Engineering of small IgG binding domains for antibody labelling and purification

Sara Kanje

Kungliga Tekniska Högskolan, KTH
Royal Institute of Technology
School of Biotechnology
Stockholm 2016
Till pappa
Our greatest weakness lies in giving up. The most certain way to succeed is always to try just one more time.

-Thomas Edison
Abstract

In protein engineering, rational design and selection from combinatorial libraries are methods used to develop proteins with new or improved features. A very important protein for the biological sciences is the antibody that is used as a detecting agent in numerous laboratory assays. Antibodies used for these purposes are often “man-made”, by immunising animals with the desired target, or by selections from combinatorial libraries. Naturally, antibodies are part of the immune defence protecting us from foreign attacks from e.g. bacteria or viruses. Some bacteria have evolved surface proteins that can bind to proteins abundant in the blood, like antibodies and serum albumin. By doing so, the bacteria can cover themselves in the host’s own proteins and through that evade being detected by the immune system. Two such proteins are Protein A from Staphylococcus aureus and Protein G from group C and G Streptococci. Both these proteins contain domains that bind to antibodies, one of which is denoted C2 (from Protein G) and another B (from Protein A). The B domain have been further engineered to the Z domain.

In this thesis protein engineering has been used to develop variants of the C2 and Z domains for site-specific labelling of antibodies and for antibody purification with mild elution. By taking advantage of the domains’ inherent affinity for antibodies, engineering and design of certain amino acids or protein motifs of the domain have resulted in proteins with new properties. A photo crosslinking amino acid, p-benzoylphenylalanine, have been introduced at different positions to the C2 domain, rendering three new protein domains that can be used for site-specific labelling of antibodies at the Fc or Fab fragment. These domains were used for labelling antibodies with lanthanides and used for detection in a multiplex immunoassay. Moreover, a library of calcium-binding loops was grafted onto the Z domain and used for selection of a domain that binds antibodies in a calcium dependent manner. This engineered protein domain can be used for the purification of antibodies using milder elution conditions, by calcium removal, as compared to traditional antibody purification.

Keywords: Antibody, labelling, purification, Protein G, Protein A, protein engineering, protein design, combinatorial selection.
Populärvetenskaplig sammanfattning


Antikroppar förekommer inte bara i våra kroppar. Eftersom de är så bra på att binda starkt till andra molekyler framställer man idag antikroppar på labbet som binder till det man vill. Till exempel finns antikroppar som kan användas för att binda till tumörceller så att man kan ta bilder på var i kroppen en tumör finns och på så sätt veta var den är när man ska operera bort den eller stråla den. Problemet är att antikroppen är så liten att den inte går att se,

När man tillverkar en egen antikropp på labb görs detta ofta genom att odla celler som man modifierat för att de ska producera ens antikropp. Informationen om hur ett specifikt protein ska byggas upp finns i dess gen. Man kan säga att genen är en ritning för proteinet. Om man sätter in genen för ens antikropp i en särskild typ av celler kan man odla, föröka, de cellerna genom att ge dem syre och näring. Då kommer cellerna producera antikroppen (och en massaa andra proteiner) och spruta ut dem från cellen. Efter någon vecka har man en soppa med sin producerade antikropp och andra proteiner som producerats av cellerna. För att kunna använda sin antikropp behöver man rena den. Man vill alltså göra sig av med de andra proteinerna i soppan, så att man bara har sin antikropp kvar. Då kan man använda ett av de här bakterieproteinerna, Protein A, som binder till nästan alla antikroppar. Om man blandar sin antikropp med Protein A, så binder de till varandra. Då kan man tvätta bort alla de andra proteinerna i soppan, som inte fastnar på Protein A. Efter att man tvättat bort de andra protein-
erna vill man att ens antikropp ska släppa Protein A. Bindningen mellan Protein A och antikroppen är stark, så för att få antikroppen att släppa måste man sänka pH, alltså göra blandningen sur, till ungefär samma pH som en tomat. Vi har framställt en variant av Protein A, som man kan få att släppa antikroppen genom att tillsätta lite mindre syra. Vår variant av Protein A är beroende av att det finns kalcium i blandningen för att kunna binda till antikroppen. Om det finns kalcium i blandningen, så binder Protein A-varianten till antikroppen. Om man istället tar bort kalciumet, så släpper vårt Protein A antikroppen. På så sätt kan man få antikroppen att släppa Protein A på ett lite snällare sätt än att göra det surt som en tomat, och då mår antikroppen bättre.
List of publications

The thesis is based on the work presented in the four articles and manuscripts listed below and referred to in the text by their roman numerals (I-IV). Full versions of the articles can be found in the appendix of the thesis.


III Kanje S.*, Müller L.*, Herrmann A. and Hober S., Next generation of labeling reagents for quantitative and multiplexing immunoassays by use of LA-ICP-MS, Submitted

IV Kanje S., Nilvebrant J. and Hober S., An engineered Protein A derived domain for calcium dependent elution in antibody purification, Manuscript

*The authors contributed equally to this work.

Published papers are printed with the permission from the copyright holders.
Contributions to the included publications

**Paper I**
Performed all experimental work. Designed experiments and wrote the paper together with co-author.

**Paper II**
Performed the experimental work together with first co-author. Designed experiments and wrote the paper together with co-authors.

**Paper III**
Performed all experiments regarding labelling of antibodies. Designed experiments and wrote the manuscript together with co-authors.

**Paper IV**
Performed all experimental work. Designed experiments and wrote the paper together with co-authors.
Defence information

This thesis will be defended on September 30 2016 at 10.00 in M2, Brinellvägen 64, KTH campus, Stockholm, for the degree of "Teknologe doktor" (Doctor of Philosophy, PhD) in Biotechnology.

Respondent:
Sara Kanje, MScEng in Biotechnology
Division of Protein Technology, School of Biotechnology
KTH, Royal Institute of Technology, Stockholm, Sweden

Faculty opponent:
Asst. Prof. Cecília Roque
Chemistry, UCIBIO, Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa, Lisbon, Portugal

Evaluation committee:
Assoc. Prof. Karin Enander
Department of Physics, Chemistry and Biology (IFM)
Linköping University, Linköping, Sweden

Assoc. Prof. Karin Lindkvist
Department of Experimental Medical Science
Lund University, Lund, Sweden

Dr. Mats Inganäs
Gyros AB
Uppsala, Sweden

Chairman:
Prof. Per-Åke Nygren
Division of Protein Technology, School of Biotechnology
KTH, Royal Institute of Technology, Stockholm, Sweden

Main supervisor:
Prof. Sophia Hober
Division of Protein Technology, School of Biotechnology
KTH, Royal Institute of Technology, Stockholm, Sweden

Co-supervisor:
Prof. Amelie Eriksson Karlström
Division of Protein Technology, School of Biotechnology
KTH, Royal Institute of Technology, Stockholm, Sweden
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aaRS</td>
<td>Aminoacyl tRNA synthetase</td>
</tr>
<tr>
<td>ADC</td>
<td>Antibody drug conjugate</td>
</tr>
<tr>
<td>ADCC</td>
<td>Antibody-dependent cell-mediated cytotoxicity</td>
</tr>
<tr>
<td>BCCP</td>
<td>Biotin carboxyl carrier protein</td>
</tr>
<tr>
<td>BirA</td>
<td>Biotin ligase</td>
</tr>
<tr>
<td>BPA</td>
<td>$p$-benzoylphenylalanine</td>
</tr>
<tr>
<td>CDR</td>
<td>Complementarity determining region</td>
</tr>
<tr>
<td>CHO</td>
<td>Chinese hamster ovary</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>Fab</td>
<td>Fragment antigen binding</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-activated cell sorting</td>
</tr>
<tr>
<td>Fc</td>
<td>Fragment crystallisable</td>
</tr>
<tr>
<td>HIC</td>
<td>Hydrophobic interaction chromatography</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish peroxidase</td>
</tr>
<tr>
<td>HSA</td>
<td>Human serum albumin</td>
</tr>
<tr>
<td>ICP-MS</td>
<td>Inductively coupled plasma mass spectrometry</td>
</tr>
<tr>
<td>IEX</td>
<td>Ion exchange chromatography</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>LA-ICP-MS</td>
<td>Laser ablation inductively coupled plasma mass spectrometry</td>
</tr>
<tr>
<td>LOD</td>
<td>Limit of detection</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>NCAA</td>
<td>Non-canonical amino acid</td>
</tr>
<tr>
<td>NHS</td>
<td>N-hydroxysuccinimide</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>POI</td>
<td>Protein of interest</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RPC</td>
<td>Reversed phase chromatography</td>
</tr>
<tr>
<td>SEC</td>
<td>Size exclusion chromatography</td>
</tr>
<tr>
<td>SPPS</td>
<td>Solid phase peptide synthesis</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface plasmon resonance</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer ribonucleic acid</td>
</tr>
</tbody>
</table>
Contents

Abstract .................................................................................. i
Populärvetenskaplig sammanfattning ....................................... ii
List of publications ............................................................... v
Contributions to the included publications ............................. vi
Defence information ............................................................. vii
Abbreviations .......................................................................... viii

Introduction .............................................................................. 1
1 Proteins .................................................................................. 1
  1.1 Antibodies ......................................................................... 3
  1.2 Antibody binding proteins ................................................. 6
  1.3 Calcium binding proteins ................................................. 11
  1.4 Summary ........................................................................... 14
2 Protein engineering ............................................................... 15
  2.1 Rational design ............................................................... 15
  2.2 Directed evolution .......................................................... 16
  2.3 Non-canonical amino acids in protein production .......... 23
  2.4 Summary .......................................................................... 28
3 Protein labelling ...................................................................... 30
  3.1 Non-specific labelling ...................................................... 30
  3.2 Site-specific labelling of antibodies ................................. 32
  3.3 Means for detection in immunoassays ............................ 34
  3.4 Summary .......................................................................... 37
4 Protein purification ............................................................... 38
  4.1 Non-specific purification ................................................ 39
  4.2 Affinity chromatography ............................................... 41
  4.3 Summary .......................................................................... 46

ix
## CONTENTS

**Present investigation**  
1. Aim of the thesis .......................... 47  
2. Site-specific labelling of antibodies (papers I-III)  ....... 49  
3. Calcium-dependent purification of antibodies (paper IV) .... 58  

**Concluding remarks and future perspectives** 63  
**Acknowledgements** 67  
**References** 70
Introduction

1 Proteins

Proteins are the building blocks of life, regulating all functions of our body. They help digest the food in our stomach (enzymes) and send signals through our body to regulate various body functions (hormones). Proteins defend our body from foreign pathogens when we are sick (immune system) and provide structure to our hair and nails (keratin) or tissues (collagen). They regulate all important functions of the body, like when a gene should be transcribed to ribonucleic acid (RNA) that can in turn be translated to a protein, or when a cell should divide into two new cells.

Proteins are encoded by our genes, located on long stretches of deoxyribonucleic acid (DNA), tightly packed as chromosomes in the nuclei of our cells. When a particular protein is required in the body, the gene encoding that protein is transcribed to a messenger molecule called messenger RNA (mRNA), a single stranded copy of the gene, that can be translated to a polypeptide chain. The polypeptide chain can subsequently be folded into a protein. This transfer of information, from DNA via mRNA to the polypeptide chain, is called the central dogma of molecular biology[1] (figure 1a).
**Introduction**

Figure 1: a) The central dogma of molecular biology. DNA encoding our genes is located in the nucleus of the cells in our body. DNA is transcribed to mRNA, a single stranded copy of the gene. The mRNA can then be translated to a polypeptide chain that can fold into a protein. b) Cartoon representation of the different orders of structure of proteins. The primary structure is the polypeptide chain of the amino acids of the protein in the order encoded by its gene. These polypeptide chains fold into secondary structure elements such as beta strands or alpha helices. Secondary elements then order themselves into folded proteins, the tertiary structure. Some proteins consist of several protein domains that organise together in order to form the full protein, this is called the quaternary structure.
The polypeptide chain that is the basis of a protein is made up from amino acids coupled together like beads on a string. This is also called the primary structure of the protein, i.e. the exact sequence of amino acids that is the basis for a given protein. The primary structure folds into ordered secondary structural elements e.g. alpha-helix or beta-strand. Depending on the protein, these secondary elements then fold together to form the tertiary structure of the protein, its final fold. This can be several beta-strands forming a beta-sheet, multiple alpha helices forming an alpha helical bundle or different other combinations of secondary structure elements. For some proteins multiple domains with tertiary structure interact to form the final protein which is called the quaternary structure (figure 1b) [2].

In most life forms there are 20 amino acids encoded for by triplet codons, three nucleotide bases of mRNA, that are the building blocks of the polypeptide chains. All amino acids have an alpha carbon connecting an amino group, a carboxylic acid and a side chain. The side chain is what makes each amino acid unique. Depending on their side chains the amino acids are often grouped into non-polar, polar or charged amino acids (figure 2)[2].

1.1 Antibodies

One of the most important proteins in our body is the antibody. It protects us from disease by binding to things that are foreign to our body, such as viruses and bacteria, and thereby directs the immune defence to clear them from the body. Antibodies are unique in the way they adapt and mutate to bind never before seen molecules with very high affinity and specificity[3]. They are stable proteins that can circulate in the body for a very long time, with a half-life of over 20 days[4].
Figure 2: Overview of the 20 canonical amino acids that are the building blocks of proteins. Depending on their side chain the amino acids can be grouped into non-polar, polar and charged amino acids. The pKₐ of the charged amino acid side chains decides at what pH the amino acid is charged. For example His has a pKₐ of 6 meaning it is uncharged at pH 7, whereas Glu with a pKₐ of 4 and Arg with a pKₐ of 12 are negatively and positively charged respectively at neutral pH.
The great properties of antibodies have been taken advantage of by researchers and now antibodies can be generated towards virtually any target [5, 6]. These "man-made" antibodies are widely used in all fields of biological, biotechnological and medical research. Antibodies are used for detection in various assays and diagnostic tests. They are used as secondary reagents to visualise different target proteins in all types of assays varying from on-cell binding in flow cytometry to complex protein samples in Western Blot. More recently antibodies have emerged as important new therapeutics for various diseases e.g. inflammatory diseases and certain cancers[7].

The most common antibody subclass, both in our bodies and used in the lab, is immunoglobulin G (IgG). It consists of two identical heavy chains and two identical light chains. A cartoon representation of IgG and its different domains and fragments is presented in figure 3. The heavy chain consists of three constant domains (C_H1-3) and one variable domain (V_H). The light chain consists of one constant domain (C_L) and one variable domain (V_L). Together the constant domains make up the constant region of the antibody and the variable domains the variable region. If an IgG molecule is cleaved by the protease papain it is divided into two identical fragment antigen binding (Fab) and one fragment crystallisable (Fc). The Fc part of the antibody is responsible for interacting with other proteins in our body, such as the FcγRIII receptor on natural killer cells, to direct them to a site of infection where they can kill the pathogen. The Fab fragments consist of two of the constant domains, C_H1 and C_L, and the two variable domains. The variable domains each have three hyper variable loops, or complementarity determining regions (CDRs), that together make up the antibody binding site and are responsible for the antibody’s bind-
Figure 3: Cartoon representation of an IgG molecule. The antibody consists of two identical heavy chains (green) and two identical light chains (orange). The different domains (constant, denoted C and variable, denoted V) are marked in the figure as well as the different regions/fragments of the antibody. The CDRs on the variable domains are responsible for antigen binding.

1.2 Antibody binding proteins

In order to increase their virulence and find an escape route around our immune system, certain bacteria have evolved surface proteins that can bind abundant proteins in the blood. By binding to the body’s own proteins, the bacteria goes incognito and can evade the immune system for a longer period of time compared to bacteria lacking these proteins[8, 9]. Some of these antibody binding proteins have been well studied and are used for different biotechnological purposes, most commonly protein purification[10, 11]. Table 1 shows an overview of two well studied antibody binding proteins, Protein A and Protein G, and their degree of binding to different antibody subclasses and species[12].
Table 1: The interaction between Protein A and Protein G with various antibody subclasses and species. Strong binding is indicated by ++, + indicates weak binding while - indicates no binding.

<table>
<thead>
<tr>
<th>Immunoglobulin</th>
<th>Protein G</th>
<th>Protein A</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG1 (human)</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>IgG2 (human)</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>IgG3 (human)</td>
<td>++</td>
<td>-</td>
</tr>
<tr>
<td>IgG4 (human)</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>IgG1 (mouse)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>IgG2a (mouse)</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>IgG2b (mouse)</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>IgG3 (mouse)</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>IgG1 (rat)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>IgG2a (rat)</td>
<td>++</td>
<td>-</td>
</tr>
<tr>
<td>IgG2b (rat)</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>IgG2c (rat)</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>IgG (rabbit)</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>IgG1 (bovine)</td>
<td>++</td>
<td>-</td>
</tr>
<tr>
<td>IgG2 (bovine)</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>IgG1 (sheep)</td>
<td>++</td>
<td>-</td>
</tr>
<tr>
<td>IgG2 (sheep)</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>IgG1 (goat)</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>IgG2 (goat)</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>IgGab (horse)</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>IgGc (horse)</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>IgGT (horse)</td>
<td>(+)</td>
<td>-</td>
</tr>
</tbody>
</table>

Protein G

Protein G is a surface protein found on Streptococci from group C and G. It consists of an N-terminal alanine rich region and sig-
nal peptide followed by three serum albumin binding domains, denoted B1, B2 and B3, three IgG binding domains, denoted C1, C2 and C3 and C-terminal cell wall and membrane anchoring domains[13]. The three IgG binding domains are highly analogous, with C1 and C2 differing by only two amino acids. There is a four amino acid difference between C2 and C3 and a six amino acid difference between C1 and C3[14]. A large part of the work of this thesis was based on the C2 domain why it will be described further in detail, but all IgG binding domains of Protein G have similar properties.

The C2 domain is a 56 amino acid protein with very high affinity for IgG of most species and subclasses (see table 1). It has affinity for both the Fc fragment[15] and the Fab fragment[16] of IgG (figure 4a), but the affinity for the different fragments varies depending on the antibody species and subclass[17, 18]. The C2 domain has the structure of a beta sheet made from four antiparallel beta strands with an alpha helix stacked on top (figure 4b). Moreover C2 is very stable with a melting point (T_m) of >80 °C and refolds easily after denaturation[19, 20].

C2's binding to the Fc fragment takes place in the region between the C_H2 and C_H3 domains, where mainly charged and polar residues on the alpha helix and the N-terminal region of beta strand three on C2 form hydrogen bonds to the IgG (figure 4c)[15]. In the Fab interaction the second beta strand of C2 forms an extended beta sheet with the last beta strand of the C_H1 domain of IgG. This binding is provided by edge-to-edge interactions, stabilised by hydrogen bonds, between the beta strands’ backbone atoms. The C-terminal region of the alpha helix on C2 interacts with the first beta strand of the C_H1 domain. This complex is stabilised by a hydrophobic core formed by buried non polar residues between the two proteins (figure 4d)[16]. The residues on the C_H1
Proteins interacting with C2 are highly conserved and the binding between C2 and Fab is mainly formed by backbone to backbone interactions. This provides an explanation to why the C2 domain has affinity for such a broad range of IgG species and subclasses compared to other antibody binding proteins[21].

**Protein A**

Protein A can be found on the surface of *Staphylococcus aureus*. It consists of an N-terminal signal sequence, five antibody binding domains denoted A, B, C, D and E, each consisting of 58–62 amino acids, and an X region that do not interact with immunoglobulins and is believed to anchor the protein in the cell wall[22, 8]. Each antibody binding domain makes up a tightly packed three helix bundle forming a strong hydrophobic core[23]. Protein A interacts not only with IgG but also weakly with human IgA, IgE and IgM[24, 25].

The B domain of protein A is 58 amino acids and have been further engineered into the Z domain, with the A1V and G29A mutations, that were designed for cloning purposes and to make the protein less sensitive to hydroxylamine treatment, used to selectively cleave Asn-Gly peptide bonds in proteins[26]. The Z domain, like all of the IgG binding domains of Protein A, forms a very stable three helix bundle (figure 4b)[23] and has high affinity (K_D 10 nM) for the IgG Fc fragment[27]. It has a T_m of 75 °C[28] and refolds easily after denaturation. Helix one and two of the Z domain are responsible for IgG interaction where it binds in the region between C_{1\,2} and C_{1\,3} on the Fc fragment through side-chain to side-chain interactions (figure 4e)[29]. This binding surface is close to, but not exactly the same as, the interaction area between the C2 domain and Fc[15, 8].
Figure 4: The C2 and Z domains and their interaction with IgG. a) Cartoon representation of an IgG molecule and the binding sites for C2 (pink) on the Fc and Fab fragment as well as Z (blue) on the Fc fragment. b) Ribbon structure representations of the C2 domain (pink) (pdb-file 1FCC[15]) and the Z domain (blue) (pdb-file 2SPZ[23]). c) Ribbon structure representation of C2 binding to the Fc fragment (pdb-file 1FCC). d) Ribbon structure representation of C2 binding to the Fab fragment (pdb-file 1QKZ[30]). e) Ribbon structure representation of Z binding to (half of) the Fc fragment (pdb-file 1FC2[29] and 2SPZ).
**Other antibody binding proteins**

While Protein G and A are very well known and well studied antibody binding proteins there are other bacteria displaying different kinds of antibody binding proteins. Certain strains of *Peptostreptoccus magnus* express Protein L on their surface. Protein L has five IgG binding domains and can bind certain \( \kappa \) light chains of IgA, IgD, IgE, IgG and IgM[31, 32, 9]. Group A streptococci expresses antibody binding proteins belonging to a supergroup called M-proteins. There are three families of M-proteins, Emm, Enn and Mrp, consisting of similar proteins differing in structure and function. Other than antibodies, M-proteins can also bind additional plasma proteins such as albumin, C4BP protein and fibrinogen [33, 34, 9]. Recently a protein from human mycoplasma denoted Protein M was discovered. It binds to all types of IgG through conserved portions of the \( \kappa \) and \( \lambda \) light chain’s variable region, blocking the antibody-antigen binding[35].

**1.3 Calcium binding proteins**

Calcium regulates many functions in the body. Thus plenty of proteins bind calcium in different ways in order to regulate various calcium dependent mechanisms. These proteins differ in structure and function but have in common their ability to bind calcium ions using different types of domains. The calcium binding proteins contribute to controlling the concentration of calcium in the cytosol, transport calcium ions across cell membranes or are sensitive to local concentration differences in calcium ions, acting as calcium sensors. There are both extracellular and intracellular calcium binding proteins[36].

Extracellular calcium binding proteins are often so called cal-
Introduction

cium sensors that undergo changes in conformation in response to changes of the calcium concentration in the local extracellular environment. These proteins are important in the modulation of vital cellular functions such as complement activation, receptor-ligand interactions and blood clotting[36].

There are a great number of different intracellular calcium binding proteins. A large group of these, including some of the most well known and important calcium binding proteins such as the calmodulin and parvalbumin family proteins, have so called EF-hand domains that bind to calcium[37]. These proteins mediate very important cellular processes including cell cycle regulation, cell division and differentiation and regulation of neuronal function[36]. The EF-hand motif is important for this thesis why it will be presented in further detail below.

Some of the intracellular calcium binding proteins lack the so called EF-hand motif and are important for binding calcium ions in the endoplasmic reticulum and the sarcoplasmic reticulum. These organelles are then (among other things) used as calcium reservoirs within the cell. Other EF-hand lacking intracellular calcium binding proteins are involved in regulating ion channel activity[36].

The EF-hand motif

The EF-hand is a calcium binding protein motif found in numerous calcium binding proteins, both extracellular and intracellular. It is a helix-loop-helix motif of a, most commonly, twelve amino acid calcium binding loop flanked by two perpendicular alpha helices. It forms a structure much like the thumb and index finger of a hand, giving the motif its name (figure 5). The canonical sequence of the EF-hand is around 30 residues and is found in more than 65 protein subfamilies, usually in pairs so that a protein con-
Figure 5: The EF-hand is a helix-loop-helix motif found in many calcium binding proteins. It consists of a twelve amino acid calcium-coordinating loop flanked by two perpendicular helices, denoted E and F, together forming a structure similar to the index finger and thumb of a hand.

contains two, four or six EF-hand motifs[36]. Upon binding to calcium, the EF-hand can undergo conformational change, but this change varies. In some cases this conformational change can lead to protein structure formation from a molten globule state, in others give a protein the ability to bind its target. The sensitivity of calcium binding differs between different EF-hand containing proteins and depend on multiple factors such as the intrinsic affinity to calcium for the loop in question, the co-operation in calcium binding with a paired EF-hand, the selectivity over the chemically similar, and much more abundant, magnesium ion as well as the protein target interaction. It is the sensitivity to changes in the sequence of the twelve amino acids of the canonical EF-hand loop that enables the calcium binding proteins to have such different affinities for calcium ions[37].

In the EF-hand loop, the calcium ion is coordinated by seven ligands, provided by the amino acid side chains, in a pentagonal
Introduction

bipyramidal way. The Ca$^{2+}$ ion has a preference to be coordinated by oxygen, why glutamic and aspartic acid residues are commonly found in the loop. The canonical loop contains nine loop amino acids and three amino acids in the exiting helix. Some of these positions are virtually the same in all canonical EF-hand loops while others are more flexible in regards to what amino acid is found there (table 2). Five of the seven coordinating groups comes from within the nine amino acids that make up the loop (position one, three, five, seven and nine), the other two from the twelfth residue (Glu or Asp). The sixth residue is as good as always a glycine, important for the loop flexibility, and the eighth position is a highly conserved hydrophobic position[37, 38].

Table 2: The amino acid preference of the canonical EF-hand loop[37]. The most common amino acids are presented with the percentage of occurrence in known EF-hand loops. Other frequently occurring amino acids are those found in the given position in more than five percent of known loops.

<table>
<thead>
<tr>
<th>EF-loop position</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most common amino acid</td>
<td>Asp (100%)</td>
<td>Lys (29%)</td>
<td>Asp (76%)</td>
<td>Gly (56%)</td>
<td>Asp (52%)</td>
<td>Gly (96%)</td>
<td>Thr (23%)</td>
<td>Ile (68%)</td>
<td>Asp (32%)</td>
<td>Phe (23%)</td>
<td>Glu (29%)</td>
<td>Glu (92%)</td>
</tr>
<tr>
<td>Other frequent amino acids</td>
<td>Ala</td>
<td>Glu</td>
<td>Thr</td>
<td>Val</td>
<td>Ile</td>
<td>Ser</td>
<td>Glu</td>
<td>Arg</td>
<td>Asn</td>
<td>Leu</td>
<td>Pro</td>
<td>Tyr</td>
</tr>
</tbody>
</table>

1.4 Summary

Proteins are very important molecules for life, built up by polypeptide chains of amino acids. The order of the amino acids for a certain protein is encoded in the DNA by the protein’s gene. In this chapter the proteins and protein motifs important for this thesis i.e. the IgG antibody, the C2 domain, The Z domain and the EF-hand calcium binding loop have been presented.
2 Protein engineering

Sometimes new features of a protein are desired. The protein can then be changed through design or random processes, i.e. engineered, in order to develop a new protein with the desired properties. Engineering of a protein is sometimes done by tweaking its properties little by little by small changes in the amino acid sequence to see if that, fully or partially, renders the wanted effect. It can also be done by drastic changes such as combining two or more proteins with different properties into one multi-purpose protein or by adding entirely new protein domains or motifs to an existing protein.

Often protein engineering is categorised either as rational design, which as the name implies is a more cognitive process based on previous knowledge of a protein and protein chemistry, or as directed evolution, where large pools of protein variants are subjected to different types of selection pressures in order to find one or several variants with the desired properties.

Sometimes the canonical 20 amino acids are not enough to attain the desired effects. There is an array of non canonical, or unnatural, amino acids with various properties. These are generally like a normal amino acid but with a designed side chain, giving the amino acid special properties. Methods to incorporate these amino acids, either chemically with peptide synthesis or in vivo in traditional host cell protein production, have been developed over the years.

2.1 Rational design

In rational design changes to a protein function or structure is made by investigation of known protein features and making changes to
the protein by educated guesses, that is thought to give the protein the desired property. Protein structures, when such have been determined, can be examined in order to figure out what amino acids changes could be done to e.g. destroy a binding site or change the affinity of a protein for its target[39, 40]. This is often investigated by changing an important amino acid to one with a side chain with different properties like a different size, charge or polarity. Another approach is alanine scanning where one amino acid at a time is changed to alanine, that has a sole methyl group as its side chain, in order to see what effect the original amino acid has for the protein’s properties[41, 42].

In recent years semi-rational approaches, using computational methods that simulate protein-protein interactions, have been used in order to suggest designed proteins with the wanted properties[43, 44]. Often these computational design tools are used to propose a pool of protein variants that may be too large to investigate one by one. Therefore this approach is often combined with directed evolution, but with a library with a more focused design, hopefully increasing the chances of finding protein variants with the desired properties[45].

2.2 Directed evolution

In directed evolution large pools, often millions or billions, of protein variants called libraries are subjected to different conditions in order to find one or a few protein variants with the desired properties. This is often described as finding a needle in haystack or trying to fish out the right protein from a massive shoal using the proper bait. The protein pool is expressed using some sort of display system that links the phenotype (the protein characteristics) to the genotype (the genetic code of the protein) in order to be able
Protein engineering
to decode the amino acid sequence of the selected proteins[46]. Di-
rected evolution mimics natural evolution in that the pool of pro-
tein variants is subjected to increasingly harsh conditions in order
to select for the proteins in the pool that have the desired proper-
ties i.e. survival of the fittest. Often these selections are performed
to find target protein binders, that could be used for therapeutic or
diagnostic purposes. The protein library is mixed with the target
protein immobilised on a solid support. Non-binders are washed
away after which binders are eluted and amplified resulting in a
new library pool with an increased number of proteins having, at
least partly, the desired properties. Over the course of a selection,
several rounds of these binding-washing-elution-amplification cy-
cles are performed[47]. Generally the number of washes are in-
creased and the amount of target is decreased in each selection
round, which increases the competition to stay in the final eluted
protein pool. This is called increasing the stringency of the selec-
tion.

Library design
To create a library, the pool of protein variants, certain positions of
a protein are chosen for randomisation. This means that in a par-
ticular chosen position two or more amino acids are allowed. By
allowing multiple amino acids in several positions of a protein the
diversity, i.e. the number of possible variants, quickly increases. If
all 20 amino acids are allowed in five different positions of a pro-
tein, this gives a theoretical library size of $20^5$, 3.2 million, possible
protein variants.

Depending on the purpose of the selection, the library used can
be designed in different ways. When searching for a new bind-
ing protein, randomisations are normally designed to occur where
the binding surface of the molecule is intended. For antibody fragments this generally means introducing randomisations in the CDRs. In so called alternative scaffolds, often small protein molecules that are designed as alternative affinity proteins, amino acids that together form a binding surface are randomised.

Randomisation can be introduced by using so called degenerate codons, allowing more than one particular base of DNA in each position of a triplet codon. Using the codon NNN, where N=A/C/T/G, gives 64 codons that encode for all twenty amino acids, as well as the three stop codons. However due to certain amino acids being encoded for by multiple codons, the NNN codon creates a bias for these amino acids as they are more likely to be encoded for. Therefore when wanting full randomisation the NNK (K=G/T) codon can be more useful as, with 32 possible codons, it encodes all 20 amino acids but without creating as much bias for certain amino acids and only allowing for one stop-codon[48]. These types of degenerate codons can also be designed in ways so that only certain amino acids are allowed in a position e.g. GAW (W=A/T) only encoding aspartic and glutamic acid or RRW (W=A/T, R=A/G) encoding Arg, Asp, Asn, Glu, Gly (x2), Lys and Ser. More sophisticated methods have emerged where entire triplet codons for given amino acids are mixed in desired proportions so that it is possible to allow certain percentages of given amino acids in given positions e.g. 5 % of each of the 20 amino acids in position x or 20 % Ala and 80 % Ile in position y[49].

Sometimes it is not known where mutations are wanted and a more random approach is desired. Then approaches such as gene shuffling or error prone PCR can be used. In gene shuffling DNA from homologous genes is fragmented by restriction enzymes or DNase followed by ligation or rounds of primer-free polymerase
protein engineering

chain reaction (PCR), annealing and extending fragments with sufficient overlap. In error prone PCR random mutations in a gene or a gene segment are introduced by the Taq polymerase’s inherent ability to make mistakes and incorporate wrong bases when replicating a gene. The error rate of the polymerase is increased by adding manganese and increasing the magnesium concentration in the reaction buffer. There is however a bias for the polymerase, where certain mismatches are more common than others. This bias can, at least partly, be decreased by varying the concentration of the different dNTPs in the buffer or by combining polymerases, resulting in less biased mutation rates[50].

Phage Display

Once a library has been established it can be used in a directed evolution selection. The most commonly used method for directed evolution is phage display, first described by George Smith in 1985[51]. In phage display the protein library is fused to a coat protein of filamentous phage, most commonly to protein III (pIII). The phenotype (protein) is then linked to its genotype (gene) which is packed inside the protein displaying phage. When fusing the protein library directly to the pIII gene on the phage genome, all copies of the protein will display the library member. As there are five copies of pIII on the filamentous phage, this will result in multivalent display, generally resulting in binders with lower affinities due to avidity effects in the selection[52]. Phagemid systems, designed to result in monovalent display of the library, i.e. one copy of the library protein per phage, have been developed. In these systems the library is fused to, typically, a truncated version of pIII on a separate plasmid, called a phagemid, that is transformed into Escherichia coli cells. These cells are then infected by so-called helper
phages providing all other proteins and components for filamentous phage production. Helper phages have a defect packing signal directing the production towards phage containing the phagemid genome[53]. Many phages produced using the phagemid/helper phage system will only contain wild type pIII and thus not display any library member[52].

Libraries can be fused also to other phage coat proteins e.g. protein VIII (pVIII). pVIII is the major coat protein of filamentous phage, where about 2700 copies makes up the phage[52]. Only short peptides (six-eight amino acids) can be fused to pVIII, as longer fusions will result in unstable phages[54]. Since there are so many copies on each phage this will cause huge avidity effects during selections, making pVIII a less suitable fusion partner for finding high affinity binders[55].

A typical set up for a phage display selection is described in figure 6. The outcome of each round of a selection is generally measured using titers, i.e. how many phages go into a round of selection and how many come out. Other than titrations it is hard to monitor a phage display selection over the course of a selection. It is generally described as a black box approach, where one cannot know whether the selection has been successful or not before screening the output after the final round of selection. In order to determine if any successful binders have been found the output of the selection is screened for clones with the desired properties, for example by enzyme-linked immunosorbent assay (ELISA), and sequencing to see if there has been any enrichment of certain sequences[56].

In recent years advances have been made using next generation sequencing, providing means to analyse the selection outcome also between rounds of selection as well as circumventing traditional colony picking and screening[57, 58]. Furthermore a method for
continuous directed evolution using phage display, denoted phage-assisted continuous evolution (PACE), have been developed\cite{59}. In PACE multiple rounds of evolution can be performed, without human intervention, in a single day \cite{50}.

**Figure 6:** Overview of a typical selection procedure using phage display. A pool of phage displaying the library variants is mixed with the target molecule on a solid phase. Non-binding phages are washed away and the remaining, binding, phages are eluted, generally by lowering the pH or by trypsination. Eluted phages are amplified by *E. coli* infection followed by superinfection with helper phages providing all necessary protein genes for filamentous phage production. The amplified phages can then be used as input in the next round of selection.
Other display systems

In order to have a better overview of what happens during a selection fluorescence-activated cell sorting (FACS) can be used. Real time events of single clones’ binding to a fluorescently labelled target can be detected and plotted against the surface display level of the clone to sort out the best binding library members. This allows for following the selection over time and see that the portion of desired library members increases over the course of the selection[50]. FACS requires particles large and spherical enough for detection and sorting, why it cannot be used with phage display. This has led to the development of different cell display systems such as yeast[60] and Staphylococcal display[61] as well as droplet based bead surface display[62]. Yeast display is the most commonly used cell display system and has the advantage of being based on eukaryotic cells, providing possibilities for post translational modifications of the proteins, not available in prokaryotic hosts. The drawback of cell display systems is that they are limited by the transformation frequency of the cells, generally resulting in smaller library sizes compared to phage display[50]. Also phage display libraries are dependent on transformation, in that case to \textit{E. coli} cells. Although these generally provide higher transformation frequencies yielding larger libraries than the cell display systems, transformation is still a limiting factor.

In order to circumvent transformation and get even larger libraries alternative, cell free, display systems have been developed. These include ribosome display[63], mRNA display[64] and CIS-display[65]. Typically a pool of library genes are translated into proteins that are linked to their genes, either through mRNA or DNA, by ribosome stalling (ribosome display), puromycin fusion (mRNA display) or RepA binding to the template DNA (CIS-display).
Not only do these systems provide a measure for larger libraries, but as the genes of the binders coming out of each selection are amplified between each round of selection, this can be done with error prone PCR. This provides a mean to introduce additional diversity into the library between each round of selection, further mimicking the natural evolution process.

2.3 Non-canonical amino acids in protein production

Sometimes the 20 canonical amino acids encoded for by the genetic code are not enough to provide the desired properties of a protein. In vivo, amino acids can be post-translationally modified e.g. by phosphorylation or glycosylation[66]. Moreover selenocysteine, a cysteine analogue with a selenium in place of the sulphur, exists naturally in all three domains of life. It is encoded for by the opal stop-codon (TGA) in presence of a selenocysteine insertion sequence in the mRNA molecule[67]. In some methanogenic archaea and bacteria pyrrolysine, a pyrrole containing amino acid, is encoded for by the amber stop codon (TAG) which is suppressed in these organisms by a special transfer RNA (tRNA)[68].

Amino acids with unnatural side chains provide a toolbox to further engineer the properties of a protein, to give them functions not feasible with the natural amino acids. These amino acids can have widely different side chains and can be used for e.g. chemoselective reactions, as fluorescent probes, provide metal-ion binding or be photo reactive. Together these, non genetically encoded, post-translationally modified or unnatural amino acids can be grouped as non-canonical amino acids (NCAA).

An NCAA important for this thesis is $p$-benzoylphenylalanine (BPA), a photo inducible crosslinker that is commonly used to investigate protein-protein or protein-DNA interactions in vivo[69].
Introduction

Figure 7: a) The non-canonical amino acid BPA that has a benzophenone on its side chain. b) Reaction mechanism of the benzophenone side chain of BPA with a methyl group upon long-wave UV irradiation[70].

It has a benzophenone group as part of its side chain (figure 7a) that can be reversibly excited upon long-wave UV irradiation, around 350 nm, and either form a covalent bond to a nearby amino acid (figure 7b) or relax back to its natural state if no reaction partner is nearby[70].

Solid Phase Peptide Synthesis

In solid phase peptide synthesis (SPPS) a polypeptide chain is built one amino acid at a time on a solid phase support. This provides the possibility to incorporate unnatural elements such as NCAAs or D-amino acids and also to make branched peptides. By taking advantage of different protecting groups on the amino acid side chains, a given amino acid can be specifically modified by orthog-
onal removal of their side chain protecting group without deprotecting the rest of the peptide[71]. Normally SPPS is done in a C to N terminal direction, although methods for synthesis in the inverse direction have been reported [72, 73].

In general the synthesis of a peptide is initiated by coupling of the C-terminal residue to the solid support (resin) with an acid labile linker. In the next step the temporary protecting group that prevents the N-terminal amine of the amino acid from reacting is cleaved off. An excess of the next amino acid, with an activated C-terminal carboxyl group, is added to the reaction. The two amino acids can react to form a peptide bond after which any unreacted reagents can be filtered away, followed by washing of the resin. Any amino groups on the resin that were not successfully coupled can be irreversibly capped to prevent deletions in the final peptide. This process is then repeated in cycles for each amino acid of the peptide until the entire peptide is synthesised[74].

The deprotection of the N-terminus of the previous amino acid before coupling must be specific, while not causing the peptide to release from the resin. In the original SPPS publication Merrifield used the Cbz protecting group[75], which was soon replaced by tert-butyloxycarbonyl (Boc). They both require the use of acid with varying strengths to be selectively cleaved off. Today the most common group to protect the N-terminus is 9-fluorenyl-methyloxy-carbonol (Fmoc)[76] that is removed by treatment with base, orthogonal to the acid used for cleaving the peptide off of the resin. Fmoc removal is generally done with 20 % piperidine that provides milder reaction conditions compared to the harsher acids[74]. The C-terminal carboxyl group is activated by an electron withdrawing group, that makes the carbonyl carbon more electrophilic and more prone to react with the amine on the previously coupled amino
acid. The side chains of the amino acids are protected by protecting groups in order to prevent them from reacting during the synthesis which would lead to undesired side reactions[74]. Certain peptides can be prone to aggregate, resulting in a lower synthesis yield. This can be avoided by using backbone protecting groups or pseudoprolines, dipeptides where Ser, Thr or Cys are cyclised to the backbone in a structure similar to proline[71]. Once the entire peptide is synthesised, concentrated acid is used to cleave of the side chain protecting groups, open up cyclised pseudoprolines and cleave the peptide from the resin[74].

A drawback of SPPS is that it is limited by coupling efficiency[71]. Even with a 99 % coupling efficiency in each step, a 60-mer peptide would have a theoretical yield of 55 %. For a 95 % coupling efficiency, theoretical yield would be less than 5 %. In reality the coupling efficiency varies for different amino acids and the primary sequence of the peptide in question, why some peptides will give much lower yields than others. SPPS is thus limited to peptides and small protein domains in order to obtain a reasonable amount of protein.

Recombinant production - expanding the genetic code

Recombinant protein production is the most commonly used method to produce proteins, due to lower costs, higher yields and the ability to produce longer proteins than in e.g. SPPS. The traditional workhorse in the field is *E. coli*, that has been used extensively and where multiple production strains, addressing different drawbacks of the host, have been developed[77]. Quite often proteins are produced in eukaryotic cells, e.g. yeast or mammalian cells, in order to produce proteins with post translational modifications such as glycosylation, or proteins containing disulphide bridges
that can be hard to produce in the reducing environment of the
*E. coli* cytosol[78].

To be able to introduce NCAAs *in vivo*, orthogonal systems for
the translational machinery have been developed. These systems
rely on the introduction of a pair of an orthogonal tRNA and its
 corresponding aminoacyl tRNA synthetase (aaRS) into a host, that
 will genetically incorporate the NCAA in response to a nonsense
codon[79, 80]. By adding the NCAA to the culture medium, it can
 be charged on the orthogonal tRNA by the orthogonal aaRS. The
 orthogonal tRNA then responds to the nonsense codon introduced
 at a specific position in the gene of interest and so the unnatural
 amino acid is introduced to a protein by the host’s natural trans-
 lational machinery (figure 8). Importantly the orthogonal tRNA
 must not react with any of the endogenous tRNA synthetases, i.e.
 it cannot be charged with any of the canonical amino acids of the
 host. Similarly the orthogonal aaRS must be able to charge the
 NCAA, but not any of the canonical amino acids, on the orthogo-
 nal tRNA. Several selection systems have been developed in order
to select for these orthogonal pairs of tRNA/aaRS, each specific for
its own NCAA[80].

For *E. coli* NCAA protein production Schultz and co-workers
developed orthogonal tRNA/aaRS pairs from the archaea *Methano-
coccus janaschii*[81]. *M. janaschii* introduces a tyrosine in response
to the amber stop codon (TAG), the stop codon least used naturally
in *E. coli*. As this particular aaRS has minimal interaction with the
anticodon of its tRNA, it could be mutated in order to accept other
amino acids than its natural tyrosine. Moreover it lacks editing
mechanisms, meaning it will not cleave off the charged NCAA from
the tRNA after acetylation, even though it is not its natural sub-
strate. Pairs for more than 70 different NCAAs have been devel-
oped using this system[79]. As the *M. janaschii* tRNA/aaRS pair is not orthogonal in eukaryotes, other orthogonal pairs have been developed from *E. coli* and certain Methanosarcina to be used in these hosts[80].

More than one NCAAs can be introduced into the same protein, by introducing two different pairs of tRNA/aaRS that responds to different nonsense codons, e.g. one responding the the TAG codon and the other to a quadruplet codon, for which a specifically evolved ribosome that can handle four-base codon translation needs to be used[66].

### 2.4 Summary

Protein engineering can be used to for example enhance the property of a protein or to give it entirely new features. This can be done using rational design, by taking advantage of previous knowledge about a protein, or by more random approaches using directed evolution. Often these approaches are combined to make focused rather than entirely random libraries of proteins to be used in a particular selection. Non canonical amino acids, outside of the normally encoded 20, can be used to further broaden the possibilities of a protein’s application. These NCAAs can readily be introduced into a protein, either by solid phase peptide synthesis, or in recombinant protein production using traditional, pro- or eu-karyotic, host cells.
Figure 8: In the natural translation machinery in a host cell the host’s endogenous tRNA is charged with a canonical amino acid by an endogenous aminoacyl tRNA synthetase. This amino acid is incorporated into the polypeptide chain in response to the tRNA’s anti-codon on the mRNA molecule. By introducing genes for an orthogonal tRNA, which responds to a codon not normally encoding an amino acid in the cell (nonsense codon), and an orthogonal aminoacyl tRNA synthetase that can charge a non-canonical amino acid on the orthogonal tRNA, the NCAA can be introduced into the polypeptide chain in response to the nonsense codon.
3 Protein labelling

Antibodies and other affinity proteins are commonly used as detecting reagents in various assays and diagnostic tests. In order to visualise detection, the detecting protein needs to be labelled, often with a fluorophore or an enzyme. Proteins used for in vivo tumor imaging are normally labelled with radionuclides. Labelling of the protein is done either directly on the affinity protein or indirectly by a secondary affinity protein, that is labelled, binding to the primary detecting protein. There are many ways to label proteins. Most common is to take advantage of certain amino acids’ side chains and couple the labelling agent to them. This often leads to random labelling where the position and number of labels per protein cannot be fully controlled. A number of different methods have been developed over the years to provide means for more controlled labelling, in particular in the field of antibody drug conjugates (ADC).

3.1 Non-specific labelling

Non-specific labelling utilises side chains of certain amino acids for labelling of proteins. Most commonly used are lysines that, together with the N-terminus of the protein, can be labelled on their primary amine via an N-hydroxysuccinimide (NHS) group. The thiol group of cysteine can be labelled using maleimide chemistry \[82\]. Less common ways to non-specifically label proteins are labelling of the carboxyl group of the C-terminus, glutamate and aspartate using N-ethyl-3-N’,N’-dimethylamino-propylcarbodiimide (EDC) chemistry, or conjugation to tyrosines using diazonium salts \[83\].

This type of labelling is generally straightforward and easy to use, however it is not possible to entirely control the level and/or
Protein labelling

position of labelling, resulting in a heterogeneously labelled protein. Lysines are generally found throughout most proteins, why they can be a good target to label. This can be positive as it is possible to obtain several labels on one protein, resulting in a higher signal readout. However, it increases the risk of heterogenous labelling and the risk of labelling the protein in or near its active site or binding surface, potentially damaging the protein function. Cysteines that are naturally present in a protein are generally paired with another cysteine, together forming a disulphide bridge. Thus, in order to label cysteines the protein needs to be reduced, and labelling on one of the cysteines will disrupt the disulphide bridge which will destabilise the protein in question. Many small alternative affinity proteins such as the Affibody[84] and the DARPin[85] do not contain cysteines. In these proteins a cysteine can easily be added genetically to allow for specific labelling.

Biotinylation

A common molecule to conjugate to proteins for labelling is biotin, a small 244 Da molecule also known as vitamin H or B7. It is naturally a cofactor for carboxylase enzymes and present in all living organisms[86]. Its binding to the proteins avidin and streptavidin is one of the strongest protein-ligand interactions known in nature, with an affinity in the femtomolar range[87]. Thus it is convenient to label a protein with the small biotin molecule to then subsequently mix it with streptavidin conjugated to the desired label e.g. a fluorophore or an enzyme, for detection.

Biotinylation can be done either with the labelling techniques described above, or by using a specific tag and enzyme. E. coli biotin holoenzyme synthetase, biotin ligase (BirA), naturally recognises the biotin carboxyl carrier protein subunit (BCCP) where it
Introduction

site-specifically biotinylates one particular lysine[86]. In order for a protein to be biotinylated by BirA it requires at least 75 residues from the BCCP to be fused to the protein. To find a smaller tag that could be specifically biotinylated, through phage display selections, a minimal 15 amino acid peptide that is recognised and site-specifically biotinylated at its lysine by BirA has been developed. It is termed the AviTag or biotin acceptor peptide and has the amino acid sequence GLNDIFEAQKIEWHE. Interestingly this peptide carries little sequence resemblance to the BCCP motif, yet it is biotinylated by BirA with similar $k_{cat}/K_m$[88]. When fused to a protein, normally N- or C-terminally, the AviTag can be biotinylated either in vivo, by co-expressing the AviTag-protein with BirA and adding biotin to the culture medium during protein expression[89], or in vitro by simply mixing the tagged protein with purified BirA and biotin[90].

3.2 Site-specific labelling of antibodies

As antibodies are commonly used, both as means for detection and as therapeutics, labelling of antibodies in a specific and controlled manner without interfering with its function is of great interest. Site-specific labelling, as the name implies, means labelling an antibody in a way so that the exact position of the label is known and controlled. These labelling methods provide means to label antibodies in a manner that do not interfere with the antibody binding site and yield more homogeneously labelled proteins. Methods for site-specific labelling of antibodies can be divided into two groups, those that require mutations of the antibody sequence, and those that can label antibodies directly, without changing the antibody itself.

Methods that require changing the antibody sequence include
introduction of a cysteine that can be labelled by partial reduction without disrupting the normal antibody disulphide bridges, so called THIOMABs (figure 9a)[91]. An alternative is to introduce a selenocysteine that is more reactive than cysteines towards electrophiles at lower pH (pH 5.2), making it possible to specifically label that particular residue via maleimide chemistry[92]. A rather new and popular method for labelling proteins, including antibodies, is sortagging[93, 94]. Sortase A is a bacterial enzyme that naturally attaches proteins to the cell wall of *S. aureus*. It can link a glycine containing label to an LPXTG motif. Sortase A cleaves the LPXTG motif between the Thr and the Gly residue, forming a enzyme-substrate thioester intermediate that can be fused to the Gly-label through a nucleophilic attack (figure 9b)[95]. Furthermore NCAAs that can be chemoselectively labelled, such as *p*-aceto-phenylalanine or *p*-azidophenylalanine, can be introduced site specifically in the antibody and subsequently be labelled using oxime ligation or click chemistry (figure 9c)[96].

When wanting to modify an already existing antibody, or when it is not possible to make a modified antibody, labelling techniques that take advantage of the antibodies natural properties can be used. These methods requires a molecule that has a natural affinity for the antibody, preferably at a site remote from the antibody binding site, and a means to covalently link the molecule to the antibody to that particular site. The methods presented below do this through photo crosslinking, i.e. they contain residues that can form radicals upon UV irradiation and by that form a covalent bond to the antibody. An example of this is indole-3-butyric acid, that can site-specifically cross link to the nucleotide binding site, a conserved region in the variable region of the Fab fragment, when subjected to shortwave UV light of 254 nm (figure
Moreover benzophenones have been introduced to the Protein A derived Z domain as well as the C3 domain of Protein G, so that they can covalently label antibodies on the antibody Fc fragment. The benzophenone is introduced in or near the antibody binding domain’s binding site to the antibody and when subjected to UV light of 365 nm, a wavelength that is generally considered non-damaging to proteins, it can covalently label the antibody. An antibody loaded with a drug or a toxin could provide a very powerful type of medical therapy that could be used in a search and destroy manner in order to specifically kill e.g. a certain cell type (typically a cancer cell). The first ADCs approved for clinical use, Kadcyla and Adcetris are made by traditional labelling of primary amines on the antibody resulting in heterogeneous labelling. Since then, research has shown that homogeneously conjugated ADCs have lower clearance and increased therapeutic index compared to their heterogeneously labelled counterpart, why methods such as the ones described above are the main focus today when making ADCs.

3.3 Means for detection in immunoassays

Often labelled antibodies are used as detecting agents in immunoassays. The antibody must then be labelled, directly or indirectly, with a mean for detection, that can provide a signal readout. Normally this is provided by an enzyme or a fluorophore, but more recent advances have also provided the possibility to detect single elements such as lanthanides.

Enzymes are commonly used to provide signal amplification in immunoassays such as ELISA, Western Blot and immunohistochemistry. Examples include alkaline phosphatase, beta-D-galactosidase
Figure 9: Examples of methods for site-specific labelling of antibodies. a) By introducing a cysteine at a specific position in the Fab fragment, it can be selectively labelled via maleimide chemistry. b) The bacterial enzyme Sortase A can conjugate an introduced LPXTG motif on the antibody to a label with one or several N-terminal glycines. c) Noncanonical amino acids such as p-acetophenylalanine can be introduced to the antibody and be labelled via chemoselective reactions such as oxime ligation, illustrated here. d) The highly conserved nucleotide binding site (NBS) is located in the variable region of the Fab. NBS is rich with aromatic residues why indole-3-butyric acid, that forms a radical upon 254 nm UV radiation, can crosslink to a tyrosine in the NBS.
and the smaller, most commonly used, horseradish peroxidase (HRP). Depending on the substrate these enzymes can produce colorimetric, luminescent or fluorescent readouts[104].

In flow cytometry and fluorescent microscopy different antigens can be detected by affinity proteins coupled to different fluorophores. These are small organic molecules that absorb light at one particular wavelength and emit light at a longer wavelength. Moreover there are smaller fluorescent inorganic nanoparticles, quantum dots, that are brighter and more stable than traditional fluorophores[105]. Additionally there are fluorescent proteins, such as the green fluorescent protein (GFP) and derivatives thereof, that are less harmful to cells than traditional fluorophores and can be used e.g. as a protein expression reporter in living cells[106]. While multiple fluorophores can be combined for multiplex experiments, emission spectral overlap often makes it hard to combine more than eight to ten fluorophores, even when using quantum dots that have narrower emission spectra[107].

In recent years methods based on inductively coupled plasma mass spectrometry (ICP-MS) have provided ways where a wide range of metal ions, metal-loaded polymers and nanoparticles can be detected with high sensitivity and low detection limits. Detecting proteins can be labelled with these metals via bifunctional ligands, combining a reactive group used for conjugation and a metal chelate complex. As the ICP-MS instruments have high spectral resolution, the ability for multiplexing is much larger than with fluorophores. There are more than 100 stable isotopes that can be detected with ICP-MS why in theory 100-plex detection could be performed[107]. Laser ablation ICP-MS (LA-ICP-MS) can be used on solid samples and have been used for multiplex tissue imaging[108] and Western Blotting[109]. Moreover individual cells can be anal-
Protein labelling

ysed with mass cytometry[110] providing a higher degree of multiplexing than traditional flow cytometers. A drawback of this method is that it is destructive to the cells that can thus only be analysed but not sorted[107].

3.4 Summary

Protein labelling is important in order to be able to detect an affinity protein used in a laboratory assay or diagnostic test. Proteins can be labelled with for example enzymes, fluorescent probes or metals. The most commonly used methods for labelling utilise chemically reactive groups to label lysines or cysteines in the protein, providing an easy and efficient way for labelling. However these methods are not fully controllable and may damage the protein or its function, why methods for site-specific labelling of proteins, in particular antibodies, have been developed. In therapy development there is a large focus on conjugating antibodies to toxins or drugs, in order to increase their therapeutic effect or to use the antibody as a homing molecule, delivering a killing agent to a particular site of disease. Research has shown that specific, homogenous labelling improves the pharmacokinetic properties of the therapeutic.
4 Protein purification

Most protein production, besides SPPS, is made by the introduction of a gene of interest to a host cell organism such as the prokaryote *E. coli*, or eukaryotic cells such as the yeast cell *Pichia pastoris* and mammalian cells, e.g. Chinese hamster ovary (CHO) cells and human embryonic kidney cells[78]. This is called recombinant protein production. Often the proteins are produced, soluble, in the cytosol of the cell or in certain cases the periplasmic space of gram negative bacteria. Some proteins are harder to produce and may precipitate during production, forming inclusion bodies that need to be resolubilised[111]. This may also happen when using strong promotors inducing high protein production. In mammalian cell cultures it is common that the soluble protein is secreted into the culture medium and can be separated from its producing cell by centrifugation.

When proteins are produced in the cytosol or periplasm, the cell membrane needs to be disrupted, lysed, after production in order to release the protein content from the cell. This can be done by chemical treatment, sonication or sheer force. Lysis is normally followed by centrifugation leaving a soluble phase, the lysate, containing the proteins and other soluble molecules, and a solid phase of cell debris[112]. Regardless of production path, the protein of interest (POI) normally ends up in a liquid phase together with other proteins, salts and medium components. From this ”soup” the POI needs to be purified, in order to end up with a pure and concentrated protein. Protein purification is normally performed on a column containing a stationary phase, resin, coupled to a ligand with certain desired properties. A general purification process consist of equilibrating the column with an appro-
Protein purification

appropriate buffer, followed by application of the protein sample to be purified diluted in equilibration buffer or in pH adjusted medium. Impurities and molecules that do not bind to the resin are washed away and the remaining protein is eluted by changing the properties of the mobile phase, causing the protein to release from the stationary phase[113].

4.1 Non-specific purification

Traditional purification methods uses general protein properties for separation. In ion exchange chromatography (IEX), proteins are separated based on their charge. It is one of the most frequently used separation techniques as it offers high resolution where it can separate two molecules differing by only one charged amino acid. Based on the charged residues of a protein, every protein has an isoelectric point (pI), the pH where the protein has a net charge of zero. At a pH above the pI, the protein will be negatively charged and conversely at a pH below the pI the overall charge of the protein will be positive. IEX can be based on anion exchange, where the resin is positively charged, or cation exchange where the resin is negatively charged. Binding to the resin is performed at low salt concentrations, where proteins that have the opposite charge of the resin will bind, and proteins with no net charge or the same charge as the resin will not interact and flow straight through the column. Elution is performed by increasing the salt concentration, where the increasing concentration of charged ions will compete with the protein for resin interaction, or by changing the pH, causing the overall charge of the protein to change[114].

In hydrophobic interaction chromatography (HIC), proteins are separated based on their hydrophobicity by a resin coupled with hydrophobic ligands such as linear alkanes, or aryls that provide
both hydrophobic and aromatic interactions. Proteins are subjected to the resin in a high salt containing buffer, favouring retention on the resin, and are eluted by decreasing the salt concentration. Less hydrophobic proteins will be eluted at higher salt concentrations compared to more hydrophobic molecules[115]. Also reversed phase chromatography (RPC) separates proteins based on hydrophobicity but using a more hydrophobic resin. In RPC elution is done by decreasing the polarity of the mobile phase. This is often done under harsh denaturing conditions using low pH buffers and organic solvents[116].

Size exclusion chromatography (SEC), or gel filtration, separates proteins based on their size. The SEC resin consists of porous, inert beads, that do not in themselves interact with the proteins. Smaller proteins or molecules will enter the pores of the beads, giving them a longer path through the column. Proteins that are too large to enter the pores will have a shorter path through the column, between the porous beads, and thus be eluted first. The basic principal of SEC is shown in figure 10. Low resolution SEC can be used for desalting, or buffer exchange, in order to separate large molecules (proteins) from small molecules (salt). By equilibrating the column with the desired storage buffer, the larger proteins will be separated from the smaller undesired salts and eluted into the storage buffer[117].

In a standard purification procedure, most often a combination of these general purification methods will be used, where primary steps may be more crude purifications while purification steps further downstream are polishing steps, making it possible to end up with a desirably pure protein.
Protein purification

Figure 10: The basic principle of size exclusion chromatography (SEC). Porous, inert beads are packed on a column. Smaller molecules and proteins will be able to enter the pores of the beads, resulting in a longer path through the column. Larger proteins that are not able to enter the beads will travel faster through the column and the molecules will thus be eluted in a size dependent manner.

4.2 Affinity chromatography

In affinity chromatography a protein’s intrinsic interaction with its target, or a fused peptide tag’s specific interaction with its affinity protein or a metal ion, is taken advantage of. This provides an efficient mean to obtain pure protein in a single purification step.

Affinity chromatography can be based on the protein’s own affinity for its target e.g. purifying an antibody from sera from an immunised animal on a resin coupled to the target used for immunisation, or purifying Protein A on an IgG column or Protein G on a human serum albumin (HSA) column based on these proteins’ natural affinity for IgG and HSA respectively.

In many cases, purification matrices with a protein’s natural ligand are not readily available, why it is more convenient and common to use tag-based affinity chromatography[118, 119]. In this
case the POI is fused to a tag, normally N- or C-terminally, that has affinity for e.g. another protein or a metal ion, readily available as chromatographic media. In a general affinity purification, a lysate containing the (tagged) POI is added to the column with the affinity resin of choice, to which the tag or POI interacts very specifically. Proteins and other molecules that do not bind the resin are washed away after which the buffer conditions are changed so that the pure POI can be released from the column. An overview of a general tag based affinity chromatography purification is shown in figure 11.

By far the most commonly used affinity purification tag is the polyhistidine tag (His-tag). It consists of two to ten consecutive His residues, where the hexahistidine (His$_6$) tag is the most common[119]. Other six residue tags where every second amino acid is a histidine such as the HEHEHE-tag have also been reported[120]. His-tagged proteins are purified by immobilised metal ion affinity chromatography where a transition metal ion, normally Ni$^{2+}$ or Co$^{2+}$ that can coordinate the His-tag, are immobilised on the resin. His-tagged proteins can be eluted both under physiological and denaturing conditions[121].

The FLAG, c-Myc, HA and HPC4 tags are all small polypeptide tags of 8-12 amino acids that have antibodies recognising them specifically. These antibodies can be immobilised on a resin for purification of proteins fused to these tags, or be used for detecting the tagged protein in various assays. While these methods are very good for obtaining pure protein in a single purification step, having a monoclonal antibody as the capture resin makes these purification tags rather expensive to use [118, 119, 122]. Both the FLAG- and the HPC4 tag have monoclonal antibodies raised against them that can release the protein in a calcium dependent manner, al-
Protein purification

1. Apply sample to purification resin
2. Wash away unbound sample
3. Elute protein of interest

Figure 11: Schematic overview of a tag-based affinity purification. 1. A sample, often a cell lysate, containing the protein of interest (POI) fused to a purification tag, is applied to a resin coupled with an affinity handle. 2. Unbound molecules of the sample are washed away followed by 3. Elution of the protein by disrupting the interaction between the purification tag and the affinity handle.
lowing for mild elution[123, 122]. All the tags described above are quite small and normally do not interfere with protein function, why they are convenient to use and generally do not require removal from the POI after purification.

Larger tags, often based on protein domains, also exist. These can be fused to the target protein to enhance expression and/or increase solubility and folding during production, with examples like the maltose binding protein (MBP), that can be purified on amylose, the glutathione S-transferase tag, purified on glutathione, or the small ubiquitin related modifier (SUMO)[118, 119]. Moreover small affinity proteins such as the Z domain can be used as a tag to purify proteins, in this case on IgG resins[124].

Some tags are based on charge, such as $Z_{\text{basic}}$[125] and the Flag-Acidic-Target Tag (FATT)[126]. These tags are highly positively ($Z_{\text{basic}}$) or negatively (FATT) charged, and can thus be used as very selective tags on IEX media.

Sometimes tags are used in combinations to be used as orthogonal purification handles, allowing for back to back affinity purifications. One domain, allowing for two orthogonal purification steps in the same protein, is the ABDz1 molecule. This small protein domain, based on an albumin binding domain of Protein G, has been selected to have affinity for the Z domain, at the same time as keeping its HSA affinity. Thus the ABDz1 domain can be fused to a protein that can be purified both on HSA and Z resins[127].

If an un-tagged protein is desired as a final product, a protease site can be introduced between the tag and the protein so that the tag can be removed by proteolytic cleavage after purification. Some tags can be cleaved without adding a proteolytic site e.g. the FLAG tag that includes the enterokinase recognition sequence DDDDKX. There are also intein tags that are self-splicing. These have been
engineered further into split inteins in order to avoid uncontrolled self cleaving during protein expression[118, 119].

**Protein A and G based affinity chromatography**

When purifying antibodies, Protein A based chromatography is by far the most commonly used purification technique[128]. For certain antibody species and subclasses as well as for Fab fragments, Protein A purification is not the most suitable choice. In those cases, Protein G based chromatography can be an eligible alternative. Both proteins offer high affinity and selectively for IgG and are used in the billion dollar monoclonal antibody industry[129]. The IgG binding domains of both proteins are highly thermostable and easily refoldable. In both cases, elution of the IgG from the affinity resin is performed by lowering the pH[11]. One reason for Protein A chromatography being more popular than Protein G is that it requires slightly less acidic pH for elution (pH 3.3 as opposed to 2.8)[11, 130]. These harsh acidic conditions may be detrimental for certain antibodies causing them to precipitate, which can result in lower yields, or even to lose their functionality[131].

Several modifications of the IgG binding domains from both Protein A and G have been made in order to alter the binding properties and subsequently the elution conditions. Examples include binding site mutations[132] and exchanging or adding amino acids in the loop between helix two and three in the Z domain which causes a destabilisation of the domain[133], both increasing the elution pH to around 4.5. Methods have been presented where addition of arginine[134] or urea[135] to the elution buffer was used to prevent aggregation and slightly increase the elution pH. Moreover, a thermo responsive Protein A that can elute antibodies by increasing the temperature to 40 °C is on the market[136]. One
of the IgG binding domains of Protein G has been modified with a metal binding site in such a way that it looses its binding to IgG upon addition of transition metals. The loss of binding can be reversed by addition of a chelating agent such as EDTA[137].

### 4.3 Summary

Most proteins used commercially or for research are produced recombinantly in host cells such as *E. coli* or CHO-cells. Once produced, the protein of interest needs to be purified from other proteins and molecules from the host organism. Protein purification is most often done by chromatography, either by utilising general protein properties such as charge or hydrophobicity, or by taking advantage of the protein’s, or an added tag’s, affinity for another protein or metal ion.
Present investigation

1 Aim of the thesis

Paper I

In paper I, "In vivo biotinylation and incorporation of a photo-inducible unnatural amino acid to an antibody-binding domain improve site-specific labelling of antibodies”, the aim of the study was to develop a small biotinylated protein domain that could covalently label IgG antibodies on the Fc fragment. A system for \textit{in vivo} incorporation of BPA and a system for \textit{in vivo} biotinylation was combined so that the rationally designed domain could be produced entirely using recombinant production in \textit{E. coli}.

Paper II

The aim of the study presented in paper II, "Site specific photo labelling of the IgG Fab fragment using a small Protein G derived domain”, was to use rational design to delete the Fc binding of the C2 domain to create an obligate Fab binding domain. Furthermore BPA was introduced at strategic positions to design a protein that could covalently label IgG at the Fab fragment through photo crosslinking.
**Present investigation**

**Paper III**

In paper III the aim of the study, ”Next generation of labeling reagents for quantitative and multiplexing immunoassays by use of LA-ICP-MS”, was to evaluate a new method for labelling antibodies with lanthanide metals, using the C2 domains developed in paper I and II. The antibodies were used in LA-ICP-MS experiments for detection and protein quantification in multiplex Western Blot experiments.

**Paper IV**

The aim of the study presented in paper IV, ”An engineered Protein A derived domain for calcium dependent elution in antibody purification”, was to develop an IgG binding domain based on the Z domain of Protein A that could bind IgG in a calcium dependent manner in order to purify antibodies with a milder elution compared to traditional Protein A chromatography.
2 Site-specific labelling of antibodies (papers I-III)

In order to detect an antibody in an assay such as ELISA or a FACS experiment, it needs to be labelled by a means for detection such as a fluorophore or an enzyme. This can be done randomly by conjugation to e.g. primary amines, or site-specifically by one of the methods presented in chapter three of the introduction of this thesis. Site-specific labelling methods provide routes for a more homogenous labelling and offer a higher control over the amount of label per antibody. Moreover these methods provide means of labelling that do not risk interfering with the antibody binding site.

In paper I and II of this thesis new and improved methods for site-specific labelling of IgG at the Fc and Fab fragment are presented. These methods are based on the introduction of the non canonical amino acid BPA to the C2 domain, that with its benzophenone side chain can covalently crosslink to nearby amino acids upon UV irradiation of 365 nm\[70\. An overview of the labelling strategy is presented in figure 12. In paper III the developed domains were used in an LA-ICP-MS application, where antibodies labelled with lanthanides could be detected in a multiplex Western Blot set up.

![Figure 12: An overview of the labelling strategy used in paper I and II. BPA is introduced at a position near the C2 binding site to the respective antibody fragment. By mixing the antibody with the BPA containing C2 variant and subjecting them to UV light the C2 can crosslink to the antibody via its BPA and thus specifically label the antibody at the Fc or Fab fragment.](image-url)
Present investigation

Rational design was used in both paper I and II to find positions suitable for BPA incorporation. The crystal structures between C2 and the Fc fragment (pdb-file 1FCC[15], see figure 4c) and the Fab-fragment (pdb-file 1QKZ[30], see figure 4d) were investigated to find amino acids close to, and facing, the interaction area between C2 and the IgG fragment but preferably not partaking in the binding themselves. Suitable positions were mutated to BPA either by using SPPS or in vivo incorporation of NCAAs and the mutated domains were tested for their crosslinking abilities to full length IgG and its respective fragments.

A modified C2 domain that has BPA incorporated in position 25 and 42 (numbering according to 1FCC.pdb), C2Fc, was shown to crosslink human IgG at the Fc fragment to >90 % of the heavy chains, resulting in on average almost two labels per antibody (figure 13). Additionally C2Fc was combined with the AviTag in order to produce the entire biotinylated domain in vivo in E. coli.

![Figure 13](image-url) a) SDS-PAGE of the biotinylated C2Fc (odd numbered lanes) showing crosslinking to human IgG (hIgG) at the Fc fragment. C2wt (even numbered lanes) was used as a control. Successful crosslinking results in an extra band above the heavy chain/Fc fragment from the added weight of the covalently linked C2Fc, indicated by an arrow. Crosslinking is seen to >90 % of the heavy chain of human IgG corresponding to on average almost two labels per antibody. b) Crosslinking to full length IgG at the Fc fragment was confirmed with streptavidin-HRP detection in Western Blot.
Before investigating the position for BPA for Fab fragment labelling in paper II, the affinity of the C2 domain for the Fc fragment was deleted. This was done in order to promote C2’s weaker Fab interaction to be able to specifically label the Fab fragment, as well as to avoid aggregation problems that might occur when the protein domain can bind antibodies on two separate binding surfaces[138]. Positions known to be important for Fc binding were mutated in order to design a variant of C2 without affinity for the Fc while still maintaining its binding to the Fab fragment. Several combinations of mutations were made in order to find the right variant. A few candidates, such as C2(N35WD40T), lost their binding to the Fc fragment, while maintaining an affinity for the Fab fragment similar to the wild type C2 domain (C2wt) (figure 14), while certain combinations, like C2(K28AD40A), did not affect the binding to Fc.
The C2 domain containing the Fc deleting mutations N35WD40T, C2_{Fab}, has BPA incorporated at position 18 for crosslinking to mouse antibodies and 29 for crosslinking to human antibodies. Both domains were shown to crosslink IgG at the Fab fragment to 50% of the heavy chains resulting in on average one label per antibody (figure 15).

Figure 15: a) SDS-PAGE of biotinylated C2_{Fab}(T18BPA) crosslinking to mouse IgG1 and IgG2b at the Fab fragment as confirmed with streptavidin-HRP detection in Western Blot (b). c) SDS-PAGE of biotinylated C2_{Fab}(V29BPA) crosslinking to human IgG1, IgG2 and IgG4 at the Fab fragment as confirmed with streptavidin-HRP detection in Western Blot (d).

In papers I and II the photo crosslinking protein domains C2_{Fc} and C2_{Fab} were tested in different types of assays. A HER2-binding
antibody (Herceptin) covalently attached to C2_Fc was used for directed immobilisation on a streptavidin coated surface plasmon resonance (SPR) chip, and compared to the same antibody traditionally labelled with biotin on its primary amines. In theory the antibody labelled with C2_Fc will be immobilised with all molecules facing the same direction, at the same time as the labelling does not interfere with the antibody binding site. The traditionally labelled antibody will have heterogeneously distributed biotins causing the antibody to be immobilised in various directions, where some labels might interfere with the antibody binding site. Both antibodies were immobilised to similar levels via their biotins to a streptavidin chip. As expected the site specifically labelled antibody showed a higher interaction signal indicating that a higher proportion of antibody on the chip was active and able to interact with the target (figure 16).

Figure 16: SPR sensograms of triplicate injections of HER2 interacting with Herceptin immobilised to similar levels on a streptavidin surface via covalently attached biotins. The antibodies were labelled either through traditional NHS-labelling to primary amines (green) or site-specifically with C2_Fc (blue). Herceptin immobilised via C2_Fc shows a higher interaction signal for the same concentration of HER2, showing that a higher proportion of antibody is active on the surface.
Present investigation

Figure 17: ADCC assay investigating whether the therapeutic antibody rituximab labelled with C2_Fc or C2_Fab could still perform effector functions like their unlabelled counterpart. Regardless of labelling, the antibodies were shown to induce ADCC.

In paper II it was investigated whether antibodies labelled with C2_Fab or C2_Fc could still perform effector functions through antibody dependent cell-mediated cytotoxicity (ADCC), by the interaction between the FcγRIII receptor on natural killer cells and antibody bound to a cell surface. Antibodies labelled with C2_Fab or C2_Fc were compared to the unlabelled antibody known to induce ADCC in this assay set up. Regardless of labelling site, the antibodies could perform effector functions (figure 17). The reason that the antibody labelled on the Fc could still perform effector functions via the Fc is thought to be that the FcγRIII receptor binds to a different part of the Fc than where C2_Fc is covalently conjugated.

Both domains were tested in an ELISA set up as detecting antibodies. In paper I, C2_Fc was compared to a traditionally labelled antibody and was shown to provide the same limit of detection
Figure 18: a) ELISA showing detection of HER2 with Herceptin labelled either through its primary amines (Herceptin-NHS) or C2_Fc (Herceptin-C2). While herceptin-NHS provides a higher signal readout in the ELISA, both antibodies show the same limit of detection. b) ELISA showing that the signal output of site-specifically labelled Herceptin can be increased by combining labelling with both C2_Fc and C2_Fab.

(LOD), although, due to a lower labelling degree, providing a lower signal readout than the traditionally labelled antibody (figure 18a). In paper II, C2_Fc was successfully combined with C2_Fab in order to increase the signal readout in ELISA (figure 18b).

In paper III both C2_Fc and C2_Fab were used to label antibodies with lanthanides. These metal ions are often used for detection in multiplex LA-ICP-MS experiments as they are similar in their physical properties and have a low natural background. Traditionally, labelling with these metals, through metal coded affinity tag (MeCAT)[139], is done by conjugation to cysteines which requires reduction of the antibody. Since most cysteines are naturally paired as disulphides, labelling on a cysteine might interfere with antibody stability and will not work with all antibodies. Here, MeCATs loaded with different lanthanides were conjugated
Present investigation

through an introduced cysteine to the C2 domain, which could in turn be used to label antibodies through photo-crosslinking. In the first experiment $\text{C2}_{\text{Fab}}$ was not yet developed why antibodies were labelled using $\text{C2}_{\text{Fc}}$. A mouse monoclonal anti-CYP1A1 antibody was labelled either with $\text{C2}_{\text{Fc}}$ or by traditional cysteine labelling and compared in a Western Blot setup. Both antibodies showed comparable signal-to-noise ratios while the traditionally labelled antibody showed a lower LOD. While $\text{C2}_{\text{Fc}}$ works well for labelling of human antibodies its labelling of mouse IgG1 is sub-optimal, resulting in on average 0.2 labels per antibody, which may explain the higher LOD. This experiment demonstrated that the C2-based labelling technique could be used in an LA-ICP-MS setup. Once the $\text{C2}_{\text{Fab}}$ domain was developed, that gives on average one label per antibody for mouse IgG, it was used to label six different mouse monoclonal antibodies. The chosen antibodies have targets present in liver lysate and were used for detection in a multiplex Western Blot experiment. Four out of six antigens could be simultaneously detected by the labelled antibodies in this setup (figure 19). While all six antibodies provided signal in singleplex Western Blots using secondary anti-mouse-HRP detection, when tested for multiplex detection with secondary antibody in a sixplex setup, only the four antibodies that were detectable in LA-ICP-MS could be detected, confirming the detection pattern seen both in singleplex and multiplex LA-ICP-MS Western Blot detection. The reason that not all six antibodies could be detected simultaneously could be differences in amount of target protein in the lysate or the different antibodies interfering with each other.

Thus, in paper III we show that the methods developed in paper I and II provide an alternative route for labelling of antibodies with lanthanides for LA-ICP-MS, which do not require reduction
Figure 19: 2D intensity profiles form a multiplex Western Blot with LA-ICP-MS detection of the lanthanide metals covalently labelled via C$_{2\text{Fab}}$ to monoclonal antibodies binding proteins in liver lysate. Anti-SDHB(Tm), anti-NAPRT1(Tb), anti-ARG1(Lu), anti-CS(Pr) and anti-SLC27A5(Er) could all be simultaneously detected while anti-STAT3(Eu) was not detected.

of the antibody. This expands the possibilities for labelling of antibodies with lanthanides, in particular for those antibodies that do not work with traditional labelling methods

Conclusions papers I-III

The two domains C$_{2\text{Fc}}$ and C$_{2\text{Fab}}$ provide novel tools for site-specific labelling of antibodies and antibody fragments that have been shown to work in laboratory assays such as ELISA as well as for oriented immobilisation in SPR. C$_{2\text{Fc}}$ gives on average almost two labels per antibody for human antibodies, and C$_{2\text{Fab}}$ gives on average one label per antibody. Combining the two domains render on average three labels per antibody which in paper II was shown to give an expected increase in signal output in ELISA. This research has shown the use of these new labelling domains in areas were traditional labelling may pose problems for certain antibodies and provide a valuable alternative for site-specific labelling without the need for antibody modification.
Present investigation

3 Calcium-dependent purification of antibodies
(paper IV)

Antibodies are routinely used in most laboratories in the biological sciences and therapeutic monoclonal antibodies is a billion dollar industry[129]. As described in chapter four of the introduction, the traditional route for antibody purification is Protein A chromatography. The need for acidic pH around 3 for elution can cause problems with aggregation why alternative purification methods with milder elution is of great interest. In paper IV a variant of the Z domain has, by insertion of a EF-loop library, been designed and selected to bind IgG in a calcium dependent manner. This domain was then used in a purification setup where antibodies could be eluted with EDTA at pH 5.5.

After studying the distances of some known EF-hand loops[140, 141, 142, 143] and the loops of the Z domain[23] a set of protein variants, with a calmodulin loop grafted between helix two and three of the Z domain, were investigated. It was found that the insertion of the loop, with or without an N-terminal glycine linker, destabilised the Z domain in regards to its IgG affinity. Further, these domains showed a difference in T_m in circular dichroism analysis when comparing buffers containing calcium or EDTA. Based on these results, a library of the Z domain was constructed with an inserted randomised loop between helix two and three. The loop library was designed based on knowledge of conserved and flexible positions in the EF-hand loop (presented in chapter one of the introduction). Further destabilisations of the scaffold by randomisation of positions known to stabilise the hydrophobic core[144, 145, 146] were also included in the library. Selections towards human polyclonal IgG were performed using phage display. Alternating
rounds of negative selections in presence of EDTA, to lose variants that could bind under such conditions, with positive selections in presence of calcium and 2 M urea, to select for binders that could subsequently be eluted using EDTA, were performed. The first selection resulted in one binder, denoted A5, that showed a partly calcium dependent binding to IgG in phage ELISA and SPR experiments (figure 20).

A maturation library based on A5 was made with error prone PCR resulting in a library with on average two to three amino acid mutations per gene. This library was used in a maturation selection with phage display, again alternating rounds of negative and positive selections similar to the first selection. In the maturation selection the positive selection had different tracks with urea present, not present or with a decreasing concentration over the rounds of selection. The outcome of the maturation selection were several binders that showed calcium dependent binding to
IgG (figure 21). The selected variants all contained mutations in positions known to be important for IgG interaction such as position K35 and H18[147, 23]. One domain, denoted Zmat8, proved to bind IgG in presence of calcium but not EDTA, shown both in phage ELISA and SPR experiments. Furthermore, in a purification experiment Zmat8 could be eluted from a column packed with IgG sepharose as desired using EDTA at pH 5.5.

Zmat8 and the original Z domain were coupled to NHS activated sepharose on pre-packed columns. The columns were used for test purifications of pure IgG, a CHO cell culture supernatant spiked with IgG and a supernatant from a CHO cell culture producing monoclonal human IgG. Antibodies purified on the Zmat8 column eluted both at traditional pH 3.2 and with EDTA (but not calcium) at pH 5.5 (figure 22a), whereas elution from the Z domain column could only be done at pH 3.2 (figure 22b). Recombinantly produced antibody from a CHO cell culture supernatant could selectively be captured and successfully be purified on the Zmat8 column with elution by EDTA at pH 5.5 (figure 22c).
Figure 22: a) Polyclonal human IgG was captured on a column coupled with Zmat8 and could be eluted at pH 3.2 (orange), and at pH 5.5 with EDTA (green) but not calcium (grey), showing a calcium dependent elution of IgG. b) Polyclonal human IgG was captured on a Z column and could be eluted at pH 3.2 (red) but not with EDTA at pH 5.5 (blue). c) SDS-PAGE (left) showing a molecular weight marker (M), the supernatant (S) from a CHO cell culture producing monoclonal human IgG, the flow through (FT) from purification on the Zmat8 column and the IgG eluted by EDTA (E). Western Blot (right) of the same gel with an anti-human IgG-HRP detection, showing antibody in the supernatant and eluate but not in the flow through.
Present investigation

Conclusions paper IV

Paper IV describes the design and selection of an antibody binding domain denoted Zmat8 that can bind IgG in a calcium dependent manner. This novel protein domain provides a valuable tool for antibody purification with elution conditions less damaging to antibodies than the traditional acidic elution used in Protein A chromatography.
Concluding remarks and future perspectives

Presented in this thesis is the development of novel methods, based on small antibody binding domains, for site-specific labelling and purification of IgG. Protein engineering of the domains C2 and Z resulted in two new tools for site-specific labelling of IgG at the Fc and Fab fragment, and a novel tool for antibody purification.

In paper I, a C2 domain containing the photo inducible amino acid BPA in two positions, C2\textsubscript{Fc}, was developed using rational design. While other Fc labelling domains based on the same or similar approaches have previously been published, C2\textsubscript{Fc}, that yields on average two labels per antibody, showed improved labelling compared to the other domains. Moreover by combining a method for incorporation of NCAAs \textit{in vivo} with the AviTag and co-production with biotin ligase, the entire biotinylated labelling domain could be produced recombinantly in \textit{E. coli}. Even though C2\textsubscript{Fc} works very well for labelling of human antibodies, the labelling of mouse antibodies was not as efficient, crosslinking to only 10 % of the heavy chains. While the C2 domain primarily interacts with the Fc domain of human IgG, its interaction with mouse antibodies is primarily through Fab interactions[18]. It is possible that increasing the affinity for the mouse Fc, through design and/or selection could improve the crosslinking ability of C2\textsubscript{Fc} to mouse antibodies.
Concluding remarks and future perspectives

In paper II the Fc binding of C2 was deleted by using rational design to create an obligate Fab binding domain. Two variants of the domain C2_{Fab} with BPA at different positions were designed. C2_{Fab}(T18BPA) crosslinks to mouse antibodies, and C2_{Fab}(V29BPA) labels human antibodies. Both domains do this to a crosslinking degree of around 50% of the heavy chains, resulting in on average one label per antibody. C2_{Fab}(T18BPA) thus shows improved crosslinking of mouse IgG compared to C2_{Fc}. Yet, in theory it should be possible to render labelling of both heavy chains also through the Fab labelling. In the investigation in paper II most, if not all, possible positions for BPA incorporation for Fab labelling were tested, yet only the two presented here were successful. This might be due to the nature of C2’s interaction with Fab, forming an extended beta-sheet, why most positions close enough for BPA to react with a nearby amino acid on Fab, are also important for the Fab interaction.

Antibody fragments are often used in the development of new antibodies when selecting from combinatorial libraries, and some Fab fragment based therapies are in the clinic. C2_{Fab} is an exciting tool for specific labelling of these fragments, something that was not possible with the previously published Fc labelling domains. Furthermore, learning that an antibody labelled either with C2_{Fc} or C2_{Fab} could still induce ADCC was intriguing. An exciting future prospect would be to test if C2_{Fc} or C2_{Fab} could be used as a method for site-specific drug conjugation to an antibody or antibody fragment.

In paper III the labelling domains developed in paper I and II were used for conjugating lanthanide metal tags to antibodies in order to be used in multiplex immunoassays using LA-ICP-MS. This technique is very promising for immunoassays as it provides
high sensitivity and unparalleled multiplexing. Metal tags used in these types of experiments are normally conjugated to cysteines which can destabilise the antibody and may not work with all antibodies. Antibodies were successfully labelled using both C2\textsubscript{Fc} and C2\textsubscript{Fab} and could be detected in multiplex Western Blot experiments. Furthermore the antibodies could be used for protein quantification in complex liver tissue samples. In the future this labelling method could hopefully be used for antibodies that have not been compatible with other labelling methods. Moreover it could be used in further multiplexing experiments and other types of immunoassays such as immunohistochemistry in LA-ICP-MS.

In paper IV protein engineering was used for design and selection of a variant of the Z domain with a calcium dependent binding to IgG. Not only could this be applicable for milder purification of antibodies but also for proteins fused to an Fc fragment, where the fused protein may be even more sensitive to low pH elution than an antibody. While this is a very promising tool for the antibody purification field, it would be interesting to investigate if the domain could be further developed to elute antibodies in a calcium dependent manner also at neutral pH, providing even milder elution. Further engineering to test and improve the domain for e.g. alkaline stability in order to use it for large scale antibody purification which requires cleaning in place procedures would also be of interest for this great potential domain.

Taken together, the four papers that this thesis is based on present the power of design and selection for engineering of small IgG binding domains to give them new features and by doing so expanding the tool box for antibody labelling and purification. These novel methods should be of high interest to those working with antibodies and to the billion dollar antibody industry.
Concluding remarks and future perspectives
Acknowledgements

Thank you to the ProNova Vinn Excellence Centre for Protein Technology for funding of this research. Thank you to ProNova, Gålös-tiftelsen, Apotekarsocieteten, Stiftelsen ÅForsk, Björns stiftelse and Styffes stiftelse for granting me scholarships that have made it possible to visit international conferences and courses during my time as a PhD student.

Till min huvudhandledare Sophia, tack för att du fångade upp mig i köket på plan3 och bad mig att höra av mig när det blev dags för exjobb. Du kanske inte visste det då, men jag hade redan innan detta en förhoppning om att få doktorera i din grupp. Tack för allt du lärt mig, för att jag fått vara en del av din roliga grupp, för allt stöd och all pepp, i med och motgång, både i forskarlivet och i det privata.

Amelie, min bihandledare, tack för att din dörr alltid står öppen trots att du har mycket på ditt bord, och för all hjälp med frågor om kemi, syntes och HPLC.

Tack Jenny för att du granskade min avhandling och för att du anställde mig i Proteinfabriken en gång i tiden så att jag fick börja min bana på denna fantastiska arbetsplats.
Acknowledgements

Tack till Per-Åke och John för att ni engagerat er i mina projekt, och alltid tar er tid för att svara på mina frågor. Tack också för trevligt konferenssällskap i Boston. Övriga PIs, tack för att även era dörrar alltid står öppna, och för att ni bidrar till den fantastiska stämningen på Plan 3.


Magnus, det är så jävla lyxigt att ha en vän som du och att få träffa varje dag på jobbet!

Tack till alla på plan3 för att ni är med och skapar världens bästa arbetsplats

Tack till mina älskade vänner, för att ni är så jädra grymma, för att ni peppar mig och för att ni håller mig vid mina sinnes fulla bruk.

Till min älskade familj: Matte, Jonas, Danne och Fia jag är sjukt glad att jag får vara er lillasyster, tack för att ni tror på mig och är så allmänt bäst. Mamma, tack för allt stöd, all pepp och att du alltid pushat och trott på mig, men mest för att du är du. Pappa, min stora förebild, tack för att du lärt mig att våga ta plats och inte be om ursäkt för min person. Jag saknar dig varje dag och önskar så att jag hade fått dela denna resan med dig.

Min älskade Viktor, tack för att du alltid intresserar dig, stöttar mig och är världens bästa. All kärlek!

❤️❤️❤️
Acknowledgements
References


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


Sara Kanje, born in Lund, Sweden, has a Master of Science in Engineering in Biotechnology from the Royal Institute of Technology, KTH, in Stockholm. She has spent the last five years at the division of Protein Technology at KTH working on the research presented in this thesis.