeting of non-desired positions and a theoretical library size too large to be conveniently screened. In Paper IV, a multiple site directed mutagenesis method (SPUX) was introduced which is based on solid phase immobilization of the gene and enzymatic creation of the mutagenic strand and degradation of the original strand. Directed affinity maturation libraries were created based on paratope information of the parent scFv and deep sequencing following selection demonstrated the viability of the approach. The SPUX method not only allows more convenient creation of site directed mutagenesis libraries, but also enables creation of numerous single mutants in an automated and high-throughput manner. We thus envision the possibility for the SPUX method to be applicable in other studies, such as precision medicine related characterization of other therapeutic antibodies in a similar manner as Paper II.

Although this thesis mainly concerns the variable domain of antibodies, there has also been progress in the engineering of the Fc domain. An antibody must not only possess desirable binding characteristics but also needs to be resistant to aggregation and degradation. Many antibody formats have now been created and the numerous ways antibodies can be engineered have shown that antibody engineering is as relevant as ever, in particular as we see more and more of these becoming accessible to patients in trials and treatments.
7. Acknowledgments

Now that we have finally arrived at the end of this thesis, I would like to express my gratitude to the following people.

Foremost, I would like to thank my supervisors Johan R and Mathias U for your excellent guidance and support. Not only did you help me with the details regarding my experiments but also encouraged me when things weren’t going too well.

Obviously, I could not have done what I did without the help and support from my parents and relatives, who also deserve a tremendous amount of gratitude.

I would also like to thank our co-authors at SOBI: Magnus B, Erik N and Patrik S for a productive collaboration in the C5 project. Likewise, I would also like to thank our co-authors at Lund University: Helena P, Anna S and Carl B for your help in providing the material in the staphylococcal display project.

Several times have I run into a dead-end but fortunately John L and Per-Åke were always able to provide me with the advice, so thank you. Other members of the Staph group also deserve a thumbs up for you valuable assistance in Staph-related issues as well as Johan N for your help with SPR related issues.

I would also like to thank both former and current members of the office (Hanna L, Lisa S, Nima, Rezan, Sarah L and Josef H) for turning our office into the most popular room on this floor. Likewise, I would also like to express my gratitude to former and current members of our group (Magnus L, Anna-Luisa, Magdalena M, Ronia, Mona M, Björn F and Niklas T) for your help and group activities.

I would also like to thank Lan-Lan and Emma L for ordering reagents, usually on a short notice as well as Inger and Kristina for all the administrative help. Finally I would like to thank everyone in the HPA MolBio, HPA Protein Factory and everyone here on Floor 3.
8. References


153. Kunik V, Ofran Y (2013) The indistinguishability of epitopes from protein surface is
explained by the distinct binding preferences of each of the six antigen-binding loops. *Protein Eng Des Sel* 26(10):599–609.


151(5):2623–2632.


92


