Specific interactions between carbohydrate-binding modules and cellulose

by

BRADLEY WILLIAM MCLEAN
M.Sc., Queen's University, 1995
B.Sc.H., Queen's University, 1991

A THESIS SUBMITTED IN PARTIAL FULFILMENT
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(DEPARTMENT OF MICROBIOLOGY AND IMMUNOLOGY
AND THE BIOTECHNOLOGY LABORATORY)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 2001
© Bradley William McLean, 2001
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Microbiology and Immunology

The University of British Columbia
Vancouver, Canada

Date April 26, 2001
Abstract

Polysaccharidases and the carbohydrate-binding modules (CBMs) have evolved to mirror the diversity inherent in the crystalline and amorphous structures of cellulose and other polysaccharides abundant in nature. In this study, the binding specificities for cellulose of CBMs representing each of the three functional types of CBMs (types A, B and C), were determined by a series of pair-wise competition binding experiments. These studies show that different binding sites of phosphoric acid-swollen cellulose (PASC) are bound by different CBMs: type A localize to crystalline regions, type B localize to the amorphous regions. The data suggest that there is heterogeneity in the amorphous regions of PASC.

The binding of the family 2a CBM from *Cellulomonas fimi* xylanase 10A (CBM2a) to insoluble cellulose was characterized in more detail. Competition experiments indicate that CBM2a binds to the crystalline regions of cellulose. It has been reported that CBM2a binds irreversibly to cellulose however, it is shown here that CBM2a bound to the cellulose surface can be displaced by CBM2a in solution. CBM2a has a number of conserved residues on the binding face, including three tryptophans that have been implicated in the binding reaction. Site-directed mutation and a Langmuir-type adsorption isotherm analysis was used to determine the individual contributions of W17, W54, W72 and a number of neighbouring residues to the overall binding affinity of CBM2a for cellulose. Each tryptophan plays a different role in binding; a tryptophan is essential at position 54, a tyrosine or tryptophan at position 17 and any aromatic residue at position 72. Other residues on the binding face, with the exception of N15, are not essential determinants of binding affinity. In an attempt to formulate a structural model of CBM2a bound to the surface of cellulose, observation of \(^{15}\)N-labelled CBM2a bound to \(^{13}\)C-enriched cellulose using solid state NMR was attempted. Given the specificity of CBM2a, the structure of crystalline cellulose and the dynamic nature
of the binding of CBM2a, a model is proposed for the interaction between the polypeptide and the crystalline surface that is also likely to apply to other type A CBMs.
# Table of Contents

ABSTRACT .......................................................................................................................... ii  
TABLE OF CONTENTS ........................................................................................................ iv  
LIST OF TABLES ................................................................................................................... viii  
LIST OF FIGURES ................................................................................................................... ix  
ABBREVIATIONS .................................................................................................................. xi  
STATEMENT OF PUBLICATION ........................................................................................... xii  
STATEMENT OF COLLABORATION ..................................................................................... xiii  
ACKNOWLEDGEMENTS .......................................................................................................... xiv  

1. INTRODUCTION .............................................................................................................. 1  
   1.1 CELLULOSE .................................................................................................................... 1  
   1.2 POLYSACCHARIDE HYDROLASES AND CELLULASES .................................................. 2  
   1.3 CARBOHYDRATE-BINDING MODULES ........................................................................ 4  
      1.3.1 CBM nomenclature ..................................................................................................... 5  
      1.3.2 Type A CBMs ............................................................................................................ 6  
         1.3.2.1 Family 1 ................................................................................................................ 6  
         1.3.2.2 Family 2a .............................................................................................................. 7  
         1.3.2.3 Family 3 ................................................................................................................ 9  
         1.3.2.4 Family 5 .............................................................................................................. 10  
         1.3.2.5 Family 10 ........................................................................................................... 11  
      1.3.3 Type B CBMs ........................................................................................................... 13  
         1.3.3.1 Family 2b ............................................................................................................. 13  
         1.3.3.2 Family 4 .............................................................................................................. 14  
         1.3.3.4 Family 17 ........................................................................................................... 15  
      1.3.4 Type C CBMs ........................................................................................................... 17  
         1.3.4.1 Family 9 ............................................................................................................... 17  
      1.3.5 Applications of CBMs .............................................................................................. 20  
   1.5 GOAL OF THIS RESEARCH .......................................................................................... 20  

2. MATERIALS AND METHODS ......................................................................................... 22  
   2.1 MATERIALS AND REAGENTS .................................................................................... 22
2.2 BACTERIAL STRAINS AND PLASMIDS

2.2.1 Escherichia coli strains and plasmids

2.2.2 Acetobacter xylinum strains

2.3 CULTURE MEDIA

2.3.1 E. coli growth media

2.3.2 A. xylinum growth media

2.4 DNA Manipulations

2.4.1 Synthesis of DNA fragment encoding CBM2a

2.4.2 Construction of CBM2a mutants

2.5 PRODUCTION AND PURIFICATION OF CBMS

2.5.1 Production of CBMs used in competition isotherms

2.5.2 Production and purification of CBM2a and mutants

2.5.2.1 M9 Media for production of \(^{15}N\)-CBM2a

2.5.2.2 Production of \(^{15}N\)-labelled CBM2a

2.6 PRODUCTION AND PROCESSING OF BACTERIAL CELLULOSE

2.6.1 Production of \(^{13}C\)-enriched Cellulose

2.6.1.1 Production of \(^{13}C\)-4-enriched Cellulose

2.6.1.2 Production of \(^{13}C\)-4-enriched cellulose

2.6.2 Processing of bacterial cellulose

2.7 DETERMINATION OF AFFINITIES

2.8 COMPETITION BINDING ISOHERMS

2.8.1 Labelling of CBMs with fluorescent probes

2.8.2 Standard curves of fluorescence vs. polypeptide concentration

2.8.3 Isotherms by fluorescence

2.8.4 Competition isotherms

2.9 ADSORPTION AND SURFACE EXCHANGE OF CBM2A TO CELLULOSE

2.9.1 Surface exchange of CBM2a

2.9.2 Kinetics of adsorption to PASC and BMCC

2.9.3 Kinetics of CBM2a exchange on BMCC
2.10 NMR                                                                                                             40
  2.10.1 Solid state NMR: Cross-polarization/magic angle spinning (CP/MAS)......................................................... 40
  2.10.2 NMR in solution ............................................................................................................................................. 41

3. RESULTS .............................................................................................................................................. 42
  3.1 PRODUCTION OF His<sub>6</sub>-CBM2A .............................................................................................................. 42
  3.2 BINDING OF CBM2A TO CELLULOSE ................................................................................................................. 43
  3.3 BINDING OF CBM2A MUTANTS TO BMCC ............................................................................................................. 49
    3.3.1 Tryptophan mutants ........................................................................................................................................... 49
    3.3.2 Other mutants .................................................................................................................................................... 50
  3.4 COMPETITION BINDING ISOTHERMS .................................................................................................................... 52
    3.4.1 CBMs bind to PASC and BMCC ......................................................................................................................... 52
    3.4.2 Labelling of CBMs ............................................................................................................................................. 53
    3.4.3 Standards and Isotherms by fluorescence ........................................................................................................ 54
    3.4.4 CBM2a adsorption to PASC in the presence of BSA .......................................................................................... 56
    3.4.5 A model of competition binding ....................................................................................................................... 61
    3.4.6 CBM2a in competition with CBM4-1, CBM17 and CBM9-2 ............................................................................ 66
    3.4.7 CBM2a in competition with Cel6A-CBM2a and CBM3 .................................................................................... 73
    3.4.8 CBM4-1 and CBM17 competition ....................................................................................................................... 78
  3.5 CBM2A BINDING ADSORPTION AND EXCHANGE .................................................................................................. 81
    3.5.1 Rate of Adsorption of CBM2a to PASC and BMCC ........................................................................................ 81
    3.5.2 Equilibrium exchange of CBM2a on BMCC ..................................................................................................... 81
    3.5.3 Rate of CBM2a exchange on BMCC ................................................................................................................ 83
  3.6 CBM2A-CELLULOSE INTERACTION OBSERVED BY SOLID STATE NMR .............................................................. 86
    3.6.1 Production and characterization of 13C-enriched bacterial cellulose .............................................................. 86
    3.6.2 CBM2a-cellulose interaction ............................................................................................................................ 87

4. DISCUSSION ........................................................................................................................................... 92
  4.1 CELLULOSE AND CBMS: A DIVERSITY OF STRUCTURE AND FUNCTION .............................................................. 93
    4.1.1 Binding specificity of CBMs for cellulose ...................................................................................................... 93
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.1.1</td>
<td>CBM2a binds to crystalline regions of cellulose</td>
<td>93</td>
</tr>
<tr>
<td>4.1.1.2</td>
<td>Heterogeneity of amorphous cellulose structure</td>
<td>97</td>
</tr>
<tr>
<td>4.1.2</td>
<td>CBMs: agents for targeting polysaccharidases and for saturating substrate</td>
<td>100</td>
</tr>
<tr>
<td>4.2</td>
<td>ADSORPTION OF CBM2A TO CELLULOSE</td>
<td>103</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Irreversibility of adsorption of CBM2a</td>
<td>103</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Amino acid residues involved in the binding of CBM2a to crystalline cellulose</td>
<td>106</td>
</tr>
<tr>
<td>4.2.3</td>
<td>CBM2a-Cellulose interaction observed by solid state NMR</td>
<td>109</td>
</tr>
<tr>
<td>4.3</td>
<td>TOWARDS A MODEL OF CBM2A-CELLULOSE INTERACTION</td>
<td>111</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Previous models of binding</td>
<td>111</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Surface area of the exposed crystalline faces</td>
<td>112</td>
</tr>
<tr>
<td>4.3.3</td>
<td>A model for the binding of CBM2a to crystalline cellulose</td>
<td>113</td>
</tr>
<tr>
<td>4.4</td>
<td>A GENERAL MODEL FOR THE BINDING OF TYPE A BINDING MODULES</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>REFERENCES</td>
<td>117</td>
</tr>
</tbody>
</table>
List of Tables

TABLE 2.1 Oligonucleotides used for the construction of the synthetic His<sub>6</sub>-CBM2A gene fragment

TABLE 2.2 Oligonucleotides used for the construction of His<sub>6</sub>-CBM2A mutants

TABLE 2.3 CBMs used in competition isotherms on PASC

TABLE 3.1 Binding affinity and capacity for the adsorption of CBM2A to insoluble cellulose

TABLE 3.2 Binding affinity of CBM2A tryptophan variants for BMCC

TABLE 3.3 Binding affinity of CBM2A variants for BMCC

TABLE 3.4 CBM binding affinity and capacity for BMCC and PASC

TABLE 3.5 CBMs, labels and number of potential labelling sites for CBMs used in competition binding experiments

TABLE 3.6 Binding of CBM2A to PASC: UV absorbance, fluorescence and BSA competition

TABLE 3.7 CBM2A-OG in competition with CBM2A-AX on BMCC

TABLE 3.8 PASC competition binding summary I: CBM2A, CBM4-1, CBM17 and CBM9-2

TABLE 3.9 Competition binding summary II: CBM2A, Cel6A-CBM2A and CBM3

TABLE 3.10 Competition binding summary III: CBM17 and CBM4-1

TABLE 3.11 Exchange at equilibrium of CBM2A bound to BMCC
List of Figures

**FIGURE 1.1** MODEL OF A PORTION OF THE PLANT CELL WALL FROM GRASSES .......................................................... 2

**FIGURE 1.2** THE CELLULASE SYSTEM OF *C. FIMI* .......................................................................................................... 3

**FIGURE 1.3** ALIGNMENT OF FAMILY 2A CBMS FROM *C. FIMI* ...................................................................................... 9

**FIGURE 1.4** STRUCTURES OF TYPE A CBMS .................................................................................................................. 12

**FIGURE 1.5** STRUCTURES OF TYPE B CBMS .................................................................................................................. 16

**FIGURE 1.6** STRUCTURES OF TYPE C CBMS .................................................................................................................. 19

**FIGURE 2.1** RESTRICTION MAP OF THE CBM2A EXPRESSION VECTOR PTUGK-H6-IEGR-CBM2A ..................... 23

**FIGURE 2.2** SCHEMATIC FOR THE TWO-STEP CONSTRUCTION OF THE HIS$_6$-CBM2A FRAGMENT ....................... 26

**FIGURE 3.1** PURIFICATION OF HIS$_6$-CBM2A FROM CONCENTRATED *E. coli* CULTURE SUPERNATANT ............... 43

**FIGURE 3.2** CP/MAS SPECTRA OF INSOLUBLE CELLULOSES ...................................................................................... 47

**FIGURE 3.3** BINDING OF CBM2A TO THE INSOLUBLE CELLULOSES PASC, BMCC AND AVICEL ...................... 48

**FIGURE 3.4** TWO VIEWS OF CBM2A HIGHLIGHTING THE MUTATED RESIDUES ......................................................... 51

**FIGURE 3.5** EMISSION SPECTRA OF CBM2A LABELLED WITH OREGON GREEN OR AMCA-X ......................... 57

**FIGURE 3.6** STANDARD CURVES OF FLUORESCENCE VS. PROTEIN CONCENTRATION: PASC EXPERIMENTS .......... 58

**FIGURE 3.7** STANDARD CURVES OF FLUORESCENCE VS. PROTEIN CONCENTRATION: BMCC ............................ 59

**FIGURE 3.8** BINDING OF CBM2A-OG TO PASC: UV, FLUORESCENCE AND BSA ...................................................... 60

**FIGURE 3.9** BINDING OF CBM2A-OG TO BMCC IN THE PRESENCE OF CBM2A-AX (1) ........................................ 64

**FIGURE 3.10** BINDING OF CBM2A-OG TO BMCC IN THE PRESENCE OF CBM2A-AX (2) .......................................... 65

**FIGURE 3.11** BINDING OF CBM2A-OG TO PASC IN THE PRESENCE OF CBM4-1 ....................................................... 69

**FIGURE 3.12** BINDING OF CBM4-1-AX TO PASC IN THE PRESENCE OF CBM2A-OG ............................................... 70

**FIGURE 3.13** BINDING OF CBM17-OG TO PASC IN THE PRESENCE OF CBM2A ...................................................... 71

**FIGURE 3.14** BINDING OF CBM2A-OG IN THE PRESENCE OF CBM9-2 ................................................................. 72

**FIGURE 3.15** BINDING OF CBM2A-OG TO PASC IN THE PRESENCE OF CBM3 .......................................................... 75

**FIGURE 3.16** BINDING OF CBM2A-OG TO PASC IN THE PRESENCE OF Cel6A-CBM2A-AX ................................. 76

**FIGURE 3.17** BINDING OF CBM2A-OG TO BMCC IN THE PRESENCE OF Cel6A-CBM2A-AX ............................ 77

**FIGURE 3.18** BINDING OF CBM17-OG TO PASC IN THE PRESENCE OF CBM4-1-AX ........................................... 79

**FIGURE 3.19** BINDING OF CBM4-1-AX TO PASC IN THE PRESENCE OF CBM17-OG .............................................. 80

**FIGURE 3.20** ADSORPTION KINETICS OF CBM2A BINDING TO PASC AND BMCC .................................................. 84

**FIGURE 3.21** SURFACE EXCHANGE OF CBM2A AT EQUILIBRIUM ON BMCC ....................................................... 85

**FIGURE 3.22** KINETICS OF CBM2A SURFACE EXCHANGE ON BMCC ............................................................... 85

**FIGURE 3.23** CP/MAS SPECTRUM OF BACTERIAL CELLULOSE FROM GLUCOSE-13C-4 ENRICHMENT ......... 89

**FIGURE 3.24** CP/MAS SPECTRA OF BMCC AND 13C-ENRICHED CELLULOSE FROM 1,3-13C-GLYCEROL SYNTHESIS ................................................................. 90

**FIGURE 3.25** 15N CP/MAS SPECTRUM OF 15N-CBM2A BOUND TO INSOLUBLE CELLULOSE AND HSQC OF 15N-CBM2A IN SOLUTION ................................................. 91

**FIGURE 4.1** CBMS BOUND TO NATIVE CELLULOSE ........................................................................................................ 97

**FIGURE 4.2** DETAILS OF THE BINDING SITES OF CBM4-1 AND CBM17 ............................................................... 99
FIGURE 4.3 A MODEL OF CBM2A SURFACE EXCHANGE .................................................. 104

FIGURE 4.4 THREE VIEWS EACH OF TWO POSSIBLE ARRANGEMENTS OF CBM2A BOUND TO THE CRYSTALLINE SURFACE OF BMCC .................................................. 116
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>amino acid</td>
</tr>
<tr>
<td>$A_n$</td>
<td>absorbance at wavelength “n”</td>
</tr>
<tr>
<td>AMCA-X or AX</td>
<td>6-((7-amino-4-methylcoumarin-3-acetyl)amino)hexanoic acid, succinimidyl ester</td>
</tr>
<tr>
<td>BMCC</td>
<td>bacterial micro-crystalline cellulose</td>
</tr>
<tr>
<td>bp</td>
<td>base pair, base pairs</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CBD</td>
<td>cellulose-binding domain</td>
</tr>
<tr>
<td>CBM</td>
<td>carbohydrate-binding module</td>
</tr>
<tr>
<td>CP/MAS</td>
<td>cross-polarization/magic-angle spinning</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethylsulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>EHEC</td>
<td>ethylhydroxyethylcellulose</td>
</tr>
<tr>
<td>$dH_2O$</td>
<td>distilled water</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxyribonucleoside triphosphate</td>
</tr>
<tr>
<td>FW-CBM2a</td>
<td>5-fluoro-L-tryptophan-labelled CBM2a</td>
</tr>
<tr>
<td>H&amp;S</td>
<td>Hestrin and Schramm medium</td>
</tr>
<tr>
<td>HEC</td>
<td>hydroxyethylcellulose</td>
</tr>
<tr>
<td>HPLC</td>
<td>high pressure liquid chromatography</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum correlation</td>
</tr>
<tr>
<td>$K_a$</td>
<td>association affinity constant</td>
</tr>
<tr>
<td>kDa</td>
<td>kiloDalton</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>NBS</td>
<td>N-bromosuccinimide</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>OG</td>
<td>Oregon Green® 514 carboxylic acid, succinimidyl ester</td>
</tr>
<tr>
<td>o/n</td>
<td>overnight</td>
</tr>
<tr>
<td>PASC</td>
<td>phosphoric acid-swollen cellulose</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>TMS</td>
<td>tetramethysilane</td>
</tr>
<tr>
<td>TYP</td>
<td>tryptone, yeast extract, phosphate medium</td>
</tr>
<tr>
<td>U</td>
<td>unit, units</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>v</td>
<td>volume</td>
</tr>
<tr>
<td>w</td>
<td>weight</td>
</tr>
</tbody>
</table>
Statement of Publication

A portion of this thesis presents data and conclusions previously issued in the following publication:


As indicated by the order of authorship, the majority of the work presented in this publication, the competition binding experiments, construction of the eleven of the fourteen CBM2a mutants, and all of the binding analysis and was done by the thesis author, B. W. McLean. The synthetic CBM2a gene fragment was designed by R. Graham and C. Rodriguez. M. R. Bray constructed three of the fourteen CBM2a mutants. The modified Langmuir-type binding model presented was developed collaboratively with A. B. Boraston. N. R. Gilkes provided details and impetus for the cellulose surface area calculations. N. R. Gilkes, C. A. Haynes and D. G. Kilburn all acted in a supervisory role.

D. G. Kilburn, Senior Author
Statement of Collaboration

The solid state NMR experimental data presented here is the product of an ongoing collaboration between the laboratories of Drs. R. A. J. Warren, Department of Microbiology and Immunology, D. G. Kilburn, Department of Microbiology and Immunology and the Biotechnology Laboratory, and Dr. Colin A. Fyfe, Department of Chemistry, UBC. Dr. C. A. Fyfe and Darren Brouwer, a Ph.D. candidate supervised by Dr. Fyfe measured all of the solid state NMR spectra presented here. All of the samples measured by NMR were prepared by B. W. McLean. It has been agreed by all collaborators that these results are appropriate to be included in this thesis.
Acknowledgements

If it takes a village to raise a child, then it takes a bustling metropolis to raise a Ph.D. candidate; I will attempt to express my gratitude to and acknowledge some of individuals involved in the life of this particular Ph.D. candidate. First of all, I must thank Tony Warren and Doug Kilburn, my supervisors, for providing guidance, wisdom, and humour; you set an example of excellent leadership and mentorship. Thanks are also extended to the other members of my supervisory committee, Chip Haynes and Colin Fyfe, for sharing their insights, expertise and for their enthusiastic guidance. Together, my committee leaps other supervisory committees in a single bound! I also must thank Peter Tomme, and Neil Gilkes for their peerless expertise and their eagerness to share it.

Thanks are extended to Dr. Colin Fyfe and Darren Brouwer for their enthusiastic collaboration resulting in the solid state NMR work presented here. The collection of the $^{15}\text{N-CBM2a HSQC}$ spectrum by Cameron Mackereth and Dr. Lawrence Mcintosh is also very much appreciated.

The Cellulase Lab has been a welcoming second home for the past few years. I thank all of the members of the Cellulase Lab, past and present, who help to make it a vital dynamic place to work, learn and be. All members of the lab have left their fingerprints on this document in some manner, I will try to name at least some of them here. To all of people in room 324, past and present, thank you all for scientific inspiration for friendship (Cheers!) especially to Al Boraston, Mark Bray, Jeff Kormos, and Dominik Stoll. I am indebted to Emily Kwan and, Emily Amandorn-Akow for all of your assistance and encouragement (technical and otherwise); somewhere in this thesis is that last isotherm, thanks in no small part to your help, and to Helen Smith who kept the lab running so smoothly; your generosity and spirit is admirable. And all of the rest of you, whom I am forgetting, my thanks as well!
I could not have done this without the foundation of personal support that I am lucky to have. Thank you to my parents, who have always been so very supportive of my scientific pursuits (starting at a very young age!). I am also very grateful for the support of my whole extended family and family-in-law.

I owe so much of this to Julie, my love and my Gibraltar, who has been a source of unfaltering support and inspiration. With you, I am always reminded that I am truly blessed and lucky.
1. Introduction

1.1 Cellulose

Cellulose, the major structural component of plant material, is the most abundant biopolymer on earth; each year plants synthesize about $4 \times 10^9$ tonnes (21). Cellulose is also produced by certain bacteria, marine invertebrates, fungi, slime moulds and amoebae (99). Cellulose is a polymer of $\beta$-1,4-linked glucose units; in nature, the cellulose molecules are aggregated together, in parallel, in the lattice forming microfibrils. The lateral dimensions of the microfibrils range in size from 3 nm to 20 nm depending on the biological source (18). Despite its chemically simple structure, cellulose can assume a diversity of complex structures or allomorphs, with unique secondary and tertiary conformations (2). Naturally produced cellulose, cellulose I, itself is a mixture of two structural forms, $I_\alpha$ and $I_\beta$ (3). Other allomorphs, such as cellulose III and the antiparallel cellulose II, arise after subjecting natural cellulose to harsh chemical treatments (2).

In the plant cell wall, cellulose is intimately associated with other polysaccharides and lignin. Collectively, the other associated polysaccharides are called hemicellulose and may include, but are not restricted to, mannans, galactomannans, xylan, xyloglucans, pectin and some glucans (Figure 1.1). All of these polysaccharides are arranged in the plant cell wall in close proximity to each other, with domains of one type of polysaccharide, such as cellulose, for example, immediately adjacent to, or embedded in a matrix of another polysaccharide, such as xylan and xyloglucan, for example. The precise arrangement and composition of cellulose depends on plant species.
Figure 1.1 Model of a portion of the plant cell wall from grasses. This model shows the complexity and intimate association of the component polysaccharides, such as xyloglucans and cellulose (15). The inset figure is a model of a fibril, with cellulose microfibrils associated with a matrix of hemicelluloses and lignin (adapted from http://www.chemistry.vt.edu/chem-dept/helm/3434WOOD/notes1/morphology.pdf)

1.2 Polysaccharide Hydrolases and Cellulases

Organisms have evolved diverse strategies for taking advantage of the rich and abundant carbon sources found in plant materials by using a complement of polysaccharide hydrolases (134). Polysaccharidase enzyme systems act in a concerted manner, often synergistically, to degrade natural substrates (26, 86, 134). An excellent example of a well characterized polysaccharidase system is the complement of enzymes from the cellulase system of the gram-positive aerobic soil bacterium Cellulomonas fimi (Figure 1.2).
The cellulase system of *C. fimi*. This schematic presents the major cellulases produced by *C. fimi*: Cel6A (formerly CenA), Cel9A (CenB), Cel9B (CenC), Cel5A (CenD), Cel6B (CbhA), Cel48A (CbhB) and Xyn10 (Cex). The key to domain and module structures of *C. fimi* polysaccharidases is provided in the figure.

The *C. fimi* cellulases, like other polysaccharidases, are modular (36). Each enzyme is comprised of a series of distinct contiguous amino acid sequences that fold into discrete modules and are often separated by linker sequences (Figure 1.2). The modules include catalytic modules and ancillary modules, the most common of which are substrate binding modules. Cellulases themselves are an important subject of study because of their the potential application in the commercial conversion of biomass into fermentable sugars, as well as other industrial processes such as the clarification of fruit juices, and as additives to laundry detergents for the removal of small fibres, thus brightening the colours.
1.3 Carbohydrate-binding modules

Amongst the ancillary modules comprising cellulases, the most common type mediates binding to substrate (36, 129). The first of the substrate-binding modules characterized bound to cellulose and were known as cellulose-binding domains (CBDs). A large number of other substrate-binding modules have been discovered with binding specificity for polysaccharides other than cellulose, so these are now collectively referred to as carbohydrate-binding modules (CBMs). Within the polysaccharidases, they can be located at either the N-or C-termini of the enzyme or internally; they can appear singly or in multiples, and may also occur in conjunction with CBMs from other families. The discrete modular structure of CBMs makes them amenable to study as functioning units isolated from the source enzyme by either limited proteolysis or genetic manipulation.

CBMs are classified into at least 17 families of related amino acid sequences (24). The families themselves can be grouped into three types based on structure, function and ligand-binding specificity. Type A CBMs (Figure 1.4) bind to insoluble surfaces, type B CBMs (Figure 1.5) bind to oligosaccharides having more than three monosaccharide units and type C (Figure 1.6) bind to mono- and disaccharides (9).

A major role of CBMs is to promote the hydrolysis of insoluble substrates by increasing the local concentration of the enzyme on the substrate (74, 82); the role of CBMs that bind to soluble sugars is not entirely clear. Removal of CBMs from enzymes reduces their activities on insoluble substrates; hydrolysis of soluble substrates is not affected (38, 47, 98, 128). CfCel6A-CBM2a has been shown to disrupt the surface of fibres of ramie cotton and release small particles and thus may further increase the available surface area accessible to an organism’s hydrolases (28). This effect, however, has not been demonstrated with other CBMs. Furthermore, CBMs may promote the hydrolysis of substrate (16) by targeting the
partner hydrolases to specific types of ligands. This diversity of CBM binding characteristics could ensure that, given a substrate that is both heterogeneous in composition and structure such as that of the plant cell wall, the maximum substrate surface area available to the organism is covered by its secreted polysaccharidases.

1.3.1 CBM nomenclature

As yet there is not a consensus on the naming of CBMs, with research groups tending to use an idiosyncratic approach. With the explosion in the number of CBMs characterized, and the advent of high throughput methods of protein discovery and characterization, it is imperative to have a standard system for naming them so that each one can be uniquely identified. Our group has proposed a system of CBM nomenclature (9) that is consistent with the accepted glycosyl hydrolase designation scheme (50); in both schemes, polypeptides are grouped into families of related amino acid sequences. The scheme for CBMs was refined recently. For the purposes of this work, the following convention will be followed: a two letter abbreviation of the source organism (such as “Cf” for \textit{C. fimi}), followed by the identifier of the parent enzyme or polypeptide designation (based on the polysaccharidase classification system (50)) then “CBM”, for “carbohydrate-binding module”, the family designation and binding module number, if more than one module of any particular family type occurs in the same enzyme or protein. That is, <two letter abbreviation of source organism><glycosyl hydrolase source name>-CBM<carbohydrate-binding module family><CBM module number if any>. For example, the family 2a binding module from the \textit{Cellulomonas fimi} family 10 xylanase A, (formerly CBD\textsubscript{Cex}) would be designated as CfXyn10A-CBM2a. The first N-terminal binding module from the \textit{C. fimi} family 9 cellulase B (formerly CBD\textsubscript{N1}) is CfCel9B-CBM4-1, the second N-terminal binding module is CfCel9B-CBM4-2 (formerly CBD\textsubscript{N2}); a construct containing both family 4 binding modules would be CfCel9B-CBM4-1.2. Once
defined, these examples could be abbreviated to CBM2a, CBM4-1, CBM4-1 and CBM4-1.2, respectively.

1.3.2 Type A CBMs

This type of CBM ranges in size from 35 aa to 140 aa and includes the CBMs from families 1, 2a, 3, 5, 10 and 12. They share an affinity for insoluble cellulose, with insignificant affinity for soluble sugars or cello-oligosaccharides. Structurally, they all have a platform of aromatic residues aligned along one face of the globular polypeptide (Figure 1.4). These aromatic residues, tryptophan, tyrosine, and occasionally histidine and phenylalanine, are often involved in the binding of the type A CBMs to cellulose (12, 14, 27, 66, 85, 94, 97, 130, 141). These residues are commonly involved in protein-carbohydrate interactions (95, 100, 131, 135).

1.3.2.1 Family 1

Among the first CBMs to be described were those from the *Trichoderma reesei* cellobiohydrolases CBHI (TrCel7A) and CBHII (TrCel6A) (128). With the exception of a polysaccharide-binding protein from the alga *Porphyra purpurea* (GenBank accession U08843), family 1 CBMs are derived entirely from fungal hydrolases. They are the smallest CBMs, typically comprising about 35 amino acids, and having a predominantly β-sheet structure with three conserved tryptophan or tyrosine residues forming a linear platform on one face of the polypeptide. The solution structures of TrCel6A-CBM1 (CBD-CbhI) (66) and TrCel7B-CBM1 (CBD(EGI)) (76) were solved by NMR. Family 1 CBMs bind to insoluble cellulose with an affinity in the range of $10^5 \text{ M}^{-1}$. They also bind to chitin, and interact weakly with cello-oligosaccharides (76, 77). TrCel7A-CBM1 (Figure 1.4-A) binds to bacterial micro-crystalline cellulose (BMCC) reversibly (73). In contrast, TrCel6A-CBM1
(CBDCbhII) binds irreversibly to cellulose(17, 91). Mutation of the three surface tyrosines of TrCel7A-CBM1 significantly reduced its affinity for insoluble cellulose(72, 97), however the extent of affinity reduction was not quantified. The exposed aromatics of other family 1 CBMs are, presumably, also essential in binding to insoluble cellulose.

1.3.2.2 Family 2a

Family 2a is, by far, the most extensively studied family of type A binding modules. Currently, it comprises 54 members, most of which come from bacterial sources, but some are from nematodes (29, 115). Family 2a CBMs comprise approximately 100 amino acids. They bind to crystalline cellulose, including bacterial microcrystalline cellulose (BMCC), and also to partially crystalline preparations, such as phosphoric acid-swollen cellulose (PASC)(88) with an affinity of $1 - 3 \times 10^6 \text{ M}^{-1}$ (this work and unpublished data). It is not clear, however, if they have significant affinity for the amorphous regions of phosphoric acid-swollen cellulose (PASC). A solution structure of CfXyn10A-CBM2a (hereafter abbreviated to CBM2a) was solved (140). It has a $\beta$-barrel type backbone with a relatively flat face on which there are several conserved tryptophan residues (Figure 1.3) involved in binding to cellulose (Figure 1.4-B). It is likely that other family 2a members have a very similar tertiary structure. The involvement of the conserved aromatic residues in binding to cellulose was clearly demonstrated for several family 2a CBMs, including the Cellulomonas fimi modules Cel6A-CBM2a (formerly CBD$_{CenA}$)(27) and CfXyn10A-CBM2a (formerly CBD$_{Cex}$)(12), and Pseudomonas cellulosa Xyn10A-CBM2a (85, 94).

Binding of CfXyn10A-CBM2a to either BMCC or PASC is apparently irreversible. Complete removal of the unadsorbed molecules from the solution phase does not lead to significant desorption of bound molecules over extended (i.e., several weeks) periods of
equilibration ((25, 57) and B. W. McLean, unpublished data). Although binding is irreversi
ble, the interaction with the cellulose surface is dynamic. Surface diffusion measurements using fluorescence recovery techniques show that CBM2a, either in its isolated form or as a module in xylanase 10A, is mobile on the surface of crystalline cellulose (57), and may therefore provide bound xylanase 10A access to a much broader field of substrate. The rate of oxidation of the three conserved surface tryptophan residues of CBM2a (W17, W54 and W72) with N-bromosuccinimide is only halved when CBM2a is bound to BMCC, suggesting a dynamic interaction of the tryptophans with the cellulose surface which, at least periodically, allows each residue to become solvent exposed and susceptible to oxidation. Furthermore, complete oxidation of the three tryptophan residues eliminates binding but not proper folding of CBM2a, indicating the importance of these hydrophobic residues to the binding interaction (12).

The thermodynamic behaviour of the binding of CBM2a to BMCC is different from that of the typical protein-carbohydrate interaction. Binding is entropically driven. The favourable enthalpy (ΔH) of binding is quite small making only a small contribution to the overall energetics relative to the change in entropy (ΔS). The thermodynamics together with a large negative heat capacity change (ΔC_p) are consistent with the binding being driven by dehydration of the sorbent and the protein surfaces, specifically the tryptophan-rich binding face(25).

The individual contributions of W17, W54, and W72 to the overall binding interaction however, are unknown. It is also unclear whether any other solvent-exposed amino-acid residues proximal to the three tryptophans make a significant contribution to the energetics of binding.
1.3.2.3 Family 3

Family 3 CBMs are common in the non-catalytic scaffolding proteins of the large, multi-component cellulosome complexes that are produced, most commonly, by cellulolytic, anaerobic bacteria. In addition to a CBM, scaffolding proteins contain dockerin modules that bind to the cohesin modules of the component cellulases. The CBM of the scaffolding ensures close interaction between the insoluble target substrates and the associated cellulosome. Family 3 CBMs are the largest of the type A CBMs, usually containing ~150 amino acids (6) and bind to insoluble cellulose with an affinity of ~10^6 M^-1 ((41, 84) and this work), similar to that of CBM2a. Two CBM3s have been characterized in some detail: one from the mesophilic Clostridium cellulovorans, the CBM from cellulose-binding protein A (CcCbpA-CBM3; formerly designated CBD_clos hereafter abbreviated CBM3) (40, 41) and another from the thermophilic C. thermocellum, the CBM from cellulose-binding protein A (CtCipA-CBM3; formerly Cip-CBD)(84).
Three crystal structures of three family 3 CBMs are known. Two of them bind to insoluble cellulose: CtCipA-CBM3 (Figure 1.4-C) (130) and a similar CBM from the mesophilic Clostridium cellulolyticum (CcelCipC-CBM3; formerly CipC-CBD)(109). Their structures are somewhat similar to that of CBM2a, consisting of two antiparallel $\beta$-sheets. Both of the family 3 CBMs have a calcium ion binding site. Residues that are conserved among family 3 members map to two regions of these CBMs: a shallow groove of underdetermined function on the top of the molecule, and a second area on a flat face on the opposite side of the molecule thought to be the region that interacts with insoluble cellulose. The flat face has conserved tyrosine, tryptophan and histidine residues that are likely the primary residues involved in the binding interaction. In addition, there are residues on this face that could form hydrogen-bonds with the cellulose surface. The role of each of these conserved residues is under investigation. The third CBM3 structure is from Thermomonospora fusca cellulase 9A (previously E4) (103). It is an internal CBM, flanked by the catalytic domain and a family 2a CBM. It does not bind to crystalline cellulose nor does it have the conserved residues that are implicated in the binding of other CBM3s to cellulose. The role of TfCel9A-CBM3 may be to guide a cellulose molecule into the active site of the catalytic module (103) while the CBM2a anchors the enzyme to the insoluble substrate.

1.3.2.4 Family 5

One structure was solved for the 62 amino acid long family 5 CBM from Erwinia crysanthermi Cel5A (formerly EGZ)(14); it is shaped like a ski boot. Two tryptophans and one tyrosine on one face of the molecule (Figure 1.4-D) are required for effective binding to insoluble cellulose(111). One of these tryptophans is also required for secretion of Cel5A(19)
1.3.2.5 Family 10

The best studied family 10 binding module is at the N-terminus of *Pseudomonas cellulosa* xylananase 10A (PcXyn10A-CBM10). It comprises 45 amino acids and has similar binding specificity for crystalline cellulose preparations as CBM2a. However, it binds with an association affinity constant of $2.5 \times 10^5 \text{ M}^{-1}$, approximately 6 times lower than that of CBM2a for similar substrates (39). Its solution structure comprises two anti-parallel $\beta$-sheets and a short $\alpha$-helix (96). Like other type A CBMs, CBM10 has a number of tyrosine and tryptophans, that are conserved within the family, three of which form a platform of aromatic residues along one face of the polypeptide (highlighted in Figure 1.4-E). Two tryptophans and the single tyrosine forming this platform are instrumental in mediating binding to insoluble cellulose (93). Investigation of the thermodynamic roles of other conserved residues located on the binding face is in progress (H.J. Gilbert, personal communication).
Figure 1.4 Structures of type A CBMs. A) *T. reesei* cellulase 7A-CBM1 B) *C. fimi* xylanase 10A-CBM2a C) *C. thermocellum* CipA-CBM3 D) *E. crysanthemi* cellulase 5A-CBM5 E) *P. cellulosa* xylanase 10A-CBM10. Ribbons indicate secondary structure; residues likely to be involved in binding are rendered as “balls and sticks”. All figures were prepared using MOLMOL (63).
1.3.3 Type B CBMs

Type B CBMs may be thought of as “chain binders”. They bind preferentially to single oligosaccharides longer than three monosaccharide units and to soluble polymers. They may also bind to insoluble substrates, such as PASC, probably to accessible single cellulose chains in these preparations. The binding of this type of CBM is both structurally and thermodynamically similar to the classic protein-carbohydrate interactions exhibited by lectins. Type B CBMs (Figure 1.5) include families 2b, 4, 17, and the recently reclassified family 22 (20). Representatives from families 2b, 4 and 17 will be discussed in more detail here.

1.3.3.1 Family 2b

The seven members of family 2b share amino acid sequence similarity with the CBMs in family 2a. The internal CBM2b from *C. fimi* xylanase 11A (CfXyn11A-CBM2b-1, hereafter CBM2b-1; Figure 1.5-A), is a typical family 2b CBM. It binds to soluble xylan, with an affinity of \(\sim 10^3 \text{M}^{-1}\). Despite the obvious amino acid sequence similarity with CBM2a (26% identity 39% similarity with two gaps comprising 14% of the aligned region) it does not bind to crystalline cellulose. Structurally, members of families 2a and 2b are strikingly similar except for the positions of the surface tryptophans. In CBM2a, the surface tryptophans are in line and coplanar, forming a flat hydrophobic ridge on one face of the molecule; in CBM2b-1 the tryptophan corresponding to W17 of CBM2a is orthogonal to the face of the protein and, with the tryptophan corresponding to W54 of CBM2a, forms a twisted binding site for the accommodation of a helical xylan polysaccharide chain (112). CBM2b-1 lacks a tryptophan corresponding to W72 of CBM2a. When the position of the tryptophan corresponding to W17 of CBM2a was altered by mutating an adjacent residue, so that it lay parallel to the face
of the polypeptide, the mutant bound to crystalline cellulose but had no appreciable affinity for xylan (113). The tryptophan residues of family 2 CBMs therefore determine both the affinity and the specificity of binding.

1.3.3.2 Family 4

The N-terminal family 4 CBM from C. fimi Cel9B (CenC) is the most thoroughly characterized type B CBM. This CBM, CfCel9B-CBM4-1 (formerly CBD\textsubscript{NI}, hereafter abbreviated CBM4-1; Figure 1.5-B) is the N-terminal repeat of the tandem family 4 CBMs of Cel9B. It was the first CBM observed to bind cellobiose, soluble cellulose derivatives, and the soluble glucopyranoside polymer, barley \(\beta\)-glucan, with significant affinity (\(K_a \sim 10^5 \text{ M}^{-1}\)) (61, 125). It binds to insoluble cellulose preparations with significant amorphous content but not to crystalline cellulosic, such as BMCC (23).

Solution structures were solved for CBM4-1 (60) and its adjacent module CfCel9B-CBM4-2 (formerly CBD\textsubscript{N2}; Figure 1.5-C) (13), in the presence of saturating amounts of cellotetraose and cellopentaose, respectively. The two CBMs share 34% amino acid sequence identity and they have very similar structures. Both comprise two antiparallel \(\beta\)-sheets folded into a jelly-roll sandwich with an obvious groove or cleft on one side. In both CBMs, the cleft is lined by a strip of non-polar amino acids and flanked by hydrogen-bonding residues. The tyrosines in the groove of CBM4-1 are essential for tight binding of insoluble and soluble substrates (64). CBM4-1 may bind the sugar chain in either orientation (58), reflecting the symmetry of the residues in the binding cleft. The crystal structure of CBM4-1 in the presence of cellopentaose confirms the positioning of the substrate within the binding cleft and the specific protein-carbohydrate interactions (A. Boraston, V Notenboom and A. Freelove, personal communication/unpublished results).
The binding of CBM4-1 and CBM4-2 is enthalpically driven, consistent with structural evidence for the formation of favorable hydrogen bonds with the equatorial hydroxyls of the cellulose oligomer. CBM4-1 binds a calcium ion at a site located away from binding groove (59). The binding of calcium does not influence the binding of ligands but more likely stabilizes the module. The two modules together in the full enzyme are not separated by any obvious linker region. The affinity of CBM4-1.2 for amorphous cellulose is roughly the sum of the affinities of each module, indicating that the binding is not cooperative (13).

1.3.3.4 Family 17

Family 17 CBMs are typified by the 200 amino acid module at the C-terminus of the non-cellulosomal *C. cellulosorans* cellulase 5A (formerly endoglucanase EngF), CcCel5A-CBM17 (hereafter CBM17; Figure 1.5-D). It binds to insoluble cellulose (8, 56, 108), cello-oligosaccharides and soluble cellulose derivatives, such as hydroxyethyl-cellulose (HEC) and ethylhydroxyethyl-cellulose (EHEC). CBM17 binds soluble cello-oligosaccharides, with the greatest affinity for cellohexaose \( K_a = 1.5 \times 10^5 \text{ M}^{-1} \). Overall, CBM17 has similar ligand specificity to CBM4-1, however, CBM17 binds with an affinity approximately twice that of CBM4-1. Even though CBM17 and CBM4-1 have similar affinities for cello-oligosaccharides comprised of six or five glucose units, respectively, PASC has approximately three times the binding capacity for CBM17 than CBM4-1 ((8) and this work). It is not known if CBM17 and CBM4-1 recognize identical regions of the amorphous cellulose in PASC and this will be investigated.
Figure 1.5 Structures of type B CBMs. A) *C. fimi* xylanase 11A CBM2b-1 B) *C. fimi* cellulase 9B-CBM4-1 C) *C. fimi* cellulase 9B-CBM4-2 D) *C. cellulovorans* cellulase 5A-CBM17 E) *C. thermocellum* xylanase 10B-CBM22-2. Ribbons indicate secondary structure; residues likely to be involved in binding are rendered as “balls and sticks”. All figures were prepared using MOLMOL (63).
1.3.4 Type C CBMs

Type C CBMs ("small sugar binders") are characterized by their ability to bind simple sugars and consequently, insoluble and soluble polysaccharides (Figure 1.6). These binding modules come from a variety of sources including animals, plants, crustaceans and microbes. They all bind to mono- and disaccharides with significant affinity. They also bind to insoluble and soluble polysaccharide chains and are often desorbed from these matrices by small soluble sugars. CBMs in families 6 (102), 9 (137), and 13 (10) are all type C. CBM family 9 is discussed in more detail below.

1.3.4.1 Family 9

The second, C-terminal family 9 binding module from xylanase 10A of the hyperthermophilic bacterium Thermotoga maritima (TmXyn10A-CBM9-2, hereafter abbreviated to CBM9-2) is comprised of ~170 amino acids. It is unique among the CBMs characterized to date in that it binds strictly to the reducing ends of sugars. CBM9-2 adsorbs to insoluble polysaccharides with an affinity of $1-3 \times 10^6 \text{ M}^{-1}$; it binds to other soluble mono- and soluble polysaccharides with affinities that range from $10^3 \text{ M}^{-1}$ to $10^6 \text{ M}^{-1}$. Cello-oligosaccharides longer than cellobiose do not bind with any greater affinity, indicating that the preferred ligand is a disaccharide. The crystal structures of CBM9-2 alone (Figure 1.6-A), and in complexes with glucose and cellobiose, show that the binding site is a shallow, blind cleft in which there are two tryptophans with the planes of their ring faces parallel to each other, roughly perpendicular to the surface of the protein. The tryptophans form the entrance to the binding cleft and, when bound to ligand, sandwich the disaccharide. Binding of the substrate is further stabilized by a network of hydrogen bonds involving a number of adjacent charged side-chains. As with other lectins (75) and CBMs with affinity for soluble sugar
ligands, binding to small soluble ligands and cellulose is exothermic, dominated by favourable enthalpy with unfavourable changes in entropy (A. B. Boraston et al.; V. Notenboom et al., in press)
Figure 1.6 Structures of type C CBMs. A) Thermotoga maritima xylanase 10A-CBM9-2 B) Streptomyces lividans xylanase 10A-CBM13. Ribbons indicate secondary structure; residues likely to be involved in binding are rendered as “balls and sticks”. All figures were prepared using MOLMOL (63).
1.3.5 Applications of CBMs

CBMs have a number of applications that take advantage of their ability to bind to cellulose, which is an inexpensive, abundant and inert matrix. CBMs can be used as affinity tags for protein purification (124, 126). Because it can be desorbed from cellulose using relatively inexpensive sugars, such as glucose, CBM9-2 is especially useful as an affinity purification tag. The irreversible binding of CBM2a allows it to be used for the immobilization of processing enzymes, such as the glycosidases, EndoF, and PNGaseF (A.B. Boraston et al., manuscript in preparation), β-glucosidase (87) and factor Xa (44). The relatively small, compact structures of CBMs make them good scaffolds for the display of libraries of combinatorial peptides (114, 117).

1.5 Goal of this research

The goal of this research is to further elucidate the nature of the interaction of CBMs with their ligand. In nature, cellulose forms a myriad of complex structures and is intimately associated with other polysaccharides. Similarly, as discussed in the previous sections, CBMs have an array of structures and a number of different binding specificities. The ligand targeted by each CBM in a complex substrate, composed of a mixture substrate structures, is not precisely known. In this study, the binding specificities of a number of CBMs for the complex cellulose PASC will be determined by a series of pair-wise competition binding experiments. These experiments will determine the type of cellulose (i.e. crystalline vs. amorphous) bound by each CBM; currently, there is no method for directly evaluating the binding specificity of CBMs for insoluble substrates, particularly complex substrates such as PASC. Overall, these experiments will demonstrate that the diversity of CBM structure and
function is merely a reflection of the diversity of structures that cellulose and cellulose-associated polysaccharides can, and do, adopt.

Because of its prominence in the *C. fimi* cellulase system and its potential applications, the interaction and the binding of CBM2a to cellulose will be investigated in more detail with a goal of working towards the formulation a functional and structural model of binding to insoluble cellulose. CBM2a binds to cellulose with apparent irreversibility; its ability for bound molecules to exchange with molecules in the solution phase, as demonstrated with other cellulases, will be investigated. CBM2a has a number of conserved residues on the binding face, including three tryptophans that have been implicated in the binding reaction. Site-directed mutation and a Langmuir-type adsorption isotherm analysis will be used to determine the individual energetic contributions of W17, W54, W72 and a number of neighbouring residues to the overall binding affinity of CBM2a for cellulose. Structural information regarding CBMs bound to the surface of insoluble cellulose is difficult to obtain using solution-based methods, such as liquids NMR. This study will work towards a structural view of CBM2a bound to the surface cellulose by the use of solid state NMR; the first attempt to directly observe a CBM bound to insoluble cellulose. In the future, this work will contribute towards a structural and functional model of the binding interaction between CBM2a and the surface of insoluble cellulose.
2. Materials and methods

2.1 Materials and reagents

All chemicals were of analytical or high pressure liquid chromatography (HPLC) grade, and purchased from Sigma (St. Louis, MO), BDH (Toronto, ON), or ICN (Aurora, OH), unless otherwise noted. Solutions were prepared as described by Sambrook et al. (104).

Avicel® PH-101 was from FMC International (Cork, Ireland). Bacterial microcrystalline cellulose (BMCC) and phosphoric acid-swollen cellulose (PASC; prepared from Avicel® PH-101) were prepared by Emily Kwan, as described previously (37, 52, 139).

2.2 Bacterial strains and plasmids

2.2.1 Escherichia coli strains and plasmids

CBM2a and its mutants were produced using the Escherichia coli strains JM101 (142), R1360, and BL21(DE3) (120) (Novagen, Inc.; Madison, WI). The expression vector used for production of CBM2a and its mutants, pTugK-H6-IEGR-CBM2a, is a derivative of the pTugA and pTugAS vectors described previously (43). It encodes resistance to kanamycin; it also carries a gene fragment encoding the leader peptide of Xyn10A at the N-terminus, followed by six histidine residues, a factor Xa protease cleavage site (amino acid sequence, IEGR), and then a synthetic gene fragment encoding CBM2a at the C-terminus (Figure 2.1).
Figure 2.1 Restriction map of the CBM2a expression vector pTugK-H6-IEGR-CBM2a. This vector has a pUC origin of replication (pUC Ori), encodes for kanamycin resistance (Kmr), and the Lac repressor. It has a gene fragment encoding the leader peptide of Xyn10A at the N-terminus, followed by six histidine residues, a factor Xa protease cleavage site (amino acid sequence, IEGR), and then a synthetic gene fragment encoding CBM2a at the C-terminus. Only unique restriction endonuclease recognition sites are indicated.

2.2.2 Acetobacter xylinum strains

A. xylinum strains 53524, 23769, 100821, and 700178 were obtained from the American Type Culture Collection (ATCC; http://www.atcc.org) and from laboratory stocks maintained by Emily Kwan.

2.3 Culture media

2.3.1 E. coli growth media

E. coli were cultured in tryptone, yeast extract, phosphate medium (TYP) (104) or M9 minimal medium. M9 minimal medium was prepared as follows: a solution containing 6 g of Na₂HPO₄, 3 g of KH₂PO₄, 0.5 g NaCl, and 1.0 g of NH₄Cl made up to 100 ml with dH₂O was
sterilized and added to 871 ml of sterile dH₂O. The following components, each dissolved in dH₂O and sterilized separately, were then added: 2 ml of 1 M MgSO₄, 1 ml of 0.01 M FeCl₃, 1 ml of 1 mg/ml thiamine and 25 ml of 40% (w/v) glucose.

2.3.2 *Acetobacter xylinum* growth media

Two media were used to culture *A. xylinum* in the production of bacterial cellulose: a rich undefined medium and a defined minimal medium. One litre of buffered H&S medium (53), the rich medium, was prepared as follows: 5 g of yeast extract, 5 g of peptone, 2.7 g of sodium phosphate and 1.15 g of citrate were dissolved in 950 ml of dH₂O, adjusted to pH 5.0 with HCl and then sterilized. 50 ml of 40% (w/v) glucose (in dH₂O, sterilized separately) were then added. One litre of the minimal medium (31) was prepared as follows: a solution was made (Stock Solution A) containing 1.15 g of citric acid, 1.0 g NH₄Cl, 2.7 g Na₂HPO₄, and 0.1 g KCl made up to 100 ml with dH₂O, and sterilized. The defined medium was made by adding Stock Solution A, 10 ml of MgSO₄·7H₂O (25-g/liter stock solution prepared and sterilized separately), 25 ml of 40% (w/v) glucose (filter sterilized) and 1 ml of sterile 7.5 mg/ml stock solution of niacin to 864 ml of sterile dH₂O for a final volume of one litre.

2.4 DNA manipulations

All DNA manipulations were performed by standard methods as described in Sambrook *et al.* (104). Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs (NEB; Beverly, MA) or Life Technologies Canada (Burlington, ON), and used in the buffers provided by the manufacturers according to their directions.
2.4.1 Synthesis of DNA fragment encoding CBM2a

The DNA fragment encoding CBM2a was synthesized by a PCR-based procedure, similar to that used in the synthesis of a xylanase gene from *Schizophyllum commune* (42). The sequence incorporated a number of unique restriction sites to facilitate manipulation of the fragment after site-directed mutation (Figure 2.1), and the codon bias was changed to that of *E. coli*, both without changing the encoded amino acid sequence. Six oligonucleotides, synthesized by the UBC Nucleic Acid and Protein Services Unit (NAPS), were used as primers in the PCR assembly of the complete fragment (Table 2.1). The synthesis was performed in two steps, as outlined in the schematic presented in Figure 2.2.

**Table 2.1** Oligonucleotides used for the construction of the synthetic His$_6$-CBM2a gene fragment.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Oligonucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCx1</td>
<td>5'GACTAGAATTTCAGGAGGAAACAGCTATGCGGCGTACTACACCAGAC-3'</td>
</tr>
<tr>
<td>SCx2</td>
<td>5'ATGCCCGGTACTACACCAGCTCCGGGTCAACCGGCCGCTCGGTGCTCGTACGCTCGTCGGGTCACCCGGCTGACCCCATGTTGTTGGTGCTACCGTTGCTTCTGCCAGCTAGGCTGCTAGCGGTCCAGCCGGCTGCCAGGTTCTGTGGGGTG</td>
</tr>
<tr>
<td>SCx3</td>
<td>5'ACCTGCTGACCCGAGGGAAAAGAGAAGGTCAGGGTCCAACCGTCAACCGGAGCGAGAGCTCGTGTTTTTAACGGTAACGTTAGCGGTGAAACCGGTGTTCCACTGGTTAACACCCCACAGAACCCTGGCAG-3'</td>
</tr>
<tr>
<td>SCx4</td>
<td>5'TAGATGACCAAGCCTGAGTTACCTGCTGACCCGAGGGAAAAGAGAAGGTCAGGGTCCAACCGTCAACCGGAGCGAGAGCTCGTGTTTTTAACGGTAACGTTAGCGGTGAAACCGGTGTTCCACTGGTTAACACCCCACAGAACCCTGGCAG-3'</td>
</tr>
<tr>
<td>SCx5</td>
<td>5'TGACAAGGGGTACCGTTCAGAGAGAAAGCGGTTGGAGCAGCGTTGGTACCGGTGAAACCGGTGTTCCACTGGTTAACACCCCACAGAACCCTGGCAG-3'</td>
</tr>
<tr>
<td>SCx6</td>
<td>5'GTTACAAGCTTTTATTAACCAACGGTGCAAGGGGTACCGTTCAGA-3'</td>
</tr>
</tbody>
</table>
Figure 2.2 Schematic for the two-step construction of the His$_6$-CBM2a fragment. The first step (A) resulted in a DNA fragment of approximately two-thirds the length (SCx(2/3)) of the final DNA fragment. SCx(2/3) was used in a second PCR reaction (B) resulting in the full length product that was cloned into pUC19 and pTugK.

The first reaction mixture contained 5 pmol each of longer primers SCx2 and SCx3, 200 μM of each deoxynucleoside triphosphate (dNTP), 3 U of Expand Hi-Fi Polymerase (Roche, Mannheim, Germany), and its recommended buffer supplemented with DMSO to 5%, in a final volume of 50 μL. After five cycles of 1 min at 95°C followed by 3 min at 72°C in a Perkin-Elmer model 2400 Thermocycler, 20 pmol each of the flanking primers SCx1 and SCx4 were added. Synthesis of the segment was completed by 25 cycles of 45 s at 95°C, 45 s at 58°C, and 1 min at 72°C. The second reaction mixture contained 1 ng of the product from the first reaction, 23.00, 2.30, or 0.23 pmol of primer SCx5 (all three primer concentrations yielded PCR product), 2.5 U of Pwo polymerase in its recommended buffer supplemented with DMSO to 5% (v/v), and 200 μM of each dNTP, in a final volume of 50 μL. After five cycles of 1 min at 95°C followed by 3 min at 72°C, 20 pmol each of the flanking primers SCx1 and SCx6 were added and synthesis of the fragment completed by twenty-five cycles of 45 s at 95°C, 45 s at 55°C, and 1 min at 72°C. The full length fragment was purified using the Qiaquick PCR purification kit (Qiagen Inc., Chatsworth CA), digested with EcoRI and
HindIII, then ligated into pUC19 (142) that had been digested with the same enzymes. After checking its sequence, the fragment was subcloned into the pTugK vector.

2.4.2 Construction of CBM2a mutants

All mutations were made by two-primer PCR mutagenesis (1) or the 'megaprimer' method (106, 116). A typical two primer reaction contained 50 ng of template DNA, 5 pmol each of the flanking (primer BG2, Table 2.2) and the mutagenic primers (Table 2.2), 200 μM of each dNTP, 1 U of VentR Polymerase (NEB) in its recommended buffer. Amplification was obtained by 30 cycles of 30 s at 96°C, 30 s at 53°C and 60 s at 72°C. The product was purified with a Qiaquick PCR purification kit (Qiagen Inc.), digested with the appropriate restriction enzymes, and ligated into a pTugK vector that had been digested with the same enzymes. Mutants of residues W17, W54, W72 and Q52 were made by this method. Mutants of residues N15, N18, Q83, and N87 were made by using a megaprimer strategy. The first product was amplified using the mutagenic primer and the BG2 flanking primer as described above. The product of this reaction was used as a megaprimer using the oligonucleotide Lax16 as a flanking primer. PCR conditions were as described previously (116). Briefly, 2 μg of megaprimer product were used with 50 ng template (pTugK-H6-IEGR-CBM2a) and 5 pmol of the Lax16 flanking primer, added after 5 cycles of denaturation at 96°C for one min, and extension at 72°C for three min. The remaining 30 cycles of the PCR program were as described above.
Table 2.2 Oligonucleotides used for the construction of His<sub>6</sub>-CBM2a mutants

<table>
<thead>
<tr>
<th>Primer</th>
<th>Oligonucleotide sequence&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>W17Y</td>
<td>5′TGGGGTGTTACCCAGTATAAACACTGGTTTCACCCTAACGCTAAC3′</td>
</tr>
<tr>
<td>W17F</td>
<td>5′TGGGGTGTTACCCAGTTTAACACTGGTTTCACCCTAACGCTAAC3′</td>
</tr>
<tr>
<td>W17A</td>
<td>5′TGGGGTGTTACCCAGGCGAACACTGGTTTCACCCTAACGCTAAC3′</td>
</tr>
<tr>
<td>W54Y</td>
<td>5′TTTCCCTCGGGTACGAGGTACACTCGGTATACTCTCTAAGTCTAACCCTAACGCTAAC3′</td>
</tr>
<tr>
<td>W54F</td>
<td>5′TTTCCCTCGGGTACGAGGTACACTCGGTATACTCTCTAAGTCTAACCCTAACGCTAAC3′</td>
</tr>
<tr>
<td>W54A</td>
<td>5′TTTCCCTCGGGTACGAGGTACACTCGGTATACTCTCTAAGTCTAACCCTAACGCTAAC3′</td>
</tr>
<tr>
<td>W72Y</td>
<td>5′CAGTCTGGATGCCGCTGTTACCGTACGTAACGCTCCGTAACCGCTAACGCTAAC3′</td>
</tr>
<tr>
<td>W72F</td>
<td>5′CAGTCTGGATGCCGCTGTTACCGTACGTAACGCTCCGTAACCGCTAACGCTAAC3′</td>
</tr>
<tr>
<td>W72A</td>
<td>5′CAGTCTGGATGCCGCTGTTACCGTACGTAACGCTCCGTAACCGCTAACGCTAAC3′</td>
</tr>
<tr>
<td>N15A</td>
<td>5′GGCTGCACCAGTACTGTTGGGTTGGGCTGCACTGGGAACACTGGTTGGACCCGCT3′</td>
</tr>
<tr>
<td>N18A</td>
<td>5′GGCTGCACCAGTACTGTTGGGTTGGGCTGCACTGGGAACACTGGTTGGACCCGCT3′</td>
</tr>
<tr>
<td>Q52A</td>
<td>5′TTTCCCTCGGGTACGAGGTACACTCGGTATACTCTCTAAGTCTAACCCTAACGCTAAC3′</td>
</tr>
<tr>
<td>Q83A</td>
<td>5′CTATCTCGGTTGGAACCGCTGACGTTGGTCTCTACAGTCTAACCCTAACGCTAAC3′</td>
</tr>
<tr>
<td>N87A</td>
<td>5′GGTTGGCCCGCGCTGGTACGAGGTACGCGGAACCACGCTGACGTTGGTCTCTACAGTCTAACCCTAACGCTAAC3′</td>
</tr>
<tr>
<td>N87H</td>
<td>5′GGTTGGCCCGCGCTGGTACGAGGTACGCGGAACCACGCTGACGTTGGTCTCTACAGTCTAACCCTAACGCTAAC3′</td>
</tr>
<tr>
<td>Lax16</td>
<td>5′TAGGACCACGCGCA3′</td>
</tr>
<tr>
<td>BG2</td>
<td>5′TGATCAGATCTTGATCCGCTAACGCTAACGCTAAC3′</td>
</tr>
</tbody>
</table>

<sup>a</sup>Underlined nucleotides indicate the loci of changes made for amino acid substitution or introduction of a restriction endonuclease recognition site

2.5 Production and purification of CBMs

2.5.1 Production of CBMs used in competition isotherms

Representative CBMs from a number of different CBM families were used in the competition binding studies (Table 2.3). All of these CBMs were produced recombinantly and provided from a number of sources, as indicated. The molar extinction coefficients, calculated from aromatic amino acid content by the method of Pace and co-workers, are also indicated (90).
### Table 2.3 CBMs used in competition isotherms on PASC

<table>
<thead>
<tr>
<th>CBM</th>
<th>Biological Source</th>
<th>Extinction Coefficient ((\varepsilon_{280} \text{M}^{-1}\text{cm}^{-1}))</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>CfCBM2a-Xyn10A</td>
<td><em>C. fimi</em> Xylanase 10A (Cex)</td>
<td>27 625</td>
<td>a</td>
</tr>
<tr>
<td>CfCBM2a-Cel6A</td>
<td><em>C. fimi</em> Cellulase 6A (CenA)</td>
<td>36 105</td>
<td>b</td>
</tr>
<tr>
<td>CcCBM3-CbpA</td>
<td><em>C. cellulovorans</em> Cellulose-binding protein A (CBD&lt;sub&gt;cl0S&lt;/sub&gt;)</td>
<td>24 300</td>
<td>c</td>
</tr>
<tr>
<td>CfCBM4-1-Cel9B</td>
<td><em>C. fimi</em> Cellulase 9B (CenC)</td>
<td>21 370</td>
<td>d</td>
</tr>
<tr>
<td>CcCBM17-Cel5A</td>
<td><em>C. cellulovorans</em> Cel5A(EngF)</td>
<td>31 310</td>
<td>e</td>
</tr>
<tr>
<td>TmCBM9-2-Xyn10A</td>
<td><em>T. maritima</em> Xylanase 10A (XynA)</td>
<td>43 430</td>
<td>e</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Note:</th>
</tr>
</thead>
<tbody>
<tr>
<td>a provided by Philip Wendler and Roberta Farrell, Sandoz Chemicals Biotech Research Corporation; b prepared by Dr. Ken Tokayasu; c provided by CBD Technologies, Rehovot Israel; d prepared by Emily Kwan and Emily Amandoron-Akow; e provided by Dr. Alisdair Boraston and Patrick Chiu</td>
</tr>
</tbody>
</table>

#### 2.5.2 Production and purification of CBM2a and mutants

The full-length fragment was ligated into the vector pTugK in frame. Inclusion of the leader peptide of Xyn10A in the fragment allows export of CBM2a to the periplasm of *E. coli* from where it leaks into the culture supernatant (42, 88). Since the aim was to obtain mutants of CBM2a with reduced affinity for cellulose, a hexahistidine sequence followed by a factor Xa site was added to the N-terminus of mature CBM2a so that the mutant polypeptides could be purified by immobilized metal-chelate affinity chromatography (IMAC) (92). This had the added advantage of avoiding the denaturing conditions required to desorb CBM2a from cellulose if the mutants, unlike the wild-type, could not be refolded following desorption (88). In the event that the hexahistidine sequence affected the binding of CBM2a to cellulose, factor Xa could be used to obtain native CBM2a following purification by IMAC.

*E. coli* strains JM101 or R1360 were used for the production of wild-type CBM2a and its mutants. Typically, 500 ml of TYP containing 1.5 mM potassium phosphate and 50 μg kanamycin ml<sup>-1</sup> were inoculated with overnight cultures of an *E. coli* strain transformed with the appropriate vector. The cultures were shaken at 200 rpm at 37°C until the OD<sub>600</sub> was ~1.0, then induced with IPTG at a final concentration of 0.3 mM. After a further 36 hours, the cells were removed by centrifugation. The proteins in the supernatant were concentrated
and exchanged into IMAC binding buffer (5 mM imidazole, 5M NaCl, 20 mM Tris-HCl, pH 7.9) using a tangential flow filtration unit (Filtron Ultrasette, 1 kDa cut-off). The solution was passed through a column of Ni$^{2+}$-Sepharose (His-Bind resin; Novagen, Milwaukee, MI) previously charged with 10 column volumes of 50 mM NiSO$_4$ and equilibrated with binding buffer. The column was washed with binding buffer. Adsorbed polypeptide was eluted with binding buffer containing stepwise increases in the concentration of imidazole (Figure 3.1). The fractions containing polypeptide were detected by running approximately 20 µl of each fraction on a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel, then staining with Coomassie brilliant blue (104). Fractions containing CBM2a were pooled, desalted and concentrated in a stirred ultrafiltration cell (Amicon; Filtron 1 kDa cut-off filter). Protein purity and approximate yield were estimated by SDS-PAGE. Final protein concentrations were determined from the $A_{280\text{nm}}$, using $\varepsilon_{280}$ of 27 625 M$^{-1}$·cm$^{-1}$ for the wild-type polypeptide, 22 125 M$^{-1}$·cm$^{-1}$ for the single tryptophan to alanine or phenylalanine mutants and 23 615 M$^{-1}$·cm$^{-1}$ for single tryptophan to tyrosine mutants (90). The wild-type and mutants of CBM2a were obtained routinely in yields of up to 100 mg of purified polypeptide per litre of culture supernatant, except for the W54A mutant, for which the yield was approximately 20 mg per litre. In each case, sufficient polypeptide was obtained for binding analysis.

2.5.2.1 M9 Media for production of $^{15}$N-CBM2a

M9 minimal medium was used for the production of the $^{15}$N-labelled form of the CBM2a polypeptide, $^{15}$N-CBM2a. For its production, 1g of $^{15}$NH$_4$Cl was substituted for the NH$_4$Cl in the salts solution.
2.5.2.2 Production of $^{15}$N-labelled CBM2a

Five millilitre cultures of TYP, supplemented with kanamycin to 50 µg/ml, were inoculated with a single colony of *E. coli* BL21(DE3) that had been freshly transformed with pTugK-H6-IEGR-CBM2a. The 5 ml cultures were incubated overnight at 30°C. The next day, the cultures were washed twice by centrifugation and resuspended in 5 ml of M9 minimal medium. 500 µl of the resuspended culture was inoculated into 500 ml of M9 minimal medium, in which $^{15}$NH$_4$Cl was substituted for the NH$_4$Cl in the salts solution, supplemented with kanamycin to 50 µg/ml. The cultures were incubated at 30°C, with shaking. When the A$_{600}$ reached ~0.6, after approximately 12 h, the culture was induced by the addition of IPTG to a concentration of 0.3 mM and incubated overnight. Cells were harvested and $^{15}$N-CBM2a was purified from culture supernatant and the cell extract as described in sections 2.5.2. Typically, production yielded 6-10 mg of purified $^{15}$N-CBM2a per litre of culture.

2.6 Production and processing of bacterial cellulose

2.6.1 Production of $^{13}$C-enriched Cellulose

2.6.1.1 Production of $^{13}$C-4-enriched Cellulose

Strains of *A. xylinum* were obtained from the American Type Culture Collection (ATCC) and from laboratory stocks. The strains, ATCC numbers 53524, 23769, and 10821, were screened for their ability to grow and produce a substantial cellulose pellicle in the defined medium. In all of the following steps, duplicate cultures of each *A. xylinum* ATCC strain were inoculated and incubated at 30°C; one culture was shaken at 200 rpm and the other was static. Two 5 ml cultures of H&S medium were inoculated from frozen stocks with each *A. xylinum* strain. After 5 days, once a visible cellulose pellicle had formed, each culture was shaken briefly on a vortex mixer and 5 µl of each culture inoculated into 5 ml tubes of *A. xylinum* minimal
medium. Only ATCC strains 53524 and 23769 grew in minimal medium; both strains produced cellulose in shaken and static culture. After 10 days, each culture was briefly shaken on a vortex mixer and 500 μl inoculated into 100 ml of defined medium in 250 ml flasks. After approximately one month, both strains, in both agitated and static conditions, had produced cellulose but ATCC 23769 grown at 30°C without agitation produced the largest amount, measured by weight, after a series of processing steps (see section 2.6.2).

To produce ^13^C-4-enriched cellulose, 5 ml of defined medium were inoculated with 5 μl of a culture of ATCC 23769, that had been inoculated from frozen stocks. After five days of static incubation at 30°C, 500 μl of this culture were inoculated into each of two 250 ml flasks containing 100 ml of defined medium with 1% D-glucose-^13^C (Cambridge Isotope Laboratories, Andover, MA). After approximately one month of static incubation at 30°C, the cellulose pellicles were harvested and processed (as described in section 2.6.2).

2.6.1.2 Production of 1,3,4,6-^13^C-enriched cellulose

The synthesis of cellulose using D-glucose-^13^C did not label just the C-4 of the resulting cellulose, there was a certain amount of scrambling of the label, it was decided that a simple enrichment of ^13^C using the undefined H&S medium and less costly labelled carbon source, glycerol-1,3-^13^C₂, would be sufficient to make a larger amount ^13^C-enriched cellulose (30). A. xylinum ATCC strains 100821, 23769, 53524 and 700178 were screened for their ability to grow and produce cellulose in H&S medium containing 1% (w/v) glycerol (hereafter H&S-glycerol) instead of 1% (w/v) glucose, at 30°C without agitation. All strains produced cellulose in H&S-glycerol. However, cultures of ATCC 53524 produced the most cellulose (191.2 mg from 100 ml of culture). Larger ^13^C enrichment cultures were prepared as follows: 500 μl of a 5 ml culture of 53524 in H&S-glycerol were inoculated into each of four 125 ml
flasks containing 40 ml of H&S containing 0.5% (w/v) glycerol-1,3-\(^{13}\)C\(_2\) (Cambridge Isotope Laboratories, Andover, MA) and 0.5% glycerol (unlabelled). After 10 days of incubation at 30°C, cellulose pellicles and associated cells were harvested and processed as described in section 2.6.2.

2.6.2 Processing of bacterial cellulose

Cellulose pellicles from \(^{13}\)C-enrichment cultures were harvested and processed by a method similar to that described by Hestrin, 1963 (52). The thick cellulose and cell pellicle was removed, blended with 100 ml of 1 M NaOH at “Frappe” in an Osterizer cycoloc™ 10 blender (Sunbeam Corporation, Boca Raton FL) for 1 min, and incubated overnight at room temperature with gentle rocking. The cellulose was then washed extensively with dH\(_2\)O through a suction filter, rinsed with 200 ml 0.5% acetic acid, then washed with dH\(_2\)O until the pH of the filtrate was neutral. The bacterial cellulose was stored in dH\(_2\)O supplemented with sodium azide to 0.02% (w/v).

2.7 Determination of affinities

The affinities of CBMs and CBM2a mutants for insoluble cellulose were determined using a solution depletion method to generate binding isotherms. CBMs were equilibrated in microcentrifuge tubes with 1 mg of cellulose (BMCC or PASC) in 50 mM potassium phosphate (pH 7.0) in a total reaction volume of 1 ml. In binding experiments with Avicel\(^{\circledR}\), 5 mg of cellulose was used. The final total polypeptide concentrations ranged from 1 \(\mu\)M to 30 \(\mu\)M. Duplicate or triplicate samples were incubated at 4°C for at least three hours while rotating end over end. The cellulose was removed by centrifugation and the concentration of protein remaining in the supernatant, termed the free protein concentration, was determined by \(A_{280\text{nm}}\) after subtracting \(A_{350\text{nm}}\) to allow for light scattering. The concentration of bound
protein for a particular concentration of the CBM was the difference between a control
incubated without cellulose (total protein) and protein remaining in the supernatant in
samples after incubation with cellulose (free protein). An isotherm of [Bound] (μmol·g⁻¹ of
cellulose) vs [Free] (μM) was generated and binding parameters were determined by non­
linear regression of the Langmuir-type isotherm (Equation 1; modified from (27)), using
GraphPad Prism 3.0 for Windows (GraphPad Software Inc., http://www.graphpad.com):

\[
[B] = \frac{[N_0] - K_a([F] - G)}{1 + K_a([F] - G)} - G
\]  

(1)

where [B] is the concentration of bound protein (μmol·g⁻¹ cellulose), [N_0] is the total
concentration of binding sites (μmol·g⁻¹ cellulose), K_a is the association affinity constant
(μmol·l⁻¹), [F] is the concentration of free protein (μmol·l⁻¹) and G is a constant, calculated
during the regression, that corrects for non-protein absorbance at the measured wavelengths.
G therefore is the sum of all of the factors that contribute to absorption measurements not
specific to the CBM.

2.8 Competition binding isotherms

2.8.1 Labelling of CBMs with fluorescent probes

A series of CBM preparations was made, each labelled with one of two different fluorescent
probes so that in a binding reaction containing more than one CBM, the concentration of
each species could be measured irrespective of the A_280. Oregon Green® 514 carboxylic acid
succinimidyl ester and 6-((7-amino-4-methylcoumarin-3-acetyl)amino)hexanoic acid
succinimidyl ester (AMCA-X, SE) (Molecular Probes, Eugene Oregon) were chosen as the
two fluorescent probes because they have distinct excitation and emission characteristics
(Figure 3.5). Oregon Green® has an absorbance maximum at 506 nm, with an emission maximum of 537 nm. AMCA-X (AX) absorbs maximally at 352 nm with an emission maximum at 447 nm. These succinimidy l ester probes react with the primary amine groups of CBMs, the N-terminus, and solvent-exposed lysine residues, to form a stable covalent linkage. Labelling reactions were performed as directed by the manufacturer. Briefly, 1 mg of the fluorescent probe dissolved in 100 μl of dimethylsulfoxide (DMSO) was added to 10 - 15 mg of CBM in approximately 1 ml of 0.1 M sodium bicarbonate buffer, pH 8.3. The solution was mixed in the dark, at room temperature, for 1 hr. The labelled polypeptide was passed twice through a 5 ml Sephadex G-25 column (Pharmacia Biotech, Uppsala Sweden), equilibrated with 50 mM potassium phosphate (pH 7.0), to separate any unbound probe from the protein. Fractions containing protein, as detected by A$_{280}$, were collected and pooled. The labelling efficiency of each reaction was calculated based on the extinction coefficients provided for each label at their maximum absorption wavelength ($ε_{352}$ for AMCA-X is 18 500 M$^{-1}$·cm$^{-1}$; $ε_{506}$ for Oregon Green® is 86 200 M$^{-1}$·cm$^{-1}$; Molecular Probes).

2.8.2 Standard curves of fluorescence vs. polypeptide concentration

Because of the sensitivity of the fluorimeter (Perkin Elmer LS 50 Luminescence Spectrometer), labelled CBM was diluted with unlabelled CBM so that maximum measurable fluorescence emission (1 000 fluorescence units) coincided with typical maximum concentrations of CBMs used in an isotherm, usually 30 μM. The appropriate dilution was determined empirically by making a series of dilutions, measuring the fluorescence of each dilution, and adjusting the excitation and emission slit widths until the measured fluorescence fell within the range measurable by the fluorimeter. For example, this ratio was approximately 373 moles of unlabelled CBM2a per mole of Oregon Green®-
labelled CBM2a and 87 moles of CBM4-1 per mole of AMCA-X-labelled CBM4-1.

Fluorescence emission at 537 nm of Oregon Green-labelled CBMs was measured for 2.0 s using an excitation wavelength of 506 nm, excitation and emission slits appropriate for each labelled CBM preparation (typically 7.5 nm), and a 530 nm emission filter. Fluorescence emission at 447 nm of AMCA-X-labelled CBMs was measured using an excitation wavelength of 352 nm, without an emission filter, with a collection time of 2.0 s at the appropriate excitation and emission slit widths.

Conversion of fluorescence to molar protein concentration required a reliable standard curve. Standard curves for fluorescence vs. free and total protein concentration were obtained by performing an isotherm with each fluorescently-labelled CBM preparation. The fluorescence and net A$_{280}$ (the A$_{280}$ corrected for light scattering estimated from the A$_{350}$) of each duplicate set of free (CBM samples equilibrated with cellulose, and centrifuged to separate the supernatant from the cellulose) and total (CBM samples equilibrated in an equal volume of buffer, without cellulose) samples were measured. The fluorescence of each sample was plotted versus the calculated molar concentration of that sample. The contribution of the fluorescent probe conjugate to the A$_{280}$nm of the polypeptide solution was negligible. Standard curves were unique to each CBM preparation and each cellulose substrate used.

2.8.3 Isotherms by fluorescence

The veracity of each standard curve, as well as the effect of the fluorescent probe on binding, was assessed for each labelled CBM preparation by comparing the binding curves and binding parameters obtained by UV absorbance (net A$_{280}$) and fluorescence. A standard curve of free and total fluorescence vs. CBM concentration (µM) was constructed from a binding experiment using unblocked microcentrifuge tubes. A second isotherm, using tubes that had
been blocked in a solution of 1% bovine serum albumin (BSA) by soaking them for 2 hours then rinsing them dH₂O, was then performed. The fluorescence of each sample was measured and the molar concentrations of each free and total sample were calculated by interpolation using the appropriate standard curves. The resulting isotherms from UV absorbance and fluorescence measurements were plotted and compared.

2.8.4 Competition isotherms

In a binding competition experiment consisted of the measurement of the binding of one CBM to cellulose (BMCC or PASC) was measured in the presence of a constant concentration of a second CBM (a competitor). The binding modules used in the competition experiments (Table 2.1) were representatives of different CBM families. In each case, the CBM and competitor were added simultaneously. After equilibration and isolation of the supernatant, the fluorescence of the CBM and its competitor, was measured. The molar concentration of each component was calculated from interpolation using the appropriate standard curves. All competition experiments were performed using BSA-blocked microcentrifuge tubes to minimize any non-specific adsorption of the CBMs to the microcentrifuge tubes. In cases where the binding of a CBM, acting as a competitor, was affected by the presence of the fluorescent probe (eg. CBM9-2), an isotherm of one CBM (not affected by being labelled) was performed in the presence of an unlabelled competitor.

2.9 Adsorption and surface exchange of CBM2a to cellulose

2.9.1 Surface exchange of CBM2a

The extent to which CBM2a in solution is able to exchange with binding-modules bound to the surface of BMCC was assessed by an exchange experiment using two preparations of
CBM2a, one labelled with Oregon Green®, the other with AMCA-X. CBM2a-AX, at a total concentration of approximately 20 μM was incubated with 1 mg of BMCC at 4°C for at least three hours in a total sample volume of 1 ml. The cellulose was pelleted by centrifugation and 900 μl of the supernatant were removed. The amount of bound CBM2a-AX was calculated from the difference between the molar concentration of samples of equal volume incubated with and without cellulose as measured by AMCA-X fluorescence and interpolated using the appropriate standard. Each BMCC pellet was washed three times with 900 μl of 50 mM potassium phosphate buffer (pH 7.0) by centrifugation and subsequent removal of the supernatant. CBM2a-OG was added to the washed pellets at three concentrations: at approximately the measured free equilibrium concentration of CBM2a-AX (~4 μM), at greater than the equilibrium concentration (~15 μM) and at less than the equilibrium concentration (~2 μM). The pellets were resuspended and incubated for a further three hours. After removal of the cellulose by centrifugation, the fluorescence of CBM2a-OG and CBM2a-AX was measured. The concentrations of remaining CBM2a-AX bound, and CBM2a-OG bound were calculated.

2.9.2 Kinetics of adsorption to PASC and BMCC

The rate at which CBM2a reaches equilibrium with sites on the surface of insoluble cellulose was measured by incubating samples of CBM2a solution with 1 mg of BMCC or PASC for different times at 4°C. Cellulose was separated from the supernatant by passing the suspension through a Qiaquick spin column (Qiagen Inc.). The fluorescence values of the free and total samples were measured, the molar concentration determined, and the amount of bound CBM calculated. The fraction bound (CBM2a bound at time t ([B]t) divided by the concentration bound at equilibrium ([B]; t>180 min) was plotted versus time (min) and the
rate of adsorption was calculated by non-linear regression using equation 2, a one site exponential association model:

\[
\frac{[B_t]}{[B]} = (1 - e^{-kt})
\]  

(2)

where \([B_t]\) is the concentration of CBM2a bound (\(\mu\text{mol}\cdot g^{-1}\) cellulose) at elapsed time \(t\) (min), \([B]\) is the concentration of CBM2a bound at equilibrium (\(\mu\text{mol}\cdot g^{-1}\) cellulose), and \(k\) is the rate constant.

2.9.3 Kinetics of CBM2a exchange on BMCC

The rate at which CBM2a in solution is able to exchange with binding-modules bound to the surface of BMCC was assessed by a kinetic exchange experiment using two preparations of CBM2a, one labelled with Oregon Green, the other with AMCA-X. A number of samples were prepared in which CBM2a-AX was equilibrated with 1 mg of BMCC at a total concentration of 20 \(\mu\text{M}\), in a total volume of 1 ml. The cellulose was pelleted by centrifugation and 900 \(\mu\text{l}\) of the supernatant were removed. As with the previously described equilibrium exchange experiments, the amount of bound CBM2a-AX was calculated from the difference between the molar concentration of samples of equal volume incubated with and without cellulose as measured by AMCA-X fluorescence and interpolated using the appropriate standard. Each BMCC pellet was washed three times with 900 \(\mu\text{l}\) of 50 mM potassium phosphate buffer (pH 7.0) by centrifugation and subsequent removal of the supernatant. CBM2a-OG was added, at a final total concentration of 20 \(\mu\text{M}\), and incubated for different times at 4°C. Cellulose was separated from the supernatant by passing the suspension through a Qiaquick spin column. The fluorescence of the free and total samples of CBM2a-AX and CBM2a-OG was measured, the molar concentrations of each determined, and the amount of bound CBM2a-AX and CBM2a-OG was calculated.
2.10 NMR

2.10.1 Solid state NMR: Cross-polarization/magic angle spinning (CP/MAS)

Solid state MAS NMR experiments were performed on a Bruker DSX-400 spectrometer using a 4 mm Bruker triple-tuned probe with zirconia rotors and Kel-F caps. The spectrometer operated at frequencies of 400.13 MHz for $^1$H, 100.61 for $^{13}$C, 40.54 for $^{15}$N and 376.50 for $^{19}$F. Chemical shifts were referenced to tetramethylsilane (TMS) using Adamantane as a secondary reference for $^{13}$C spectra. The $^1$H$\rightarrow^{13}$C CP Hartmann-Hahn match condition was set up using Adamantane; the $^1$H$\rightarrow^{13}$C$\rightarrow^{15}$N match was set up using $^{13}$C,$^{15}$N-glycine. Unless otherwise specified, all spectra were collected at room temperature on wet samples (40%-60% water content) at a spinning rate of 5 kHz. $^{13}$C CP/MAS spectra of unlabelled cellulose ($^{13}$C, 1.1% natural abundance) were the sum of 40 000 scans, collected using a standard CP pulse sequence, using a 4.5 μs proton 90° pulse, a 1000 μs contact pulse and a 2 s delay between repetitions. $^{13}$C CP/MAS spectra of $^{13}$C-enriched cellulose samples were the sum of 400 scans, collected using a standard CP pulse sequence, using a 3.0 μs proton 90° pulse, a 1000 μs contact pulse and a 2 s delay between repetitions for the cellulose from 1,3-$^{13}$C-glycerol synthesis and a contact pulse of 2000 μs and 10 s delay between repetitions for the cellulose from the $^{13}$C-4-glucose synthesis. $^{15}$N-CBM2a CP/MAS spectra were the sum of 12 000 scans, collected using a standard CP pulse sequence, using a 5.25 μs proton 90° pulse, a 2000 μs contact pulse and a 5 s delay between repetitions. For the CP-drain experiments, spectra were the sum of 31 232 scans, using 6.0 μs proton 90° pulse, a 1000 μs contact pulse, a 1 s delay between pulses with a 10 ms drain time. No line broadening or resolution enhancement was applied during processing of the data for these CP experiments.
2.10.2 NMR in solution

NMR spectra were recorded at 27°C on a Varian Unity 600 MHz spectrometer equipped with a gradient triple resonance probe. $^1$H chemical shifts were reference to an internal standard of DSS at 0.00 pp, and $^{15}$N chemical shifts were referenced to external 2.9 M $^{15}$NH$_4$Cl in 1 M HCl at 24.93 ppm (71). This latter reference yields $^{15}$N chemical shifts 1.6 ppm greater than those obtained using liquid NH$_3$ (138). Spectra were processed an analyzed using FELIX software (Biosym Technologies; San Diego, CA).
3. Results

In the following section, I present data to show that CBM2a binds to the insoluble cellulose preparations PASC, BMCC, and Avicel. The residues involved in the binding of CBM2a to BMCC were investigated by site-directed mutation. Solvent-exposed tryptophans are shown to be essential determinants of binding affinity and, with one exception, the other residues that are prominent on the tryptophan-rich binding face are not essential determinants of binding affinity. Additionally, each tryptophan plays a different role in binding. CBM2a is shown to bind specifically to the crystalline regions of cellulose; it shares sites with other binding modules that bind to crystalline cellulose (Cel6A-CBM2a and CBM3) and does not share sites with binding modules with specificity for cello-oligosaccharides and amorphous cellulose (CBM17 and CBM4-1) or the reducing ends of cellulose chains (CBM9-2). CBM17 and CBM4-1 compete for some sites on amorphous cellulose (PASC), but there are sites that are uniquely recognized by each module. The binding of CBM2a was investigated in more detail. The rate of adsorption of CBM2a to PASC is much slower than to BMCC and appears to be limited by mass transfer. Binding of CBM2a is apparently irreversible; the desorption isotherm is not equivalent to the adsorption isotherm, yet at equilibrium, CBM2a bound to the surface of BMCC exchanges with CBM2a in the solution phase, a rapid process where the binding equilibrium is re-established in less than 30 seconds. In preliminary experiments of the samples and technique, it was found that without further refinement, specific interactions between $^{15}$N-CBM2a and $^{13}$C-enriched cellulose could not be detected by solid state NMR.

3.1 Production of His$_6$-CBM2a

His$_6$-CBM2a was produced by *E. coli* using the expression vector pTugKH6-IEGR-CBM2a. Most of it leaked from the periplasm into the culture supernatant. It was purified from the
supernatant by IMAC to >95% homogeneity, as estimated by Coomassie blue-stained SDS-PAGE gels (Figure 3.1). Yields routinely exceeded 100 mg of purified product per litre of culture.

Figure 3.1 Purification of His$_6$-CBM2a from concentrated _E. coli_ culture supernatant. Lane 1, MW standards (size indicated in kDa); lane 2, concentrated supernatant loaded onto the charged IMAC column; lane 3, column eluent; lanes 4, 5, 6, 7, and 8 are fractions eluted in IMAC binding buffer with step-wise increases of imidizole: 30 mM, 50 mM, 100 mM, 250 mM, and 500 mM respectively. Lane 9 is the fraction eluted with 100 mM EDTA.

### 3.2 Binding of CBM2a to Cellulose

Crystallinities, relative to a highly crystalline preparation from _Valonia ventricosa_, for the cellulose preparations used in this work are 0.76 for BMCC (68) and 0.50 for Avicel (139) as measured by X-ray diffraction techniques. The crystallinity index for PASC is not known, but is likely to be significantly less than that of _Valonia_ due to the acid-swelling process used in its preparation. $^{13}$C CP/MAS spectra of these celluloses (Figure 3.2) also confirm some of these structural features. The peaks that correspond to C-4 (81 – 93 ppm) are indicative of cellulose structure (2, 4, 32). C-4 atoms that are part of the internal, highly ordered crystalline structure give rise to a sharp downfield peak at approximately 89 ppm; C-4 atoms at the cellulose surface or part of amorphous structures (4), are less highly ordered and
produce the broader, less intense, upfield peak at 84 ppm. Based on relative areas of these two C-4 peaks, BMCC has a large proportion of internal C-4 atoms and PASC has a larger proportion of C-4 exposed to the solvent. The spectrum of Avicel appears to be a mixture of the other two spectra. Paradoxically, the bulk crystallinity of PASC appears very low, as evidenced by the predominant broad C-4 peak and the much less discernible downfield sharp peak from the crystalline bulk, yet CBM2a binds to PASC with high capacity.

The affinity of His₆-CBM2a for these insoluble cellulose preparations decreased in the order BMCC>PASC>Avicel (Table 3.1; Figure 3.3) and ranged from 3.2 x 10⁶ M⁻¹ to 1.0 x 10⁶ M⁻¹, suggesting preferential binding of CBM2a to crystalline cellulose. It should be noted that the affinities reported here are average values based on the mass of each cellulose sample, not the available surface area. Furthermore, it is assumed that binding can be validly described by a Langmuir-type binding expression. The binding properties of the wild-type CBM2a, produced by proteolytic cleavage and separation of the binding module from the intact enzyme (Xyn10A), on these substrates were indistinguishable from the His₆-CBM2a construct (data not shown). Hereafter, CBM2a indicates either CBM2a or His₆-CBM2a because the molecules are functionally indistinguishable. Since the hexahistidine tag does not affect binding, it was not removed from any of the wild-type and mutant polypeptide constructs.

**Table 3.1** Binding affinity and capacity for the adsorption of CBM2a to insoluble cellulose at 4°C in 50 mM potassium phosphate, pH 7.0.

<table>
<thead>
<tr>
<th>Insoluble cellulose</th>
<th>Affinity (Kₐ) (10⁶ M⁻¹)</th>
<th>Capacity ([N₀]) (μmol g⁻¹ cellulose)</th>
<th>ΔG (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMCC</td>
<td>3.2 ± 0.3</td>
<td>11.7 ± 0.3</td>
<td>-34.5 ± 0.2</td>
</tr>
<tr>
<td>PASC</td>
<td>1.5 ± 0.1</td>
<td>14.3 ± 0.4</td>
<td>-32.8 ± 0.2</td>
</tr>
<tr>
<td>Avicel</td>
<td>1.0 ± 0.1</td>
<td>2.4 ± 0.1</td>
<td>-31.8 ± 0.3</td>
</tr>
</tbody>
</table>
Based on isothermal titration calorimetry (ITC) data, highly crystalline BMCC preparations offer at least two classes of binding sites to CBM2a, both of relatively high affinity and characterized by binding constants ($K_a$) within the range reported here ($10^6$-10$^7$ M$^{-1}$) (25). Thus, possibly due to differences in the surface energies of its various crystal faces, crystalline cellulose appears to present a heterogeneous array of binding sites to CBM2a.

The $K_a$ values reported in table 3.1, however, assume that the cellulose surface is uniform and offers only a single class of binding site. This approach is justified by the fact that the shape of the cumulative isotherm generated by the depletion method (used here) is relatively insensitive to the energetics and occupancy of lower affinity/lower occupancy sites (54, 70) as opposed to that of the differential binding isotherm, provided by ITC. The $K_a$ regressed from the cumulative depletion isotherm using the Langmuir equation (Equation 1, section 2.7) is therefore an averaged value which reflects, in part, the relative contributions of each class of binding site on the cellulose surface.

Based on this mode of analysis, the lower $K_a$ values reported for binding of CBM2a to Avicel and PASC, compared with the affinity for BMCC, suggest that these forms of insoluble cellulose present a higher fraction of low affinity sites to CBM2a resulting in a lower $K_a$ calculated from the cumulative isotherm. The nature of these low-affinity sites, however, is unclear. They could reflect natural differences in the abundance of high and low energy cellulose crystal faces. They could also originate from the ability of CBM2a to bind both crystalline and amorphous morphologies of cellulose, with the affinity of the module for amorphous regions being somewhat weaker. The latter possibility was tested by competition isotherms measuring the binding of CBM2a to PASC in the presence and absence of CBMs.
that recognize only the amorphous regions of PASC, namely CBM4-1 and CBM17 (see section 3.4.6).
Figure 3.2 CP/MAS Spectra of insoluble cellulose: (A) Avicel, (B) BMCC, and (C) PASC. The horizontal bars above the spectrum of Avicel indicate the spectral ranges of the corresponding carbon atoms of the glucose monomer unit of cellulose and apply to BMCC and PASC as well.
Figure 3.3 Binding of CBM2a to insoluble cellulose PASC (O), BMCC (●) and Avicel (▲). Binding data (Table 3.1) is regressed using the Langmuir binding expression (Equation 1).
3.3 Binding of CBM2a mutants to BMCC

Using the synthetic CBM2a-encoding DNA (pTugK-H6-IEGR-CBM2a) as a template and appropriate primers (Table 2.2), fourteen mutants of CBM2a were constructed by two-primer PCR cassette mutagenesis (see Section 2.4), including conservative (phenylalanine and tyrosine) and non-conservative (alanine) mutants of each of the three tryptophan residues, W17, W54, and W72, on the binding face. Other residues on the binding face of CBM2a that have the potential to participate in hydrogen bonds with the cellulose surface were substituted individually with alanine (Figure 3.4). Since CBM2a binds well to BMCC, it was chosen as the matrix for comparison of the affinities of the wild-type and mutant CBM2a polypeptides. In a later section, it will be shown that CBM2a localizes to crystalline regions of cellulose.

3.3.1 Tryptophan mutants

Substitution of each of the surface tryptophan residues with tyrosine, phenylalanine or alanine reduced the affinity for BMCC, with alanine having a greater effect than the more conservative phenylalanine substitution (Table 3.2). The affinities of the tyrosine mutants were W72Y>W17Y>W54Y; those of the phenylalanine mutants were W72F>W17F>W54F; and those of the alanine mutants were W17A>W72A>W54A. The W54A mutant had the lowest affinity, two orders of magnitude lower than the wild-type, whereas the W72Y mutant had the highest affinity of the nine tryptophan mutants, about two-thirds of the wild type (Table 3.2). Most of the binding energy (ΔG) is provided by the tryptophan residues. The sum of the ΔΔG values (Table 3.2) of the tryptophan to alanine mutants is −21.2 kJ/mol and accountss for approximately 61% of the total binding energy of the wild-type binding module.
Table 3.2 Binding affinity of CBM2a tryptophan variants for BMCC at 4°C in 50 mM potassium phosphate, pH 7.0.

<table>
<thead>
<tr>
<th>CBM2a Variant</th>
<th>$K_a$ ($10^6$ M$^{-1}$)</th>
<th>$\Delta G$ (kJ/mol)</th>
<th>$\Delta G$ (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>His$_6$-CBM2a</td>
<td>3.2 ± 0.3</td>
<td>-34.5 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>W17F</td>
<td>0.6 ± 0.1</td>
<td>-30.8 ± 0.2</td>
<td>-3.7 ± 0.4</td>
</tr>
<tr>
<td>W17Y</td>
<td>1.6 ± 0.2</td>
<td>-32.3 ± 0.3</td>
<td>-2.2 ± 0.5</td>
</tr>
<tr>
<td>W17A</td>
<td>0.5 ± 0.1</td>
<td>-30.1 ± 0.6</td>
<td>-4.1 ± 0.8</td>
</tr>
<tr>
<td>W54F</td>
<td>0.36 ± 0.03</td>
<td>-29.5 ± 0.2</td>
<td>-5.0 ± 0.4</td>
</tr>
<tr>
<td>W54Y</td>
<td>0.34 ± 0.01</td>
<td>-28.7 ± 0.1</td>
<td>-5.8 ± 0.3</td>
</tr>
<tr>
<td>W54A</td>
<td>0.03 ± 0.02</td>
<td>-24.0 ± 3.0</td>
<td>-11.0 ± 3.0</td>
</tr>
<tr>
<td>W72F</td>
<td>1.7 ± 0.1</td>
<td>-33.1 ± 0.2</td>
<td>-1.4 ± 0.4</td>
</tr>
<tr>
<td>W72Y</td>
<td>2.3 ± 0.2</td>
<td>-33.7 ± 0.3</td>
<td>-0.8 ± 0.5</td>
</tr>
<tr>
<td>W72A</td>
<td>0.21 ± 0.01</td>
<td>-28.2 ± 0.1</td>
<td>-6.3 ± 0.3</td>
</tr>
</tbody>
</table>

3.3.2 Other mutants

Alanine was substituted for other residues on the binding face that could potentially form hydrogen bonds with cellulose. The mutation N87A reduced binding to half that of the native polypeptide; the N15A mutation reduced binding to one-fifth of native affinity. The mutations, N18 and Q52A reduced binding only slightly; the mutation Q83A did not affect binding (Table 3.3). The sum of the $\Delta G$ values of these mutants is $-6.6$ kJ/mol and, when added to the contribution made by the tryptophan to alanine mutants, is $-27.8$ kJ/mol, accounts for $80\%$ of the total binding energy. There are a few other residues on this binding face that contribute to the binding of CBM2a to cellulose; these could be discovered by continued site-directed mutation studies.

Table 3.3 Binding affinity of CBM2a variants for BMCC at 4°C in 50 mM potassium phosphate, pH 7.0.

<table>
<thead>
<tr>
<th>CBM2a Variant</th>
<th>$K_a$ ($10^6$ M$^{-1}$)</th>
<th>$\Delta G$ (kJ/mol)</th>
<th>$\Delta G$ (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>His$_6$-CBM2a</td>
<td>3.2 ± 0.3</td>
<td>-34.5 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>N15A</td>
<td>0.47 ± 0.02</td>
<td>-30.1 ± 0.2</td>
<td>-4.4 ± 0.4</td>
</tr>
<tr>
<td>N18A</td>
<td>2.9 ± 0.2</td>
<td>-34.3 ± 0.2</td>
<td>-0.2 ± 0.4</td>
</tr>
<tr>
<td>Q52A</td>
<td>2.7 ± 0.2</td>
<td>-34.1 ± 0.2</td>
<td>-0.3 ± 0.4</td>
</tr>
<tr>
<td>Q83A</td>
<td>3.4 ± 0.3</td>
<td>-34.6 ± 0.2</td>
<td>0.1 ± 0.4</td>
</tr>
<tr>
<td>N87A</td>
<td>1.6 ± 0.1</td>
<td>-32.8 ± 0.2</td>
<td>-1.7 ± 0.4</td>
</tr>
</tbody>
</table>
Figure 3.4 Two views of CBM2a. The mutated residues are rendered using balls and sticks and are highlighted; surface tryptophans are green and the other polar residues are red. This figure was created with MolMol (63) based on the NMR solution structure (141).
3.4 Competition Binding Isotherms

The binding specificities of many CBMs are known. Types B and C bind to soluble ligands; type A CBMs bind to insoluble cellulose and to PASC, but it is not precisely known if this type also binds to amorphous cellulose and soluble oligosaccharides. Using the complex substrate PASC, that binds all types of CBM (Table 3.4) and has crystalline and amorphous regions, binding competition experiments, involving two CBMs at a time, were performed to determine whether CBMs from the three types compete for binding sites.

3.4.1 CBMs bind to PASC and BMCC

CBMs from each of the three types were used in competition experiments. Type A CBMs were CBM2a, Cel6A-CBM2a and CBM3; type B CBMs were CBM4-1 and CBM17; type C was CBM9-2. All of the CBMs used in the competition binding experiments bound to PASC (Table 3.4); type A CBMs (CBM2a, Cel6A-CBM2a and CBM3) also bind to BMCC (Table 3.4). Binding capacity of PASC for the CBMs varied from 6.1 μmol·g⁻¹ for CBM4-1 to 24.0 μmol·g⁻¹ for CBM17. It should be noted however that type A CBMs had a similar capacity for PASC (~15 μmol·g⁻¹) and a similar capacity for BMCC (~11 μmol·g⁻¹). The CBM binding parameters reported in Table 3.4 were each determined by a depletion isotherm, as described in section 2.7. The concentration of each CBM was calculated from its $A_{280}$ (corrected for light scattering) and molar extinction coefficient (Table 2.3).
Table 3.4 CBM binding affinity and capacity for BMCC and PASC at 4°C in 50 mM potassium phosphate, pH 7.0.

<table>
<thead>
<tr>
<th>CBM</th>
<th>PASC</th>
<th>BMCC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Affinity ((K_a) \times 10^6 M^{-1})</td>
<td>Capacity ([N_0] (\mu mol \cdot g^{-1} cellulose))</td>
</tr>
<tr>
<td>CBM2a</td>
<td>1.5 (± 0.1)</td>
<td>14.3 (± 0.4)</td>
</tr>
<tr>
<td>Cel6A-CBM2a</td>
<td>0.9 (± 0.3)</td>
<td>14.9 (± 1.1)</td>
</tr>
<tr>
<td>CBM3</td>
<td>1.1 (± 0.2)</td>
<td>17.4 (± 0.8)</td>
</tr>
<tr>
<td>CBM4-1</td>
<td>0.25 (± 0.02)</td>
<td>6.1 (± 0.1)</td>
</tr>
<tr>
<td>CBM17</td>
<td>1.1 (± 0.1)</td>
<td>24.0 (± 0.8)</td>
</tr>
<tr>
<td>CBM9-2</td>
<td>0.63 (± 0.04)</td>
<td>8.4 (± 0.1)</td>
</tr>
</tbody>
</table>

3.4.2 Labelling of CBMs

Five to ten milligrams of each CBM used in the competition binding experiments were labelled with either AMCA-X or Oregon Green®. These probes react with primary amines, the N-terminus or lysine residues, of polypeptides. Each binding module therefore will have at least one potential labelling site (Table 3.5). The labelling efficiency of each reaction was calculated based on the extinction coefficients at their maximum absorption wavelength for each label (18 500 M\(^{-1}\)·cm\(^{-1}\) at 352 nm for AMCA-X and 86 200 M\(^{-1}\)·cm\(^{-1}\) at 506nm for Oregon Green®; Molecular Probes). These two fluorescent probes were chosen because they have well separated excitation and emission properties (Figure 3.5). Labelling efficiency ranged from approximately 0.42 to 1.21 moles of fluorescent probe per mole of polypeptide.

In competition experiments, due to the sensitivity of the fluorimeter, each labelled CBM was diluted with unlabelled CBM so that maximum fluorescence measured (~1000 fluorescence units) would coincide with the maximum protein concentration used in a binding isotherm, typically 30 μM. In other words, labelled CBM was diluted in unlabelled CBM so that every CBM sample fell roughly into the same reading range of the fluorimeter.
### Table 3.5 CBMs, labels and number of potential labelling sites for CBMs used in competition binding experiments.

<table>
<thead>
<tr>
<th>CBM</th>
<th>Number of potential labelling sites</th>
<th>Label</th>
<th>Extent of Labelling (Molar ratio Label/CBM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBM2a</td>
<td>2</td>
<td>Oregon Green</td>
<td>0.42</td>
</tr>
<tr>
<td>CBM2a</td>
<td>2</td>
<td>AMCA-X</td>
<td>1.21</td>
</tr>
<tr>
<td>CBM4-1</td>
<td>1</td>
<td>AMCA-X</td>
<td>0.60</td>
</tr>
<tr>
<td>Cel6A-CBM2a</td>
<td>2</td>
<td>AMCA-X</td>
<td>0.65</td>
</tr>
<tr>
<td>CBM17</td>
<td>9</td>
<td>Oregon Green</td>
<td>0.65</td>
</tr>
<tr>
<td>CBM3</td>
<td>11</td>
<td>not labelled</td>
<td>-</td>
</tr>
<tr>
<td>CBM9-2*</td>
<td>16</td>
<td>AMCA-X</td>
<td>0.78</td>
</tr>
</tbody>
</table>

*CBM9-2 labelled with AMCA-X did not bind to cellulose

### 3.4.3 Standards and Isotherms by fluorescence

A standard or calibration curve for each of the labelled CBM preparations and each cellulose substrate used in the competition experiments was created so that the concentration of each CBM could be measured by fluorescence. To generate each curve, a binding experiment was performed for each labelled CBM and cellulose substrate, using untreated microcentrifuge tubes (ie. without preadsorption with BSA). A control sample containing CBM but no cellulose was also prepared in an identical microcentrifuge tubes. The fluorescence and net A280 of each sample, including the control, was measured; the molar concentration protein in the solution phase of each sample was then used to construct calibration curves for both cellulose-containing and cellulose-free samples (PASC, Figure 3.6; BMCC, Figure 3.7). There was a linear relationship between fluorescence and the solution concentration of free or total CBM samples. Depletion-type isotherms were then constructed from fluorescence data. The free-protein concentration was determined by using the calibration curve constructed in the presence of cellulose; total-protein concentration was determined by using the calibration curve constructed in the absence of cellulose. Each of the two calibration curves were linear, but the calibration curve for free-protein did not fall on the same line as
the total-protein standard. Unless otherwise specified, a CBM preparation (*e.g.* CBM2a-OG) contains a mixture of labelled and unlabelled CBM.

In previous work, it was shown that the binding of CBM2a to insoluble cellulose is not affected by the presence of FITC (57). Similarly, the Oregon Green and AMCA-X labels should not perturb binding of CBM2a. This was verified by determining that the Langmuir binding constants for labelled and unlabelled CBM2a preparations, where the concentration was calculated by UV absorbance (net A$_{280}$), in binding to PASC were essentially equivalent (Fig 3.8-A; Table 3.6). The effect of competition between the specific binding of CBM2a to cellulose and the non-specific binding of protein to cellulose was assessed by determining the binding constant of CBM2a in the presence of approximately 20 µM BSA. The binding of CBM2a was not affected (see section 3.4.4; Table 3.6). It was observed that CBM2a bound, non-specifically, to the sides of the microcentrifuge tubes. To minimize the effect of non-specific adsorption of CBMs affecting competition results, BSA was used to block the tubes used in competition experiments. The BSA pre-treatment did not affect the CBM; binding parameters of CBM2a measured by fluorescence (Table 3.6) using treated tubes was essentially equivalent to those calculated for CBM2a isotherms performed in untreated tubes, measured by UV absorbance.

For the other CBMs used in the competition experiments (Table 3.5) the residues on each of the polypeptides that may potentially react with the fluorescent probes appeared well removed from the putative ligand binding sites highlighted in figures 1.4, 1.5 and 1.6. In experiments similar to the one described above, the binding of CBM2a, Cel6A-CBM2a, CBM4-1, or CBM17 to cellulose was not affected by the presence of a fluorescent probe (data not shown). The labelling of CBM9-2 with AMCA-X blocked its binding to PASC. Competition experiments involving CBM9-2 were therefore performed using unlabelled
polypeptide. Because of solubility problems with the protein preparation, unlabelled
CBM3 was also used unlabelled.

3.4.4 CBM2a adsorption to PASC in the presence of BSA

Microcentrifuge tubes used in competition experiments were treated with BSA to minimize
the amount of non-specific binding of CBMs to the walls of tubes. The binding of CBM2a-
OG was not affected by the presence of BSA (Figure 3.8-B; Table 3.6). The association
affinity constants of CBM2a alone and in the presence of 20 μM BSA, as measured by
fluorescence were similar and within the standard error of regression. Additionally, the 20
μmol-g\(^{-1}\) capacity ([N\(_0\)]) of PASC to bind CBM2a-OG was not affected by the presence of
BSA. Since the non-specific adsorption of BSA to the microcentrifuge tubes, or the cellulose,
did not affect the specific adsorption of CBM2a to cellulose, the BSA was used to block the
microcentrifuge tubes in the subsequent competition experiments.

Table 3.6 Binding of CBM2a to PASC: UV absorbance, fluorescence and BSA competition
at 4°C in 50 mM potassium phosphate, pH 7.0.

<table>
<thead>
<tr>
<th></th>
<th>CBM2a Affinity (K(_a): 10(^6) M(^{-1}))</th>
<th>Capacity ([N(_0]): μmol·g(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>UV Absorbance (unlabelled)</td>
<td>1.5 (± 0.1)</td>
<td>14.3 (± 0.4)*</td>
</tr>
<tr>
<td>UV Absorbance (CBM2a-OG)</td>
<td>1.6 (± 1.0)</td>
<td>22.0 (± 3.0)</td>
</tr>
<tr>
<td>Fluorescence</td>
<td>2.3 (± 0.4)</td>
<td>20.0 (± 1.0)</td>
</tr>
<tr>
<td>Fluorescence with 20 μM BSA</td>
<td>2.2 (± 0.4)</td>
<td>20.8 (± 0.9)</td>
</tr>
</tbody>
</table>

*The capacity measured for CBM2a unlabelled was for a preparation of PASC different from that used for the
CBM2a-OG experiments listed. PASC preparations differ in total available surface area, therefore only [N\(_0\)] is
affected.
Figure 3.5 Emission spectra of CBM2a labelled with Oregon Green® or AMCA-X. CBM2a-OG was excited at 506 nm, emission was measured from 510 nm to 600 nm using a 530 nm emission filter; CBM2a-AX was excited at 352 nm, emission was measured from 360 nm to 600 nm. For comparison of the two labels, the fluorescence emission was normalized to the maximum fluorescence value measured for each label.
Figure 3.6 Standard curves of fluorescence vs protein concentration (µM) for each of the preparations of CBMs used in the PASC binding competition experiments (clockwise from top left): CBM2a-OG, CBM2a-AX, CBM4-1-AX and CBM17-OG. For each CBM, two calibration curves were plotted: fluorescence vs. CBM concentration for the total protein samples (incubated without cellulose) (O); fluorescence vs. CBM concentration for the free protein samples (incubated with cellulose) (●). Free CBM concentration was calculated using the calibration curve constructed using free protein samples; total CBM concentration was calculated using the calibration curve constructed using total protein samples.
Figure 3.7 Standard curves of fluorescence vs. protein concentration (μM) for CBM2a-OG (top left), and CBM2a-AX (top right) and Cel6A-CBM2a-AX (lower right) preparations used in the BMCC binding competition experiments. Standards of free protein (●) and total polypeptide (○) were considered separately.
Figure 3.8 Binding of CBM2a-OG to PASC. (A) CBM2a-OG adsorption to PASC measured by UV absorbance (●) and by fluorescence (○). (B) CBM2a-OG adsorption to PASC alone (●) and in the presence of 20 μM BSA.
3.4.5 A model of competition binding

Knowing the binding parameters of two binding modules competing for a single class of binding sites on a sorbant surface, the extent of competition can be predicted using derivatives of the Langmuir-like binding model. When two binding modules compete for the same binding site, it is the apparent affinity of each that is affected. For example, in an hypothetical competition binding experiment involving two CBMs, called CBM-1 and CBM-2, have an affinity, $K_{a1}$ and $K_{a2}$, respectively, for the same binding site (capacity $[N_0]$), then the apparent $K_{a1}$ ($K_{a1\text{apparent}}$) would be lower in the presence of CBM-2 than the $K_{a1}$ measured in the absence of CBM-2. The capacity ($[N_0]$) is the sum of the concentration of unoccupied binding sites ($[N]$) and the concentration of sites occupied by CBM-1 is $[B_1]$ the concentration of sites occupied CBM-2 is $[B_2]$.

$$[N_0] = [N] + [B_1] + [B_2]$$  \hspace{1cm} (3)

The affinity of CBM-1 ($K_{a1}$) is described by:

$$K_{a1} = \frac{[B_1]}{[F_1][N]}$$  \hspace{1cm} (4)

substituting equation 3 into equation 4, gives:

$$K_{a1} = \frac{[B_1]}{[F_1]([N_0] - [B_1] - [B_2])}$$  \hspace{1cm} (5)

By definition, $[B_1] = \frac{[N_0]K_{a1\text{apparent}}[F_1]}{(1 + K_{a1}[F_1])}$

$$\Rightarrow [B_1] = \frac{[N_0]K_{a1\text{apparent}}[F_1]}{(1 + K_{a1}[F_1])}$$  \hspace{1cm} (6)

At low concentrations of CBM-1 (i.e. $[F_1]$→0)

$$K_{a1\text{apparent}} = K_{a1}\left(1 - \frac{[B_2]}{[N_0]}\right)$$  \hspace{1cm} (8)
In the presence of a large concentration of CBM-2, CBM-1 will appear to bind with a lower apparent affinity ($K_{a\text{Apparent}}$). Similarly, if the $K_{a\text{Apparent}}$ of a CBM is approximately equal to its $K_a$ in the absence of a competitor, then the two CBMs do not compete for the same binding sites. Thus, changes in $K_a$ are a sensitive indicator of binding competition. An effective plot to aid interpretation of binding competition is the fraction of saturation ($[B_2]/[N_0]$) of a CBM vs. the fraction of saturation of the second CBM ($[B_1]/[N_0]$). At a $[B_1]/[N_0]$ equal to 1, the surface is fully saturated with CBM-1; at $[B_1]/[N_0]$ equal to 0, no CBM-1 is bound. If two CBMs bind to separate sites, and do not compete, this plot is a horizontal line, with a slope of zero; CBM-2 bound to the surface is not affected by the binding of CBM-1 and, therefore, the CBMs are recognizing different sites. In the case where CBM-1 and CBM-2 compete for exactly the same sites, then the plot has a negative slope, with a x-intercept of 1; as the concentration of CBM-1 approaches saturation, CBM-2 is displaced and the fraction of CBM-2 bound approaches zero. The actual slope of the $[B_2]/[N_0]$ vs. $[B_1]/[N_0]$ plot depends on the starting concentration of the competitor (CBM-2 in this example) and the percentage of binding sites common to both CBMs.

To validate this model of competition, two binding experiments were performed involving the competition between two labelled CBM2a preparations (CBM2a-AX and CBM2a-OG) on BMCC (Figures 3.9 and 3.10). In this case, the fraction of saturation of CBM2a-AX ($[B]/[N_0]$) vs. the fraction of saturation of CBM2a-OG ($[B]/[N_0]$) is plotted (Figures 3.9-B and 3.10-B); the competitor (CBM2a-AX at constant concentration of either 11 µM or 15 µM) is plotted on the Y-axis. In both cases, where the competitor CBM2a-AX concentration was held constant at either 11 µM (Figure 3.9) or 15 µM (Figure 3.10), as CBM2a-OG approached saturation (CBM2a-OG $[B]/[N_0]$ approaches 1), the fraction of CBM2a-AX
bound to the surface of BMCC (CBM2a-AX [B]/[N₀]) approached zero, as would be expected.

At high concentrations of CBM2a-OG ([B_{CBM2a-OG}]/[N₀] ~0.5) the data depart slightly from the line predicted by ideal competition conditions (shown by a dashed line in Figures 3.9-B and 3.10-B). At these concentrations of CBM2a-OG, the total concentration of CBM2a (CBM2a-AX plus CBM2a-OG) is virtually at saturation (>40 μM) and secondary binding effects, such as protein-protein layering, are likely occurring and could account for the deviation from the predicted competition. Additionally, the capacity of CBM2a-OG, when in competition with CBM2a-AX, is expected to eventually reach the [N₀] of CBM2a-OG that is observed without competition (Figures 3.9-A and 3.10-A); this is not seen at the range of CBM2a-OG concentrations tested when in competition.

The apparent affinity of CBM2a-OG can be predicted using this theory. At 11 μM, bound CBM2a-AX is at ~82% saturation; at 15 μM, the BMCC substrate is ~95% saturated. The apparent affinity of CBM2a-OG in competition with CBM2a-AX, predicted by the fraction of saturation (equation 8), is 0.78 x 10⁶ M⁻¹ and 0.22 x 10⁶ M⁻¹, respectively, and agrees remarkably well with the observed values (Table 3.7).

**Table 3.7 CBM2a-OG in competition with CBM2a-AX on BMCC at 4°C in 50 mM potassium phosphate, pH 7.0.**

<table>
<thead>
<tr>
<th>CBM</th>
<th>Competitor</th>
<th>CBM2a-OG Kₐ apparent (10⁶ M⁻¹)</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBM2a-OG³</td>
<td>-</td>
<td>4.3 (± 0.5)</td>
<td>3.9-A</td>
</tr>
<tr>
<td>CBM2a-OG³</td>
<td>CBM2a-AX (11 μM)</td>
<td>0.72 (± 0.05)</td>
<td>3.9-A</td>
</tr>
<tr>
<td>CBM2a-OG³</td>
<td>CBM2a-AX (15 μM)</td>
<td>0.45 (± 0.03)</td>
<td>3.10-A</td>
</tr>
</tbody>
</table>
Figure 3.9 Binding of CBM2a-OG to BMCC in the presence of CBM2a-AX. A) Binding of CBM2a-OG alone (●) and in the presence of 11 μM CBM2a-AX (○); CBM2a-AX bound vs. CBM2a-OG free is also plotted (▲). B) Fraction bound CBM2a-AX ([B]/[N₀]) vs. fraction bound CBM2a-OG ([B]/[N₀]); solid line is the linear regression of the data, the dashed line is the theoretical regression assuming CBM2a-AX saturation of 0.82 and ideal competition between CBM2a-AX and CBM2a-OG
Figure 3.10 Binding of CBM2a-OG to BMCC in the presence of CBM2a-AX. A) Binding of CBM2a-OG alone (●) and in the presence of 15 μM CBM2a-AX (○); CBM2a-AX bound vs. CBM2a-OG free is also plotted (▲). B) Fraction bound CBM2a-AX ([B]/[N₀]) vs. fraction bound CBM2a-OG ([B]/[N₀]); solid line is the linear regression of the data, the dashed line is the theoretical regression assuming CBM2a-AX saturation of 0.95 and ideal competition between CBM2a-AX and CBM2a-OG.
3.4.6 CBM2a in competition with CBM4-1, CBM17 and CBM9-2

CBM4-1 binds specifically to amorphous cellulose, showing no significant affinity for crystalline preparations of cellulose such as V. ventricosa cellulose and BMCC (23, 64, 125). The $K_a$ for CBM4-1 binding to either soluble cellulose polymers or PASC at 4°C (50 mM potassium phosphate buffer, pH 7.0) is $2.5 \times 10^5$ M$^{-1}$ (Table 3.4), an order of magnitude less than the average $K_a$ for CBM2a binding to PASC. CBM17 has specificity that is similar to CBM4-1; it binds cello-oligosaccharides or amorphous cellulose with a maximum affinity for PASC of $7.6 \times 10^5$ M$^{-1}$ (Table 3.4). Competition experiments should determine whether the family 2a and the family 4 and 17 CBMs share the same sites on cellulose. Thus, in a competition experiment at fixed total concentration of CBM4-1, increasing concentrations of CBM2a should lead to the progressive exclusion of bound CBM4-1 if CBM2a shows an affinity for amorphous regions of PASC. Similarly, for a fixed total CBM2a concentration, increasing concentrations of CBM4-1 or CBM17 should lead to a progressive displacement of CBM2a.

Binding of CBM2a to PASC was not significantly affected by the presence of CBM4-1 (Figure 3.11; Table 3.8). Alone, CBM2a binds to PASC with a $K_a$ of $1.2 \pm 0.1 \times 10^6$ M$^{-1}$. When CBM4-1 is present at a total concentration of 19 µM (an amount that is approximately 95% of CBM4-1 saturation), the apparent $K_a$ for binding of CBM2a to PASC is reduced only slightly to $1.0 \pm 0.1 \times 10^6$ M$^{-1}$; in the presence of 14.8 µM CBM4-1, approximately 80% of saturation, the apparent $K_a$ of CBM2a was again only slightly lower than its corresponding CBM4-1 free value (Table 3.8). The amount of bound CBM4-1 decreased no more than 10% (Figure 3.11-C) as CBM2a approached saturation, indicating that the vast majority of bound CBM4-1 is unaffected by the presence of CBM2a.
The binding of neither CBM4-1 nor CBM17 was affected by the presence of CBM2a. The affinity of CBM4-1 was 0.16 (± 0.02) x 10^6 M^-1 and did not change in the presence of CBM2a (Figure 3.12; Table 3.8). As CBM4-1 approached saturation, only a very small population of bound CBM2a (<5%) was displaced by the binding of CBM4-1 ([Figure 3.12-B]). The affinity of CBM17 decreased insignificantly from 0.76 (± 0.08) x 10^5 M^-1 to 0.67 (± 0.04) x 10^5 M^-1 in the presence of CBM2a at approximately 85% of saturation ([Figure 3.13; Table 3.8). At a high concentration of CBM17 (near saturation) only about 10% of the bound CBM2a was displaced by the binding of CBM17 (Figure 3.13-B). These results suggest that CBM2a occupies separate sites distinct from those occupied by CBM4-1 and CBM17, with a possible 5-10% overlap.

CBM9-2 binds specifically to the reducing ends of sugars (A. B. Boraston et al.; V. Notenboom et al., in press); its binding site accommodates only two sugar rings, such as cellobiose. It binds to PASC with an affinity of 0.63 (± 0.04) x 10^6 M^-1 at a sorbant capacity of 8.4 (± 0.1) μmol·g^-1. The binding of CBM2a was not affected by the presence of CBM9-2 (Figure 3.14; Table 3.8) at a lower saturation; in the presence of more CBM9-2, both the affinity and binding capacity of CBM2a-OG decreased only slightly (Figure 3.14; Table 3.8). Together, these results suggest that CBM2a binding is specific to crystalline regions of PASC. CBM2a does not share binding sites with binding modules with specificity for cello-oligosaccharides or amorphous cellulose (CBM4-1 and CBM17); it also does not share binding sites with CBM9-2, which binds to the reducing ends of cellulose polymers. The small amount of competition between bound CBM2a and these other CBMs presumably occurs only at boundaries between crystalline and amorphous microstructures or reducing ends of cellulose molecules.
Table 3.8 PASC competition binding summary I: CBM2a, CBM4-1, CBM17 and CBM9-2 at 4°C in 50 mM potassium phosphate, pH 7.0.

<table>
<thead>
<tr>
<th>CBM</th>
<th>Competitor</th>
<th>$K_a$ Apparent of CBM $(10^6 \text{M}^{-1})$</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBM2a-OG</td>
<td>(none)</td>
<td>1.2 (± 0.1)</td>
<td>3.11</td>
</tr>
<tr>
<td>CBM2a-OG</td>
<td>CBM4-1-AX (19 µM)</td>
<td>1.0 (± 0.1)</td>
<td>3.11-A</td>
</tr>
<tr>
<td>CBM2a-OG</td>
<td>CBM4-1-AX (14.8 µM)</td>
<td>0.96 (± 0.04)</td>
<td>3.11-B</td>
</tr>
<tr>
<td>CBM4-1-AX</td>
<td>(none)</td>
<td>0.16 (± 0.02)</td>
<td>3.12</td>
</tr>
<tr>
<td>CBM4-1-AX</td>
<td>CBM2a-OG (18 µM)</td>
<td>0.16 (± 0.02)</td>
<td>3.12</td>
</tr>
<tr>
<td>CBM17-OG</td>
<td>(none)</td>
<td>0.76 (± 0.08)</td>
<td>3.13</td>
</tr>
<tr>
<td>CBM17-OG</td>
<td>CBM2a-AX (18.8 µM)</td>
<td>0.67 (± 0.04)</td>
<td>3.13</td>
</tr>
<tr>
<td>CBM2a-OG</td>
<td>CBM9-2 (8 µM)</td>
<td>0.98 (± 0.09)</td>
<td>3.14</td>
</tr>
<tr>
<td>CBM2a-OG*</td>
<td>(none)</td>
<td>2.3 (± 0.4)</td>
<td>3.14-A</td>
</tr>
<tr>
<td>CBM2a-OG*</td>
<td>CBM9-2 (18.6 µM)</td>
<td>1.7 (± 0.2)</td>
<td>3.14-B</td>
</tr>
</tbody>
</table>

*a second preparation of CBM2a-OG
Figure 3.11 Binding of CBM2a-OG to PASC in the presence of CBM4-1. (A) CBM2a-OG adsorption to PASC alone (●) and in the presence of 95% saturating concentration of CBM4-1 (○) where the fraction of CBM4-1-AX bound vs. CBM2a-OG Free (▲) is also plotted. (B) CBM2a-OG adsorption to PASC alone (●) and in the presence of 80% saturating concentration of CBM4-1 (○) where the fraction of CBM4-1-AX bound vs. CBM2a-OG Free (▲) is also plotted. (C) Fraction CBM4-1 bound ([B]/[N]) vs. Fraction CBM2a-OG bound for both saturation concentrations of CBM4-1-AX: 95% (●), and 80% (○).
Figure 3.12 Binding of CBM4-1-AX to PASC in the presence of CBM2a-OG. A) CBM4-1-AX adsorption to PASC alone (●) and in the presence of CBM2a-OG (○) at 65% of saturation concentration. B) Fraction CBM2a-OG bound vs. fraction CBM4-1-AX bound (●).
Figure 3.13 Binding of CBM17-OG to in the presence of CBM2a. CBM17 binding to PASC alone (●) and in the presence of CBM2a (○) at approximately 85% of saturation; CBM2a-AX bound vs. CBM17-OG Free(▲) is also plotted. B) Fraction CBM2a-AX bound ([B]/[N₀]) vs. Fraction CBM17-OG bound.
Figure 3.14 Binding of CBM2a-OG in the presence of CBM9-2. A) CBM2a binding to PASC alone (●) and in the presence of 8 μM CBM9-2 (○). B) CBM2a binding to PASC alone and in the presence of 18.6 μM CBM9-2. These two competition experiments were done using two separate preparations of CBM2a-OG, hence the slight differences in affinity of CBM2a-OG alone.
3.4.7 CBM2a in competition with Cel6A-CBM2a and CBM3

The binding module from *C. cellulovorans* cellulose-binding protein A (CcCbpA-CBM3, hereafter CBM3), binds to insoluble cellulose (40, 41). A crystal structure has been solved for the related *C. thermocellum* scaffoldin subunit CtCipC-CBM3 (130); the CBM3 amino acid sequence is very similar to that of CtCipC-CBM3 (130), has similar binding characteristics, and presumably a similar structure. CtCipC-CBM3 has an array of exposed aromatic residues characteristic of CBMs that bind to crystalline cellulose. The affinity of CBM3 for PASC and BMCC is very similar to that of CBM2a ($\sim 10^6 \text{M}^{-1}$). With CBM3 present as a competitor (not labelled) at a constant total concentration of 15 μM (approximately 72% of saturation) the apparent affinity of CBM2a-OG for PASC is significantly decreased. This demonstrates that there are sites shared by both CBM3 and CBM2a (Figure 3.15; Table 3.9). In this experiment, the affinity of CBM2a-OG was slightly lower than that measured in other experiments likely due to incomplete separation of free Oregon Green from the CBM2a-OG conjugates in the preparation of the labelled CBM2a-OG. However, the free fluorescent probe would not account for the apparent reduction in affinity of CBM2a observed in the presence of CBM3.

Cel6A-CBM2a (previously referred to as CBD$_{CelA}$) and Xyn10A-CBM2a (CBM2a) share similar binding properties and apparent binding specificity for cellulose with a significant degree of crystallinity (27, 37). Xyn10A-CBM2a and Cel6A-CBM2a share 52% amino acid sequence identity with 67% aa sequence similarity, including the three surface tryptophan residues implicated in the binding of family 2a CBMs to insoluble cellulose (12, 27, 94). These two family 2a binding modules bind similarly to the insoluble cellulose preparations PASC and BMCC (Table 3.1 and Table 3.4) with an affinity for PASC of approximately $1 \times 10^6 \text{M}^{-1}$ and an affinity for BMCC of approximately $3 \times 10^6 \text{M}^{-1}$. In competition experiments
using a constant amount of Cel6A-CBM2a-AX as a competitor, the apparent affinity of
CBM2a-OG for both PASC (Figure 3.16) and BMCC (3.17) was significantly reduced (Table
3.9). A majority of Cel6A-CBM2a bound to the cellulose surfaces directly competes with
CBM2a (Figure 3.16-B and Figure 3.17-B); as CBM2a approached saturation, the fraction of
bound Cel6A-CBM2a decreased to approximately 10%.

Table 3.9 Competition binding summary II: CBM2a, Cel6A-CBM2a and CBM3 at 4°C in 50
mM potassium phosphate, pH 7.0.

<table>
<thead>
<tr>
<th>CBM*</th>
<th>Competitor</th>
<th>Cellulose</th>
<th>CBM2a $K_a$ apparent $(10^6 M^{-1})$</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBM2a-OG$^1$</td>
<td>(none)</td>
<td>PASC</td>
<td>0.50 (± 0.02)</td>
<td>3.15</td>
</tr>
<tr>
<td>CBM2a-OG$^1$</td>
<td>CBM3 (15 μM)</td>
<td>PASC</td>
<td>0.17 (± 0.02)</td>
<td>3.15</td>
</tr>
<tr>
<td>CBM2a-OG$^2$</td>
<td>(none)</td>
<td>PASC</td>
<td>1.2 (± 0.1)</td>
<td>3.16</td>
</tr>
<tr>
<td>CBM2a-OG$^2$</td>
<td>Cel6A-CBM2a-AX (15.8 μM)</td>
<td>PASC</td>
<td>0.39 (± 0.03)</td>
<td>3.16</td>
</tr>
<tr>
<td>CBM2a-OG$^3$</td>
<td>(none)</td>
<td>BMCC</td>
<td>4.3 (± 0.5)</td>
<td>3.17</td>
</tr>
<tr>
<td>CBM2a-OG$^3$</td>
<td>Cel6A-CBM2a-AX (12.5 μM)</td>
<td>BMCC</td>
<td>1.2 (± 0.1)</td>
<td>3.17</td>
</tr>
</tbody>
</table>

*superscript number delineates the CBM2a-OG preparation used in the competition experiment
Figure 3.15 Binding of CBM2a-OG to PASC in the presence of CBM3. Binding of CBM2a alone (●) and in the presence of 15 μM CBM3 (○).
Figure 3.16 Binding of CBM2a-OG to PASC in the presence of Cel6A-CBM2a-AX. A) Binding of CBM2a-OG alone (●) and in the presence of Cel6A-CBM2a (○); Cel6A-CBM2a-AX bound vs. CBM2a-OG free is also plotted (▲). B) Fraction bound CBM2a-AX ([B]/[N_0]) vs. fraction bound CBM2a-OG ([B]/[N_0]).
Figure 3.17 Binding of CBM2a-OG to BMCC in the presence of Cel6A-CBM2a. A) Binding of CBM2a-OG alone (●) and in the presence of Cel6A-CBM2a (○); Cel6A-CBM2a-AX bound vs. CBM2a-OG free is also plotted (▲). B) Fraction bound CBM2a-AX ([B]/[N₀]) vs. fraction bound CBM2a-OG ([B]/[N₀]).
3.4.8 CBM4-1 and CBM17 competition

As mentioned previously, CBM4-1 and CBM17 have similar substrate specificity binding to cello-oligosaccharides, greater than 5 or 6 glucose units in length and to amorphous cellulose (8, 61, 125, 127); CBM17, however, binds to PASC with approximately 4-fold higher affinity than CBM4-1 (Table 3.4).

In competition experiments using a constant amount of CBM4-1-AX as a competitor, the apparent affinity of CBM17-OG for PASC was reduced (Table 3.10; Figure 3.18-A). A significant number of the sites bound by CBM17-OG are also bound by CBM4-1-AX. However, not all of the CBM4-1-AX was displaced as the concentration of CBM17 increased (Figure 3.18-B). In the complementary experiment, the apparent affinity of CBM4-1-AX was reduced in the presence of a constant amount of CBM17-OG (Table 3.10, Figure 3.19). Only a small fraction of the CBM17-OG molecules bound were displaced as the CBM4-1 approached saturation; a majority of the sites that are bound by CBM17-OG are not recognized by CBM4-1-AX. The amorphous regions of PASC are therefore heterogeneous, composed of sites recognized by only CBM4-1, CBM17 or both CBMs.

<table>
<thead>
<tr>
<th>CBM</th>
<th>Competitor</th>
<th>$K_a^{apparent}$ ($10^9$ M$^{-1}$)</th>
<th>$[N_0]$ (µmol·g$^{-1}$ cellulose)</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBM17-OG</td>
<td>(none)</td>
<td>0.81 (± 0.05)</td>
<td>27.8 (± 0.6)</td>
<td>3.18</td>
</tr>
<tr>
<td>CBM17-OG</td>
<td>CBM4-1-AX</td>
<td>0.31 (± 0.01)</td>
<td>27.8 (± 0.6)</td>
<td>3.18</td>
</tr>
<tr>
<td>CBM4-1-AX</td>
<td>(none)</td>
<td>0.24 (± 0.05)</td>
<td>7.1 (± 0.4)</td>
<td>3.19</td>
</tr>
<tr>
<td>CBM4-1-AX</td>
<td>CBM17-OG</td>
<td>0.003 (± 0.007)</td>
<td>7.1 (± 0.4)</td>
<td>3.19</td>
</tr>
</tbody>
</table>
Figure 3.18 Binding of CBM17-OG to PASC in the presence of CBM4-1-AX. A) Binding of CBM17-OG alone (●) and in the presence of 15 μM CBM4-1-AX (○); CBM4-1-AX bound vs. CBM17-OG free is also plotted (▲). B) Fraction bound CBM4-1-AX ([B]/[N]) vs. fraction bound CBM17-OG ([B]/[N]).
Figure 3.19 Binding of CBM4-1-AX to PASC in the presence of CBM17-OG. A) Binding of CBM4-1-AX alone (●) and in the presence of 15 μM CBM17AX (○); CBM17-OG bound vs. CBM4-1-AX free is also plotted (▲). B) Fraction bound CBM17-OG ([B]/[N]) vs. fraction bound CBM4-1-AX ([B]/[N]).
3.5 CBM2a binding adsorption and exchange

3.5.1 Rate of Adsorption of CBM2a to PASC and BMCC

CBM2a bound to PASC at a similar rate at the two concentrations tested. At 19.6 μM, CBM2a bound to PASC with a t_{1/2} of 4 min; at 27.8 μM, CBM2a binds with a t_{1/2} of 6 min. The rate of adsorption at both concentrations is within experimental and regression standard error (Figure 3.20-A). The rate of CBM2a adsorption to BMCC was too fast to measure by this method; CBM2a reached binding equilibrium in less than 30 s (Figure 3.20-B).

Assuming that CBM2a binds to crystalline regions of BMCC and PASC by a similar mechanism, then the binding of CBM2a to PASC is likely limited by mass transfer through the porous cellulose matrix of PASC.

3.5.2 Equilibrium exchange of CBM2a on BMCC

CBM2a binds irreversibly to cellulose (25, 57, 88). Binding exhibits hysteresis; the desorption isotherm, generated from dilution, is not congruent with the adsorption isotherm. Two cellulases with family 2 CBMs (11) (*Thermomonospora fusca* Cel6B, formerly E3 and Cel5A, formerly E5) also show binding hysteresis yet they exhibit surface exchange. The cellulases bound to the surface are displaced by those free in solution, without apparent disturbance of the apparent binding equilibrium.

Based on observations from the competition experiments, CBM2a also exhibited surface exchange; the extent and rate of the surface exchange of CBM2a was investigated. BMCC was loaded with CBM2a-AX at a total concentration of 18.2 μM. After equilibration at 4°C, the average concentration of free CBM2a-AX was 4.2 μM and the average bound concentration of CBM2a-AX was 14.1 μmol·g⁻¹ (approximately the maximum loading of
BMCC with CBM2a). The concentration of bound CBM2a removed during the three washing steps was less than 1 \( \mu M \), indicating that almost all of the bound CBM2a is bound irreversibly. After washing, CBM2a-OG was added at different concentrations: below equilibrium (1.4 \( \mu M \)), at approximately equilibrium (3.2 \( \mu M \)) and greater than equilibrium (14 \( \mu M \)). After equilibration, the concentration of CBM2a-AX and CBM2a-OG in solution was measured and the concentration of the labelled species bound to BMCC calculated. The total concentration of CBM2a bound to BMCC (CBM2a-AX plus CBM2a-OG) remained constant at approximately 14 \( \mu mool \cdot g^{-1} \) regardless of the concentration of CBM2a-OG added (Figure 3.21). The proportion of the CBM2a-AX and CBM2a-OG bound to the surface did however depend on the overall proportion of each labelled CBM2a present (Table 3.11). For example, at high concentration addition of CBM2a-OG (14 \( \mu M \)), CBM2a-OG represented 50% of the total CBM2a present, and 42% of the total CBM2a bound to BMCC CBM2a-OG. A large fraction of the surface-bound CBM2a molecules can therefore exchange with CBM2a molecules in the solution phase. The quantification of the extent of exchange is much more reliable at highest concentration of added CBM2a-OG; at lower concentrations (<3.2 \( \mu M \)), the concentrations of CBM2a-AX and CBM2a-OG are difficult to measure with confidence.

<table>
<thead>
<tr>
<th>CBM2a-OG Added (( \mu M ))</th>
<th>Sum CBM2a-OG/Total CBM2a</th>
<th>Bound CBM2a-OG/Total Bound CBM2a</th>
<th>Sum CBM2a-AX/Total CBM2a</th>
<th>Bound CBM2a-AX/Total Bound CBM2a</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (Buffer)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.4</td>
<td>0.09</td>
<td>0.06</td>
<td>0.91</td>
<td>0.94</td>
</tr>
<tr>
<td>3.2</td>
<td>0.18</td>
<td>0.15</td>
<td>0.82</td>
<td>0.85</td>
</tr>
<tr>
<td>14.0</td>
<td>0.50</td>
<td>0.42</td>
<td>0.50</td>
<td>0.58</td>
</tr>
</tbody>
</table>

Note: The Sum CBM2a-OG equals the total CBM2a-OG present (sum of free CBM2a-OG and bound CBM2a-OG), and similarly CBM2a-AX equals the total CBM2a-AX present. Total bound CBM2a equals the sum of bound CBM2a-AX and bound CBM2a-OG.
3.5.3 Rate of CBM2a exchange on BMCC

The rate that surface exchange occurs was investigated. Ten samples, each with 1 mg of BMCC, were loaded with a total of 16 μM of CBM2a-AX and equilibrated at 4°C. At equilibrium, the average bound CBM2a-AX was 12.5 μmol·g⁻¹, the average free CBM2a-AX was 3.4 μM. Each sample was washed three times with potassium phosphate buffer. CBM2a-OG was then added and incubated with BMCC for different times, the amount of CBM2a-AX and CBM2a-OG remaining in the solution phase was measured and the concentration of each labelled species bound to BMCC was calculated. CBM2a-OG re-established equilibrium by exchanging with the CBM2a-AX bound to the cellulose surface, in less than 30 seconds, faster than could measured by this method (Figure 3.22).
Figure 3.20 Adsorption kinetics of CBM2a binding to PASC and BMCC. A) Rate of adsorption to PASC (fraction of equilibrium bound; $[B]/[B_{\text{max}}]$ vs. time (min)) at a total concentration of 19.6 μM CBM2a-OG (●, solid line) and 27.8 μM CBM2a-OG (○, dashed line). B) Rate of CBM2a-OG adsorption to BMCC at a total concentration of 19.7 μM (●).
Figure 3.21 Surface exchange of CBM2a, at equilibrium, on BMCC. Concentration of CBM2a populations versus the total free concentration of CBM2a; CBM2a-OG (O, dashed line), CBM2a-AX (▲, dotted line) and total CBM2a (a sum of bound CBM2a-AX and CBM2a-OG; ●, solid line) is plotted; the connecting lines are added for visual reference only.

Figure 3.22 Kinetics of CBM2a surface exchange on BMCC. Bound CBM2a populations CBM2a-AX (O, dashed line), CBM2a-OG (▲, dotted line) and total CBM2a (a sum of bound CBM2a-AX and CBM2a-OG; ●, solid line) vs. time are plotted; connecting lines are added for visual reference only.
3.6 CBM2a-cellulose interaction observed by solid state NMR

This study is the first documented attempt to directly observe the CBM-cellulose interaction using solid state NMR. Detection of this interaction relies upon observing interacting nuclei that are visible by NMR, such as $^{13}$C and $^{15}$N. Here, $^{13}$C-enriched cellulose was produced using A. xylinum and CBM2a was homogeneously labelled with $^{15}$N ($^{15}$N-CBM2a).

3.6.1 Production and characterization of $^{13}$C-enriched bacterial cellulose

A sample of CBM2a bound to bacterial cellulose was prepared for the solid state NMR studies using cellulose produced by A. xylinum and processed as described in a previous section (Section 2.6.2). The CBM2a had binding affinity and capacity ($K_a = 1.7 \times 10^6 \text{ M}^{-1}$, $[N_0] = 12.9 \mu\text{mol}\cdot\text{g}^{-1}$) for binding this preparation of bacterial cellulose that was similar to CBM2a binding to BMCC. $^{13}$C-4-enriched bacterial cellulose, synthesized from D-glucose-$^{13}$C in minimal medium, yielded 25 mg of cellulose (dry weight) from 100 ml of culture. The C-4 of the bacterial cellulose was not labelled completely uniformly or uniquely. Signals corresponding to C-1 and C-2,3,5 regions of cellulose were evident in the spectrum of $^{13}$C-4-enriched cellulose (Figure 3.23). Through unanticipated metabolic processes, the bacteria scrambled some of the label that was incorporated into the bacterial cellulose. Since the bacterium rearranges the $^{13}$C label from glucose, A. xylinum was grown in a rich medium using glycerol, because, the $^{13}$C-labelled glycerol preparations are significantly less expensive than $^{13}$C-glucose preparations. Growing A. xylinum 53524 in H&S medium, using glycerol-1,3-$^{13}$C$_2$, resulted in higher cellulose yields than in minimal medium, (191 mg / 100 ml of culture, dry weight) with significant labelling (~50%) at C-1, C-3, C-4 and C-6 (Figure 3.24). The main advantage of the labelled cellulose preparations is the greatly improved signal to noise ratio as evidenced by the much fewer number of scans needed to collect a
high resolution spectrum. Using the $^{13}\text{C}$ natural abundance (~1.1%), the spectrum of bacterial cellulose was the sum of 40 000 scans; spectra of $^{13}\text{C}$-enriched cellulose, with comparable resolution, could be collected with only 400 scans (Figure 3.24). It is very unlikely that CBM-cellulose interaction could be detected by relying on $^{13}\text{C}$ natural abundance alone.

### 3.6.2 CBM2a-cellulose interactions

Approximately 7 mg of $^{15}\text{N}$-CBM2a was bound, in solution, to ~50 mg (dry weight) $^{13}\text{C}$-enriched cellulose from the glycerol-$1,3^{-13}\text{C}_2$ synthesis; excess solution was removed so that the sample was approximately 40-60% water content, by weight. The interaction between the $^{15}\text{N}$ nuclei of CBM2a and $^{13}\text{C}$ of the cellulose should be detectable by a CP-drain experiment if there are a sufficiently large number of $^{13}\text{C}/^{15}\text{N}$ interactions for the sensitivity of the experiment to detect. The CP-drain experiments rely upon observing the CP/MAS spectrum of one nucleus (e.g. $^{13}\text{C}$ of the enriched cellulose), after transfer of magnetism to a second nucleus (e.g. $^{15}\text{N}$ of $^{15}\text{N}$-CBM2a). If the two nuclei are in close proximity (i.e. the $^{15}\text{N}$ of the CBM bound to the surface of the $^{13}\text{C}$-labelled cellulose), there will be transfer and a decrease in signal will be observed; if they are too distant, no transfer will occur. This experiment is relatively robust and since it measures a decrease in the $^{13}\text{C}$ signals caused by only the interactions with the $^{15}\text{N}$ nuclei, it should therefore unambiguously detect only surface binding of $^{15}\text{N}$-CBM2a to $^{13}\text{C}$-enriched cellulose. Moreover, nuclear interactions between specific $^{13}\text{C}$ atoms of the cellulose (eg. C-4), and specific $^{15}\text{N}$ can be distinguished, in theory. Unfortunately, we did not detect CBM-cellulose interaction by the CP-drain experiments; the spectrum of the labelled cellulose after CP-drain was virtually indistinguishable from the pre-CP-drain spectrum (shown in Figure 3.24-B). The system was working properly as the $^{15}\text{N}$-
CP/MAS spectrum of $^{15}$N-CBM2a bound to bacterial cellulose could be measured and collected (Figure 3.25).

This is the first time a CBM has been observed bound to cellulose. The $^{15}$N-CP/MAS spectrum can be approximately correlated with a one-dimensional projection of the $^{15}$N signal from the HSQC of $^{15}$N-CBM2a (Figure 3.25); the spectra are somewhat similar yet, characteristically, the peaks of the $^{15}$N-CP/MAS spectrum are broader than that of the spectrum measure of CBM2a in solution. CBM2a itself appears to be folded when bound to the surface and is not denatured.
Figure 3.23 $^{13}$C-CP/MAS Spectra of bacterial cellulose and $^{13}$C-enriched cellulose from $^{13}$C-4-glucose synthesis. (A) Bacterial cellulose ($^{13}$C natural abundance, 1.1%) spectrum was the sum of 40,000 scans. (B) $^{13}$C-enriched bacterial cellulose spectrum is the sum of 400 scans. Horizontal bars indicate the spectral regions of the corresponding carbon atoms in the glucose monomer unit of cellulose.
**Figure 3.24** CP/MAS Spectra of bacterial cellulose and $^{13}$C-enriched cellulose from 1,3-$^{13}$C-glycerol synthesis. (A) Bacterial cellulose ($^{13}$C natural abundance, 1.1%) spectrum was the sum of 40,000 scans, collected as described in the materials and methods. (B) $^{13}$C-enriched (~50%) bacterial cellulose spectrum is the sum of 400 scans.
Figure 3.25 $^{15}$N CP/MAS spectrum of $^{15}$N-CBM2a bound to insoluble cellulose (A) and HSQC spectrum of $^{15}$N-CBM2a in solution (B). The scales of the solid (A) and liquid spectra have been aligned. Peaks of $^{15}$N-CP/MAS of $^{15}$N-CBM2a bound are roughly aligned to the corresponding peaks of the HSQC peaks of $^{15}$N-CBM2a in solution. All spectra are referenced using liquid ammonia; the $^{15}$N-CP/MAS spectra had a secondary reference of solid ammonium chloride. The full $^{15}$N CP/MAS of $^{15}$N-CBM2a, showing upfield spectral features is shown in (C).
4. Discussion

The purpose of this study was to further elucidate the nature of the interaction of CBMs with their ligand and in doing so, further understand the biological role that CBMs play as components in polysaccharide degradation systems and investigate the diversity of CBM structure and function. CBM specificities for the binding to BMCC and PASC was studied by a series of competition experiments using six CBMs representing the three CBM types. As a consequence of knowing the binding specificities of this complement of CBMs, insights as to the structure of the complex structure of PASC were gained using the competition binding technique. It was shown that PASC, has distinct regions of both crystalline and amorphous cellulose.

Family 2a CBMs are the most common CBM found in the cellulases and xylanases of C. fimii. The binding of CBM2a to cellulose was investigated in more detail to further understand its function in substrate binding, as component of an organism’s polysaccharidases, and for the potential enhancement of properties useful in specific applications such as affinity purification and enzyme immobilization. Although it has been reported that CBM2a binds with apparent irreversibility, during the course of the competition experiments, it was observed that CBM2a molecules that are bound to cellulose exchange with free CBM2a molecules in solution. The extent and kinetics of surface exchange was investigated. Knowing that CBM2a binds to crystalline cellulose, the residues that are important determinants of binding affinity to BMCC were explored by site-directed mutation. A number of residues on the binding face of the molecule were explored: conserved tryptophan residues were mutated to tyrosine, phenylalanine and alanine; other polar residues with the potential to interact with cellulose were substituted with alanine.
For the first time, the direct observation of CBM2a bound to insoluble cellulose was attempted using the technique of solid state NMR. $^{15}$N-labelled CBM2a was produced, bound to $^{13}$C-enriched cellulose and observation was attempted by a CP-drain experiment, a robust and sensitive technique that can detect nuclei that are in close proximity. Further refinement of solid state NMR technique should yield more structural information about CBM2a bound to cellulose in its native state.

4.1 Cellulose and CBMs: a diversity of structure and function

4.1.1 Binding specificity of CBMs for cellulose

The competition binding isotherm experiments, detailed in this study, are a new approach for the investigation of CBM binding specificity for complex substrates. Solution-dependent methods, such as binding studies with soluble ligands (123), NMR, or crystallization of CBM with ligand, are not applicable to studying the specificities of CBMs for complex insoluble substrates such as PASC. As a consequence of knowing the binding specificity of CBMs, competition experiment can be used to gain information about the structures, and mixtures of structures, present in complex substrates. For example, as the current work shows, PASC is comprised of different forms of amorphous cellulose as well as a large surface area of crystalline cellulose. This technique could be applied to characterize surfaces of other substrates, such as wood pulp.

4.1.1.1 CBM2a binds to crystalline regions of cellulose

Until now, there has been little direct evidence that CBM2a binds strictly to the crystalline regions of insoluble cellulose. In competition with type B CBMs with specificities for amorphous cellulose and soluble ligands, it was shown that CBM2a binds to sites not
recognized by the type B CBMs (79). The type B CBMs used in competition experiments, CBM4-1 and CBM17, bind soluble oligosaccharides and amorphous cellulose; they have no significant affinity for highly crystalline BMCC (8, 56, 60, 61, 108, 125). CBM4-1 binds saccharide chains within an obvious binding groove, and has the greatest affinity for cellopentaose, or longer oligosaccharides, and amorphous cellulose. Binding is enthalpically driven and is fully reversible (125). CBM17 also binds reversibly to cello-oligosaccharides and amorphous cellulose. The shallow, twisted binding groove of CBM17 binds cello-oligosaccharides and has the greatest affinity for cellohexaose and PASC. In contrast to these two modules, the type A module CBM2a has a more planar binding face and does not bind small sugars or cello-oligosaccharides. CBM2a has a similar affinity for all forms of insoluble cellulose tested, BMCC, Avicel and PASC (K_a ~10^6 M^-1; Table 3.1) suggesting that it binds to the same component in each of these celluloses. Binding is entropically driven, presumably due to dehydration of apolar surface residues (25). Since the binding of CBM17 and CBM4-1 to PASC was not significantly affected by the presence of CBM2a, and the binding of CBM2a to PASC is largely unaffected by the presence of CBM4-1, CBM2a must bind to regions not recognized by either CBM4-1 or CBM17. Presumably, CBM4-1 and CBM17 bind to individual cellulose molecules in amorphous regions of PASC, whereas CBM2a binds to crystalline surfaces (Figure 4.1).

Like CBM2a, the type A binding modules CBM3 and Cel6A-CBM2a, also bind to highly crystalline celluloses, such as BMCC (Table 3.4), but not to cello-oligosaccharides or soluble sugars. All three CBMs have a linear platform of aromatic residues exposed on one face of the module that are major mediators of the binding interaction. CBM2a and Cel6A-CBM2a have very similar amino acid sequences, including three exposed tryptophans on the binding face. They have similar affinities for insoluble cellulose (Table 3.4). Type A CBMs
all had a similar capacity for PASC (~15 μmol·g⁻¹) and a similar capacity for BMCC (~11 μmol·g⁻¹) (Table 3.4). As might be expected, CBM2a binds to the same sites on BMCC and PASC as Cel6A-CBM2a. (Figures 3.16 and 3.17; Table 3.9). Although the CBM3 amino acid sequence is dissimilar to those of the family 2a binding modules, it has a similar anti-parallel β-sheet structure with a platform of conserved exposed aromatic amino acids. Competition experiments show that CBM3 binds to the same sites of PASC as CBM2a (Figure 3.15; Table 3.9).

The binding of CBMs with affinity for crystalline cellulose is a function of the total solvent exposed surface area of the crystalline cellulose that is available to CBM2a. The binding capacity, [N₀], does not directly correlate with the bulk crystallinity of the cellulose. Overall, BMCC has a larger proportion of crystalline cellulose than PASC; however, based on the larger capacity of PASC to bind CBM2a, a much larger surface area of the crystalline regions of PASC is available to the binding module and, as shown by competition binding experiments, CBM2a binds exclusively to the crystalline regions of cellulose. Paradoxically in CP/MAS spectrum of PASC (Figure 3.2), the predominant broad C-4 peak at 81-83 ppm, characteristic of disordered, surface exposed C-4 atoms, and the extremely small peak at ~89 ppm arising from the highly ordered crystalline regions (4), shows that PASC, an acid-swollen preparation of Avicel, has significant amorphous characteristic, yet CBM2a, which binds to crystalline cellulose, binds to PASC with high capacity (Table 3.1). On the basis of weight, the crystalline surface area made available to CBM2a is increased by approximately six-fold by the acid swelling process that converts Avicel to PASC. This process is accompanied by with a drastic decrease in the bulk crystallinity, or total volume of the crystalline regions, as characterized by the disappearance of the sharp downfield C-4 peak at ~89 ppm in the ¹³C-CP/MAS spectrum of PASC compared with Avicel(Figure 3.2). The acid
swelling process partially dissolves the cellulose; the cellulose is re-precipitated by dilution in dH2O (37, 139). Cellulose longer than seven glucose monomer units is not very soluble in water and forms aggregates. The crystallinity of PASC can be envisioned to be a large number of microcrystalline regions formed by the aggregation of a number of cellulose chains. These microcrystalline regions have sufficient surface area to be bound by CBM2a, are extremely numerous, manifested in the increased capacity for CBM2a, yet are extremely small in volume so are not seen in bulk measurements of crystallinity.

Crystalline cellulose may also be heterogeneous, in terms of binding specificity (16, 25). Although I did not detect such heterogeneity, there is some evidence, based on ITC data (25), that crystalline cellulose presents two classes of high affinity binding sites to CBM2a. The affinities of CBM2a for the crystalline regions of BMCC, PASC, and Avicel are similar; the relatively small differences (Table 3.1) likely reflect the proportion of high and low affinity binding sites on each cellulose preparation. The structural differences between these two apparently different binding regions are unclear. It could be a combination of the relative proportion of the cellulose allomorphs, such as cellulose Iα, Iβ and cellulose II. Further competition experiments, like those performed in this study, involving an expanded library of type A CBMs binding to cellulose with different proportions of cellulose allomorphs may reveal further details of the complex surface structure of crystalline cellulose, and the basis for CBMs distinguishing the various sites.
Figure 4.1 CBMs bound to native cellulose. A) Native cellulose with crystalline and amorphous regions. Individual cellulose molecules are represented by black lines. The cellulose molecules are aggregated into crystalline regions, which are highly ordered, and more disordered amorphous regions with semi-soluble, single cellulose chains. The reducing ends of the parallel cellulose molecules are also indicated. B) CBMs bound to cellulose. CBM2a (grey-shaded) and CBM3 (light-grey) bind to the crystalline regions, CBM9-2 (unshaded) binds to the available reducing ends and CBM4-1 (chequered) and CBM17 (cross-hatched) bind to amorphous regions. As demonstrated by competition binding experiments, some of the sites of amorphous cellulose are bound only by CBM17 or CBM4-1; some amorphous cellulose sites can be bound by both.

4.1.1.2 Heterogeneity of amorphous cellulose structure

The two type B binding modules, CBM4-1 and CBM17, bind cello-oligosaccharides and amorphous cellulose; they compete for binding sites on PASC. The apparent affinity of CBM4-1 for PASC is reduced from $0.24 \times 10^6 \text{ M}^{-1}$ to $0.003 \times 10^6 \text{ M}^{-1}$ in the presence of CBM17 (Figure 3.19, Table 3.10). Similarly, the apparent affinity of CBM17 for PASC is reduced from $0.81 \times 10^6 \text{ M}^{-1}$ to $0.31 \times 10^6 \text{ M}^{-1}$ in the presence of CBM4-1 (Figure 3.18, Table 3.10). There are, however, selected regions of the cellulose that are recognized uniquely by one and not the other CBM. The fraction of CBM4-1 bound ([B]/[N₀]) decreased to only $\sim0.5$ as CBM17 approached saturation (Figure 3.18-B) and only a small fraction of the bound
CBM17 was displaced by saturating amounts of CBM4-1 (Figure 3.19-B). The data suggest that there are at least three distinct sites on PASC, one recognized only by CBM4-1, one recognized only by CBM17, and a third class of sites recognized by both of them. The structures of the binding sites on CBM17 and CBM4-1 may help to reveal the differences between these sites on PASC (Figure 4.2). In the crystal structure of CBM4-1 bound to cellopentaose, binding occurs in a relatively narrow groove on the polypeptide; the sugar chain is almost fully encompassed by the module. The CBM contacts both faces of the cellobiosaccharide molecule. By contrast, the binding groove of CBM17 is much shallower, somewhat twisted and makes contact with only one face of the sugar chain. In a preparation of PASC with a number of complex amorphous structures, CBM4-1 therefore could bind only fully accessible individual cellulose molecules, whereas, CBM17 could bind to some of the fully accessible chains recognized by CBM4-1, as well as some partially accessible single molecules. The capacity of PASC is approximately three times greater for CBM17 than for CBM4-1, suggesting a much larger proportion of amorphous cellulose comprises only partially accessible molecules. These subtle differences in the specificity of CBM17 and CBM4-1 underline the heterogeneity of amorphous cellulose in PASC (Figure 4.1; Figure 4.2).
Figure 4.2 Details of the binding sites of CBM4-1 (A) and CBM17 (B) with bound cellopentaose and cellotetraose, respectively. The cellobioisaccharides are rendered in ball and stick form. To emphasize the shape of the respective ligand-binding sites, the protein surface is rendered to display the solvent-accessible surface; it is shaded according to the electrostatic charge. In both figures, the structure of the complex is viewed parallel to the long axis of the cellobioisaccharide (A. B. Borason, V. Notenboom, A. Freelove, unpublished results)
4.1.2 CBMs: agents for targeting polysaccharidases and for saturating substrate

Cellulose can have different crystalline structures (2, 3). For example, cellulose I, the form of cellulose produced by cellulose synthase, comprises a mixture two distinct crystalline types, designated I$_a$ and I$_p$, the proportion of each type depending on the source. These competition experiments show that amorphous cellulose is also heterogeneous. In the plant cell wall, cellulose is intimately associated with hemicellulose which may include mannans, galactomannans, xylan, xyloglucans, pectin and some glucans (Figure 1.1). Mirroring the diversity of cellulose and polysaccharide polymers, various organisms have evolved an array of modular polysaccharidases to degrade them. The CBMs display an equally diverse array of substrate specificities. Currently, among the CBMs that have been characterized, are those that bind cellulose (9), xylan (20, 112), and mannan (119).

Based upon overall architecture and substrate specificity, there are three types of CBM. Within each type, and often within CBM families, the binding specificities of individual CBMs are different. CBMs are thought to promote the hydrolysis of polysaccharides by increasing the concentration of enzyme on insoluble and semi-soluble substrates (38, 47, 98, 128) or, in the case of TfCel9A-CBM3 (103), by guiding substrate into the active site of the catalytic module. CBMs could also promote substrate hydrolysis by targeting their associated hydrolases to specific sites on the substrates, especially in heterogeneous material like the plant cell wall. The specificity of the CBM can influence the overall specificity and activity of the associated cellulase. For example, when the family 2a CBM from CfCel6A (CenA) was exchanged for the tandem family 4 CBMs of CfCel9B (CenC), the resulting enzyme, with two CBM4-1 modules and the Cel6A catalytic module, was less active on Avicel and BMCC than Cel6A, yet more active on CMC , a soluble derivative of cellulose, and cellulose azure, a regenerated form of cellulose (22). In this case, the specificity or enzyme activity of
the catalytic module itself is not changed, however the types of cellulose for which the enzyme is most active changes and this change is mediated by the associated CBM.

The specificity of the CBM may, or may not, correlate with the overall substrate specificity of the associated enzyme. The binding specificities of CBM2a modules do not correlate with the associated enzyme’s substrate. In *C. fimi*, CBM2a modules are associated with enzymes that cleave β-1,4-linked glycans: celllobiohydrolases (Cel48A and Cel6B) (80, 107), endoglucanases with high activity with crystalline (Cel9A; Cel5A) or amorphous cellulose and soluble cellulose derivatives (Cel6A, Cel9A, Cel9B), or xylan (Xyn10A) (127). All of the Family 2a CBMs studied bind to crystalline cellulose only.

There is some evidence that binding specificity of family 4 CBMs does correlate with overall enzyme substrate specificity (144). Among the polysaccharidases with family 4 CBMs are three celllobiohydrolases (35, 133, 143), a xylanase (62), β-1,4-glucanases (55, 127), and a laminarinase (β-1,3-glucanase) (145). The family 4 CBMs from *C. fimi* Cel9B, as discussed previously, bind to amorphous cellulose and soluble sugar polymers. Cel9B itself has activity on amorphous cellulose. Unlike CfCel9B-CBM4-1 and CfCel9B-CBM4-2, the family 4 binding modules from *Thermotoga neapolitana* laminarinase (Lam16A), potentially members of a sub-family within family 4, do not bind β-1,4-glycans; they have affinity for soluble (TnLam16A-CBM4-1 and TnLam16A-CBM4-2) and insoluble (TnLam16A-CBM4-2) 1,3-β-glucans. The binding specificities of these two binding modules is virtually identical to the preferred substrates of the catalytic module (144). More information about the binding specificities of other family 4 CBMs and their associated glycosyl hydrolases is neccessary to unequivocally establish a correlation.
There are eight polysaccharidases with CBM17 modules and each is associated with a family 5 glycosyl hydrolase. Only three of the enzymes, and one associated CBM17, have been characterized. The endo-1,4-glucanase from *Bacillus* sp. 22-28 is active with CMC (83); the endo-1,4-glucanase from *Bacillus* sp. KSM-S237 also has activity with CMC, it does not hydrolyze crystalline cellulose (46). The binding characteristics of the associated CBM17s are not known. The *C. cellulovorans* cellulase 5A (EngF) has the greatest activity with CMC, followed by PASC and lichenan. Little or no activity was observed with laminarin, xylan and Avicel were tested as substrates (108); the binding module also binds to CMC and amorphous cellulose. With only these examples, there is not enough information to assess any apparent correlation between CBM17 binding specificity and its associated enzyme activity.

In many cases, there is little apparent correlation between CBM binding specificity and enzyme substrate specificity, it is unlikely that CBMs act only to target the enzyme to a particular substrate. CBMs could be increasing the efficiency of the degradation of insoluble, and heterogeneous substrates (Figure 1.1) by binding to one area of the complex substrate while the catalytic module hydrolyzes substrate in an area located away from the site bound by the CBM. For example, the family 2a binding module of CfCel6A binds to a crystalline region of cellulose while the enzyme, separated from the CBM by a linker sequence, may access available amorphous regions of the cellulose (see Figure 4.1; crystalline and amorphous cellulose regions are adjacent). Polysaccharides are intimately associated in the plant cell wall, the CBM2a of CfXyn10A may bind crystalline cellulose, while the associated catalytic module digests xylan that is within close proximity. CBMs, therefore, may also ensure adequate coverage of the complex substrate by enzymes with different specificities.
4.2 Adsorption of CBM2a to cellulose

Six of the eight cellulolytic enzymes of C. fimi have family 2a CBMs (Figure 1.2). This module is an important component of the polysaccharidases of system of this organism. Additionally, as a fusion partner, CBM2a has been used as an affinity tag for the purification or immobilization of a number of polypeptides (124). Both an understanding of CBM2a in its biological context and for the alteration or customization of CBM2a for specific applications motivate the need to have more detailed understanding of the specific interactions of CBM2a with cellulose. To do so, the binding interaction of CBM2a with BMCC and PASC was characterized, the irreversibility of binding to BMCC was assessed, the amino acids important in binding were probed and direct observation by solid state NMR CBM2a bound to cellulose was attempted.

4.2.1 Irreversibility of adsorption of CBM2a

It has been reported that CBM2a adsorbs irreversibly to insoluble cellulose (25, 87); complete removal of CBM2a from the solution phase does not result in any appreciable desorption of CBM2a from the surface. Binding exhibits hysteresis; the isotherm that describes the adsorption of CBM2a to cellulose is not equivalent to the desorption isotherm generated by of removal of CBM2a from solution. The equilibrium described by the adsorption isotherm is not re-established. Although apparently irreversible, binding is dynamic: under conditions where all CBM2a in the solution phase has been removed, CBM2a moves in two-dimensions on the surface without desorbing from it (57). Paradoxically, the current work shows that a high percentage of CBM2a molecules bound to the cellulose surface can exchange with CBM2a molecules in solution; binding is not strictly irreversible. In these exchange experiments, most (>93%) of CBM2a-AX was not desorbed from BMCC by incubation with buffer alone however, any addition of free CBM2a-OG
displaced CBM2a-AX bound to the surface. Most of the CBM2a molecules bound to
BMCC are available to exchange. The process of surface exchange of CBM2a is rapid.
Binding equilibrium is re-established in less than 30 s (Figure 3.22) faster than the rate of
exchange observed for the family 1 CBM, TrCel7A-CBM1 (73). The adsorption of CBM2a,
and other binding modules, can be described by the Langmuir-like binding expression ((37)
and equation 1), but the observed binding hysteresis when protein is removed from the
solution phase and the surface exchange indicates that the binding reaction is not a simple
two-state equilibrium. The actual mechanism of surface exchange is not known, however, the
data is consistent with a model requiring the interaction of three components: CBM2a in
solution, CBM2a bound to the surface, and the sorbant surface itself. CBM2a in the solution
phase is required to displace of CBM2a bound to the surface, indicating a protein-protein
interaction drives the direct, one to one exchange of bound and free CBM2a. This process
can be envisioned using the following speculative model (Figure 4.3):

Figure 4.3 A model of CBM2a surface exchange. CBM2a establishes an apparent adsorption
equilibrium (1). CBM2a in the solution phase is removed, bound CBM2a does not desorb
because of the energy barrier involved with the hydration of the tryptophan-rich binding face.
CBM2a added to the solution phase (3). By diffusion, solution-phase CBM comes in contact
with bound CBM and the surface and there is a protein-protein interaction (4) with two
possible results: free CBM2a binds to the surface, bound CBM2a released into the solution
phase (5a) or free CBM2a separates from bound CBM2a without displacement of the bound
CBM2a (5b)
A CBM2a molecule is bound to the surface of cellulose; it does not desorb from it when CBM2a in the solution phase is removed. CBM2a added to the solution diffuses to the surface and is involved, in a three-way interaction involving the bound CBM2a, the free CBM2a and the surface itself. There are two possible outcomes of this interaction. The bound CBM2a is displaced, and the formerly free CBM2a binds to the surface, resulting in a one-to-one exchange of bound and free CBM2a. Or, the free CBM2a dissociates from the bound CBM2a and the surface without exchanging. The nature of the binding interaction of CBM2a and the cellulose surface could aid in explaining the mechanism of the protein-protein-surface interaction resulting in exchange. Dehydration of the tryptophan-rich hydrophobic binding ridge drives binding; the desorption of CBM2a directly into solution is prevented by the energy required to rehydrate the hydrophobic binding face. Exchange or displacement requires a protein-protein interaction, perhaps involving contact of two CBM2a hydrophobic binding faces. This protein-protein, interaction may be a lower energy, intermediate step between CBM2a bound to the surface and CBM2a free in solution that allows for the desorption of one CBM2a molecule, with the concomitant adsorption of the second CBM2a molecule.

CBM2a is not unique in its apparent irreversibility of binding and ability to exchange. Cellulases with either family 1 (69) or family 2 CBMs (11) also demonstrate surface exchange. Of two Thermomonospora fusca cellulases with family 2a CBMs (Cel6B, formerly E3 and Cel5A, formerly E5), only about 75% of the bound population of these cellulases were available to exchange, whereas most of the C. fimi CBM2a molecules bound to BMCC were available for surface exchange (Table 3.11). Additionally, there is no obvious correlation between the apparent binding irreversibility, and surface exchange (11, 69, 73); CBMs that bind reversibly or irreversibly demonstrate surface exchange.
There are a number of biologically relevant reasons for CBMs and cellulases in solution to have the ability to exchange with those bound to the surface of insoluble substrates. Exchange may ensure that enzymes bound to the surface of substrate, but no longer functional, are continually turned over and replaced by functional enzymes. Exchange might be a mechanism for swapping one for an enzyme with another having different catalytic specificity. For example, plant material is heterogeneous, so it is possible that if one enzyme has locally exhausted its substrate in a particular area, it is replaced with another enzyme, with different specificity. Efficient and complete degradation of an heterogeneous substrate is ensured. Some of the cellulases (Cel9A, Cel5A, Cel6B, and Cel48A) produced in cultures of *C. fimic* grown on Avicel are subject to proteolysis, resulting in the separation of the CBM from the catalytic module (105). Surface exchange could assist the replacement of a substrate-bound, non-catalytically productive CBM, resulting from proteolytic cleavage, by an intact and active cellulase.

4.2.2 Amino acid residues involved in the binding of CBM2a to crystalline cellulose

The single-site mutation studies indicate that binding of CBM2a involves the concerted interaction with crystalline cellulose of a cluster of solvent-exposed residues located on a planar surface region (Figure 3.4) defined by the presence of three tryptophan residues (W17, W54, and W72). Each of the tryptophan residues makes a significant but different contribution to binding. W72A has 15-fold lower affinity for BMCC compared to the wild-type CBM2a. Replacement of W72 with either phenylalanine or tyrosine, however, results in relatively little loss of binding affinity, suggesting that the contact formed between W72 and the sorbent surface may involve hydrophobic interactions between the side chain and the apolar face of a glucopyranoside ring within the crystalline cellulose matrix.
The 100-fold loss in affinity exhibited by W54A is an order of magnitude more than for any other mutant and reflects the essential role of this residue in binding. The more conservative substitution of W54 with either phenylalanine or tyrosine reduces binding affinity only 10-fold. Thus, like W72, W54 appears to couple to the crystalline cellulose matrix, at least in part, through favorable hydrophobic interactions. The importance of the apolar aromatic character of W54 and W72 therefore supports ITC studies (25) which indicate that binding is primarily driven by dehydration of residues in close contact with the crystalline cellulose surface. Dehydration of contacting sorbent and protein surfaces and the concomitant formation of strong van der Waals contacts are known to drive a wide range of specific and non-specific protein adsorption processes (48). As observed for binding of CBM2a to BMCC (25), the thermodynamic signature of an adsorption process dominated by dehydration effects is a large positive change in entropy and a large negative change in heat capacity, both of which are due to the release of ordered water in the first and second solvation shells.

Replacement of the tryptophan at position 54 with phenylalanine leads to 10-fold drop in affinity, in contrast to W72F which has only a two-fold reduction from the wild-type. This suggests that the aromatic hydrophobic characteristic of W54 is essential for favourable interaction with the crystalline cellulose surface. Both W54, through its pyrrolic amine, and the cellobiose repeating unit of cellulose have the potential to form hydrogen bonds. The arrangement of β-linked glucopyranoside units in cellulose presents a uniform surface distribution of hydroxyl groups on the outside of each molecule. In native crystalline cellulose, the rigid β-1,4 glycosidic chain linkage positions these hydroxyl groups to allow all but one of the hydrogen bonds (per glucopyranoside) to be satisfied at both the (110) and (1-10) crystal faces through in-plane inter-chain interactions. (see Figure 4.4). Thus, in the
absence of local structural perturbation, cellulose I appears to offer a suitable proton acceptor for hydrogen bond formation with the pyrrolic amine of W54. The potential to form an intermolecular hydrogen bond suggests that W54 may provide binding specificity in addition to contributing to the overall driving force for adsorption to the crystalline cellulose surface due, in large part, to dehydration of the hydrophobic indole ring and the underlying crystalline cellulose surface, which exhibits a pronounced hydrophobic character.

The importance to binding affinity of proper formation of an intramolecular hydrogen bond involving the pyrrolic amine of a surface tryptophan is suggested by mutants of W17 and N15. Replacement of the tryptophan at position 17 with either alanine or phenylalanine, neither of which have hydrogen bonding potential, results in an order of magnitude reduction in binding affinity. An equivalent reduction in affinity is observed for N15A. The similarity in the affinities of both of these mutants could be due to the necessity of an intramolecular hydrogen bond, likely involving an essential bridging water molecule between N15 and the pyrrolic amine of W17, locking W17 into a specific orientation. This argument is supported by the fact that the mutant W17Y maintains native-like affinity, presumably because the intramolecular hydrogen bonding requirement for proper orientation of the side chain is at least weakly satisfied by the phenolic group of tyrosine.

Collectively, the properties of the site-directed mutants suggest that effective binding of a family 2a CBM to crystalline cellulose requires an aromatic group at position 72, a tryptophan at position 54, and either tyrosine or tryptophan, for possible hydrogen bond formation, at position 17. The tryptophan residues are well conserved among members of family 2a, consistent with the results obtained here by site specific substitution. W72 is the least well conserved of the three surface tryptophans. In 6 of approximately 45 members, this position is occupied by tyrosine. The interchangeability of tryptophan and tyrosine at this
position in CBM2a, also mirrored in other family 2a members, reinforces the role of hydrophobic interaction of an aromatic ring with the glucopyranoside rings of the cellulose surface. W54 is not effectively substituted by tyrosine, phenylalanine or alanine, and it is invariant among members of family 2a, further demonstrating that this residue likely contributes to binding both through dehydration and by providing sorbent specificity via its hydrogen bonding potential. Tryptophans corresponding to W17 are also strictly conserved; the asparagine corresponding to N15 is also highly conserved (>60%), supporting the hypothesis that an intramolecular hydrogen bond between these residues is an important functional feature of family 2a modules.

4.2.3 CBM2a-Cellulose interaction observed by solid state NMR

In biological applications, solid state NMR is an excellent tool for obtaining structural information for proteins that are not amenable to solution-based NMR techniques, or cannot be crystallized for X-crystallographic studies. Solid state NMR has been used to study prion proteins (49), membrane proteins (89), bacteriorhodopsins and gramicidin (110). The application of solid state NMR to the study of CBM-cellulose interactions was tried in an attempt to directly collect structural information about CBM2a bound to the surface of insoluble cellulose. Ideally, specific information regarding protein amino acid residues involved in binding and the allomorphs of cellulose and which are interacting can be gleaned. This study is the first attempt to observe the interaction between a CBM and insoluble cellulose and it lays the foundation for further, more sophisticated, solid state NMR measurements.

Since NMR relies upon observing nuclei that have nuclear spin, cellulose, enriched in $^{13}$C, and $^{15}$N-labelled CBM2a were produced. The C-4 carbon peaks of the $^{13}$C-CP/MAS
spectrum are particularly indicative of cellulose structure, so it was decided that producing cellulose specifically labelled at C-4 would provide definitive information regarding the regions of the cellulose with which the CBM was interacting. The cellulose produced by *A. xylinum* cultures, however, was not uniquely labelled at C-4 (Figure 3.23); the bacterium, likely through a pentose phosphate pathway (30), scrambled a portion of the label. The signal to noise benefit for using 13C-enriched cellulose prompted the production of larger amount of labelled material using glycerol-1,3,\(^{13}\)C\(_2\) as a carbon source for the *A. xylinum* cultures resulting in cellulose that was had a significant proportion of 13C in positions 1, 3, 4, 6 of the glucose monomer component of cellulose. This cellulose was used to attempt the observation of CBM2a-cellulose interaction.

Observation of CBM2a-cellulose interaction was attempted by a CP-drain experiment. This experiment is relatively robust and since it measures a decrease in the \(^{13}\)C signals caused by only the interactions with the \(^{15}\)N nuclei, it should therefore unambiguously detect only surface binding of \(^{15}\)N-CBM2a to \(^{13}\)C-enriched cellulose. The interaction between \(^{15}\)N-CBM2a and \(^{13}\)C-enriched cellulose could not be directly observed by this experiment for a number of possible reasons. For example, \(^{15}\)N-\(^{13}\)C dipolar couplings are weak, effective to an upper limit distance of 4-5 Å (118). \(^{19}\)F-\(^{13}\)C interactions however, are effective over a much greater distance. To take advantage of this fact, CBM2a has been produced, by a two-step expression method (101), that incorporates 5-fluoro-L-tryptophan. Immediate future study will focus on detecting the FW-CBM2a-cellulose interaction. Since CBM2a moves on the cellulose surface (with a diffusion coefficient of \(1.2 \times 10^{-10} \text{ cm}^2 \text{s}^{-1}\), (57)), any internuclear interaction will be transient, only occurring for a fraction of the 10 ms CP-drain duration. The diffusion could be slowed by performing the experiments at a lower temperature and repeating the CP-drain experiment.
Some valuable information was obtained in this study; a $^{15}$N-CP/MAS spectrum of $^{15}$N-CBM2a bound to the surface of cellulose was collected. A comparison of the $^{15}$N-CP/MAS spectrum with the projected $^{15}$N spectrum of $^{15}$N-CBM2 in solution (Figure 3.25), shows that $^{15}$N-CBM2a bound to the surface is folded and has a conformation that is roughly similar to that in solution. The binding of CBM2a to the cellulose therefore, does not involve a significant conformational change, also supported by calorimetric evidence (C. A. Haynes, personal communication). With further refinement, solid state NMR should be successful in detecting and characterizing CBM interaction with insoluble cellulose.

4.3 Towards a model of CBM2a-cellulose interaction

4.3.1 Previous models of binding

The specificity of CBM2a to bind crystalline cellulose and the identification of amino acid residues involved in binding provide a sound basis for formulating a putative binding model. Previous models for the binding of CBMs viewed crystalline cellulose as comprising layers of parallel cellulose molecules, with the sugar rings in successive layers perfectly superimposed. The surface to which the CBM binds would then be an ordered array of parallel cellulose chains, with the planes of the pyranose rings parallel to the surface and fully exposed to the solvent. Binding was postulated to occur by direct stacking of the exposed aromatic residues on the binding face of the CBM onto the exposed pyranose rings forming the (020) face of the cellulose crystal (77, 98, 130). The surface of crystalline cellulose, however, is more accurately described as a staircase (Figure 4.4) (51). In cross-section, viewed from one end, the crystal comprises layers with decreasing numbers of parallel cellulose molecules moving out to the vertices; the molecules in successive layers are off-set slightly with respect to those in the layer above it (5, 34, 45, 65).
4.3.2 Surface area of the exposed crystalline faces

The largest solvent-exposed surfaces, the two (110) and the two (1-10) crystal faces, comprise mostly the edges of the sugar rings (Figure 4.4). In a perfect cellulose crystal, only the sugar rings of the apical cellulose molecules at two opposite vertices of the rectangular crystal have fully exposed faces. BMCC comprises bundles of microfibrils. Each microfibril presents two solvent-exposed crystalline, roughly rectangular, faces with cross-sectional dimensions estimated to be 15 nm ((110) face) by 40 nm ((1-10) face) (67, 136). Using the dimensions of the BMCC microfibril and the density of crystalline cellulose (1.5 g/cm$^3$) (81), we calculate the (110) crystal face to have a surface area of $3.34 \times 10^5$ cm$^2$/g cellulose and the (1-10) face to have a surface area of $8.91 \times 10^5$ cm$^2$/g. As calculated from the solution structure, the maximum area shadowed by a bound CBM2a molecule is $1.32 \times 10^{13}$ cm$^2$ (calculated by D. Lim using the coordinates determined by Xu et al. (141)). At maximum capacity, assuming monolayer surface coverage and confluent packing of the CBM on the sorbent surfaces, the (110) crystalline face could accommodate 4.2 μmoles CBM2a/g BMCC and the (1-10) face could accommodate 11.2 μmole CBM2a/g BMCC. The predicted capacity of the two binding faces is therefore 15.4 μmoles/g. Given that the estimates of available cellulose surface area are based on a maximal fiber dispersal and monolayer surface coverage, the observed capacity of 11.7 μmoles/g cellulose (Table 3.1) agrees remarkably well with this estimated maximum value. The fully exposed cellulose chains, located at the two vertices of the rectangular fibril (the obtuse edges), could accommodate at most 0.91 μmol/g cellulose, far less than the observed capacity. Therefore, CBM2a must be binding to one or both of the crystalline faces ((110) and (1-10)) of BMCC by interacting with the partially occluded sugar rings. The previous model based on stacking of fully exposed sugar rings leads to a predicted capacity far less than observed experimentally.
4.3.3 A model for the binding of CBM2a to crystalline cellulose

Tryptophan residues are involved in many protein-carbohydrate interactions (95, 100, 122, 131, 132, 135). The non-covalent interactions between the aromatic residues of a protein and the sugar ring are not restricted to parallel stacking; there are likely many other conformations that are both thermodynamically stable and able to mediate ligand binding in crystalline cellulose. Since the great majority of pyranose rings of the glucose units are partially occluded in crystalline cellulose, the tryptophans can not stack directly on the sugar rings without a disruption of the cellulose structure. Other conformations, involving angled relationships or off-set parallel stacking that allow interaction between tryptophan and the sugars rings are therefore more likely to occur (78, 121, 135). The angle at which aromatic side chains are positioned with respect to the pyranose ring in protein-carbohydrate interactions varies from 17° to 52°; flat parallel ring stacking is not observed as frequently (135). With minimal movement from their equilibrium positions in the solution state, the surface tryptophans could form non-covalent interactions with the individual cellulose chains without substantial disruption of the crystal lattice. In CBM2a, the tryptophans essential for binding form a ridge along one face of the module. I propose that the tryptophans on this ridge bind to partially exposed cellulose chains, or ‘steps’, on the face of the cellulose crystal (Figure 4.4). The binding module tilts toward the ‘staircase’ so that other residues such as Q52 and N87, located alongside but slightly removed vertically from the vertex of the tryptophan ridge, are oriented in positions to potentially hydrogen bond or form van der Waal’s interactions with the groups at the edges of the cellulose chains. The module may bind along one cellulose molecule, or straddle several ‘steps’ that comprise the cellulose surface. The arrangement of the tryptophan residues appears to allow contact with cellulose via only two pairs of the three tryptophans at once: either W17 and W54, or W54 and W72. This model is also supported by the ability of CBM2a to bind to BMCC when it has a bulky
carbohydrate moiety attached to N24, a residue adjacent to the row of tryptophans on the binding face (7). N24 is glycosylated when the polypeptide is produced in *Pichia pastoris*. CBM2a binds to the surface in such a way that the large, hydrophilic carbohydrate group is accommodated between the protein and cellulose surfaces.

The binding of CBM2a to BMCC is dominated by entropic effects (25). Dehydration of the hydrophobic, tryptophan-rich protein surface and the crystalline cellulose surface drives binding. The specificity of the module for the cellulose surface, via the formation of hydrogen bonds and van der Waals contacts, is provided by both the surface tryptophans (especially W54) and other residues on the binding surface. The model proposed here agrees with this mechanism of binding. With the exception of N15, residues of CBM2a immediately adjacent to the tryptophan binding ridge that have hydrogen bonding potential have little overall effect on binding affinity.

The binding of CBM2a to crystalline cellulose apparently irreversible, removal of CBM2a from solution does not result in desorption of CBM2a from the surface, yet binding is dynamic (12, 25, 88). The domain moves in two dimensions over the cellulose surface without ever fully dissociating from it (57). Sufficient multiple contacts must be maintained to prevent desorption. Since the module moves on the surface, the number of residues directly interacting with the cellulose surface must be in constant flux. For example, once CBM2a is bound to cellulose, none of the surface tryptophans is protected from oxidation by N-bromosuccinimide, although they are more difficult to oxidize (12). The accessibility to N-bromosuccinimide is consistent with a model for binding in which the tryptophans do not overlap the pyranose rings completely, and only two of the three conserved tryptophans bind to BMCC at any instant. The mutation W54A reduced affinity the most, suggesting that this residue plays the major role in binding. The NMR structure shows that W54 is not in the
same plane as W17 and W72; part of the importance of W54 could be its role in allowing
pivoting between binding by W17/W54 and W54/W72 pairs. CBM2a could be envisioned to
“waddle” across the cellulose, using W54 as the essential residue contacting the surface.

4.4 A general model for the binding of type A binding modules

The abrogation or reduction of binding affinity when surface tryptophan residues of this and
other CBMs of family 2a are modified by site-directed mutation or chemical modification
clearly emphasizes the substantial role that tryptophans play in their binding to insoluble
cellulose (12, 27, 93, 94). Additionally, the conservation of aromatic residues, such as
tyrosine and phenylalanine, on an exposed surface is common in CBMs from families 1, 3, 5
and 10. Representatives from each of these families bind to crystalline cellulose and, as with
CBM2a, mutation of the exposed aromatic residues reduces the affinity for cellulose (27, 85,
93, 94, 96, 97, 111). Many of the CBM2a variants constructed here with conservative
substitutions maintain appreciable affinity for crystalline cellulose. It is likely that all of the
type A CBMs, those binding binding modules with affinity for crystalline cellulose having a
platform of exposed aromatic side chain residues, share a mode of binding where the surface
aromatics drive binding and mediate contact with the staircase-like cellulose surface.
Figure 4.4 Three views each of two possible arrangements of CBM2a bound to the (110) face of crystalline cellulose. I) CBM2a bound roughly along a single cellulose chain and II) CBM2a bound across a number of cellulose chains. (A) cross section of the microfibril, the internal cellulose chains are omitted for simplicity, (B) side view of the solvent-exposed surface and (C) view along the cellulose surface. The arrangements shown allows for at least two of the tryptophan residues (rendered as balls and sticks) instrumental in binding, to interact with the edge of the pyranose rings of the staircase-like surface of crystalline cellulose. One side of the molecule is tilted towards the surface, allowing for other residues (also rendered as balls and sticks) to interact transiently with the cellulose surface forming hydrogen bonds with the available cellulose hydroxyl groups. The cellulose surface was constructed by using the structure of repeating cellobiose units arranged to approximate the proposed crystalline surface of BMCC (5, 33, 34, 45, 65).
References


112. **Simpson, P. J., D. N. Bolam, A. Cooper, A. Ciruela, G. P. Hazlewood, H. J. Gilbert, and M. P. Williamson.** 1999. A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIa cellulose-binding domain but different ligand specificity. Structure Fold Des. 7:853-864.


