Functional studies and engineering of family 1 carbohydrate-binding modules

Janne Lehtio

Royal Institute of Technology
Department of Biotechnology
Stockholm 2001
Functional studies and engineering of family 1 carbohydrate-binding modules

Janne Lehtiö

Royal Institute of Technology
Department of Biotechnology
Stockholm 2001
Cover photograph: Recombinant Staphylococcus carnosus, expressing Trichoderma reesei Cel6A CBM1 on the surface, immobilized on cotton fibers.

© Janne Lehtiö

Department of Biotechnology
Royal Institute of Technology (KTH)
SE-106 91
Sweden

Printed at Universitetsservice US AB, Stockholm 2001

ISBN 91-7283-152-9
TABLE OF CONTENTS

INTRODUCTION 4

1. STRUCTURE-FUNCTION RELATIONS OF CARBOHYDRATE-PROTEIN INTERACTIONS 4
   1.1 Group I carbohydrate-binding proteins 5
   1.2 Group II carbohydrate-binding proteins 7
      1.2.1 Carbohydrate-binding modules 7

2. CELLULOSE-BINDING MODULES 8
   2.1 Cellulose Structure 11
   2.2 Binding-modules with affinity towards crystalline cellulose 12
   2.4 Role of the cellulose-binding modules 19

3. FAMILY I CELLULOSE-BINDING DOMAINS 20
   3.1 The fold of the family 1 cellulose-binding modules 21
   3.2 Characterization and Function of family 1 CBMs 22

4. AIMS OF THIS STUDY 24

RESULTS AND DISCUSSION 24

5. FAMILY I CELLULOSE-BINDING MODULE FUNCTION AND INTERACTION WITH CELLULOSE (I, II) 24
   5.1 The importance of being tryptophan: mutation studies of family 1 CBMs 25
   5.2 Binding specificity of family 1 CBMs: electron microscopic experiments 29
   5.3 CBM function as a part of an intact enzyme 32

6. CELLULOSE-BINDING MODULES IN AFFINITY APPLICATIONS (III, IV) 33
   6.1 Role of the linker peptide in native and engineered enzymes 33
   6.2 Bacterial cell immobilization 34

7. PROTEIN ENGINEERING OF FAMILY 1 CARBOHYDRATE-BINDING MODULES (V, VI) 36
   7.1 Phage display of CBM1 for novel binding properties 38
      7.1.1 Selection of a protein scaffold 38
      7.1.2 Construction and characterization of the CBM1 library 39
      7.1.3 Selection of binders from the randomized CBM1 library 40

8. CONCLUDING REMARKS 41

9. ABBREVIATIONS 43

10. ACKNOWLEDGEMENTS 44

11. REFERENCES 46
Abstract

The family 1 cellulose-binding modules (CBM1) form a group of small, stable carbohydrate-binding proteins. These modules are essential for fungal cellulose degradation. This thesis describes both functional studies of the CBM1s as well as protein engineering of the modules for several objectives.

The characteristics and specificity of CBM1s from the *Trichoderma reesei* Cel7A and Cel6A, along with several other wild type and mutated CBMs, were studied using binding experiments and transmission electron microscopy (TEM). Data from the binding studies confirmed that the presence of one tryptophan residue on the CBM1 binding face enhances its binding to crystalline cellulose. The two *T. reesei* CBM1s as well as the CBM3 from the *Clostridium thermocellum* CipA were investigated by TEM experiments. All three CBMs were found to bind in linear arrangements along the sides of the fibrils. Further analyses of the bound CBMs indicated that the CBMs bind to the exposed hydrophobic surfaces, the so-called (200) crystalline face of *Valonia* cellulose crystals.

The function and specificity of CBM1s as a part of an intact enzyme were studied by replacing the CBM from the exo-acting Cel7A by the CBM1 from the endoglucanase Cel7B. Apart from slightly improved affinity of the hybrid enzyme, the module exchange did not significantly influence the function of the Cel7A. This indicates that the two CBM1s are analogous in their binding properties and function during cellulose hydrolysis.

The CBM1 was also used for immobilization studies. To improve heterologous expression of a CBM1-lipase fusion protein, a linker stability study was carried out in *Pichia pastoris*. A proline/threonine rich linker peptide was found to be stable for protein production in this host. For whole bacterial cell immobilization, the *T. reesei* Cel6A CBM1 was expressed on the surface of the gram-positive bacteria, *Staphylococcus carnosus*. The engineered *S. carnosus* cells were shown to bind to cellulose fibers.

To exploit the stable CBM1 fold as a starting point for generating novel binders, a phage display library was constructed. Binding proteins against an amylase as well as against a metal ion were selected from the library. The amylase-binding proteins were found to bind and inhibit the target enzyme. The metal binding proteins selected from the library were cloned on the surface of the *S. carnosus* and clearly enhanced the metal binding ability of the engineered bacteria.

Keywords: cellulose-binding, family 1 carbohydrate-binding module, phage display, bacterial surface display, combinatorial protein library, metal binding, protein engineering, *Trichoderma reesei*, *Staphylococcus carnosus*.

© Janne Lehtiö
LIST OF PUBLICATIONS

This thesis is based on the following papers, which in text are referred to by their Roman numerals:

I. Lehtiö J., Sugiyama J., Gustavsson M., Fransson L. and Teeri T.T. The binding specificity and affinity determinants of the family I cellulose-binding domains from *Trichoderma reesei* Cel6A and Cel7A. Manuscript.


INTRODUCTION

Carbohydrates play a key part in essentially all living organisms where their roles expand from structural elements of cell walls to the delicate control of protein folding. The protein-carbohydrate interactions regulate countless activities in living cells. Glycoproteins in eucaryotic cells maintain and regulate cell-cell interactions, control the cellular immune system, are involved in protein targeting as well as in glycoprotein turnover in the blood stream, to name a few examples of the different functions. In all living organisms carbohydrates are used for energy storage and in many cases as a structural component of the cells. In plants, the energy captured by photosynthesis is converted to chemical energy and stored in carbohydrates. Marine and terrestrial plant cell walls, composed of mainly cellulose and hemicellulose, form the greatest source of biomass on earth.

The ability of proteins to create specific interactions with soluble and solid carbohydrates is a key to the sugar metabolism. Proteins that specifically recognize various sugar structures maintain the above mentioned functions and are essential in the metabolic pathways for the generation and degradation of carbohydrates. Lectins, monosaccharide transport proteins and substrate binding modules of carbohydrate-active enzymes form a large and heterogeneous group of proteins capable of specific recognition and binding of saccharides and polysaccharides. This introduction will concentrate on non-catalytic carbohydrate-protein interactions. The focus will be on carbohydrate-binding modules, which form a defined part of a carbohydrate-active enzyme, with special emphasis on cellulose-binding modules. The carbohydrate-active enzymes will be only briefly discussed as far as the substrate binding is concerned.

1. Structure-function relations of carbohydrate-protein interactions

Carbohydrate-binding proteins can be divided into two broad groups on the basis of their binding site topology and location (Drickamer, 1989; Quiocho, 1986; Rini, 1995). The first group consists of proteins which have a buried binding site for carbohydrates. This group contains for example periplasmic monosaccharide-binding proteins, membrane bound sugar transporters and enzymes such as hexokinases and exoglucanases. The second group of
binding proteins is composed of binders that feature a more open binding site topology on the surface of the protein. This second group contains proteins such as lectins, endoglucanases, immunoglobulins and substrate-binding modules of enzymes. Numerous proteins and domains in this second category bind to insoluble carbohydrates. The binding sites of such binders are by and large a shallow groove or a flat surface on the binding domain. The topology of the binding site reflects the binding affinity and function of the protein. High affinity binders tend to have sugar binding sites in a deep pocket or tunnel. This is typically seen for example in the processive enzymes, which need to hold on to the substrate during the hydrolytic process. On the other hand, proteins with a broad specificity, such as lectins, have a shallow binding site.

Despite the various binding-site topologies of the sugar-binding proteins, shared principles as to how they achieve molecular recognition and affinity towards saccharides can be found. Van der Waals interactions arising from stacking aromatic residues of the protein onto the hydrophobic pyranose rings of the sugar often have a critical role in the binding. The other common binding interaction is based on hydrogen bonding between the polar groups on saccharides and amino acid side chains.

1.1 Group I carbohydrate-binding proteins

Proteins with a deep cleft or a tunnel for sugar binding exhibit somewhat different binding interactions than the proteins with a shallow cleft or a planar binding surface. Hydrogen bonding is extensive and the hydroxyl groups of the oligosaccharides are often involved in multiple hydrogen bonds with several amino acid side chains (Quiocho et al., 1997). Aromatic stacking is also involved, often resulting in high affinity and in some cases also contributing to specificity by steric exclusion. The dissociation constants ($K_d$) of this type of carbohydrate-binding interactions are typically in the nanomolar range. This is significantly higher than that observed with other types of mono- and disaccharide-binders, such as lectins (Table I). For example, studies of the bacterial maltodextrin/maltose-binding protein have revealed a high affinity sugar protein interaction (Quiocho et al., 1997; Spurlino et al., 1992). In this case, two of the studied sugar molecule substrates, maltose and maltotriose, are almost totally engulfed by the protein, exposing less than 5% of the bound molecule to the solvent. A relatively large substrate induced conformational change in the protein is postulated to occur upon binding. Both, extensive hydrogen bond networks and aromatic stacking contribute to
the binding. Very similar binding interactions are reported with other sugar-binding protein receptors, such as D-allose binding protein (Chaudhuri et al., 1999) and D-glucose D-galactose-binding protein (Appleyard et al., 2000). The binding event in these proteins seems to be enthalpy driven and steric fitting of the substrate in the tight binding pocket is an important asset in creating specificity.

**Table I:** Examples of carbohydrate-protein interactions and their affinities

<table>
<thead>
<tr>
<th>Protein</th>
<th>Substrate</th>
<th>Affinity $K_d$ (µM)</th>
<th>Referens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maltodextrin-binding protein</td>
<td>E. coli</td>
<td>Maltotriose</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Quijcho, et al., 1997)</td>
</tr>
<tr>
<td>L-Arabinose-binding protein</td>
<td>E. coli</td>
<td>L-Arabinose</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Quijcho, et al., 1997)</td>
</tr>
<tr>
<td>Convalin A</td>
<td>D. grandiflora</td>
<td>Mannotriose</td>
<td>3900</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Dam et al., 2000a)</td>
</tr>
<tr>
<td>CBM17Cel5A</td>
<td>C. cellulovorans</td>
<td>PASA</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Boraston et al., 2000)</td>
</tr>
<tr>
<td>CBM22Xyn10B</td>
<td>C. Thermocellum</td>
<td>Xylotriose</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Charnock et al., 2000)</td>
</tr>
<tr>
<td>CBM9Xyn10A</td>
<td>T. maritima</td>
<td>PASC</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Boraston et al., 2001)</td>
</tr>
<tr>
<td>CBM2aXyn10A</td>
<td>C. fimicola</td>
<td>BMCC</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(McLean et al., 2000)</td>
</tr>
</tbody>
</table>

Carbohydrate-active enzymes synthesizing and degrading sugars bind the substrate in their active sites following the same principles for the sugar binding. Their active sites often contain aromatic and polar residues that create the affinity to the substrate. As far as the binding site topology is concerned, similar structure-function relationship evident for the non-catalytic carbohydrate-binding proteins can also be detected on carbohydrate active enzymes. This is clearly demonstrated in cellulose-degrading enzymes, for which several structures are available for detailed structure-function studies of these enzymes (reviewed by Davies & Henrissat, 1995). In cellulases, the topology of the substrate-binding site typically reflects the functional characteristics of the enzyme. In the processive exoglucanases, the substrate-binding site is buried in a long tunnel within the catalytic module. Such an active site allows the enzyme to maintain the substrate chain attached to the active site hence enabling the processive hydrolysis (Divne et al., 1994; Divne et al., 1998). According to the current hypothesis, a sugar chain is guided into and held in the active site tunnel by several aromatic residues and an extensive hydrogen bonding network. In related endoglucanases, the binding...
site is located in a cleft. With the more open active site, the enzyme can cleave the substrate from the middle of the chain and then supposedly dissociate to find the next cleavage site. The β-glucosidases frequently have a pocket-like binding site topology, which is optimal for more accessible chains and facilitates more specific recognition of small soluble substrates.

1.2 Group II carbohydrate-binding proteins

The second major group of proteins binding carbohydrates consists of lectin families, viral proteins (e.g. hemagglutinin, foot-and-mouth disease capsid proteins), toxins, for example the cholera toxin, and substrate-binding modules, such as the cellulose-binding modules. The topology of the sugar-binding site in these proteins is a shallow groove or a flat surface. Lectins form a large group of proteins with diverse biological functions. They are known to be essential in cell adhesion, cell activation, growth, apoptosis, to initiate immune response among other functions. The evolutionary diverse lectins are a good example of the common principles of the carbohydrate-protein interactions achieved by proteins with different origin and fold. Whatever the origin, the binding site of lectins is typically a shallow groove. Similar to the monosugar-binding proteins, the affinity and specificity in ligand binding is achieved through Van der Waals interactions using aromatic side chains and hydrogen bonds. Metal coordination is also used for carbohydrate interactions in some lectins, such as the mammalian type C-lectins (Feinberg et al., 2000). The affinity of lectins is relatively low, often in a millimolar range. For example, the type 1 galectins have affinities around $1 \times 10^3 - 5 \times 10^4 \text{ M}^{-1}$ (Ramkumar et al., 1995). Even with a relatively low affinity, lectins can contribute to a tight interaction in cases when several lectins on the cell surface bind simultaneously to numerous glycans on the surface of another cell (Dam et al., 2000b; Sacchettini et al., 2001). Substrate-binding modules constitute another major subgroup of the group II carbohydrate-binding proteins. These modules are linked to the catalytic modules of carbohydrate-active enzymes. Characteristic to the binding modules so far studied is that they have a shallow binding groove or a flat binding surface.

1.2.1 Carbohydrate-binding modules

Carbohydrate-active enzymes hydrolyzing insoluble substrates commonly have a modular structure. The catalytic module or modules are coupled to one or often to several other modules. In some cases more than ten modules with either a known or an unknown function are linked together in one enzyme. One of the most frequently occurring modules in these
enzymes are carbohydrate-binding modules. By definition, a carbohydrate-binding module (CBM) is a contiguous amino acid sequence within a carbohydrate-active enzyme with a discreet fold and independent carbohydrate-binding activity (Coutinho & Henrissat, 1999). Modules that are part of a structural protein, such as the non-catalytic scaffolding protein in the cellulosomes, represent an exception to this definition (for a recent review look Shoham et al., 1999). The first characterized CBMs showed affinity towards cellulose (Gilkes et al., 1988; Meinke et al., 1991; Tomme et al., 1988). Sequencing of an ever-growing number of cellulose-binding modules (previously called domains) has led to their classification into families (Coutinho et al., 1992; Gilkes et al., 1991a and b; Tomme et al., 1995a). Since numerous new modules binding to a variety of carbohydrates have been discovered in recent years, the classification has been broadened by Henrissat and Coutinho to include all carbohydrate binding modules (Coutinho and Henrissat, 1999). Today, the CBMs are classified in 28 families based on amino acid sequence similarities among over 800 sequences of known or putative CBMs (Table II) (For the new carbohydrate-binding module nomenclature see page 9). The known CBMs bind to various carbohydrates, ranging from crystalline cellulose to soluble sugar residues and monosaccharides (Table II). Some of the families are further divided into subclasses. For example the subclass CBM2a binds crystalline cellulose while CBM2b binds oligosaccharides, amorphous cellulose and xylan (Bolam et al., 2001; Simpson et al., 2000; Xie et al., 2001). The CBMs are usually linked to the enzyme catalytic module by a flexible linker peptide. The CBMs show independent binding to the substrate when separated from the rest of the protein (Din et al., 1994b; Linder et al., 1995b). In some cases, a binding module forms a continued substrate-binding site with the catalytic module. Such arrangements are seen for example in starch-binding modules of amylases (MacGregor et al., 2001) and in the Thermofibida fusca endo/exo-acting cellulase (Sakon et al., 1997).

2. Cellulose-binding modules

Insoluble polycarbohydrates, such as cellulose, represent challenging substrates for enzymatic degradation. This is in part due to the poor accessibility of the tightly packed substrate to the catalytic site of the enzyme and partly due to heterogeneity of the physical structure, which varies from highly crystalline cellulose to isolated sugar chains. These substrate characteristics set special requirements for the organisms and enzymes hydrolyzing it. One of the evolutionary solutions for the enzymes degrading insoluble substrates has been to generate
distinct substrate-binding modules. Examples of such modules can be found commonly in carbohydrate-active enzymes such as cellulases, hemicellulases and chitinases, but solid substrate-binding domains are also found in other type of enzymes, such as collagenases (Matsushita et al., 2001; Overall, 2001). Several structures of the carbohydrate-binding modules have been solved (Table III and IV). So far, predominantly the β-sheet secondary structure is found in cellulose-binding modules. With known structures and substrate specificities, studies of the structure-function relations of the CBM-cellulose interactions have become possible.

### CBM nomenclature

<table>
<thead>
<tr>
<th>Old name</th>
<th>New name</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>T. reesei</em> CBHI CBD (CBD_{CBHI})</td>
<td><em>T. reesei</em> Cel7A CBM1 (CBM1_{Cel7A})</td>
</tr>
<tr>
<td><em>T. reesei</em> CBHII CBD (CBD_{CBHII})</td>
<td><em>T. reesei</em> Cel6A CBM1 (CBM1_{Cel6A})</td>
</tr>
<tr>
<td><em>T. reesei</em> EgI CBD (CBD_{EgI})</td>
<td><em>T. reesei</em> Cel7B CBM1 (CBM1_{Cel7B})</td>
</tr>
<tr>
<td><em>N. patriciarum</em> CBHII</td>
<td><em>N. patriciarum</em> Cel6A CBM1</td>
</tr>
<tr>
<td><em>C. thermocellum</em> CipA CBD</td>
<td><em>C. thermocellum</em> CipA CBM3</td>
</tr>
</tbody>
</table>

The nomenclature of cellulose-binding modules has been changed recently (previously called cellulose-binding domains). In the new nomenclature, the family number follows the CBM-abbreviation (for example family 1 modules are CBM1). For enzymes with more than one CBM, the CBMs get a second serial number starting from the N-terminus of the protein (for example CBM4-1 and CBM4-2). Some of the families are divided into sub-families. This is indicated by a lower case letter following the family number (such as CBM3a or CBM2b).
### Table II: Carbohydrate binding module classification

<table>
<thead>
<tr>
<th>Family</th>
<th>Present number of entries</th>
<th>Approximate size of the CBM (aa)</th>
<th>3D structures</th>
<th>Demonstrated binding substrates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71</td>
<td>30-40</td>
<td>yes</td>
<td>Cellulose, Chitin</td>
</tr>
<tr>
<td>2a</td>
<td>88</td>
<td>110</td>
<td>yes</td>
<td>Cellulose, Chitin</td>
</tr>
<tr>
<td>2b</td>
<td>55</td>
<td>90</td>
<td>yes</td>
<td>Xylan</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>150</td>
<td>yes</td>
<td>Cellulose, chitin</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>60</td>
<td>yes</td>
<td>Cellulose, Amorphous cellulose, Xylan</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>120</td>
<td>yes</td>
<td>Amorphous cellulose, Xylan</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>50</td>
<td>yes</td>
<td>Cellulose, Cellulose, mono-, di-, and Ologosaccharides</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>150</td>
<td>no</td>
<td>Cellulose</td>
</tr>
<tr>
<td>9</td>
<td>16</td>
<td>170</td>
<td>yes</td>
<td>Cellulose, mono-, di-, and Ologosaccharides</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>50</td>
<td>yes</td>
<td>Cellulose</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>180-200</td>
<td>no</td>
<td>Cellulose</td>
</tr>
<tr>
<td>12</td>
<td>63</td>
<td>40-60</td>
<td>no</td>
<td>Chitin, Cellulose</td>
</tr>
<tr>
<td>13</td>
<td>94</td>
<td>150</td>
<td>yes</td>
<td>Xylan, Mannose, GalNac</td>
</tr>
<tr>
<td>14</td>
<td>91</td>
<td>70</td>
<td>no</td>
<td>Chitin, Cellulose</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>40</td>
<td>no</td>
<td>Cellulose</td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>160</td>
<td>no</td>
<td>Xylan, Amorphous cellulose</td>
</tr>
<tr>
<td>17</td>
<td>8</td>
<td>200</td>
<td>no</td>
<td>Amorphous cellulose, Celloooligosaccharides</td>
</tr>
<tr>
<td>18</td>
<td>144</td>
<td>40</td>
<td>no</td>
<td>Chitin, Cellulose</td>
</tr>
<tr>
<td>19</td>
<td>4</td>
<td>60-70</td>
<td>no</td>
<td>Chitin</td>
</tr>
<tr>
<td>20</td>
<td>60</td>
<td>100</td>
<td>yes</td>
<td>Granular starch, Cyclodextrins</td>
</tr>
<tr>
<td>21</td>
<td>21</td>
<td>100</td>
<td>no</td>
<td>Starch</td>
</tr>
<tr>
<td>22</td>
<td>38</td>
<td>160</td>
<td>yes</td>
<td>Xylan</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>200</td>
<td>no</td>
<td>Soluble mannan</td>
</tr>
<tr>
<td>24</td>
<td>3</td>
<td>80</td>
<td>no</td>
<td>α-1,3-glucan</td>
</tr>
<tr>
<td>25</td>
<td>9</td>
<td>110</td>
<td>yes</td>
<td>Starch</td>
</tr>
<tr>
<td>26</td>
<td>9</td>
<td>100</td>
<td>no</td>
<td>Starch, Mannan</td>
</tr>
<tr>
<td>27</td>
<td>5</td>
<td>100</td>
<td>no</td>
<td>Mannan</td>
</tr>
<tr>
<td>28</td>
<td>8</td>
<td>200</td>
<td>no</td>
<td>Amorphous cellulose, Celloooligosaccharides</td>
</tr>
<tr>
<td>Summary</td>
<td>841</td>
<td>30-200</td>
<td>Referens:</td>
<td>(Coutinho and Henrissat, 1999)</td>
</tr>
</tbody>
</table>
2.1 Cellulose Structure

Cellulose is a homopolymer of β-1,4-linked D-glucose. The smallest repeating unit in cellulose is cellobiose, due to a rotation of every glucose ring by 180° relative to the adjacent glucose ring (Figure 1). The degree of polymerization of cellulose varies from about 100 to as many as 15 000 glucose units (Hon, 1994). The long linear glucose polymers crystallize to form microfibrils. Several electron microscopic studies of the microfibrils demonstrate that one microfibril can be considered as a single crystalline entity (Helbert et al., 1998; Sugiyama et al., 1985). Apart from highly crystalline cellulose, more disordered semicrystalline and amorphous cellulose are also present in various cellulose preparations and in the plant cell walls. Recent data has confirmed that naturally occurring cellulose crystals have a chain polarity with all reducing ends pointing in the same direction in the crystal (Koyama et al., 1997). Relatively pure cellulose is produced by bacteria such as Acetobacter xylium.

![Cellulose structure](image)

**Figure 1.** Cellulose structure. The glucose units are linked by β-1,4-linkages, with cellobiose as the repeating unit. The glucose chains are packed into cellulose crystal. The crystal faces are indicated on the figure. The gray short lines illustrate the packing of the individual chains within the crystal.

In nature, cellulose is most often found as a part of the plant cell walls where it forms complex structures with other wall polysaccharides, lignin and structural proteins. This heterogeneity of the cellulose and cellulose composites presents a structurally varying target for the binding modules and the enzymes acting on the cell walls.
2.2 Binding-modules with affinity towards crystalline cellulose

Cellulose-binding modules (CBM) can be roughly divided into two groups according to their substrate specificity: crystalline cellulose-binding modules and modules binding to amorphous cellulose and cellooligosaccharides. Some CBMs, such as the CBM2a from *P. fluorescens* and *C. fimi*, bind crystalline cellulose with high affinity but are also able to recognize amorphous cellulose (McLean *et al.*, 2000; Nagy *et al.*, 1998). With several experimental CBM structures available, interesting features of the binding site topology related to substrate specificity can be noted.

At present, there are several 3D structures for binding modules, which have demonstrated affinity towards crystalline cellulose (families 1, 2, 3, 5, 9 and 10) (Table III). The structures of all of these CBMs exhibit a planar binding face containing three aromatic amino acids, mainly tryptophans and tyrosines. The importance of these aromatic side chains for the binding to cellulose has been demonstrated by numerous studies (Bray *et al.*, 1996; Din *et al.*, 1994; Linder *et al.*, 1995a; Linder *et al.*, 1995b; McLean *et al.*, 2000; Nagy *et al.*, 1998; Ponyi *et al.*, 2000; Tormo *et al.*, 1996). Apart from the aromatic amino acids, a number of potential hydrogen bonding polar amino acids, such as glutamines and asparagines, are often conserved at close proximity to the aromatic side chains. In alanin scanning experiments these polar amino acids have been shown to play a part in the binding interaction. However, the effects of changing the polar residues are not as dramatic as the replacement in one of the aromatic amino acids (Linder *et al.*, 1995a; Linder *et al.*, 1995b; McLean *et al.*, 2000).

An interesting observation concerning the CBMs binding to crystalline cellulose is that some of them exhibit seemingly irreversible binding or at least a very slow off-rate from crystalline cellulose. This phenomenon has been shown for certain family 1, family 2 and family 3 CBMs, when the bound CBM has been eluted in non-equilibrium dilution experiments or in exchange studies using radioactively labeled and non-labeled CBMs (Carrard & Linder, 1999; Creagh *et al.*, 1996; Ong *et al.*, 1991). Some of the studied cellulose-binding modules, such as *T. reesei* CBM1Cel7A and *T. maritima* CBM9, bind fully reversibly and return to equilibrium as soon as the bound CBM is diluted with buffer (Boraston *et al.*, 2001; Linder & Teeri, 1996). In one study, where the *C. fimi* CBM2a was bound to cellulose, the tryptophans were nevertheless found to be accessible to oxidation with N-bromosuccinimide, although with a lowered reaction rate than those in an unbound CBM (Creagh *et al.*, 1996). The fact that these
aromatic side chains are not totally protected indicates that the binding is not fully irreversible. The authors postulate that, at a given time point, two of the three aromatic side chains are bound simultaneously, while the third aromatic side chain is not bound and therefore remains more accessible to oxidation. In another study, the *C. fimi* CBM2a was labeled with a fluorescent compound and adsorbed onto the crystalline cellulose surface, thereafter a part of the surface was photo-bleached. It was found that the bound CBMs diffused laterally along the surface, moving to the bleached parts, but without clear dissociation from the cellulose (Jervis *et al*., 1997). These experiments indicate that the crystalline cellulose binding modules are maybe capable of “sliding” over a surface containing a very high density of overlapping binding sites.

**Table III**: Published crystalline cellulose-binding module structures

<table>
<thead>
<tr>
<th>Family</th>
<th>CBM</th>
<th>Crystal/solution structure</th>
<th>Referens</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>T. reesei</em> Cel7A CBM1</td>
<td>NMR</td>
<td>(Kraulis <em>et al</em>., 1989)</td>
</tr>
<tr>
<td></td>
<td><em>T. reesei</em> Cel7B CBM1</td>
<td>NMR</td>
<td>(Mattinen <em>et al</em>., 1998)</td>
</tr>
<tr>
<td>2a</td>
<td><em>C. fimi</em> Xyn10A CBM2a</td>
<td>NMR</td>
<td>(Xu <em>et al</em>., 1995)</td>
</tr>
<tr>
<td>3a</td>
<td><em>C. thermocellum</em> CipC CBM3</td>
<td>Crystal structure</td>
<td>(Tormo <em>et al</em>., 1996)</td>
</tr>
<tr>
<td>5</td>
<td><em>E. crysantheni</em> Cel5A CBM5</td>
<td>NMR</td>
<td>(Brun <em>et al</em>., 1997)</td>
</tr>
<tr>
<td>10</td>
<td><em>P. fluorescens</em> Xynl0A CBM10</td>
<td>NMR</td>
<td>(Raghothama <em>et al</em>., 2000)</td>
</tr>
</tbody>
</table>

There are currently several hypothesis on how the CBM-crystalline cellulose interaction is achieved. It seems that many of these modules need a continuous crystalline surface for effective binding. It is generally accepted that the planar ring-ring stacking interaction between the aromatic side chains of the protein and the apolar face of the sugar pyranose rings governs the binding interaction. According to one binding model, the crystalline cellulose binding modules interact with the 200-surface of the crystal, where the pyranose rings are fully exposed to solvent (Brun *et al*., 1997; Mattinen *et al*., 1997; Reinikainen *et al*., 1995; Tormo *et al*., 1996). However, owing to the small estimated surface area of the (200) surface on a perfect crystal, another model has been proposed recently. According to this model, the CBMs bind onto the (110) and (1-10) faces of the cellulose crystal (McLean *et al*., 2000). On these faces, the pyranose rings are only partially solvent-exposed and form a sort of staircase, which might represent a binding site for the CBMs.
It has been demonstrated that the orientation of the aromatic side chains is critical for the binding interaction of the family 2b xylan-binding domain. In this case, a single mutation (R262G) changing the orientation of a binding site tryptophan changes substrate specificity of the CBM from xylan to cellulose (Simpson et al., 2000). This demonstrates the importance of structural complementary of the binding site on the protein and on the substrate. Also, when examining the structures, it seems compelling that the aromatic amino acids aligned on the planar surface of a CBM have almost exactly the same spacing as every second glucose unit on the cellulose surface (Reinikainen et al., 1992; Tormo et al., 1996) (Figure 2).

Thermodynamics of the *C. fimi* CBM2a-crystalline cellulose interaction has been studied by microcalorimetry experiments (Creagh et al., 1996). It has been postulated that the dehydration of the CBM binding surface and the cellulose crystal surface are the entropic driving force of the binding (Creagh et al., 1996). Furthermore, the rigidity of the CBM binding surface seems to be an important asset, since a disulfide deletion in the *T. reesei* CBM1<sub>Cel6A</sub> weakens the affinity to crystalline cellulose (Carrard and Linder, 1999). This disulfide bridge stabilizes the loop carrying one of the essential aromatic amino acids. The deletion may thus alter the orientation of the tryptophan and reduce rigidity of the binding surface. The entropy loss in a binding event, where a rigid planar CBM surface binds to a flat crystalline substrate, is probably small. Therefore, the entropy gain in the dehydration of the surfaces can give an overall beneficial entropy in the binding. Further, crystalline cellulose presents a contiguous surface of binding sites for the CBM. This can give a motional freedom for the bound CBM, which may also contribute to the entropy of the binding. The evidence on the surface diffusion of the CBM2a supports this theory (Jervis et al., 1997).
**Figure 2.** Structures of crystalline cellulose-binding modules indicating how different folds nevertheless maintain the same orientations of the essential aromatic side chains. A) *T. reesei* Cel7A CBM1 (Kraulis *et al*., 1989), B) *E. crysanthesi* Cel5A CBM5 (Brun *et al*., 1997), C) *P. fluorescens* Xynl0A CBM10 (Raghothama *et al*., 2000), D) a superposition of the three structures. Side chains are shown only for the aromatic amino acids on the binding surface of the molecules. In the panels A, B and C β-sheets are coloured green while loops and a short α-helix in the family 10 CBM (panel C) are shown in yellow. In panel D, each structure is coloured using the colour of the aromatic side chains in the panels A, B and C to indicate its identity.
Figure 3. Solvent accessible structures of CBMs with different binding site topologies: A) The structure of the *C. fimi* Cel9a CBM4-2 (Brun *et al*., 2000). The sugar-binding site is located in the groove visible on top of the molecule. B) The structure of *T. maritima* Xyn10A CBM9 (Notenboom *et al*., 2001). This family 9 CBM binds to reducing end of the sugar chain and the binding site topology is a "blind canyon".
2.3 Binding modules recognizing amorphous cellulose and cello-oligosaccharides.

The modules that bind to soluble carbohydrates or amorphous parts of the solid substrate seem to have evolved binding sites in shallow grooves accommodating the interacting amino acid side chains (Figure 3A). Such binding grooves are generally able to accommodate a single sugar chain. The CBMs in the families 2, 4, 9, 13, 17 and 28 have so far shown binding to amorphous cellulose or cello-oligosaccharides. Currently there are several structures solved of CBMs with affinity towards isolated chains (Table IV).

The first CBM discovered with a clearly demonstrated binding specificity towards soluble sugar chains was the *C. fimi* Cel9B CBM4-1. This binding module binds to soluble cellulose derivatives such as HEC, CMC and β-glucan as well as cellooligosaccharides with a DP of 3 or greater. No added affinity was observed beyond DPs above 5, indicating that the binding site can accommodate 5 sugar residues (Johnson *et al.*, 1996b; Kormos *et al.*, 2000; Tomme *et al.*, 1996a). Alanine scanning of the amino acids in the binding groove demonstrates the importance of two tyrosines (Y19, Y84) for the carbohydrate-protein interaction: mutation of these residues decreases the affinity by 50-fold (Kormos *et al.*, 2000). Some of the asparagines and glutamines have also been shown to contribute for the affinity of the CBM4-1. Similar results were obtained with the homologous CBM4-2 from the same enzyme (*C. fimi* Cel9B CBM4-2) (Brun *et al.*, 2000). It is noteworthy, that effectively the same amino acid side chains that contribute in the affinity for the crystalline cellulose are also essential for the binding of isolated sugar chains. As shown earlier, the specificity is achieved by adjusting the topology of the binding site and the orientation of the aromatic side chains (Simpson, *et al.*, 2000). The newly characterized family 17 CBM from the *Clostridium cellulovorans* cellulase 5A has similar binding specificity as the *C. fimi* CBM4s, but exhibits an order of magnitude greater affinity (Boraston *et al.*, 2000). Interestingly, two newly characterized binding modules, in family 4, found in the *Rhodothermus maritus* Xyn10A, bind both to xylan and to amorphous cellulose (Abou Hachem *et al.*, 2000). In a xylan chain, the adjacent sugar residues rotate 120° (as compared to the 180° in cellulose) creating a helical backbone structure and the structural basis of this wide binding specificity has not been investigated in detail. There is another case, where the two family 2 xylan-binding modules from the *C. fimi* Xyn11A bind synergistically, although quite weakly, to amorphous cellulose, even though neither of the isolated domains have affinity to this substrate (Bolam *et al.*, 2001).
An interesting structure was recently published for the Xyn10A family 9 CBM from *Thermotoga maritima* (Figure 3B) (Notenboom *et al*., 2001). This CBM9 shows specificity for a wide range of substrates: BMCC, Avicel, amorphous cellulose (PASA), β-glucan, cellobiose and also weak binding to xylan (Boraston *et al*., 2001). Binding experiments and an enzyme complex structure with cellobiose indicate that the binding site can accommodate two residues (Notenboom *et al*., 2001). When the cellulose was treated with sodium borohydride, which reduces the hemiacetal linkage at C-1 of the glucose ring, the binding was reduced or erased totally. This shows that the CBM9 binds to the reducing end of the sugar chain, explaining its broad specificity (Boraston *et al*., 2001). The solved crystal structure of the module reveals that the sugar-binding site is located in a “blind canyon”, a groove that is blocked at one end (Figure 3B) (Notenboom *et al*., 2001). The complex structure with cellobiose confirms that the reducing end of the saccharide is involved in several hydrogen bonds thus contributing to the affinity (Notenboom *et al*., 2001).

**Table IV:** Published amorphous cellulose / cello-oligosaccharide binding module structures

<table>
<thead>
<tr>
<th>Family</th>
<th>CBM</th>
<th>Crystal/solution structure</th>
<th>Referens</th>
</tr>
</thead>
<tbody>
<tr>
<td>2b</td>
<td><em>C. fimi</em> Xyn11A CBM2b</td>
<td>NMR</td>
<td>(Simpson <em>et al</em>., 1999)</td>
</tr>
<tr>
<td>3</td>
<td><em>T. fusca</em> Cel9 CBM3</td>
<td>Crystal structure</td>
<td>(Sakon, <em>et al</em>., 1997)</td>
</tr>
<tr>
<td>4</td>
<td><em>C. fimi</em> Cel9B CBM4-1</td>
<td>NMR</td>
<td>(Johnson <em>et al</em>., 1996a)</td>
</tr>
<tr>
<td></td>
<td><em>C. fimi</em> Cel9B CBM4-2</td>
<td>NMR</td>
<td>(Brun, <em>et al</em>., 2000)</td>
</tr>
<tr>
<td>9</td>
<td><em>T. maritima</em> Xyn10A CBM9</td>
<td>Crystal Structure</td>
<td>(Notenboom, <em>et al</em>., 2001)</td>
</tr>
<tr>
<td>22</td>
<td><em>C. Thermocellum</em> Xyn10B CBM22*</td>
<td>Crystal Structure</td>
<td>(Charnock, <em>et al</em>., 2000)</td>
</tr>
</tbody>
</table>

* Xylan binding module

The thermodynamics of the protein-carbohydrate binding seems to gradually shift from the entropy-driven binding of the crystalline cellulose binders to the enthalpy-driven interaction of the proteins using pocket- or deep cleft-like binding sites. Calorimetric studies of the *C. fimi* Cel9B CBM4-1 and CBM4-2 indicate enthalpy-driven binding interactions with a favorable standard enthalpy upon binding (Brun *et al*., 2000; Tomme *et al*., 1996a). In such binding interactions the hydrogen bond network has an increased importance for generating affinity compared with CBMs recognizing crystalline cellulose. However, several different
mutation studies have shown that the aromatic amino acids also contribute significantly to the binding interactions when the CBMs bind to soluble substrates. On the other hand, mutations on the hydrogen bonding polar residues of CBMs do not decrease the binding in a very dramatic fashion (Kormos et al., 2000; Xie et al., 2001). This could be explained by the fact that the CBMs binding to cellobiose saccharides have a quite shallow binding groove, which allows a certain degree of motional freedom for the substrate. This might contribute for the enthalpy-entropy compensation with loss of a hydrogen bond as postulated in the study of family 2 xylan-binding module (Xie et al., 2001). Thermodynamic studies of the family 9 CBM, with a lectin like binding site topology, show that the binding event is highly exothermic with favorable enthalpy changes, as indeed seen in several sugar-lectin interactions (Boraston et al., 2001; Dam et al., 2000). The irreversible binding, which is discussed with the crystalline cellulose binding modules, has not been reported with soluble saccharide binding modules (Boraston et al., 2001; Tomme et al., 1996a).

2.4 Role of the cellulose-binding modules

Arguably the best-characterized binding modules are those with specificity to cellulose. Cellulose binding modules have been found in at least 15 of the over 28 currently known CBM families (Table II). Cellulose-binding modules range from small fungal binding domains of approximately 36 amino acids in family 1, to modules consisting of over 200 residues in families 11 and 17. These CBMs have been shown to have spectrum of specificities from strictly crystalline cellulose to amorphous cellulose and hemicellulose as well as monosaccharides, as in lectins (reviewed by Linder & Teeri, 1997). In many cases, the binding specificity of a CBM is not definite, but it is rather a spectrum of affinities towards various substrates.

The obvious function of cellulose-binding modules is to bring and maintain the catalytic module close to the substrate hence increasing the local substrate concentration for the enzyme active site. This is evident in the CBM deletion experiments where the ability of the isolated catalytic module to hydrolyze insoluble substrates is significantly decreased, while the activity towards soluble substrate is not affected (Bolam et al., 1998; Gilkes et al., 1988; Reinikainen et al., 1992; Reinikainen et al., 1995; Tomme et al., 1988). These experiments indicate that the CBMs do not take part in the actual hydrolysis event but merely increase the effective substrate concentration and prolong the enzyme association with the substrate.
surface. Similar effects have been observed with other carbohydrate-active enzymes when the domain function has been studied. For example, in the *C. thermocellum* Xyn10B, the substrate-binding module enhances the activity towards insoluble xylan (Charnock *et al.*, 2000). However, there are also indications that some CBMs can have other functions and properties during cellulose degradation. The isolated *C. fimi* CBM2 seems to promote enzyme activity on solid substrates even when not covalently linked to the catalytic core (Din *et al.*, 1994a). One of the postulated mechanisms in this synergistic action of the CBM, is that it disrupts or disperses the cellulose fibers and may thereby assist the hydrolysis by increasing the effective cellulose surface area. There is also evidence that the family 3 CBM increases the processivity of the endo/exo-acting *T. fusca* Cel9A, possibly by helping to feed the cellulose chain into the active site (Irwin *et al.*, 1998). It has been observed that purified *C. fimi* family 2 CBM inhibits the flocculation of the BMCC in the buffer solution, probably by weakening the intercrystalline interactions (Gilkes *et al.*, 1993). A similar effect has been observed with the intact *T. reesei* Cel7B endoglucanase (Srisodsuk *et al.*, 1993).

Most of the characterization of the CBM binding properties have been done on model substrates such as BMCC, Avicel and *Valonia* cellulose. Relatively little work has been done to study the CBM interactions with polycarbohydrate composites and plant cell walls, which are the natural substrates for the enzymes containing cellulose binding modules. It should be noted that CBMs with affinity to cellulose are also present in enzymes with the catalytic domain active on another substrate than cellulose (Black *et al.*, 1996; Bolam *et al.*, 1998; McKie *et al.*, 2001; McLean *et al.*, 2000). It is likely that the cellulose-binding modules in these cases will facilitate the degradation of complex plant cell wall substrates instead of cellulose *per se* (Black, et al., 1996; Bolam, et al., 2001). It is quite possible that more diverse roles of the CBMs will be found when the enzyme action on the complex composite materials, such as cell walls, will be more extensively studied.

### 3. Family I cellulose-binding domains

Family 1 CBMs form a distinctive group of small peptides with demonstrated activity towards crystalline cellulose and, at least in one case, to chitin (Linder *et al.*, 1996). The CBM1s are almost exclusively of fungal origin. Only one algal cell wall protein has been found with four repeats of putative CBM1s (Liu *et al.*, 1996). Furthermore, all CBMs found in aerobic fungi are exclusively from family 1. Structures solved for the family 1 CBMs have revealed a fold
CARBOHYDRATE-BINDING MODULES

composed of three short β-sheets, which are held together by 2 or 3 disulfide bridges (Kraulis et al., 1989; Mattinen et al., 1998). The cellulose-binding surface of the family 1 CBMs is a planar surface with three aromatic amino acids and few conserved polar residues (Figure 2A and Figure 5) (Kraulis et al., 1989; Linder et al., 1995a; Linder et al., 1995b; Mattinen et al., 1998; Reinikainen et al., 1992). The sequence alignments show that the three aromatic amino acids and four to six cysteins are well conserved within the family 1 CBMs (Figure 4). Some members of the family 1 CBMs, such as Agaricus bisporus CBM1, have a phenylalanine at the “tip” of the wedge shaped molecule (position 31 in T. reesei CBM1Cel7A) instead of a tryptophan or a tyrosine. It remains to be seen, how well these putative CBMs with a phenylalanine bind to carbohydrates and what is their substrate specificity.

<table>
<thead>
<tr>
<th>Cellulase</th>
<th>Family 1 CBM</th>
<th>Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tr Cel7A</td>
<td>PTQSHYGQCGGI...GYSGPTVCASGTTCQLVLNPYYSQCL</td>
<td></td>
</tr>
<tr>
<td>Cel7B</td>
<td>CTQTWHGQCGGI...GYSGCKTCTGTTCQYNSNDYYSQCL</td>
<td></td>
</tr>
<tr>
<td>Cel16B</td>
<td>ACSSWVGQCGGQ...NWGPTCAGSTCVYNSNDYYSQCL</td>
<td></td>
</tr>
<tr>
<td>Cel15A</td>
<td>QTTWGQCGGQ...GWGPTNCAGSTCLINYYAQCL</td>
<td></td>
</tr>
<tr>
<td>Cel145A</td>
<td>QTLYGQCGGQ...GWTGPTTCAPGTCKVQNYWSQCL</td>
<td></td>
</tr>
<tr>
<td>Hi Cel145A</td>
<td>CTAEWQAQCGGN...GWGCTTTCVAGSTCKINDWYHQC</td>
<td></td>
</tr>
<tr>
<td>Cel15A</td>
<td>AQGGAWQCGGQ...GFGSSTCSCVYTCVLNDWYQSCQ</td>
<td></td>
</tr>
<tr>
<td>Cel17A</td>
<td>PKAGRWQCGGQ...GETGPTQCEEPYIYCKTNWDYQSCQ</td>
<td></td>
</tr>
<tr>
<td>Cel17A</td>
<td>GSVQWGQCGGQ...NYSGPTTCSPFCTCKINDFYSQCQ</td>
<td></td>
</tr>
<tr>
<td>Cel145A</td>
<td>SVVPAYYQCGGKSAYPNGLNACATGSKCCVQNEYYSQCV</td>
<td></td>
</tr>
<tr>
<td>Cel110</td>
<td>AQPAPWQCGGQ...GWGATCCGLKEKINDWYQSCQ</td>
<td></td>
</tr>
<tr>
<td>Cel16A</td>
<td>CSNGVWQAQCGGQ...NWGTPCTSGNCKVLNDFYQSCQ</td>
<td></td>
</tr>
<tr>
<td>Np Cel16A</td>
<td>CGGAWAQCQGGE...NEFGDKCCVSGHCVSINQWYSQCL</td>
<td></td>
</tr>
<tr>
<td>Xyn10</td>
<td>NCAAKWQCGGNN...GFNGPTTCCQNGSRQVPNWYQSCQ</td>
<td></td>
</tr>
<tr>
<td>Pc Cel17A</td>
<td>VTVQWGQCGGQ...GYTGSTCTASCYTVLNPYYSQCY</td>
<td></td>
</tr>
<tr>
<td>Bg13</td>
<td>AQSLGVWQCGGQ...GWTGATCCSVGATCLNPYYQSCQ</td>
<td></td>
</tr>
<tr>
<td>Pp Dom1</td>
<td>ACGLVEYQCGGQ...GFDGVTCEGLMCMKMGPGYQSCR</td>
<td></td>
</tr>
<tr>
<td>Dom2</td>
<td>QVKPPQYQCGGQ...NYSGKTMCSPFGBKCVLNEFFQSCD</td>
<td></td>
</tr>
<tr>
<td>Dom3</td>
<td>VCGKGYAAACGGE...FMFGAKCCKFGLVCYESGKQYSQ</td>
<td></td>
</tr>
<tr>
<td>Dom4</td>
<td>GEVGRYAQCGGQ...GYMGSTMBCGGKCMASEGQMYKQ</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.** Sequence alignment of CBMs and putative CBMs in family 1. The abbreviations are Tr: Trichoderma reesei, Hi: Humicola insolens, Fo: Fusarium oxysporum, Np: Neocallimastix patriciarum, Pc: Phanerochaete chrysosporium, Pp: Porphyra purpurea. The conserved aromatic amino acids on the binding surface are indicated in bold.

### 3.1 The fold of the family 1 cellulose-binding modules

Family 1 cellulose-binding modules occur as structurally independent, well-defined domains usually attached to a catalytic module via a flexible linker peptide at either the N- or C-terminal end of the catalytic module. Their molecular structure very much resembles that
found in the inhibitory cystine knot (ICK) family (Norton & Pallaghy, 1998; Pallaghy et al., 1994). Distinctive characteristics in the ICK-motifs are short \( \beta \)-sheets, which are held together by up to four disulfide bridges. There is no significant hydrophobic core in these proteins, the disulfide bridges alone seem to stabilize the structure. The proteins found in this structural family are predominantly toxins interacting with ion channels or as enzyme inhibitors (Norton & Pallaghy, 1998). Interestingly, a few lectins have also been found to adopt the ICK motif structure. The venom of the Chinese bird spider *Selenocosmia huwena* is a lectin with three disulfide bridges and a three-stranded antiparallel \( \beta \)-sheet (Lu et al., 1999). There are also two chitin binding proteins, which have similar structural features: the hevein, which is a chitin-binding protein from the rubber tree *Hevea brasiliensis* (Asensio et al., 1995), and the chitin-binding protein from *Amaranthus caudatus* (Martins et al., 1996). However, these proteins do not require crystalline chitin for binding, but bind to oligochitin with five or more residues with micromolar affinity (Asensio et al., 2000a; Asensio et al., 2000b). The similarity of these proteins with the family 1 CBMs is limited to their structure being stabilized by disulfide bridges and forming small compact and very stable domains. The spacing of the cysteins and the pattern of the disulfide bridges seems to be conserved in the ICK-family and similar conservation can be seen in the family 1 CBM sequence alignments (Flinn et al., 1999; Norton and Pallaghy, 1998). However, many of the CBMs contain only two disulfide bridges, instead of the four found in ICK-family.

3.2 Characterization and Function of family 1 CBMs
The best-characterized members of the family 1 CBMs are the *Trichoderma reesei* Cel7A CBM1 and Cel6A CBM1 (for the new and the old nomenclature see page 9). The importance of the aromatic side chains on the planar surface is well demonstrated in several studies (Table V, in the results part) (Linder et al., 1995a; Linder et al., 1995b; Reinikainen et al., 1992). Studies of these CBMs show that the binding characteristics rely on a few amino acids. Single point mutations on the aromatic amino acids of the CBM1Cel7A can elevate the affinity close to that of the significantly better binding of the CBM1Cel7B, or to erase it almost completely (Linder et al., 1995a; Linder et al., 1995b). Mutation of the aromatic amino acid (Y31A) to an alanine in *T. reesei* CBM Cel7A is the least disturbing for the overall structure of the molecule, but does nevertheless abolish its binding affinity to cellulose (Linder et al., 1995b; Mattinen et al., 1997). Histidines occur on the binding face of the family 3 bacterial CBMs but have not been observed in the family 1 CBMs (Sakon et al., 1997; Tormo et al.,...
CARBOHYDRATE-BINDING MODULES

1996). However, a histidine residue has been successfully introduced on the binding surface of \textit{T. reesei} CBM1\textsubscript{Cel7A} by mutagenesis, resulting in a pH-dependent binding module (Linder et al., 1999; Reinikainen et al., 1992). Furthermore, when the binding surface of the Cel7A CBM1 was randomized and displayed on a bacteriophage, one of the binders obtained, after biopanning against Avicel, contained a histidine residue as a part of the binding surface (Smith et al., 1998).

Irreversible binding behavior reported on crystalline cellulose binding of some bacterial CBMs, has also been observed with the \textit{T. reesei} CBM1\textsubscript{Cel6A} (Carrard & Linder, 1999). The mechanism of such irreversible binding is not yet clear, but this property can be altered to fully reversible binding by single point mutations (Carrard & Linder, 1999; Linder and Teeri, 1996). Additionally, it has been shown that the two \textit{T. reesei} CBMs (CBM1\textsubscript{Cel7A} and CBM1\textsubscript{Cel6A}) demonstrate synergistic binding when the CBMs are covalently coupled to each other with a linker peptide (Linder et al., 1996). The catalytic module and CBM1 have also been shown to have synergistic binding onto cellulose, especially at concentrations resulting in a lower surface coverage (Palonen et al., 1999; Srisodsuk et al., 1997).

A CBM1 is clearly critical for crystalline cellulose degradation by the fungal enzymes: deletion of the CBM1 does not have an effect on the enzyme activity on soluble substrate but reduces the enzyme activity on crystalline substrate considerably (Reinikainen et al., 1992; Reinikainen et al., 1995; Tomme et al., 1988). It is clear that these CBMs enhance the local concentration of the catalytic module in the proximity of the substrate, similar to some bacterial CBMs (Gilkes et al., 1988). It has also been proposed, that the CBM1s contribute to the disruption of the cell wall texture or even on cellulose crystals. So far, the only evidence supporting such theory is a recent report showing that the \textit{T. reesei} endoglucanase CBM1\textsubscript{Cel7B} weakens the hydrogen bonding of the cellulose crystals, which leads to detectable changes on an NMR-spectra (Xiao et al., 2001). Earlier reports have shown that the same intact enzyme prevents the flocculation of BMCC in the solution, which may also be mainly a result of the CBM binding (Srisodsuk et al., 1998). There is no clear evidence that CBM1s would promote enhanced hydrolysis on cellulose when not covalently linked to the catalytic module. A relatively long linker peptide coupling the binding- and catalytic module seems to be needed for effective enzyme action on BMCC (Srisodsuk et al., 1993). An interesting observation concerning the \textit{T. reesei} Cel7A and Cel6A enzymes, is that the amount of bound enzyme increases as a function of the hydrolysis on BMCC (Palonen et al., 1999).
4. Aims of this study

The general aim of this work was to study family 1 carbohydrate-binding modules in order to understand their binding mechanism, function as a part of a hydrolytic enzyme and their engineering for different applications. The specific aims were:

1) To investigate the adsorption of the CBM1\textsubscript{Cel7A} and CBM1\textsubscript{Cel6A} on cellulose crystals (publication I).
2) To study the function of family 1 CBMs as a part of an intact cellulase (publication II).
3) To investigate the use of a CBM1 as an immobilization tag for another enzyme (publication III) and for immobilization of microbial cells (publication IV).
4) To use phage display technology to engineer variants of the CBM1 with novel binding specificities (publications V and VI).

RESULTS AND DISCUSSION

5. Family 1 cellulose-binding module function and interaction with cellulose (I, II)

The solution structure of \textit{T. reesei} CBM1\textsubscript{Cel7A} reveals a compact wedge shaped molecule, which is stabilized by two disulfide bridges (Figure 5). Mutational studies have demonstrated that the aromatic amino acids 5, 31 and 32 (numbering according to Kraulis \textit{et al.}, 1989) on the planar surface are crucial for binding to crystalline cellulose (Linder \textit{et al.}, 1995b; Reinikainen \textit{et al.}, 1992). In addition, it has been shown that the molecular rigidity can play an essential role in the CBM1 binding, since removal of one of the disulfide bridges in the Cel6A CBM1 reduces its binding (Carrard & Linder, 1999). In publication I, complementary mutation studies have been carried out for the Cel7A CBM1. Mutation work performed in this and other studies of the \textit{T. reesei} CBM1\textsubscript{Cel7A} and CBM1\textsubscript{Cel6A} are summarized in Table V.

In order to study the CBM1 binding to crystalline cellulose, several CBM1s were fused to the double Z-domain of the staphylococcal protein A (SpA) (Nilsson \textit{et al.}, 1987). The ZZ-tag
CARBOHYDRATE-BINDING MODULES

was used for several reasons. Firstly, the antibody binding ability of the ZZ-tag was utilized in the affinity purification of the ZZ-CBM fusion proteins and in detection of the proteins during the cultivations using the western blot technique. Furthermore, the tag allowed direct immunogold labeling of the CBM fusions in the TEM studies (publication I). The wild type *T. reesei* CBM1\textsubscript{Cel7A}, CBM1\textsubscript{Cel6A} and CBM1\textsubscript{Cel7B}, *Neocallimastix patriciarum* CBM1\textsubscript{Cel6A} and two mutated CBM1\textsubscript{Cel7A} variants (CBM1\textsubscript{Cel7A}Y5W and CBM1\textsubscript{Cel7A}Y5W\textsubscript{Y31W}) were cloned and produced as fusion to the ZZ-domain (Figure 1, publication I). Also the family 3 CBM from the *C. thermocellum* scaffolding protein CipA (CBM3\textsubscript{CipA}) was produced and tested in some of the experiments (publication I). Binding of the CBM1-ZZ fusion proteins was tested on BMCC and *Valonia* cellulose (equilibrium binding isoterms, Figure 2, publication I) and the reversibility of the binding was investigated in non-equilibrium dilution experiments (Table VI). To determine the binding site of the CBM1 on the cellulose crystal surfaces the fusion proteins were immunogold labeled and studied by transmission electron microscopy (TEM).

5.1 The importance of being tryptophan: mutation studies of family 1 CBMs

In the family 1 CBMs the three aromatic amino acids, that interact directly with cellulose, are on a planar surface of the wedge shaped molecule (Figure 5) (Linder *et al*., 1995b; Reinikainen *et al*., 1992). The Cel7A CBM1 interaction with cellohexose, shown in NMR studies, seems to rely on the glucose residues 1, 3 and 5 interacting with the three aromatic side chains on the CBM1 (Mattinen *et al*., 1997). The combined surface area of both the aromatic side chains and the pyranose rings are postulated to contribute directly to the binding energy in the hydrophobic interaction (Doig & Williams, 1992; Williams *et al*., 1993; Xie *et al*., 2001). Apart from the aromatic amino acids, a glutamine (Q34) in CBM\textsubscript{Cel7A} contributes to the affinity (Linder *et al*., 1995b).

A single point mutation (Y5W) in the CBM1\textsubscript{Cel7A} increases its affinity (publication I) (Linder *et al*., 1995b). Mutating two of the tyrosines to tryptophans (CBM1\textsubscript{Cel7A}Y5W\textsubscript{Y31W}) does not further enhance the affinity. All the CBM1s with at least one tryptophan (*T. reesei* Cel6A, Cel7B and *N. patriciarum* Cel6A as well as the mutated Cel7A CBM1s) on their binding face feature similar binding (publication I). Better binding of the CBM1 containing one tryptophan might be due to increased surface area of the hydrophobic face stacking onto the pyranose
rings on the cellulose surface. The fact that the second tryptophan on the double mutant does not further enhance the binding is somewhat puzzling. Perhaps the exposed hydrophobic surfaces are unable to form complementary interactions because the glucose residue spacing on the cellulose chain does not perfectly match that of the aromatic residues on the CBM. The spacing seems to be an important asset, since similar spacing can be seen in all the crystalline cellulose-binding modules that have their 3D structure available (Figure 2) (Tormo et al., 1996). Structural changes of the double mutant binding surface can not be ruled out in the absence of the experimental structure. However, the binding of the N. patriciarum Cel6A, containing two tryptophans, was shown to be similar to the other CBM1 with one or more tryptophans.

While the T. reesei Cel7A CBM1 is fully reversible, the CBM1 Cel6A has shown irreversible binding on BMCC (Carrard & Linder, 1999). All of the CBM1s tested, fused to a double Z domain, show fully reversible binding and returned to equilibrium in less than 30 minutes (Table VI). In the case of Cel6A CBM1, the irreversible binding earlier observed could be canceled by a single point mutation (W7Y), by deleting one of the disulfide bridges or by increasing the temperature (Carrard & Linder, 1999). Deletion of the disulfide bridge probably makes the binding surface less rigid and would thus have a negative effect on the entropy of the binding interaction. Changing a tryptophan to a tyrosin diminishes the surface area needed to create the Van der Waals interaction and has been shown to reduce the affinity of Cel6A CBM1 (Carrard & Linder, 1999). It is interesting that, while the isolated Cel6A CBM binds irreversibly, the intact Cel6A is partially reversible (Carrard & Linder, 1999; Palonen et al., 1999). This difference in the binding characteristics is puzzling since the two modules of the intact enzyme bind synergistically in low concentrations (Palonen et al., 1999). It is possible that the added mass of the catalytic domain nevertheless weakens the binding from seemingly irreversible to partially reversible. In the case of the CBM-ZZ fusion proteins, the double Z domain does not bind to cellulose at all. Again, it might be that the added mass of the ZZ-fusion partner leads to the fully reversible binding of the protein (Table VI). The fact that the CBM1 reversibility is temperature dependent and seems to be sensitive to the mass changes raise questions of its biological relevance. In nature T. reesei grows in a tropical climate, at temperatures where the CBMs show reversible binding. Furthermore, the mass of a catalytic module changes the characteristics of the binding towards reversibility.
**Table V**: CBM1 mutations and their relative effects on affinity

<table>
<thead>
<tr>
<th>CBM1&lt;sub&gt;Cel7A&lt;/sub&gt;</th>
<th>( Y5A )</th>
<th>( Q34A )</th>
<th>( Y32A )</th>
<th>( N29A )</th>
<th>( Y31A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y5W )</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Y5H )</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Y5H )</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( H4V )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P16R )</td>
<td>-</td>
<td>( T17K )</td>
<td></td>
<td>( V27Y )</td>
<td></td>
</tr>
<tr>
<td>( V18T )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( A20T )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( CBM1&lt;sub&gt;Cel6A&lt;/sub&gt; )</td>
<td>( W7Y )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Delta S3 )</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6. The structure of the Cel7A CBM1 shown in two different orientations (Kraulis et al., 1989). In the combinatorial CBM1 library eleven positions were randomized. The amino acids at the randomized positions are coloured red on the wild type structure, illustrating the binding surface topology.
Consequently, the irreversible binding phenomenon is not likely to be relevant with intact enzymes in nature. This could explain why the processively acting T. reesei Cel6A has a seemingly irreversible CBM.

Table V: Non-equilibrium dilution experiment of the ZZ-CBM fusion proteins

<table>
<thead>
<tr>
<th>Fusion protein</th>
<th>Recovery to isotherm in 30 minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. reesei Cel7A ZZ-CBM1</td>
<td>99,0±1,1</td>
</tr>
<tr>
<td>T. reesei Cel6A CBM1-ZZ</td>
<td>99,4±0,2</td>
</tr>
<tr>
<td>T. reesei Cel7B ZZ-CBM1</td>
<td>99,8±0,1</td>
</tr>
<tr>
<td>N. patriciarum Cel6A CBM1-ZZ</td>
<td>99,5±0,3</td>
</tr>
<tr>
<td>T. reesei Cel7A ZZ-CBM1Y5W</td>
<td>99,8±0,9</td>
</tr>
<tr>
<td>T. reesei Cel7A ZZ-CBM1Y5W:Y31W</td>
<td>99,0±0,5</td>
</tr>
</tbody>
</table>

5.2 Binding specificity of family 1 CBMs: electron microscopic experiments

There is an ongoing discussion concerning possible specific binding sites of different CBMs on the crystalline cellulose. It is difficult to detect any obvious differences on the binding faces of the family 1 CBMs when comparing the sequences (alignment, Figure 4) or experimental or modeled structures of the family 1 CBMs (Hoffren et al., 1995; Kraulis et al., 1989; Mattinen et al., 1998). However, there is one study suggesting that CBMs may have at least partly separate binding sites on crystalline cellulose (Carrard et al., 2000). In that study, several CBMs were linked one by one to the CelD endoglucanase catalytic module from C. thermocellum. It was shown that a complex with a new CBM promotes enhanced hydrolysis after 50 h, whereas the addition of the original CBM-complex did not have this boost effect (Carrard et al., 2000). However, it was found that the added uncoupled catalytic module, used as a control, also promoted the boost effect in some cases, but not all. Furthermore, with the error of 5-12% the differences were quite small and it remains to be seen what is the actual specificity difference of these CBMs.

To address the question of the specificity of CBMs, binding of different CBM-ZZ fusion proteins on Valonia cellulose was studied by transmission electron microscopic experiments. The CBMs studied were family 1 CBMs from T. reesei Cel7A and Cel6A and the family 3 CBM from C. thermocellum. In the TEM visualisation, distribution of the gold particles on isolated crystals indicated binding of the CBMs in one or two rows along the crystal edges (publication I). For a more detailed picture of the CBM adsorption to Valonia cellulose crystals, a set of TEM experiments was carried out with isolated, centred crystals carrying the
bound immuno-gold labeled CBMs. Two-dimensional coordinates of the gold particles were recorded with three different rotations of individual crystals (0°, -45° and 45°) and this data was used to calculate the three dimensional location of the labeled proteins around the cellulose crystal (Figure 4, publication I). The residual error for the coordinate transformation, as defined in Eq. 1 (publication I) was always less than 0.02, indicating that the experimentally measured locations of the labels fit acceptably in the mathematical equation, and that the tilting and centering of the fiber were accurate. The calculated coordinates of the labels indicated that the CBMs were bound in two distinct groups. The fiber orientation of the same crystals were determined by diffraction experiments as described in (Imai et al., 1998; Koyama et al., 1997). The diffraction pattern is distinctive to the direction of the glucose chains in the crystal, hence it is possible to determine the orientation of the isolated crystal. The mapping of the labels combined with the diffraction pattern analysis, indicate that the CBMs adsorbed onto the surface with fully exposed pyranose rings, i.e. the (200) crystal face of the cellulose surface.

The intriguing question is the amount of (200) face exposed on the crystal surface. In a perfectly shaped crystal this face is only one glucose chain wide, which would lead to a very limited surface area for the binding of the CBMs. The electron microscopic experiments of Halocynthia papillosa and Valonia cellulose reveal that cellulose crystals are not perfectly shaped. Instead, the (200) face is often blunt exposing a larger area of this surface to the solvent (Helbert et al., 1998; Sugiyama et al., 1985). This observation could explain why earlier surface coverage calculations of the CBMs on crystalline cellulose disagree with the experimental binding capacity of the CBMs (Mattinen et al., 1997; Reinikainen et al., 1995; Tormo et al., 1996). In case of imperfect crystals, the exposed (200) surface area increases dramatically. The distribution of the labels into two distinct groups this experiment is obvious. Knowing that the distance of the gold particle to the actual binding site of the CBM can be anything from 0.5 nm up to 5 nm judging from the sizes of the CBM-ZZ fusion protein and the gold-conjugated antibody, the grouping of labeling was significant. The estimated size of the isolated Valonia crystal is 20 x 20 nm and the clear grouping of the labels implies that the specific binding site is to a limited surface on the crystal. We did not detect any significant binding on the (110) or (1-10) surfaces as proposed by (McLean et al., 2000).

In the experiments where T. reesei enzymes Cel6A and Cel7A binding was monitored as a function of the hydrolysis, the relative amount of bound enzyme increased despite the fact
that the total cellulose surface area diminishes (Palonen *et al*., 1999). This agrees with our TEM visualization data since the cellobiohydrolases peeling off chains from the (200) face would simultaneously reveal more of the (200) surface for the CBM binding. Other experiments have suggested that bacterial CBM2s have primary and secondary binding sites on the BMCC (Creagh *et al*., 1996) (personal communication with Larry Walker). It may be that some CBMs first bind to (110) face of the crystal as suggested in (McLean *et al*., 2000). Binding on the (110) surface should conceivably have lower affinity than the binding on the (200) face. In the above mentioned recent publication by (Carrard *et al*., 2000), the CBM3*$_{\text{CipA}}$* fused to a catalytic module by cohesin-dockerin fusion, is postulated to cover a wide range of binding sites on the cellulose surface. In the case of CBM3, an addition of the same or another type of CBM, fused to the same catalytic module, does not promote an additional boost effect. We also tested the same *C. thermocellum* CBM3*$_{\text{CipA}}$* fused to the ZZ-domains in the TEM experiments (publication I). The labeling pattern obtained with CBM3*$_{\text{CipA}}$* on the *Valonia* cellulose crystals was very similar to the *T. reesei* CBM1$_{\text{Cel7A}}$ and CBM1$_{\text{Cel6A}}$. Also the CBM3*$_{\text{CipA}}$* mapping with rotation experiment gave similar results, indicating binding on the (200) face of the crystal. This suggests, that the crystalline cellulose-binding modules use a universal strategy for creating affinity to crystalline cellulose.

Linking of the *T. reesei* Cel7A and Cel6A CBM1s together with a linker peptide results in significant synergistic binding, indicating that the binding sites of these two CBMs on the BMCC have to be able to bind very close to one another on the crystal surface (Linder *et al*., 1996). We have also constructed a different double CBM with the two CBM1s separated by a double Z-domain (unpublished data). This double CBM was found to exhibit synergistic binding to BMCC similar to the first double CBM. A domain exchange experiment, where the CBM1 of a processive cellobiohydrolase (*T. reesei* Cel7A) was replaced with the higher affinity CBM1 from endo-acting enzyme (*T. reesei* CEL7B), indicated no great change or even slightly enhanced hydrolytic capacity of the hybrid enzyme (Figure 3a, publication II). This indicates that the different endoglucanase CBM1 did not direct the cellobiohydrolase catalytic module to unfavorable locations on cellulose (publication II). Furthermore, the presence of the Cel7B CBM did not hinder the processivity of the exo-acting enzyme with its higher affinity. Also these experiments suggest that the family 1 CBMs bind to similar sites on a cellulose crystal.
5.3 CBM function as a part of an intact enzyme

The function of the CBMs in the overall process of cellulose hydrolysis is an interesting question. A CBM obviously enhances the effective local concentration of the catalytic module, hence allowing it to react with the insoluble substrate. In the case of crystalline cellulose-degrading enzymes, the task of breaking the crystal is likely to be the rate-limiting step instead of the bond cleavage event. The CBM might act by maintaining the catalytic module in close proximity of the crystal allowing the substrate to enter the active site (Teeri et al., 1998). The domain switch experiment in publication II raises the question why T. reesei Cel7A has CBM that binds weaker when the chimeric enzyme with the higher affinity CBM1 of Cel7B can hydrolyse BMCC somewhat more effectively (Figure 3a, publication II). Why has this domain switch not occurred in nature? The natural substrate of these enzymes is the plant cell wall and the fungus produces a series of enzymes to degrade the wall polysaccharides. The probable answer lies in the interaction of the synergistically acting enzyme mixture on the complex substrates like plant cell walls.

The action of the Cel7A and Cel6A has been shown to be processive, with Cel6A acting on the non-reducing and with Cel7A reducing ends of that crystal (Boisset et al., 2000; Brooks et al., 1992; Imai et al., 1998). However, we could not detect any distinct preference of the studied cellobiohydrolase CBM1s to the cellulose chain ends in the microscopic studies (publication I). Instead, as the TEM experiments indicated that both CBM1s bind to the exposed (200) face along the crystal, which is likely to be exposed on the disrupted parts of plant cell wall. In that way, the proposed binding specificity could in fact direct the processive exoglucanases towards accessible chain ends for productive binding. Interestingly, the recently discovered family 9 T. maritima CBM9Xyn10A has been shown to bind selectively to the reducing end of the cellulose (Boraston et al., 2001; Notenboom et al., 2001). This binding specificity could indeed bring a catalytic module in close proximity of the chain ends. However, in the case of the processive cellobiohydrolases, with the active site in a tunnel (Divne et al., 1993; Divne et al., 1994; Divne et al., 1998), it is probable that this kind of binding would block the productive binding or stop the processive action of the enzyme.
6. Cellulose-binding modules in affinity applications (III, IV)

Today, there are countless studies in which a specific binding ability of a protein has been used for various biotech applications. The traditional uses of affinity proteins include ligands for affinity purification, different detection methods employing antibodies and other binding proteins, enzymes and cell immobilization. The cellulose binding modules present an attractive immobilization tag for biotechnical applications due to their natural ability to spontaneously bind to cellulose across a wide range of conditions. Cellulose provides an inexpensive, biodegradable and medically safe immobilization matrix, and proteins fused to CBMs can be immobilized and purified in a one step procedure. There is a vast number of studies where a CBM has been used as an affinity tag and an immobilization anchor (Bayer et al., 1994; Ong, et al., 1991) (Linder et al., 1998; Rotticci-Mulder et al., 2001; Tomme et al., 1998; Tomme et al., 1996b). In publication III, we have carried out a study to optimize a linker peptide for heterologous protein expression of the Candida antarctica lipase B (CALB) in yeast. We have also studied the direct immobilization of recombinant Staphylococcus carnosus cells onto cellulose fibers by surface expression of the CBM1Cel6A (publication IV).

6.1 Role of the linker peptide in native and engineered enzymes

Carbohydrate active enzymes are typically modular proteins with a catalytic module attached to various other modules (Henrissat, 1999; Henrissat & Davies, 1997). Especially in enzymes active on insoluble substrates, these modules are for carbohydrate binding. These CBMs are frequently linked to the rest of the enzyme through a long and probably flexible linker peptide (Beguin & Aubert, 1994; Gilkes et al., 1991b; Teeri et al., 1987). The role of the linker peptide in natural enzymes has been studied in several linker deletion studies (Srisodsuk, et al., 1993) (Black et al., 1997; Ferreira et al., 1990). In the T. reesei Cel7A, both the absorption and the activity were reduced by deleting major parts of the linker (Srisodsuk et al., 1993). It seems in this case, that for effective hydrolysis of the insoluble substrate, the catalytic module has to be anchored to the cellulose by a CBM with a relatively long linker to allow productive binding and hydrolysis. Similar long linker sequences are found in cellulose-binding modules with catalytic modules active on hemicellulose. The length of the linker is shown to be important when degrading the complex cell wall or composite structures, but the importance of the linker length is not always apparent when the enzyme action is studied.
using model substrates (Black et al., 1997; Ferreira et al., 1990). There are also some reports that truncated forms of the enzymes occur in the culture supernatant under certain culture conditions. Under these conditions, the catalytic core of the enzyme is probably released by linker-specific proteases (Knowles et al., 1988; Miller et al., 1988). It might be that in some cases it is beneficial to release the catalytic module from the binding module to allow additional mobility. It has also been speculated that the linker sequences are beneficial for a rapid evolution of these modular enzymes (Ferreira et al., 1990).

Protein engineering of enzymes for biotechnical applications often requires linker peptides as spacers to facilitate protein-protein fusion. The most important feature of such linkers is proteolytic stability. Another requirement of these linkers is solubility, in order to sustain the extended conformation in solution and flexibility, in order to allow independent function of the fused protein domains. Linker peptides found in nature frequently consist of small polar residues such as glycine, presumably to give flexibility and enhance the solubility (Argos, 1990). In fungal cellulases O-glucosylated linkers are common, probably to improve the solubility and proteolytic stability (Srisodsuk et al., 1993). In an earlier study, a lipase linked to the CBM1_Cel6A from Neocallimastix patriciarum showed good binding on cellulose and the CBM-lipase fusion protein was shown to maintain full lipase activity (Rotticci-Mulder et al., 2001). However, the linker peptide used for the fusion suffered significant proteolytic degradation during cultivations of the Pichia pastoris host. As described in publication III, we tested a series of engineered linkers to find a stable, well-expressed linker peptide for protein production in P. pastoris (publication III, Table I). The native N. patriciarum linker (linker construct 2) was susceptible to proteolysis like the earlier linker derived from the T. reesei Cel6A linker (construct 1). We found that the proline/threonine rich linker, which mimicked the C. fimi cellulase linker, was stable in P. pastoris protein production. This linker was somewhat shorter (13 amino acids) than the linkers 1, 2 and 3, which were degraded during the cultivation. The 13 amino acid long linker is, in all probability, enough to give sufficient spatial separation from the cellulose surface and allows optimal orientation of the catalytic site for interaction with the substrate.

6.2 Bacterial cell immobilization

Immobilization of microorganisms and recombinant surface expression of selected affinity proteins have considerable potential in biotechnology, applied microbiology and
immunology. In addition to immobilization of various enzymes and antibodies, CBMs have been employed also for whole cell immobilization (Francisco et al., 1993; Wierzba et al., 1995). In recent years there have also been several reports of bacterial surface display aiming for various applications (for reviews see Georgiou et al., 1997; Ståhl et al., 2000; Benhar, 2001; Hansson et al., 2001). A widely studied area of bacterial surface display is live vaccine-delivery systems, but biotech applications such as whole-cell bioabsorbents for environmental use (Samuelson et al., 2000; Sousa et al., 1996), diagnostic tools development (Fuchs et al., 1991; Gunneriusson et al., 1999; Gunneriusson et al., 1996) and display of combinatorial libraries on bacteria are also growing areas (Daugherty et al., 1999; Lu et al., 1995). Here we have used the T. reesei CBM1Cel6A for immobilizing the gram-positive Staphylococcus carnosus on cellulose fibers (publication IV). This food-grade bacteria is traditionally used for food biotech applications like starter cultures in meat and fish fermentations as well as in numerous vaccine studies. Consequently, it is regarded safe in food products and environmental applications (Hammes et al., 1995). Immobilization of recombinant S. carnosus cells on to inert and medically safe cellulose matrixes could be used for example in the development biofilters, fermentation immobilization or possibly in development of fast and easy diagnostic tools using a “dip-stick” format.

The CBM1Cel6A from T. reesei has been well characterized. It has a typical family 1 CBM-structure with three disulfide bridges stabilizing the domain and a flat surface featuring three exposed aromatic amino acids (Carrard & Linder, 1999; Hoffren et al., 1995). The Cel6a CBM1 provides a robust and stable immobilization tag with specific affinity towards cellulose. The CBM1Cel6A has higher affinity to cellulose than CBM1Cel7A and the isolated CBM1Cel6A does not dilute from the cellulose in non-equilibrium experiments at low temperatures (Carrard & Linder, 1999). These features are beneficial for proteins utilized as immobilization tags.

The Cel6A CBM1 was cloned into the surface expression vector (Samuelson et al., 1995) and transformed to S. carnosus cells. Surface accessibility of the CBM, and the albumin binding protein (ABP) fused to CBM on the surface receptor chimera, were detected in whole-cell ELISA assays, as described in the materials and methods (publication IV). The chimeric surface receptor was also purified from the cell wall of the S. carnosus and shown to exhibit the expected molecular weight on western blot analyses. Functionality of the displayed CBM1Cel6A was detected by cellulose binding experiments on cotton fibers as well as by light
microscopic experiments (publication IV). The recombinant *S. carnosus* cells bound readily to the cotton fibers, and in higher bacterial densities, the fibers were covered with a monolayer of bacteria. The bacteria lacking surface expressed CBM were not associated with fibers. The good immobilization of the bacteria is probably due to a avidity effect of the CBM1Cel6A, since these chimerical receptors are shown to have high surface density on *S. carnosus* (Robert *et al.*, 1996). Synergistic, high affinity binding of a double CBM has also been demonstrated in a study where two family 1 CBMs were coupled with a linker peptide, (Linder *et al.*, 1996). In an earlier study recombinant *S. carnosus* cells were immobilized successfully on fibronectin by surface expression of different fibronectin-binding domains (Liljeqvist *et al.*, 1999). These two studies propose that this heterologous surface display method provides a good general way to create whole-cell immobilization for various applications.

7. **Protein engineering of family 1 carbohydrate-binding modules (V, VI)**

Traditionally, the immune system has been employed for the generation of high-affinity antibodies to diverse target molecules for vast application areas including therapeutic, diagnostic and biotechnological applications. More than $10^8$ antibodies with different binding specificities circulate in the human body (Skerra, 2000). This diversity has the potential to recognize virtually any foreign structure that the body encounters. However, the stability, heterologous production, size and complexity of the antibodies can present some difficulties and limitations for their use in biotechnology applications. This has led to efforts to recruit alternative protein scaffolds for creating diversity for the selection of novel binders (for reviews see Benhar, 2001; Nygren & Uhlen, 1997). In such work, proteins with the desired binding properties are selected from pools (libraries) of related but different candidates using a suitable selection system. Libraries are typically constructed using combinatorial mutagenesis, at the DNA level, of the molecular surface of the chosen parental protein scaffold.
**Phage display**

A) Principle of the phagemid system

1. **Transformation of** *E. coli* **with phagemid DNA**
2. **Infection with helper phage**

**Phagemid DNA**

- [P] S X gene III
- Protein X (phagemid encoded)
- pVIII (helper phage encoded)
- 50 aa/ ca 2700 copies

**single stranded DNA**
- 6.4 kb

**E. coli** cells are transformed with the phagemid containing the foreign gene, in this case a CBM1 containing the randomized residues (X), fused to pIII-gene. Thereafter, the growing cells are coinfected with a helper phage to produce new phage particles, which are now decorated with the library variants, and contain the respective gene in the phagemid. The helper phages are required for phage assembly. Schematic drawing of phage particle is shown on the right.

B) A schematic presentation of the selection procedure (biopanning).

- **Production of phage particles**
- **Selection**
- **Wash**
- **Identify binder from insert DNA sequence**
- **Reinfect E. coli**

- **Target**
- **Elute binding phages**

**Cultivation of** *E. coli* **cells** containing phagemid DNA

A standard selection for novel binders involves multiple rounds of selection (typically 4-5 biopanning rounds) combined with amplification. The variants with affinity towards the desired target molecule are selected and enriched. After several rounds, the selected variants are identified by DNA sequencing. Thereafter, the variants are produced for further characterization and application.
7.1 Phage display of CBM1 for novel binding properties

Combinatorial protein libraries employing the phage display technology are probably the most widely used system to generate binders with novel binding specificity (The M13 phage display principle is shown on the page 38). The display of heterologous proteins fused to filamentous phage coat protein was first described in the 1980s (Smith, 1985). This technique has since been combined with combinatorial protein chemistry to create a powerful means of selecting peptides and proteins with novel properties (Geysen et al., 1984). Phage display enables the coupling of the genotype and phenotype allowing the selection of the binding protein together with the respective gene encoding the protein. Numerous new techniques have been developed in recent years for peptide and protein display (for recent reviews see Benhar, 2001; Forrer et al., 1999; Nygren & Uhlen, 1997; Skerra, 2000).

7.1.1 Selection of a protein scaffold

The selection of a suitable scaffold for combinatorial mutagenesis is not trivial and there are several criteria that set limitations. The fold should be stable enough to tolerate an extensive randomization of the surface, typically involving 10-20 positions. To overcome the problems associated with the antibody binders, the scaffold for creating novel binders should be based on one polypeptide chain and be easily produced in a microbial hosts. The shape and size of the randomized surface is also important. In this sense, no single protein fold is ideal for all applications. Use of a folded protein as a starting point for the combinatorial approaches can provide certain advantages over the use of linear peptides. The smaller loss of motional freedom for a structured protein compared to that of a flexible peptide during target binding can give a smaller entropy loss (Ladner, 1995). Also, the display of hydrophobic side chains required for the hydrophobic interactions with the target might be easier to achieve in a pre-fixed protein surface than on a peptide where the side chains may become buried upon display.

We have investigated the use of the non-toxic, ICK-like fold of the CBM1\textsubscript{Cel7A} from \textit{Trichoderma reesei} as a protein scaffold in a combinatorial library to create novel binding proteins (publication V). The ICK fold is characterized by the virtual absence of a hydrophobic core and is instead stabilized by disulfide bridges (Figure 7). The family 1 CBMs are very stable and tolerate organic solvents as well as boiling. Sequence alignments of the family 1 CBMs show that the amino acids forming the planar cellulose-binding surface...
and the cysteins forming the disulfide bridges are well conserved, while the rest of the sequence is generally highly variable (Figure 4). The ICK-fold seems to be robust and not very sensitive for sequence variation and gap size between the half cysteins (Norton & Pallaghy, 1998; Pallaghy et al., 1994). There is also one study where the cellulose-binding surface of this CBM has been randomized successfully, indicating that the fold is tolerant to changes if only the positions of the disulphide bridges are maintained (Smith, et al., 1998). In nature this fold is found in proteins with diverse biological functions. Many of the ICK-structural motif proteins are inhibitors and receptor binders interacting with the target proteins active site or binding pocket (Norton & Pallaghy, 1998).

When selecting a scaffold for the library construction we considered the properties of the CBM1 fold and especially the binding surface topology. The rationale was to randomize a surface area around the tip of the wedge-shaped molecule in order to obtain a topology of the novel binders, which could interact with the binding cavities or active sites of the target proteins.

![Figure 7](image_url). A schematic view of the general pattern of the postulated disulfide bridge formation of the family 1 CBMs.

### 7.1.2 Construction and characterization of the CBM1 library

A phage display library of CBM1Cel7A was constructed by randomization of eleven positions located around the “tip” (Figure 6) (publication V). The estimated library size was 42-46 million variants (publication V). We exploited the phagemid system allowing monovalent display of the library variant thus minimizing avidity effects (Nord et al., 1995). 50 different variants of yet unselected clones were analyzed on the DNA level by sequencing (results indicating over 92% correctly constructed genes) of which five randomly picked variants were produced and purified from *E. coli* periplasmic fractions. None of the produced variants showed interdomain disulfide bridge formation. Subsequently, 32 different variants after the different pannings were produced and none of the proteins formed dimers, indicating that the formation of the correct intramolecular disulfide bridges.
7.1.3 Selection of binders from the randomized CBM1 library

To select novel binders, we first used porcine pancreatic α-amylase (PPA) as a target protein, since the structure was known and the active site is located in a groove-like binding pocket. The novel binding molecules obtained demonstrated specific binding towards the PPA. No cross reactivity was detected upon binding to another amylase (barley α-amylase) with a similar structure (publication V). Two of the PPA-binders were characterized further and shown to inhibit the enzyme to a certain degree. Furthermore, acarbose, an active site binding inhibitor, did interfere with the binding of the variants. It thus seems likely that the novel binders bound in close proximity of the enzyme active site.

We also characterized the two PPA binding variants by creating structural molecular models of the modules using the wild type as a template. The models indicated no significant main chain conformational perturbations between the mutants and the wild type. In the variant CBD$_{PPA1.5}$ one non-randomised residue A20 has been deleted, but even this deletion did not seem to have a severe impact on the backbone conformation. The subsequent residues, S21 (now S20) and G22 (now G21), which are loop residues, are positionally shifted to compensate for the deletion, thereby shortening the loop that they constitute. Also in this variant the C19-C35 SS-bond is still maintained in the model. However, the C19 C-alpha atom is shifted towards the deletion by 0.6Å nevertheless not significantly affecting the SS-bond lengths.

In a different set of experiments, biopanning against metal ions (Ni$^{2+}$) was performed, using the library, in order to test the scaffold for metal capture purposes (publication VI). Several Ni$^{2+}$-binders were obtained after five panning rounds. Since the wild type CBM$_{Cel6A}$ had been successfully displayed on the surface of S. carnosus (publication IV), we choose to investigate metal capture in a whole bacterial cell format by genetically engineering the selected metal binders on the surface of the gram-positive bacteria. All of the variants were readily displayed on the bacterial surface and two of the variants showed significantly enhanced Ni$^{2+}$-binding capacity (publication VI). The selected metal binders featured numerous histidine residues on the surface indicating that the imidazole side chains coordinated the Ni$^{2+}$-ions. The same principle of creating affinity is used in immobilized metal affinity chromatography. In the eight selected variants, histidine residues appear in all of the randomized positions suggesting that all of the side chains point towards the solution.
The side chain orientation is important for effective display of various chemical properties and for the formation of a stable folded protein. This type of recombinant bacteria, with enhanced metal binding capacity, has potential use in biofilters or biosensors for environmental applications.

In the different selections we obtained novel binding proteins exploiting the CBM1 fold. Variants with novel affinities could be isolated from the library and that all of the produced variants formed single soluble proteins. The randomized positions form a relatively large surface area on the 36 amino acid residue CBM1$_{Cel7A}$ and it is remarkable that the fold tolerates such an extensive randomization. The selection of novel binders towards two widely different types of epitopes shows that the created library is an excellent source of novel binding molecules.

8. Concluding remarks

Following years of detailed structure-function studies of different CBMs, a detailed picture on the structural basis of their affinity to polycarbohydrates and their role in enzymatic carbohydrate degradation is beginning to emerge. Nevertheless, several open questions remain to be studied. In this work, the function and applications of the small, stable family 1 cellulose-binding modules were studied. To address the questions on CBM1 specificity on the crystalline cellulose as well as their function in module interactions, several mutagenesis, electron microscopy and domain exchange studies were performed. For the first time, direct experimental evidence was obtained for the postulated CBM-cellulose interaction on the fully exposed sugar rings on crystalline cellulose. Further experiments will be needed to verify whether this is the only interaction or one of the possible modes of binding. Details of the mechanisms of this and the other types of protein-carbohydrate interaction also require continued efforts. Especially interesting will be comparisons between the constantly growing numbers of different CBMs as well as studies addressing the mechanism of the seemingly irreversible binding of some CBMs.

In recent years, several interesting applications exploiting CBM have been reported. CBMs have not only been used successfully in the traditional affinity applications but also for example in transgenic trees to enhance cellulose production. In this work, the use of a CBM as an immobilization anchor was studied for two different purposes: protein immobilization
and live bacterial cell immobilization. The CBM1 was also used successfully as a protein scaffold for protein engineering. The CBM1 fold was proven to be very stable and novel binders were selected against two very different targets. One of these novel binders was cloned on the surface of bacteria for metal capture purposes. There is potential to further develop the recombinant CBM1-bacteria created for many biotechnological and environmental applications as well as to use the CBM-library to discover further binding molecules with a variety of affinities.
9. Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBM</td>
<td>carbohydrate binding module</td>
</tr>
<tr>
<td>CBD</td>
<td>cellulose binding domain (old nomenclature)</td>
</tr>
<tr>
<td>PASC</td>
<td>phosphoric acid swollen cellulose</td>
</tr>
<tr>
<td>BMCC</td>
<td>bacterial microcrystalline cellulose</td>
</tr>
<tr>
<td>HEC</td>
<td>hydroxyl methyl cellulose</td>
</tr>
<tr>
<td>CMC</td>
<td>carboxyl methyl cellulose</td>
</tr>
<tr>
<td>TEM</td>
<td>transmission electron microscopy</td>
</tr>
<tr>
<td>CALB</td>
<td>Candida antarctica lipase B</td>
</tr>
<tr>
<td>ABP</td>
<td>albumin binding protein</td>
</tr>
<tr>
<td>PPA</td>
<td>porcine pancreatic α-amylase</td>
</tr>
<tr>
<td>CipA</td>
<td>Clostridium thermocellum scaffolding protein</td>
</tr>
<tr>
<td>Cel</td>
<td>cellulase enzyme</td>
</tr>
</tbody>
</table>
10. Acknowledgements

This thesis is the result of many late lonely hours in the lab, as well as joyful teamwork and collaborations. I would like to sincerely thank all of you who I have worked with during these years.

Most of all I would like to thank Professor Tuula T. Teeri for enthusiastic supervising of this thesis and for your never-ending optimism, always believing that a great scientific discovery is just around the corner. I had no idea what the future would bring when I, as an undergraduate student, called to the VTT telephone exchange.

I would like to thank Professor Per-Åke Nygren for your supervision of the phage work and your ever so philosophic approach to biochemistry. Professor Stefan Ståhl, thanks for your great, uplifting optimism and positive attitude. Without you I never would have expressed CBD on staff “for the first time ever”. I also thank Professor Mathias Uhlén for running the great DNAcorner with great style.

I will also like to thank Professor Junji Sugiyama for letting me peek into the electron microscope. It turned out to be: YATAA!

I have fantastic friends! Thank you Linda, Jussi, Christin, Henrik, Malin and Mark for you valuable help during these last crazy weeks. Without you I never would have managed to print this work in time.

Thanks to all of my colleagues at the Department of Biotechnology for your welcoming attitude. You really made me feel like home from the first day when I arrived from Finland.

I am especially grateful to:

Henrik W for the successful staff-stuff, but most of all for being a good friend. One of these days you might even beat me on a tennis court. ☺ Malin G, you are always as enthusiastic and helpful, no matter if it is work or sushi-lunch. Patrik for introducing me to the world of gram-positive bacteria as well as really making me feel welcome to DNAcorner. Linda F –not only for your wonderful, sacrifising help with the thesis, but also for the concept-days. Malin E for great fun both in the lab and outside, maybe we’ll get to have a real business lunch one of these days. Jenny O for ”skrivarstugan”, lets pretend we still have to do it in future.

I would also like to thank Aspe, Åsa, Ulla, Soraya, Tarja N., Markus L., Elin, Marianne, Jenny R., Eva, Peter, Savo, Jakob, Karin, Maria W., C. och E, Nina, Anna G. och B, Martin L, Martin B, Alex, Stuart, Sofia, Fiona, Fredrik, Joke, Kristoffer, Ulla, Urban, Tornbjörn, Kristina, Joakim W and L and all the rest of you whom I’ve worked with during these five years at VTT, DNAcorner and at ex-XVe.
I want to say my thanks to all of the good friends who helped me to stay in touch with the outside world. Jussi for being my best friend for over 25 years. There just isn’t a thing in this world that I do which you would not understand. Christin, five years ago I would have never guessed how much laughter and fun there was in store for me when I met you. You are my favorite swede! Katarina for without you the Troikka would not be complete. Thanks also Uwe, Jouko, Timo, Pelle, Pekka, Mark, Al, Jaakko and all the others for the many superb moments.

I would also like to thank my grandparents for keeping my nutrition well balanced during my study years with great dinners!

Jyri and Mona and especially Eemeli and Oskari for the lively moments. I hope I can spend more time with you now from now on.

Finally I would like to thank my parents for your full support. You have provided me with a shelter and freedom at same time and most of all a steady platform from where I could jump as high as I possibly can.
11. References


CARBOHYDRATE-BINDING MODULES


