Department of Immunotechnology
Lund University

NOVEL BINDING SPECIFICITIES ENGINEERED INTO THE SCAFFOLD OF A CARBOHYDRATE BINDING MODULE

LAVINIA CICORTAS GUNNARSSON

ACADEMIC THESIS which, by due permission of the Faculty of Engineering at Lund University, will be publicly defended in Seminar room, Segerfalkssalen, Wallenberg Neurocentrum, Sölvegatan 17, Lund,
Friday 12th of January 2007, at 13.15 a.m.

Faculty opponent is Professor Arne Skerra, Lehrsthul für Biologische Chemie, Technische Universität München, Germany

LUND UNIVERSITY

ISBN 13: 978-91-628-7033-1
ISBN 10: 91-628-7033-5

Organization
LUND UNIVERSITY

Department of Immunotechnology
BMC D13
SE-22184 Lund, SWEDEN

Document name
DOCTORAL DISSERTATION

Date of issue
2007-01-12

Author(s)
Lavinia Cicortas Gunnarsson

Sponsoring organization

Title and subtitle

Novel binding specificities engineered into the scaffold of a carbohydrate binding module

Abstract

The growing field of biotechnology is in constant need of proteins that can function as recognition tools for separational, analytical and therapeutic purposes. Different molecular engineering approaches are applied on natural proteins in order to create variants with desired properties. This thesis is based on five original papers that deal with selection, characterisation and application of novel binding specificities engineered into the scaffold of a carbohydrate binding module that originates from a xylanase in the thermophilic bacterium *Rhodothermus marinus*. Molecular evolution studies on this scaffold allowed for the generation of variants that bind specifically to the carbohydrate targets xylan, Avicel™, mannan and xyloglucan. In addition, the scaffold employed in this work was also able to adopt specific protein recognition to a human IgG4 molecule. Apart from high binding specificities, the engineered proteins have additional properties such as high thermal stability and ease of production in *Escherichia coli*, which are advantageous in most applications. One of the papers in this thesis demonstrates the potential use of the created xylan-binding variants for detection of their target in wood fibres and plant sections. Also, generated variants with other binding specificities have the potential to find similar use as bioanalytical tools. In conclusion, the scaffold of the carbohydrate binding module engaged in the engineering studies of this thesis proved to be suitable for carrying diversity and has thus allowed for the creation of novel variants with diverse binding specificities useful in biotechnological applications.

Key words
Scaffold, protein engineering, combinatorial library, selection, binding specificity, carbohydrate binding module

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title

ISBN
ISBN 13: 978-91-628-7033-1
ISBN 10: 91-628-7033-5

Recipient´s notes

Number of pages

Price

Security classification

Distribution by (name and address)

Signature Lavinia CG

Date 2006-12-01

CONTENTS

ABBREVIATIONS

AE	affinity electrophoresis
CBM	carbohydrate binding module
CD	circular dichroism
CDR	complementarity determining region
DARPins	designed ankyrin repeat proteins
DNA	deoxyribonucleic acid
ELISA	enzyme-linked immunosorbent assay
Fab	antibody binding fragment
FACS	fluorescently activated cell sorting
FIND	fragment induced diversity
ITC	isothermal titration calorimetry
mRNA	messenger RNA
pIII	phage-coat protein III
pVIII	phage-coat protein VIII
PCR	polymerase chain reaction
PDB	protein data bank
RACHITT	random chimeragenesis on transient templates
RNA	ribonucleic acid
scFv	single chain antibody fragment variable
VH	variable heavy chain
VL	variable light chain

ORIGINAL PAPERS

This thesis is based on the following papers, which are referred to in the text by their Roman numerals (I-V).

I. A carbohydrate binding module as a diversity-carrying scaffold
Cicortas Gunnarsson L., Nordberg Karlsson E., Albrekt A.-S., Andersson M., Holst O. and Ohlin M.
Protein Eng Des Sel (2004), 17, 213-221

II. Engineered xyloglucan specificity in a carbohydrate-binding module
Cicortas Gunnarsson L., Zhou Q., Montanier C., Nordberg Karlsson E., Brumer 3rd H. and Ohlin M.
Glycobiology (2006), 16, 1171-1180

III. Evolution of a carbohydrate binding module into a protein-specific binder
Cicortas Gunnarsson L., Dexlin L., Nordberg Karlsson E., Holst O. and Ohlin M.
Biomol Eng (2006), 23, 111-117

IV. Novel xylan-binding properties of an engineered carbohydrate-binding module
Cicortas Gunnarsson L., Montanier C., Gilbert H.J., Tunnicliffe R.B., Williams M.P., Nordberg Karlsson E. and Ohlin M.
(manuscript)

V. Synthetic xylan-binding modules for mapping of pulp fibres and wood sections
Filonova L., **Cicortas Gunnarsson L.**, Daniel G. and Ohlin M.
(manuscript)

1. INTRODUCTION

Proteins that can serve as specific recognition tools are needed in many biotechnological applications including bioprocessing, diagnostics and therapy. In order to find binders that fulfil the application demands on binding specificity, molecular stability and additional properties, engineering approaches are often applied on natural proteins. For this purpose, the traditionally employed antibody scaffold has lately been complemented with other alternative scaffolds, listed by Binz *et al.* (2005). Large molecular libraries are created by introducing changes into the protein sequence, which allows for the selection of variants with desired properties.

Although the antibody scaffold is able to harbour diversity that enables recognition of a range of ligands, including low-molecular weight molecules, DNA, peptides and proteins, generation of antibodies specific towards polysaccharides has been a challenging task (Willats *et al.*, 2000). The relatively low number of antibodies engineered for specific oligosaccharide recognition further indicates the difficulty associated with the creation of specific carbohydrate binders. Carbohydrates are not only major constituents in plants but many different glycans can also be attached to proteins (glycoproteins), changing their biological properties and specific glycosylation patterns can often be associated to different diseases. Specific binders towards carbohydrate molecules would allow for the study of plant material composition and could also serve as recognition markers in diagnostics. At the time of initiation of this project, there were to my knowledge no reports on specific binders towards such targets engineered into any other alternative scaffolds.

In Nature there are proteins specialised for binding to carbohydrates including the carbohydrate binding modules (CBMs) most often found attached to glycoside hydrolysing or modifying enzymes. The molecular scaffold of such a molecule is naturally fitted to recognise carbohydrate targets and therefore we considered it to constitute a promising starting point for engineering novel carbohydrate binders. This thesis, which is based on five original papers (I-V),

deals with molecular engineering of a CBM-scaffold (PAPER I and II) and characterisation (PAPER III and IV) and application studies (PAPER V) of selected CBM-variants.

The scaffold investigated in our studies was that of CBM4-2, which has been identified as one of two CBMs in the modular thermostable xylanase Xyn10A from *Rhodothermus marinus* (Nordberg Karlsson *et al.*, 1997). In PAPER I we showed that this scaffold is well suited for carrying the diversity introduced by our combinatorial library design. Engineered CBM-variants were selected for binding to plant cell wall carbohydrates, such as xylan, glucan, mannan and xyloglucan (PAPER I and II) and some for binding to a human IgG4 molecule (PAPER I). The latter ones were found to be very specific for the glycoprotein target and in PAPER III we demonstrated that they bound to the protein structure itself and not to a carbohydrate structure on the immunoglobulin molecule. This finding further extended the potential use of the CBM4-2 scaffold for engineering not only carbohydrate specific but also protein specific binding reagents.

Characterisation studies in PAPER IV revealed the generation of a CBM-variant with higher binding specificity for the polysaccharide xylan as compared to the wild type CBM4-2. This module had developed an original mode of binding as compared to other xylan binding CBMs, which only involves one aromatic residue that can form the important stacking interaction to the sugar rings of xylan. Engineering work on the CBM4-2 scaffold has thus allowed for the evolution of binding features beyond those found in Nature.

In PAPER V, CBMs selected for binding to xylan and the wild type CBM4-2 were investigated for specific detection of xylan in pulp fibres and wood sections. This study confirmed that CBMs generated by the selection of the CBM-library on different targets could function as valuable molecular probes for analysis of the carbohydrate composition in plant materials. In addition to proper target recognition, advantageous properties of the wild type CBM4-2, such as high thermal stability and ease of production in *Escherichia coli*, that were preserved in the engineered CBMs (PAPER I) constitute important characteristics for their use in various applications.

2. PROTEINS AND CARBOHYDRATES

Proteins are very complicated molecules essential to the structure, function, and regulation of all living organisms. The DNA of each cell holds genes that encode for the appropriate amino acid sequences of proteins and thus for their structures and functions. The building blocks of proteins are twenty unique amino acids that can be arranged in many ways to yield proteins of different sizes, up to thousands of amino acids long, and also different structures. We know today that the human genome contains about 25,000 different genes (Pennisi, 2005), somewhat less than the first figures reported from the human genome project (Lander *et al.*, 2001; Venter *et al.*, 2001). Taking into consideration the alternative modes of processing gene transcripts and the posttranslational modifications that occur in the body, the number of functioning human proteins is expected to be significantly higher, more than one million (Wingren and Borrebaeck, 2004). In addition, there are many other organisms with their own unique sets of proteins, which give an immense number of molecules that belong to the diverse group of natural proteins.

Increased accessibility to genomic sequence data and structural information has led to a growing understanding of the relationships between sequence, structure and function of proteins. All types of vital processes like molecular signalling, enzymatic degradation, intra- and inter-molecular transportation, involve proteins (e.g. hormones, enzymes, and antibodies) interacting with other molecules, not necessarily proteins. So whether binding is their main function or not, all proteins have built in recognition of other molecules.

Proteins with particular binding characteristics are of great need in the growing field of biotechnology. Apart from biomedical utilization in diagnostics or therapy, there are several areas where proteins are used for targeting, purification and/or immobilisation of different biologically active molecules. Complementary to suitable binding, the proteins often have to fulfil other application requirements such as size limitation, tolerance to high or low pH, high temperature stability and ease of production and purification. The lack of

natural proteins that match all these criteria, has forced scientists to engineer novel binders. The binding characteristics of a protein molecule and especially the achievement of modified binding properties through protein engineering are essential for this thesis.

Like proteins, carbohydrates constitute a complex group of molecules. These are the most abundant organic molecules in Nature and include monosaccharides such as glucose, xylose and fructose, oligosaccharides that are built up by a few monosaccharides and polysaccharides, or glycans, which consists of many residues joined by glycosidic linkages. Furthermore, carbohydrate chains can be branched and/or substituted yielding matrices that are either crystalline or amorphous. Protein-carbohydrate interactions, the main topic of the thesis, play a vital role in many biological processes such as host-pathogen recognition, cell-cell communication, protein trafficking and microbial recycling of carbon from the plant cell wall. Carbohydrates have also great importance in life as structural constituents in plants and through glycosylation, carbohydrates affect the biological properties of both proteins and lipids. Studies of the localisation and distribution of carbohydrates in plant cell walls or determination of the occurrence and extent of protein glycosylation typical for certain diseases, as well as several other areas of research, are in need of proteins that specifically recognise carbohydrates. In Nature there are groups of specialised carbohydrate-binding proteins like the periplasmic sugar transport proteins, lectins from both the plant and animal kingdom that bind to mono- and oligosaccharides, and plant, fungal or bacterial CBMs that can recognise most types of carbohydrates. The CBM4-2 module engaged in the work of this thesis has a natural capacity for interaction with various carbohydrates molecules and through molecular engineering, variants of this protein module were generated that only recognise specific target ligands (PAPER I, II and IV).

3. CARBOHYDRATE BINDING MODULES

CBMs are non-catalytic proteins most commonly representing parts of modular glycoside hydrolysing or modifying enzymes found in Nature. The CBMs contain 30 to about 200 amino acids and exist as single, double, or triple domains of an enzyme but also free as the recently reported small olive pollen protein, Ole e 10 (Barral *et al.*, 2005). The first evidences of existence of these binding proteins were reported about 20 years ago (Van Tilbeurgh *et al.*, 1986; Gilkes *et al.*, 1988; Tomme *et al.*, 1988) and still there is much left to explore about these proteins specialised in binding to carbohydrates. Currently, several hundreds of CBMs have been identified and classified into 46 sequence-related families (http://afmb.cnrs-mrs.fr/CAZY/). The binding specificity of CBMs differs greatly between and also within these families (Figure 1) and some individual modules may even show affinity for more than one type of carbohydrate target (Boraston *et al.*, 2001a; Charnock *et al.*, 2002; Najmudin *et al.*, 2006) which is also the case for the wild type CBM4-2 (Abou Hachem *et al.*, 2000) described in this thesis.

The structural biology of CBMs is of great interest since it helps researchers to understand the mechanism by which CBMs bind to their ligands. Progresses in this field have been nicely summarised in a recent review by Boraston *et al.* (2004) and it is not the scope of this thesis to repeat that information. However, in order to better comprehend the structure-function relationships for the CBMs studied in this thesis, an effort will be made to describe the principal binding characteristics of this group of protein modules. In addition, an overview of their biological role as well as current applications of CBMs will be given.

3.1 Structure-function relationships of CBMs

3.1.1 Binding site topography and ligand conformation

Structural data is available for representative members of 34 CBM families and in some cases not only for the protein molecule but also for its complex with the carbohydrate ligand. Three-dimensional structures of CBMs have shown that the topography of the binding sites in these proteins reflects the nature of the target molecule. Consequently, another classification of CBMs into types A, B and C has been proposed (Boraston *et al.*, 1999). Type A CBMs have a planar architecture of the binding site, which is thought to be complementary to the flat surfaces of the insoluble crystalline glycans like cellulose and chitin (Tormo *et al.*, 1996; McLean *et al.*, 2000; Simpson *et al.*, 2000). The most common class of CBMs is that of type B, the members of which bind a wide range of single soluble carbohydrate chains into grooves that extend along the entire length of the proteins. CBMs of type C contain small binding sites that can fit mono-, di- or tri-saccharides.

CBM4-2 of *R. marinus* used in the studies of this thesis is a type B module and therefore more focus will be put on this type of CBMs in the proceeding of this chapter. CBM4-2 has a β-sandwich fold (Figure 1) (Simpson *et al.*, 2002), which is the dominant fold among CBMs and comprises two β-sheets, each consisting of three to six antiparallel β-strands. Loop structures in concert with the orientation of the side chains of aromatic residues (see below) determine the topography of the binding site. The shape of the binding site of CBMs mirrors the conformation of the target ligand leading to a reduced cost in energy upon binding otherwise required for achieving a good fit (see chapter 3.2).

The conformation of oligosaccharides or polymeric carbohydrate chains depends on the monosaccharide components of the backbone as well as the type of the glycosidic bonds (commonly, α-1,4; α-1,6; β-1,3; β-1,4 or β-1,6) that link these monomers. Further substitutions of the backbone sugar rings are also of importance for the flexibility of the chain and have thus an impact on the conformation. These factors plus the environmental influence, such as hydrogen bonding with water molecules in the solution, lead to carbohydrate chains adopting a planar, helical, twisted or U-shaped conformation that needs to be fitted into the binding site of the CBM (Figure 1).

Although not the only binding determining factor, the binding site topography complementary to the ligand conformation is vital for the formation of proper binding interactions between the CBM and its target molecule. Structural features of CBM4-2, such as positioning of important interacting residues at the tips of highly exposed loops, suggest that this binding site is more flexible than most other CBMs (Simpson *et al.*, 2002). This might explain its unusual capacity to bind both to xylan and to amorphous cellulose, although the latter one with much lower affinity (Abou Hachem *et al.*, 2000). Some of the mutations introduced in the CBM4-2 scaffold in PAPER I might have led to a reduced flexibility of the binding site and an increased binding specificity of engineered CBMs. This hypothesis will have to be confirmed by future structural studies of the CBMs selected in PAPER I and II.

The high degree of shape complementarity described in this chapter is not an unique feature of CBMs but has been seen also for other types of protein-target interactions like those between an antibody and its antigen (Webster *et al.*, 1994). It can be argued though that the importance of binding site topography is higher in the interaction between a protein and a carbohydrate due to the nature of the latter. The sugar residues in a carbohydrate molecule have more limited possibilities to establish interactions than the diverse amino acids have in peptides or at protein surfaces. Similarly, binding to low molecular weight targets has a greater demand on the binding site topography, which should be of a cavity shape like that of lipocalins and related proteins (Flower *et al.*, 1993).

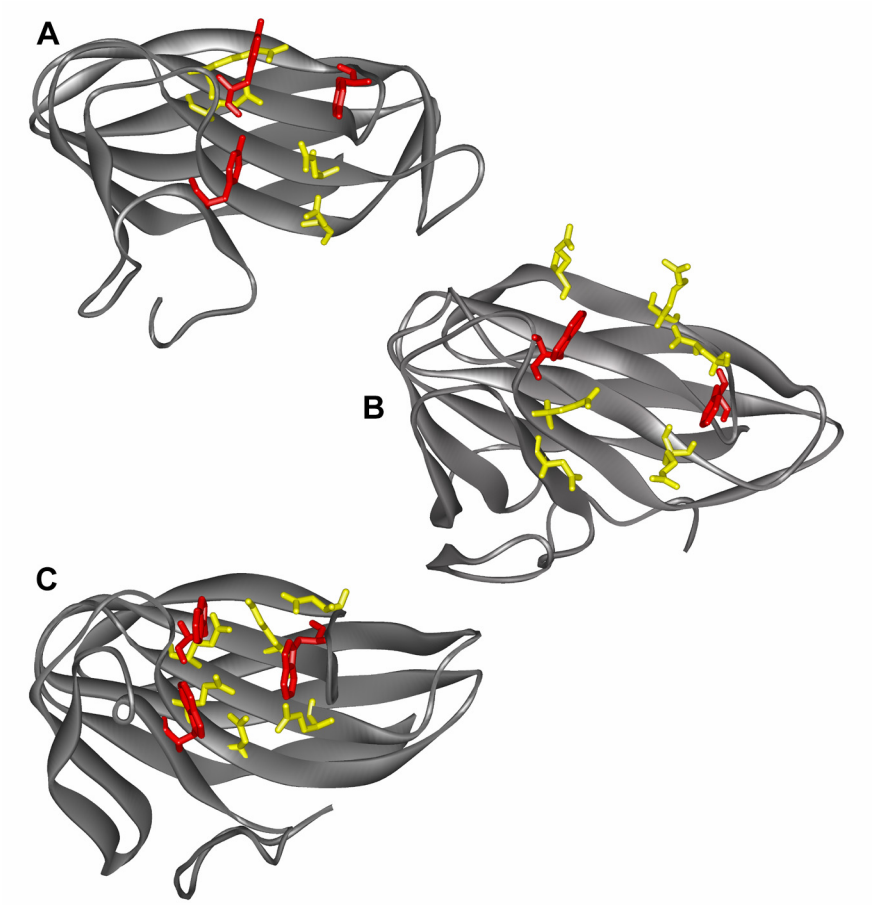

Figure 1. Structures of three CBMs that belong to family 4 and display different binding specificities. They have a similar β-sandwich structure, but different binding site topographies determined by the side chains of interacting amino acids. Aromatic residues are highlighted in red and hydrogen bonding residues in yellow. (**A**) CBM4-1 of endo-glucanase Cel9B from *Cellulomonas fimi* is specific for the linear chain of β-1,4-glucan (soluble cellulose) (PDB code 1ULO). (**B**) CBM4-2 of xylanase Xyn10A from *R. marinus* binds preferentially to xylan, which has a helical structure (PDB code 1K45). (C) CBM4-1 of laminarinase Lam16A from *Thermotoga maritima* binds to the U-shaped β-1,3-glucan (laminarin) (PDB code 1GUI).

3.1.2 Aromatic residues

The aromatic side chains of tyrosine, tryptophan and, less commonly, phenylalanine are believed to form hydrophobic stacking interactions with the sugar rings of the target ligand. Replacement of such amino acids at the binding site of a CBM has been shown by us in PAPER IV and by others (Nagy *et al.*, 1998; Ponyi *et al.*, 2000; Pell *et al.*, 2003) to reduce or completely abolish the affinity for the carbohydrate target suggesting that aromatic residues play a pivotal role in the CBM-carbohydrate recognition.

The orientation of aromatic side chains is critical for both binding affinity and specificity of the CBM. This was evident in the case of the xylan-binding CBM2b of *Cellulomonas fimi* xylanase 11A, which was converted into a cellulose binder by introducing a mutation that resulted in the reorientation of the side chain of a tryptophan residue (Simpson *et al.*, 2000). Consequently, the tryptophans of the mutated CBM-variant formed a planar hydrophobic surface, which is typical for the type A CBMs that bind to the flat crystalline cellulose (Figure 2A). Aromatic residues at the binding sites of type B CBMs are commonly situated on each side of the binding cleft (Figure 1) forming a sandwich together with the bound ligand chain as in the case of CBM4-2 from *R. marinus* Xyn10A (Figure 2B). However, the xylan-binding CBM15 derived from *Pseudomonas cellulosa* xylanase Xyn10C and most other members of this CBM family have the tryptophans located on one side of the binding cleft leading to formation of stacking interactions with only one face of the ligand chain (Figure 2C) (Szabo *et al.*, 2001). These data demonstrate the importance of aromatic residues in carbohydrate binding sites in general. Findings in PAPER IV nevertheless define that also a reduced number of aromatic residues, like in the case of X-2 module where only one such residue is found in the binding site (Figure 2D), has the potential to mediate substantial high affinity interaction with polymeric substrates. *In vitro* evolution can thus find solutions to establish a protein-carbohydrate interaction different from those commonly used by Nature.

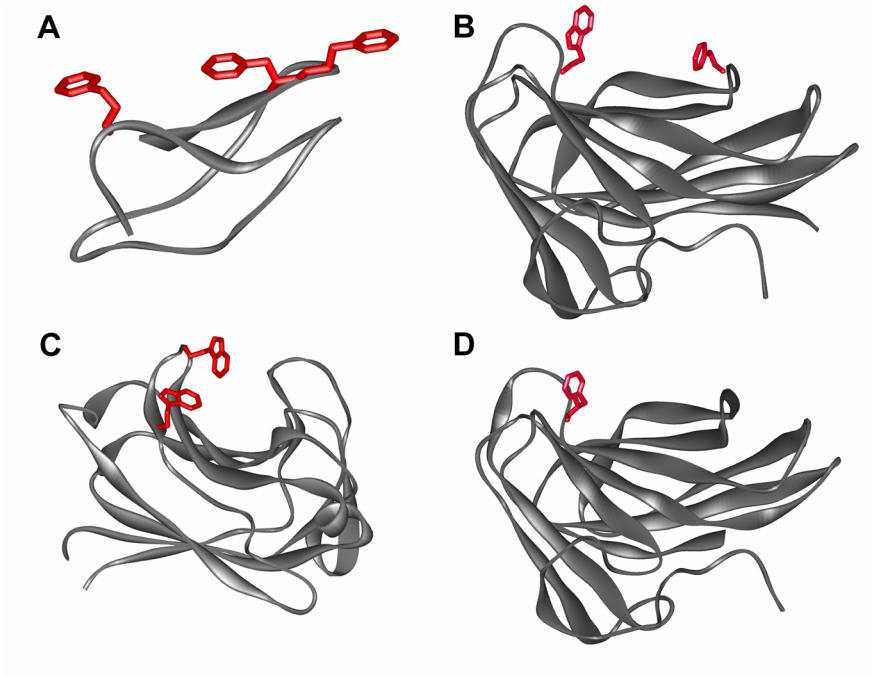

Figure 2. Structures of CBMs displaying variations in number, positioning and side chain orientation of aromatic residues at the binding site. (**A**) CBM1 of cellobiohydrolase CBH1 from *Trichoderma reesei* (PDB code 1CBH) has three tyrosine residues that form a planar surface for interaction with crystalline cellulose. (**B**) CBM4-2 of *R. marinus* Xyn10A (PDB code 1K45) has two aromatic residues displayed on each side of the cleft entrance forming a sandwich with the bound in ligand. (**C**) CBM15 from xylanase Xyn10C of *Cellvibrio japonicus* (PDB code 1GNY) has two tryptophans positioned on one side of the binding cleft and their aromatic rings make an angle close to 120° to match the helical structure of xylan. (**D**) A model of X-2 module (obtained using CPHmodels 2.0 at http://www.cbs.dtu.dk/services/CPHmodels/) shows only one aromatic residue present at the binding site of this xylan-specific CBM.

3.1.3 Hydrogen bonding

Although not as important as the aromatic residues, hydrogen bonding amino acids in the binding sites of CBMs have been shown to play a significant role in ligand binding. Mutation to alanine of polar residues involved in hydrogen bonding had only minor effect on the binding affinity of type A CBMs (McLean *et al.*, 2000), whereas larger decreases in binding affinity (Notenboom *et al.*,

2001; Pell *et al.*, 2003) and even abolished binding (Xie *et al.*, 2001b) have been observed for type B CBMs. Hydrogen bonds present in deep binding clefts seem thus to have larger impacts on the binding affinities of CBMs as compared to those present on more flat binding sites.

The exact position of the polar amino acids in the binding site is also of vital importance for the ligand specificity of CBMs since these residues participate in defining the ligand complementary shape of the binding site (Figure 1). Hydrogen bonds contribute to the fixation of the carbohydrate ligand into the binding cleft explaining why both the binding enthalpy and entropy (see below) are affected by the replacement of direct bonding residues.

3.2 Thermodynamics of ligand binding of CBMs

Protein-carbohydrate binding is a complex process that requires correct orientation of the side chains of the aromatic residues, the formation of direct interactions such as hydrogen bonds, and a ligand complementary shape of the binding site. Precisely how these factors, individually or together, influence the thermodynamics of the interaction is still not completely understood. The association constants for the ligand binding of CBMs are determined by both enthalpic and entropic changes associated with binding. Enthalpic changes are connected with polar interactions such as ionic or hydrogen bonds and van der Waals interactions while entropic changes arise from alterations in the translational and rotational freedom of the protein and/or carbohydrate in addition to the solvent reorganisation (Gilbert *et al.*, 2002). These changes are closely related and compensate each other as can be seen in the case of removal of a hydrogen bond. The loss in binding energy, resulted from a lower change in enthalpy, is then compensated by the favourable change in entropy linked to the higher freedom of the ligand in the absence of the hydrogen bond (Xie *et al.*, 2001a).

In accordance to what was discussed in chapter 3.1, the nature of both protein and carbohydrate has an impact on the type of interaction that is established and consequently on the kind of energy involved in the binding process. A fixed conformation of the crystalline cellulose leads to a rather low

entropic loss upon binding to type A CBMs. Furthermore, the large increase in entropy gained from dehydration of the CBM-ligand contact surface makes this type of binding entropically driven (Creagh *et al.*, 1996). Binding to type B and C CBMs is on the other hand enthalpically driven, meaning that the energy evolved by direct interactions between the CBM and the target ligand has to be greater than the entropy cost associated with CBM binding to soluble, flexible carbohydrates (PAPER IV) (Charnock *et al.*, 2000; Boraston *et al.*, 2001a; Notenboom *et al.*, 2001; Boraston *et al.*, 2002; Pell *et al.*, 2003). The decrease in entropy upon binding to a single carbohydrate chain suggests that conformational changes occur in either the ligand or the protein or both. These changes serve several purposes such as allowing the essential residues to be properly oriented for binding and may occur also in other types of carbohydrate binding proteins like the periplasmic sugar binding proteins (Quiocho, 1986).

The wild type CBM4-2 and the xylan-specific X-2 module evolved in this thesis have in comparison with other CBM-carbohydrate interactions relatively high affinities for xylan (6.67×10^5 M^{-1} and 1.15×10^5 M^{-1}, respectively) (PAPER IV). CBMs in general have binding constants in the range of 10^3-10^6 M^{-1} for their carbohydrate ligands (Boraston *et al.*, 1999; Boraston *et al.*, 2004). In comparison with other binding proteins, like antibodies, engineered variants of which can have affinity constants as high as 10^{-13} M^{-1} (Boder *et al.*, 2000), the affinities of CBMs cannot be classified as high. As previously mentioned, carbohydrate specific antibodies are however difficult to make (Willats *et al.*, 2000) resulting in the relative rarity of variants with high binding affinities. It can thus be argued that the limitations in finding high affinity binders towards carbohydrate ligands are linked to the ligand rather than to the protein scaffold. However, Puhlmann et al. have reported on the generation of a monoclonal antibody, CCRC-M1, that has an affinity constant in the nanomolar range for sycamore maple xyloglucan (Puhlmann *et al.*, 1994). This shows that it is not impossible to achieve a strong binding between proteins and carbohydrates, although such interactions are not common.

3.3 The biological role of CBMs

The biological role of CBMs as parts of modular hydrolysing enzymes might explain their weaker binding affinities ($K_A<10^7$ M^{-1}), although it is poorly investigated exactly how the affinity of the CBMs affect the activities of enzymes. Studies comparing the intact enzymes with enzymes that had their CBM(s) deleted by genetic engineering have shown that in some cases CBMs have only small effects on the enzymatic activity (Ali *et al.*, 2005) whereas most commonly, CBMs clearly enhance the hydrolytic activity of modular enzymes on insoluble and/or soluble substrates (Bolam *et al.*, 1998; Ali *et al.*, 2001; Zverlov *et al.*, 2001; Kittur *et al.*, 2003; Araki *et al.*, 2004). Also, addition of a CBM, by genetic engineering, to a single catalytic domain can increase the hydrolytic activity, which has been shown by using CBM4-2 inserted in frame with an endoglucanase from *R. marinus* (Nordberg Karlsson *et al.*, unpublished data). The mechanism by which the CBMs influence the enzyme activity is not understood in detail. One accepted explanation for the observed effects is that binding of CBMs to the polysaccharide increases the local enzyme concentration on the substrate surface, thereby prolonging the contact between the enzyme and the target substrate (Bolam *et al.*, 1998). CBMs have also been shown to disrupt the crystalline structure of cellulose (Din *et al.*, 1994) or the helical structure of starch (Southall *et al.*, 1999), in that way accelerating hydrolysis by making the substrate more accessible to the enzyme.

It has been demonstrated that CBMs recognise microstructures of their targets (McLean *et al.*, 2002), suggesting that CBMs are able to direct the enzymes to specific sites at the substrate surface. Consequently, mixtures of complexes containing the same catalytic domain but different CBMs may hydrolyse different sites on the substrate (Carrard *et al.*, 2000). Degradation of carbohydrates in Nature is thus not only dependent on the presence of a panel of different enzymes but also the diverse specificities of their CBMs are of importance.

CBMs are not only connected to glycoside hydrolysing enzymes although those are the most studied ones, but they can also be connected to other modifying enzymes or even exist as independent entities, not directly appended to enzymes. Human laforin is one example where a CBM is connected to a

phosphatase and the CBM's function is to direct the enzyme to glycogen. Mutations in this CBM leads to mis-targeting of the phosphatase to intracellular glycogen, causing the so called Lafora disease which is a type of epilepsy (Wang *et al.*, 2002c). The highly allergenic olive pollen protein, Ole e 10, is an example of an independent CBM not linked to another polypeptide module. This CBM is suggested to interact with 1,3-β-glucans and regulate enzymatic activity of the cell wall synthesis/degradation pathways during pollen germination, as a part of a multi-protein complex (Barral *et al.*, 2005). These are two examples of CBMs identified during the last five years, a period when many naturally diverse CBMs have been isolated, some of which belong to 18 newly established CBM families. This rapid identification of new CBMs and the intensified characterisation studies of those modules and the existing ones by using highly developed techniques (see chapter 5.3) will undoubtedly shed more light on the biological roles of CBMs in the future.

3.4 Applications of CBMs

The wide diversity in binding selectivities and affinities of CBMs for different variants of polysaccharides creates a large scope of applications for bioseparations and bioprocessing as reviewed by many (Tomme *et al.*, 1998; Levy and Shoseyov, 2002; Shoseyov *et al.*, 2006). Most commonly CBMs are fused to other biologically active proteins and used as tags, such as affinity tags, immobilisation tags or targeting tags. Furthermore, these proteins have found uses as analytical tools in research and diagnostics.

Since type A CBMs have high binding affinities for cheap and safe matrixes such as cellulose, these are well suitable as tags for affinity purification purposes (Boraston *et al.*, 2001b; Kavoosi *et al.*, 2004; Rodriguez *et al.*, 2004). In scientific research today, screening tools such as phage-display (see chapter 5.2) and protein microarray are used for the simultaneous analysis of the binding properties of many proteins. Both of these methods are under constant development and the use of fusion proteins with CBMs have been shown to overcome some of the technical obstacles faced by these systems (Berdichevsky *et al.*, 1999; Ofir *et al.*, 2005). For instance, a novel approach for high-

throughput screening of recombinant single-chain antibody fragments variable (scFvs) was developed, based on the immobilisation of CBM-scFv fusion proteins onto cellulose-based supports (Azriel-Rosenfeld *et al.*, 2004). CBMs are thus valuable as affinity tags not just for the purification but also for screening purposes and a combination of the two processes is also feasible.

CBMs have been successfully used as tags for immobilisation of both antibodies (Reinikainen *et al.*, 1997), enzymes (Kwan *et al.*, 2002; Kwan *et al.*, 2005) and more spectacularly, whole bacterial cells by surface display of a functional CBM (Wang *et al.*, 2002a). Also in these cases, type A modules are the most suitable and the most frequently used ones.

In line with the identification of CBMs highly specific for plant cell wall polysaccharides, these have recently found applications as molecular probes in wood ultrastructure research (Hildén *et al.*, 2003; McCartney *et al.*, 2004; McCartney *et al.*, 2006). As mentioned in the introduction to this chapter, natural CBMs often show binding to more than one carbohydrate ligand. Engineering of a CBM permits the fast isolation of CBM-variants with increased specificity for a particular selection target. Xylan-specific CBMs created by the work in this thesis have successfully been used for detection of xylan in wood sections and in treated and untreated wood fibers (PAPER V). CBMs selected for in PAPER II for binding to xyloglucan are also believed to find use in similar ways for detection of their target in plant cell walls in particular as these are the first markers available that also can recognise non-fucosylated variants of xyloglucan.

The wild type CBM4-2 and two variants created in this thesis have also been investigated for function in weak affinity chromatographic separation of oligosaccharides. In addition to the N-terminal amino group, all these CBM-variants had one free amino group displayed by a lysine residue, both of which are situated on the "backside" of the binding site. This allowed for the immobilisation of the CBMs to the separation columns through amino-coupling without risking any loss of their binding ability. The various binding properties engineered into the CBM4-2 scaffold allowed for generation of distinct separation profiles for the xylo- and cello-oligosaccharides assessed (Johansson *et al.*, 2006). Furthermore, the inherited thermal stability of these proteins was exploited by examining a wide range of column temperatures (up to 65°C) for

the optimisation of separation and retention times. The engineered CBMs could thus be used in chromatographic processes as specific recognition tools and not just as fusion tags for other proteins, as mentioned for other natural CBMs above.

Proteins based on one molecular scaffold with various engineered binding specificities, like recombinant antibodies (Holt *et al.*, 2000; Wingren and Borrebaeck, 2004) or the CBM-variants in this thesis, can be used in protein microarray platforms in a multiplex manner to assess sample composition, e.g. of biological fluids associated with health and disease. The stability of molecules based on the CBM4-2 scaffold at least with respect to denaturation by heat, which is much higher than that of antibodies, and their ability to target carbohydrates, make them attractive candidates for use in such platforms. Current evaluation of the engineered CBMs utility in such applications has initially demonstrated their capacity to act as stable and functional probes on an array surface (Wingren *et al.*, unpublished data). Thus, variants derived from the CBM4-2 scaffold have properties that allow their incorporation in high-throughput protein microarray platforms further expanding the range of applications for CBMs.

4. MOLECULAR SCAFFOLDS

In the search for molecules with desired properties, including binding specificities and affinities, certain molecules serve as scaffolds into which novel properties are engineered as described in chapter 5. Skerra (2000) has defined three major criteria that characterise an applicable scaffold: (i) structurally separated stability and functional properties; (ii) well-defined hydrophobic core and (iii) solvent-accessible active site. Traditionally antibodies have played a central role in this type of studies and the antibody and its derivates are still the most common scaffolds used for molecular engineering purposes. Indeed, this scaffold has been created by Nature to allow for the development of molecular libraries *in vivo*. Accumulated data demonstrates that the principles for generation of libraries that apply to antibodies (both *in vivo* and *in vitro*) can also be applied to other molecular scaffolds. The successful attempts in engineering (see chapter 5) alternative scaffolds, some of which are described in chapter 4.2, have generated specific binders with beneficial properties that make them more suitable than antibodies for certain application purposes.

4.1 Antibodies and antibody-related scaffolds

Antibodies (or immunoglobulins) are a group of glycoproteins secreted by B-lymphocytes upon antigen encounter and they serve as key effector molecules of the humoral adaptive immune system. An antibody is a Y-shaped molecule that consists of two heavy and two light chains. The antibody repertoire is capable of recognising an almost infinite number of antigens, from small organic compounds to large macromolecular complexes. The structural basis for this recognition is a robust part consisting of one variable light chain (VL) domain and one variable heavy chain (VH) domain that together support six hypervariable loops, also called complementarity determining regions (CDRs). These CDRs combine into highly diverse binding sites that are complementary in

shape to the antigens. Like in the case of CBMs (see chapter 3.1.1), the binding site topography of antibodies has been classified into three types: cavity, groove and planar (Webster *et al.*, 1994). The immune system has developed these molecules and biological principles that allow for their perfection *in vivo*. Many of these principles are now used in the laboratory to develop molecules with perfect functions from a diverse range of scaffolds, including CBM4-2.

The emergence of various methods in both recombinant DNA-technology and selection/screening strategies, has led to the isolation of antibodies with various binding specificities suitable for use in diagnostics, therapy and many other biotechnological applications (Holt *et al.*, 2000; Kretzschmar and von Ruden, 2002; Hudson and Souriau, 2003). In the case of carbohydrate targets, which are essential in this thesis, the number of specific antibodies is more limited. Oligosaccharide-specific antibodies have however been generated by different engineering techniques (see chapter 5) using a starting material of either immunised (Deng *et al.*, 1994), natural (Mao *et al.*, 1999) or semi-synthetic (Söderlind *et al.*, 2000) origin (see chapter 5.3). In the case of antibodies specific towards polysaccharides of plant cell walls, these have mostly been prepared by animal immunisation (Puhlmann *et al.*, 1994; Willats *et al.*, 2000; McCartney *et al.*, 2005). This way an antibody specific for the fucosylated variant of xyloglucan was generated (Puhlmann *et al.*, 1994), whereas there are no reports on antibodies that can bind also non-fucosylated xyloglucan. Engineering of an alternative scaffold (see below) in this thesis, that of CBM4-2, was however able to generate specific xyloglucan binders that also recognised non-fucosylated variant of the target. This demonstrates that although the antibody scaffold is well fitted to harbour diversity enough for creation of variants with specificities towards a wide range of target molecules, some type of ligands might be very difficult to address.

One, more pronounced, limitation/drawback associated with the use of antibodies in various applications is their large size (≈150 kDa), which is also one of the reasons why smaller variants only containing the variable regions of antibodies have been created. The fact that entire antibody molecules cannot be produced in *E. coli*, was also a determining factor for engineering antibody fragments that could be produced in such prokaryotic systems. The most

common of these antibody-related scaffolds is the scFv (>250 residues) that comprises of one VH and one VL domain and a 15-25 amino acid linker that connects the two into a single domain (Bird *et al.*, 1988; Skerra and Plückthun, 1988). Also the antibody binding fragment (Fab) comprised of the two variable domains and two constant domains (Better *et al.*, 1988), although larger than scFv, has been used successfully in engineering studies. Other examples of antibody-related scaffolds are the single domain antibodies, small units (~ 11-15 kDa) corresponding to either VH or VL domains of human antibodies (Holt *et al.*, 2003) or naturally existing from camelids (Muyldermans, 2001). Both these scaffolds are commercialised by Domantis (www.domanits.com) and Ablynx (www.ablynx.com), respectively, and are mainly used in the search for therapeutic recognition tools. The antibody-derivates have to some extent solved the size problem and the difficulty in production of the original antibody molecule, which makes them more advantageous for use in many applications. Still there are purposes for which the antibody scaffold derivates does not fulfil the criteria and for which other alternative scaffolds have to be investigated.

4.2 Alternative protein scaffolds

Characteristics such as small size, high molecular stability, lack of disulfide bridges in the structure and ease of production in simple prokaryotic systems are preferable when using a protein scaffold for molecular evolution. In order to overcome the limitations of the antibody scaffold regarding these issues, other alternative scaffolds have been investigated, most often with great success as reviewed by (Nygren and Uhlén, 1997; Skerra, 2000a; Ladner and Ley, 2001; Nygren and Skerra, 2004; Binz *et al.*, 2005; Binz and Plückthun, 2005; Hosse *et al.*, 2006). Even for therapeutic use, where antibodies have a dominating position, several alternative binding proteins are in preclinical studies or clinical trials (Gill and Damle, 2006) (e.g. affibodies, DARPins, anticalins and fibronectins; see below). The great potentials of many alternative scaffolds have thus led to their commercial exploitation for pharmaceutical as well as other industrial applications (Hey *et al.*, 2005).

In the constantly growing list of antibody alternatives (Binz *et al.*, 2005) the scaffolds with β-sheet frameworks are the most dominating but there are also examples of scaffolds with irregular secondary structures or with α-helical frameworks. One of the most successful alternative scaffolds is that of "affibodies" (or the synthetic domain Z), which derives from an α-helical bacterial receptor domain B from *Staphylococcus aureus* protein A (Nord *et al.*, 1995). Affibodies are composed of three α-helices and they are small (58 residues) and stable molecules that lack disulfide bonds. Their stability has been successfully exploited in affinity chromatography (Nord *et al.*, 2000; Rönnmark *et al.*, 2002). In addition, specific affibodies have been generated towards several protein targets (Nord *et al.*, 1997; Sandström *et al.*, 2003; Wikman *et al.*, 2004) including some of medical relevance, such as the human CD28 (Sandström *et al.*, 2003) and the human epidermal growth factor receptor 2 (Wikman *et al.*, 2004). Together, these properties suggest that affibodies are suitable for use in many types of applications including diagnostics and therapy.

Another scaffold example where the ligand interaction is mediated by α-helices (together with β–turns) is that of designed ankyrin repeat proteins (DARPins). These well-expressed, stable and soluble proteins are composed of four to six repeat units, each of them consisting of 33 residues that form a β–turn followed by two anti-parallel α-helices and a loop reaching the β–turn of the next repeat (Binz *et al.*, 2003). High affinity binders have been generated towards the *E. coli* maltose binding protein and two eukaryotic mitogen-activated protein kinases by using the DARPins scaffold (Binz *et al.*, 2004; Amstutz *et al.*, 2005). The possibility to adapt the ligand interaction surface, not only by randomisation but also by adding more repeat units, is unique for this type of scaffolds.

Lipocalins are a family of relatively small (160-180 residues) and robust β–barrel proteins that include ten human members of which the plasma retinol binding protein is the most well known. The scaffold of lipocalins, consisting of a β–sheet that forms a conical barrel and supports four loops at the open end of the structure, has been engineered into artificial binding proteins termed "anticalins" (Skerra, 2000b; Skerra, 2001). Initial work focused on generation of specific binders towards low-molecular-weight ligands such as fluorescein

(Beste *et al.*, 1999) and digoxigenin (Schlehuber *et al.*, 2000) by reshaping the ligand binding pocket in the β-barrel structure. Recently, targeting the diversity to the four loops of a lipocalin have allowed the generation of protein specific anticalins, some of which are believed to find use in medical applications (Vogt and Skerra, 2004; Schlehuber and Skerra, 2005). Engineering of the four loop structures is analogous to the engineering of the six CDRs in the antibody scaffold. However, the fact that lipocalins are smaller than antibodies and consists of a single polypeptide chain makes them easier to manipulate at the gene level.

The fibronectin type III domain is a small (94 residues) immunoglobulin-like folding unit that occurs in a wide variety of natural proteins involved in ligand binding. This fold is highly stable without relying on disulfide bonds and has successfully been engineered to recognise various targets like ubiquitin (Koide *et al.*, 1998) and TNF-α (Xu *et al.*, 2002), the latter with affinities in the low picomolar range.

Prior to our work, another CBM molecule, the cellulose-specific CBM from cellobiohydrolase Cel7A in *Trichoderma reesei* (a type A CBM; see chapter 3.1.1), had been engineered for binding to bovine alkaline phosphatase (Smith *et al.*, 1998). In another study, variants of the same CBM scaffold were isolated for inhibiting the activity of porcine alpha-amylase (Lehtiö *et al.*, 2000). Engineering the CBM4-2 scaffold in this thesis represents the first approach for generation of specific binders towards different carbohydrates (xylan, Avicel, mannan and xyloglucan; see PAPER I and II), by engineering an alternative scaffold. Findings in PAPER III demonstrated the ability of this scaffold to also harbour binding specificities towards protein structures. The fact that CBM4-2 and variants thereof are highly thermostable proteins without possessing any disulfide bonds and are very well-expressed in *E. coli* (PAPER I) further increases their potential for use as specific carbohydrate probes in different applications.

In general, alternative scaffolds subjected to engineering approaches are thus composed of single polypeptide chains, which facilitates the genetic engineering (see chapter 5) and have small sizes and high stability, which are beneficial in most applications. The rapidly expanding field of biotechnology seeks binders with these kinds of features why the list of antibody alternatives will surely continue to grow also in the future.

5. SCAFFOLD ENGINEERING

Protein engineering deals with the modification of proteins either by chemical or enzymatic methods or by introducing changes at the gene level. The latter approach, also called genetic engineering, involves the substitution, deletion or insertion of single amino acids or larger segments into the gene, thereby creating new protein variants. Protein engineering by genetic means was made possible by a series of revolutionary discoveries in the 1970s and 1980s in the area of recombinant DNA technology. These include gene cloning techniques for precise cutting (by restriction enzymes) and rejoining (by ligase) of DNA pieces (Linn and Arber, 1968; Arber and Linn, 1969), methods for DNA sequencing (Sanger and Coulson, 1975; Maxam and Gilbert, 1977; Sanger *et al.*, 1977) and the Nobel Prize awarded polymerase chain reaction (PCR) invention (Mullis and Faloona, 1987). One very successful approach of protein engineering involves creation of combinatorial libraries that contain large numbers of sequence-similar proteins with altered functionalities. Different screening and selection methods are then used for isolation of library members with desired properties. Many proteins have served as scaffolds for engineering variants with desired properties such as novel ligand recognition (Skerra, 2000a), enzymatic activity (Bloom *et al.*, 2005) or enhanced thermal stability (Lehmann and Wyss, 2001). Although engineering work, additional to that on the CBM4-2 scaffold, has been preformed on the catalytical domain of the same enzyme to enhance its thermal stability (see chapter 5.1.2), the main focus of this thesis and therefore also of this chapter is the generation of novel binding proteins.

5.1 Combinatorial libraries

5.1.1 Site-directed mutagenesis in library construction

The Nobel Prize in chemistry 1993 was shared between Kary B. Mullis, who developed the PCR, and Michael Smith for his pioneering work on site-directed mutagenesis. This technique that allows for *in vitro* synthesis of mutant DNA by using oligonucleotides (Hutchison *et al.*, 1978) was a major breakthrough for the field of molecular biology. Degenerated primers, which allow more than one nucleotide to be incorporated at one, two or all three positions of a given codon, are often engaged in site-directed library construction for directing the diversity to a specific part of the protein sequence. In this way, a number of gene fragments are first amplified, allowing the incorporation of the site-directed mutations, and later assembled to yield the whole gene, which is further amplified prior to cloning into a vector. The proofreading capability of the DNA polymerase used for these extensive PCRs has a vital importance for avoiding the introduction of un-designed random mutations. The low-fidelity DNA polymerase from *Thermus aquaticus* (*Taq* polymerase) (Tindall and Kunkel, 1988) was used for the CBM4-2 library construction in this thesis (PAPER I) consequently allowing for such mutations to occur. During the investigation of clones selected for binding to human IgG4 in PAPER III, we found this kind of random substitutions to be critical for their specificity. Additional mutations, not introduced by will, can thus further increase the diversity of a combinatorial library and allow for novel specificities to be established.

The variation allowed for each residue diversified in a combinatorial library can be either completely random (all 20 amino acids are allowed) (Nord *et al.*, 1995; Koide *et al.*, 1998; Beste *et al.*, 1999) or more restricted like in the library based on the CBM4-2 scaffold in PAPER I (Table 1). Complete diversity in many positions (x) of a scaffold leads to a dramatic increase in library size (20^x). Due to limitations faced by such large libraries during the actual library construction and further handling, including transformation of the cloned material into a proper host, often only fractions of the designed library are subjected to selection (Nord *et al.*, 1997; Vogt and Skerra, 2004). There is also a risk of secondary structure disturbance when introducing diversity of inappropriate nature into a protein sequence. This phenomena was observed in

the case of an affibody variant selected for binding to protein A, which behaved as an aggregation-prone molten globe and folded into the original three-helix bundle scaffold first upon binding to the ligand (Wahlberg *et al.*, 2003). Although this structure was stable enough to survive through the selection and analysis process, destabilised protein structures are unfavourable for binding affinity since some of the free energy gained from the complex formation upon binding needs then to be used for protein folding. The diversity introduced in PAPER I was mainly restricted to amino acids related to those found in the wild type module in each position in order to preserve structural stability and still allow for modified interactions with the ligand. Furthermore, the diversity generated this way does not exceed the practical limits of library size, faced during the library construction and the selection processes (see chapter 5.2).

Table 1. Diversity introduced in the combinatorial library of CBM4-2

Target residues[a]	**Designed modifications**
W69	W F Y H L
E72	E Q D H
F76	F Y H L
F110	F Y H L
Q111	Q E D H
E112	E Q D H
R115	R K
H117	H F Y L
E118	E Q D H
R142	R S
H146	H F Y L
Y149	Y F H L

[a] Residue numbering is in accordance with Simpson *et al.* (2002).

Site-directed library design requires the access to the scaffold sequence and preferably its 3D-stucture, which allows for more qualified predictions on the effects of certain amino acid substitutions. As previously mentioned, one of the criteria for an applicable scaffold in engineering studies is the possession of

structurally separated functionality and stability (see chapter 4). Consequently, changes at the binding surface of a protein scaffold are not expected to destabilise the molecule. Recently, a method for estimating the preserved structural stability of a combinatorial library was reported (Wiederstein and Sippl, 2005). By calculating the percentage of stabilised mutants, using knowledge-based potentials, this study confirmed for all the investigated scaffolds, including CBM4-2, that certain protein surfaces, the ligand binding sites, are more suitable for diversification than others.

There are however studies where other surfaces than those known to constitute the original binding site of a scaffold have been successfully engineered for target recognition. One example is that of the CBM2a of Cel7A from *T. reesei* where instead of the flat binding site, a loop region at one end of the molecule was engineered to create variants that could inhibit the activity of porcine alpha-amylase by binding to the active site of the enzyme (Lehtiö *et al.*, 2000). Another example is the diversification of the four loops at the entrance to the binding pocket of the lipocalin-scaffold, which led to the creation of a library with variants suitable for specific protein recognition (Vogt and Skerra, 2004). These examples have thus demonstrated the ability of some scaffolds to harbour diversity relevant for binding interactions with different types of ligands.

5.1.2 Randomised protein libraries

Random point mutations can be introduced into DNA sequences through error-prone PCR that utilises low-fidelity DNA polymerases and non-standard PCR reaction conditions (Leung *et al.*, 1989). The type of DNA polymerase employed for the library construction has an impact on the random diversity introduced by PCR. Unwanted mutational bias towards specific substitutions (A to G and T to C) of the commonly used *Taq* DNA polymerase (see above) can be reduced by mixing this enzyme with Stratagene's Mutazyme, a combination that displays a near-uniform mutational spectrum (Vanhercke *et al.*, 2005). Libraries randomised through error-prone PCR are less common for the generation of novel binders. However, it is a successful approach for affinity maturation of binding proteins (see below) and for enzyme engineering (for review see: Bloom *et al.*, 2005). An error-prone library has been created on the catalytical module of

the xylanase 10A of *R. marinus*, from which also the CBM4-2 originates. At the time, there was no structural information available on this module and the nearest homologue shared only around 30 percent sequence-identity. Random engineering for enhanced stability in this case allowed for the isolation of variants with only slightly increased thermal stability and preserved activity at 95°C (Cicortas Gunnarsson *et al.*, unpublished data). It is not unusual that protein variants with very enhanced or diverse functionality are rare in randomly constructed libraries. In order to improve the quality of randomised protein libraries, computational tools have been evolved to facilitate library design based on predictive structure-guided evolution (Patrick and Firth, 2005). The establishment of such semi-rational approaches is facilitated by the rapid expansion of the RCSB Protein Data Bank (PDB).

As many engineered proteins are intended for diagnostic and therapeutic purposes, which demand both high binding specificities and affinities, several methods have been evolved for enhancing the binding affinity of specificity-selected proteins. Such technologies attended for antibody engineering are summarised in a recent review (Wark and Hudson, 2006) and most of them are also applicable on other protein scaffolds. One approach for the creation of a second-generation library involves a random mutagenesis, taking after the somatic hypermutations that occur during affinity maturation of antibodies in the immune response. Random point mutations can be introduced by amplifying the vector, that contains DNA encoding for the protein to be engineered, in a mutator strain of *E. coli* (Fowler *et al.*, 1986), which has a higher spontaneous mutation frequency than the wild type strain. However, random diversity is most commonly generated through error-prone PCR, in the scaffold of one selected variant (Hawkins *et al.*, 1992; Boder *et al.*, 2000) or a pool of selected variants (Gram *et al.*, 1992; Xu *et al.*, 2002). In addition, the error-prone PCR method can also be combined with DNA shuffling of a pool of selected mutant genes (Zahnd *et al.*, 2004) in order to increase the library diversity. A pool of homologous genes, either created by error-prone PCR or selected from a combinatorial library is randomly digested and the generated fragments are then reassembled by PCR into full-length genes (Stemmer, 1994). In this way an increased diversification of the gene is achieved and some of the variants might have accumulated a

number of advantageous mutations and avoided deleterious ones. Since Stemmer first introduced the DNA shuffling in 1994, a wide range of alternative recombination methods, such as RACHITT (RAndom CHImeragenesis on Transient Templates) (Coco *et al.*, 2001) and FIND™ (Fragment INduced Diversity) (Knecht *et al.*, 2006), have been evolved and successfully applied on different projects mostly involving enzyme engineering but also binding affinity maturation. Shuffling techniques were never applied in the work of this thesis but may well be considered in future engineering of the CBM4-2 scaffold, not the least for affinity maturation of selected clones.

5.1.3 Antibody libraries

Beside the engagement of the mutagenesis and shuffling techniques mentioned in previous chapters, the access to natural sequences, from the immune system of individuals, as a source of variability, further increases the number of approaches for antibody library construction (Holt *et al.*, 2000; Söderlind *et al.*, 2001; Hoogenboom, 2002; Kretzschmar and von Ruden, 2002). The combinatorial antibody libraries can be divided into immunised, naïve and synthetic, depending on the source of immunoglobulin genes.

Repertoires of antibody genes collected from immunised animals or immune donors can be used for creation of libraries with diversity determined by the natural immune responses. In this way, a combinatorial library of Fab fragments was created by using blood samples of vaccinated donors and could be used for isolation of human antibodies specific to hepatitis B surface antigen (Zebedee *et al.*, 1992). Similarly, assembling of VH and VL regions (with addition of a linker; see chapter 4.1) isolated from lymphocytes of cancer patients or donor with increased autoantibody titer generated scFv-libraries from which binders specific towards tumor-associated targets were isolated (Kupsch *et al.*, 1999; Mao *et al.*, 1999). Immunised antibody libraries can thus generate specific antibodies against antigens to which an immune response had been induced. However, these libraries are often small and face limitations both in the type of antigens recognised and in the binding affinities of selected antibodies.

Naïve and synthetic libraries have been created in an attempt to allow selection of antibodies specific for a large variety of antigens, preferentially

having high binding affinities. Construction of a naïve antibody library is much the same as for the immunised library except for using repertoires of genes from unimmunised human donors (Marks *et al.*, 1991; Vaughan *et al.*, 1996). Some of the synthetic approaches on the other hand involve *in vitro* assembly after introducing randomisation into the third complementary determining region (CDR3) of germline (Hoogenboom and Winter, 1992) or rearranged (Barbas *et al.*, 1992) VH genes. Fully synthetic combinatorial antibody libraries (HuCAL) can be achieved by randomisation of both VH and VL in the CDR3 (Knappik *et al.*, 2000) or all six CDRs (MorphoSys, Martinsried, Germany) encoding regions in a set of master genes. Another example of synthetic library design follows the n-CoDeR concept where sequences encoding *in vivo*-formed CDRs of different germline origin are recombined into a master framework (Söderlind *et al.*, 2000; Azriel-Rosenfeld *et al.*, 2004). These kinds of library designs lead to a broader appliance of the scaffold by the increased library size due to novel CDR-combinations and thereby increased probability of finding diverse binders of interest.

Since much is known about the antibody binding site topography (Webster *et al.*, 1994; MacCallum *et al.*, 1996) and the contact residues important in recognition of various types of targets (Padlan *et al.*, 1995), libraries biased for antigens of predefined sizes/structures should be possible. This is, however, not a very used approach and in fact the first focused antibody library was just recently reported. The introduced diversity was restricted to cavity-lining residues in a hapten-specific scFv, thereby constructing a library for improved hapten recognition (Persson *et al.*, 2006). This library design was similar to the one in this thesis, where diversity was directed to the original binding site of the CBM4-2 scaffold, in a manner that is likely to generate variants with maintained the binding site topography suitable for carbohydrate recognition.

5.2 Selection of combinatorial libraries

5.2.1 Selection methods

Construction of libraries with billions of protein mutants is relatively straightforward using the recombinant DNA-technology while selection and down-stream analysis of the best protein candidate is more challenging. The selection of an engineered phenotype requires its direct linkage to the associated genotype which can be achieved either by utilising some kind of display format or compartmentalisation format. The cell-like aqueous compartments of water-in-oil emulsions (Tawfik and Griffiths, 1998) are well suited for selection of catalytic activity and in combination with DNA display also for the selection of binding proteins (Bertschinger and Neri, 2004). Selection approaches that engage some kind of display can be divided into: (i) systems that require transformation of the library-DNA into a cellular host like in phage-display (Smith, 1985; McCafferty *et al.*, 1990), bacterial surface display (Fuchs *et al.*, 1991; Francisco *et al.*, 1993), yeast surface display (Boder and Wittrup, 1997) and (ii) systems that imply *in vitro* protein translation such as polysome/ribosome display (Mattheakis *et al.*, 1994; Hanes and Plückthun, 1997), mRNA display (Nemoto *et al.*, 1997; Roberts and Szostak, 1997) or DNA display (Reiersen *et al.*, 2005). There are of course advantages and limitations associated with all these methods and apart from local access, properties of the protein scaffold to be displayed as well as the library size are determining factors when choosing a selection system.

Since it was first described by Smith in 1985, the display of foreign molecules on filamentous bacteriophage particles has been intensively used for many applications (Smith and Petrenko, 1997; Willats, 2002; Kehoe and Kay, 2005) and it is by far the major tool for isolation of engineered recombinant proteins. The library based on the CBM4-2 scaffold reported in this thesis was constructed in a phagemid vector as a fusion to the phage-coat protein III (pIII), of which 3-5 copies are displayed on phage particles. These vectors were transformed into an *E. coli* strain that possesses F-pilus (PAPER I). Production of phages that displayed CBM-variants on their pIII was induced by infection of the host cells, through the F-pilus, with helper phages. The phagemid vector carries a replication origin that allows the phage to replicate normally in *E. coli* (Bass *et al.*, 1990) and the helper phage supplies the other phage genes. The

employment of the phagemid vector yields a primarily single copy (monovalent) display and consequently binders with higher affinities can be selected this way as compared to using the true phage vector, which results in a multivalent display and thus the selection of lower affinities due to the avidity effects (O'Connell *et al.*, 2002). However, high affinity might not always be the major goal of the selection and in some cases multivalent display is preferred since it favours the identification of rare and/or low-affinity binders from a large combinatorial library. Multivalency can also be achieved by fusing the library members to the major coat protein VIII that is present in approximately 2700 copies on the wild type M13 and by using the phagemid vector 100-200 copies of displayed fusion can be achieved, at least in the case of peptide display (Hoess, 2001).

The bio-panning process used in our studies is the most common selection scheme based on affinity selection. Typically, the phage-displayed library is exposed to the target immobilised on a solid support; i.e. chromatography columns (McCafferty *et al.*, 1990), or the surface of wells of microtitre plates (Barbas *et al.*, 1992) or paramagnetic beads (Hawkins *et al.*, 1992) (PAPER I and II). Phage-displayed proteins that bind to cell surface targets can also be selected directly on cells (Siegel *et al.*, 1997; Kupsch *et al.*, 1999; Fransson *et al.*, 2004). The insoluble form of the carbohydrates used for selection in PAPER I allowed for phages displaying proteins specific for these ligands to be pelleted by centrifugation. After a suitable incubation time followed by washing steps to remove unbound phages, the selectively captured binding phages are eluted (Figure 3). Elution can be carried out by various means, including incubation at low pH, competitive elution or proteolytic cleavage like that applied in PAPER I and II, which makes use of an engineered trypsin sensitive site in the linker that connects the displayed protein to the phage coat protein (Johansen *et al.*, 1995). The eluted phages are used to infect *E. coli* and are thus amplified for use in additional rounds of selection or for identification and characterisation purposes.

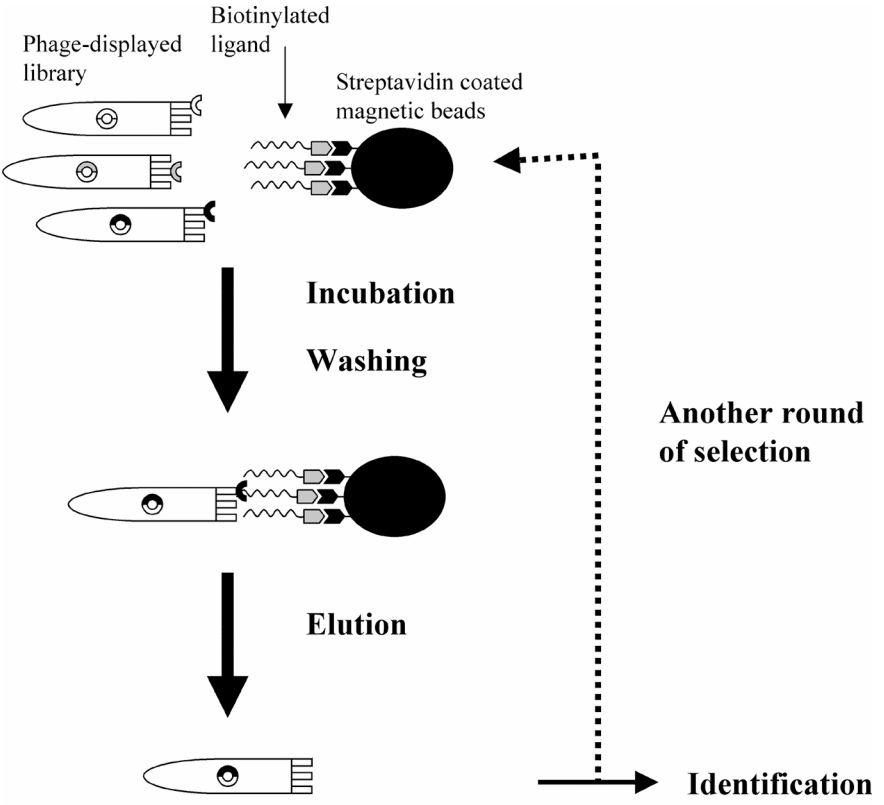

Figure 3. Schematic view of the phage-display selection procedure (the bio-panning) used to fish out specific binders from a combinatorial library. A modified version from Cicortas Gunnarsson *et al.* (2006)

The wild type CBM4-2 and the combinatorial library based on this scaffold were shown in PAPER I to be efficiently displayed on phages. The lack of required posttranslational modifications, due to its prokaryotic origin, and the absence of disulfide bonds in the structure, which enables production both in reducing and oxidising environments, surely facilitate the display of this scaffold. It is however not necessary so that all kinds of protein sequences are suitable for selection by phage-display, due to toxicity effects or improper folding in *E. coli*. In order to solve the latter issue, especially for mammalian

proteins, the yeast display system has been evolved since the protein folding and secretory machinery of yeast is strikingly homologous to that of mammalian cells. Selection of scFvs using this display format has yielded high affinity binding variants with dissociation constants down in the femtomolar range (Boder *et al.*, 2000). The matter of toxicity and the library size limitation ($\approx 10^9$ members) due to the transformation of library-DNA into host cells are the main disadvantages of the display systems that rely on living organisms in the selection process.

Several *in vitro* display technologies have emerged by circumventing the need to introduce DNA into a cellular host and thus promising an access to displayed libraries of increased sizes. The DNA or mRNA is physically added to a cell free protein synthesis system where potentially 10^{14}-10^{15} molecules can be translated and remain linked to their encoding gene sequence. Additional diversity can easily be introduced between selection rounds during the inherent PCR step by applying error-prone PCR and/or shuffling techniques (see chapter 5.1.2). This kind of "affinity maturation while selecting" process has led to the isolation of high affinity antibodies (scFvs) (Hanes *et al.*, 2000) and fibronection type III domain-variants (Xu *et al.*, 2002), both types of binders with dissociation constants in the picomolar range. The most successful *in vitro* display systems are the ribosome and mRNA display, which employ mRNA that lacks a stop codon thereby giving rise to the formation of a mRNA-ribosome-protein complex (Lipovsek and Plückthun, 2004). Potential dissociation of the complex is a limitation associated with the ribosome display system and it is overcome in the mRNA display system by covalently coupling the protein to the mRNA via puromycin after translation (Roberts and Szostak, 1997). The difficulty in handling mRNA because of their susceptibility to ribonucleases and the reducing environments required for the activity of the RNA polymerase activity are also limitations that apply for both these display systems. Technical refinements of the original methods have however provided a broad applicability of the *in vitro* display as a selection tool (FitzGerald, 2000; Dower and Mattheakis, 2002). Although never exploited for selection of CBM, the built in maturation capacity of in vitro selection technologies affords an attractive choice in particular when specific CBMs with higher binding affinity are being developed.

5.2.2 Selection conditions

An expression often used in the selection context is "you always get what you select for". This applies to all kinds of selection systems (see above) and means that the selection conditions are crucial for the outcome. As suggested above the phage-display system, and other cell-surface displays, may be biased towards proteins that are well expressed and not toxic to the host cell. This does not have to be a disadvantage, although functional library size may be lower than expected, since production of high amounts of these proteins is probably desirable for their intended applications. Apart from high productivity there might be other properties, such as thermal stability, high specificity and high affinity that are beneficial for the use of a selected binder. Demands of such properties can be addressed by using proper selection pressures.

The stability of the phage particle gives the phage display system an advantage over most other systems. Selections preformed at elevated temperatures, at least up to 60°C after which the phage itself can be harmed, or in the presence of different concentrations of guanidinium chloride have enabled the selection of binding proteins with for improved stability (Jung *et al.*, 1999). Although the wild type CBM4-2 and the xylan-specific variant X-6 selected in PAPER I are both thermostable (wt CBM4-2 has a melting transition temperature T_m= 89.3°C and X-6 has T_m= 78.5°C) and have similar affinities at room temperature (PAPER V), the wild type shows a higher affinity for xylo-oligos than the X-6 module at elevated temperatures (Johansson *et al.*, 2006). One explanation for this is that CBM4-2 originates from a thermophilic bacterium and has thus naturally been selected for good binding at high temperatures, in contrast to the X-6 module, which was selected at room temperature. High selection temperatures should thus be applied for the isolation of binders intended for applications at elevated temperatures.

For most applications high binding specificities are required and in order to enhance the selection of such binders and to eliminate cross-reactivity, competing molecules or cells can be included in the selection mixture (Dennis and Lazarus, 1994; Siegel *et al.*, 1997). In PAPER II we showed that presence of soluble xylan during the selection for xyloglucan-specific CBM-variants had an influence on the selection process and in the end resulted in isolation of highly

target-specific binders with no binding affinity for xylan. Another strategy for increasing the specificity of the selection outcome is to pre-select the displayed libraries on targets that should not be recognised by the selected binders. Altogether, there are ways of suppressing the selection of variants with binding specificities additional to the one towards the selection target.

Applications such as therapy and diagnostics require proteins with high binding affinities while other purposes can be found for proteins with moderate binding affinities. By elaborating with the incubation time, number of washing steps and the concentration of selection target (like in PAPER II), library variants with different affinities for the same ligand can be isolated. In order to allow for enrichment of rare binders, it is common that the selection conditions are mild during the first rounds of selections and more stringent in the following rounds. The higher stringency applied and the more selection rounds performed, the less diverse pool of clones is being selected. This facilitates the screening work (see below) following the selection in order to isolate the library variant that best fits the intended application.

5.3 Screening of combinatorial libraries

In contrast to the selection methods described above, which examine the properties of a pool of library members in one go, screening for functionality is performed individually for each member of the selected population. Screening for binding specificities can be applied either on a library with reasonably small diversity (Abedi *et al.*, 1998; Malabarba *et al.*, 2001) or on a sub-library of pre-selected clones as in PAPER I and II.

An advantage of phage-display over the *in vitro* selection systems is that selected clones can easily be displayed on phages and used for screening without requiring the cloning step necessary in the case of *in vitro* selection. Phage stocks of individual protein variants are most commonly analysed by phage-ELISA (enzyme linked immunosorbent assay) (McCafferty *et al.*, 1990; Mattheakis *et al.*, 1994). Phage supernatants from selected clones are added to target-immobilised wells of a 96-well plate and binding is detected by using an anti-phage antibody. Specificities and relative affinities can be determined by using

this assay in a competitive format. When coating the target on to a surface is impossible, high enough output phages compared to input phages might serve as a confirmation for specific binding of phage-displayed proteins to for instance the insoluble carbohydrate targets used in PAPER I.

There are several other screening methods apart from the phage-display based ones including filter screening where expressed proteins from lysed bacterial colonies are captured onto a filter and examined for binding to biotinylated ligands using horseradish peroxidase labelled streptavidin for detection (Malabarba *et al.*, 2001). Cell-surface display, such as yeast-display, allows the use of fluorescent activated cell sorting (FACS) for detection of binding by either employing fluorescently labelled antibodies (Starwalt *et al.*, 2003) or by fusing peptide libraries with engineered green fluorescent protein prior to the display (Abedi *et al.*, 1998).

The screening set-up might affect the outcome just like the selection conditions discussed above do. For instance proteins might show higher binding to coated targets due to avidity effects or proteins can be stabilised by the fusion to the phage-pIII or by the immobilisation of the protein onto filters used for screening. The latter effect was experienced during the screening process for enhanced stability of the engineered catalytic module (see chapter 5.1.1), where protein-immobilised filters were incubated at elevated temperatures. Some of the isolated variants were later found to be less stable when analysed in solution (unpublished data) proposing that their stability was enhanced by immobilisation to the filter. In conclusion, it is vital to optimise not only selection but also screening in order to identify variants with optimal characteristics.

Optimal screening work should allow for thorough investigation of many library members. Different methods of course require various efforts and for instance phage-display screening as performed in PAPER I and II by manual means severely limits the number of investigated clones from each selection. However, by engaging robotic systems that can handle some or preferentially all steps involved, a high-throughput screening can be performed both in solution and by using cell-based systems (Sundberg, 2000). Automatic screening not only allows larger libraries to be screened but also in more than one way thus assessing the specificity of the protein variants even more. The era of array-

based technologies has generated new ways/formats for addressing the specificity determining issue of protein variants. Protein microarrays like the one containing about 5,000 yeast proteins (Michaud *et al.*, 2003) or the first nonredundant human protein chip (Lueking *et al.*, 2003) or the commercial Protoarray™ (Invitrogen) that today include 8,000 human proteins are valuable tools in specificity studies of proteins, mostly antibodies so far. Even though this type of screening is preferable, the high cost of the arrays limits the afforded number of proteins that are screened. Instead such systems are more utilised for further characterisation of a few selected and/or screened variants, as discussed in the next chapter. Carbohydrate microarrays, another strongly emerging set of technologies, are well suited for screening of carbohydrate-binding proteins (Feizi *et al.*, 2003). Arrays of both polysaccharides (and glycoproteins) (Wang *et al.*, 2002b) and oligosaccharides (Fukui *et al.*, 2002) have been developed and used for specificity mapping of proteins (Schwarz *et al.*, 2003). Future selections/screening of CBMs engineered from the CBM4-2 scaffold would greatly benefit from collaborations with laboratories preparing these kinds of arrays.

5.3 Characterisation of binding proteins

More thorough investigations are normally performed on a few promising binders isolated through selection and/or screening of a combinatorial library. Depending on the application purpose both biochemical and structural information has to be gathered on a particular binding protein. The characterisation work can be divided into: (i) functional studies, (ii) structural studies and (iii) structure-function studies.

For the functional studies, there are several techniques available that assess the binding specificities and affinities of soluble and purified proteins. Two such methods, also used for characterising the specificities of engineered CBM-variants in this thesis, are ELISA (Engvall *et al.*, 1971) (the phage-display format of this method is described in chapter 5.3) and affinity electrophoresis (AE) (Takeo, 1984). These assays can however only give semi-quantitative values on binding affinities why other techniques like Biacore, a surface plasmon

resonance based technology (www.biacore.com), or isothermal titration calorimetry (ITC) have to be used for more reliable values on the association and dissociation constants of binding. Apart from affinity constants these methods also generate kinetic and thermodynamic data, respectively. The choice of method to use depends on several factors like the access to the instrumentation, the amount of protein required (certain recombinant proteins can be hard to produce and purify in large amounts), the applicability of the technique for studying a particular binding complex and the type of additional data generated. In PAPER IV, binding characteristics of the wild type CBM4-2 and the engineered X-2 module were studied using ITC. The generated data revealed that mutations in X-2 have contributed to a higher specificity for xylose containing carbohydrates and preference of binding to the polysaccharide xylan rather than to xylopentaose. One advantage of the ITC method used in this thesis to analyse the binding interaction, is that it lacks the requirement of immobilisation found in other techniques. The relatively high amount of purified protein consumed in each titration is on the other hand a drawback of ITC, although this is not a big concern for variants selected from the library based on the CBM4-2 scaffold since they all showed a high productivity in *E. coli* (PAPER I).

Structural studies can reveal changes in the protein structure caused by engineering that probably are responsible for the novel binding properties or for the loss or gain in molecular stability. Circular dichroism (CD) spectroscopy, a method for determining the secondary structure of a protein sequence, was used for some of the selected mutants in this thesis (PAPER I). The results of that investigation showed that the CBM4-2 scaffold was able to harbour diversity necessary for novel binding properties while maintaining the secondary structure. The 3D-structure of a protein can be solved either in solution by using NMR spectroscopy or more accurately by x-ray crystallography, which requires the protein to be crystallised at first. Proteins are sometimes co-crystallised with their ligands and structural information from this type of studies brings a larger contribution to the structure-function determination. The NMR-based structure of the wild type CBM4-2 (Simpson *et al.*, 2002) was solved just prior to the start of the engineering work and currently some of the generated CBM-variants are in the process of being crystallised. Information about the structure of the

mutated proteins will help to better understand their interactions with the ligands and hopefully confirm some of the structure-function theories raised in PAPER II, III and IV.

Even without the access to the 3D-structure of an engineered protein, structure-function studies can be conducted by comparing its protein sequence, which is easily determined after the selection and/or screening, with that of the wild type and/or other isolated variants. Mutated residues identified during such comparisons are most probably responsible for the displayed differences in functionality. The importance of individual residues are most commonly analysed by alanine scanning which implies the replacement of those amino acids, one at the time, with alanine followed by functional analysis of the single-mutants created. The choice of alanine as a substitution candidate relies on the fact that this amino acid does not alter the main-chain conformation and has a short non-interacting side chain eliminating any extreme or electrostatic effects (Cunningham and Wells, 1989). Substitutions to alanine as well as to other amino acids (found among the library members) performed in this thesis have revealed binding affinity and/or specificity determining residues in both wild type CBM4-2 and in human IgG4-specific variants (PAPER III and IV). Further studies of this kind as a complement to solved 3D-stuctures are necessary for more complete understanding the structure dependence of the novel binding specificities engineered into the CBM4-2 scaffold.

6. CONCLUDING REMARKS

In this thesis, engineering of the CBM4-2 **scaffold** has generated binders with novel **binding specificities** and **affinities**, which have and may find utility in several biotechnological **applications**.

The **scaffold** of CBM4-2 was shown to be an excellent carrier of the diversity introduced by the library design in PAPER I, as well as additional random mutations brought in through the engineering process. Clones selected in PAPER I and II had a reduced but still high stability as compared to the thermostable wild type CBM4-2. Furthermore, these were all easily produced in *E. coli* and purified by chromatographic means yielding high amounts (average 75 mg purified protein/l shake flask culture) of soluble proteins. The choice of the CBM4-2 scaffold, naturally prone to bind carbohydrate molecules, proved to be successful in the creation of specific carbohydrate binders, which are normally hard to generate by the use of other scaffolds (Willats *et al.*, 2000).

Modules with various **binding specificities** were identified from the combinatorial library based on the CBM4-2 scaffold in PAPER I and II. Both the xyloglucan-specific CBM XG-34 in PAPER II and the xylan-specific CBM X-2 in PAPER IV had retained and/or enhanced the wild type binding to one particular target and completely abolished the binding to other ligands. The modules isolated for binding to a human IgG4 molecule in PAPER I were later shown in PAPER III to bind to the protein sequence of this glycoprotein target. Engineering of the CBM4-2 scaffold yielded thus not only binders specific for different carbohydrate targets but also protein-specific ones. The establishment of the binding specificity towards human IgG4 molecule into the CBM4-2 scaffold required not only changes introduced by the library design but also additional random mutations outside the original binding cleft were found to be critical for the target binding (PAPER III). This implies that further diversification of the CBM4-2 scaffold both by rational and random approach will increase the probability of finding variants with binding specificities for different types of targets.

In the search for novel binding specificities in this thesis there was very little selection pressure on the **binding affinities** of the engineered modules. Considering the rare existence of proteins with high binding affinities for carbohydrates (usually $K_A < 10^7$ M^{-1}; see chapter 3.2) and the small size of the library created in this thesis, low-affinity binders were thus more likely to be generated. Some of the isolated binders showed different affinities for a particular target (e.g. xylan) both in comparison with each other and with the wild type CBM4-2 (PAPER IV and V). The binding affinities of isolated variants specific for xyloglucan (PAPER II) or a human IgG4 molecule (PAPER I and III) were indeed too low to be determined accurately but nevertheless, these modules efficiently recognised their targets in a number of analytical applications. If future applications of these proteins require increased binding affinities, affinity maturation (see chapter 5.1.1) will have to be performed. Another possibility is to fuse two or more copies of the same protein, which will lead to an avidity effect during binding that is equivalent to a higher binding affinity of the fusion protein.

There are several **application** areas for the engineered CBM-variants of this thesis. The results presented in PAPER V demonstrate the successful use of xylan-specific CBMs, selected in PAPER I, for detection of xylan in both wood sections and wood fibers. These CBMs and presumably those selected in PAPER II for binding to xyloglucan can thus serve as specific biomarkers in the analysis of the composition of plant cell walls and other wood material, an area that has been poorly assessed in the past. Chromatographic separation of oligosaccharides has been shown to be another potent application for selected CBMs (Johansson *et al.*, 2006). The nice performance of CBM-variants for detection of binding to human IgG4 in an array format (Wingren *et al.*, unpublished data) holds great promise for the use of such scaffold-variants in biosensoring applications. Glycan-specific CBMs engineered in the future can find utility in diagnostics for detection of specific glycosylations on medically important targets. Apart from limitations in therapeutic use, due to the non-human origin, engineered variants of the CBM4-2 scaffold can thus be found suitable for most biotechnological applications.

Results of this thesis clearly demonstrate the utility of the CBM4-2 scaffold for engineering proteins with novel binding specificities of great interest in biotechnology. Future engineering of this scaffold holds great promise in generating specific binders towards vital ligands that are difficult to target by other means.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Varje cell innehåller gener som kodar för proteiner, organismens styrmolekyler i många livsviktiga processer. Genetiska förändringar, s.k. mutationer, kan leda till ändrade egenskaper/funktioner hos proteiner. Darwins evolutionsteori bygger på att det naturliga urvalet gynnar existensen av de förbättrade proteinvarianterna och tar bort varianterna med oönskade egenskaper.

Forskare har lyckats härma och påskynda den naturligt långsamma molekylära evolutionen och kan på så sätt tillgodose det ökade kravet på tillgång av proteiner. Dessa molekyler fungerar nämligen som viktiga igenkänningsverktyg i många biotekniska tillämpningar, bl.a. inom diagnostik och terapi. Utvecklingen av specifikt bindande proteiner med hjälp av molekylärbiologiska metoder har varit framgångsrik när det gäller många typer av målmolekyler, såsom DNA, peptider och proteiner. Att skapa bindare mot kolhydrater har dock visat sig vara svårare. Kolhydrater är de mest förekommande organiska molekylerna i naturen och utgör huvudbeståndsdelen i växter. Med hjälp av specifika bindare skulle man lättare kunna detektera dessa molekyler i processer i både livsmedel och pappersindustrin. Vidare kan kolhydrater påträffas på ytan av proteiner (glykoproteiner) och leda till förändrade biologiska egenskaper. Detta är vanligt förekommande vid sjukdomstillstånd varför identifiering av sådana specifika kolhydrater kan vara av stort diagnostiskt värde.

Avhandlingens syfte har varit att generera specifika bindare mot kolhydrater genom att modifiera ett naturligt kolhydratbindande protein. Detta var första gången som den typen av bindande molekyl har använts som utgångspunkt för att skapa specifika kolhydrat bindare. Genom att introducera flera olika mutationer i genen som kodar för detta kolhydratbindande protein skapades ett s.k. molekylbibliotek med ett stort antal olika proteinvarianter. Biblioteksmedlemmarna sattes fast på speciella bakterievirus (fager) m h a en teknik som heter ”fag-display” och som är lämplig att använda för att hitta proteinvarianter som binder starkt till en särskild målmolekyl. På så sätt har specifikt bindande proteiner mot ett antal olika kolhydrater samt proteindelen i ett glykoprotein kunnat isoleras. Proteinerna genererade i denna studie har

förutom specifik igenkänning också andra fördelaktiga egenskaper, såsom god stabilitet vid höga temperaturer och de kan lätt produceras i stora mängder i bakterier, vilket ökar deras användningspotential. Vissa av de utvecklade proteinerna har redan visats fungera väl i bioseparations och bioanalytiska applikationer.

ACKNOWLEDGEMENTS

My dear supervisor Mats Ohlin, thank you for your trust when you let me start out on this project, and for your continuing faith in me, even though I spent the first half-year preparing electrocompetent cells. Thank you very much for all your advice, comments and help over these years. I envy your boundless knowledge, your calm and your eye for detail (e.g., Table 3 in paper V).

Eva Nordberg Karlsson has been my greatest source of information concerning the CBMs. Thank you for always taking the time to talk about my project and for your great company at meetings and conferences.

Olle Holst and Mats Andersson, I would like to thank you for those fruitful discussions during our project meetings. If it weren't for you Olle, I might never have found this department, so thank you for suggesting me to Mats for the PhD position.

Collaborations are both fun and productive. I would like to thank all my collaborators over these years: Reine Johansson and Sten Ohlsson from Kalmar University, Harry Brumer and Qi Zhou from the Royal Institute of Technology in Stockholm, Lada Filonova and Geoffrey Daniel from WURC in Uppsala, Harry Gilbert and Cedric Montanier from the University of Newcastle upon Tyne, Mike Williamson and Richard Tunnicliffe from the University of Sheffield, and Christer Wingren at the department.

I would like to thank the senior scientists Carl, Mats and Christer for the way you lead the research at the department, which results in a great working atmosphere and continuing success within the research field.

Marianne deserves many thanks for taking care of all the administrative matters, and so does Tommie for all his practical help.

Fredrik (Kalle) Karlsson, my former, calm and lovely roomie, and Kristina Lundberg, my present, happy and snack-loving roomie: thank you for all the chats and laughs (and sometimes tears).

Many thanks and hugs to my dear co-workers and friends, Malin, Helena and Sara, for all the moments of joy and comfort and the nice conversations about science and private matters. Extra thanks to Malin and Helena for your comments on my thesis.

The present and former staff at the Department of Immunotechnology, those mentioned above and Lantto, Bengt, Åsa, Carl-Magnus, Cornelia, Anabel,

Ingrid, Ann-Sofie, Linda, Johan I, Jakob, Niklas, Eva O, Ann-Charlott, Johan F, Kristina La, Anders (extra thanks for the lovely cover background), Ulrika, Peter, Elin, Fredrika and Jonas: thank you all for the great times we had together at Wednesday-cakes, Thursday-buns, retreats, excursions and all kinds of festivities.

To all my friends in Skåne, Närke, Romania or wherever you are, I would like to thank you for making my life outside of work pleasant and joyful.

Simona, my dear sister, thank you so much for listening, being so understanding and sorting out so many things for me. It's such a shame you live so far away.

Vreau sã-mi multumesc pãrinti, pentru cã m-au încurajat ca intotdeauna sã lupt cît mai mult pentru a ajunge cît mai departe. Uite cã am ajuns destul de departe. Vã multumesc din tot sufletul pentru ajutorul si dragostea voastrã.

My beloved husband Johan, I cannot find words to describe how happy and lucky I feel that I have you. I thank you for all your help and support over these years, and especially in the last couple of months. Thank you also for taking such good care of our son Axel. My baby boy, Axel, you are such a sweet little child, and your joy has certainly eased the stress of writing the thesis. The two of you mean the world to me.

REFERENCES

Abedi, M.R., Caponigro, G. and Kamb, A. (1998) Green fluorescent protein as a scaffold for intracellular presentation of peptides. *Nucleic Acids Res*, **26**, 623-630.

Abou Hachem, M., Nordberg Karlsson, E., Bartonek-Roxa, E., Raghothama, S., Simpson, P.J., Gilbert, H.J., Williamson, M.P. and Holst, O. (2000) Carbohydrate-binding modules from a thermostable Rhodothermus marinus xylanase: cloning, expression and binding studies. *Biochem J*, **345**, 53-60.

Ali, E., Araki, R., Zhao, G., Sakka, M., Karita, S., Kimura, T. and Sakka, K. (2005) Functions of family-22 carbohydrate-binding modules in Clostridium josui Xyn10A. *Biosci Biotechnol Biochem*, **69**, 2389-2394.

Ali, M.K., Hayashi, H., Karita, S., Goto, M., Kimura, T., Sakka, K. and Ohmiya, K. (2001) Importance of the carbohydrate-binding module of Clostridium stercorarium Xyn10B to xylan hydrolysis. *Biosci Biotechnol Biochem*, **65**, 41-47.

Amstutz, P., Binz, H.K., Parizek, P., Stumpp, M.T., Kohl, A., Grutter, M.G., Forrer, P. and Pluckthun, A. (2005) Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins. *J Biol Chem*, **280**, 24715-24722.

Araki, R., Ali, M.K., Sakka, M., Kimura, T., Sakka, K. and Ohmiya, K. (2004) Essential role of the family-22 carbohydrate-binding modules for beta-1,3-1,4-glucanase activity of Clostridium stercorarium Xyn10B. *FEBS Lett*, **561**, 155-158.

Arber, W. and Linn, S. (1969) DNA modification and restriction. *Annu Rev Biochem*, **38**, 467-500.

Azriel-Rosenfeld, R., Valensi, M. and Benhar, I. (2004) A human synthetic combinatorial library of arrayable single-chain antibodies based on shuffling in vivo formed CDRs into general framework regions. *J Mol Biol*, **335**, 177-192.

Barbas, C.F., 3rd, Bain, J.D., Hoekstra, D.M. and Lerner, R.A. (1992) Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. *Proc Natl Acad Sci U S A*, **89**, 4457-4461.

Barral, P., Suarez, C., Batanero, E., Alfonso, C., Alche Jde, D., Rodriguez-Garcia, M.I., Villalba, M., Rivas, G. and Rodriguez, R. (2005) An olive pollen protein with allergenic activity, Ole e 10, defines a novel family of carbohydrate-binding modules and is potentially implicated in pollen germination. *Biochem J*, **390**, 77-84.

Bass, S., Greene, R. and Wells, J.A. (1990) Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins*, **8**, 309-314.

Berdichevsky, Y., Ben-Zeev, E., Lamed, R. and Benhar, I. (1999) Phage display of a cellulose binding domain from Clostridium thermocellum and its

application as a tool for antibody engineering. *J Immunol Methods*, **228**, 151-162.

Bertschinger, J. and Neri, D. (2004) Covalent DNA display as a novel tool for directed evolution of proteins in vitro. *Protein Eng Des Sel*, **17**, 699-707.

Beste, G., Schmidt, F.S., Stibora, T. and Skerra, A. (1999) Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proc Natl Acad Sci U S A*, **96**, 1898-1903.

Better, M., Chang, C.P., Robinson, R.R. and Horwitz, A.H. (1988) Escherichia coli secretion of an active chimeric antibody fragment. *Science*, **240**, 1041-1043.

Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grutter, M.G. and Pluckthun, A. (2004) High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat Biotechnol*, **22**, 575-582.

Binz, H.K., Amstutz, P. and Plückthun, A. (2005) Engineering novel binding proteins from nonimmunoglobulin domains. *Nat Biotechnol*, **23**, 1257-1268.

Binz, H.K. and Plückthun, A. (2005) Engineered proteins as specific binding reagents. *Curr Opin Biotechnol*, **16**, 459-469.

Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P. and Pluckthun, A. (2003) Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J Mol Biol*, **332**, 489-503.

Bird, R.E., Hardman, K.D. et al. (1988) Single-chain antigen-binding proteins. *Science*, **242**, 423-426.

Bloom, J.D., Meyer, M.M., Meinhold, P., Otey, C.R., MacMillan, D. and Arnold, F.H. (2005) Evolving strategies for enzyme engineering. *Curr Opin Struct Biol*, **15**, 447-452.

Boder, E.T., Midelfort, K.S. and Wittrup, K.D. (2000) Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc Natl Acad Sci U S A*, **97**, 10701-10705.

Boder, E.T. and Wittrup, K.D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol*, **15**, 553-557.

Bolam, D.N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M.P., Rixon, J.E., Boraston, A., Hazlewood, G.P. and Gilbert, H.J. (1998) Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J*, **331** 775-781.

Boraston, A.B., Bolam, D.N., Gilbert, H.J. and Davies, G.J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J*, **382**, 769-781.

Boraston, A.B., Creagh, A.L., Alam, M.M., Kormos, J.M., Tomme, P., Haynes, C.A., Warren, R.A. and Kilburn, D.G. (2001a) Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from Thermotoga maritima xylanase 10A. *Biochemistry*, **40**, 6240-6247.

Boraston, A.B., Ghaffari, M., Warren, R.A. and Kilburn, D.G. (2002) Identification and glucan-binding properties of a new carbohydrate-binding module family. *Biochem J*, **361**, 35-40.

Boraston, A.B., McLean, B.W., Guarna, M.M., Amandaron-Akow, E. and Kilburn, D.G. (2001b) A family 2a carbohydrate-binding module suitable as an affinity tag for proteins produced in Pichia pastoris. *Protein Expr Purif*, **21**, 417-423.

Boraston, A.B., McLean, B.W., Kormos, J.M., Alam, M., Gilkes, N.R., Haynes, C.A., Tomme, P., Kilburn, D.G. and Warren, R.A.J. (1999) Carbohydrate binding modules: Diversity of structure and function. In Gilbert, H.J., Davies, G.J., Svensson, B. and Henrissat, B. (eds), *Recent Advances in Carbohydrate Engineering*. Royal Society of Chemistry, Cambridge, pp. 202-211.

Carrard, G., Koivula, A., Soderlund, H. and Beguin, P. (2000) Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *Proc Natl Acad Sci U S A*, **97**, 10342-10347.

Charnock, S.J., Bolam, D.N., Nurizzo, D., Szabo, L., McKie, V.A., Gilbert, H.J. and Davies, G.J. (2002) Promiscuity in ligand-binding: The three-dimensional structure of a Piromyces carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose. *Proc Natl Acad Sci U S A*, **99**, 14077-14082.

Charnock, S.J., Bolam, D.N., Turkenburg, J.P., Gilbert, H.J., Ferreira, L.M., Davies, G.J. and Fontes, C.M. (2000) The X6 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the Clostridium thermocellum X6b domain. *Biochemistry*, **39**, 5013-5021.

Cicortas Gunnarsson, L., Karlsson, E.N., Andersson, M., Holst, O. and Ohlin, M. (2006) Molecular engineering of a thermostable carbohydrate-binding module. *Biocatalysis and Biotransformation*, **24**, 31-37.

Coco, W.M., Levinson, W.E., Crist, M.J., Hektor, H.J., Darzins, A., Pienkos, P.T., Squires, C.H. and Monticello, D.J. (2001) DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat Biotechnol*, **19**, 354-359.

Creagh, A.L., Ong, E., Jervis, E., Kilburn, D.G. and Haynes, C.A. (1996) Binding of the cellulose-binding domain of exoglucanase Cex from Cellulomonas fimi to insoluble microcrystalline cellulose is entropically driven. *Proc Natl Acad Sci U S A*, **93**, 12229-12234.

Cunningham, B.C. and Wells, J.A. (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science*, **244**, 1081-1085.

Deng, S.J., MacKenzie, C.R., Sadowska, J., Michniewicz, J., Young, N.M., Bundle, D.R. and Narang, S.A. (1994) Selection of antibody single-chain

variable fragments with improved carbohydrate binding by phage display. *J Biol Chem*, **269**, 9533-9538.

Dennis, M.S. and Lazarus, R.A. (1994) Kunitz domain inhibitors of tissue factor-factor VIIa. II. Potent and specific inhibitors by competitive phage selection. *J Biol Chem*, **269**, 22137-22144.

Din, N., Damude, H.G., Gilkes, N.R., Miller, R.C., Jr., Warren, R.A. and Kilburn, D.G. (1994) C1-Cx revisited: intramolecular synergism in a cellulase. *Proc Natl Acad Sci U S A*, **91**, 11383-11387.

Dower, W.J. and Mattheakis, L.C. (2002) In vitro selection as a powerful tool for the applied evolution of proteins and peptides. *Curr Opin Chem Biol*, **6**, 390-398.

Engvall, E., Jonsson, K. and Perlmann, P. (1971) Enzyme-linked immunosorbent assay. II. Quantitative assay of protein antigen, immunoglobulin G, by means of enzyme-labelled antigen and antibody-coated tubes. *Biochim Biophys Acta*, **251**, 427-434.

Feizi, T., Fazio, F., Chai, W. and Wong, C.H. (2003) Carbohydrate microarrays - a new set of technologies at the frontiers of glycomics. *Curr Opin Struct Biol*, **13**, 637-645.

FitzGerald, K. (2000) In vitro display technologies - new tools for drug discovery. *Drug Discov Today*, **5**, 253-258.

Flower, D.R., North, A.C. and Attwood, T.K. (1993) Structure and sequence relationships in the lipocalins and related proteins. *Protein Sci*, **2**, 753-761.

Fowler, R.G., Schaaper, R.M. and Glickman, B.W. (1986) Characterization of mutational specificity within the lacI gene for a mutD5 mutator strain of Escherichia coli defective in 3'----5' exonuclease (proofreading) activity. *J Bacteriol*, **167**, 130-137.

Francisco, J.A., Campbell, R., Iverson, B.L. and Georgiou, G. (1993) Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. *Proc Natl Acad Sci U S A*, **90**, 10444-10448.

Fransson, J., Ek, S., Ellmark, P., Soderlind, E., Borrebaeck, C.A. and Furebring, C. (2004) Profiling of internalizing tumor-associated antigens on breast and pancreatic cancer cells by reversed genomics. *Cancer Lett*, **208**, 235-242.

Fuchs, P., Breitling, F., Dubel, S., Seehaus, T. and Little, M. (1991) Targeting recombinant antibodies to the surface of Escherichia coli: fusion to a peptidoglycan associated lipoprotein. *Biotechnology (N Y)*, **9**, 1369-1372.

Fukui, S., Feizi, T., Galustian, C., Lawson, A.M. and Chai, W. (2002) Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate-protein interactions. *Nat Biotechnol*, **20**, 1011-1017.

Gilbert, H.J., Bolam, D.N. et al. (2002) An update on carbohydrate binding modules. In Teeri, T.T., Svensson, B., Gilbert, H.J. and Feizi, T. (eds),

Carbohydrate Bioengineering Interdisciplinary Approaches. Royal Society of Chemistry, Cambridge, pp. 89-98.

Gilkes, N.R., Warren, R.A., Miller, R.C., Jr. and Kilburn, D.G. (1988) Precise excision of the cellulose binding domains from two Cellulomonas fimi cellulases by a homologous protease and the effect on catalysis. *J Biol Chem*, **263**, 10401-10407.

Gill, D.S. and Damle, N.K. (2006) Biopharmaceutical drug discovery using novel protein scaffolds. *Curr Opin Biotechnol*.

Gram, H., Marconi, L.A., Barbas, C.F., 3rd, Collet, T.A., Lerner, R.A. and Kang, A.S. (1992) In vitro selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. *Proc Natl Acad Sci U S A*, **89**, 3576-3580.

Hanes, J. and Plückthun, A. (1997) In vitro selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci U S A*, **94**, 4937-4942.

Hanes, J., Schaffitzel, C., Knappik, A. and Pluckthun, A. (2000) Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat Biotechnol*, **18**, 1287-1292.

Hawkins, R.E., Russell, S.J. and Winter, G. (1992) Selection of phage antibodies by binding affinity. Mimicking affinity maturation. *J Mol Biol*, **226**, 889-896.

Hey, T., Fiedler, E., Rudolph, R. and Fiedler, M. (2005) Artificial, non-antibody binding proteins for pharmaceutical and industrial applications. *Trends Biotechnol*, **23**, 514-522.

Hildén, L., Daniel, G. and Johansson, G. (2003) Use of a fluorescence labelled, carbohydrate-binding module from Phanerochaete chrysosporium Cel7D for studying wood cell wall ultrastructure. *Biotechnology Letters*, **25**, 553-558.

Hoess, R.H. (2001) Protein design and phage display. *Chem Rev*, **101**, 3205-3218.

Holt, L.J., Enever, C., de Wildt, R.M. and Tomlinson, I.M. (2000) The use of recombinant antibodies in proteomics. *Curr Opin Biotechnol*, **11**, 445-449.

Holt, L.J., Herring, C., Jespers, L.S., Woolven, B.P. and Tomlinson, I.M. (2003) Domain antibodies: proteins for therapy. *Trends Biotechnol*, **21**, 484-490.

Hoogenboom, H.R. (2002) Overview of antibody phage-display technology and its applications. *Methods Mol Biol*, **178**, 1-37.

Hoogenboom, H.R. and Winter, G. (1992) By-passing immunisation. Human antibodies from synthetic repertoires of germline VH gene segments rearranged in vitro. *J Mol Biol*, **227**, 381-388.

Hosse, R.J., Rothe, A. and Power, B.E. (2006) A new generation of protein display scaffolds for molecular recognition. *Protein Sci*, **15**, 14-27.

Hudson, P.J. and Souriau, C. (2003) Engineered antibodies. *Nat Med*, **9**, 129-134.

Hutchison, C.A., 3rd, Phillips, S., Edgell, M.H., Gillam, S., Jahnke, P. and Smith, M. (1978) Mutagenesis at a specific position in a DNA sequence. *J Biol Chem*, **253**, 6551-6560.

Johansen, L.K., Albrechtsen, B., Andersen, H.W. and Engberg, J. (1995) pFab: a new, efficient vector for expression of antibody Fab fragments displayed on phages. *Protein Eng.*, **8**, 1063-1067.

Johansson, R., Gunnarsson, L.C., Ohlin, M. and Ohlson, S. (2006) Thermostable carbohydrate-binding modules in affinity chromatography. *J Mol Recognit*, **19**, 275-281.

Jung, S., Honegger, A. and Pluckthun, A. (1999) Selection for improved protein stability by phage display. *J Mol Biol*, **294**, 163-180.

Kavoosi, M., Meijer, J., Kwan, E., Creagh, A.L., Kilburn, D.G. and Haynes, C.A. (2004) Inexpensive one-step purification of polypeptides expressed in Escherichia coli as fusions with the family 9 carbohydrate-binding module of xylanase 10A from T. maritima. *J Chromatogr B Analyt Technol Biomed Life Sci*, **807**, 87-94.

Kehoe, J.W. and Kay, B.K. (2005) Filamentous phage display in the new millennium. *Chem Rev*, **105**, 4056-4072.

Kittur, F.S., Mangala, S.L., Rus'd, A.A., Kitaoka, M., Tsujibo, H. and Hayashi, K. (2003) Fusion of family 2b carbohydrate-binding module increases the catalytic activity of a xylanase from Thermotoga maritima to soluble xylan. *FEBS Lett*, **549**, 147-151.

Knappik, A., Ge, L. et al. (2000) Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J Mol Biol*, **296**, 57-86.

Knecht, W., Willemse, J. et al. (2006) Limited mutagenesis increases the stability of human carboxypeptidase U (TAFIa) and demonstrates the importance of CPU stability over proCPU concentration in down-regulating fibrinolysis. *Febs J*, **273**, 778-792.

Koide, A., Bailey, C.W., Huang, X. and Koide, S. (1998) The fibronectin type III domain as a scaffold for novel binding proteins. *J Mol Biol*, **284**, 1141-1151.

Kretzschmar, T. and von Ruden, T. (2002) Antibody discovery: phage display. *Curr Opin Biotechnol*, **13**, 598-602.

Kupsch, J.M., Tidman, N.H., Kang, N.V., Truman, H., Hamilton, S., Patel, N., Newton Bishop, J.A., Leigh, I.M. and Crowe, J.S. (1999) Isolation of human tumor-specific antibodies by selection of an antibody phage library on melanoma cells. *Clin Cancer Res*, **5**, 925-931.

Kwan, E., Guarna, M.M., Boraston, A.B., Gilkes, N.R., Haynes, C.A., Kilburn, D.G. and Warren, R.A. (2002) Self-activating factor X derivative fused to the C-terminus of a cellulose-binding module: Production and properties. *Biotechnol Bioeng*, **79**, 724-732.

Kwan, E.M., Boraston, A.B., McLean, B.W., Kilburn, D.G. and Warren, R.A. (2005) N-Glycosidase-carbohydrate-binding module fusion proteins as immobilized enzymes for protein deglycosylation. *Protein Eng Des Sel*, **18**, 497-501.

Ladner, R.C. and Ley, A.C. (2001) Novel frameworks as a source of high-affinity ligands. *Curr Opin Biotechnol*, **12**, 406-410.

Lander, E.S., Linton, L.M. et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.

Lehmann, M. and Wyss, M. (2001) Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Curr Opin Biotechnol*, **12**, 371-375.

Lehtiö, J., Teeri, T.T. and Nygren, P.A. (2000) Alpha-amylase inhibitors selected from a combinatorial library of a cellulose binding domain scaffold. *Proteins*, **41**, 316-322.

Leung, D.W., Chen, E. and Goeddel, D.V. (1989) A method for random mutagenesis of a defined DNA segment using modified polymerase chain reaction. *Technique* **1**, 11-15.

Levy, I. and Shoseyov, O. (2002) Cellulose-binding domains: biotechnological applications. *Biotechnol Adv*, **20**, 191-213.

Linn, S. and Arber, W. (1968) Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. *Proc Natl Acad Sci U S A*, **59**, 1300-1306.

Lipovsek, D. and Plückthun, A. (2004) In-vitro protein evolution by ribosome display and mRNA display. *J Immunol Methods*, **290**, 51-67.

Lueking, A., Possling, A., Huber, O., Beveridge, A., Horn, M., Eickhoff, H., Schuchardt, J., Lehrach, H. and Cahill, D.J. (2003) A nonredundant human protein chip for antibody screening and serum profiling. *Mol Cell Proteomics*, **2**, 1342-1349.

MacCallum, R.M., Martin, A.C. and Thornton, J.M. (1996) Antibody-antigen interactions: contact analysis and binding site topography. *J Mol Biol*, **262**, 732-745.

Malabarba, M.G., Milia, E., Faretta, M., Zamponi, R., Pelicci, P.G. and Di Fiore, P.P. (2001) A repertoire library that allows the selection of synthetic SH2s with altered binding specificities. *Oncogene*, **20**, 5186-5194.

Mao, S., Gao, C., Lo, C.H., Wirsching, P., Wong, C.H. and Janda, K.D. (1999) Phage-display library selection of high-affinity human single-chain antibodies to tumor-associated carbohydrate antigens sialyl Lewisx and Lewisx. *Proc Natl Acad Sci U S A*, **96**, 6953-6958.

Marks, J.D., Hoogenboom, H.R., Bonnert, T.P., McCafferty, J., Griffiths, A.D. and Winter, G. (1991) By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J Mol Biol*, **222**, 581-597.

Mattheakis, L.C., Bhatt, R.R. and Dower, W.J. (1994) An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proc Natl Acad Sci U S A*, **91**, 9022-9026.

Maxam, A.M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proc Natl Acad Sci U S A*, **74**, 560-564.

McCafferty, J., Griffiths, A.D., Winter, G. and Chiswell, D.J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature*, **348**, 552-554.

McCartney, L., Blake, A.W., Flint, J., Bolam, D.N., Boraston, A.B., Gilbert, H.J. and Knox, J.P. (2006) Differential recognition of plant cell walls by microbial xylan-specific carbohydrate-binding modules. *Proc Natl Acad Sci U S A*, **103**, 4765-4770.

McCartney, L., Gilbert, H.J., Bolam, D.N., Boraston, A.B. and Knox, J.P. (2004) Glycoside hydrolase carbohydrate-binding modules as molecular probes for the analysis of plant cell wall polymers. *Analytical Biochemistry*, **326**, 49-54.

McCartney, L., Marcus, S.E. and Knox, J.P. (2005) Monoclonal antibodies to plant cell wall xylans and arabinoxylans. *Journal of Histochemistry & Cytochemistry*, **53**, 543-546.

McLean, B.W., Boraston, A.B., Brouwer, D., Sanaie, N., Fyfe, C.A., Warren, R.A., Kilburn, D.G. and Haynes, C.A. (2002) Carbohydrate-binding modules recognize fine substructures of cellulose. *J Biol Chem*, **277**, 50245-50254.

McLean, B.W., Bray, M.R., Boraston, A.B., Gilkes, N.R., Haynes, C.A. and Kilburn, D.G. (2000) Analysis of binding of the family 2a carbohydrate-binding module from Cellulomonas fimi xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. *Protein Eng*, **13**, 801-809.

Michaud, G.A., Salcius, M., Zhou, F., Bangham, R., Bonin, J., Guo, H., Snyder, M., Predki, P.F. and Schweitzer, B.I. (2003) Analyzing antibody specificity with whole proteome microarrays. *Nat Biotechnol*, **21**, 1509-1512.

Mullis, K.B. and Faloona, F.A. (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol*, **155**, 335-350.

Muyldermans, S. (2001) Single domain camel antibodies: current status. *J Biotechnol*, **74**, 277-302.

Nagy, T., Simpson, P., Williamson, M.P., Hazlewood, G.P., Gilbert, H.J. and Orosz, L. (1998) All three surface tryptophans in Type IIa cellulose binding domains play a pivotal role in binding both soluble and insoluble ligands. *FEBS Lett*, **429**, 312-316.

Najmudin, S., Guerreiro, C.I. et al. (2006) Xyloglucan is recognized by carbohydrate-binding modules that interact with beta-glucan chains. *J Biol Chem*, **281**, 8815-8828.

Nemoto, N., Miyamoto-Sato, E., Husimi, Y. and Yanagawa, H. (1997) In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett*, **414**, 405-408.

Nord, K., Gunneriusson, E., Ringdahl, J., Stahl, S., Uhlen, M. and Nygren, P.A. (1997) Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nat Biotechnol*, **15**, 772-777.

Nord, K., Gunneriusson, E., Uhl inverted question marken, M. and Nygren, P.A. (2000) Ligands selected from combinatorial libraries of protein A for use in affinity capture of apolipoprotein A-1M and taq DNA polymerase. *J Biotechnol*, **80**, 45-54.

Nord, K., Nilsson, J., Nilsson, B., Uhlen, M. and Nygren, P.A. (1995) A combinatorial library of an alpha-helical bacterial receptor domain. *Protein Eng*, **8**, 601-608.

Nordberg Karlsson, E., Bartonek-Roxa, E. and Holst, O. (1997) Cloning and sequence of a thermostable multidomain xylanase from the bacterium Rhodothermus marinus. *Biochim Biophys Acta*, **1353**, 118-124.

Notenboom, V., Boraston, A.B., Chiu, P., Freelove, A.C., Kilburn, D.G. and Rose, D.R. (2001) Recognition of cello-oligosaccharides by a family 17 carbohydrate-binding module: an X-ray crystallographic, thermodynamic and mutagenic study. *J Mol Biol*, **314**, 797-806.

Nygren, P.A. and Skerra, A. (2004) Binding proteins from alternative scaffolds. *J Immunol Methods*, **290**, 3-28.

Nygren, P.A. and Uhlén, M. (1997) Scaffolds for engineering novel binding sites in proteins. *Curr Opin Struct Biol*, **7**, 463-469.

O'Connell, D., Becerril, B., Roy-Burman, A., Daws, M. and Marks, J.D. (2002) Phage versus phagemid libraries for generation of human monoclonal antibodies. *J Mol Biol*, **321**, 49-56.

Ofir, K., Berdichevsky, Y., Benhar, I., Azriel-Rosenfeld, R., Lamed, R., Barak, Y., Bayer, E.A. and Morag, E. (2005) Versatile protein microarray based on carbohydrate-binding modules. *Proteomics*, **5**, 1806-1814.

Padlan, E.A., Abergel, C. and Tipper, J.P. (1995) Identification of specificity-determining residues in antibodies. *Faseb J*, **9**, 133-139.

Patrick, W.M. and Firth, A.E. (2005) Strategies and computational tools for improving randomized protein libraries. *Biomol Eng*, **22**, 105-112.

Pell, G., Williamson, M.P., Walters, C., Du, H., Gilbert, H.J. and Bolam, D.N. (2003) Importance of Hydrophobic and Polar Residues in Ligand Binding in the Family 15 Carbohydrate-Binding Module from *Cellovibrio japonicus* Xyn10C. *Biochemistry*, **42**, 9316-9323.

Pennisi, E. (2005) Why do humans have so few genes? *Science*, **309**, 80.

Persson, H., Lantto, J. and Ohlin, M. (2006) A focused antibody library for improved hapten recognition. *J Mol Biol*, **357**, 607-620.

Ponyi, T., Szabo, L., Nagy, T., Orosz, L., Simpson, P.J., Williamson, M.P. and Gilbert, H.J. (2000) Trp22, Trp24, and Tyr8 play a pivotal role in the binding of the family 10 cellulose-binding module from Pseudomonas xylanase A to insoluble ligands. *Biochemistry*, **39**, 985-991.

Puhlmann, J., Bucheli, E., Swain, M.J., Dunning, N., Albersheim, P., Darvill, A.G. and Hahn, M.G. (1994) Generation of monoclonal antibodies against plant cell-wall polysaccharides. I. Characterization of a monoclonal antibody to a terminal alpha-(1-->2)-linked fucosyl-containing epitope. *Plant Physiol*, **104**, 699-710.

Quiocho, F.A. (1986) Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. *Annu Rev Biochem*, **55**, 287-315.

Reiersen, H., Lobersli, I. et al. (2005) Covalent antibody display--an in vitro antibody-DNA library selection system. *Nucleic Acids Res*, **33**, e10.

Reinikainen, T., Takkinen, K. and Teeri, T.T. (1997) Comparison of the adsorption properties of a single-chain antibody fragment fused to a fungal or bacterial cellulose-binding domain. *Enzyme Microb Tech*, **20**, 143-149.

Roberts, R.W. and Szostak, J.W. (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc Natl Acad Sci U S A*, **94**, 12297-12302.

Rodriguez, B., Kavoosi, M., Koska, J., Creagh, A.L., Kilburn, D.G. and Haynes, C.A. (2004) Inexpensive and generic affinity purification of recombinant proteins using a family 2a CBM fusion tag. *Biotechnol Prog*, **20**, 1479-1489.

Rönnmark, J., Gronlund, H., Uhlen, M. and Nygren, P.A. (2002) Human immunoglobulin A (IgA)-specific ligands from combinatorial engineering of protein A. *Eur J Biochem*, **269**, 2647-2655.

Sandström, K., Xu, Z., Forsberg, G. and Nygren, P.A. (2003) Inhibition of the CD28-CD80 co-stimulation signal by a CD28-binding affibody ligand developed by combinatorial protein engineering. *Protein Eng*, **16**, 691-697.

Sanger, F. and Coulson, A.R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*, **94**, 441-448.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, **74**, 5463-5467.

Schlehuber, S., Beste, G. and Skerra, A. (2000) A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin. *J Mol Biol*, **297**, 1105-1120.

Schlehuber, S. and Skerra, A. (2005) Lipocalins in drug discovery: from natural ligand-binding proteins to "anticalins". *Drug Discov Today*, **10**, 23-33.

Schwarz, M., Spector, L., Gargir, A., Shtevi, A., Gortler, M., Altstock, R.T., Dukler, A.A. and Dotan, N. (2003) A new kind of carbohydrate array, its use for profiling antiglycan antibodies, and the discovery of a novel human cellulose-binding antibody. *Glycobiology*, **13**, 749-754.

Shoseyov, O., Shani, Z. and Levy, I. (2006) Carbohydrate binding modules: biochemical properties and novel applications. *Microbiol Mol Biol Rev*, **70**, 283-295.

Siegel, D.L., Chang, T.Y., Russell, S.L. and Bunya, V.Y. (1997) Isolation of cell surface-specific human monoclonal antibodies using phage display and magnetically-activated cell sorting: applications in immunohematology. *J Immunol Methods*, **206**, 73-85.

Simpson, P.J., Jamieson, S.J., Abou-Hachem, M., Karlsson, E.N., Gilbert, H.J., Holst, O. and Williamson, M.P. (2002) The solution structure of the CBM4-2 carbohydrate binding module from a thermostable Rhodothermus marinus xylanase. *Biochemistry*, **41**, 5712-5719.

Simpson, P.J., Xie, H., Bolam, D.N., Gilbert, H.J. and Williamson, M.P. (2000) The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. *J Biol Chem*, **275**, 41137-41142.

Skerra, A. (2000a) Engineered protein scaffolds for molecular recognition. *J Mol Recognit*, **13**, 167-187.

Skerra, A. (2000b) Lipocalins as a scaffold. *Biochim Biophys Acta*, **1482**, 337-350.

Skerra, A. (2001) 'Anticalins': a new class of engineered ligand-binding proteins with antibody-like properties. *J Biotechnol*, **74**, 257-275.

Skerra, A. and Plückthun, A. (1988) Assembly of a functional immunoglobulin Fv fragment in Escherichia coli. *Science*, **240**, 1038-1041.

Smith, G.P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, **228**, 1315-1317.

Smith, G.P., Patel, S.U., Windass, J.D., Thornton, J.M., Winter, G. and Griffiths, A.D. (1998) Small binding proteins selected from a combinatorial repertoire of knottins displayed on phage. *J Mol Biol*, **277**, 317-332.

Smith, G.P. and Petrenko, V.A. (1997) Phage Display. *Chem Rev*, **97**, 391-410.

Söderlind, E., Carlsson, R., Borrebaeck, C.A. and Ohlin, M. (2001) The immune diversity in a test tube--non-immunised antibody libraries and functional variability in defined protein scaffolds. *Comb Chem High Throughput Screen*, **4**, 409-416.

Söderlind, E., Strandberg, L. et al. (2000) Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nat Biotechnol*, **18**, 852-856.

Southall, S.M., Simpson, P.J., Gilbert, H.J., Williamson, G. and Williamson, M.P. (1999) The starch-binding domain from glucoamylase disrupts the structure of starch. *FEBS Lett*, **447**, 58-60.

Starwalt, S.E., Masteller, E.L., Bluestone, J.A. and Kranz, D.M. (2003) Directed evolution of a single-chain class II MHC product by yeast display. *Protein Eng*, **16**, 147-156.

Stemmer, W.P. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, **370**, 389-391.

Sundberg, S.A. (2000) High-throughput and ultra-high-throughput screening: solution- and cell-based approaches. *Curr Opin Biotechnol*, **11**, 47-53.

Szabo, L., Jamal, S., Xie, H., Charnock, S.J., Bolam, D.N., Gilbert, H.J. and Davies, G.J. (2001) Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation. *J Biol Chem*, **276**, 49061-49065.

Takeo, K. (1984) Affinity electrophoresis: Principles and applications. *Electrophoresis*, **5**, 187-195.

Tawfik, D.S. and Griffiths, A.D. (1998) Man-made cell-like compartments for molecular evolution. *Nat Biotechnol*, **16**, 652-656.

Tindall, K.R. and Kunkel, T.A. (1988) Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. *Biochemistry*, **27**, 6008-6013.

Tomme, P., Boraston, A. et al. (1998) Characterization and affinity applications of cellulose-binding domains. *J Chromatogr B Biomed Sci Appl*, **715**, 283-296.

Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T. and Claeyssens, M. (1988) Studies of the cellulolytic system of Trichoderma reesei QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur J Biochem*, **170**, 575-581.

Tormo, J., Lamed, R., Chirino, A.J., Morag, E., Bayer, E.A., Shoham, Y. and Steitz, T.A. (1996) Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *Embo J*, **15**, 5739-5751.

Van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R. and Pettersson, G. (1986) Limited proteolysis of cellbiohydrolase I from *Trichoderma reesei*: Separation of functional domains. *FEBS Lett*, **204**, 223-227.

Vanhercke, T., Ampe, C., Tirry, L. and Denolf, P. (2005) Reducing mutational bias in random protein libraries. *Anal Biochem*, **339**, 9-14.

Vaughan, T.J., Williams, A.J. et al. (1996) Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nat Biotechnol*, **14**, 309-314.

Venter, J.C., Adams, M.D. et al. (2001) The sequence of the human genome. *Science*, **291**, 1304-1351.

Vogt, M. and Skerra, A. (2004) Construction of an artificial receptor protein ("anticalin") based on the human apolipoprotein D. *Chembiochem*, **5**, 191-199.

Wahlberg, E., Lendel, C., Helgstrand, M., Allard, P., Dincbas-Renqvist, V., Hedqvist, A., Berglund, H., Nygren, P.A. and Hard, T. (2003) An affibody in complex with a target protein: structure and coupled folding. *Proc Natl Acad Sci U S A*, **100**, 3185-3190.

Wang, A.A., Mulchandani, A. and Chen, W. (2002a) Specific adhesion to cellulose and hydrolysis of organophosphate nerve agents by a genetically

engineered Escherichia coli strain with a surface-expressed cellulose-binding domain and organophosphorus hydrolase. *Appl Environ Microbiol*, **68**, 1684-1689.

Wang, D., Liu, S., Trummer, B.J., Deng, C. and Wang, A. (2002b) Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. *Nat Biotechnol*, **20**, 275-281.

Wang, J., Stuckey, J.A., Wishart, M.J. and Dixon, J.E. (2002c) A unique carbohydrate binding domain targets the lafora disease phosphatase to glycogen. *J Biol Chem*, **277**, 2377-2380.

Wark, K.L. and Hudson, P.J. (2006) Latest technologies for the enhancement of antibody affinity. *Adv Drug Deliv Rev*, **58**, 657-670.

Webster, D.M., Henry, A.H. and Rees, A.R. (1994) Antibody-antigen interactions. *Curr Opin Struct Biol*, **4**, 123-129.

Wiederstein, M. and Sippl, M.J. (2005) Protein sequence randomization: efficient estimation of protein stability using knowledge-based potentials. *J Mol Biol*, **345**, 1199-1212.

Wikman, M., Steffen, A.C., Gunneriusson, E., Tolmachev, V., Adams, G.P., Carlsson, J. and Stahl, S. (2004) Selection and characterization of HER2/neu-binding affibody ligands. *Protein Eng Des Sel*, **17**, 455-462.

Willats, W.G. (2002) Phage display: practicalities and prospects. *Plant Mol Biol*, **50**, 837-854.

Willats, W.G.T., Steele-King, C.G., McCartney, L., Orfila, C., Marcus, S.E. and Knox, J.P. (2000) Making and using antibody probes to study plant cell walls. *Plant Physiology and Biochemistry*, **38**, 27-36.

Wingren, C. and Borrebaeck, C.A. (2004) High-throughput proteomics using antibody microarrays. *Expert Rev Proteomics*, **1**, 355-364.

Xie, H., Bolam, D.N., Nagy, T., Szabo, L., Cooper, A., Simpson, P.J., Lakey, J.H., Williamson, M.P. and Gilbert, H.J. (2001a) Role of hydrogen bonding in the interaction between a xylan binding module and xylan. *Biochemistry*, **40**, 5700-5707.

Xie, H., Gilbert, H.J. et al. (2001b) Clostridium thermocellum Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding. *Biochemistry*, **40**, 9167-9176.

Xu, L., Aha, P. et al. (2002) Directed evolution of high-affinity antibody mimics using mRNA display. *Chem Biol*, **9**, 933-942.

Zahnd, C., Spinelli, S., Luginbuhl, B., Amstutz, P., Cambillau, C. and Pluckthun, A. (2004) Directed in vitro evolution and crystallographic analysis of a peptide-binding single chain antibody fragment (scFv) with low picomolar affinity. *J Biol Chem*, **279**, 18870-18877.

Zebedee, S.L., Barbas, C.F., 3rd et al. (1992) Human combinatorial antibody libraries to hepatitis B surface antigen. *Proc Natl Acad Sci U S A*, **89**, 3175-3179.

Zverlov, V.V., Volkov, I.Y., Velikodvorskaya, G.A. and Schwarz, W.H. (2001) The binding pattern of two carbohydrate-binding modules of laminarinase Lam16A from Thermotoga neapolitana: differences in beta-glucan binding within family CBM4. *Microbiology*, **147**, 621-629.