STRUCTURAL AND DYNAMIC ANALYSIS OF OLIGOSACCHARIDE BINDING BY CBD\textsubscript{N1}

by

PHILIP EDWARD JOHNSON

B.Sc., Simon Fraser University, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY in THE FACULTY OF GRADUATE STUDIES DEPARTMENT OF CHEMISTRY

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

November 1997

© Philip Edward Johnson, 1997
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 Chemistry
The University of British Columbia
Vancouver, Canada

Date 22-1-98
Abstract

This thesis describes the biophysical and structural characterization of CBD$_{N1}$, the amino-terminal cellulose-binding domain (CBD) from the *Cellulomonas fimi* $\beta$-1,4-glucanase CenC. Heteronuclear multidimensional nuclear magnetic resonance (NMR) methods were used to determine the three dimensional structure of CBD$_{N1}$ in the presence of saturating amounts of cellotetraose. It was found that CBD$_{N1}$ is composed of 10 $\beta$-strands, folded into two antiparallel $\beta$-sheets with the topology of a jelly-roll $\beta$-sandwich. The dominant feature of the CBD$_{N1}$ structure is a cleft which runs along the length of one face of the molecule.

On the basis of perturbations of the NMR spectrum of CBD$_{N1}$ due to the addition of sugar, it was shown that CBD$_{N1}$ binds soluble cellobiooligosaccharides within the cleft. Intermolecular nuclear Overhauser enhancements (NOEs) between residues in the cleft and the bound sugar confirmed that this face is responsible for ligand binding. Association constants, determined from the dependence of the amide $^1$H and $^{15}$N chemical shifts upon added sugar, were found to increase with increasing sugar length, reaching a maximum at a ligand length of five glucose units. This corresponds approximately to the length of the binding cleft. The binding cleft is composed of a strip of hydrophobic residues, flanked by hydrophilic residues. It is proposed that the pyranose rings of the sugar lie over the hydrophobic residues while the polar side-chains are involved in hydrogen bonding interactions with the equatorial hydroxyl groups of the glucose rings. CBD$_{N1}$ binds a calcium ion at a site opposite the oligosaccharide binding face. The binding affinity of oligosaccharides is unaffected by the presence of calcium.

CBD$_{N1}$ binds nitroxide spin-labelled oligosaccharides in multiple orientations. In one orientation the spin-label group lies near residue alanine 18, in the other near glycine 86. This ability to bind the same ligand in several orientations indicates that different hydrogen bonding combinations between the sugar and protein must occur. This is likely related to the motional disorder observed for residues present in the binding face as determined by NMR relaxation methods.
# Table of Contents

Abstract ........................................................................................................... ii

Table of Contents .......................................................................................... iii

List of Tables ................................................................................................... vii

List of Figures ................................................................................................ viii

Abbreviations ................................................................................................. xii

Acknowledgments ........................................................................................... xiv

## Chapter 1- Cellulose Binding Domains: Structures and Binding Domains ................. 1

Introduction ..................................................................................................... 1

CBDs that bind crystalline cellulose ................................................................. 5
  Structures .................................................................................................... 5
  Identification of binding face .................................................................... 6
  (i) CBD<sub>BHI</sub> ..................................................................................... 6
  (ii) CBD<sub>Cex</sub> ................................................................................... 8
  Mechanism for binding crystalline cellulose ............................................. 9
  Structural considerations of CBDs binding to crystalline cellulose .......... 10

CBDs that do not bind crystalline cellulose ...................................................... 13
  Structure .................................................................................................. 13
  Binding specificity and thermodynamics of CBD<sub>N1</sub> .............................. 16

Summary ......................................................................................................... 17

## Chapter 2- Cellooligosaccharide Binding by CBD<sub>N1</sub> ........................................... 18

Abstract ......................................................................................................... 18

Introduction ..................................................................................................... 19
  Background ............................................................................................... 19
  Ligand binding studied by NMR .............................................................. 19

Experimental methods .................................................................................. 21
  Expression and Purification of CBD<sub>N1</sub> ................................................. 21
  Characterization of CBD<sub>N1</sub> ............................................................... 26
  CD and FTIR spectroscopy ...................................................................... 26
  NMR Spectroscopy ................................................................................... 27
  Titration of CBD<sub>N1</sub> with Cellooligosaccharides Monitored by NMR ....... 28
  NMR studies of reduced CBD<sub>N1</sub> ......................................................... 29
Experimental Methods .......................................................... 102
Sample preparation ............................................................. 102
NMR spectroscopy .............................................................. 103
Determination of binding constants ........................................ 104
Analysis of binding kinetics .................................................. 105
Structure calculations of calcium-loaded CBDN1 ...................... 105

Results ................................................................................. 106
CBDN1 binds metal ions ....................................................... 106
Calcium equilibrium binding association constant and binding stoichiometry ........................................ 106
Effect of calcium on cellooligosaccharide affinity ....................... 112
Kinetics of calcium binding .................................................... 114
Identification of calcium binding site ....................................... 115

Discussion ........................................................................... 121
Location of the metal binding site and structure of calcium-loaded CBDN1 ........................................ 121
Binding affinity and kinetics ................................................. 124
A structural role for calcium binding by CBDN1 ...................... 126

Chapter 5- Dynamic Analysis of Ligand Binding ...................... 128

Abstract ................................................................................ 128

Introduction ......................................................................... 129
Background ........................................................................... 129
15N relaxation and dynamics ................................................ 130
Methyl group deuteron relaxation and dynamics ...................... 134

Experimental Methods ......................................................... 136
Sample preparation for 15N relaxation experiments .................... 136
Preparation of fractionally deuterated CBDN1 ......................... 136
15N relaxation experiments ..................................................... 137
Methyl group deuteron relaxation experiments ......................... 139
Data processing ...................................................................... 140
Resonance assignment of the calcium-loaded CBDN1 ................. 142
Ultracentrifugation and viscosity measurements ....................... 143

Results .................................................................................. 144
15N T1, T2 and NOE values ...................................................... 144
Estimation of the overall correlation time \(\tau_m\) ......................... 150
T1/T2 analysis ....................................................................... 151
Model-free analysis of backbone 15N relaxation data ................ 153
Ultracentrifuge and viscosity measurements ............................ 157
Resonance assignment of calcium-loaded CBDN1 .................... 157
Methyl deuteron relaxation and analysis ................................... 159

Discussion ............................................................................ 169
Comments on the model-free analysis of 15N relaxation ............. 169
Global trends ........................................................................ 170
Effect of calcium binding ...................................................... 173
Effect of cellooligosaccharide binding on the dynamics of CBDN1 .................................................. 176
i) Backbone $^{15}\text{N}$ relaxation data ........................................... 176  
ii) Methyl containing deuteron relaxation ...................................... 178  
Changes in $\tau_m$ with viscosity ...................................................... 181

Chapter 6- Structural Analysis of Ligand Binding .................................. 183  

Abstract .............................................................................................. 183  

Introduction .......................................................................................... 184  
Background ........................................................................................... 184  
Theory on the use of spin labels ............................................................. 184  

Experimental methods .......................................................................... 187  
Spin labelled celloseligosaccharides ....................................................... 187  
Preparation of CBD$_{N1}$ samples and titration of nitrooxide-labelled celloseligosaccharides ................................................................. 187  
NMR spectroscopy of nitrooxide-labelled celloseligosaccharide-CBD$_{N1}$ complexes ................................................................. 188  
Data analysis and calculation of $\Delta(R_1)$ and $\Delta(R_2)$ ...................... 189  
Titration of mutant CBD$_{N1}$ protein samples monitored by NMR .......... 190  

Results .................................................................................................... 191  
Binding of the nitrooxide-labelled celloseligosaccharides to CBD$_{N1}$ ...... 191  
CBD$_{N1}$ binds nitrooxide-labelled celloseligosaccharides in two distinct orientations ................................................................. 192  
Measurement of $\Delta R_1$ and $\Delta R_2$ values ......................................... 197  
Calculation of $\tau_c$ values and electron-proton distances ...................... 201  
Celloseligosaccharide binding of CBD$_{N1}$ variants .................................. 204  

Discussion ............................................................................................. 207  
Binding affinity of the nitrooxide-labelled celloseligosaccharides .......... 207  
Measurement of the relaxation parameters $\Delta R_1$ and $\Delta R_2$ and calculation of $\tau_c$ and $r$ values ................................................................. 207  
Ligand binding by CBD$_{N1}$ mutants ..................................................... 211  
Changes in the line-width of binding face residues in the CBD$_{N1}$ mutants ................................................................. 212  
Implications for the association of CBD$_{N1}$ with celloseligosaccharides .... 214

Appendix .................................................................................................. 220

Bibliography ............................................................................................ 227
List of Tables

Table 1.1  Biochemical data for CBDs of known structure ...........................................5

Table 2.1  Association constants for the binding of cellobiose saccharides by CBDN1 ......38

Table 3.1  Acquisition and processing parameters for NMR experiments recorded on CBDN1 for use in the structure determination process .................................62

Table 3.2  Coupling constants and \( \chi_1 \) assignments for valine residues in CBDN1 ....74

Table 3.3  Coupling constants and \( \chi_1 \) assignments for isoleucine and threonine residues in CBDN1 .................................................................75

Table 3.4  Structural statistics and rms differences .........................................................85

Table 5.1  Description of samples used in \( ^{15} \)N relaxation experiments ..................138

Table 5.2  Summary of model-free parameters ......................................................142

Table 5.3  Acquisition and processing parameters for NMR experiments used to assign calcium-loaded CBDN1 ..............................................................143

Table 5.4  Number of residues per sample for which \( ^{15} \)N relaxation data was obtained ....147

Table 5.5  Values for the overall correlation time (\( \tau_m \)) determined for CBDN1 .........151

Table 5.6  Summary of the number of resonances occurring for each spectral density model used to fit \( ^{15} \)N \( T_1 \), \( T_2 \) and \( ^1H\)-\( ^{15} \)N NOE data for the different CBDN1 samples ........................................153

Table 5.7  Average \( S^2_{axis} \) values calculated for each methyl type in CBDN1 ..........166

Table 5.8  Comparison of the spectral density models used to fit the experimental data with original and introduced minimum errors .............................................170

Table 6.1  Association constants for CBDN1 binding TEMPO-labelled cellobiose saccharides ...........................................................................192

Table 6.2  Order of disappearance of residues in \( ^1H\)-\( ^{15} \)N HSQC spectra induced by TEMPO-labelled cellotetraose binding .................................................195

Table 6.3  Order of disappearance of residues in \( ^1H\)-\( ^{15} \)N HSQC spectra induced by TEMPO-labelled cellotriose binding .................................................196
## List of Figures

| Figure 1.1 | Cellulases produced by *Cellulomonas fimi* | 2 |
| Figure 1.2 | Ribbon diagrams of CBDs of known structure | 4 |
| Figure 1.3 | Surface representation of CBDs | 7 |
| Figure 1.4 | Schematic view of the cross section of the structure of cellulose | 11 |
| Figure 1.5 | Widths of the binding faces of the CBDs that bind to crystalline cellulose | 14 |
| Figure 2.1 | Plasmid map of pTugN1n | 22 |
| Figure 2.2 | SDS PAGE gel showing the overexpression of CBD\(_N_1\) | 24 |
| Figure 2.3 | SDS PAGE gel showing the purification of CBD\(_N_1\) on Avicel | 25 |
| Figure 2.4 | FTIR spectrum of CBD\(_N_1\) | 30 |
| Figure 2.5 | CD spectra of CBD\(_N_1\) | 32 |
| Figure 2.6 | Titration of CBD\(_N_1\) with cellotetraose monitored by \(^1\text{H}\text{-}^{15}\text{N}\) HSQC spectra | 34 |
| Figure 2.7 | Overlaid fits for CBD\(_N_1\) binding cellotetraose | 35 |
| Figure 2.8 | Association constants for CBD\(_N_1\) binding cellobiose monitored by \(^1\text{H}\text{-}^{15}\text{N}\) HSQC spectra | 36 |
| Figure 2.9 | Pattern of \(^{15}\text{N}\) and \(^{1}\text{H}\) chemical shift perturbation with cellotetraose binding | 40 |
| Figure 2.10 | Schematic fold of CBD\(_N_1\) | 42 |
| Figure 2.11 | Aromatic region of DQF-COSY spectra of CBD\(_N_1\) in the presence and absence of cellotetraose | 44 |
| Figure 2.12 | Histogram summarizing the perturbation of aromatic ring \(^1\text{H}\) chemical shifts with cellotetraose binding | 45 |
| Figure 2.13 | \(^1\text{H}\text{-}^{15}\text{N}\) HSQC spectra of native CBD\(_N_1\) and CBD\(_N_1\) denatured by the reduction of the disulphide and by urea | 47 |
| Figure 3.1 | Flowchart of the structure determination process | 54 |
| Figure 3.2 | Three-dimensional heteronuclear experiments used to assign the resonances of CBD\(_N_1\) | 56 |
| Figure 3.3 | HNCACB and CBCA(CO)NH spectra demonstrating the sequential assignment of CBD\(_N_1\) | 57 |
Figure 3.4  (H)C(CO)NH-TOCSY and H(CCO)NH-TOCSY spectra demonstrating the assignment side-chain resonances of CBDN1  58
Figure 3.5  Labelled 1H-15N HSQC spectrum of cellotetraose-bound CBDN1  68
Figure 3.6  Spectra of the side-chain NH2 assignment process  70
Figure 3.7  Portion of the constant time 1H-13C HSQC spectrum and 1H-13C HSQC spectrum of CBDN1 fractionally 10% 13C labelled  73
Figure 3.8  Strip plot of the long range 13C-13C coupling experiment  76
Figure 3.9  Summary of data used to determine the secondary structure of CBDN1  78
Figure 3.10  Cross strand NOE s between the β-strands of sheets A and B in CBDN1  80
Figure 3.11  Strip plot of the 15N HSQC-NOESY spectrum of cellotetraose-bound CBDN1  81
Figure 3.12  Schematic diagram showing the fold of CBDN1  83
Figure 3.13  Cα traces of the ensemble of 25 structures of CBDN1  86
Figure 3.14  Structural statistics for CBDN1  87
Figure 3.15  Ramachandran plot of the mean minimized CBDN1 structure  88
Figure 3.16  Superposition of the side-chains of residues that lie in the hydrophobic core of CBDN1  89
Figure 3.17  Diagram of CBDN1 showing the side-chains present on the oligosaccharide binding face  91
Figure 3.18  Alignment of the amino acid sequences of family IV CBDs  96
Figure 3.19  Ribbon diagrams of CBDN1 and the hybrid Bacillus 1,3-1,4-β-glucanase  98
Figure 3.20  Superposition of CBDN1 and the hybrid Bacillus 1,3-1,4-β-glucanase showing the catalytic residues of the glucanase and the corresponding residues in CBDN1  100

Figure 4.1  Screen for metal binding using one-dimensional 1H NMR  107
Figure 4.2  Titration of CBDN1 with calcium monitored by one-dimensional 1H NMR  108
Figure 4.3  Titration of CBDN1 with calcium monitored by 1H-15N HSQC spectra  109
Figure 4.4  Determination of Kd values for calcium-binding by CBDN1  111
Figure 4.5  Determination of Kd values for cellopentaose binding by calcium-loaded CBDN1  113
| Figure 4.6 | Comparison of experimental and simulated lineshapes to obtain the $k_{off}$ value for calcium association/dissociation | 116 |
| Figure 4.7 | Pattern of $^{15}$N and $^1$H chemical shift perturbation with calcium binding | 117 |
| Figure 4.8 | C\(\alpha\) worm diagram of CBD\(_{N1}\) showing the location of residues that experience the most significant changes in chemical shift with calcium binding | 118 |
| Figure 4.9 | Calcium titration of CBD\(_{N1}\) monitored by $^1$H-$^{13}$C HSQC spectra | 120 |
| Figure 4.10 | Structure of *Bacillus macerans* 1,3-1,4-\(\beta\)-glucanase showing the location of the calcium and ligating groups | 123 |
| Figure 4.11 | Structures of CBD\(_{N1}\) calculated with and without a calcium ion | 125 |
| Figure 5.1 | Plots of $T_1$, $T_2$, and heteronuclear NOE vs. $\tau_m$ | 132 |
| Figure 5.2 | Portion of spectra used to measure $T_2$ values | 145 |
| Figure 5.3 | Two-parameter monoexponential curve fitting to obtain $T_1$ and $T_2$ values | 146 |
| Figure 5.4 | $^{15}$N $T_1$, $T_2$ and $^1$H-$^{15}$N NOE data for calcium-bound CBD\(_{N1}\) in the absence and presence of cellopentaose | 148 |
| Figure 5.5 | $^{15}$N $T_1$, $T_2$ and $^1$H-$^{15}$N NOE data for CBD\(_{N1}\) in the absence and presence of calcium | 149 |
| Figure 5.6 | $^{15}$N $T_1/T_2$ values for CBD\(_{N1}\) | 152 |
| Figure 5.7 | Model-free parameters for calcium-bound CBD\(_{N1}\) in the absence and presence of cellopentaose | 155 |
| Figure 5.8 | Model-free parameters for CBD\(_{N1}\) in the absence and presence of calcium | 156 |
| Figure 5.9 | Values of viscosity and $t_m$ plotted against protein concentration | 158 |
| Figure 5.10 | $^1$H-$^{15}$N HSQC spectrum of calcium-loaded cellooligosaccharide-free CBD\(_{N1}\) | 160 |
| Figure 5.11 | Spectrum illustrating the $^{15}$N-$^1$H correlations of residues following methyl containing amino acids | 161 |
| Figure 5.12 | $^1$H-$^{13}$C HSQC spectrum of calcium-loaded cellooligosaccharide-free CBD\(_{N1}\) | 162 |
| Figure 5.13 | Cellopentaose titration of calcium-loaded CBD\(_{N1}\) monitored by $^1$H-$^{13}$C HSQC spectra | 163 |
| Figure 5.14 | Portion of spectra used to measure the decay rates of $I_zC_zD_z$ magnetization | 164 |
| Figure 5.15 | Examples of the decay of $I_zC_zD_z$, $I_zC_zD_y$ and $I_zC_z$ magnetization | 165 |
| Figure 5.16 | $T_1(D)$ and $T_{1p}(D)$ values of methyl groups | 167 |
Figure 5.17 Order parameters for methyl groups ..........................................................168
Figure 5.18 Comparison between rms deviation of the ensemble of 25 structures and 
$^{1}H-^{15}N$ NOE values .............................................................................................172
Figure 5.19 Structure of CBD$_{N1}$ showing methyl groups present in the 
oligosaccharide-binding face ......................................................................................179

Figure 6.1 $^{1}H-^{15}N$ HSQC spectra of CBD$_{N1}$ bound to TEMPO-labelled cellotetraose .....193
Figure 6.2 Schematic view of the structure of CBD$_{N1}$ .........................................................194
Figure 6.3 $\Delta R_1$, $\Delta R_2$ and $\Delta LW$ values resulting from the binding of TEMPO-labelled 
cellotriose .......................................................................................................................198
Figure 6.4 $\Delta R_1$, $\Delta R_2$ and $\Delta LW$ values resulting from the binding of TEMPO-labelled 
cellotetraose .....................................................................................................................199
Figure 6.5 $\alpha^\alpha$ worm diagram of CBD$_{N1}$ showing the location of residues with the largest 
$\Delta R_1$, $\Delta R_2$ and $\Delta LW$ values with TEMPO-labelled cellotetraose binding ............200
Figure 6.6 $\tau_c$ values for the electron-proton vector ............................................................202
Figure 6.7 Values for the electron-proton distance (r) ..........................................................203
Figure 6.8 $^{1}H-^{15}N$ HSQC spectra of wild-type CBD$_{N1}$ and the variants Y19A and Y85A ..205
Figure 6.9 $^{1}H-^{15}N$ HSQC spectra from the titration of Y85A with cellopentaose ...............206
Figure 6.10 Possible orientations of TEMPO-labelled cellotriose bound to CBD$_{N1}$ ............209
Figure 6.11 Model of the cellopentaose-CBD$_{N1}$ complex ...................................................211
Figure 6.12 Orientations of cellotetraose showing the location of oxygen atoms ...................215
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMCC</td>
<td>bacterial microcrystalline cellulose.</td>
</tr>
<tr>
<td>CBD</td>
<td>cellulose-binding domain.</td>
</tr>
<tr>
<td>CBD_{CBHI}</td>
<td>cellulose-binding domain from <em>Trichoderma reesei</em> cellobiohydrolase 1.</td>
</tr>
<tr>
<td>CBD_{Cex}</td>
<td>cellulose-binding domain from <em>Cellulomonas fimi</em> xylanase-glucanase Cex.</td>
</tr>
<tr>
<td>CBD_{Cip}</td>
<td>cellulose-binding domain from <em>Clostridium thermocellum</em> scaffoldin subunit Cip B.</td>
</tr>
<tr>
<td>CBD_{E4}</td>
<td>cellulose-binding domain from <em>Thermomonospora fusca</em> endo/exocellulase E4.</td>
</tr>
<tr>
<td>CBD_{EGZ}</td>
<td>cellulose-binding domain from <em>Erwinia chrysanthemi</em> cellulase EGZ.</td>
</tr>
<tr>
<td>CBD_{N1}</td>
<td>N-terminal cellulose-binding domain from <em>Cellulomonas fimi</em> β-1-4-glucanase CenC.</td>
</tr>
<tr>
<td>CBD_{N2}</td>
<td>cellulose-binding domain from <em>Cellulomonas fimi</em> β-1,4-glucanase CenC following CBD_{N1} in sequence.</td>
</tr>
<tr>
<td>CBD_{N1N2}</td>
<td>the tandem cellulose-binding domains from <em>Cellulomonas fimi</em> β-1,4-glucanase CenC.</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism.</td>
</tr>
<tr>
<td>CT-HSQC</td>
<td>constant-time heteronuclear single quantum correlation.</td>
</tr>
<tr>
<td>DQF-COSY</td>
<td>double quantum filtered correlation spectroscopy.</td>
</tr>
<tr>
<td>DSS</td>
<td>2,2-dimethyl-2-silapentane-5-sulphonic acid, sodium salt.</td>
</tr>
<tr>
<td>DTNB</td>
<td>5,5'dithio-bis(2-nitrobenzoic acid).</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol.</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediamine tetraacetate.</td>
</tr>
<tr>
<td>FITC</td>
<td>fluorescein isothiocyanate.</td>
</tr>
<tr>
<td>FTIR</td>
<td>fourier transform infra red.</td>
</tr>
<tr>
<td>HMQC</td>
<td>heteronuclear multiple quantum correlation.</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum correlation.</td>
</tr>
<tr>
<td>INEPT</td>
<td>insensitive nuclei enhanced by polarization transfer</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-thiogalactopyranoside.</td>
</tr>
</tbody>
</table>
J coupling constant.
NMR nuclear magnetic resonance.
NOE nuclear Overhauser effect.
NOESY nuclear Overhauser effect spectroscopy.
PASA phosphoric acid-swollen Avicel.
PASC phosphoric acid-swollen cellulose.
pH* the observed pH meter reading without correction for isotope effects.
ppm parts per million.
rms root mean square.
rmsd root mean square deviation.
SDS sodium dodecylsulphate.
sw spectral width.
TEMPO 2,2,6,6-tetramethylpyrrolidine-1-oxyl.
TOCSY total correlation spectroscopy.
Acknowledgments

I would like to thank my supervisor, Lawrence McIntosh, for all his great ideas and help during the last four and a half years. Many thanks also go to Peter Tomme who first showed me around a biochemistry lab, and provided invaluable assistance throughout my doctoral studies. I also acknowledge the help of the other people I have collaborated with on this project; Tony Warren, Doug Kilburn, Jeff Kormos, Louise Creagh, Chip Haynes, Steve Withers and Lloyd McKenzie. Many thanks go to Lewis Kay for providing his great NMR expertise and most of the pulse sequences I used. I also thank past and present members of the McIntosh lab for the help they have given me; especially Logan Donaldson for his computer expertise and willingness to rock early in the morning, Manish Joshi and Koman Joe for their help on the CBD_{N1} project, and Emmanuel Brun for enlightening discussions about CBDs. For helping to make the past five years so enjoyable I thank Claire, and for encouragement through all of my years at university I thank my Mom. Most of all I would like to acknowledge my wife, Anne, for all the support and encouragement she has given me.
Chapter 1

Cellulose-Binding Domains: Structures and Binding Mechanisms

Introduction

Enzymes involved in the degradation of cellulose generally have a modular architecture with binding to the substrate mediated by one or more cellulose-binding domains (CBDs) (figure 1.1). Currently, over 180 different putative CBD sequences have been identified and classified into 13 families based on their sequence similarities (Tomme et al., 1995; Dr. Peter Tomme, pers. comm.). Only a small fraction of these sequences have been shown experimentally to bind cellulose, but the binding ability of representatives from each of these families have been demonstrated.

Although chemically homogeneous, cellulose is a structurally complex and heterogeneous substrate and poses a challenge for enzymes to hydrolyze. In nature, cellulose exists in several crystalline forms, as well as in disordered or "amorphous" states. Paralleling this heterogeneity, CBDs from various families exhibit a range of specificities and affinities toward the various allomorphs of cellulose. Most notably, CBD_{N1} binds only amorphous cellulose (and soluble β-1,4-glucans, including cellooligosaccharides), whereas the other CBDs of known structure bind preferentially to crystalline cellulose. Calorimetric studies reveal that the binding of CBD_{N1} to amorphous cellulose is enthalpically-driven (Tomme et al., 1996a), whereas the association of CBD_{Cex} with crystalline cellulose is favored entropically (Creagh et al., 1996).

Cellulose is composed of repeating units of β-1,4-linked glucose and is the primary structural component of plant cell walls. It is the most abundant natural polymer on earth, with an estimated
Figure 1.1. Cellulases produced by *Cellulomonas fimi*. Each cellulase has a multi-domain architecture. Roman numeral after the CBDs indicates which family that it belongs to.
4 x 10^{12} kg biosynthesized and degraded annually (Coughlan, 1990). As a result of the abundance and low cost of cellulose, a number of potential biotechnology applications for CBDs have been devised. CBDs have been used as affinity tags for protein purification in batch, on a column or in an aqueous two phase system (Ong et al., 1989; Tomme et al., 1996b). CBD fusion proteins can also be used to immobilize proteins on a cellulose matrix with the activity of the heterogeneous protein often comparable to that of native protein.

The binding of CBDs to cellulose also represents an excellent model system to study protein carbohydrate interactions. A number of structures of lectins, both free and complexed with oligosaccharide are known (Rini, 1995). But, in general, these only interact with a mono or disaccharide. In contrast CBDs interact with multiple sugar subunits of the oligosaccharide and thus possess a much larger binding site.

The CBD my research has focused on is the family IV N-terminal CBD from *Cellulomonas fimi* β-1,4-glucanase CenC (CBD_{N1}) (figure 1.1). CenC has a family 9 catalytic domain, and is a semi-processive enzyme with both endo- and exoglucanase activity. This enzyme works with inversion of configuration at the anomeric carbon and is most active on soluble sugars, though it also degrades Avicel and BMCC (Tomme et al., 1996c). A previous study has shown that CBD_{N1} has the unique ability to bind only phosphoric acid-swollen cellulose, and not crystalline cellulose (Coutinho et al., 1992). The goal of my studies was to determine the three-dimensional structure of CBD_{N1} to explain its binding selectivity.

When I started this research in 1993, the only published structure of a cellulose-binding domain was the 36 residue family I CBD from the fungus *Trichoderma reesei* cellobiohydrolase I (CBD_{CBHI}) (Kraulis et al., 1989) (figure 1.2). A preliminary structure of the family II CBD from *Cellulomonas fimi* xylanase/β-1,4-glucanase Cex (CBD_{Cex}) (Xu et al., 1995) was also known (figure 1.2). Over the course of this thesis, structures of three other CBDs have been determined. These are the family III CBDs from *Clostridium thermocellum* scaffoldin subunit Cip B (CBD_{Cip}) (Tormo et al., 1996) and *Thermomonospora fusca* endo/exocellulase E4 (CBD_{E4}) (Sakon et al., 1997), and the family V CBD from plant pathogenic bacterium *Erwinia chrysanthemi* cellulase
Figure 1.2. Ribbon diagrams of five CBDs with known structure. Shown in green are the aromatic rings and hydrophobic residues on the binding faces of the CBDs. Selected hydrophilic residues are in red. Figure drawn using Molscript (Kraulis, 1991) and Raster3D (Merritt & Murphy, 1994).
EGZ (CBD\textsubscript{EGZ}) (Brun et al., 1997) (figure 1.2). Complementing these structural studies are several detailed biochemical and functional studies, including the thermodynamics of CBD-cellulose interactions (Creagh et al., 1996; Tomme et al., 1996a). A summary of the biochemical and structural data for these CBDs is given in table 1.1. The aim of this chapter is to introduce to the reader the field of CBDs. This will serve as a background to my work, presented in chapters 2 to 6.

<table>
<thead>
<tr>
<th>Table 1.1. Biochemical data for CBDs of known structure.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBD\textsubscript{CBHI}</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>Family</td>
</tr>
<tr>
<td># Residues</td>
</tr>
<tr>
<td>Size (kDa)</td>
</tr>
<tr>
<td>Calculated pI</td>
</tr>
<tr>
<td>Location\textsuperscript{a}</td>
</tr>
<tr>
<td># Disulfides</td>
</tr>
<tr>
<td>Fold</td>
</tr>
<tr>
<td>Binding Site</td>
</tr>
<tr>
<td>Ca\textsuperscript{2+} binding</td>
</tr>
<tr>
<td>Parental Enzyme</td>
</tr>
</tbody>
</table>

\textsuperscript{a} C: carboxy terminus, N: amino terminus, I: internal.

CBDs that bind crystalline cellulose

Structures

The three-dimensional structures of representative members of four families of CBDs that bind crystalline cellulose have been determined by NMR or x-ray crystallography (figure 1.2). Each of these CBDs shares a common feature of being composed entirely of anti-parallel β-strands connected by turns and loops. The three β-strands of CBD\textsubscript{CBHI} fold into a wedge shape, while
Chapter 1-CBDs: Structures and Binding Mechanisms

CBD_{Cex} has a nine-stranded \( \beta \)-barrel fold. CBD_{Cip} is a jelly-roll \( \beta \)-sandwich, and CBD_{EGZ} is a boot-shaped molecule composed of a large loop region and five \( \beta \)-strands that form two \( \beta \)-sheets.

The dominant structural feature of these CBDs is the presence of three solvent-exposed aromatic rings aligned along a flat face of each molecule (figure 1.2, 1.3). In CBD_{CBHI}, these are three tyrosine side-chains (Tyr5, Tyr31, and Tyr32), while CBD_{Cex} has three tryptophan residues solvent-exposed (Trp17, Trp54, and Trp72). For CBD_{Cip}, these residues are two tryptophans and a histidine (Trp118, His57, and Tyr67), and in CBD_{EGZ} they are two tryptophans and a tyrosine (Trp18, Trp43, and Tyr44). A common theme in protein-carbohydrate interactions is the stacking of aromatic side-chains on sugar rings (Quiocho, 1986, 1989; Vyas, 1991). Although all four structures were determined in the absence of cellulose or cellulose-derivatives, evidence from site-directed mutagenesis and cellohexaose-binding studies strongly indicates that these aromatic residues, and others nearby, mediate the association of CBDs with crystalline cellulose.

Identification of binding face

(i) CBD_{CBHI}: Extensive studies of the effects of amino acid substitutions on the binding of the family I CBD_{CBHI} to cellulose have been reported (Reinikainen et al., 1992, 1995; Linder et al., 1995a, b). These studies show that the three solvent-exposed tyrosines (Tyr5, Tyr31, Tyr32), and a polar residue (Gln34) are important for retaining binding affinity. Mutating these residues to alanines greatly reduces binding to bacterial micro-crystalline cellulose (BMCC). The side-chains of these four residues form a strip across one face of CBD_{CBHI} (Figure 1.3). In contrast, the mutation of Asn29 to alanine reduced BMCC binding slightly (Linder et al., 1995a). This residue is found on the binding face, but off to the side of the strip formed by the tyrosines and Gln34 (figure 1.3). Changing Pro16 to Arg, present on the "rough" face of the protein, opposite to the tyrosine-containing face, has only a small effect on binding (Reinikainen et al., 1992).

The structures of all three tyrosine-to-alanine mutant CBD_{CBHI} proteins have been determined by NMR spectroscopy (Mattinen et al., 1997a). The three-dimensional structures of the Y31A and Y32A variants are very similar to that of wild type. These are also well defined, having a low rms
Figure 1.3. GRASP (Nicholls et al., 1991) surface representation looking down onto the binding face or cleft of each CBD. The CBDs that bind crystalline cellulose, CBD_{CBHI}, CBD_{Cex}, CBD_{Cip} and CBD_{EGZ}, all have flat binding faces with three exposed aromatic residues. CBD_{N1} does not bind crystalline cellulose and has a binding groove. Shown in yellow are the aromatic and hydrophobic residues (W, L, V, A, Y, F, H, I, P, M, C) with the hydrophilic residues in red (N, D, E, Q, R, S, T, K). Selected residues are labelled to orient the reader. The different CBDs are not drawn to scale relative to each other in this figure.
deviation amongst the ensemble of final structures. This indicates the elimination of binding by these mutations results directly from the removal of the tyrosine, and not indirectly by causing structural changes in the protein. In contrast, the structure of the Y5A variant is significantly altered from that of wild type. The ensemble of structures for this mutant have a higher rms deviation, especially in the N-terminal region near position 5. Consequently, it is difficult to separate the structural and functional roles of Tyr5 in CBDCBHI, except to state that its presence is required for cellulose binding.

Chemical shift perturbations of several other mutant CBDCBHI proteins relative to the wild-type have been reported (Linder et al., 1995a). The variants Q34A and N29A show little change in their chemical shifts, indicating that the structural effects of these mutations are minimal. In the case of the P16R variant, changes in chemical shift were significant, suggestive of conformational changes in this variant. However, any such structural changes have no effect on binding.

The importance of Tyr31 in ligand binding was also shown when the association of various T. reesei CBDs with cellohexaose was studied (Mattinen et al., 1997b). The NMR line widths of cellohexaose increased when bound weakly by the CBDs from cellobiohydrolase I and II and endoglucanase I. For the Y31A variant of CBDCBHI, no increase in the NMR line width was observed, indicating that this protein does not appreciably bind cellohexaose.

(ii) CBDCex: The involvement of the three solvent-exposed tryptophans of CBDCex in ligand binding is supported by chemical shift data, a chemical modification study, and by data for mutant forms of a related family II CBD. The $^{1}H$ and $^{15}N$ chemical shifts of the indole rings of Trp54 and Trp72 are perturbed upon the addition of cellohexaose to CBDCex (Xu et al., 1996). Both of these residues are solvent-exposed (figure 1.3). Additionally, when the equivalent of two of the three solvent-exposed tryptophans of CBDCex are oxidized by N-bromosuccinimide (NBS), binding to BMCC is essentially eliminated (Bray et al., 1996). Finally, the mutation of Trp68 in the family II CBD from C. fimi endoglucanase A (CBDCenA) to alanine reduces binding by 30% (Din et al., 1994). From sequence alignments (Tomme et al., 1995), this residue corresponds to Trp72 of CBDCex.
Together, evidence for CBD_{CBHI} and CBD_{Cex} indicates that only residues on the face containing the solvent-exposed aromatics are involved in cellulose-binding. Although experimental evidence is not available for CBD_{Cip} and CBD_{EGZ}, it can be safely assumed that similar results would be obtained.

**Mechanism for binding crystalline cellulose**

Thermodynamic data defining the mechanism of binding of CBD_{Cex} to insoluble BMCC were reported by Creagh et al. (1996). The dominant thermodynamic driving force of binding is a net increase in entropy (ΔS°). This is accompanied by a small decrease in enthalpy (ΔH°), and a large negative ΔC_p. The increase in entropy, as well as the negative ΔC_p, indicates a dehydration or hydrophobic effect taking place during cellulose binding.

A structural explanation for the increase in entropy is the exclusion of water when the three indole rings on the binding surface of CBD_{Cex} contact cellulose. This results in a smaller net number of ordered water molecules present in the system, as there is less net polar surface area present to hydrate. Some hydrogen bond formation between the CBD and cellulose can be envisioned to account for the small negative enthalpy of binding. An increase in entropy as the driving force for complex formation is different from the interactions of most other carbohydrate-binding proteins (Vyas, 1991). Though it remains to be experimentally proven, the presence of three exposed aromatic rings in the family I, III and V CBDs indicates that they likely share a similar binding mechanism with CBD_{Cex}.

It is expected that a tryptophan side-chain, being larger than a tyrosine, would cover more surface area of cellulose, and thereby displace more water molecules upon binding. This in turn would result in a larger increase in entropy and a higher affinity for cellulose. In support of this argument, the family I CBD from Endoglucanase 1 of T. reesei (CBD_{EGI}) has a significantly higher affinity for crystalline (tunicate) cellulose than does CBD_{CBHI}. A major difference between these two CBDs is the presence of a tryptophan residue at position 5 in CBD_{EGI}, as compared to a tyrosine in CBD_{CBHI}. The Y5W variant of CBD_{CBHI} has substantially higher binding affinity for
crystalline cellulose over wild-type, although not equivalent to CBD\textsubscript{EGI} (Linder \textit{et al}., 1995b). It would be possible to test further the effect of creating a larger surface area of dehydration by also looking at the binding of a Y31W variant of CBD\textsubscript{CBHI}. A double CBD\textsubscript{CBHI} variant, Y5W and Y31W, could also be created to see if an additive effect on binding affinity is observed.

In addition to the tryptophans, there are a number of hydrophilic residues on the binding face of CBD\textsubscript{Cex} which could be involved in the hydrogen bond formation to cellulose, as suggested by the thermodynamic data. These include Asn15, Gln52, Gln83 and Asn87 (figure 1.3). The hydrophilic character of these residues tends to be conserved among family II CBDs (Tomme \textit{et al}., 1995). Other CBDs also contain hydrophilic residues that punctuate the hydrophobic character of the binding face (figure 1.3). In some cases, these residues are thought to form intramolecular salt bridges, for example R112 and D56 in CBD\textsubscript{Cip} and K16 and D17 in CBD\textsubscript{EGZ}. These hydrophilic residues could also stabilise the orientation of the aromatic rings by forming hydrogen bonds to atoms on the aromatic ring.

None of the CBD structures determined at present contain a phenylalanine among the exposed aromatic rings on the binding face. From the sequence alignments (Tomme \textit{et al}., 1995) of families I, II, III and V, very few phenylalanines align with residues corresponding to their solvent-exposed aromatics. This suggests that an electronegative atom on the aromatic ring is important, enabling hydrogen bond formation either to the cellulose or to another residue on the binding face. Alternatively, a phenylalanine may have a detrimental effect on the solubility of the protein due to its hydrophobic character.

\textit{Structural considerations of CBDs binding to crystalline cellulose}

Cellulose is an insoluble polymer of \(\beta\)-1,4-linked glucopyranosyl units. Through interchain hydrogen bonds, cellulose strands of 100 to over 10 000 glucose units form fibers. These in turn wrap into bundles to form the main structural component of plant cell walls. Highly ordered regions of cellulose are referred to as crystalline cellulose, whereas regions of the cellulose crystal lattice that have been disrupted are termed amorphous cellulose. Cellulose that CBDs encounter in
nature exist in a spectrum of states ranging from highly crystalline to amorphous, and also contain other compounds, such as lignin and xylan.

A number of possible allomorphs of crystalline cellulose have been identified. In all forms, alternate glucose residues are rotated 180° relative to each other, making cellobiose the repeating subunit. The allomorphs differ primarily in the arrangement of hydrogen bonds between individual cellulose chains and sheets. It is proposed that neighboring chains can be oriented parallel (cellulose I) or antiparallel (cellulose II) to each other. In all forms, cellulose chains are hydrogen bonded together to form sheets that are held together by van der Waals interactions. The sheets of cellulose are staggered, resulting in a diamond arrangement with only the top layer, the (0,2,0) face, having its pyranose rings completely exposed (figure 1.4). A recent crystal structure of β-D-cellotetraose has been solved (Gessler et al., 1994; Gessler et al., 1995), providing an atomic resolution model of the cellulose II polymer.

Figure 1.4. Schematic view of the cross section of the structure of crystalline cellulose I according to Gardner and Blackwell (1974). This cross section shows the strands of cellulose coming out of the page.
A commonly used form of cellulose is bacterial microcrystalline cellulose (BMCC). This is a highly crystalline form of cellulose I prepared from *Acetobacter xylinum* (Gilkes et al., 1992). Cellulose that has significant amorphous character can be generated by treatment with phosphoric acid. This is referred to as phosphoric acid-swollen cellulose (PASC), or phosphoric acid-swollen avicel (PASA), depending on the starting material. A limitation of this compound is the inability to measure directly the amount of residual crystalline regions present. Without this measure, caution should be used in interpreting binding studies using PASC, as the CBD may interact selectively with residual crystalline regions and not in the disordered, or amorphous regions. Other types of cellulose such as Avicel, filter paper and cotton are used for the purification of CBDs, but are not very useful for quantitative binding studies due to their varied composition.

It remains to be established where CBDs bind on the cellulose crystal. The dehydration mechanism of binding, suggested by thermodynamic data, only indicates that water is displaced upon binding. Reinikainen et al. (1995) and Tormo et al. (1996) both propose that CBDs bind to the (0,2,0) face of cellulose (figure 1.4). The rings of the exposed aromatic side-chains of the CBDs are spaced to directly overlap the pyranose rings of a single cellulose chain. In a perfect cellulose crystal, this face is only one chain wide. Since the binding faces of the CBDs are not concave, this implies a CBD molecule would only make contact with one strand of cellulose and only residues in line with the aromatic rings would be able to interact with this strand. Residues off to the side of the binding face would likely not be able to make contact with cellulose without significant conformational change on binding.

In the binding model outlined in Tormo et al. (1996), CBDs are proposed to interact with three adjacent cellulose chains. Hydrophilic residues off to the side of the strip of aromatic residues act as anchor points for hydrogen bonding to cellulose. This idea of anchor points has yet to be supported by studies where the proposed anchor residues are mutated and the effect on binding measured. As mentioned earlier, the mutation N29A in CBD<sub>CBI</sub> has a small, but significant, effect on the BMCC binding ability of CBD<sub>CBI</sub> (Linder et al., 1995a). Asn29 is one of the proposed anchor points for CBD<sub>CBI</sub> (Tormo et al., 1996).
It is consistent with a binding mechanism involving the displacement of water, for CBDs to bind the cellulose crystal in regions other than where a flat planar array of cellulose strands is present. Binding could occur on any of the cellulose crystal faces shown in figure 1.4. Here the CBD could be aligned along the cellulose chain or perpendicular to it. As shown in figure 1.5 the binding faces of the CBDs are not all completely flat. CBD_{CBHI}, CBD_{Cex} and CBD_{EGZ} all have binding faces where hydrophilic residues lie above and or below the plane formed by the linear arrangement of aromatic rings. This slope of the binding site could enable the CBD to simultaneously interact with cellulose chains on multiple sheets. The aromatic rings on the binding face of the CBD could still align with the partially exposed pyranose rings of cellulose. Although smallest in molecular weight, CBD_{CBHI} has a binding face just as wide, if not wider, than those of CBD_{Cex} or CBD_{EGZ}. It is clear from figure 1.5 that the binding face of CBD_{Cip} is wider and flatter than the other CBDs. The width of the binding face of CBD_{Cip} may hinder its binding to regions of the cellulose crystal that are not also fairly flat.

Experiments on CBD_{Cex} labelled with fluorescein isothiocyanate (FITC) show that at low protein concentrations FITC-CBD_{Cex} binds uniformly onto BMCC (Creagh et al., 1996). At saturating concentrations FITC-CBD_{Cex} does concentrate in one area of BMCC. This is attributed to a region of high surface area, not a region of preferential binding. It has also been shown that FITC-CBD_{Cex} rapidly diffuses across the cellulose surface without becoming unbound from the cellulose crystal (Jervis et al., 1997).

CBDs that do not bind crystalline cellulose

Structure

The family IV CBDs CBD_{N1}, CBD_{N2}, and the native tandem CBD_{N1N2} are unique among CBDs in that they bind only to PASC and soluble β-1,4 linked glucans, and not to crystalline cellulose (chapter 2; Coutinho et al., 1992; Tomme et al., 1996a, Johnson et al., 1996a). The
Figure 1.5. Head-on view of the putative binding faces for the four CBDs that bind crystalline cellulose. These views are rotated 90° compared to those shown in figure 1.2. Shown are ribbon diagrams of the backbone with aromatic and hydrophobic residues on the putative binding faces shown in green, hydrophilic residues are shown in red. Superimposed on this is the solvent accessible surface area produced by GRASP.
structure of CBD$_{N1}$ is composed of two $\beta$-sheets each containing 5 $\beta$-strands. These two sheets then form a jelly-roll sandwich structure (figure 1.2) (chapter 3; Johnson \textit{et al.}, 1996b). In contrast with the CBDs that bind crystalline cellulose, the binding site of CBD$_{N1}$ is a cleft, or groove, that runs along one face of the molecule. The structure of CBD$_{N1}$ was determined in the presence of cellotetraose, but this ligand was not included in the structure calculations. Nevertheless, binding is known to occur in this cleft based on the perturbation of chemical shifts upon the addition of cellooligosaccharides (chapter 2; Johnson \textit{et al.}, 1996a) and the detection of NOEs between the protein and unassigned protons on the sugar (chapter 3; Johnson \textit{et al.}, 1996b). This presence of a groove readily explains the binding selectivity of this CBD. The cleft prevents the protein from interacting with the flat, rigid surface of crystalline cellulose. Instead, single cellulose chains, as might be encountered in the amorphous regions of cellulose, can fit into the binding cleft. A preliminary structure of CBD$_{N2}$ shows this protein has a very strong structural similarity with CBD$_{N1}$, and that its binding site is also a cleft (Dr. E. Brun, personal communication).

The structure of CBD$_{N1}$ resembles that of CBD$_{Cip}$, as both proteins have a jelly-roll fold. The structure of CBD$_{Cip}$ also has a groove running along one $\beta$-sheet face of the molecule. This region of the protein is not thought to be the site where binding to crystalline cellulose occurs. Instead it has been suggested that this tyrosine rich cleft might be involved in protein-protein interaction between CBD$_{Cip}$ and another module of the cellulosome complex (Tormo \textit{et al.}, 1996). This face could also be involved in protein-carbohydrate interactions as either a second binding site for cellulose, or for interaction with a glycosylated region of another protein (Tormo \textit{et al.}, 1996).

CBD$_{N1}$ and CBD$_{Cip}$ also share the ability to bind calcium ions, although not at similar locations in the structure. In both proteins the calcium-binding site is located away from the cellulose-binding sites. Calcium increases the stability of CBD$_{N1}$, but does not alter its ability to bind cellooligosaccharides (chapter 4).

A calorimetric study shows that CBD$_{N1}$ is a marginally stable protein, having a maximum stability of 33 kJ mol$^{-1}$ at 1 °C and pH 6.1 (Creagh \textit{et al.}, 1997). CBD$_{N1}$ can be reversibly
denatured by either thermal or chemical means. Such reversibility may be common among CBDs, many of which are purified by binding to Avicel, with a denaturing concentration of guanidinium hydrochloride used to elute the protein. Another structural feature of CBD_{NI} is that it contains a single disulphide bond. Reduction of this disulphide results in complete unfolding of the protein (chapter 2; Creagh et al., 1997). As shown in table 1.1 disulphides are commonly found in CBDs. However the reduction of the disulphide bond in CBD_{EGZ} does not cause unfolding of the protein, and reduced CBD_{EGZ} retains its cellulose-binding ability (Dr. E. Brun, personal communication).

**Binding specificity and thermodynamics of CBD_{NI}**

The binding of CBD_{NI} to soluble cellooligosaccharides provides an opportunity to characterize the interaction of this domain with well defined model ligands. In parallel with the structural studies of CBD_{NI}, a detailed calorimetric analysis of the binding of CBD_{NI} to soluble ligands and PASC was conducted (Tomme et al., 1996a). CBD_{NI} was found to bind long polymeric substrates, such as hydroxyethyl cellulose and barley- and oat-\(\beta\)-glucan, with equal affinity to that of cellohexaose and cellopentaose. For shorter substrates, affinity decreases with chain length. This maximum binding affinity for cellopentaose correlates well with the binding cleft being approximately the length of five glucose rings.

Oligosaccharide binding by CBD_{NI} is accompanied by a favorable enthalpic change (\(\Delta H^\circ\)), compensated in part by a decrease in entropy (\(\Delta S^\circ\)). This implies that a predominance of polar interactions, such as hydrogen bonding and van der Waals interactions, provide the primary driving force for binding. This is similar to other carbohydrate-binding proteins (Quiocho, 1986, 1989), and in sharp contrast with that of CBD_{Cex} where an increase in entropy is the thermodynamic driving force for the binding of BMCC.

This proposed binding mechanism is supported by the structural features of the binding cleft of CBD_{NI} (Figures 1.2, 1.3). The weak van der Waals interactions take place between the strip of hydrophobic residues (Val17, Tyr19, Val48, Leu77, Tyr85, Ala126) that span all five strands of the binding face, and the relatively non-polar pyranose rings of the glucose sugars. On both sides
of this hydrophobic strip are residues (Asn50, Gln124, Gln128, Asn81, Arg75, Asp90) which can form hydrogen bonds to the equatorial hydroxyl groups on the sugars.

Summary

Two distinct structural classes of CBDs exist. CBDs that bind crystalline cellulose have a binding face that contains three solvent-exposed aromatic rings. The proposed binding mechanism for these CBDs involves a displacement of the ordered water molecules from the surfaces of the protein and cellulose, thereby increasing the entropy of the system. For the family IV CBDs that do not bind crystalline cellulose, their structures are characterized by the presence of a binding groove. Single sugar chains fit into this groove and binding is mainly driven by hydrogen bond formation.

This thesis describes my studies on the family IV cellulose-binding domain CBDN1. In chapter 2, the means of expressing and purifying CBDN1 is outlined, and binding to cellooligosaccharides characterized. Chapter 3 details the structure of CBDN1, while chapter 4 examines the calcium binding properties of this protein. The dynamics of apo, calcium-bound and cellooligosaccharide-bound CBDN1, as studied by 15N and 2H relaxation, are presented in chapter 5. Chapter 6 examines how CBDN1 binds oligosaccharides.
Chapter 2

Cellooligosaccharide Binding by CBD_{N1}

Abstract

The N-terminal cellulose-binding domain (CBD_{N1}) from *Cellulomonas fimi* β-1,4-glucanase CenC binds amorphous but not crystalline cellulose. To investigate the structural and thermodynamic bases of cellulose binding, NMR spectroscopy was used in parallel with calorimetry (Tomme *et al.*, 1996a) to characterize the interaction of soluble cellooligosaccharides with CBD_{N1}. Association constants, determined from the dependence of the amide $^1$H and $^{15}$N chemical shifts of CBD_{N1} upon added sugar, increase from $180 \pm 60 \text{ M}^{-1}$ for cellotriose to $4200 \pm 720 \text{ M}^{-1}$ for cellotetraose, $34000 \pm 7600 \text{ M}^{-1}$ for cellopentaose, and an estimate of $50000 \text{ M}^{-1}$ for cellohexaose. This implies that the CBD_{N1} cellulose-binding site spans approximately five glucosyl units. Based on the observed patterns of amide chemical shift changes, the cellooligosaccharides bind along a five-stranded β-sheet that forms a concave face of the jelly-roll β-sandwich structure of CBD_{N1}. This β-sheet contains a strip of hydrophobic side-chains flanked on both sides by polar residues. NMR measurements demonstrate that tyrosine, but not tryptophan, side-chains may be involved in oligosaccharide binding. These results lead to a model in which CBD_{N1} interacts with soluble cellooligosaccharides, and by inference, single polysaccharide chains in regions of amorphous cellulose, primarily through hydrogen bonding to the equatorial hydroxyl groups of the pyranose rings. van der Waals stacking of the sugar rings against the apolar side-chains may augment binding. CBD_{N1} stands in marked contrast to previously characterized CBDs that absorb to crystalline cellulose via a flat binding surface dominated by exposed aromatic rings.
Introduction

Background

An absolute requirement for an NMR oriented study of protein structure is the availability of milligram quantities of material. For a protein the size of CBD$_{N1}$, 152 residues or 15 kDa, the production of isotopically $^{15}$N and $^{13}$C labelled protein is also essential. This chapter outlines the method for expressing and purifying samples of CBD$_{N1}$ that I used throughout my thesis work. Included in this chapter is the initial biophysical characterization of CBD$_{N1}$ performed using circular dichroism and fourier transform infrared spectroscopy. Work related to the stability of CBD$_{N1}$ to thermal and chemical denaturation is also presented.

Very early in this study I found that CBD$_{N1}$ binds short chains of $\beta$-1,4-linked glucose monomers called cellooligosaccharides. These sugars, which are chemically equivalent to cellulose, provide a well defined substrate for investigating sugar-binding by CBD$_{N1}$. The studies outlined in this chapter were performed before the resonance assignments of CBD$_{N1}$ were known. However, in this chapter, the results are interpreted with the knowledge of both the assignments and structure, which will be presented in chapter 3. Most of the work discussed in this chapter has been published (Johnson et al., 1996a). With the exception of the cloning and difference ultraviolet spectroscopy measurements, I performed all the work that led to this publication.

Ligand binding studied by NMR

Throughout this thesis, NMR spectroscopy is used to measure association binding constants ($K_a$). In this chapter, the binding of a series of cellooligosaccharides to CBD$_{N1}$ is investigated. In chapter 4 the association constants of calcium with CBD$_{N1}$ and calcium-loaded CBD$_{N1}$ with cellopentaose are determined. In chapters 5 and 6, results of CBD$_{N1}$ binding to spin-labelled cellooligosaccharides and the binding of CBD$_{N1}$ variants with cellooligosaccharides is presented. In all these studies, the same equations and strategy for determining equilibrium association binding constants was followed, and is described here.
For binding events monitored by NMR spectral changes that occur in the fast exchange limit of the NMR timescale, the fraction of protein in the bound form at each point \( i \) in the titration \((f_{bi})\) is calculated using the observed \(^1\)H or \(^{15}\)N chemical shift \((\delta_i)\) compared to that of the free \((\delta_f)\) and fully bound forms \((\delta_b)\) by equation (2.1):

\[
f_{bi} = \frac{(\delta_i - \delta_f)}{(\delta_b - \delta_f)} \tag{2.1}
\]

\(\delta_f\) is determined from the spectrum of the protein without added ligand and \(\delta_b\) obtained from the fitting routine. For a single binding site, the association constant \(K_a\) for the binding of free protein \((P_f)\) and free ligand \((L_f)\) to yield the bound complex \((PL)\) is given by equation (2.2):

\[
K_a = \frac{[PL]}{[P_f][L_f]} \tag{2.2}
\]

By definition, for each titration point, the fraction of protein in the bound form is given by equation (2.3), the total protein concentration \([P_t]\) by equation (2.4), and the total ligand concentration \([L_t]\) by equation (2.5).

\[
f_b = \frac{[PL]}{[P_f] + [PL]} \tag{2.3}
\]

\[
[P_t] = [P_f] + [PL] \tag{2.4}
\]

\[
[L_t] = [L_f] + [PL] \tag{2.5}
\]

Rearranging (2.4) and (2.5) using (2.2), followed by substitution into (2.3) results in equation (2.6).

\[
f_{bi} = \frac{K_a[L_f]_i - K_a f_{bi}[P_f]_i}{1 + K_a[L_f]_i - K_a f_{bi}[P_f]_i} \tag{2.6}
\]

Equation (2.6) is solved for \(f_{bi}\) using the real root of the quadratic equation, allowing equation (2.1) to be recast in terms of the experimental parameters \([P_f]_i\), \([L_t]_i\), \(\delta_f\), and \(\delta_i\) and the variables \(K_a\) and \(\delta_b\). The latter two terms are determined by the fitting routine. An initial estimate for \(\delta_b\) is supplied from the \(^1\)H or \(^{15}\)N chemical shift at the final titration point, and the concentration of protein at each point \(i\) is corrected for dilution resulting from the addition of ligand. For every study presented in this thesis the programme PLOTDATA (TRIUMF, UBC) was used for the non-linear least squared fitting of the chemical shift data versus total ligand concentration.
In all the quantitative studies of ligand association by CBD$_{N1}$, binding is monitored by changes in the two-dimensional $^1$H-$^{15}$N HSQC spectrum of uniformly $^{15}$N-labelled CBD$_{N1}$. However, the method presented here can be applied to any measurable property of the system that changes in an incremental manner with binding. For binding processes studied by NMR it is necessary that association between the ligand and protein is in the fast exchange limit on the NMR chemical shift timescale to apply this formalism. For processes that are in the slow exchange limit, $f_{bi}$ can be estimated from peak intensities of the free and bound forms of the protein.

**Experimental methods**

*Expression and Purification of CBD$_{N1}$*

The gene encoding the 152 residue CBD$_{N1}$ from *Cellulomonas fimi* $\beta$-1,4-glucanase CenC was obtained from Dr. Peter Tomme (Department of Microbiology, UBC). To achieve efficient overexpression in minimal media, Dr. Tomme recloned the gene encoding CBD$_{N1}$ from the pTZ vector used in previous studies (Coutinho *et al.*, 1992) into the high expression vector pTug (figure 2.1) (Graham *et al.*, 1995). This expression vector encodes a gene for kanamycin resistance. The gene encoding CBD$_{N1}$ is fused to the Cex leader peptide to transport the expressed protein into the periplasm of *E. coli*. The gene fragment encoding CBD$_{N1}$ used throughout this thesis is termed pTugNln.

The vector pTugNln was expressed in *E. coli* JM101 cells (Yanish-Perron *et al.*, 1983). Unlabelled protein was produced in liquid tryptone / yeast extract / phosphate medium (TYP) (Sambrook, 1989). Biosynthetically $^{15}$N labeled protein was prepared using M9 media (Miller, 1972) containing 1g/L 99% $^{15}$NH$_4$Cl (Cambridge Isotopes) and 1 g/L of 99% $^{15}$N labeled Isogro (Isotec Inc.) as the sole sources of nitrogen. The Isogro (algae extract) supplement is necessary for efficient growth of *E. coli* JM101 and protein production in M9 media. The CBD$_{N1}$ samples with selectively deuterated aromatic rings were obtained from a synthetic medium containing 100 mg/L
of L-δ₁,ε₂,ζ₁,2,η₂-[²H₅]tryptophan and 100 mg/L of either L-δ₁,2,ε₁,2,ζ-[²H₅]phenylalanine or δ₁,2,ε₁,2-[²H₄]tyrosine (Cambridge Isotope Laboratories and Isotec Inc.) (McIntosh et al., 1990; McIntosh & Dahlquist 1990). One litre bacterial cultures were grown in 1.8 L Fernbach flasks at 30 °C to an A₆₀₀ of 0.6, induced with 0.5 mM IPTG, and incubated with shaking for approximately 24 hours. The leader peptide, which is cleaved upon translocation, targets the expressed CBD₅₁ to the periplasm of E. coli. However, over this extended incubation period, CBD₅₁ leaks out into the culture supernatant (Ong et al., 1993).

To increase yield when preparing labelled protein, the cell pellet was also subjected to osmotic shock by resuspending the cells in 30 mM Tris-HCl (pH 8.0) buffer containing 20% sucrose. After 15 min, the cells were centrifuged, resuspended in 5 mM magnesium sulphate at 4
°C, stored on ice for a further 15 min, and finally repelleted. The supernatants from this procedure were combined with the culture supernatant for purification. Figure 2.2 is an SDS gel showing the expression of CBD\textsubscript{N1}, and also the presence of CBD\textsubscript{N1} in the osmotic shock fractions.

CBD\textsubscript{N1} is initially purified based on its affinity to a commercial type of cellulose called Avicel. Avicel is heterogeneous, containing regions of both crystalline and amorphous cellulose. To each litre of culture supernatant 35 g of dry Avicel PH-101 (Fluka Chemika) is added, and the resulting slurry is adjusted to 1 M sodium chloride and 50 mM potassium phosphate at pH 7. After standing at 4 °C for 4 hours, the Avicel is collected by vacuum filtration on a Whatman GF/A glass filter and washed with 250 mL of 1 M sodium chloride, 50 mM potassium phosphate buffer at pH 7. CBD\textsubscript{N1} is eluted from Avicel with 250 mL of distilled water. To increase the recovery of labelled protein, the initial filtrate and the salt wash solution were recombined with the Avicel and left overnight at 4 °C. The above procedure was repeated and the two protein fractions combined. Figure 2.3 shows fractions from the first purification procedure on Avicel. This figure shows that some protein remains in the flow through and salt wash, and not all the protein is eluted from Avicel, the amount lost here is reduced with a second binding to Avicel. It is also seen that CBD\textsubscript{N1} in the water elution fraction is very pure at this stage.

CBD\textsubscript{N1} is purified further by anion exchange FPLC at pH 5 and then pH 7 using Q Sepharose (Pharmacia). In both cases, CBD\textsubscript{N1} was eluted with a gradient of 0-1 M sodium chloride in 50 mM potassium phosphate buffer. Often CBD\textsubscript{N1} eluted from ion exchange columns in two highly overlapped, but distinct, peaks. Electrospray mass spectroscopy showed that both peaks have the same molecular mass. Also, both peaks bind cellooligosaccharides with equal affinity. I am unable to explain the cause of this difference in mobility on the ion exchange column.

Finally, the purified CBD\textsubscript{N1} was de-salted, exchanged into sample buffer, and concentrated by ultrafiltration using non-cellulose membranes (Filtron Technology Corp., Northborough, MA.). The yields of the unlabelled, \textsuperscript{15}N labelled, and selectively deuterated CBD\textsubscript{N1} samples were approximately 80, 25, and 30 mg/L of culture supernatant, respectively.
Figure 2.2. SDS PAGE gel showing the overexpression of CBD$_{N1}$. Most of the protein leaks out into the culture supernatant. Protein in the periplasm can be recovered by osmotic shock (sucrose and MgSO$_4$ fractions).
Figure 2.3. SDS PAGE gel showing the purification of CBD$_{N1}$ on Avicel.
Characterization of CBD$_{NI}$

Protein concentrations were measured using absorption spectroscopy. The molar absorptivity $\varepsilon_{280}$ of CBD$_{NI}$ was determined to be 21370 M$^{-1}$ cm$^{-1}$ (or 1.39 mL mg$^{-1}$ cm$^{-1}$) using the method of Edelhoch (1967), as reviewed by Gill & von Hippel (1989) and Pace (Pace et al., 1995). The small contribution of a disulfide was included in the calculations. This molar absorptivity value, which differs from that originally published by Coutinho et al. (1992), was confirmed by quantitative amino acid analysis. The average molar absorptivity determined from three amino acid analyses, each run in duplicate, agreed within 4% of that obtained using the Edelhoch method.

The molecular mass of the unlabelled CBD$_{NI}$ is 15425.3 ± 0.9 Da as measured by electrospray mass spectroscopy. This is in excellent agreement with the expected value of 15425.8 Da based on the sequence of CBD$_{NI}$, after post-translational cleavage of the secretory leader peptide and corrected for the presence of a disulfide bond. The disulfide bond was identified by the lack of reactivity of CBD$_{NI}$ with 5,5'-dithiobis-(2-nitrobenzoic acid), and subsequently confirmed by NMR $^{13}$C chemical shifts (Wishart & Sykes, 1994). Using Edman degradation, the ten N-terminal residues of CBD$_{NI}$ expressed in $E. coli$ were confirmed to be ASPIGEGTFD. This matches that of native CenC from $C. fimi$.

Based on sedimentation equilibrium measurements, CBD$_{NI}$ (0.19 mM) in 50 mM sodium chloride, 50 mM potassium phosphate, 0.02% sodium azide, pH 5.9 at 20 °C is a monomeric protein with an apparent mass of 14130 Da (assuming a partial specific volume of 0.73 cm$^3$/gm; Chervenka, 1969). No evidence of higher-order association was detected under the conditions employed for the ultracentrifugation analysis. This conclusion is supported qualitatively by the relatively narrow linewidths observed in the NMR spectra of CBD$_{NI}$.

$CD$ and $FTIR$ spectroscopy

Circular dichroism (CD) spectroscopy was performed on 0.2 mg mL$^{-1}$ samples of CBD$_{NI}$ in a buffer of 50 mM sodium chloride, 50 mM potassium phosphate (pH 5.9), 0.02% sodium...
azide in the absence and presence of a 20 fold molar amount of cellohexaose, or a 10 fold molar amount of barley-β-glucan. Spectra were acquired with a Jasco J-730 CD spectropolarimeter using a 0.1 cm pathlength jacketed quartz cell. The spectra were solvent subtracted and processed using Jasco software. The midpoint temperature at which the tertiary structure of CBD$_{N1}$ unfolds was obtained by monitoring the ellipticity of CBD$_{N1}$ at 204 nm as the temperature increased from 35 °C to 60 °C at a rate of 1 °C/minute.

Fourier transform infrared (FTIR) spectra of CBD$_{N1}$ using a ca. 1 mM sample of CBD$_{N1}$ in a buffer of 50 mM sodium chloride, 50 mM potassium phosphate (pH* 5.9), 0.02% sodium azide were acquired with a Perkin-Elmer Model 2000 instrument. 6 μL of protein solution was needed to fill the calcium fluoride cell of 6 μm pathlength used. A spectral range of 2200-1200 cm$^{-1}$ and a spectral resolution of 2 cm$^{-1}$ was used, 1000 scans were signal averaged for each spectrum. Post processing of the spectra involved subtracting the spectral contributions of the buffer and the water vapor, then baseline flattening.

**NMR Spectroscopy**

The CBD$_{N1}$ samples were exchanged into 50 mM sodium chloride, 50 mM potassium phosphate (pH* 5.9), 0.02% sodium azide, 10% D$_2$O/90% H$_2$O. NMR spectra were recorded on a Varian Unity 500 MHz spectrometer equipped with a pulse field gradient triple-resonance probe. All spectra were collected at 35 °C and processed using FELIX v2.30 (Biosym Technologies). $^1$H chemical shifts were referenced to an internal standard of DSS at 0.00 ppm, and $^{15}$N chemical shifts to external 2.9 M $^{15}$NH$_4$Cl in 1 M HCl at 24.93 ppm (Levy & Lichter, 1979).

$^1$H-$^{15}$N HSQC spectra were recorded using the enhanced sensitivity pulsed field gradient experiment of Kay *et al.* (1992). A selective water flip back pulse was incorporated to ensure minimum perturbation of the water magnetization (Zhang *et al.*, 1994). The assignments of the amide $^1$H and $^{15}$N resonances of CBD$_{N1}$ were obtained using triple resonance correlation experiments (Bax & Grzesiek, 1993). The $^1$H resonances from the aromatic rings were identified from DQF-COSY, TOCSY, and NOESY spectra recorded with samples of 1.9 mM unlabelled, 2.0
mM (δ1,2, ε1,2-[2H4]-Tyr and δ1, ε2, ζ1,2, η2-[2H5]-Trp)-labelled, and 1.9 mM (δ1,2, ε1,2, ζ-
[2H5]-Phe and δ1, ε2, ζ1,2, η2-[2H5]-Trp)-labelled CBDNI, alone and in the presence of 80 mM
cellotetraose. These assignments were confirmed using the (Hβ)Cβ(CγCδ)Hδ and
(Hβ)Cβ(CγCδCε)Hε experiments to connect the ring 1H resonances to the corresponding 13Cβ
resonance of each aromatic residue in uniformly 13C/15N enriched CBDNI (Yamakazi et al.,
1993). The protein samples were lyophilized from the above sample buffer and resolubilized in
99% D2O for several of these measurements. The NMR assignments and structural analysis of
CBDNI are described in chapter 3.

**Titration of CBDNI with Cellooligosaccharides Monitored by NMR**

The binding of soluble cellooligosaccharides to CBDNI at 35 °C and pH* 5.9 was
measured quantitatively using 1H-15N NMR spectroscopy. Stock weight per volume solutions of
cellotriose, cellotetraose, cellopentaose, and cellohexaose (Seikagaku Corp.) were prepared in the
identical buffer used for the CBDNI. The initial concentration of protein was 0.5 mM, except in
the case of the titration with cellotriose for which a 0.18 mM sample was used. Aliquots of the
sugar solution were added directly to the protein in a NMR tube and mixed using 1 mm inner
diameter PE-160 polyethylene tubing attached to a Gilson Pipetman. To avoid excessive dilution
of the CBDNI, additions of greater than approximately 1 mg of sugar were made by removing the
protein from the NMR tube and dissolving the solid oligosaccharide directly in the sample solution.
For each titration, 8 to 10 1H-15N HSQC spectra were recorded consecutively with increasing
concentrations of sugar. The spectra were measured in 1.5 hours with 1024 and 128 complex
points obtained in the 1H and 15N dimensions (spectral widths 6000 Hz and 1450 Hz), and
processed with zero filling to a final digital resolution of 2.93 and 1.42 Hz/point, respectively.

Equilibrium association constants were determined by non-linear least squares fitting of the
chemical shift data versus sugar concentration to the Langmuir isotherm describing the binding of
one ligand molecule to a single protein site as outlined in the introduction to this chapter. The
Chapter 2-Cellooligosaccharide Binding by CBD$_{N1}$

programme PLOTDATA (TRIUMF, UBC) was used for the analyses. In the case of cellohexaose, an estimate of $K_a$ was made by simulation of the binding isotherm.

**NMR studies of reduced CBD$_{N1}$**

A 1.6 mM sample of uniformly $^{15}$N labelled CBD$_{N1}$ was split into two portions. To one urea was added and its $^1$H-$^{15}$N HSQC spectrum was recorded (unfolded oxidised). The concentration of urea was 6.9 M. as measured by refractive index. A 100 fold excess of D,L-dithiothreitol (DTT) was added and another $^1$H-$^{15}$N HSQC spectrum was obtained (unfolded reduced). With the other portion of CBD$_{N1}$, a $^1$H-$^{15}$N HSQC spectrum was acquired (folded oxidised) then a 100 fold excess of DTT was added and a final $^1$H-$^{15}$N HSQC spectrum was obtained (unfolded reduced). As the reduction of a disulphide is via the thiolate anion of DTT, this reaction is slow at pH 6. To overcome this kinetic barrier the samples with added DTT were unfolded by heating to 90 °C. Upon cooling $^1$H-$^{15}$N HSQC spectra of the urea-free reduced sample was recorded. The validity of this approach was confirmed by repeating these experiments at pH 8 using a sample of unlabelled CBD$_{N1}$. At pH 8 heating was not necessary to rapidly obtain reduced CBD$_{N1}$.

**Results**

**CD and FTIR spectroscopy**

The stretching and bending vibrations found in the Amide I region (1700-1620 cm$^{-1}$) of an FTIR spectrum of a protein contain features that are indicative of secondary structure content (Surewicz et al., 1993). Figure 2.4 shows the FTIR spectrum and the second derivative of the spectrum for the Amide I region for CBD$_{N1}$. The peak at 1633 cm$^{-1}$ is indicative of β-sheet conformation. The absence of peaks at 1658 and 1550 cm$^{-1}$ which are indicative of the presence of α-helical conformation indicates that CBD$_{N1}$ does not contain helical regions.
Figure 2.4. FTIR spectrum of CBD$_{N1}$. In the second derivative (lower trace) of the Amide I region of the FTIR spectrum (upper trace) the peak at 1633 cm$^{-1}$ arises from β-sheet conformation. The absence of peaks at 1658 and 1550 cm$^{-1}$ which are indicative of the presence of α-helical conformation indicates that CBD$_{N1}$ does not contain helical regions.
Chapter 2 - Cellooligosaccharide Binding by CBD$_{N1}$

The circular dichroism effect observed for an amide chromophore in the far UV (190-250 nm) depends on its secondary structure (Johnson, 1988). The CD spectra of CBD$_{N1}$ free and cellohexaose and barley-β-glucan-bound are shown in figure 2.5. The presence of the observed trough at 220 nm is diagnostic of a protein that contains entirely β-sheets. The CD effect of α-helical conformations, local minima at 208 nm and 222 nm, is much stronger than for β-sheets. The absence of these two local minima strongly suggest the absence of α-helices in CBD$_{N1}$. Also, the close similarity of the CD spectra of both free and cellohexaose and barley-β-glucan-bound CBD$_{N1}$ indicates there is very little change in secondary-structure content upon sugar binding.

Figure 2.5 also shows CD melts of CBD$_{N1}$ free, as well as bound to cellohexaose and barley-β-glucan. Free protein thermally unfolds at 47.5 °C, cellohexaose-bound protein unfolds at 50.7 °C and barley-β-glucan-bound CBD$_{N1}$ unfolds at 50.6 °C. The higher unfolding temperature of the latter two indicates both cellohexaose and barley-β-glucan bind CBD$_{N1}$ stabilising the folded structure. A similar increase in the unfolding temperature of CBD$_{N1}$ when cellohexaose-bound was seen in one-dimensional NMR spectra (data not shown). Cooling of the thermally-denatured protein results in an identical CD spectrum to that found before denaturation. There is little difference in the CD spectrum and denaturation temperature of CBD$_{N1}$ that has been repeatedly unfolded. This indicates that the thermal unfolding of CBD$_{N1}$ is a reversible process. Thermodynamic characterisation of the stability of CBD$_{N1}$ and the effect of oligosaccharide binding is given in Creagh et al. (1997).

CBD$_{N1}$ Binds Soluble Cellooligosaccharides

The binding of CBD$_{N1}$ to cellotriose, cellotetraose, cellopentaose, and cellohexaose was detected initially by the observation of numerous changes in the 1H-NMR spectrum of the protein resulting from the addition of these soluble sugars (data not shown). In contrast, the spectrum of CBD$_{N1}$ remained unperturbed in the presence of cellobiose, indicating that the protein does not bind this disaccharide appreciably. Cellooligosaccharides longer than cellohexaose were not
Figure 2.5. (Top) Far UV CD spectra of free and oligosaccharide-bound CBD\textsubscript{NI}. The broad trough at 220 nm indicates the presence of $\beta$-sheet conformation. The close similarity of the CD spectra of both free and cellohexaose and barley-$\beta$-glucan-bound CBD\textsubscript{NI} indicates that there is very little change in secondary-structure content upon sugar binding. (Bottom) CD melts of CBD\textsubscript{NI} free, as well as bound to cellohexaose and barley-$\beta$-glucan. Free protein thermally unfolds at 47.5 °C, cellohexaose-bound protein unfolds at 50.7 °C and barley-$\beta$-glucan-bound CBD\textsubscript{NI} unfolds at 50.6 °C. The higher unfolding temperature of the latter two indicates that both cellohexaose and barley-$\beta$-glucan bind CBD\textsubscript{NI}, stabilising the folded structure.
investigated as these compounds are not commercially available and have limited solubility in aqueous buffers.

The interaction of CBD\textsubscript{N1} with cellotriose, cellotetraose, cellopentaose, and cellohexaose was quantified using two-dimensional \textsuperscript{1}H-\textsuperscript{15}N correlation spectroscopy. Figure 2.6 shows an overlay of ten \textsuperscript{1}H-\textsuperscript{15}N HSQC spectra recorded as uniformly \textsuperscript{15}N enriched CBD\textsubscript{N1} was titrated with cellotetraose. It is readily seen that many peaks, arising from the backbone amide groups of the protein, show significant changes in both \textsuperscript{1}H and \textsuperscript{15}N chemical shifts with the progressive addition of the sugar. In the case of each cellobiosaccharide investigated, the free and bound forms of the protein are in fast exchange on the NMR timescale, resulting in the observation of population weighted average chemical shifts throughout the titration series.

 Binding Stoichiometry

Cellotriose, cellotetraose, cellopentaose, and cellohexaose bind to CBD\textsubscript{N1} with a stoichiometry of one sugar molecule per protein molecule. This conclusion is supported by the following evidence. First, as exemplified in figure 2.7 for cellotetraose, all the amides with \textsuperscript{1}H and \textsuperscript{15}N chemical shifts that are perturbed by sugar binding show co-incident titration curves. This discounts the possibility of multiple binding sites on the CBD with differing affinities for the cellobiosaccharides. Given that CBD\textsubscript{N1} is monomeric in solution, it is unlikely to have two or more distinct binding sites with equal affinities for the sugar ligands. Therefore the simplest interpretation of the co-incident titration curves is that cellobiosaccharides bind to CBD\textsubscript{N1} at a single site and that each perturbed \textsuperscript{1}H-\textsuperscript{15}N group reports the same association event. Second, plots of CBD\textsubscript{N1} amide chemical shifts versus added cellopentaose or cellohexaose show a plateau at approximately equal concentrations of total sugar and protein, indicating a binding stoichiometry of 1:1 (figure 2.8). Cellotetraose and cellotriose do not exhibit such pronounced titration end points due to their lower affinities for CBD\textsubscript{N1}. However, all four sugars cause similar changes in the \textsuperscript{1}H-\textsuperscript{15}N spectrum of the labelled protein, strongly suggesting that each binds to CBD\textsubscript{N1} with the same stoichiometry and at the same site. Third, the titration curves measured for the four
Figure 2.6. Cellotetraose binds to CBD$_{N1}$. A portion of ten $^1$H-$^{15}$N HSQC spectra of uniformly $^{15}$N labelled protein in the presence 0, 0.04, 0.08, 0.16, 0.36, 0.51, 0.76, 1.26, 2.24, and 10.5 mM total cellotetraose are overlayed. The arrows indicate the directions in which the amide $^1$H-$^{15}$N peaks shift with added sugar. Although all $^1$H-$^{15}$N resonances have been assigned, the peaks in the crowded region of the spectrum are not labelled for clarity. Q80$^{ε2}$ designates the $^{15}$NH$^{ε2}$ of Gln80. The sample buffer was 50 mM sodium chloride, 50 mM potassium phosphate (pH* 5.9), and 0.02% sodium azide in 10% D$_2$O/90% H$_2$O at 35° C.
Figure 2.7. Cellotetraose binds to CBD$_{N1}$ at a single site. The co-incident plots of the normalized $H^N$ chemical shift changes for residues Tyr19, Gly44, Thr87, and Gly130 versus total added cellotetraose demonstrate that each amide group in CBD$_{N1}$ monitors the same binding event. The solid lines represent the titration isotherms obtained by fitting the observed data points (○) to the equation describing the association of cellotetraose and CBD$_{N1}$ to form a 1:1 protein-sugar complex. For clarity, the overlapping data and fits for the four residues are not labelled individually.
Figure 2.8. The association constants of CBD$_{N1}$ for soluble cellooligosaccharides were determined from titration curves monitored by $^1$H-$^{15}$N NMR spectroscopy. The amide H$^N$ chemical shift of Tyr19 is plotted as a function of the total concentration of added cellohexaose (circle), cellopentaose (square), cellotetraose (diamond), and cellotriose (star). The solid lines represent the best fit of the experimental data to the equilibrium equation describing binding to a single protein site, or in the case of cellohexaose, a simulated titration curve based on an estimation of the association constant. The arrow marking the plateau in the titration curves for cellopentaose and cellohexaose falls at the point where the total sugar concentration equals the total protein concentration (~0.5 mM CBD$_{N1}$) indicating a 1:1 binding stoichiometry.
cellooligosaccharides are adequately fit to the binding isotherm describing the simple equilibrium expressed by equation (2.2) (figures 2.7 and 2.8). Finally, this conclusion is confirmed by isothermal titration microcalorimetry (Tomme et al., 1996a).

**Association Constants Determined by NMR Spectroscopy**

The association constants ($K_a$) describing the interactions of cellotriose, cellotetraose, and cellopentaose to CBD$_{NI}$ were determined by non-linear least squares fitting of the chemical shift titration data to the binding isotherm for a protein with a single ligand recognition site. Figure 2.8 shows an example of the analysis of data measured for Tyr19, as well as a comparison of the binding curves for each cellooligosaccharide. Two $K_a$ values were determined independently for each amide showing a significant chemical shift change upon sugar binding, one using the data for the amide H$_N$ and the other for the $^{15}$N nucleus. In this manner, association constants were determined from titration curves measured for 10 to 16 amides in the protein. These included Val17, Tyr19, Val34, Tyr43, Gly44, Val45, Gly46, Leu49, Asn81, Gly82, Thr87, Ala126, Gly130, Leu139, Leu141, and Ala145. The average association constants for the three cellooligosaccharides were calculated using the values obtained individually from the analyses of the H$_N$ and $^{15}$N data recorded for each amide (Table 2.1). In the case of cellohexaose, binding to the CBD$_{NI}$ was sufficiently tight that, given the concentration of protein required for NMR analysis, it was not possible to determine accurately the $K_a$. Therefore an estimate of the upper limit of the association constant was obtained by a visual comparison of simulated titration curves with the observed NMR data (figure 2.8 and Table 2.1).

The association constants of CBD$_{NI}$ for the cellooligosaccharides measured by NMR agree with those determined independently by isothermal titration calorimetry under similar experimental conditions (Tomme et al., 1996a). Due to the different concentration windows necessary for each method, NMR provided a good measure of the relatively weak binding of cellotriose, whereas calorimetry yielded a more accurate value of the association constant of cellohexaose with the CBD.
Table 2.1. Association Constants $K_a$ for the Binding of Soluble Cellooligosaccharides to CBD$_{N1}$

<table>
<thead>
<tr>
<th>Ligand</th>
<th>$K_a$ (M$^{-1}$)$^a$</th>
<th>$K_a$ (M$^{-1}$)$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellotriose</td>
<td>180±60</td>
<td>not determined</td>
</tr>
<tr>
<td>Cellotetraose</td>
<td>4200 ± 720</td>
<td>3200 ± 500</td>
</tr>
<tr>
<td>Cellopentaose</td>
<td>34000 ± 7600</td>
<td>21000 ± 3000</td>
</tr>
<tr>
<td>Cellohexaose</td>
<td>(50000)</td>
<td>22000 ± 4000</td>
</tr>
</tbody>
</table>

$^a$ Data obtained at 35 °C and pH* 5.9 in 50 mM sodium chloride, 50 mM potassium phosphate, 0.02% sodium azide and 10% D$_2$O/90% H$_2$O. The reported $K_a$ values are the average of the those determined from the $^1$H and $^{15}$N chemical shift perturbations of Val17, Tyr19, Val34, Tyr43, Gly44, Val45, Gly46, Leu49, Asn81, Gly82, Thr87, Ala126, Gly130, Leu139, Leu141, and Ala145. The error range is one standard deviation. The $K_a$ reported for cellohexaose is an estimate of the upper limit of the association constant based upon simulations of the titration curves.

$^b$ Values determined by isothermal titration calorimetry at 35 °C and pH 7.0 in 50 mM potassium phosphate, 0.02% sodium azide (Tomme et al., 1996a).

Identification of the Cellooligosaccharide Binding Site

In addition to providing a means for quantitating the association of the cellooligosaccharides to CBD$_{N1}$, NMR also yields information regarding the location and structure of the binding site. Perturbations of the resonances of main chain $^1$H$_N$ and $^{15}$N nuclei upon sugar binding may arise due to the direct interaction of an amide with the ligand, or indirectly due to conformational changes resulting from the formation of the protein-sugar complex. These conformational changes are likely subtle as circular dichroism spectra reveal that the global secondary structure of CBD$_{N1}$ is unaffected in the presence of saturating quantities of cellotetraose. Bearing this in mind, structural insights into the binding site are provided by the pattern of NMR chemical shift changes observed with sugar binding. Figure 2.9 summarizes the changes of the $^1$H$_N$ and $^{15}$N chemical shifts of each amide in CBD$_{N1}$ due to the binding of cellotetraose. Similar
shift perturbations result from the binding of cellotriose, cellopentaose and cellohexaose (not shown).

As will be shown in chapter 3 CBD$_{N1}$ is composed of 10 $\beta$-strands, folded into two antiparallel $\beta$-sheets identified as A and B (figure 2.9). As illustrated schematically in figure 2.10, the global topology of CBD$_{N1}$ is that of a jelly-roll $\beta$-sandwich. Residues with exposed side-chains that form $\beta$-sheet A are identified on this diagram.

The average change in the absolute value of the H$^N$ chemical shifts of all resolved amides in CBD$_{N1}$ due to cellotetraose binding was 0.04 ppm with a standard deviation of 0.05 ppm. In the $^{15}$N dimension, the average absolute change in shift was 0.3 ppm, with a standard deviation of 0.4 ppm. Of the 24 residues with a perturbation in amide H$^N$ or $^{15}$N chemical shift greater than one standard deviation above the average, 18 lie within or immediately adjacent to the $\beta$-sheet of CBD$_{N1}$ composed of strands A1-A5 (figures 2.9 and 2.10). This strongly indicates that the binding site for the cellobiosaccharides lies on the face of the CBD formed by these strands. Furthermore, residues showing pronounced chemical shift changes upon the addition of sugar are located in all five of the $\beta$-strands that form this sheet. Therefore, the cellobiosaccharides are likely to bind across, and not parallel to, strands A1-A5. This conclusion is supported by the observation of intermolecular proton NOE interactions between bound cellotetraose and protein side-chains located within $\beta$-sheet A in CBD$_{N1}$ (chapter 3) and by the attenuation of the intensities of residues in this sheet by the binding of nitroxide-labelled cellobiosaccharides (chapter 6).

A limited number of amides located outside of $\beta$-strands A1-A5 also show changes in H$^N$ or $^{15}$N chemical shift upon sugar binding that are greater than one standard deviation from the average. All 6 of these amides either flank the Cys33-Cys140 disulfide bond (Val34, Leu139, Leu141, and Ala145), or lie in $\beta$-strand B3, which is adjacent to the strands containing this cysteine group (Thr67 and Ala68). As will be discussed in chapter 4 many of these perturbations arise from a contaminating amount of calcium present in the cellotetraose used in this study. However, some residues near the disulphide bond are still perturbed by cellopentaose binding in the fully calcium-bound CBD$_{N1}$. It is possible that sugar-binding indirectly influences a structural...
Chapter 2 - Celluloligosaccharide Binding by CBD$_N$1
Figure 2.9. (A) (previous page) Patterns of NMR chemical shift perturbations due to sugar binding demonstrate that the cellobiose--recognition site of CBD

is formed by a five-stranded anti-parallel β-sheet. The absolute values of the differences between the H\textsuperscript{N} and \textsuperscript{15}N chemical shifts of the main chain amides in the free and cellotetraose-bound forms of CBD\textsubscript{N1} are indicated as positive and negative numbers, respectively. Blank spaces identify residues for which no difference could be determined. The locations of the ten β-strands in the protein are shown by the arrows above the amino acid sequence. Two β-sheets identified in CBD\textsubscript{N1} are composed of strands A1–A5 (open) and B1–B5 (solid). Circled residues have a change in H\textsuperscript{N} and \textsuperscript{15}N chemical shift upon binding greater than one standard deviation from the mean absolute value change observed for all measured residues. Based on the pattern of chemical shift perturbations, the oligosaccharide-binding site lies across the face of the protein formed by strands A1–A5. Residues not detected in the \textsuperscript{1}H–\textsuperscript{15}N HSQC spectrum of CBD\textsubscript{N1} in the absence of added cellotetraose are boxed.

(B) (this page) Cα worm diagram of CBD\textsubscript{N1} with residues that experience the largest change in chemical shift with cellotetraose binding coloured red. The top panel shows β-sheet A, the binding face. The lower panel is a 90° rotation from the view shown in the top panel, the binding cleft is seen at the top of the structure in this bottom view. The amino and carboxyl termini are denoted by the labels N and C, respectively. This figure was made using the programme GRASP (Nicholls et al., 1991).
Figure 2.10. Schematic representation of the jelly-roll β-sandwich fold of CBD$_{N1}$. Based on the observed patterns of chemical shift perturbations resulting from sugar binding, the recognition site for the cellobiosaccharides lies across the face of the protein formed by β-strands A1-A5 (open arrows). Residues in these β-strands that have exposed side-chains are labelled. Cys33 and Cys140 in strands B2 and B5, respectively, form a disulfide bridge. The loops connecting the β-strands are not drawn to scale.
feature of the disulfide bond, such as disulfide isomerism, resulting in the observed chemical shift perturbations.

The changes in chemical shifts of remaining amides located in the β-sheets and loops of the protein are less than one standard deviation from the average. It is likely that these chemical shift differences reflect subtle structural perturbations associated with ligand binding that are propagated through the core of the protein.

**Effect of Cellotetraose Binding on the Aromatic Residues in CBD\textsubscript{NI}**

To investigate the possible roles that the aromatic side-chains play in sugar binding, the \(^1\text{H}\) resonances of the phenylalanine, tyrosine, and tryptophan residues in CBD\textsubscript{NI} were assigned in the absence and presence of saturating amounts of cellotetraose. Samples of the protein in which either tryptophan and tyrosine or tryptophan and phenylalanine were uniformly deuterated at all ring positions were prepared to simplify the NMR spectra of CBD\textsubscript{NI} and to identify unambiguously the \(^1\text{H}\) resonances from the phenylalanine or tyrosine residues, respectively. Near complete aromatic \(^1\text{H}\) assignments were obtained for both the unbound and bound forms of CBD\textsubscript{NI} (figure 2.11). An exception is Phe106, for which no resonances are observed in the homonuclear spectra of the protein. This residue, as well as Tyr19 and Tyr85, which show extensive line broadening, may undergo conformational averaging on a timescale of intermediate exchange due to partially hindered ring flipping (Wüthrich, 1986).

The differences between the aromatic \(^1\text{H}\) chemical shifts of the free and cellotetraose-bound forms of the protein are summarized in figure 2.12. The average change in the absolute values of the \(^1\text{H}\) chemical shifts for all aromatic side-chains was 0.07 ppm with a standard deviation of 0.09 ppm. Of the assigned aromatic residues, only Phe9, Trp16 and Tyr19 show above average ring chemical shift changes upon binding the oligosaccharide. In addition, the lineshapes of the aromatic proton resonances of Tyr85 are clearly altered in the present of cellotetraose (figure 2.11). As illustrated in figure 2.9, Trp16, Tyr19, and Tyr85 are located within β-sheet A and also exhibit significant changes in amide \(^1\text{H}\) and \(^{15}\text{N}\) chemical shifts upon sugar binding. This provides
Figure 2.11. Aromatic region of the DQF–COSY spectrum of $[^2H_5]$-Phe–, $[^2H_5]$-Trp labelled CBD$_{NI}$ in the (a) absence and (b) presence of 80 mM cellotetraose. Only the resonances from the six tyrosine residues are detected because the aromatic rings of the tryptophan and phenylalanine residues in the protein were biosynthetically deuterated and the amide $^{1}H$N signals eliminated by reversibly unfolding the protein in 99% D$_2$O buffer.
Figure 2.12. Histogram summarizing the perturbations of the $^1$H chemical shifts of the aromatic rings of the tryptophan, tyrosine, and phenylalanine residues in CBD$_{N1}$ due to the binding of cellotetraose. The data represent the average of the absolute value of the chemical shift changes observed for all protons in the aromatic side-chain of each residue. The average change for all assigned aromatic rings is 0.07 ppm with a standard deviation of 0.09 ppm. The chemical shifts of Tyr85 could not be confidently determined due to linebroadening and are not included in the figure. The resonances of Phe106 were not observed in the $^1$H-NMR spectra of CBD$_{N1}$. 
further support for the identification of the sugar binding site in CBD\textsubscript{NI}. Although Phe9 and Trp16 show pronounced NMR chemical shift changes upon sugar binding, structural studies reveal that the side-chains of these residues are located within the interior of CBD\textsubscript{NI}. Additionally, these residues are located close to the calcium-binding site in CBD\textsubscript{NI} (chapter 4). Their displayed chemical shift changes probably arise from the binding of the calcium present in the cellotetraose used in this study, and not due to cellotetraose itself. This leaves Tyr19 and Tyr85 as the only two residues with exposed aromatic rings that are likely to be involved in the binding of cellooligosaccharides.

\textit{Effect of High Concentrations of Cellotetraose}

In the absence of added cellotetraose, resonances from the amides of 13 residues are not observed in the $^1$H-$^15$N HSQC spectrum of CBD\textsubscript{NI}. These residues are located at the N-terminus of the protein and in $\beta$-strands B2 and B5, next to the disulfide bond (figure 2.9). The H\textsuperscript{N} and $^15$N resonances from the amides in question are observed in the presence of high concentrations of cellotetraose. As will be discussed in chapter 4, the appearance of these amide resonances is due to a calcium contamination in the cellotetraose used in this study. CBD\textsubscript{NI} tightly binds calcium at a site located on the opposite side of the protein from the binding face. The affinity of CBD\textsubscript{NI} for oligosaccharides is unaffected by the binding of calcium. Instead, calcium stabilises the tertiary structure of CBD\textsubscript{NI}. The presence of calcium in the cellotetraose used in this study therefore does not affect any of the results presented in this chapter.

\textit{Reduction of the disulphide in CBD\textsubscript{NI} results in the protein unfolding}

CBD\textsubscript{NI} contains two cysteine residues located at positions 33 and 140. As shown by the lack of reactivity to 5,5'-dithiobis-(2-nitrobenzoic acid), by electrospray mass spectroscopy and by structural data (chapter 3), these residues form a disulphide bond in the native protein. The effect of reducing this disulphide to the free sulphydryls was studied by $^1$H-$^15$N HSQC spectroscopy. Figure 2.13 shows spectra of oxidised and reduced CBD\textsubscript{NI} in the presence and absence of a
Figure 2.13. $^1$H-$^1$N spectra at 35°C and pH6.0 of native, oxidised CBD$_{N1}$, and protein denatured by reduction, addition of 7 M urea, and reduction in the presence of 7 M urea.
denaturing amount of urea. It is clearly seen in this figure that the only condition in which the resonances are well dispersed in the $^1$H dimension, indicating that the protein is folded, is in the absence of urea and when the disulphide is present. There is no evidence of folded reduced protein being present. Exactly the same results were found with spectra collected at 7 °C.

All three forms of denatured protein remain soluble in these experiments, even at the high concentrations used (1.6 mM). Upon exposure to air, the reduced sample reoxidises and completely refolds. Similarly with the urea denatured forms, exchanging the protein into buffer without urea results in refolding of the protein. There are slight differences for each of the three forms of denatured CBD$_{N1}$, probably reflecting differences in solution conditions, and the presence of a disulphide in the urea denatured CBD$_{N1}$.

**Discussion**

*CBD$_{N1}$ Binds Soluble Cellooligosaccharides*

I have shown using heteronuclear NMR spectroscopy that CBD$_{N1}$ binds soluble cellooligosaccharides to form 1:1 protein-carbohydrate complexes. From quantitative titration measurements, it was found that the affinity of CBD$_{N1}$ for cellooligosaccharides increases in the order cellotriose $<$ cellotetraose $<$ cellopentaose $\sim$ cellohexaose. The basis for the differences in the affinities for these sugars likely reflects the size of the CBD$_{N1}$ binding site. If this site is the same length as cellopentaose, then cellopentaose and cellohexaose would be expected to bind approximately equally well, while shorter ligands would exhibit diminished affinity. As will be shown in chapter 3, the binding site is approximately the same length as cellopentaose.

*Binding Site is Formed by a 5-Stranded $\beta$-sheet*

The chemical shift of a nucleus is very sensitive to changes in its local environment, thus providing an avenue for identifying the binding site of a ligand on a protein (Otting, 1993). When
mapped onto the sequence of CBD$_{N1}$, it is clear that most amides with chemical shifts perturbed significantly due to oligosaccharide binding are located within one of the two $\beta$-sheets of this jelly-roll $\beta$-sandwich protein (figure 2.9). Furthermore, changes in chemical shifts greater than one standard deviation above average occur for amides in each of the five strands of this sheet. I therefore conclude that the binding site for cellobiose lies across the face of CBD$_{N1}$ that is composed of strands A1-A5 (figure 2.10).

With the exception of the anomalous behaviour observed for a few amides in the presence of high concentrations of cellotetraose (resulting from calcium binding, chapter 4), the effects of all four cellobiose on the $^1$H-$^{15}$N HSQC spectra of CBD$_{N1}$ are very similar. That is, saturating amounts of cellobiose and cellotriose shift approximately equally the amide H$^N$ and $^{15}$N resonances of the protein. This implies that the four sugars bind CBD$_{N1}$ at the same location and that binding results in a common perturbation of the backbone structure of the CBD. A detailed discussion of the binding mechanism of CBD$_{N1}$ with cellobiose in relation to the side-chains present on the binding face is given in chapter 3.

Comparison to Other CBDs

CBD$_{N1}$ has the distinctive feature of binding solely to phosphoric-acid swollen cellulose (PASC), and not crystalline cellulose (Coutinho et al., 1992). In contrast, all other known CBDs bind both crystalline cellulose and PASC (chapter 1; Tomme et al., 1995). The binding of cellobiose by CBD$_{N1}$ reflects its binding specificity for cellulose. PASC is a heterogeneous disordered form of cellulose. Its surface contains large areas where the interstrand hydrogen bonds are broken, resulting in regions that resemble the single chain cellobiose used in this study.

The binding of cellobiose by two other CBDs has also been studied. Evidence for the role of the tryptophan residues of CBD$_{Cex}$ in sugar binding was obtained using NMR spectroscopy to monitor the effects of cellobiose on the spectrum of the protein (Xu et al., 1995). When four equivalents of this ligand were added to a uniformly $^{15}$N-labelled sample of
CBD\textsubscript{Cex}, the indole H\textsuperscript{e1} and 15N\textsuperscript{e1} resonances of two exposed tryptophans were perturbed slightly. Although this supported the role of the surface tryptophans in cellulose binding, it also demonstrated that CBD\textsubscript{Cex} interacts only weakly with soluble cellooligosaccharides. Recently, a study of the interaction between \textit{T. reesei} CBDs and cellohexaose was published (Mattinen \textit{et al.}, 1997b). No mention was made of changes in the chemical shift of resonances of the protein. Instead binding was monitored by observing changes in the line width of resonances of cellohexaose. A dissociation constant of 350 ± 90 μM, corresponding to a $K_d$ of ~3000 M\textsuperscript{-1}, was found for CBD\textsubscript{CBHI} binding cellohexaose.

In contrast to the studies with CBD\textsubscript{Cex} and CBD\textsubscript{CBHI}, CBD\textsubscript{N1} tightly binds cellohexaose, and the effect of ligand binding on the NMR spectrum of the protein is extensive. As shown in figure 2.9, approximately 40% of the main chain amides in CBD\textsubscript{N1} are perturbed upon addition of the soluble sugars. The weak binding to cellohexaose by CBD\textsubscript{Cex} and CBD\textsubscript{CBHI} indicates this cellooligosaccharide is a poor model for crystalline cellulose. These CBDs probably need a less flexible substrate or have to interact with multiple chains of cellulose to obtain tight binding. In contrast, cellohexaose closely resembles the natural binding substrate for CBD\textsubscript{N1} and is bound tightly.

\textit{Stability of CBD\textsubscript{N1}}

From the CD melt data it was shown that the thermal unfolding of CBD\textsubscript{N1} is a completely reversible process. Though CBD\textsubscript{N1} has a relatively low denaturation temperature, 47 °C, it increased with the addition of both cellohexaose and barley-β-glucan. This indicates that these substrates are bound by CBD\textsubscript{N1} and oligosaccharide-binding stabilises the folded structure of CBD\textsubscript{N1}. These results provided the ground work for a detailed calorimetric study of the stability of CBD\textsubscript{N1}, as presented in Creagh \textit{et al.} (1997).

The fact that CBD\textsubscript{N1} can be unfolded simply by reducing its single disulphide is surprising. Similar unfolding of proteins upon reduction has been observed before, but generally with proteins that contain multiple disulphides such as bovine pancreatic trypsin inhibitor (BPTI), hen lysozyme.
and ribonuclease (Creighton, 1993). In general, disulphides act to decrease the entropy of an unfolded protein so a smaller entropy penalty of folding is encountered. The greater the number of residues in the loop created by a disulphide, the greater the decrease in entropy. For CBD\textsubscript{NI} the disulphide forms a loop of 106 residues out of a total of 152. The results of Creagh \textit{et al.} (1997) show CBD\textsubscript{NI} has a $\Delta G$ of unfolding of 32 kJ mol\textsuperscript{-1} at 7 °C. This implies the disulphide imparts at least this much free energy to the stability of the protein.

This work was performed on CBD\textsubscript{NI} not calcium bound. As will be shown in chapter 4 CBD\textsubscript{NI} binds a calcium ion and calcium binding stabilises the tertiary structure of the protein (L.Creagh & C. Haynes, personal communication). It is possible that reduction of the disulphide in calcium-bound CBD\textsubscript{NI} might not result in unfolding of the protein.
Chapter 3

Structure Determination of CBD$_{N1}$

Abstract

Multidimensional heteronuclear NMR spectroscopy was used to determine the tertiary structure of the 152 amino acid N-terminal cellulose-binding domain from \textit{Cellulomonas fimi} $\beta$-1,4-glucanase CenC (CBD$_{N1}$). CBD$_{N1}$ was studied in the presence of saturating concentrations of cellotetraose, but due to spectral overlap, the oligosaccharide was not included in the structure calculations. A total of 1705 interproton NOE, 56 $\phi$, 88 $\psi$, 42 $\chi_1$, 9 $\chi_2$ dihedral angle and 88 hydrogen bond restraints were used to calculate 25 final structures. These structures have an rmsd from the average of 0.79 $\pm$ 0.11 Å for all backbone atoms excluding disordered termini, and 0.44 $\pm$ 0.05 Å for residues with regular secondary structures. CBD$_{N1}$ is composed of 10 $\beta$-strands, folded into 2 anti-parallel $\beta$-sheets with the topology of a jelly-roll $\beta$-sandwich. The strands forming the face of the protein previously determined by chemical shift perturbations to be responsible for cellooligosaccharide binding (chapter 2; Johnson \textit{et al.}, 1996a) are shorter than those forming the opposite side of the protein. This results in a 5-stranded binding cleft, containing a central strip of hydrophobic residues that is flanked on both sides by polar hydrogen-bonding groups. The presence of this cleft provides a structural explanation for the unique selectivity of CBD$_{N1}$ for amorphous cellulose and other soluble oligosaccharides, and the lack of binding to crystalline cellulose. The tertiary structure of CBD$_{N1}$ is strikingly similar to that of the bacterial 1,3-1,4-$\beta$-glucanases, as well as other sugar-binding proteins with jelly-roll folds.
Chapter 3 - Structure Determination of CBD_N1

Introduction

This introductory section is designed to familiarise the reader with some of the procedures used in the process of resonance assignment and structure calculation of CBD_N1 (figure 3.1). Specific details on these NMR experiments will be given in the experimental methods section. The work presented in this chapter has been published (Johnson et al., 1996b). With the exception of the resonances assignment of the aromatic rings, and the preparation of some of the samples that aided in the resonance assignment, I performed all the work presented in this publication.

Resonance assignment of proteins

The development of multi-dimensional triple-resonance spectroscopy in the late 1980’s and early 1990’s helped revolutionize the field of protein NMR (Bax & Grzesiek, 1993). Providing it is possible to obtain a 13C/15N labelled sample of about 1 mM that is stable for at least three weeks, it is a routine matter to assign the backbone and side-chain resonances of proteins up about ~20 kDa. Recently, with the use of 2H, 13C and 15N labelling, this molecular weight limit has been extended significantly (Sattler & Fesik, 1996; Kay & Gardner, 1997).

The key to the success of the triple-resonance experiments is the exclusive use of through-bond couplings to connect resonances from nuclei within adjacent amino acids. These experiments exploit the large coupling constants between 13C and 15N nuclei and their directly attached protons for efficient magnetization transfer. Previously, with the use of homonuclear experiments, potentially ambiguous through-space NOE connections were necessary to link the spin-systems for each amino acid (Wüthrich 1986).

A second development that has improved the sensitivity of experiments, and also helped to remove artifacts, is the incorporation of pulsed field gradients into pulse sequences (Keeler et al., 1994; Kay, 1995). These benefits result from pulsed field gradients reducing the need for phase-cycling, selecting for transfer pathways that do not suffer from sensitivity loss, and improved water suppression.
Figure 3.1. Flowchart outlining the steps in the structure determination process used for CBD$_{N1}$.
An especially effective method to assign the backbone amide chemical shifts is the combination of HNCACB (Wittekind & Mueller, 1993) and CBCA(CO)NH (Grzesiek & Bax, 1992) experiments. The HNCACB experiment correlates the $^{15}$N and $^1$H resonances of an amide with its own C$\alpha$ and C$\beta$ nuclei, as well as those of the preceding residue. In contrast, the CBCA(CO)NH experiment correlates the $^{15}$N and $^1$H resonances of an amide exclusively with the C$\alpha$ and C$\beta$ nuclei of the previous residue (figure 3.2). Also, for the HNCACB experiment, when transformed, the signals due to the C$\alpha$ are 180° out-of-phase with respect to the C$\beta$ signals. As shown in figure 3.3 it is possible from inspecting the dataset for peaks at the appropriate frequency to connect the spin systems together. This results in the assignment of the backbone $^{15}$N and $^1$H as well as the side-chain C$\alpha$ and C$\beta$ resonances.

Near complete assignment of the remaining side-chain $^1$H and $^{13}$C resonances can be obtained from the H(CCO)NH-TOCSY and (H)C(CO)NH-TOCSY experiments, figure 3.4 (Logan et al., 1992; Montelione et al., 1992; Grzesiek et al., 1993a). These experiments connect the side-chain $^1$H or $^{13}$C resonances with the $^{15}$N and $^1$H resonances of the following residue. To assign residues that precede a proline, and to correlate resonances from directly bonded $^1$H and $^{13}$C nuclei HCCH-TOCSY (Bax et al., 1990; Kay et al., 1993) and HCCH-COSY (Kay et al., 1990; Ikura et al., 1991) experiments are also run. Once the protein has been assigned as completely as possible the process of obtaining restraints for structure calculations begins.

Structure calculations

The determination of a three dimensional structure of a protein by the NMR method relies on the assignment of a large number of relatively short range, 5 Å or less, structural restraints. These distance restraints are based on the observation of nuclear Overhauser effects (NOEs) between pairs of protons. NOEs are assigned from two, three and four-dimensional NOESY experiments. The distances derived from NOEs are not exact, but rather used as a range with the lower limit being the van der Waals radii of the two protons (1.8 Å) and the upper limit either 2.9 Å, 3.5 Å or 5 Å for strong medium and weak NOEs, respectively.
Figure 3.2. Three dimensional heteronuclear experiments used to assign the resonances of CBD$_{N1}$. Boxed atoms are detected in the experiments.
Sequential assignments of proteins can be obtained by the identification of common Cα and Cβ resonances in the HNCACB (left) and CBCA(CO)NH (right) experiments. Shown here are strips through the amide resonances of F106 to Y112 of CBDN1. The CBCA(CO)NH experiment provides only inter-residue correlations between 15N(i) and HN(i) resonances and Cα(i-1) and Cβ(i-1), while both inter-residue Cα(i-1) and Cβ(i-1), and intra-residue Cα(i) and Cβ(i), correlations to 15N(i) and HN(i) resonances are identified in the HNCACB spectrum. Intra-residue Cα and Cβ correlations are indicated. Cα correlations (green) in the HNCACB spectrum appear 180 out-of-phase with respect to the Cβ correlations (red).
Figure 3.4. The assignment of side-chain resonances can be obtained by analysis of (H)C(CO)NH-TOCSY and H(CCO)NH-TOCSY spectra. The (H)C(CO)NH-TOCSY experiment (A) provides correlations between the $^{15}$N(i) and HN(i) resonances and the carbon resonances of the previous residue (i-1), while the H(CCO)NH-TOCSY experiment (B) correlates $^{15}$N(i) and HN(i) resonances with the proton resonances of the previous residue (i-1). Shown here are strips through the amide resonances of F106 to Y112 of CBDN1. Peaks below the dotted line in B indicate they appear aliased in this spectrum. Peaks marked (*) are examples of spectral overlap of different residues with similar $^{15}$N and HN frequency. Peaks marked (***) are due to residues that are not overlapped but have such strong intensities that they appear in the strips of nearby residues.
Torsion angle restraints are used for protein structure calculations. Values of the various dihedral angles are deduced by measuring three-bond proton-proton, nitrogen-carbon and carbon-carbon coupling constants (Bax et al., 1994). Backbone phi ($\phi$) as well as side-chain chi1 ($\chi_1$) and chi2 ($\chi_2$) angles are the most amenable dihedral angles for this process.

Hydrogen bond restraints are also very useful. These restraints are deduced from hydrogen-deuterium exchange experiments that identify slowly exchanging amide protons. After the secondary structure of the proton has been elucidated, these amides are paired with an oxygen atom. Hydrogen bonds are included as restraints between O and N, and O and H$_N$ atoms.

Initially, it is not possible to unambiguously assign most of the peaks present in the NOESY spectra. This is due to chemical shift degeneracy that is not resolved even in four-dimensional experiments. A starting set of dihedral and distance restraints derived from a limited number of unambiguously assigned NOEs are used to calculate initial structures. Based on these preliminary structures, additional restraints are assigned in a reiterative fashion, discarding possible assignments for NOEs that are far apart in the preliminary structures. This is a manual process that takes many months.

Distance geometry (DG) is a common method for generating a starting structure. The DG algorithm chooses exact distance restraints at random from the experimental range used as input and fits them to a mathematical solution while maintaining normal bond lengths in the molecule. Following distance geometry the structures contain many sub-optimal bond lengths and angles and therefore need to be corrected. This regularisation is done by simulated annealing (Nilges et al., 1988). Simulated annealing (SA) is a term used to describe a minimization / molecular dynamics schedule where the contributions due to van der Waals interactions, bond lengths, angles and peptide-bond planarity as well as the experimentally determined distance and dihedral angle restraints are included with different weightings. Each of these factors contributes an energy term to a target function that the system tries to minimize.

Using this method it is possible to obtain different structures, all of which satisfy the experimental data. Therefore a large number of structures are calculated with a subset, or
ensemble, of low energy ones that satisfy the experimental restraints as well as having regular geometry chosen to represent the structure of the protein.

**Experimental methods**

**Sample Preparation.**

Samples of uniformly (~99%) $^{15}$N and $^{13}$C/$^{15}$N labelled CBD$_{N1}$ were produced by expression of the plasmid pTugN1n in *E. coli* JM101 cells, as described previously (Chapter 2; Johnson *et al.*, 1996a). In the case of the $^{13}$C/$^{15}$N labelled protein, the growth media contained 2 g/L of $[^{13}$C$_6]$-glucose and 1 g/L of $^{13}$C/$^{15}$N Isogro algal extract (Isotec, Inc.). Since the JM101 cells containing pTugN1n grew poorly in minimal media, CBD$_{N1}$ non-randomly fractionally $^{13}$C labelled at a level of 10% (Neri *et al.*, 1989) was produced by initially growing bacteria in M9 media (Miller, 1972) supplemented with 1 g/L unlabelled Isogro. At an OD$_{600}$ of 1.0, the cells were spun down, washed with M9 media, and resuspended in M9 media containing 0.3 g 99% $[^{13}$C$_6]$-glucose and 2.7 g unlabelled glucose as the sole carbon sources. IPTG (0.5 mM) was then added to induce expression of the gene encoding CBD$_{N1}$. The cells were grown at 30 °C for an additional 16 hours, and the secreted protein was purified as outlined previously (Chapter 2; Johnson *et al.*, 1996a). The amides of Tyr, Leu, Asp/Asn, and Val were selectively [$\alpha$-$^{15}$N] enriched using the protocol of McIntosh and Dahlquist (1990). CBD$_{N1}$ samples with selectively deuterated aromatic rings were obtained from a synthetic media containing 100 mg/L of L-$\delta_1,\epsilon_2,\zeta_1,\eta_2-[^{2}$H$_5]$ tryptophan and 100 mg/L of either L-$\delta_1,\epsilon_1,\zeta_2-[^{2}$H$_5]$ phenylalanine or $\delta_1,\epsilon_1,\zeta_2-[^{2}$H$_4]$ tyrosine (Cambridge Isotope Laboratories and Isotec Inc.) (McIntosh *et al.*, 1990; McIntosh & Dahlquist 1990).

Samples of CBD$_{N1}$ for NMR analysis were exchanged into 50 mM sodium chloride, 50 mM potassium phosphate (pH* 5.9), 0.02% sodium azide, 10% D$_2$O/90% H$_2$O using ultrafiltration through a cellulose free membrane (Filtron). Samples in deuterated buffer were
Chapter 3-Structure Determination of CBD$_{N1}$

obtained by twice lyophilizing the CBD$_{N1}$, and redissolving in an equivalent amount of 99.9% D$_2$O. Typical protein concentrations were 2 mM as determined by $\varepsilon_{280} = 21370 \text{ M}^{-1} \text{ cm}^{-1}$. With the exception of the selectively $^{15}$N labelled proteins, CBD$_{N1}$ samples contained up to a 40 fold molar excess of cellotetraose (Seikagaku Corp.) to facilitate the observation of all the resonances of CBD$_{N1}$ (Chapter 2). As will be discussed in chapter 4, the appearance of these signals is due to a calcium contaminant in the cellotetraose. It was unknown at the time the work presented in this chapter was performed that CBD$_{N1}$ bound calcium.

\textit{NMR Spectroscopy.}

NMR spectra were recorded on a Varian Unity 500 MHz spectrometer equipped with a triple resonance probe and a pulsed field gradient accessory. $^1$H chemical shifts were referenced to an internal standard of DSS at 0.00 ppm, $^{13}$C chemical shifts were referenced to an external DSS standard at 0.00 ppm, and $^{15}$N was referenced to external 2.9 M $^{15}$NH$_4$Cl in 1 M HCl at 24.93 ppm (Levy & Lichter, 1979). This latter reference yields $^{15}$N chemical shifts 1.6 ppm greater than those obtained using liquid NH$_3$ (Wishart \textit{et al.}, 1995).

All spectra were collected at 35 °C and analysed using a combination of FELIX v2.30 (Biosym Technologies; San Diego, Ca.), NMRPipe (Delaglio \textit{et al.}, 1995), and PIPP (Garrett \textit{et al.}, 1991). Experiments with $^1$HN detection were recorded using the enhanced sensitivity pulsed field gradient approach of Kay \textit{et al.} (1992) and Muhandiram and Kay (1994). Selective water flip back pulse was incorporated to minimize the perturbation of the bulk water magnetization (Grzesiek & Bax, 1993; Zhang \textit{et al.}, 1994). Quadrature detection was accomplished using the States-TPPI method (Marion \textit{et al.}, 1989a). The initial delays in most of the indirectly detected dimensions were set to $1/(2*sw)$, resulting in a $180^\circ$ first order phase shift across the transformed spectrum and the inversion of aliased peaks (Bax \textit{et al.}, 1991). A summary of the data collection and processing parameters for the NMR experiments used to determine the structure of CBD$_{N1}$ is given in Table 3.1.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Method</th>
<th>Temperature</th>
<th>Frequency (MHz)</th>
<th>Chemical Shift (δ)</th>
<th>J Coupling (Hz)</th>
<th>Experiment</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winkler &amp; Müller, 1993</td>
<td></td>
<td></td>
<td>256</td>
<td></td>
<td>143.0</td>
<td>4.5</td>
<td>13C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>128</td>
<td></td>
<td>60.9</td>
<td>9</td>
<td>1H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>128</td>
<td></td>
<td>71.4</td>
<td>11</td>
<td>1H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>256</td>
<td></td>
<td>61.9</td>
<td>9</td>
<td>1H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>256</td>
<td></td>
<td>61.9</td>
<td>9</td>
<td>1H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>256</td>
<td></td>
<td>61.9</td>
<td>9</td>
<td>1H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>256</td>
<td></td>
<td>61.9</td>
<td>9</td>
<td>1H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>256</td>
<td></td>
<td>61.9</td>
<td>9</td>
<td>1H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>256</td>
<td></td>
<td>61.9</td>
<td>9</td>
<td>1H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>256</td>
<td></td>
<td>61.9</td>
<td>9</td>
<td>1H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>256</td>
<td></td>
<td>61.9</td>
<td>9</td>
<td>1H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>256</td>
<td></td>
<td>61.9</td>
<td>9</td>
<td>1H</td>
</tr>
</tbody>
</table>

**Table 1:** Acquisition and processing parameters for NMR experiments recorded on CBD.
Chapter 3-Structure Determination of
A,

O

2

O
ol O
o
w

O
n
-<
o

n
W

o
n
i

o
o

S- o

z

o
b O
3 O
cn O
o>

n
n

IO

a

P

Co

63

CBDNy

o

GO
to

>

to
LU

o

Z Q

K nS3

ON
Ul
O
O

b

TD

^~

1—*

4^ CO
Ul ON
0
4^

4^ t o
ON p
s ] Ul
NO

si
ON
4^
-J

-<

ON
Ul
O
O

4^
ON
si

O

^

«

4^

4^

O J

O J

b b
4^

4^

LU

iL 'L II_
ffi n s

Ul . . >—'
W tO

to ^ 00

4^
Ul O
0
to
Ul
0
O N

4^
O N

-J

4^
O J
L ^

4^
0
Ul
0

Ul J i
H O
to o

NO
ON

4^
Ul O
0
to
0
Ul

-1^
0
Ul
0

O N

O J

4^

, 1—*

O N

4^

Ul

o

s)

to
^3

s]

10

^1

o_

Ul

M

to

to

00 00

a

O

c
i-t

1-1 EU

P

^
NO
NO
O J

f» NO

S

NO

4^

vo
4^

Ul

O N

1—I

4^

1—»

O N

I

O

3

3

o

O

T3

00

U J

TD

13

ON

O

to
N

si

a—

o

ui to to
to ui ui
O

O N

O N

2,0

to to to
Ul Ul Ul
O N

O N

O N

2-

L<-^<

«

H
Ul U I

O N

ON
Ul

o
o

ON

Ul
ON

Ui

,0 V

ON
Ul
O
O

Ul
0

i-. .
Ul

5*
E»

NO
NO
O J

»-t

E»

p
l

<
CD

CD

P

^<
CD

NO
NO
O

1—.

tO

O J

Ul
o

Ul
O J

1—. 00
4^
Ul
0

4^
O
N
O
ON
O
N
•
s i Ul
so

0
Ul
0

"

g
"^

w

^

TD" TD

S-J

^
T3
O

TD

cn

cn

NO NO
O
O

cn

O J

tO

O

ON
Ul
0
0

1—1

t— Ul

4^
Ul
0

4i. t o
ON P
s ) Ul
NO

Cu
TD

ON
O si
TD

Ul

o

TD

o

cn

TD

cr NO

NO

o

o • 00
o

TD

O

O

to ^- to
Ul to Ul

tO H->
Ul t o
ON 00

LTT

O N

to

to

ui to ui

ON 00 ON

«
p
X

l-l

to

to

to

O N

O N

ui to ui
O N

I—I

O
cn OQ
CD

CD

1—f
P ,

Ea

IzZ

o

00

to

ui to ui
00

O N

o
c

NJ

II II
1 Ln

-

LU

"DO

o

to to to
Ul Ul Ul
O N

o

so

cr
^ Q o o

TD

Ul ON

||

Z fi

O J

P

II

•GO

tO

4^ >— ON
4^ Ui
0
Ul Ui O
0

o

10

'L Ti

^

4^

b

Ul
NO 4^

io ^

T3

>

00

'L L T i

to

z

>

tO

OJ

ON

Ul 4^ 0
0
Ul Ul
0
0
0

^1

1

f

O J

£

O J

2

c tr.

CD

H

W

^

4^ Ul

»o

z

o F

H
O
O
oo

SL?

O

T3

c
o sr
a fo t i

00

3

K)

O J

CD

^ Z ffi

cr S-°*

o

M

4^.

3
O

i

cr

o

Ul

to

'L ~ 'L
33 (-j K

si

s j

^

OO
O 00
Ul O N
4^
0

T3 T3

— 13 T3
V3

cr "

-<

LU

^
b£ OJ
^
ON on
0 0
Is) S

to

00

Q

2:

4^
O
O
O

Ul

"CO

II

Ul
_ . t. o1
n r t "Tr
*—'k .CO
(0
0 + ^

O J

Ti_

II II

O

OJ
tO

LU

II

|s

z
a

o
w
o
>
o
o

O
W
n
>
n
o
n

cn

S°
ft) fJQ

CD'

K

CD

CD

X

P

rr,

P

2

NO vO
NO NO
O J
0

P O
N

p

j

p

o

S

3

f

3

o
CD

to

»
NO
g
L
NO
"

ON OO ON

N
CD

cn

p
NO
NO
O J

CD'

7?

9P
C
d
p
xNO
NO

to


Chapter 3: Structure Determination of 2

<table>
<thead>
<tr>
<th>Author</th>
<th>Date</th>
<th>Method</th>
<th>Temperature</th>
<th>Crystal Data</th>
<th>Refinement Details</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vissers et al.</td>
<td>1993</td>
<td>NMR</td>
<td>200 K</td>
<td>6500, 300</td>
<td>312 C</td>
<td></td>
</tr>
<tr>
<td>Czerny et al.</td>
<td>1996</td>
<td>X-ray</td>
<td>173 K</td>
<td>6500, 300</td>
<td>312 C</td>
<td></td>
</tr>
<tr>
<td>Bou et al.</td>
<td>1992</td>
<td>NMR</td>
<td>173 K</td>
<td>6500, 300</td>
<td>312 C</td>
<td></td>
</tr>
<tr>
<td>Clore et al.</td>
<td>1991</td>
<td>NMR</td>
<td>173 K</td>
<td>6500, 300</td>
<td>312 C</td>
<td></td>
</tr>
<tr>
<td>Pascali et al.</td>
<td>1991</td>
<td>NMR</td>
<td>173 K</td>
<td>6500, 300</td>
<td>312 C</td>
<td></td>
</tr>
<tr>
<td>Macura et al.</td>
<td>1984</td>
<td>NMR</td>
<td>173 K</td>
<td>6500, 300</td>
<td>312 C</td>
<td></td>
</tr>
<tr>
<td>Oehme et al.</td>
<td>1988</td>
<td>NMR</td>
<td>173 K</td>
<td>6500, 300</td>
<td>312 C</td>
<td></td>
</tr>
<tr>
<td>Yamazaki et al.</td>
<td>1993</td>
<td>NMR</td>
<td>173 K</td>
<td>6500, 300</td>
<td>312 C</td>
<td></td>
</tr>
</tbody>
</table>
Amide Hydrogen Exchange.

Amide hydrogen exchange rates were determined by recording a series of sensitivity-enhanced gradient $^1$H-$^{15}$N HSQC spectra at 10, 29, 59, 95, 131, 266, 451, 688, 962, 5377, 10790, and 25211 min after dissolving lyophilized CBD$_{N1}$ in D$_2$O. To minimize the quantity of residual H$_2$O, the uniformly $^{15}$N-labelled protein was lyophilized twice, being resuspended in D$_2$O after the first freeze-drying step. The buffer concentration and pH were held constant by maintaining the sample volume.

Structure Calculations.

All structure calculations were performed using X-PLOR 3.1 (Brünger, 1992). Distance restraints from NOE experiments were tabulated with in-house programmes available from http://otter.biochem.ubc.ca/www/nmrtools.html. Initially, a preliminary fold for CBD$_{N1}$ was calculated following the DGSA protocol using NOE restraints involved in $\beta$-strand pairing, unambiguous NOE restraints identified from 3 and 4D heteronuclear NOESY spectra, and dihedral angle restraints. This preliminary structure was used in a reiterative fashion to assign additional NOE interactions. A total of 1705 NOE derived distance restraints were used in the final calculation of an ensemble of 60 structures. This data set was comprised of 711 nontrivial intraresidue, 411 sequential, 90 short-range ($1 < |i-j| \leq 4$), and 463 long range ($|i-j| > 4$) distance restraints.

Interproton distances were assigned to three strengths following a square-well potential energy function: weak 1.8 - 5.0 Å, medium 1.8 - 3.5 Å, and strong 1.8 - 2.9 Å. A correction of 0.5 Å was added to the upper bounds of restraints involving methyls (Clore et al., 1987). The distance ranges for the $^{15}$N NOESY-HSQC were calibrated using the intensities of the $^1$H$^N_i$-$^1$H$^\alpha_i$. $^1$H$^N_j$ NOEs, which should be strong (~2.2 Å) in regions of $\beta$-strand conformation, and cross strand $^1$H$^\alpha_i$-$^1$H$^N_j$ NOEs, which in anti-parallel $\beta$-sheets should have a medium intensity (~3.2 Å). For the simultaneous 3D $^{13}$C/$^{15}$N NOESY-HSQC spectra (Pascal et al., 1994), these sequential NOEs, as well as the strong cross-strand $^1$H$^\alpha_i$-$^1$H$^\alpha_j$ NOEs (~2.3 Å), were used to calibrate the intensity
ranges. For the 2D NOESY spectra involving Phe and Tyr aromatic rings, the $^1\text{H} \delta - ^1\text{H} \epsilon$ NOE was used for calibration. In the case of the 4D $^{13}\text{C} - ^{13}\text{C}$ NOESY, a 150 msec mixing time was used, resulting in extensive spin diffusion. As a result of this complication, all NOEs extracted from this experiment were classified as weak.

In addition, 88 hydrogen bond restraints (44 hydrogen bonds), 56 $\phi$-angle restraints, 88 $\psi$-angle restraints, 42 $\chi_1$-angle restraints, and 9 $\chi_2$-angle restraints were included in structure calculations. Hydrogen bonds, deduced from patterns of amide hydrogen exchange and NOE interactions involving main chain protons, were restrained to 2.5 - 3.5 Å between O and N atoms, and 1.5 - 2.5 Å between HN and O atoms. The $\phi$ torsion angles were restrained to $60^\circ \pm 30^\circ$ for $J < 5.5$ Hz, $-120^\circ \pm 30^\circ$ for $8$ Hz < $J < 9$ Hz, and $-140^\circ \pm 20^\circ$ for $J > 9$ Hz. These $^{3}\text{J}_{\text{HN}-\text{H} \alpha}$ couplings were determined from a $^{1}\text{H} - ^{15}\text{N}$ HMQC-J spectrum (Kay & Bax, 1990) using software provided by Lewis Kay, as described by Foreman-Kay et al. (1990). The $\psi$ angles were restrained to $120^\circ \pm 100^\circ$ or $-30^\circ \pm 110^\circ$ based on the ratio of $^{1}\text{H} \alpha_{i-1} - ^{1}\text{H} \text{N}_i$ and $^{1}\text{H} \alpha_{i} - ^{1}\text{H} \text{N}_i$ NOE intensities (Gagné et al., 1994). The $\chi_1$ and $\chi_2$ angles were restrained to $\pm 30^\circ$ from their assigned rotamer values.

Results

For this current study, all spectra of $\text{CBD}_{\text{NI}}$ were obtained in the presence of a 40 fold excess of cellotetraose (typically 80 mM) in order to obtain complete resonance assignments, as well as to investigate the effects of ligand binding. At 35 °C, the free and cellotetraose-bound forms of $\text{CBD}_{\text{NI}}$ are in fast exchange on the NMR timescale, resulting in the observation of population weighted average chemical shifts. With the concentration of cellotetraose used, $\text{CBD}_{\text{NI}}$ is fully saturated with this ligand ($K_a = 4200 \pm 720$ M$^{-1}$; Chapter 2; Johnson et al., 1996a). However, the cellotetraose was not specifically included in the structure calculations due to the
degeneracy of its NMR spectrum, which prevented the unambiguous assignment of intermolecular NOEs.

**Main Chain Resonance Assignments.**

The $^1$H$^N$ and $^{15}$N resonances from the main chain amides of CBD$_{N1}$ were assigned using the combination of HNCACB (Wittekind & Mueller, 1993) and CBCA(CO)NH (Grzesiek & Bax, 1992) experiments to correlate the $^{15}$NH$_i$ and $^{13}$C$_i$/$^{13}$C$_{i-1}$, $^{13}$C$_i$/$^{13}$C$_{i-1}$ and $^{15}$NH$_i$ and $^{13}$C$_i$/$^{13}$C$_{i-1}$ resonances, respectively (figure 3.3). Selectively [α-$^{15}$N]-labelled Tyr, Leu, Val, and Asp/Asn CBD$_{N1}$ provided amino acid-specific starting points for this assignment procedure. A modified version of the CBCACO(CA)HA experiment that detects only the $^1$H$^\alpha$ and $^{13}$C$_\alpha$/$^{13}$C$_{\alpha}$ resonances of residues preceding prolines was also extremely useful for providing unambiguous reference points (Olejniczak & Fesik, 1994; Lewis Kay, pers. comm.). The assignments of the resonances from the remaining backbone residues were obtained using the HNCO (Ikura et al., 1990; Muhandiram & Kay, 1994) and CBCACO(CA)HA experiments (Kay, 1993). Figure 3.5 shows the assigned $^1$H-$^{15}$N HSQC spectrum of CBD$_{N1}$. With the exception of the 13 amides not detected in the absence of added cellotetraose, near complete assignments of the resonances from the main chain $^1$H$^N$ and $^{15}$N nuclei of uncomplexed CBD$_{N1}$ were also obtained by following the progressive titration of the protein with cellotetraose (chapter 2; Johnson et al., 1996a).

**Aliphatic Side-Chain Resonance Assignments.**

Virtually complete side-chain assignments of the resonances from $^1$H, $^{13}$C, and $^{15}$N nuclei in CBD$_{N1}$ were obtained using a combination of $^{15}$N TOCSY-HSQC (Marion et al., 1989c), H(CCO)NH-TOCSY, (H)C(CO)NH-TOCSY (Logan et al., 1992; Montelione et al., 1992; Grzesiek et al., 1993a), HCCH-TOCSY (Bax et al., 1990; Kay et al., 1993), and HCCH-COSY (Kay et al., 1990; Ikura et al., 1991) experiments (figure 3.4). The HCCH-COSY and HCCH-
Figure 3.5. Sensitivity-enhanced gradient $^1$H-$^{15}$N HSQC spectrum of CBD$_{N1}$ at 35°C and pH 5.9 showing the assignments of the resonances from backbone amide and tryptophan indole $^{15}$Ne$^1$H groups, and those of the observable signals from the side-chain $^{15}$N$^5$H$_2$ and $^{15}$NeH$_2$ of asparagine and glutamine, respectively. The aliased peak from Thr$_8$ is denoted by an asterisk. The spectrum in (B) is the expansion of the boxed central region of (A).
TOCSY experiments were recorded with the protein in D$_2$O buffer. These assignments are reported in the appendix.

Initial inspection of the $^1$H-$^{15}$N HSQC spectrum of CBD$_{N1}$ revealed that four of seven expected peaks from the side-chain $^{15}$NH$_2$ resonances were either missing, very weak or degenerate. The strong resonances from the $^{15}$N$^2$H$_2$ groups of Gln42, Gln80, and Gln101 were readily assigned from the HNCACB experiment and confirmed using the $^{15}$N NOESY-HSQC spectrum. Therefore, the unidentified side-chain amides resonances were those of Asn50, Asn81, Gln124, and Gln128. All four of these residues lie within the oligosaccharide-binding cleft of CBD$_{N1}$ and could possibly play an important role in interacting with sugar. As a result of their potential importance, a special effort was made to assign them. Spectra from this process are shown in figure 3.6.

First, a modified version of the CBCACO(CA)HA experiment that links the side-chain carbonyls to the previously assigned $^{13}$C$^\alpha$/13C$^\beta$ and $^1$H$^\beta$2,$^3$ of Asp/Asn and $^{13}$C$^\beta$/$^{13}$C$^\gamma$ and $^1$H$^\gamma$2,$^3$ of Glu/Gln was recorded (Kay, 1993). Next an HNCO experiment, tuned using a total delay of $1/(4J_{NH})$ during the first reverse INEPT sequence to favor AX$_2$ spin systems (Schleucher et al., 1994), was used to correlate the resonances of these side-chain carbonyls to the corresponding $^{15}$NH$_2$ groups. Finally, an HSQC spectrum, with similar delays to enhance AX$_2$ spin systems, was recorded overnight to help identify the $^{15}$NH$_2$ resonances. This strategy yielded assignments for all seven side-chain $^{15}$NH$_2$ groups, as well as those of the carbonyls of almost all Asp/Asn and Glu/Gln residues (Appendix).

Both side-chain amide $^1$H$^\delta$2 protons as well as the $^{15}$N$^\delta$2 of Asn50 are completely degenerate with those of Gln101. Fortunately, the difference in the chemical shift of the side-chain carbonyl resonances of these two residues allowed them to be distinguished (figure 3.6, panel C). The chemical shift of the $^{13}$C$^\gamma$ carbonyl of Asn50 is unusually upfield shifted to 173.5 ppm. In the absence of structural information for the bound cellotetraose, the reason for this perturbed shift this is not immediately apparent. The side-chain $^{15}$NH$_2$ groups of Asn81, Gln124, and Gln128, all of which are likely to participate in hydrogen bonding to the cellotetraose, exhibit very weak
Figure 3.6 Spectra used to assign side-chain NH2 resonances. (A) The side-chain carbonyl resonances are assigned. This panel shows a plane from the modified version of the CBCACO(CA)HA experiment used to connect the side-chain carbonyls to the previously assigned 13Cα/13Cβ and 1Hβ2,β3 of Asp/Asn and 13Cβ/13Cγ and 1Hγ2,γ3 of Glu/Gln. Peaks resulting from Q80 are marked by an asterisk to indicate these signals lie on a different plane, and overlap onto the one shown in this figure. (B) Using the assignments of the side-chain carbonyls the NH2 groups are assigned from an HNCO experiment, tuned using a total delay of 1/(4JNH) during the first reverse INEPT sequence to favor AX2 spin systems. (C) Nitrogen plane from the modified HNCO experiment shows the NH2 resonances of Q101 and N50 are overlapped. They are assigned on the basis of their different side-chain carbonyl chemical shifts. Peaks marked with an asterisk result from resonances that lie on another plane.
resonances. Only one $^1$H$^2$ for the amide of Asn81 was found. It is not known if the resonances from the two amide protons are degenerate or if the second is very weak. It is likely that the anomalous behavior of these side-chain amide resonances results from conformational exchange broadening, possibly due to unfavorable kinetics of cellotetraose association and dissociation from CBD$_{N_1}$ or due to mobility of the sugar ligand within the binding cleft. In support of this suggestion, I note that the side-chain of Gln80, which lies on a $\beta$-strand that is part of the binding face but does not point into the binding cleft, has a strong $^{15}$N$^5$H$_2$ signal. In contrast, the side-chain of the adjacent Asn81, which points into the binding cleft, has a very weak $^{15}$N$^8$H$_2$ signal.

Aromatic Side-Chain Assignments.

Due to spectral overlap, the use of protein samples with selectively deuterated aromatic rings was an extremely useful tactic for the assignment of these side-chain spin systems. The $^1$H resonances from the aromatic rings were identified from homonuclear DQF-COSY, 70 msec mixing time TOCSY, and 150 msec mixing time NOESY spectra. These were recorded with samples of 1.9 mM unlabelled, 2.0 mM ([$^2$H$_4$]-Tyr and [$^2$H$_5$]-Trp)-labelled, and 1.9 mM ([$^2$H$_5$]-Phe and [$^2$H$_5$]-Trp)-labelled CBD$_{N_1}$, each in the presence of 80 mM cellotetraose. The unlabelled CBD$_{N_1}$ sample was recorded in both H$_2$O and D$_2$O buffers, while the two deuterated CBD$_{N_1}$ samples were recorded solely in D$_2$O buffer. The $^1$H$^5$ and $^1$H$^6$ of the aromatic ring spin systems were then directly connected to the previously assigned $^{13}$C$^\beta$ nuclei using the (H$\beta$)C$\beta$(C$\gamma$C$\delta$)H$\delta$ and (H$\beta$)C$\beta$(C$\gamma$C$\delta$C$\epsilon$)He experiments (Yamakazi et al., 1993). Finally the assignment of the $^{13}$C resonances of the aromatic rings were obtained from $^1$H-$^{13}$C HSQC and CT-HSQC spectra acquired using the $^{13}$C/$^{15}$N-labelled CBD$_{N_1}$ (Santoro & King, 1992; Vuister & Bax, 1992). The CT-HSQC was particularly useful for distinguishing the $^{13}$C$^6$ of tryptophan residues due to their inverted signals relative to those of other $^{13}$C nuclei with an even number of neighboring carbons.
**Stereospecific Assignments and Side-Chain Torsion Angle Restraints.**

Stereospecific assignment of 23 of the 67 residues in CBDN₁ with prochiral $^1\text{H}^\beta$ protons was obtained from a conservative analysis of the HNHB (Archer et al., 1991), 40 msec and 72 msec mixing time $^{15}\text{N}$ TOCSY-HSQC (Marion et al., 1989c), and 50 msec $^{13}\text{C}/^{15}\text{N}$ NOESY-HSQC (Pascal et al., 1994) spectra of CBDN₁. These assignments and the corresponding $\chi_1$ restraints were determined based on the staggered rotamer model, as outlined by Powers et al. (1993).

Near complete stereospecific assignments of the diastereotopic methyls of valine and leucine residues were obtained using the elegant approach of Neri et al. (1989). As demonstrated by these authors, in high resolution HSQC spectra of 10% non-randomly fractionally $^{13}\text{C}$ enriched proteins, the Pro-R (Leu$^{51}$, Val$^{11}$) methyls are doublets due to $^{13}\text{C}-^{13}\text{C}$ couplings, while the Pro-S (Leu$^{52}$, Val$^{12}$) methyls are singlets. An additional level of discrimination is provided by the use of a constant time $^1\text{H}-^{13}\text{C}$ HSQC experiment with a total evolution delay of $1/J_{\text{CC}} = 1/34$ Hz (Lewis Kay, pers. comm.). In the resulting CT-HSQC spectrum, the signals due to Val$^{11}$, Leu$^{51}$, and Ile$^{12}$ and the signals due to Val$^{12}$ and Leu$^{52}$ all appear as singlets but with opposite sign (Figure 3.7). Also, the peaks due to the Thr$^{12}$ and Ile$^{51}$ methyls, which are apparent triplets in the $^{13}\text{C}$ HSQC due to approximately equal levels of $^{13}\text{C}-^{13}\text{C}$ and $^{13}\text{C}-^{12}\text{C}$ labeling (Szyperski et al., 1992), are nulled in the constant time $^{13}\text{C}$ HSQC. Together, these two factors help simplify crowded regions of the spectrum, allowing previously overlapped resonances to be assigned.

$\chi_1$ restraints for 9 of the 14 valines, 6 of the 20 threonines, and all 3 isoleucines were determined on the basis of $^3J_{\text{NC}}\gamma$ and $^3J_{\text{C-C'}}\gamma$ coupling constants (Tables 3.2 and 3.3) and intraresidue NOE interactions, according to the staggered rotamer model. The coupling constants were determined quantitatively from $^{13}\text{C}-\{^{15}\text{N}\}$ and $^{13}\text{C}-\{^{13}\text{C}\}$ spin echo difference CT-HSQC spectra (Grzesiek et al., 1993b; Vuister et al., 1993). Peak volumes were used in the calculation of the coupling constants using programmes I wrote, as described by Grzesiek et al. (1993b) and Vuister et al. (1993). The $\chi_1$ analysis of valine was aided by the previously determined stereospecific assignments, particularly in the case of resonances obscured by spectral overlap.
Chapter 3 - Structure Determination of CBD$_{N1}$

Figure 3.7. A portion of the (A) constant time $^1$H-$^{13}$C HSQC and (B) $^1$H-$^{13}$C HSQC spectra of 10% fractionally $^{13}$C enriched CBD$_{N1}$. Due to the biosynthetic labeling pathways, the Pro-R (Leu$_1^{51}$, Val$^{71}$) methyls are doublets and Pro-S (Leu$_2^{52}$, Val$^{72}$) methyls are singlets in the regular HSQC (B). In the CT-HSQC (A), all peaks are singlets but the Pro-R (Leu$_1^{51}$, Val$^{71}$) methyls (filled) have the opposite sign compared to the Pro-S (Leu$_2^{52}$, Val$^{72}$) methyls (open). The use of constant time results in further simplification of the spectra by the elimination of Thr$^{12}$ and Ile$^{81}$ methyls (not shown in the spectral window) that are apparent triplets in the non-constant time HSQC. For clarity, assignments are indicated only for selected peaks.
Table 3.2: Coupling Constants\textsuperscript{a} and $\chi 1$ Assignments\textsuperscript{b} for Valine Residues in CBD\textsubscript{N1}.

<table>
<thead>
<tr>
<th>Residue</th>
<th>$^{3}J_{NCY}$ (Hz)</th>
<th>$^{3}J_{CCY}$ (Hz)</th>
<th>$^{3}J_{NCY}$ (Hz)</th>
<th>$^{3}J_{CCY}$ (Hz)</th>
<th>$\chi 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>V17</td>
<td>c</td>
<td>c</td>
<td>0.7</td>
<td>1.7</td>
<td>d</td>
</tr>
<tr>
<td>V34</td>
<td>0.4</td>
<td>3.4</td>
<td>0.6</td>
<td>0.8</td>
<td>-60°</td>
</tr>
<tr>
<td>V36</td>
<td>c</td>
<td>c</td>
<td>0.6</td>
<td>3.4</td>
<td>180°</td>
</tr>
<tr>
<td>V45</td>
<td>c</td>
<td>c</td>
<td>0.3</td>
<td>3.0</td>
<td>180°</td>
</tr>
<tr>
<td>V47</td>
<td>1.5</td>
<td>1.1</td>
<td>0.7</td>
<td>2.9</td>
<td>180°</td>
</tr>
<tr>
<td>V48</td>
<td>0.7</td>
<td>2.7</td>
<td>c</td>
<td>c</td>
<td>-60°</td>
</tr>
<tr>
<td>V52</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td></td>
</tr>
<tr>
<td>V72</td>
<td>0.2</td>
<td>1.8</td>
<td>1.5</td>
<td>0.5</td>
<td>d</td>
</tr>
<tr>
<td>V74</td>
<td>f</td>
<td>3.3</td>
<td>0.8</td>
<td>1.2</td>
<td>-60°</td>
</tr>
<tr>
<td>V78</td>
<td>2.1</td>
<td>0.7</td>
<td>0.7</td>
<td>1.7</td>
<td>180°</td>
</tr>
<tr>
<td>V88</td>
<td>c</td>
<td>c</td>
<td>0.9</td>
<td>0.9</td>
<td>d</td>
</tr>
<tr>
<td>V102</td>
<td>1.9</td>
<td>0.7</td>
<td>0.5</td>
<td>3.6</td>
<td>180°</td>
</tr>
<tr>
<td>V144</td>
<td>1.7</td>
<td>1.1</td>
<td>c</td>
<td>c</td>
<td>180°</td>
</tr>
<tr>
<td>V150</td>
<td>1.6</td>
<td>2.7</td>
<td>0.8</td>
<td>2.8</td>
<td>d</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Coupling constants were determined from a $^{13}$C-$^{15}$N and $^{13}$C-$^{13}$C' spin echo CT-HSQC spectra (Grzesiek \textit{et al}., 1993b; Vuister \textit{et al}., 1993). The coupling constants were corrected for a systematic underestimate in the determined coupling constant value inherent in the method by multiplication of the measured value by 1.06 ($^{3}J_{CCY}$) and 1.08 ($^{3}J_{NCY}$) (Damberger \textit{et al}., 1994).

\textsuperscript{b} $\chi 1$ angles were determined on the basis of $^{3}J_{CCY}$ and $^{3}J_{NCY}$ values, as well as intraresidue NOEs determined with a 50 msec mixing time $^{13}$C/$^{15}$N NOESY-HSQC and a 125 msec $^{15}$N NOESY-HSQC.

\textsuperscript{c} not determined due to spectral overlap.

\textsuperscript{d} not assigned due to conflicting evidence for the presence of a single $\chi 1$ angle amongst $^{3}J_{CCY}$ values and/or $^{3}J_{NCY}$ values and/or intraresidue NOEs.

\textsuperscript{e} not determined due to degeneracy of the two V52 methyls in both $^{13}$C and $^{1}$H dimensions.

\textsuperscript{f} coupling constant value too small to accurately measure.
Table 3.3: Coupling Constants\textsuperscript{a} and \(\chi_1\) Assignments\textsuperscript{b} for the Ile and Thr Residues of CBD\textsubscript{NI}.

<table>
<thead>
<tr>
<th>Residue</th>
<th>(3^J_{\text{NC}2}) (Hz)</th>
<th>(3^J_{\text{CC}2}) (Hz)</th>
<th>(\chi_1)</th>
<th>Residue</th>
<th>(3^J_{\text{NC}2}) (Hz)</th>
<th>(3^J_{\text{CC}2}) (Hz)</th>
<th>(\chi_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T8</td>
<td>0.7</td>
<td>2.9</td>
<td>60°</td>
<td>T73</td>
<td>(d)</td>
<td>(d)</td>
<td>(e)</td>
</tr>
<tr>
<td>T21</td>
<td>0.9</td>
<td>2.4</td>
<td>60°</td>
<td>T77</td>
<td>(d)</td>
<td>(d)</td>
<td>(e)</td>
</tr>
<tr>
<td>T27</td>
<td>(c)</td>
<td>3.3</td>
<td>(e)</td>
<td>T91</td>
<td>0.3</td>
<td>2.4</td>
<td>60°</td>
</tr>
<tr>
<td>T29</td>
<td>1.0</td>
<td>2.1</td>
<td>(e)</td>
<td>T96</td>
<td>(c)</td>
<td>2.4</td>
<td>60°</td>
</tr>
<tr>
<td>T58</td>
<td>2.0</td>
<td>0.7</td>
<td>(e)</td>
<td>T103</td>
<td>(d)</td>
<td>(d)</td>
<td>(e)</td>
</tr>
<tr>
<td>T59</td>
<td>(d)</td>
<td>(d)</td>
<td>(e)</td>
<td>T105</td>
<td>(d)</td>
<td>(d)</td>
<td>(e)</td>
</tr>
<tr>
<td>T61</td>
<td>1.7</td>
<td>(c)</td>
<td>(e)</td>
<td>T107</td>
<td>(d)</td>
<td>(d)</td>
<td>(e)</td>
</tr>
<tr>
<td>T65</td>
<td>(d)</td>
<td>(d)</td>
<td>(e)</td>
<td>T111</td>
<td>1.6</td>
<td>0.6</td>
<td>-60°</td>
</tr>
<tr>
<td>T67</td>
<td>(d)</td>
<td>(d)</td>
<td>(e)</td>
<td>T115</td>
<td>1.0</td>
<td>2.6</td>
<td>(e)</td>
</tr>
<tr>
<td>T70</td>
<td>(c)</td>
<td>1.8</td>
<td>(e)</td>
<td>T138</td>
<td>(d)</td>
<td>(d)</td>
<td>(e)</td>
</tr>
<tr>
<td>I4</td>
<td>0.8</td>
<td>3.3</td>
<td>60°</td>
<td>I125</td>
<td>2.2</td>
<td>0.6</td>
<td>-60°</td>
</tr>
<tr>
<td>I54</td>
<td>2.0</td>
<td>0.8</td>
<td>-60°</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a} Coupling constants were determined from a \(^{13}\text{C}-^{15}\text{N}\) and \(^{13}\text{C}-^{13}\text{C}'\) spin echo CT-HSQC spectra (Grzesiek et al., 1993b; Vuister et al., 1993). The coupling constants were corrected for a systematic underestimate in the determined coupling constant value inherent in the method by multiplication of the measured value by 1.06 (\(3^J_{\text{CC}2}\)) and 1.08 (\(3^J_{\text{NC}2}\)) (Damberger et al., 1994).

\textsuperscript{b} \(\chi_1\) angles were determined on the basis of \(3^J_{\text{CC}2}\) and \(3^J_{\text{NC}2}\) values, as well as intraresidue NOEs using a 50 msec mixing time \(^{13}\text{C}/^{15}\text{N}\) NOESY-HSQC and a 125 msec \(^{15}\text{N}\) NOESY-HSQC.

\textsuperscript{c} coupling constant value too small to accurately measure.

\textsuperscript{d} not determined due to spectral overlap.

\textsuperscript{e} not assigned due to conflicting evidence for the presence of a single \(\chi_1\) angle amongst \(3^J_{\text{CC}2}\) values and/or \(3^J_{\text{NC}2}\) values and/or intraresidue NOEs.

Qualitative analysis of the 3D \(^{13}\text{C}-^{13}\text{C}\) long range correlation experiment (Bax et al., 1992) provided \(\chi_2\) restraints for 7 of the 12 leucine residues and 2 of the 3 isoleucine residues. As shown in Figure 3.8, a strong \(^1\text{H}^{81}\text{C}^{13}\text{C}^{\alpha}\) cross peak was observed for Ile4 and Ile125, reflecting a large \(3^J_{\text{C}81\text{C}^{\alpha}}\) coupling indicative of a \textit{trans} conformation (\(\chi_2=180°\)). In the case of Ile54, there
Figure 3.8. Strip plot of a portion of the long range $^{13}\text{C}-^{13}\text{C}$ coupling constant experiment recorded for CBD$_{N1}$ (Bax et al., 1992). Shown are $\omega 1(^{13}\text{C})$-$\omega 3(^{1}\text{H})$ strips for Ile4 and Ile125, which exhibit a strong $^{1}\text{H}^{\delta 1}-^{13}\text{C}^{\alpha}$ cross peaks relative to the $^{1}\text{H}^{\delta 1}-^{13}\text{C}^{\delta 1}$ auto-correlation peaks. This reflects large $^{3}\text{J}^{\text{C}^{\delta 1}\text{C}^{\alpha}}$ couplings, indicative of trans (180°) $\chi_2$ dihedral angles. In contrast, no $^{1}\text{H}^{\delta 1}-^{13}\text{C}^{\alpha}$ cross peak is observed for Ile54 (a box indicates the position at which it is expected). Also, only a weak $^{1}\text{H}^{\delta 1}-^{13}\text{C}^{\gamma 2}$ cross peak is observed. Together, this reflects both small $^{3}\text{J}^{\text{C}^{\delta 1}\text{C}^{\gamma 2}}$ and $^{3}\text{J}^{\text{C}^{\delta 1}\text{C}^{\alpha}}$ coupling constants, indicative of a gauche$^+$ conformation ($\chi_2 = -60^\circ$). The $^{13}\text{C}^{\delta 1}$ of Ile54 is marked with an asterisk indicating it is aliased in the $^{13}\text{C}$ dimension shown in this plot. Cross peaks and auto-correlation peaks are shown without distinction of their opposite signs.
was no $^1H\delta_1$-$^13C\alpha$ peak and a weak $^1H\delta_1$-$^13C\gamma^2$ peak. This indicates that both the $^3J_{C\delta 1C\alpha}$ and the $^3J_{C\delta 1C\gamma^2}$ couplings are small, suggesting the side-chain adopts a gauche$^+$ conformation ($\chi^2 = -60^\circ$). In the absence of positive evidence for this conformation, no $\chi^2$ torsion restraint was included for Ile54 during the structure calculations. However, in the final ensemble of structures, this $\chi^2$ angle was indeed well defined at -65$^\circ$ with a chi2 angular order parameter of 0.99 ($S(\chi^2)$; Hyberts et al., 1992).

**Secondary Structure Determination.**

Initial studies using CD and FT-IR spectroscopy indicated that CBD$_{N1}$ is composed of $\beta$-strands and devoid of helices (chapter 2). This global analysis was confirmed when the regular secondary structural elements of CBD$_{N1}$ were determined based on patterns of amide hydrogen exchange rates, sequential and cross strand NOEs, $^3J_{HN-H\alpha}$ coupling constants, and $^13C\alpha$, $^13C\beta$, $^13C\gamma$, and $^1H\alpha$ chemical shifts. The information defining the $\beta$-strands in CBD$_{N1}$ is summarized in Figure 3.9. In general, residues in $\beta$-strand conformations are characterized by $^3J_{HN-H\alpha}$ $\geq$ 8 Hz, positive $^1H\alpha$ and $^13C\beta$ secondary chemical shifts ($\delta_{\text{observed}} - \delta_{\text{random coil}}$; Wishart et al., 1992; Wishart & Sykes, 1994), negative $^13C\alpha$ and $^13C\gamma$ secondary chemical shifts, strong sequential $^1H\alpha_i-^1H\alpha_{i+1}$ NOEs, cross-strand $^1H\alpha_i-^1H\alpha_j$, $^1H\alpha_i-^1H\gamma_j$, $^1H\alpha_i-^1H\gamma_{j+1}$ NOEs, and protection from exchange due to cross-strand hydrogen bonding (Wüthrich, 1986).

Figures 3.10 summarizes the interresidue NOEs observed in the NOESY spectra of CBD$_{N1}$ (Figure 3.11), as well as the hydrogen bonds used in the structure calculations. Based on the patterns of cross-strand $^1H\alpha_i-^1H\alpha_j$, $^1H\gamma_i-^1H\gamma_j$, and $^1H\gamma_i-^1H\gamma_{j+1}$ NOEs, CBD$_{N1}$ contains two $\beta$-sheets, denoted A and B, each composed of five anti-parallel $\beta$-strands. Defining the exact boundaries of the regular elements of secondary structure in a protein is often difficult. For example, in the case of CBD$_{N1}$, there is some ambiguity in identifying the ends of strands B3 and B4. As seen in Figure 3.10, interstrand NOEs indicate that strand B3 could be defined to start at Ile54. Correspondingly strand B4 would end at Tyr112. The lack of cross strand $^1H\alpha_i-^1H\alpha_j$ NOEs is a result of extensive degeneracy of the $^1H\alpha$ resonances in this region. However, the
Chapter 3-Structure Determination of CBD_{NJ}
Figure 3.9. Summary of amide hydrogen exchange rates, $^{3}J_{HN-H\alpha}$ coupling constants, and $^{1}H\alpha$, $^{13}C\alpha$, $^{13}C\beta$, and $C'$ chemical shifts used to deduce the $\beta$-strand secondary structure of CBD$_{N1}$. The locations of the 5 $\beta$-strands of sheet A are indicated by open arrows, and those of the 5 $\beta$-strands of sheet B are indicated by solid arrows. (i) Hydrogen exchange: Filled circles indicate residues with slow hydrogen-deuterium exchange kinetics ($t_{1/2} > 1000$ min), half-filled circles indicate those with intermediate hydrogen-deuterium exchange kinetics ($10 \text{ min} < t_{1/2} < 1000$ min), and open circles indicate those with fast hydrogen-deuterium exchange kinetics ($t_{1/2} < 10$ min) at $35 ^\circ\text{C}$ and pH* 5.90. (ii) $^{3}J_{HN-H\alpha}$: measured values are reported in Hz. S denotes couplings that are too small to be determined reliably using the HMQC-J experiment. (iii) Main chain $^{1}H\alpha$, $^{13}C\alpha$, $^{13}C\beta$, and $^{13}C'$ chemical shifts are plotted as the difference from the random coil values ($\delta_{\text{observed}} - \delta_{\text{random coil}}$). Residues in $\beta$-strands have a positive change in $^{1}H\alpha$ and $^{13}C\beta$ shift, and a negative change in $^{13}C\alpha$ and $^{13}C'$ chemical shift (Wishart et al., 1992; Wishart & Sykes, 1994).
Figure 3.10. Alignment of the β-strands to form sheets A and B present in CBD$_{N1}$. The NOEs used to deduce these alignments are shown by arrows. Dotted lines indicate the hydrogen bonds included in the structure calculations. Boxed amide hydrogens have slow hydrogen-deuterium exchange kinetics (filled circles in figure 3.9), while circled amide hydrogens have intermediate hydrogen-deuterium exchange kinetics (half-filled circles in figure 3.9). The positions of β-bulges are indicated by jagged lines.
Figure 3.11. $\omega_1(^1H)-\omega_3(^1H)$ strip plots of a portion of the $^{15}N$ HSQC-NOESY spectrum of CBD$_{N1}$, recorded with a mixing time of 125 msec. Selected NOE interactions are labelled, and solid lines connect the strong NOEs of an $^1H_N$ to the $^1H_\alpha$ of the previous residue, indicative of an extended $\beta$-strand conformation. This is an example of the data obtained to derive distance restraints for the structure generating process using the hybrid distance geometry/simulated annealing protocols in XPLOR 3.1.
classification of these residues as having a β-sheet conformation is not supported by the secondary chemical shifts of these residues (Figure 3.9). From the tertiary structure of CBD\textsubscript{N1}, it is found that the polypeptide backbone in these regions turn sharply, linking sheets A and B. Therefore, these sequences were not defined as part of the β-strands in CBD\textsubscript{N1}.

Two strands of CBD\textsubscript{N1} are broken by β-bulges. One bulge, classified by Promotif (Hutchinson & Thornton, 1996) as being classical (Chan et al., 1993), begins at residue Thr87, following which Val88 and Leu89 both lie in the hydrophobic core of the protein. The second bulge starts at Leu141, after which Asp142 and Asp143 both have their side-chains on the exterior of the protein. The presence of both of these bulges is evident from the secondary chemical shifts shown in Figure 3.9. This emphasizes the potential wealth of structural information contained in NMR chemical shifts.

From the topological arrangement of the β-strands of CBD\textsubscript{N1}, it is evident that this protein adopts a jelly-roll β-sandwich structure (Figure 3.12; Brandon & Tooze, 1991). The jelly-roll comprises strands A2 to A5 and B2 to B5, with the two short strands A1 and B1 appended along one side of this core motif. Strands A1 / B1 and A4 / B4 are not connected together by hydrogen bonding, thus defining the structure as a β-sandwich as opposed to a continuous β-barrel. This is evident from the protection patterns of the backbone amide $^1$HN protons in the hydrogen-deuterium exchange experiments (Figures 3.8 & 3.9). The outer edges of the β-sheets do not show protection for the backbone amide $^1$HN protons.

**Tertiary Structure.**

A total of 1988 distance, hydrogen bond, and dihedral restraints were used to calculate 60 structures following the hybrid distance geometry/simulated annealing protocol (Nilges et al., 1988) with X-PLOR 3.1 (Brünger, 1992). The 25 structures with the lowest total energy and fewest NOE violations were selected for comparison. None of these had NOE violations greater than 0.4 Å, and, except for the two $\chi_1$ restraints involving Cys33 and Cys140, none of the 25
Figure 3.12. Schematic diagram showing the jelly-roll $\beta$-sandwich topology of \( \text{CBD}_{N1} \). Sheets A and B are indicated by open and solid arrows, respectively, and the position of the disulphide between Cys33 and Cys140 is indicated. Strands A2-A5 and B2-B5 comprise the jelly-roll motif, with the two short strands A1 and B1 appended along one side. The global structure of CBD$_{N1}$ can be envisioned by folding the figure such that sheet B lies below sheet A, and that sheet A is concave. The lengths of the strands and loops are not drawn to scale.
structures had dihedral violations greater than 4.3°. Statistics for the 25 accepted structures are listed in Table 3.4.

The superposition of the final ensemble of structures calculated for CBD_{N1} is shown in Figure 3.13. The structural ensemble is clearly consistent with the jelly-roll β-sandwich topology deduced at the level of secondary structure analysis. The backbone conformation of the strands in each of the sheets is well determined, having an rmsd of 0.44 ± 0.05 Å with respect to the average structure. Apart from the N- and C-termini, the regions which have the highest rms deviation from the average structure are between residues 20-30, which contains the short β-strand B1 involving residues 25-27, and in the loops between residues 37-44, 81-85, and 113-121. All these stretches contain no, or few, long distance restraints (Figure 3.14A). As shown in Figures 3.14B and 3.14C, these stretches are also regions where the angular order parameters S(φ) and S(ψ) are the lowest, indicative of local disorder (Hyberts et al., 1992).

The stereochemical quality of the backbone coordinates for the ensemble of 25 structures was checked using the programmes Procheck and Procheck-NMR (Morris et al., 1992; Laskowski et al., 1993). For this ensemble, 98% of the residues lie in the allowed regions of the Ramachandran plot (figure 3.15). The few residues with main chain dihedral angles that often fall outside the allowed regions, namely Glu14, Ala83, and Tyr85, are all found in the parts of CBD_{N1} that exhibit high rms deviations, low angular order parameter, and the lack of regular secondary structure.

As shown in Figure 3.16, the side-chains that make up the hydrophobic core of CBD_{N1} are well defined structurally. This is reflected by both low rms deviations and high angular order parameters S(χ1) (Figure 3.14D, E). Of the 33 side-chains that comprise the hydrophobic core of CBD_{N1}, 32 have values of S(χ1) greater than 0.98. The one exception is for Leu25 which has a value of 0.62. This residue lies in the strand that is at one edge of face B, in an area that is not well defined as discussed above. In addition, the Leonard-Jones energy of each of the accepted structures is large and negative (Table 3.4) indicating that no unfavorable van der Waals contacts exist.
### Table 3.4: Structural Statistics and Atomic RMS Differences

<table>
<thead>
<tr>
<th>Description</th>
<th>&lt;SA&gt;</th>
<th>&lt;SA&gt;_{av}</th>
</tr>
</thead>
<tbody>
<tr>
<td>rmsd from experimental distance restraints(^b) (Å) (1793)</td>
<td>0.013 ± 0.002</td>
<td>0.011</td>
</tr>
<tr>
<td>rmsd from experimental dihedral restraints(^c) (deg.) (195)</td>
<td>0.45 ± 0.13</td>
<td>0.29</td>
</tr>
</tbody>
</table>

**Deviations from idealized geometry**

<table>
<thead>
<tr>
<th>Description</th>
<th>&lt;SA&gt;</th>
<th>&lt;SA&gt;_{av}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonds (Å) (2129)</td>
<td>0.0019 ± 0.0002</td>
<td>0.0016</td>
</tr>
<tr>
<td>Angles (deg) (3841)</td>
<td>0.54 ± 0.01</td>
<td>0.52</td>
</tr>
<tr>
<td>Improper(^d) (deg) (1102)</td>
<td>0.41 ± 0.02</td>
<td>0.39</td>
</tr>
</tbody>
</table>

**XPLOR Energies\(^e\) (kcal mol\(^{-1}\))**

<table>
<thead>
<tr>
<th>Energy Type</th>
<th>Value</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>E_{NOE}</td>
<td>16.53 ± 3.9</td>
<td>10.72</td>
</tr>
<tr>
<td>E_{cdih}</td>
<td>2.64 ± 1.4</td>
<td>1.01</td>
</tr>
<tr>
<td>E_{imp}</td>
<td>28.2 ± 2.3</td>
<td>25.5</td>
</tr>
<tr>
<td>E_{angle}</td>
<td>168.2 ± 6.6</td>
<td>158.6</td>
</tr>
<tr>
<td>E_{bond}</td>
<td>7.5 ± 1.3</td>
<td>5.7</td>
</tr>
<tr>
<td>E_{vdw}</td>
<td>13.9 ± 3.4</td>
<td>10.3</td>
</tr>
<tr>
<td>E_{L-J}(^f)</td>
<td>-569.2 ± 19.8</td>
<td>-612.5</td>
</tr>
</tbody>
</table>

**Atomic rms differences\(^g\) (Å)**

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>residues 4-148</td>
<td>backbone(^h)</td>
<td>all heavy atoms</td>
</tr>
<tr>
<td>β-sheet regions(^i)</td>
<td>0.79 ± 0.14</td>
<td>1.60 ± 0.20</td>
</tr>
<tr>
<td></td>
<td>0.44 ± 0.05</td>
<td>1.61 ± 0.20</td>
</tr>
</tbody>
</table>

\(^a\) <SA> represents the final ensemble of 25 simulated annealing structures; <SA>\_{av} is the restrained minimized average structure obtained by averaging the 25 structures over residues 4-148. Errors reported are ± one standard deviation. The number of restraints is given in parentheses.

\(^b\) This includes 1705 NOE derived distance restraints, and 88 hydrogen bond restraints (44 hydrogen bonds).

\(^c\) Torsion angle restraints include 56 φ-angle restraints, 88 ψ-angle restraints, 42 χ\(_1\)-angle restraints, and 9 χ\(_2\)-angle restraints based on \(^3\)J\(_{HN-H\alpha}\), \(^3\)J\(_{NC\gamma}\) and \(^3\)J\(_{CC\gamma}\) measurements, qualitative analysis of \(^3\)J\(_{C\delta C\gamma}\) values, and intraresidue and sequential NOEs from NOESY spectra.

\(^d\) Improper torsion angle restraints maintain peptide planarity and chirality.

\(^e\) The square well NOE (E\_{NOE}) using center averaging, the restrained dihedral (E\_{cdih}), the improper torsion angle (E\_{imp}), the angle (E\_{angle}), the bond (E\_{bond}), and the quartic van der Waals repulsion energies (E\_{vdw}) were calculated using force constants of 50 kcal mol\(^{-1}\) Å\(^{-2}\), 200 kcal mol\(^{-1}\) rad\(^{-2}\), 500 kcal mol\(^{-1}\) rad\(^{-2}\), 1000 kcal mol\(^{-1}\) Å\(^{-2}\), and 4 kcal mol\(^{-1}\) Å\(^{-4}\). The van der Waals repulsion energy was calculated using sphere radii set to 0.75 times that supplied with XPLOR (parallhdg.pro).

\(^f\) E\_{L-J} is the Lennard-Jones van der Waals energy. This term was not included in any of the structure generating steps but was calculated for the final 25 structures and the restrained minimized average structure.

\(^g\) Atomic rms differences were calculated using the average structure before restrained minimization.

\(^h\) Atoms used were N, C\(^\alpha\), and C\(^\beta\).

Figure 3.13. Two views of the Cα traces of the final ensemble of 25 structures calculated for CBD\textsubscript{N1}. The co-ordinates were superposed using the backbone atoms from residues 4-148. The jelly-roll β-sandwich topology is apparent in view (A), whereas the presence of a binding cleft formed by β-sheet A is clearly evident in view (B). View A looks down on the binding cleft, with view (B) rotated by approximately 90° compared to (A). The figure was produced using the programme Molscript (Kraulis, 1991).
Figure 3.14. (A) Distribution of NOE restraints per residue. Filled bars represent the number of long range NOEs (|i - j| > 4), open bars represent the sum of non-trivial intraresidue, sequential and short range NOEs (1 < |i - j| ≤ 4). For every restraint the originating and destination residues were each counted once. (B, C, D) Angular order parameters $S_\phi$, $S_\psi$, and $S_{\chi_1}$ for the $\phi$ and $\psi$ main chain, and side chain $\chi_1$ dihedral angles, respectively, observed in the final ensemble of 25 structures (Hyberts et al., 1992). (E, F) The rms deviation for all heavy atoms and main chain atoms, respectively, for the ensemble of 25 structures, aligned against the average structure obtained by superimposing residues 4-148. The locations of the ten $\beta$-strands are indicated on the top of the figure.
Figure 3.15. Ramachandran plot of the mean minimized CBDN1 structure generated by PROCHECK (Laskowski et al., 1988). The degrees of shading indicate the most favored, allowed and generously allowed regions of phi / psi space from an analysis of 118 high-resolution crystal structures. Glycine residues are represented as triangles, all other non-proline residues as squares.
Figure 3.16. Superposition of the co-ordinates of the side-chains that make up the hydrophobic core of CBD$_{N1}$ on the C$\alpha$ trace of the minimized average structure (thick line) of this protein. All heavy atoms between residues 4 and 148 were used to superimpose the 25 structures. The CBD is oriented with the binding cleft towards the page and $\beta$-sheet B closest to the reader. This figure was produced using the programme Molscript (Kraulis, 1991).
The presence of a cis-peptide bond for Ala83-Pro84 in CBD$_{N1}$ was initially identified based on the $^{13}$C$_7$ chemical shift of the proline at 24.9 ppm. This is approximately 3 ppm upfield from the value found for prolines with trans-peptide linkages (Stanczyk et al., 1989). The cis-peptide bond was further confirmed by the observation of a strong $^1$H$^\alpha_{A83}$ to $^1$H$^\delta_{P84}$ NOE, and the lack of an $^1$H$^\alpha_{A83}$ to $^1$H$^\delta_{P84}$ NOE in the 50 msec mixing time $^{13}$C/$^{15}$N NOESY-HSQC of CBD$_{N1}$ (Wüthrich, 1986). This NOE pattern is reversed for every other proline in CBD$_{N1}$, all of which are trans-linked. Pro84 lies in a turn between strands A3 and A4, classified as type VIa1 by Promotif (Hutchinson & Thornton, 1996). This type of turn requires a cis-linked proline at position i+2.

A disulphide bond between Cys33 and Cys140 was identified initially from the fact that the $^{13}$C$^\beta$ chemical shift of a cysteine is indicative of its oxidation state. The $^{13}$C$^\beta$ shift of Cys33 is 47.4 ppm and that of Cys140 is 47.5 ppm. These are within the expected range for an oxidized Cys, which has a random coil value of 41.8 ppm, as opposed to 28.6 ppm for the reduced form (Wishart & Sykes, 1994). This conclusion is consistent with the results of a DTNB titration of CBD$_{N1}$, demonstrating that there are no free thiols present in the protein. Cross-strand NOEs, illustrated in Figure 3.10, also confirm the pairing of these two cysteine residues.

The disulphide bond connecting Cys33 and Cys140 is unusual in that it bridges two $\beta$-strands (Figures 3.10 & 3.17). Such covalent bridges between paired $\beta$-strands are not observed commonly, although examples are found in azurin and chymotrypsin (Thornton, 1981). The disulphide in CBD$_{N1}$ is classified as a short right-handed hook by Promotif (Hutchinson & Thornton, 1996). Although the ($\phi$, $\psi$) angles of the cysteines and their neighboring amino acids do deviate slightly from those expected for residues in ideal anti-parallel $\beta$-strands, the presence of the disulphide bond does not produce any pronounced distortions in strands B2 and B5. The backbone dihedral angles of the cysteines and adjacent residues still lie within the most favored regions of the Ramachandran plot and are well defined. In contrast, I observe that the $\chi_1$ restraints for Cys33 and Cys140 were often violated during the structural calculations. This may reflect conformational isomerization of the disulphide, particularly since the side-chain dihedral angles
A

B

Figure 3.17. Molscript ribbon diagram of the minimized average structure of CBD\textsubscript{N1} with residues from β-sheet A that are implicated in oligosaccharide binding (chapter 2; Johnson et al., 1996a) shown in ball-and-stick format. Hydrophilic residues are identified with black atoms and hydrophobic residues with white atoms. The disulphide bond between β-strands B2 and B5 is also presented using gray atoms and dark gray bonds. View (A) looks directly down on the binding face, whereas view (B) is rotated by 90° to emphasize the binding cleft. These are the same molecular orientations used in figure 3.13. Strands B1-B5 are labelled in italics.
were determined based upon a staggered rotamer model using qualitatively estimated coupling constants.

**Discussion**

**CBD\textsubscript{NI} Structure.**

The structure of CBD\textsubscript{NI} in the presence of saturating concentrations of cellotetraose was determined using NMR methods. At the time this work was completed it was the third CBD structure published, the first of a Family IV CBD (Tomme *et al.*, 1995). CBD\textsubscript{NI}, like all other CBDs of known conformation, is composed of anti-parallel $\beta$-strands. However, in contrast to CBD\textsubscript{CBH1} and CBD\textsubscript{CEX}, CBD\textsubscript{NI} has a jelly-roll $\beta$-sandwich topology with a disulphide bond bridging two adjacent $\beta$-strands. Furthermore, as a consequence of the strands of sheet A being shorter than those of sheet B, CBD\textsubscript{NI} contains a prominent cleft that runs across one face of the protein. As discussed below, this cleft is the binding site for cellooligosaccharides and amorphous cellulose. The individual $\beta$-strands in CBD\textsubscript{NI} are very well ordered, while the protruding loops are much less defined. In particular, the loops formed by residues 37-44, 82-84, and 113-121 are the most disordered, as evident by their high rms deviations and low angular order parameters (Figures 3.12 and 3.13). These loops form the extreme edges of the binding face, and thus a range of depths and widths are found for the dimensions of the binding cleft within the set of 25 accepted structures. An estimate of the observed variation in groove width is provided by the distances between the C$^\alpha$ atoms of Tyr43 and Ala118, which range from 24 to 31 Å in the ensemble of structures. As will be discussed in chapter 5 using backbone $^{15}$N relaxation methods, this is a function of true mobility and not a result of having few long-range restraints for residues in these loop regions of CBD\textsubscript{NI}. 
Binding Mechanism.

Previously, I determined that CBD<sub>N1</sub> binds soluble cellobiooligosaccharides in order of increasing affinity cellotriose < cellotetraose < cellopentaose ~ cellohexaose (chapter 2; Johnson <i>et al.</i>, 1996a; Tomme <i>et al.</i>, 1996a). In addition to binding amorphous cellulose, CBD<sub>N1</sub> also associates with oat- and barley-β-glucans with affinities equal to that of cellopentaose (Tomme <i>et al.</i>, 1996a). Together, these results implied that the binding site spans approximately the length of a cellopentaose molecule. From the patterns of amide <sup>15</sup>N and <sup>1</sup>H N chemical shift perturbations resulting from the addition of cellobiooligosaccharides to CBD<sub>N1</sub>, it was also demonstrated that residues in strands A1-A5 all interact with these soluble ligands (chapter 2; Johnson <i>et al.</i>, 1996a). Therefore, β-sheet A was identified as the binding face of the N-terminal CBD of CenC. The involvement of the aromatic side-chains of Tyr19 and Tyr85 in sugar binding was also suggested based on changes in the NMR chemical shifts and lineshapes of these residues and on perturbations of the near ultraviolet absorption spectrum of CBD<sub>N1</sub> due to addition of cellobiooligosaccharides (Johnson <i>et al.</i>, 1996a).

Expanding upon these initial studies, I have now shown that the binding face of CBD<sub>N1</sub> is in fact a groove or cleft that extends across β-sheet A (Figure 3.17). The identification of this β-sheet as the ligand binding cleft of CBD<sub>N1</sub> is confirmed by the observation of intermolecular NOEs between bound cellotetraose and residues located within this region of the protein. Although the <sup>1</sup>H resonances of the cellotetraose are currently unassigned, NOEs from unlabelled sugar protons to <sup>13</sup>C-labelled Tyr19 <sup>1</sup>H<sup>δ</sup> and <sup>1</sup>H<sup>ε</sup>, V48 <sup>1</sup>H<sup>γ</sup> and <sup>1</sup>H<sup>δ2</sup>, L77 <sup>1</sup>H<sup>δ2</sup>, Y85 <sup>1</sup>H<sup>δ</sup> and <sup>1</sup>H<sup>ε</sup>, and A126 <sup>1</sup>H<sup>β</sup> were detected using a <sup>13</sup>C 01-edited, 03-filtered HMQC-NOESY experiment (Lee <i>et al.</i>, 1994). Consistent with the reported dependence of binding affinity on the degree of oligosaccharide polymerization, the length of the CBD<sub>N1</sub> binding groove is approximately equal to that of an extended cellopentaose chain.

Using the NMR-derived tertiary structure of CBD<sub>N1</sub>, it is now possible to look more closely at the residues involved in oligosaccharide binding, particularly in light of the thermodynamic studies of the interactions between CBD<sub>N1</sub> and various sugars (Tomme <i>et al.</i>,
1996a). Figure 3.17 shows the exposed side-chains present on the amphipathic binding face of the averaged minimized structure of CBD$_{N1}$. Two distinct features of CBD$_{N1}$ are evident from this figure. First, a strip of hydrophobic residues, comprised of Val17, Tyr19, Val48, Leu126, and Leu77, runs along the center of the binding cleft. Tyr 85, which may also be involved in binding, is located in a loop region near this hydrophobic strip. These non-polar residues may contact the pyranose rings of the oligosaccharide, providing favorable hydrophobic and van der Waals interactions. Second, there are numerous hydrophilic groups, including Asn50, Arg75, Asn81, Asp90, Thr87, Gln124, and Gln128, that flank the hydrophobic strip by lining the sides of the binding cleft. The polar residues likely provide hydrogen bonds to the equatorial hydroxyl groups of the sugar rings. In accordance with the relative insensitivity of sugar binding to pH and ionic strength (Tomme et al., 1996a), only two of these seven polar groups are ionizable. Although CBD$_{N1}$ was investigated in the presence of saturating cellotetraose, the sugar was not included in the structural calculations. Thus, it is not possible to state confidently that all of these polar and non-polar residues are involved directly in ligand binding. To address this issue, the specific roles played by these amino acids are being studied by site-directed mutagenesis (Jeff Kormos and Peter Tomme, Department of Microbiology, UBC).

This proposed structural mechanism of oligosaccharide binding is entirely consistent with the thermodynamic parameters characterizing the association of soluble oligosaccharides to CBD$_{N1}$ (Tomme et al., 1996a). Based upon detailed calorimetric analyses, it was reported that sugar binding results in a favorable enthalpic change, compensated in part by a decrease in entropy. This implies that a predominance of polar interactions, such as hydrogen bonding, provides the primary driving force for binding. The dominant role of hydrogen bonding is observed with most other carbohydrate-binding proteins (Quiocho, 1986, 1989).

Finally, and perhaps most importantly, the presence of a binding cleft in the structure of CBD$_{N1}$ provides a simple explanation as to why this protein has the ability to bind soluble oligosaccharides and amorphous cellulose, but not crystalline cellulose. The residues that mediate binding are all located within this cleft, and as a result, are unable to interact with the flat surface
presented by crystalline cellulose. In contrast, soluble sugars and single strands of amorphous cellulose can bind to CBD$_{N1}$ by lying within this groove.

Comparison to Other Family IV CBDs.

CBD$_{N1}$ is the first Family IV CBD for which a structure has been determined. Figure 3.18 presents the sequences of four members of this family (Tomme et al., 1995), aligned with the secondary structural elements identified in CBD$_{N1}$. This alignment shows that the residues forming the eight strands of the jelly-roll β-sandwich motif of CBD$_{N1}$ are in general well conserved, with deletions/insertions found in the intervening loop regions and the small β-strands A1 and B1. The two cysteines that form the disulphide bridge are invariant, and most of the residues involved in ligand binding by CBD$_{N1}$ are conserved. Accordingly, it is reasonable to postulate that CBD$_{N2}$ from CenC, as well as the related CBDs from Thermomonospora fusca E1 and Streptomyces reticuli Cell, adopt secondary and tertiary structures similar to those determined for CBD$_{N1}$. Although it remains to be demonstrated if the T. fusca E1 and S. reticuli Cell CBDs exhibit the same specificity for soluble forms of cellulose, NMR studies of isolated CBD$_{N2}$ confirms that the second CBD from CenC does indeed bind cellooligosaccharides (E. Brun, pers. comm.). It has also been shown that the three dimensional structure of CBD$_{N2}$ is very similar to that of CBD$_{N1}$ (E. Brun, pers. comm.). This leads to the more tantalizing questions as to whether or not inter-domain interactions exist between CBD$_{N1}$ and CBD$_{N2}$ within the native enzyme and why, when linked together in CBD$_{N1N2}$, do these two domains appear to bind phosphoric-acid swollen cellulose in an independent and not co-operative manner (Tomme et al., 1996a).

Comparison to CBDs that bind crystalline cellulose.

The dominant feature of the structure of CBD$_{N1}$ is the presence of a binding cleft. This distinguishes the N-terminal CenC CBD from all other characterized CBDs. CBD$_{CBHI}$ (Kraulis et al., 1989), CBD$_{Cex}$ (Xu et al., 1995), CBD$_{Cip}$ (Tormo et al., 1996) and CBD$_{EGZ}$ (Brun et al., 1997) all have flat binding surfaces that contain three exposed aromatic rings. Paralleling this
Figure 3.18. Alignment of the amino acid sequences of four family IV CBDs (Tomme et al., 1995), C. fimi CenC (CBD_{N1} and CBD_{N2}), Thermomonospora fusca E1 and Streptomyces reticuli Cell. The alignment was obtained using the program PHD (Rost, 1996). Boxes highlight positions where residues are conserved in three or more family members. The secondary structure of CBD_{N1} is shown as open and filled boxes to represent the β-strands of sheet A and sheet B, respectively. The residues in sheet A that are implicated in cellobioisaccharide binding are shaded.
structural difference, the other CBDs also stand apart from $CBD_N1$ in their affinity for crystalline cellulose and in the different thermodynamic forces that lead to carbohydrate binding by the $CBD_{Cex}$ protein domain. Whereas the association of $CBD_N1$ with soluble oligosaccharides and phosphoric acid-swollen cellulose is enthalpically driven (Tomme et al., 1996a), the affinity of $CBD_{Cex}$ for insoluble bacterial micro-crystalline cellulose results primarily from a favorable increase in entropy, indicative of a hydrophobic interaction (Creagh et al., 1996). Furthermore, it is known that the exposed tryptophans of $CBD_{Cex}$ are involved in the binding event (Poole et al., 1993; Din et al., 1994; Bray et al., 1996). I postulate that $CBD_{Cex}$ relies on hydrophobic stacking of these aromatic rings with the flat surface of crystalline cellulose as the hydroxyl groups of the glucosyl residues are involved primarily in interactions with adjacent polysaccharide chains. In contrast, $CBD_N1$ associates with a single strand of cellulose by exploiting a binding cleft in which polar residues are positioned to hydrogen bond to the exposed equatorial hydroxyl groups of the glucopyranose rings. These distinct structural and thermodynamic mechanisms highlight the complexity of cellulose as a substrate for enzymatic recognition and degradation.

**Structural Similarity with 1,3-1,4-β-Glucanase.**

The tertiary structure of $CBD_N1$ closely resembles those of *endo*-1,3-1,4-β-glucanases from *Bacillus macerans* (Hahn et al., 1995a) and *Bacillus licheniformis* (Hahn et al., 1995c) as well as a hybrid *Bacillus* 1,3-1,4-β-glucanase (Keitel et al., 1993; Hahn et al., 1995b). Although the larger, 1,3-1,4-β-glucanases are composed of two β-sheets of seven strands each, as opposed to five strands each for $CBD_N1$, all these proteins share a common jelly-roll β-sandwich fold (Figure 3.19). Using the programme DALI (Holm & Sanders, 1995), *C. fimi* $CBD_N1$ and the hybrid *Bacillus* 1,3-1,4-β-glucanase were found to have 118 residues aligned with an rmsd of 3.7 Å based on superposition of Cα co-ordinates. This alignment occurs despite only 8% sequence identity of these residues.
Figure 3.19. Molscript ribbon diagrams of (A) the minimized average structure of *C. fimi* CBD$_{N1}$ and (B) the crystal structure of the hybrid *Bacillus* 1,3-1,4-$\beta$-glucanase (Keitel et al., 1993; Hahn et al., 1995b). These two proteins share a common jelly-roll $\beta$-sandwich topology.
As pointed out by Hahn et al. (1995b), the prokaryotic 1,3-1,4-β-glucanases are topologically related to other polysaccharide degrading enzymes such as cellobiohydrolase I (Devine et al., 1994) and 1,4-β-xylanase II (Törrönen et al., 1994), both from Trichoderma reesei. Thus, it is not surprising that the programme DALI identified these as well as Bacillus circulans xylanase (Campbell et al., 1993) and an S-lectin (Liao et al., 1994) as having tertiary structures similar to that of CBD$_{N1}$. The lectin also has a jelly-roll fold. It is intriguing that these polysaccharide-binding domains resemble structurally the catalytic domains of several polysaccharide-degrading enzymes.

The 1,3-1,4-β-glucanases are a distinct family of glucanohydrolases that specifically cleave 1,4-β-D-glucosidic bonds that are adjacent to the β-1,3 linkages in mixed 1,3- and 1,4-linked β-glucans (Anderson & Stone, 1975). In addition, the hybrid 1,3-1,4-β-glucanase hydrolyses cellobiohexaose, but with poor efficiency (Hahn et al., 1995b). All bacterial endo-1,3-1,4-β-glucanases known to date share sequence similarities with endo-1,3-β-glucanases and have been classified into glycosyl hydrolase family 16 (Henrissat, 1991; Henrissat & Bairoch, 1993). The observation that CBD$_{N1}$ appears structurally related to these β-glucanases is particularly interesting given that this CBD also binds the mixed 1,3- and 1,4-linked oat- and barley-β-glucans (chapter 2; Tomme et al., 1996a). These two soluble oligosaccharides, along with lichenan from the Icelandic moss Cetraria islandica, are natural substrates for the 1,3-1,4-β-glucanases.

Furthermore, the catalytic residues Glu105 and Glu109 of the hybrid 1,3-1,4-β-glucanases are located on opposite sides of the active site of the enzyme, whereas the similarly spaced Gln124 and Gln128 lie on the two sides of the CBD$_{N1}$ binding cleft (figure 3.20). The structural similarities between CBD$_{N1}$ and the 1,3-1,4-β-glucanases, combined with their binding to a common class of polysaccharides, suggests that these proteins are evolutionarily related, perhaps through divergence from a common ancestor with a jelly-roll sugar-binding motif. The unusual function of CBD$_{N1}$ in binding soluble or amorphous, but not crystalline glucans, may stem from an evolutionary relationship with the 1,3-1,4-β-glucanases, and not an ancestral binding-domain shared by the members of other CBD families.
Figure 3.20. Two views of the superimposition of the structures of CBD$_{N1}$ (purple) and the hybrid *Bacillus* 1,3–1,4–β–glucanase (green). Shown in red are the side-chains of the catalytic glutamate residues of the 1,3–1,4–β–glucanase (E105, E109) and the corresponding glutamine residues in CBD$_{N1}$ (Q124, Q128).
Chapter 4

Calcium Binding By CBD$_{N1}$

Abstract

The interaction of CBD$_{N1}$ with calcium was studied by NMR spectroscopy. The association constant for calcium binding by CBD$_{N1}$ is $(1.1 \pm 0.5) \times 10^5 \text{ M}^{-1}$ at $35 \, ^\circ\text{C}$ pH 6.0. The oligosaccharide binding ability of CBD$_{N1}$ is not affected by the presence of this metal ion, as determined by the similarity of binding constants of calcium-loaded and calcium-free CBD$_{N1}$ with cellopentaose. On the basis of the observed patterns of amide chemical shift changes, and similarity with the family of Bacillus 1,3-1,4-$\beta$-glucanase structures, the putative calcium ligating atoms on CBD$_{N1}$ are T8 O, G30 O, D142 O and a side-chain oxygen of D142. From analysis of lineshapes of resonances that are in intermediate exchange on the NMR chemical shift timescale, the on-rate for calcium association is determined to be $(5 \pm 2) \times 10^7 \text{ s}^{-1} \text{ M}^{-1}$. This is within a couple of orders of magnitude of the diffusion limit, suggesting that the calcium-binding site on CBD$_{N1}$ is largely preformed.

Introduction

Background

When binding studies of CBD$_{N1}$ with cellooligosaccharides were initially performed (chapter 2) it was noted that the resonances of approximately 13 amides near the N-terminus and adjacent to the disulphide bond were not observed in the $^1\text{H}-^{15}\text{N}$ HSQC spectrum of the protein.
These resonances were detected only upon addition of excess cellotetraose, but not with any other cellooligosaccharide studied. The appearance of these resonances with the addition of cellotetraose was observed for at least seven different samples of CBD$_{NI}$ over a period of 1.5 years, using several sugar samples with the same lot number. However, when subsequent relaxation experiments (chapter 5) were performed, it was observed that these peaks no longer appeared upon addition of cellotetraose from a new lot. The identity of the new sample of cellotetraose was checked by mass spectrometry and it was found to have the expected molecular weight. Also, the new cellotetraose eluted with the same HPLC retention time as a sample of the previously used cellotetraose, obtained from an old NMR sample.

With confirmation of the identity of the new cellotetraose, I concluded that the original sample of cellotetraose, and not any of the other cellooligosaccharides, was contaminated with a substance that bound CBD$_{NI}$ and caused the missing amide peaks to appear. Given that the hybrid Bacillus, the Bacillus macerans and the Bacillus licheniformis 1,3-1,4-β-glucanase structures and CBD$_{Cip}$, all of which share a similar β-jelly roll sandwich fold as CBD$_{NI}$, binds calcium (Keitel et al., 1993; Hahn et al., 1995a, Hahn et al., 1995c, Tormo et al., 1996) it was strongly suspected that this metal ion could be the contaminant. More recently the structure of the family III CBD from the Thermomonospora fusca endo/exocellulase E4 was reported (CBD$_{E4}$; Sakon et al., 1997). This CBD, which has a similar structure as CBD$_{NI}$, also contains a bound calcium ion.

**Experimental Methods**

*Sample preparation*

The unlabelled and $^{15}$N-labelled protein samples were overexpressed and purified as described previously (chapter 2). The sample of CBD$_{NI}$ nonrandomly fractionally $^{13}$C enriched at a level of 10% is that used previously for the resonance assignment of CBD$_{NI}$ (chapter 3). All buffers were passed over a Chelex-100 column to remove metal ions. Plasticware, glassware and
NMR tubes were soaked overnight in 4N HCl, then extensively rinsed with deionized water before use to remove any calcium present. The only exception was for the microsep concentrators (Filtron) which were soaked in 1N HCl, according to manufacturers recommendations. Ethylenediamine tetraacetate (EDTA), added to the unlabelled and $^{15}$N-labelled CBD$_{N1}$ sample to remove metal ions bound to the protein, was subsequently removed by extensive buffer exchange using a microsep concentrator. Some protein is lost with numerous buffer exchanges, and thus in an effort to conserve material, the 10% $^{13}$C-labelled CBD$_{N1}$ sample was not treated with EDTA. As a result, it was approximately 20% calcium-loaded. The titration of this sample was only performed to obtain qualitative data on chemical shift changes with calcium binding. The buffer used for the NMR experiments presented in this chapter was 50 mM sodium chloride, 50 mM sodium acetate ($^2$H$_3$), 0.02% sodium azide, pH 6.0, 10% D$_2$O / 90% H$_2$O. Acetate only binds calcium weakly, with a low $K_a$ of about 100 M$^{-1}$ (Brian Sykes, personal communication), and thus does not compete significantly with CBD$_{N1}$.

**NMR spectroscopy**

One dimensional NMR spectra of 0.4 mM unlabelled CBD$_{N1}$ samples were used to screen for metal binding. $^1$H-$^{15}$N HSQC spectra were obtained using a 1.4 mM sample of uniformly $^{15}$N enriched CBD$_{N1}$. Regular and constant-time versions of $^1$H-$^{13}$C HSQC experiments were recorded using a 0.5 mM sample of CBD$_{N1}$ nonrandomly fractionally $^{13}$C enriched at a level of 10%. Experiments were performed on a Varian Unity 500 MHz spectrometer equipped with a triple resonance probe and pulsed field gradients. All spectra were recorded at 35 ºC and analyzed using a combination of NMRPipe (Delaglio et al., 1995), FELIX v2.30 and FELIX95 (Biosym Technologies). The $^1$H-$^{15}$N HSQC experiments were recorded using the enhanced-sensitivity pulsed field gradient of approach of Kay et al. (1992). Selective water flip back pulses were incorporated to minimize the perturbation of the bulk water magnetization (Grzesiek & Bax, 1993; Zhang et al, 1994). The initial delays in the indirectly detected dimensions were set to $1/(2*sw)$,
resulting in a 180° first order phase shift across the transformed spectrum and inversion of aliased peaks (Bax et al., 1991).

**Determination of binding constants**

The binding of calcium to CBDN1 at 35 °C, pH 6.0 was monitored quantitatively using 1H-15N NMR spectroscopy. A stock 0.5 M solution of calcium chloride was made by weight in exactly the same buffer as for CBDN1. Dilute solutions were made subsequently from this stock. The pH values of the CaCl2 solutions were adjusted as necessary to pH 6.0. The initial concentration of CBDN1 was 1.4 mM. Ten 1H-15N HSQC spectra with increasing concentrations of calcium were acquired as 512 x 96 complex points in the 1H and 15N dimensions with spectral widths of 7000 and 1650 Hz respectively.

Equilibrium association constants were determined by nonlinear least-squared fitting of the chemical shift data versus total calcium concentration to the Langmuir isotherm describing the binding of one ligand molecule to a single binding site. The data were fit using the programme PLOTDATA (TRIUMF, UBC, Vancouver). Only data from peaks in the 1H-15N HSQC spectra that are in fast exchange between the free and calcium-bound forms of CBDN1 on the NMR chemical shift time scale were used in the analysis. The same equations and procedures (chapter 2) used to analyse the binding of soluble celooligosaccharides to CBDN1 were applied here.

To establish whether calcium binding has an effect on celooligosaccharide binding, the association constant of fully calcium-loaded CBDN1 with cellopentaose was determined. Ten 1H-15N HSQC spectra with increasing concentrations of cellopentaose were acquired using 512 x 96 complex points in the 1H and 15N dimensions with spectral widths of 6500 and 1450 Hz respectively. The association binding constant of calcium-loaded CBDN1 with cellopentaose was determined using the methods outlined in chapter 2.
Analysis of binding kinetics

Calcium association/dissociation kinetics were determined by lineshape analysis of residues showing intermediate uncoupled two-site exchange during the titration of CBD$_{N1}$ with calcium. This was done manually by comparing the experimental lineshape and frequency with that from simulated spectra. For the experimental spectra used in these comparisons, data were processed using only a linebroadening window function. Spectral simulations were obtained using values for the fraction of protein bound ($f_b$), the total change in chemical shift of a given nucleus between the bound and free forms ($\Delta \delta_{\text{total}}$), the dissociation rate constant ($k_{\text{off}}$) and the transverse relaxation times for the $^1$H or $^{15}$N nucleus in the free ($T_{2f}$) and bound ($T_{2b}$) protein. From the previously determined calcium-binding constant, values of $f_b$ and $\Delta \delta_{\text{total}}$ were calculated. Values of $T_{2f}$ and $T_{2b}$ were estimated from the full linewidths at half height, and checked by comparison of the experimental and calculated lineshapes for free and fully calcium-saturated protein. Spectra were then simulated as a function of $k_{\text{off}}$ using a programme written by Michael Strain (University of Oregon) run as a macro in Felix version 2.3. The spectral simulations produced by this programme are based on the formalism described by Sandström (1982).

Originally, the equations for the lineshapes were derived for one species in equilibrium with another (A $\rightleftharpoons$ B). The case of CBD$_{N1}$ association with calcium differs in that it is a bimolecular process with both free protein (P$_f$) and calcium (Ca$^{2+}$) in equilibrium with bound protein (PCa$^{2+}$) (P$_f$ + Ca$^{2+}$ $\rightleftharpoons$ PCa$^{2+}$). As a result, the association rate is the product of the on-rate constant and the free calcium concentration. To avoid this complication, the data were fit by determining the off-rate ($k_{\text{off}}$), which is a unimolecular process. Finally, $k_{\text{on}}$ can be determined as it is the product of $k_{\text{off}}$ and $K_a$ ($k_{\text{on}}$ = $k_{\text{off}}$ $\times$ $K_a$).

Structure calculations of calcium-loaded CBD$_{N1}$

Structure calculations of calcium-bound CBD$_{N1}$ were performed using X-PLOR v3.8. Torsion angle and distance restraints used were the same as those used earlier for the structure calculations of non-calcium-loaded CBD$_{N1}$ (chapter 3). A calcium atom was explicitly included in
Chapter 4-Calium Binding by CBD\textsubscript{N1}

the structure generation process with distance restraints of 2.45 ± 0.15 Å to its putative ligating atoms (T8 O, G30 O, D142 O, D142 O\textsubscript{5}). This restraint was determined from analysis of the distances between the ligating atoms and the calcium ion in the Bacillus macerans 1,3-1,4-β-glucanase structure (pdb code: 2AYH). Fifty structures were determined following the simulated annealing protocol in X-PLOR (Brünger, 1993) with the previously determined minimized average structure of CBD\textsubscript{N1} used as the starting model.

Results

CBD\textsubscript{N1} binds metal ions

The binding of CBD\textsubscript{N1} to a series of diamagnetic metal ions (Ca\textsuperscript{2+}, Zn\textsuperscript{2+}, Cd\textsuperscript{2+} and Mg\textsuperscript{2+}) was studied by one-dimensional NMR spectroscopy (Figure 4.1). Based on the numerous changes in the spectra of CBD\textsubscript{N1} with the addition of each ion, it is concluded that all these metals are bound by CBD\textsubscript{N1}. However, it is only with the addition of calcium that the complete appearance of the two most downfield resonances, W16 H\textsubscript{E1} and G7 H\textsubscript{N}, previously thought to be characteristic of cellotetraose binding, are observed. As the protein concentration in each of these samples is identical, and the same ten fold molar excess of metal was added to each, it is concluded that CBD\textsubscript{N1} binds Ca\textsuperscript{2+} with the highest affinity, followed by Zn\textsuperscript{2+}, Cd\textsuperscript{2+} and Mg\textsuperscript{2+}, in order of decreasing affinity. These last two bind with approximately equal affinity. Also, as each metal ion produces similar changes in the spectra of CBD\textsubscript{N1}, it is likely that all bind at the same location in the protein.

Calcium equilibrium binding association constant and binding stoichiometry

After this initial screen for metal binding, the interaction of calcium with CBD\textsubscript{N1} was further studied. The effects of calcium binding by CBD\textsubscript{N1} was monitored by one-dimensional \textsuperscript{1}H spectra (figure 4.2), and \textsuperscript{1}H-\textsuperscript{15}N HSQC spectra (figure 4.3) recorded for ten samples with
Figure 4.1. Screen for metal binding by CBDN1 using one-dimensional 1H NMR spectroscopy. Perturbation of the resonances W16 He1 and G7 HN are indicative of metal association. A 10 fold molar excess of metal ion was used for each spectrum.
Figure 4.2. Titration of CBDN1 with calcium monitored by 15N-decoupled one-dimensional NMR spectroscopy. The molar ratio of calcium-to-protein is indicated. Resonances W16 He1 and G7 HN exhibit intermediate two-site exchange.
Figure 4.3. Titration of CBDN1 with calcium monitored by 1H-15N HSQC spectra. Shown is a portion of ten overlaid spectra with the same molar ratios as indicated on figure 4.2. Arrows indicate the directions in which the amide peak shifts with added calcium. For the sake of clarity, not all the peaks are assigned. The peaks labelled * are due to impurities.
increasing amounts of added calcium. The peaks most indicative of calcium binding in one-dimensional $^1$H spectra are those of W16 H$_{el}$ and G7 H$_N$. Based on these shifts it is estimated that, without steps being taken to remove calcium, freshly prepared samples of CBD$_{N1}$ are approximately 20% calcium-loaded.

From analysis of figure 4.4A, it can be determined that the stoichiometry of calcium binding by CBD$_{N1}$ is 1:1. This arises from the fact that the titration curves have a plateau at a total calcium concentration equal to the protein concentration used in this experiment, 1.4 mM. From this data it is possible to rule out a model which describes the binding of two or more molecules of calcium, each binding with approximately equal affinity. This latter situation would have resulted in a series of curves which plateau at a total calcium concentration at least twice that of the protein, 2.8 mM.

An equilibrium association constant ($K_a$) for calcium binding was determined by a nonlinear least-squares fit of the $^1$H-$^{15}$N HSQC chemical shift data to a binding isotherm for a protein with a single ligation site. For each amide, two $K_a$ values were obtained, one using the data for the H$_N$ and another for the $^{15}$N nucleus. Figure 4.4A shows 28 fits for the data from 14 different amides that exhibit the greatest change in chemical shift with calcium binding, and who have measurable chemical shifts at each titration point. These amides show a coincident titration indicating they all are effected by the same binding event. In total 45 fits from 28 different amides (D10, G12, G15, W16, V17, T29, G30, A31, G39, V48, L49, G51, V52, A53, R63, T65, A66, S69, D71, T73, L95, S97, R100, V102, T105, L139, L141 and A145) were used to obtain an average $K_a$ value of $1.1 \times 10^5$ M$^{-1}$ with a standard deviation of $0.5 \times 10^5$ M$^{-1}$ at 35 °C, pH* 6.0.

This is a relatively high $K_a$ value to be determined at the CBD$_{N1}$ concentrations used for NMR studies. To test its validity, a series of curves were calculated with varying $K_a$ values. These simulated curves were plotted against the data obtained from the titration of R100 H$_N$, which has a regressed $K_a$ value of $1.0 \times 10^5$ M$^{-1}$ (Figure 4.4b). This graph emphasizes that it is hard to determine accurately a $K_a$ value under these conditions as large changes in $K_a$ produce relatively subtle variations in the shape of the calculated curves. This graph does show however, that the $K_a$
Figure 4.4. (A) Overlay of the titration data and best-fit curves of 28 15N and HN resonances from 14 different amides (D10, G15, W16, T29, G30, A31, A53, R63, T65, A66, S69, D71, R100, L139 and L141) in the titration of CBDN1 with calcium. Diamonds indicate the experimental data, the solid lines the fit to a model that describes the binding of calcium to protein in a 1:1 complex. Coincident data indicate the amides all monitor the same binding event. The arrow marks the plateau in the curve where the total sugar concentration equals the protein concentration (~1.4 mM) indicating 1:1 binding stoichiometry. (B) Simulations of the binding curves to the data for R100 HN. Squares indicate the experimental data. The lines are the simulated curves with Ka values of 50 000 M$^{-1}$, 70 000 M$^{-1}$, 100 000 M$^{-1}$, 200 000 M$^{-1}$, 300 000 M$^{-1}$, 500 000 M$^{-1}$ and 1 000 000 M$^{-1}$. 
value for calcium binding by CBD$_{N1}$ is in the range of $1 \times 10^5$ M$^{-1}$ to $3 \times 10^5$ M$^{-1}$. This value is supported by the determination of a $K_a$ value of $(8.3 \pm 0.2) \times 10^4$ M$^{-1}$ for the association of calcium with CBD$_{N1}$ under identical buffer conditions, as determined by isothermal titration calorimetry (Louise Creagh, personal communication).

Three amide resonances, I4, G7 and F9, exhibit a nonlinear titration during the last two or three titration points (not shown). This could be indicative of these residues sensing a second binding event. This second binding site would have a very much weaker association constant, in the order of less than 4000 M$^{-1}$. Since this is weak binding and only detected for three amides, its significance is unclear.

**Effect of calcium on cellooligosaccharide affinity**

Calcium-saturated CBD$_{N1}$ was titrated with cellopentaose to determine the effect that metal binding has on the protein's affinity for cellooligosaccharides. The equilibrium affinity constant for cellopentaose binding was calculated in a similar manner to that described earlier (chapter 2) and in the previous section. As was found previously for cellopentaose-binding by ~20% calcium-loaded CBD$_{N1}$ (chapter 2), the stoichiometry of cellopentaose binding by calcium-loaded CBD$_{N1}$ is 1:1. This is shown by the data having a plateau at a total cellopentaose concentration equal to that of the protein, 1.4 mM (figure 4.5A).

Fits using H$^N$ and $^{15}$N chemical shift data from 14 different amides (V17, Y19, G20, V34, G44, V45, L49, R75, N81, G83, T87, G130, G131 and L139) resulted in an average $K_a$ for cellopentaose binding of $3.1 \times 10^4$ M$^{-1}$ with a standard deviation of $0.8 \times 10^4$ M$^{-1}$. The value obtained previously for a sample of CBD$_{N1}$ approximately 20% calcium-bound is $(3.4 \pm 0.8) \times 10^4$ M$^{-1}$ (Chapter 2; Johnson et al., 1996b). This shows that the oligosaccharide-binding ability of CBD$_{N1}$ is not affected by calcium. Fits of the data for both the H$^N$ and $^{15}$N chemical shift data from 11 amide resonances are shown in figure 4.5A. Theoretical curves calculated with varying $K_a$ values plotted against the data obtained for the titration of Y19 H$^N$ are shown in figure 4.5B. In this case of cellopentaose binding, the $K_a$ value is lower than that for calcium binding and more
Figure 4.5. (A) Overlay of the titration data and best-fit curves of 22 15N and HN resonances from 11 different amides (V17, Y19, G20, V34, L49, R75, N81, T87, G130, G131 and L139) in the titration of calcium-loaded CBDN1 with cellopentaose. Diamonds indicate the experimental data, the solid lines the fit to model that describes the binding of cellopentaose to protein in a 1:1 complex. Coincident fits indicate that the amides all monitor the same binding event. (B) Simulations of the binding curves to the data for Y19 HN. Squares indicate the experimental data. The lines are the simulated curves with Ka values of 5 000 M⁻¹, 10 000 M⁻¹, 30 000 M⁻¹, 50 000 M⁻¹, 100 000 M⁻¹ and 1 000 000 M⁻¹.
confidence can be placed in its accuracy as changes in $K_a$ values produce a greater effect on the calculated curves than was the case for calcium binding.

**Kinetics of calcium binding**

NMR spectroscopy provides a powerful method to monitor the exchange of a nucleus between different environments due to a chemical reaction, ligand binding or conformational transitions. The exchange process can be monitored by NMR even if the sites are chemically equivalent as long as they are magnetically distinct and in chemical equilibrium.

To qualitatively understand the effect of exchange on an NMR spectrum, consider a nucleus that is experiencing conformational exchange between two magnetically distinct sites. The exchange rate between the two sites is $k$, and the two sites differ in frequency by $\Delta\nu$. The resonance frequency of the spin in each site can be observed for a time in the order of $1/k$ before the spin jumps to the other site and begins to precess with a different frequency. The finite observation time places a lower limit on the magnitude of $\Delta\nu$ required to distinguish the sites. When $k \ll \Delta\nu$, distinct signals from are observed for the nucleus at each if the two frequencies. This is the slow exchange limit. At the fast exchange limit, $k \gg \Delta\nu$, a single resonance appears at the population-weighted average chemical shift of the nuclei at the two sites. This was the case for the binding of cellooligosaccharides by CBD$_{N1}$ presented in this chapter and chapter 2.

A third case arises when the exchange rate is of the order of the chemical shift separation between the two sites ($k \sim \Delta\nu$). This situation, where the lines become very broadened and begin to coalesce, is known as intermediate exchange. In this section, detailing the binding of calcium to CBD$_{N1}$, some resonances whose chemical shifts are perturbed by calcium binding are in the fast exchange limit, while others experience intermediate exchange. The NMR chemical-shift timescale varies by residue, as it is defined by the difference in chemical shift between the resonance frequencies of the two environments of a given nucleus. The equations that describe the lineshape, or bandshape, of resonances experiencing two-site exchange are complex and are summarised in Sandström (1982).
The kinetics of calcium association/dissociation were obtained from analysis of lineshapes from residues that are in intermediate exchange on the NMR timescale. As expected from the classical model of two-site exchange, these residues appear as sharp peaks in the calcium-free form then broaden and decrease in intensity with added calcium. In the fully bound form these peaks are once again sharp. An example of this behaviour is shown for W16 Hε1 in figures 4.2 and 4.3.

Seven different lineshapes from six different residues were analysed by the formalism of Sandström (1982). These were the 1H lineshapes of W16 Hε1 obtained from one-dimensional spectra, and the 1H lineshapes of G30, A31, L139, L141, and A145 and the 15N lineshapes of G30 obtained from the 1H-15N HSQC spectra collected during the titration. Only residues whose amide broadened but did not disappear entirely were selected for analysis. Values for $k_{\text{off}}$ were obtained by simulation of the spectrum at as many of the titration points as possible to yield an average $k_{\text{off}}$ value for each peak. From these individual points a global average of $4.5 \times 10^2$ s$^{-1}$ with a standard deviation of $0.6 \times 10^2$ s$^{-1}$ was determined. Figure 4.6 shows an example of the data from the calcium titration used to determine the $k_{\text{off}}$ rates compared with the calculated lineshapes. A value for $k_{\text{on}}$ of $(5 \pm 2) \times 10^7$ M$^{-1}$ s$^{-1}$ was determined from the experimentally determined $K_a$ and $k_{\text{off}}$ values.

*Identification of calcium-binding site*

When discussing patterns of chemical shift perturbations with ligand binding, it is important to bear in mind that changes can arise either from direct interaction of the nucleus with the ligand or indirectly from conformational changes resulting from formation of the protein-ligand complex. Chemical shifts are also a very sensitive indicator of structure, with subtle conformational changes often resulting in large changes in shift. Keeping this in mind, it is possible to gain insight into the location of the calcium-binding site in CBD$_{N1}$ from the patterns of chemical shift perturbations upon addition of calcium. Figure 4.7 summarises the change in H$^N$ and 15N chemical shift of each amide in CBD$_{N1}$ due to the binding of calcium.
Figure 4.6. Comparison of the experimental (A) and simulated lineshapes (B) using the data for W16 He1. The experimental data is an expanded region of the 1-D 1H NMR spectra shown in figure 2. The simulated lineshapes were produced using a koff value of 500 s⁻¹, T₂ values of 0.02 s (free) and 0.015 s (bound), and frequency values of 5539 Hz (free) and 5256 Hz (bound). The numbers indicate the fraction of protein in the calcium-bound form at that point in the titration, as calculated using a Ka of 110 000 M⁻¹.
Figure 4.7. Patterns of chemical shift perturbation due to calcium binding by CBD_{N1}. The absolute value in the difference between the HN and 15N chemical shifts of the backbone amides in the free and calcium-bound protein are plotted as negative and positive numbers, respectively. The region that shows the greatest change in chemical shift involves the putative binding site residues T8, G30 and D142.
Figure 4.8. Cα worm diagram of CBD_{N1} with residues that experience the most significant change in chemical shift with calcium binding (see text), or whose amide resonances are not observed in the absence of added calcium and appear upon calcium addition (T8, V144, L146) coloured red. The top panel shows β-sheet B, which lies opposite to the binding face. The lower panel is a 90° rotation from the view shown in the top panel, the binding cleft is seen at the top of the structure in this view. Selected residues are labelled. The amino and carboxyl termini are denoted by the labels N and C, respectively. This figure was made using the programme GRASP (Nicholls et al., 1991).
Chapter 4 - Calcium Binding by CBD

The average change in the absolute value of $^{15}$N chemical shift due to calcium binding is 0.4 ppm with a standard deviation of 0.7 ppm. For the $^{1}$H$^{N}$ dimension the average absolute value in chemical shift change is 0.05 ppm with a standard deviation of 0.08 ppm. Shown in figure 4.8 is a backbone worm diagram of CBD$_{N_{1}}$ with residues that experience a change in either $^{15}$N or $^{1}$H$^{N}$ chemical shift one standard deviation greater than average in red. In addition, those residues not observed in the absence of added calcium (T8, V144, L146) are similarly coloured. Although this analysis is statistically questionable, as the distribution of the absolute values of the change in shifts is skewed and not Gaussian, it does provide a criterion for defining what level of perturbation is “significant”. An alternative to this gauge as to what changes are significant would be to indicate the location of an arbitrary number (~20%) of residues that titrate the most.

Figure 4.9 shows the four overlapped $^{1}$H-$^{13}$C HSQC spectra from the titration of the 10% $^{13}$C labelled CBD$_{N_{1}}$ with calcium. It is clear that, when compared to the $^{1}$H-$^{15}$N HSQC spectra (figure 4.3), the methyl resonances are less sensitive to calcium addition. Far fewer resonances change their chemical shift, and the ones that do titrate experience a much smaller perturbation. The resolved methyl resonances that shift upon calcium binding are, in order of greatest shift, T8$^{\gamma_{2}}$, I4$^{\gamma_{2}}$, L32$^{\delta_{1}}$, L49$^{\delta_{2}}$, L49$^{\delta_{1}}$ and A68$^{\beta}$. With the exception of L49, all lie in regions of the protein which contain backbone amides that also show significant chemical shift change. The methyls of L49 are in the hydrophobic core of the protein lying close to the side-chain of W16. As is shown in figures 4.2 and 4.3, W16 H$^{e_{1}}$ also has a large change in chemical shift with added calcium. Due to spectral overlap, it is not possible to obtain data for the methyl groups of residues T65$^{\gamma_{2}}$, T67$^{\gamma_{2}}$, A31$^{\beta}$ or T29$^{\gamma_{2}}$, which lie close to additional residues whose amide resonances titrate markedly.

From figures 4.7 and 4.8 it is clear that four regions of CBD$_{N_{1}}$ are affected the most by calcium binding. These include the entire N-terminus up to residue 17, residues 30-33 ($^{\beta}$-strand B2), residues 65-68 ($^{\beta}$-strand B3), residues 139-146 ($^{\beta}$-strand B5) and residue 100 ($^{\beta}$-strand B34). Aside from residues 16 and 17, no amides on the binding face of the protein show a significant change in chemical shift with calcium binding. Instead all the residues with the largest
Figure 4.9. Titration of CBDN1 with calcium monitored by 1H-13C HSQC spectra. Shown is the region of the spectrum where the observed peaks are due to methyl groups. Four spectra with increasing amounts of added calcium are overlaid. Only residues whose methyl resonances experience a change in chemical shift with calcium binding are assigned.
change in chemical shift are on the opposite face. Each of β-strand B2 to B5 of this face has at least one residue that shows larger than average changes in chemical shift with calcium binding, with greatest changes concentrated on β-strands B2, B4 and B5.

Discussion

Location of the metal binding site and structure of calcium-loaded CBD$_{N1}$

Chemical shift perturbations provide only qualitative data for determining the location of the calcium-binding site on CBD$_{N1}$, but do not unambiguously identify the coordinating atoms. Therefore I examined the CBD$_{N1}$ structure for possible regions which contain residues that commonly interact with calcium. In general, these are areas high in negative charge, and thus rich in aspartic and glutamic acid residues.

There are two regions that qualify as potential binding sites. The first is the loop region involving residues 6 to 11, containing the potential ligating residues E6, D10 and D11. In the minimized average structure of CBD$_{N1}$ (Chapter 3; Johnson et al., 1996b) the side-chain of E6 points away from D10 and D11. However, this is also the most disordered region of the structure and the exact location, or locations, of the side-chains may not be accurately represented with this structure. Thus the involvement of E6 cannot be ruled out based on its apparent orientation. The other possible location for calcium binding is the area around D142, D143, residues 29-32, and the N-terminal residues 7-9.

In contrast to CBD$_{N1}$, CBD$_{N2}$ does not bind calcium (E. Brun, personal communication). Thus, it is expected that the calcium-binding residues are not conserved in the sequences of these two proteins. If the residues of one of the possible binding site were conserved and another not, this would provide evidence for its location. Unfortunately, residues in both possible calcium-binding sites exhibit low sequence similarity. E6, D10 and D11 in CBD$_{N1}$ correspond to a histidine, a serine and a glutamate in CBD$_{N2}$, while D142 and D143 in CBD$_{N1}$ align with a serine
and a glutamine in CBD_{N2}. This does not differentiate the two sites, but it is consistent with the lack of binding in CBD_{N2}.

Strong evidence for the location of the calcium-binding site comes from a comparison of CBD_{N1} with the 1,3-1,4-β-glucanase structures. As discussed earlier (chapter 3) these proteins are structurally similar, all having a β-jelly roll fold. Both the hybrid Bacillus 1,3-1,4-β-glucanase and the Bacillus macerans 1,3-1,4-β-glucanase bind calcium octahedrally, using P9 O, G45 O, D207 O and D207 O⁸ (Keitel et al., 1993; Hahn et al., 1995a) (figure 4.10). Additionally, two water molecules in the hybrid 1,3-1,4-β-glucanase are coordinated to the calcium ion in that crystal structure. Presumably two waters molecules also coordinate the calcium in the Bacillus macerans 1,3-1,4-β-glucanase structure to complete the octahedral geometry. The structure of the Bacillus licheniformis 1,3-1,4-β-glucanase is slightly different, binding calcium in a pentahedral-bipyramidal manner with the atoms P9 O, D207 O, D207 O⁸ and two water molecules forming the pentahedral plane. Another water and the carbonyl of G45 form the apical positions (Hahn et al., 1995c).

CBD_{N1} and these 1,3-1,4-β-glucanase structures are remarkably similar in this calcium-binding region. All posses a bulge in the β-sheet at residue D142 for CBD_{N1} and D207 for the 1,3-1,4-β-glucanase structures. In light of these two facts, it is therefore very likely that the calcium-binding site of CBD_{N1} is at the same position, involving T8 O, G30 O, D142 O and a side-chain oxygen of D142 as the metal coordinating ligands. As will be discussed below, this location for the calcium-binding site is also supported by similar calcium-binding constants, and a common biological role of stabilising the folded structure for both CBD_{N1} and the hybrid 1,3-1,4-β-glucanase.

The structure of CBD_{N1} bound to calcium was determined using these atoms to coordinate the calcium ion. As there is no unambiguous experimental evidence these are the coordinating atoms, this structure should be considered a model. For this reason the coordinates were not deposited in the Brookhaven Protein Data Bank. One way to unambiguously experimentally determine the location of the metal binding site would be to substitute the calcium ion with ^{113}\text{Cd}. 
Figure 4.10. Two views of the structure of *Bacillus macerans* 1,3–1,4–β–glucanase showing its calcium–binding site. Drawn in red are the ligands that coordinate calcium in this structure (P9 O, G45 O, D207 O and D207 O\(^5\)). Based on this structure, the corresponding atoms T8 O, G30 O, D142 O and D142 O\(^8\) in CBD\(_N1\) are assigned as the putative calcium coordinating ligands. The calcium atom is coloured in green.
A $^{13}$C-$^{13}$Cd HSQC could then be run to assign the resonances and identity of the backbone carbonyls or side chain carboxyls that ligates the metal.

In almost all calcium binding proteins coordination is either octahedral or pentagonal bipyramidal (McPhalen et al., 1991). It is therefore likely that two or three water molecules also ligate calcium in the CBD$_{N1}$ structure. It is slightly atypical that only one carboxylate group and three backbone carbonyls ligate calcium. The presence of two or more Asp or Glu sidechains is usually the case (McPhalen et al., 1991). The involvement of three backbone carbonyl atoms ligating calcium probably accounts for the octahedral coordination. The steric constraint of having the backbone of the protein so close to the calcium probably excludes a higher coordination number.

As expected for structures calculated using the same restraints, the only region of CBD$_{N1}$ that exhibits a difference between the structures calculated with calcium and the structures calculated previously (chapter 3) is in the vicinity of the calcium ion, figure 4.11. Specifically, in the calcium bound structure the backbone carbonyl of T8 twists around 180° so as to ligate the calcium ion. The backbone carbonyl, and the side-chain of D142 also twists to point toward the calcium ion (figure 4.11).

**Binding affinity and kinetics**

In this study it was found that CBD$_{N1}$ binds calcium fairly tightly, with a $K_a$ of $(1.1 \pm 0.5) \times 10^5$ M$^{-1}$ at pH* 6.0, 35 °C. This value matches, within the error range, that determined by isothermal titration calorimetry of $(8.3 \pm 0.2) \times 10^4$ M$^{-1}$ (L. Creagh & C. Haynes, pers. comm.). Also, the $K_a$ of CBD$_{N1}$ for calcium is similar to that determined for the hybrid *Bacillus* 1,3-1,4-β-glucanase H(A16-M) which was reported as $(1.0 \pm 0.4) \times 10^5$ M$^{-1}$, also at pH 6.0. (Keitel, et al., 1994). This similarity in $K_a$ values supports the proposal that the calcium-binding site of CBD$_{N1}$ lies at the same location, and involves the same ligands, as the hybrid *Bacillus* 1,3-1,4-β-glucanase H(A16-M).
Figure 4.11. Model of the structure of the calcium-binding site in CBD$_{N1}$. Shown in dark grey are the putative calcium-ligating ligands (T8 O, G30 O, D142 O and the side-chain of D142) of the lowest energy structure from the ensemble calculated in the presence of calcium. Shown in light grey are the same ligands in the mean minimized structure calculated without calcium (chapter 3). In black is the calcium ion, drawn using its ionic radius. In the calcium-bound model the backbone carbonyl of T8 twists 180 degrees to ligate calcium. The backbone carbonyl of D142 twists 90 degrees, and its side chain rotates to point towards the calcium ion. The position of the carbonyl of G30 is relatively unchanged. In addition to these ligands, one or more water molecules probably also coordinate calcium. This figure was made using the programme Molscript (Kraulis, 1991).
When produced by *Cellulomonas fimi*, the CenC cellulase, from which CBD$_{N1}$ is derived, is secreted into the environment. It has been estimated that the concentration of calcium extracellularly is 2 mM (McPhalen *et al.*, 1991) and 10 mM in sea water (Glusker, 1991). The calcium concentration present in the natural environment of *C. fimi* is not known, but it is likely to also be in the millimolar range. At these concentrations, CBD$_{N1}$ would be fully calcium saturated. The calcium-binding association constant is also fairly high when compared to that of other extracellular proteins, though much lower than some cellular calcium binding proteins (Strynadka & James, 1989; McPhalen *et al.*, 1991).

The on-rate for calcium binding of $(5 \pm 2) \times 10^7$ M$^{-1}$ s$^{-1}$ is within two orders of magnitude of the diffusion-controlled limit of $\sim 10^9$ M$^{-1}$ s$^{-1}$ (Fersht, 1985). This implies that the calcium-binding site of CBD$_{N1}$ is likely mostly preformed with no major structural rearrangement necessary to properly orient the atoms that ligate calcium. The presence of this preformed site is consistent with the relatively high binding affinity of CBD$_{N1}$ for calcium. The dynamics of this region were studied by $^{15}$N relaxation methods, these results are presented in chapter 5.

**A structural role for calcium binding by CBD$_{N1}$**

The apparent biological role of calcium binding by CBD$_{N1}$ is to stabilise its folded structure. As determined by differential scanning calorimetry (DSC), calcium binding increases the temperature of denaturation of CBD$_{N1}$ by 8 °C, from 49 °C to 57 °C (Louise Creagh & Charles Haynes, personal communication). The putative calcium-binding site of CBD$_{N1}$ ties together three segments of the protein chain that are remote in primary sequence. Having the calcium ion coordinated by ligands from various regions of the protein is common among proteins that are stabilised by calcium binding (Strynadka & James, 1989). The stabilisation of the folded structure by calcium binding also supports the choice of potential calcium-binding sites in CBD$_{N1}$. The alternate site considered contains ligands solely in the N-terminal region. Binding by these residues does not seem as likely to stabilise the protein as ligands from different regions of the
protein. It is also of note that CBD$_{N2}$, which does not bind calcium, has a higher denaturation temperature than both the apo and calcium-loaded forms of CBD$_{N1}$.

This stabilising role of calcium binding is supported by the finding that there is no significant difference between the $K_a$ values for cellopentaose binding by CBD$_{N1}$ 20% calcium-loaded or fully saturated. This indicates that the binding of oligosaccharides is not dependent upon whether or not calcium is bound. Although the effect calcium has on the binding of other oligosaccharides by CBD$_{N1}$, and the effects of other metals were not checked, it is safe to assume that the results would be similar. This is not too surprising in light of the fact that the binding site for calcium is on the opposite side of the protein from the sugar binding site. Large structural changes with calcium binding would have to occur in order to affect the orientation of the residues on the binding face. For the class of lectins whose binding to their sugar ligand is calcium dependent, many of the structures show that the calcium ion is bound close to the sugar binding site. Often in these proteins, calcium mediates hydrogen bonding between the protein and sugar (Rini, 1995)

The biological role of calcium stabilising the tertiary structure of CBD$_{N1}$ is shared by the Bacillus 1,3-1,4-β-glucanase enzymes. Seven native and hybrid 1,3-1,4-β-glucanase studied by Welfle et al. (1994; 1995; 1996) all had higher denaturation temperatures and higher Gibbs free energies in the presence of calcium than EDTA. It was also found that the hybrid Bacillus 1,3-1,4-β-glucanase H(A-16-M) was more stable to guanidinium chloride denaturation calcium-loaded than calcium free (Keitel et al., 1994).

Very little is known about the role calcium binding has in other CBDs. As mentioned in the introduction to this chapter, a calcium ion has been identified in the crystal structures of both CBD$_{Cip}$ (Tormo et al., 1996) and CBD$_{E4}$ (Sakon et al., 1997). As was found for CBD$_{N1}$, in both these structures the calcium ion is located far from the putative binding face. It is therefore likely that calcium has the role of stabilising the structure of these proteins, and does not influence cellulose binding.
Chapter 5

Dynamic Analysis of Ligand Binding

Abstract

Backbone amide $^{15}$N and side-chain methyl-containing $^2$H relaxation techniques were used to investigate the effects of calcium binding and oligosaccharide binding on the dynamics of CBD$_{N1}$. $^{15}$N spin-lattice relaxation times ($T_1$), spin-spin relaxation times ($T_2$) and heteronuclear NOEs were determined for the uniformly $^{15}$N labelled protein. $^2$H $T_1$ and $T_{1p}$ values were determined for a sample of fractionally deuterated $^{15}$N $^{13}$C CBD$_{N1}$. The $^{15}$N data were analysed using the model-free formalism to derive the model-free parameters ($S^2$, $\tau_e$, $R_{ex}$, $S_I^2$ and $\tau_s$) for individual backbone N-H bond vectors and values for the overall rotational correlation time ($\tau_m$). Calcium binding dramatically reduces the values of $R_{ex}$ terms found for residues in the N-terminal region of CBD$_{N1}$ that forms one part of the calcium-binding site. This implies that motion on the millisecond time scale is reduced upon calcium binding. Upon cellopentaose binding, $R_{ex}$ terms for a set of residues that lie at the end of the $\beta$-strands that form the binding face decrease in value. This also indicates that motion on the millisecond time scale of these is reduced upon cellopentaose binding. The deuteron relaxation techniques probe the nanosecond-picosecond time scale dynamics of methyl containing side-chain residues. It is established that methyl groups present on the binding face have a high degree of mobility in both the free and cellopentaose-bound form. It is hypothesized that this mobility is needed for CBD$_{N1}$ to recognize and bind different orientations of the same ligand, as well as different oligosaccharides.
Introduction

Background

The determination of a three-dimensional structure of a protein, by either x-ray crystallography or NMR methods, can give the impression that its folded structure is static. In reality proteins are dynamic entities. Motion occurs on a wide variety of time scales, from rapid fluctuations of bonds and torsion angles, to conformational transitions involving the collective motion of a large number of atoms. The fast motions are characterized by frequencies on the order of gigahertz while the slower motions occur in the range of milliseconds to microseconds.

The dynamics, or internal motions that occur in proteins are dramatically shown in the case of T4 lysozyme. A benzene molecule rapidly enters an engineered solvent-inaccessible hydrophobic cavity in variants of this protein that are 5-6 Å from the surface (Feher et al., 1996). Benzene stabilises the protein by binding in the hydrophobic cavity created by the mutations Leu99Ala and Met102Ala (Eriksson et al., 1992a; Eriksson et al., 1992b). This implies that relatively large fluctuations in the conformation of the protein must occur for the benzene molecule to enter the interior of this protein.

The dynamics of proteins can be an important part of their functional behaviour. The dynamic behavior of proteins has been related to protein folding, ligand binding, enzymatic ability and allosteric regulation (Creighton, 1993). In the case of a binding protein, like CBD₉₁, a number of different possibilities can be imagined where the mobility of a region, or of specific residues, in the protein plays an important role. For instance, residues in the binding site of a protein can be disordered in the unbound state. Here the residues involved in binding are exploring many different conformations. If this region becomes more rigid upon binding, an entropic cost is entailed that must be compensated by other favorable interactions.

In this chapter the dynamic behaviour of CBD₉₁ is studied using the well developed NMR method of amide ⁱ⁵N relaxation and the recently developed technique of methyl containing
deuteron relaxation. In particular, the changes that occur in the dynamic properties of CBD$_N$ upon calcium and oligosaccharide binding are studied.

15$^N$ relaxation and dynamics

When a population of spins is disturbed by an RF pulse, the system will return to the equilibrium state. This process is the phenomenon of relaxation. Two time constants, $T_1$ and $T_2$, are used to describe the relaxation process. $R_1 (1/T_1)$, spin-lattice or longitudinal relaxation, is the rate by which the spins return to the equilibrium population. $T_1$ relaxation is associated with population differences and involves an exchange of energy with the surroundings. $R_2 (1/T_2)$, or spin-spin relaxation, describes the rate by which the spins lose their phase coherence in the transverse plane.

NMR relaxation is caused by fluctuating local magnetic fields experienced by the nuclei that arise from molecular motion. In liquids the most important relaxation mechanism is the dipole-dipole interaction between neighboring spins. Chemical shift anisotropy (CSA) can also act as an efficient relaxation mechanism. CSA arises from the fact that different orientations of a nucleus with respect to the external magnetic field have different chemical shifts. Incomplete averaging of these shifts by molecular tumbling results in relaxation. The power available to cause relaxation resulting from motion at a certain frequency ($\omega$), is given by the spectral density, $J(\omega)$.

The $T_1$ and $T_2$ relaxation times and the steady-state heteronuclear NOE enhancement of an amide 15$^N$ nucleus are dominated by the dipolar interaction of the 15$^N$ nucleus with its directly attached proton. A significant contribution due to chemical shift anisotropy is also present. The rates of these relaxation processes are given by Abragam (1961) (after Farrow et al., 1994):

$$R_1 = \frac{1}{T_1} = d^2[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2J(\omega_N)$$ 

(5.1)
Chapter 5-Dynamic Analysis of Ligand Binding

\[
R_2 = \frac{1}{T_2} = \frac{d^2}{2} [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + \frac{c^2}{6} [3J(\omega_N) + 4J(0)] \tag{5.2}
\]

\[
\text{NOE} = 1 + \left(\frac{\gamma_H}{\gamma_N}\right) d^2 [6J(\omega_H + \omega_N) + 3J(\omega_N) - J(\omega_H - \omega_N)] T_1 \tag{5.3}
\]

The constants \(d^2\) and \(c^2\) represent the dipolar and chemical shift anisotropy contributions, defined as:

\[
d^2 = -\frac{1}{4\pi^2} \left(\frac{1}{r_{NH}^3}\right)^2
\]

and

\[
c^2 = \frac{2}{15} \gamma_N^2 H_0^2 (\Delta \sigma^2)
\]

In these equations \(h\) is Plank’s constant, \(\gamma_H\) and \(\gamma_N\) are the gyromagnetic ratios of the \(^1\text{H}\) and \(^{15}\text{N}\) nuclei, respectively, \(\omega_H\) and \(\omega_N\) are the \(^1\text{H}\) and \(^{15}\text{N}\) Larmor frequencies, \(r_{NH}\) is the internuclear \(^1\text{H}^{15}\text{N}\) distance (1.02 Å, a value that is essentially invariant), \(H_0\) is the magnetic field strength and \(\Delta \sigma\) is the difference between the parallel and perpendicular components of the axially symmetrical \(^{15}\text{N}\) chemical shift tensor. The assumption of an axially symmetric chemical shift tensor has been shown to be valid for peptide bonds, with \(\Delta \sigma\) equal to -160 ppm (Hiyama et al., 1988). Figure 5.1 shows the dependence of \(T_1\), \(T_2\) and heteronuclear NOE for a \(^{15}\text{N}^{1}\text{H}\) pair on the motion of a rigid molecule. The molecular motion is described by the rotational correlation time, \(\tau_m\). It is defined as the average time for the molecule to rotate by one radian. This figure clearly shows the motional dependency of these experimentally measurable parameters. The aim of this chapter is to relate \(T_1\), \(T_2\) and heteronuclear NOE values to the dynamics, or motions, of \(\text{CBDN}_1\).

In the absence of any further assumptions, the three experimentally measured parameters (\(T_1\), \(T_2\) and \(^{15}\text{N}^{1}\text{H}\) NOE) do not provide enough information to enable direct determination of the spectral density function at the five frequencies (0, \(\omega_N\), \(\omega_H\), \((\omega_H - \omega_N)\) and \((\omega_H + \omega_N)\)) of equations 5.1 to 5.3. To relate relaxation data to the motional properties of the proton-heteronucleus bonds, specific models have been employed to provide an analytical form of the spectral density function (Peng & Wagner, 1994). The most widely adopted, and simplest,
method of analysing relaxation data is the model-free formalism of Lipari and Szabo (1982a,b). This method employs a minimum number of parameters to describe the overall tumbling of the molecule, and the internal motions of the $^1\text{H}-^{15}\text{N}$ amide bond vector using the following expression:

$$J(\omega) = \frac{2}{5} \left[ \frac{S^2 \tau_m}{(1+\omega^2\tau_m^2)} + \frac{(1-S^2)\tau}{(1+\omega^2\tau^2)} \right]$$  \hspace{1cm} (5.6)$$

$S^2$ is the order parameter squared which describes the degree of spatial restriction of the internal motion of the $^1\text{H}-^{15}\text{N}$ bond vector. $S^2$ can vary from 0 for completely isotropic motion, to 1 for...
completely restricted motion. An effective correlation time for the more rapid local motion of the $^{15}$N-$^1$H vector, as compared to $\tau_m$, is described by $\tau_e$ where:

$$\frac{1}{\tau} = \frac{1}{\tau_m} + \frac{1}{\tau_e} \quad (5.7)$$

Equation 5.6 assumes that the overall tumbling of the molecule is isotropic. In the case where $\tau_e \ll \tau_m$ equation 5.6 reduces to

$$J(\omega) = \frac{2}{5} \left[ \frac{S^2 \tau_m}{1 + \omega^2 \tau_m^2} \right] \quad (5.8)$$

The model-free analysis has been extended by Clore et al. (1990a) to approximate internal motions on two time scales differing by at least one order of magnitude; it is described by the following equation:

$$J(\omega) = \frac{2}{5} \left[ \frac{S^2 \tau_m}{1 + (\omega \tau_m)^2} + \frac{(S_f^2 - S^2)\tau}{1 + (\omega \tau)^2} \right] \quad (5.9)$$

Here $\tau_s$ is the effective correlation time for slow internal motions, and is included in the relationship $\frac{1}{\tau} = \frac{1}{\tau_s} + \frac{1}{\tau_m}$. $S^2$ is expressed as the product of two order parameters ($S^2 = S_f^2 S_s^2$) describing fast and slow internal motions, $S_f^2$ and $S_s^2$, respectively.

An additional term, $R_{ex}$, can be incorporated when modeling the observed $T_2$ values to take into account the contributions from processes other than those from dipole-dipole (DD) and chemical shift anisotropy (CSA). In many cases, these contributions arise from conformational exchange averaging occurring on the microsecond to millisecond time scale.

$$\frac{1}{T_2} = \frac{1}{T_{2\text{(DD)}}} + \frac{1}{T_{2\text{(CSA)}}} + R_{ex} \quad (5.10)$$
**Methyl group deuteron relaxation and dynamics**

Recently, Kay and co-workers have developed experiments to study nanosecond-to-picosecond side-chain dynamics based on the fractional incorporation of deuterium into $^{15}$N and $^{13}$C-labelled proteins (Muhandiram et al., 1995). The key to the use of deuterium as a probe of molecular dynamics lies in the fact that the energy of a deuteron depends critically on its local environment. This quadrupolar interaction is well understood and the interpretation of deuterium relaxation is simpler than is the case for many other nuclei.

These deuterium-based relaxation experiments specifically select for $^{13}$CH$_2$D methyl groups present in the protein. The data are recorded as a series of two-dimensional constant-time $^1$H-$^{13}$C HSQC experiments in which the methyl peak is dependent upon the deuterium $T_1(D)$ and $T_{1p}(D)$ times. These values are determined by measuring relaxation of terms of the form $I_z C_z D_z$ and $I_z C_z D_y$, as well as the relaxation rate of $I_z C_z$, where $I_z$, $C_z$ and $D_z$ denote the z magnetization of the methyl proton, carbon, and deuteron, respectively. $D_y$ is the y component of deuterium magnetization. The theoretical validity of these experiments, and the justification for measuring deuterium relaxation in CH$_2$D versus CHD$_2$ groups, are discussed in Muhandiram et al. (1995) and Yang & Kay (1996).

Since the decay of the triple spin terms $I_z C_z D_z$ and $I_z C_z D_y$ are dominated by deuterium relaxation, the following relations are excellent approximations for $T_1(D)$ and $T_{1p}(D)$:

\[ \frac{1}{T_1(D)} = \frac{1}{T_1(I_z C_z D_z)} - \frac{1}{T_1(I_z C_z)} \quad (5.11) \]

\[ \frac{1}{T_{1p}(D)} = \frac{1}{T_{1p}(I_z C_z D_y)} - \frac{1}{T_1(I_z C_z)} \quad (5.12) \]

The measured $T_1(D)$ and $T_{1p}(D)$ times are related to motional properties at specific sites in the protein through their dependence on the power spectral density function, $J(\omega)$, according to the following equations (Abragam, 1961; after Kay et al., 1996):
\[
\frac{1}{T_1} = \frac{3}{16} \left( \frac{e^2 q Q}{h} \right)^2 \left[ J(\omega_D) + 4J(2\omega_D) \right]
\] (5.13)

\[
\frac{1}{T_{1p}} = \frac{1}{32} \left( \frac{e^2 q Q}{h} \right)^2 \left[ 9J(0) + 15J(\omega_D) + 6J(2\omega_D) \right]
\] (5.14)

\(\frac{e^2 q Q}{h}\) is the quadrupole coupling constant, which is 165 kHz for methyl group deuterons.

Similarly as was described in the previous section for \(^{15}\)N relaxation, \(J(\omega)\) can be expressed as (Lipari & Szabo, 1982a,b):

\[
J(\omega) = \frac{2}{5} \left[ \frac{S_i^2 \tau_m}{1 + (\omega \tau_m)^2} + \frac{(1 - S_i^2) \tau_i}{1 + (\omega \tau_i)^2} \right]
\] (5.15)

Here \(S_i\) is an order parameter for the methyl group \(i\). \(S_i\) describes the spatial restriction of motion of the \(^{13}\)C-\(^2\)H bond vector in the methyl group on the nanosecond-picosecond time scale, \(\tau_m\) is the overall correlation time, and \(\frac{1}{\tau_i} = \frac{1}{\tau_m} + \frac{1}{\tau_{e,i}}\), with \(\tau_{e,i}\) the effective correlation time describing the internal motions for the \(^{13}\)C-\(^2\)H bond vector \(i\). The order parameter of the bond vector about which the methyl group rotates is denoted \(S_{axis}\) and is related to \(S_i\) by \(S_i^2 = 0.111 S_{axis}^2\), assuming tetrahedral geometry for the methyl group (Nicholson et al., 1992). A value of 1 for \(S_{axis}^2\) indicates complete restriction of motion of the methyl averaging axis, while an \(S_{axis}^2\) of 0 indicates complete freedom of motion.

This technique of measuring methyl deuteron relaxation rates has been previously applied to study the interaction of a Src-homology-2 (SH2) domain with a phosphorylated tyrosine-containing peptide (Kay et al., 1996). In this study, the authors showed that motion of methyl groups in the phosphotyrosine binding pocket were restricted upon binding. Methyl groups in other regions of the binding site remained disordered in both the bound and free state. Recently, the methyl deuteron relaxation rates as a function of temperature were measured for the N-terminal SH3 domain from drk (Yang et al., 1997).
Experimental Methods

Sample preparation for $^{15}\text{N}$ relaxation experiments

$^{15}\text{N}$ labelled CBD$_{N1}$ samples for the $^{15}\text{N}$ relaxation experiments were produced as outlined in Chapter 2. In many cases, the same $^{15}\text{N}$ CBD$_{N1}$ sample was recycled by removing added cellooligosaccharides. This was accomplished by unfolding the protein with 6 M urea directly in a microsep concentrator (Filtron) with a 1K cut-off membrane. With the protein completely unfolded, all the sugar is free in solution and passes through the membrane. Extensive buffer exchange removes the urea, allowing protein refolding. The validity of this approach was confirmed by NMR spectroscopy. Spectra of the original and refolded samples are identical with no evidence of any residual soluble unfolded protein or sugar remaining after buffer exchange.

Preparation of fractionally deuterated CBD$_{N1}$

A sample of $^{15}\text{N}$, $^{13}\text{C}$ CBD$_{N1}$ fractionally deuterated to a level of approximately 40% (subsequently referred to as $^{15}\text{N}/^{13}\text{C}/^{2}\text{H}(40\%)$ CBD$_{N1}$), was prepared by expressing the plasmid pTugNln in Escherichia coli JM101 cells grown at 30 °C using a 1L fermentor. The M9 minimal media contained 2g/L $[^{13}\text{C}]_6$ glucose, 1g/L $^{15}\text{NH}_4\text{SO}_4$ and 1g/L $^{13}\text{C}$ (99%)/$^{15}\text{N}$ (99%)/$^{2}\text{H}$ (50%) Celtone algal extract (Martek) in 400 mL D$_2$O/600 mL H$_2$O.

The inoculant for the 1 L culture was obtained by culturing a single colony from a plate in 5 mL TB media (100% H$_2$O) for 6 hours to an OD$_{600}$ of 1.1. One mL of this culture was spun down, resuspended in 10 mL of TYP media in 40% D$_2$O/60% H$_2$O, and grown for 2.5 hours to an OD$_{600}$ of 1.0. The entire culture was spun down and resuspended in 100 mL of M9 media containing 1g/L of unlabelled Celtone in 40% D$_2$O/60% H$_2$O. After growth for 4 hours to an OD$_{600}$ of 0.9, all the cells were spun down and used to inoculate 1 L of the labelled media. This 1 L bacterial culture was grown for 5 hours to an OD$_{600}$ of 1.5, at which point protein expression was induced with the addition of IPTG to a final concentration of 0.5 mM. The culture was incubated for 31 hours, the final OD$_{600}$ being 5.6. All the culturing was performed at 30 °C.
Purification of $^{15}\text{N}/^{13}\text{C}/^{2}\text{H}(40\%)$ CB$D_{\text{NI}}$ was as described previously (Johnson et al., 1996a; chapter 2) with the exception that the osmotic shock fractions were not subjected to Avicel binding. Instead they were combined with the $\text{H}_2\text{O}$ elution fractions from Avicel binding and loaded onto the ionic exchange column. After purification, 19 mg of $^{15}\text{N}/^{13}\text{C}/^{2}\text{H}(40\%)$ CB$D_{\text{NI}}$ was obtained from 1 L of culture.

$^{15}\text{N}$ relaxation experiments

All spectra measuring $^{15}\text{N}$ T$_1$, $^{15}\text{N}$ T$_2$ and $^1\text{H}$$-^{15}\text{N}$ steady-state NOE values were acquired using the pulse sequences described in Farrow et al. (1994). These enhanced sensitivity pulse sequences employ pulsed field gradients to minimize artifacts, suppress the solvent signal and select for the coherence transfer pathway from $^{15}\text{N}$ to $^1\text{H}$ for observation. Additionally, a selective water flip back pulse is incorporated to ensure minimum perturbation of the water magnetization.

In total 11 different $^{15}\text{N}$ relaxation series are reported here (table 5.1). All spectra were recorded at 35 °C using either a Varian Unity 500-MHz spectrometer or a 600 MHz Varian Inova spectrometer (Prof. Lewis Kay, University of Toronto). Two different sample buffers were used. Initially, the buffer was 50 mM sodium chloride, 50 mM potassium phosphate (pH* 5.9), 0.02% sodium azide, 10% D$_2$O / 90% H$_2$O. When it was found that CB$D_{\text{NI}}$ bound calcium (chapter 4), the buffer used in subsequent experiments was 50 mM sodium chloride, 50 mM sodium acetate (d3) (pH* 6.0), 0.02% sodium azide, 10% D$_2$O / 90% H$_2$O. Samples 6 to 11 are calcium loaded to a level of approximately 20% (chapter 4).

The data for apo CB$D_{\text{NI}}$ (sample 1) were obtained by collecting spectra on celloooligosaccharide-free CB$D_{\text{NI}}$ in the presence of 3.8 molar equivalents of sodium ethylenediamine tetraacetate (EDTA). Samples 2 and 4, 3 and 5, and 6 and 7 were the same samples, respectively, with solid cellopentaose added to prevent dilution of the protein sample. Samples 9, 10 and 11 were from a common pool of protein that was split in two. Using one half of the sample the unbound relaxation series was run, then solid cellopentaose added. To the other half solid cellotetraose was added.
Table 5.1. $^{15}$N relaxation experiments performed on $CBD_{N1}$.

<table>
<thead>
<tr>
<th>sample #</th>
<th>sample description</th>
<th>protein conc. (mM)</th>
<th>buffer $^{c}$</th>
<th>protein conc. (mM)</th>
<th>buffer $^{c}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Apo; Calcium and sugar-free</td>
<td>1.3</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Calcium-bound</td>
<td>1.3</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Calcium-bound</td>
<td>1.3</td>
<td>A</td>
<td>1.9</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>Calcium and Cellopentaose-bound</td>
<td>1.3</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Calcium and Cellopentaose-bound</td>
<td></td>
<td></td>
<td>1.9</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>Dilute, sugar-free</td>
<td>0.3</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Dilute, Cellopentaose-bound</td>
<td>0.3</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Concentrated sample</td>
<td>5.3</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Sugar-free</td>
<td>1.6</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Cellopentaose-bound</td>
<td>1.6</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Cellotetraose-bound</td>
<td>1.6</td>
<td>B</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^{a}$ Data collected using $^{15}$N-labelled $CBD_{N1}$ samples. $^{b}$ Data collected using $^{15}$N / $^{13}$C / $^{2}$H (40%) $CBD_{N1}$. $^{c}$ (A): 50 mM sodium chloride, 50 mM sodium acetate (d3) (pH* 6.0), 0.02% sodium azide, 10% D$_2$O / 90% H$_2$O. (B): 50 mM sodium chloride, 50 mM potassium phosphate (pH* 5.9), 0.02% sodium azide, 10% D$_2$O / 90% H$_2$O. Samples in buffer B are approximately 20% calcium bound.

For sample 3, the ratio of protein-to-calcium was 1:2; for sample 2 this ratio was 1:5. Using the experimentally determined association constant of $CBD_{N1}$ with calcium (113 000 M$^{-1}$; chapter 4) it is calculated that 99.5% of $CBD_{N1}$ is calcium-loaded in sample 3, and 99.8% calcium-loaded in sample 3. In the case of the cellobio oligosaccharide-bound samples, sugar was added to such a level that the fraction of protein that was ligand-bound was 95%. The weight of sugar needed to achieve this level was calculated using the association constants previously determined (Tomme et al., 1996a; Johnson et al., 1996a; Chapter 2).

$^{15}$N-labelled $CBD_{N1}$ samples were used for experiments collected at 500 MHz. $^{15}$N $T_1$ values were recorded with nine different durations of delay $T$ = 11.1, 33.3, 55.5, 99.9, 166.5, 266.4, 410.7, 643.8 and 999.0 ms. $T_2$ values were determined from spectra recorded with the
delays T = 16.7, 33.4, 50.1, 66.8, 83.5, 100.2, 116.9, 133.6 and 150.3 ms. For each series, the spectra were acquired in mixed order. To test the reliability of the data, duplicate spectra were initially recorded for T = 55.5 ms (T1) and T = 50.1 ms (T2). 1H-15N NOE values were determined from spectra recorded in the presence and absence of a proton presaturation period of 3 s. In the case of the control no-NOE spectra, a net relaxation delay of 5 s was used, while a relaxation delay of 2 s prior to a 3 s proton presaturation period was used for the NOE spectra. Spectra were acquired using spectral widths of 6500 Hz and 1450 Hz in the 1H and 15N dimensions, respectively. Spectra were acquired as either 96 x 1024 or 96 x 512 complex matrices.

For experiments collected at 600 MHz, data were acquired using the 15N / 13C / 2H (40%) CBDN1 sample. Prior to data collection this deuterated sample was heated to 90 °C to unfold the protein. This allowed any amides which might be residually deuterated from expression in 40% D2O to fully exchange with protons in the sample buffer. Thermal unfolding of CBDN1 is a completely reversible process (chapter 2; Creagh et al., 1997). 15N T1 values were collected with delays T = 11.02, 66.12, 132.25, 220.42, 308.58, 396.75, 517.98, 650.23 and 804.52 ms. T2 values were determined from spectra recorded with the delays T = 16.66, 33.32, 49.98, 66.64, 83.30, 99.96, 116.61, 133.27 and 149.93 ms. In the case of the control no-NOE spectra, a net relaxation delay of 6.5 s was used, while a relaxation delay of 3.5 s prior to a 3 s proton presaturation period was used for the NOE spectra. Data were acquired using spectral widths of 9000 Hz and 1720 Hz in the 1H and 15N dimensions, respectively. Spectra were 13C decoupled and acquired as 128 x 576 complex matrices.

Methyl group deuteron relaxation experiments

NMR experiments were performed using a 600 MHz Varian Inova spectrometer (Prof. Lewis Kay, University of Toronto) on a 1.9 mM sample of 15N/13C/2H(40%) CBDN1 in 50 mM sodium chloride, 50 mM sodium d3-acetate, 0.02% sodium azide, 4.2 mM calcium chloride, 10% D2O / 90% H2O, pH* 6.0, 35 °C in the presence and absence of 3.2 mM cellopentaose. Pulse sequences used for the measurement of the relaxation properties of methyl group deuterons were
those described by Muhandiram et al. (1995). To obtain $T_1(I_2C_2D_2)$ values, nine two-dimensional $^1H$-$^{13}C$ correlation spectra were acquired with delays of 0.05, 4.1, 8.6, 13.5, 19.0, 25.1, 32.1, 40.3, and 50.0 ms. Values of $T_1(I_2C_2)$ were obtained with identical delays. Values of $T_1(I_2C_2D_y)$ were obtained using delays of 0.20, 1.3, 2.8, 4.3, 6.1, 8.0, 10.3, 12.9, and 16.0 ms. The same delays were used when collecting data on both the free and cellopentaose bound forms of CBD$_{N1}$. All spectra were obtained as 128 x 576 complex points with spectral widths of 5000 and 9000 Hz in the $^{13}C$ and $^1H$ dimensions, respectively.

Data processing

Both amide $^{15}N$ relaxation and methyl group deuteron relaxation experiments were processed using NMRPipe (Delaglio et al., 1995). Lorentzian-to-Gaussian apodization functions were applied in both dimensions. For the constant time $^1H$-$^{13}C$ HSQC experiments acquired to measure methyl group deuteron relaxation, mirror image linear prediction was employed in the t1 ($^{13}C$) dimension (Zhu & Bax, 1992). Relative peak volumes were measured using the nlinLS module built into NMRPipe. A very conservative approach was employed when peak picking the spectra. Data from peaks that lie close to each other, but not totally overlapped, were not used if their measured $T_1$ or $T_2$ values were similar. This results in a relatively large number of missing data points, but a high level of confidence can be placed in the final data.

Using relative peak volumes and the delay values (T) as input for the programme lmquick (Dr. Neil Farrow, University of Toronto), per-residue $^{15}N$ $T_1$ and $T_2$ and methyl group $^2H$ $T_1(I_2C_2)$, $T_1(I_2C_2D_2)$ and $T_1(I_2C_2D_y)$ values were obtained. lmquick fits the relative peak volumes to a function of the form $I(T) = I(0) \exp(-T/T_1)$, where $1/T_1$ is the relevant relaxation rate, $I(T)$ is the relative peak volume at time T and $I(0)$ is the intensity at time T=0. Errors in the measured relaxation rates are estimated using Monte Carlo procedures as described in Farrow et al. (1994).
Chapter 5-Dynamic Analysis of Ligand Binding

Steady-state $^1\text{H}-^{15}\text{N}$ NOE values were determined from the ratios of the intensities of the peaks with ($I_{\text{sat}}$) and without proton saturation ($I_{\text{unsat}}$).

$$\text{NOE} = \frac{I_{\text{sat}}}{I_{\text{unsat}}}$$

The standard deviation of the NOE value ($\sigma_{\text{noe}}$) was determined on the basis of the measured background noise levels using the following relationship

$$\frac{\sigma_{\text{noe}}}{\text{NOE}} = \left(\left(\frac{\sigma_{\text{sat}}}{I_{\text{sat}}}\right)^2 + \left(\frac{\sigma_{\text{unsat}}}{I_{\text{unsat}}}\right)^2\right)^{1/2}$$

A routine in NMRPipe was used to measure the rms noise of background regions of the spectra ($\sigma_{\text{sat}}$ and $\sigma_{\text{unsat}}$).

For the $^{15}\text{N}$ relaxation analysis, a value for the overall correlation time for the molecule ($\tau_m$) was calculated using the programme tm_f77 (Dr. Neil Farrow, University of Toronto). This method uses a grid search in $\tau_m$, optimizing $\tau_e$ and $S^2$ for each residue (equation 5.6). Data with a $T_1/T_2$ ratio differing by more than one standard deviation from average were not used in the process of determining $\tau_m$ values. Once an overall correlation time was found, the data were fit to the different spectral density functions outlined in the introduction to this chapter (equations 5.6 and 5.10). The programme j_f77 (Dr. Neil Farrow, University of Toronto) was used for this process.

Five different models of the spectral density function were fit to the experimentally determined $^{15}\text{N}$ backbone $T_1$, $T_2$, and $^1\text{H}-^{15}\text{N}$ NOE values. These are: (1) a function of the form of equation 5.8 with $\tau_e$ fixed at zero; (2) a function of the form of equation 5.6 with $\tau_e$ used as a fitting parameter; (3) a function of the form of equation 5.6 with $\tau_e$ fixed at zero and incorporating an $R_{\text{ex}}$ term; (4) a function of the form of equation 5.6 with $\tau_e$ and $R_{\text{ex}}$ terms used as fitting parameters; and (5) the two time scale version of the spectral density function given in equation 5.9. Table 5.2 summarises these models and the parameters fit in each of them.

The method for model selection was to proceed from the simplest (model 1) to the more complex model (model 4) as follows; (i) a model was discarded if all the experimental parameters ($T_1$, $T_2$, $^1\text{H}-^{15}\text{N}$ NOE) could not be fit to within the 95% confidence limit; (ii) the value of the
fitting parameter must exceed its calculated error; (iii) if both (i) and (ii) are satisfied the model with
the lowest $\chi^2$ for the fits to the experimental data was chosen; (iv) finally, if none of models 1 to 4
(Table 5.2) satisfied criteria (i) and (ii) the two time scale model (model 5) was chosen provided
that points (i) and (ii) were met.

| Table 5.2. Summary of the model-free parameters for $^{15}$N relaxation analysis |
|---------------------------------|--------------------------------|
| model  | fitting parameters | description |
| 1      | $S^2$              | Fast isotropic motion of the N-H bond. |
| 2      | $S^2$, $\tau_e$   | A fast motion ($\tau_e$) affects $\tau_{in}$. |
| 3      | $S^2$, $R_{ex}$   | The amide experiences chemical exchange line broadening. |
| 4      | $S^2$, $\tau_e$, $R_{ex}$ | Fast motions and chemical exchange affect relaxation. |
| 5      | $S^2$, $S_f^2$, $\tau_s$ | Motions occur on two time scales. |

For the analysis of the methyl group deuteron data the values of $T_1(D)$ and $T_{1p}(D)$ were fit
only to the first two models of table 5.2 according to equation 5.15. The criteria to select a model
was the same as points (i) to (iii) as outlined above.

Resonance assignment of the calcium-loaded CBD$_{N1}$

The backbone $^{15}$N and H$^N$ resonances of calcium-loaded oligosaccharide-free CBD$_{N1}$ were
assigned from the analysis of the HNCACB (Wittekind & Mueller, 1993) and CBCA(CO)NH
(Grzesiek & Bax, 1992) experiments acquired on the $^{13}$C / $^{15}$N / $^2$H(40%) CBD$_{N1}$ sample. The
assignment process was facilitated by the use of a pulse sequence to correlate the $^{15}$N and H$^N$
resonances of residues that immediately follow methyl-containing amino acids in $^{15}$N, $^{13}$C and
fractionally deuterated proteins (Muhandiram et al., 1997).

Assignments of the methyl groups of calcium-loaded oligosaccharide free CBD$_{N1}$ were
obtained from analysis of H(CCO)NH-TOCSY (Montelione et al., 1992; Grzesiek et al., 1993a)
and (H)C(CO)NH-TOCSY (Logan et al., 1992; Grzesiek et al, 1993a) experiments. The data
collection and processing parameters for the NMR experiments to assign calcium-loaded sugar-free
CBD_{N1} is given in table 5.3. All assignment experiments were processed and analysed using Felix version 2.3 (Biosym).

<table>
<thead>
<tr>
<th>Experiment</th>
<th>nucleus (complex)</th>
<th>acq. points</th>
<th>spectral width (Hz)</th>
<th>carrier freq. (ppm)</th>
<th>processing (real)</th>
<th>matrix dim.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HNCACB</td>
<td>t1=^{13}Cαβ</td>
<td>38</td>
<td>9179</td>
<td>43.0</td>
<td>lp, ss (90°)</td>
<td>256</td>
</tr>
<tr>
<td></td>
<td>t2=^{15}N</td>
<td>30\text{ct}</td>
<td>1720</td>
<td>119.0</td>
<td>lp, ss (80°)</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>t3=^{1}HN</td>
<td>576</td>
<td>9000</td>
<td>4.68</td>
<td>tdc, gm (15), poly</td>
<td>256</td>
</tr>
<tr>
<td>CBCA(CO)NH</td>
<td>t1=^{13}Cαβ</td>
<td>48\text{ct}</td>
<td>9179</td>
<td>43.0</td>
<td>lp, ss (80°)</td>
<td>256</td>
</tr>
<tr>
<td></td>
<td>t2=^{15}N</td>
<td>30\text{ct}</td>
<td>1720</td>
<td>119.0</td>
<td>lp, ss (80°)</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>t3=^{1}HN</td>
<td>576</td>
<td>9000</td>
<td>4.68</td>
<td>tdc, gm (15), poly</td>
<td>256</td>
</tr>
<tr>
<td>H(CCO)-TOCSY-NH</td>
<td>t1=^{1}H</td>
<td>58</td>
<td>9179</td>
<td>4.68</td>
<td>lp, ss (90°)</td>
<td>256</td>
</tr>
<tr>
<td></td>
<td>t2=^{15}N</td>
<td>30\text{ct}</td>
<td>1720</td>
<td>119.0</td>
<td>lp, ss (80°)</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>t3=^{1}HN</td>
<td>576</td>
<td>9000</td>
<td>4.68</td>
<td>tdc, gm (15), poly</td>
<td>256</td>
</tr>
<tr>
<td>C(CO)-TOCSY-NH</td>
<td>t1=^{13}C</td>
<td>58</td>
<td>9179</td>
<td>43.0</td>
<td>lp, ss (90°)</td>
<td>256</td>
</tr>
<tr>
<td></td>
<td>t2=^{15}N</td>
<td>30\text{ct}</td>
<td>1720</td>
<td>119.0</td>
<td>lp, ss (80°)</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>t3=^{1}HN</td>
<td>576</td>
<td>9000</td>
<td>4.68</td>
<td>tdc, gm (15), poly</td>
<td>256</td>
</tr>
</tbody>
</table>

a Data were collected at 600 MHz using a Varian Inova spectrometer.
ct=constant time; lp=linear prediction (or mirror image linear prediction with constant-time);
ss=sinebell squared (degrees shifted); sb=sinebell (degrees shifted); gm=Lorentzian-to-Gaussian multiplication with a maximum at approximately 0.1 of the acquisition time (Hz line broadening);
poly=polynomial baseline flattening; tdc=time domain convolution.

Ultracentrifugation and viscosity measurements

Ultracentrifuge sedimentation velocity runs were performed at 20 °C on 0.35 mM samples of free and cellopentaose-bound CBD_{N1}. A 3.4 fold molar excess of cellopentaose was used, resulting in CBD_{N1} being 94% bound. Sedimentation equilibrium was performed on the bound
Chapter 5-Dynamic Analysis of Ligand Binding

form of CBD$_{N1}$ at a protein concentration of 0.23 mM. This complements the measurement run previously performed on an unbound sample of CBD$_{N1}$ (chapter 2).

Viscosity measurements were performed on free and cellopentaose-bound forms of CBD$_{N1}$ using Cannon-Manning Semi-Micro type viscometers at 35 °C. Viscosity was initially measured using 1.9 mM protein samples. The cellopentaose-bound sample of CBD$_{N1}$ contained an 1.4 fold excess of cellopentaose, resulting in 94% binding. Three dilutions were then made, with the viscosity measured at each step. The buffer used for dilution of the ligand-bound form of CBD$_{N1}$ contained cellopentaose at the same concentration as in the undiluted sample. Thus, upon dilution, the relative concentration of cellopentaose-to-protein increased, resulting in 98% saturation after the final dilution. The buffer used for both ultracentrifuge and viscosity measurements was 50 mM sodium chloride, 50 mM potassium phosphate (pH 5.9), 0.02% sodium azide. It is estimated that under these conditions, the protein is approximately 20% calcium loaded (chapter 4). The ultracentrifuge and viscosity measurements were carried out by Les Hicks of the University of Alberta.

Results

$^{15}$N $T_1$, $T_2$ and NOE values

$T_1$, $T_2$ and $^1H$-$^{15}$N NOE values were determined for eleven different samples of CBD$_{N1}$ (Table 5.1). Shown in figure 5.2 are portions of the 2D $^1H$-$^{15}$N HSQC spectra used to determine the $T_2$ values of apo CBD$_{N1}$ (sample 1). Examples of the decay curves to determine $T_1$ and $T_2$ values are shown in figure 5.3. In the analysis of the data obtained for the various samples of CBD$_{N1}$, it was not possible to determine the relaxation parameters for each backbone amide. This is mainly due to resonance overlap, but also, in the case of sample 1, due to extreme line broadening, resulting in weak peaks for a number of resonances. Table 5.4 summarises the number of residues for which data was obtainable for each sample discussed here.
Figure 5.2. Portion of the $^1$H-$^{15}$N correlation spectra recorded to measure T2 values for calcium-free and oligosaccharide-free CBDN1 (sample 1) recorded with delay values of (A) 16.7 ms, (B) 50.1 ms, (C) 100.2 ms and (D) 150.3 ms.
Figure 5.3. Two parameter monoexponential curve fitting of the (A) $T_1$ and (B) $T_2$ data. $T_1$ data points for residues D71 (cross), S40 (circle), N50 (diamond) and V48 (square), and $T_2$ data points for residues D10 (cross), V45 (circle), E15 (diamond) and G44 (square) are shown.
Table 5.4. Number of residues per sample for which $^{15}$N relaxation data was obtained.$^a$

<table>
<thead>
<tr>
<th>sample$^b$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of residues analysed</td>
<td>89</td>
<td>81</td>
<td>84</td>
<td>94</td>
<td>80</td>
</tr>
</tbody>
</table>

$a$ The expected number is 141. This is based on the number of residues in CBD$_{NI}$ less the number of proline residues and the N-terminal amino group.

$b$ Refer to Table 5.1 for sample descriptions.

The value of the relaxation parameters $T_1$, $T_2$ and $^{1}$H-$^{15}$N NOE are shown in figure 5.4 for calcium-loaded CBD$_{NI}$ free and cellopentaose-bound at both 500 MHz and 600 MHz (samples 2-5). The average value of $T_1$ for sugar-free CBD$_{NI}$ is $0.53$ s at 500 MHz and $0.65$ s at 600 MHz. The average $T_2$ value is $0.12$ s at 500 MHz and $0.11$ s at 600 MHz. For cellopentaose-bound CBD$_{NI}$ the average $T_1$ values are $0.53$ s and $0.68$ s for data at 500 MHz and 600 MHz, respectively. The average $T_2$ value is $0.12$ s for the data collected at both field strengths. For sample 1, the calcium- and sugar-free sample, $T_1$, $T_2$ and $^{1}$H-$^{15}$N NOE values are shown in figure 5.5. Also shown in this figure, for comparison purposes, are the values for the calcium-loaded form of CBD$_{NI}$, sample 2. The average value of $T_1$ for sample 1 is $0.52$ s, while the average $T_2$ value is $0.11$ s; this compares to $0.53$ and $0.12$ s for the $T_1$ and $T_2$ values of the calcium-loaded CBD$_{NI}$ (sample 2), respectively.

The $T_1$ values are very uniform throughout the protein for all the samples. Besides the termini the only values that significantly differ from average in all samples are residues G12, G23, D71, F132 and S133. Residue L32 has a high $T_1$ value in sample 2, but its value was not measurable in any other sample.

In contrast to the relative uniformity of the $T_1$ data, the $T_2$ values show a greater variation. Residues with a large $T_2$ value in all samples are G12 and G23. Residue L32 has a large $T_2$ value in sample 2, but is not measurable in any other sample. Residues with $T_2$ values that are significantly smaller than average in all samples are T8, V17, and V34. For sample 1, residues I4 and G7 have low $T_2$ values, yet an average value in sample 2. Residues E14 and A66 in sample 1
Figure 5.4. $^{15}$N $T_1$, $T_2$ and $^1$H-$^{15}$N NOE data for: (left) calcium-bound (samples 2 and 3); (right) calcium-bound cellopentaose-bound (samples 4 and 5) CBD$_{N1}$. Open circles show data collected at 500 MHz, filled circles show data collected at 600 MHz. Bars indicate error values that are greater than the radius of the circles. The secondary structure of CBD$_{N1}$ is shown at the top of the diagram. Open and closed boxes indicate β-strands on sheet A (the oligosaccharide binding face) and sheet B, respectively.
Figure 5.5. $^{15}$N $T_1$, $T_2$ and $^1$H-$^{15}$N NOE data for calcium-free CBD$_{N1}$ (open circles; sample 1) and calcium-bound CBD$_{N1}$ (closed circles; sample 2). Bars indicate error values that are greater than the radius of the circles. The secondary structure of CBD$_{N1}$ is shown at the top of the diagram, the β-strands are labelled in this diagram. Arrows indicate the putative calcium-binding residues.
also have low $T_2$ values but no data was obtainable for these residues in sample 2. Residues L139 and L141 in sample 1 have a low $T_2$, whereas the value of L141 is average in sample 2, while that of L139 was not measurable in sample 2. Residue T87 has a low $T_2$ value in sample 2 and 3, but is not measurable in samples 4 and 5. Residue G44 has a low $T_2$ value in samples 2 and 3 and an average values in samples 4 and 5.

For all samples, the greatest variation in steady-state $^1$H-$^{15}$N NOE values lie with the first 23 residues. The only noticeable variations after this point arise in the loop regions near residues 100 and 120. The C-terminal region, residues E149 to L152 have very low $^1$H-$^{15}$N NOE values in all samples. When comparing the $^1$H-$^{15}$N NOE values of the calcium-free and calcium-loaded forms of CBD$_N$ (samples 1 and 2), G12 has a low $^1$H-$^{15}$N NOE value in both sample 1 and 2. In sample 1 residues I4 and D11 have $^1$H-$^{15}$N NOE values much lower than average, with their corresponding values in sample 2 being average. Those of residues D10 and G23 are much lower than average for sample 1, but were not measurable in sample 2. In comparing the $^1$H-$^{15}$N NOE results with and without cellulose (samples 2 to 5), residues G12 and G23 are low in all cases, residue D22 is low in sample 3 but not measurable in the other samples. It also appears that there is less scatter in the $^1$H-$^{15}$N NOE values of the first 23 residues of the samples bound to cellulose than in the free form of the protein.

*Estimation of the overall correlation time $\tau_m$*

Values for the overall correlation time ($\tau_m$) of CBD$_N$ were determined by fitting the relaxation data on a per-residue basis using the model-free spectral density function given in equation 5.6 (table 5.5). The values of the correlation time reported are the average of the optimal per-residue $\tau_m$ values, using only residues with a $T_1/T_2$ ratio within one standard deviation of the mean $T_1/T_2$ value. The error reported is the standard deviation of the $\tau_m$ values considered in this data set.
Table 5.5. Values for the overall correlation times ($\tau_m$) determined for CBD$_{N1}$ at 35 °C.

<table>
<thead>
<tr>
<th>sample</th>
<th>sample description</th>
<th>$\tau_m$ (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Apo; Calcium-free, sugar-free, 1.3 mM</td>
<td>7.6 ± 0.4</td>
</tr>
<tr>
<td>2</td>
<td>Calcium-bound, 1.3 mM, 500 MHz</td>
<td>7.5 ± 0.5</td>
</tr>
<tr>
<td>3</td>
<td>Calcium-bound, 1.9 mM, 600 MHz</td>
<td>7.8 ± 0.3</td>
</tr>
<tr>
<td>4</td>
<td>Calcium and Cellopentaose bound, 1.3 mM, 500 MHz</td>
<td>7.2 ± 0.4</td>
</tr>
<tr>
<td>5</td>
<td>Calcium and Cellopentaose bound, 1.9 mM, 600 MHz</td>
<td>7.1 ± 0.4</td>
</tr>
<tr>
<td>6</td>
<td>Dilute sample; sugar-free, 0.3 mM</td>
<td>7.2 ± 0.4</td>
</tr>
<tr>
<td>7</td>
<td>Dilute sample; cellopentaose-bound, 0.3 mM</td>
<td>7.3 ± 0.4</td>
</tr>
<tr>
<td>8</td>
<td>Concentrated sample; sugar-free 5.3 mM</td>
<td>8.9 ± 0.4</td>
</tr>
<tr>
<td>9</td>
<td>Sugar-free, 1.6 mM</td>
<td>7.4 ± 0.4</td>
</tr>
<tr>
<td>10</td>
<td>Cellopentaose-bound, 1.6 mM</td>
<td>7.3 ± 0.3</td>
</tr>
<tr>
<td>11</td>
<td>Cellotetraose-bound, 1.6 mM</td>
<td>7.2 ± 0.4</td>
</tr>
</tbody>
</table>

Equation 5.6 assumes that the tumbling of the molecule is isotropic. This assumption, in the case of CBD$_{N1}$, is valid when the three-dimensional structure of this protein is considered. CBD$_{N1}$ (Johnson et al., 1996b; Chapter 3), to a first approximation, is a sphere, with no one axis being significantly longer than another. This implies that the molecule rotates isotropically in solution.

$T_1/T_2$ analysis

The $T_1/T_2$ ratios of CBD$_{N1}$ samples 1 to 5 are shown in figure 5.6. Throughout much of the protein, the $T_1/T_2$ values are similar, suggesting isotropic motion of the protein in solution. $T_1/T_2$ data can be easily analysed in a qualitative way. $T_1/T_2$ values lower than average tend to indicate motions are occurring on two separate time scales, or there are fast motions present (equation 5.10; model 5, table 5.2). $T_1/T_2$ values higher than average usually indicate that the $T_2$ values are decreased by a chemical exchange process ($R_{ex}$) (Clore et al., 1990b).

In all samples, residues with the highest $T_1/T_2$ ratios tend to occur in the N-terminal region of the protein. The highest overall $T_1/T_2$ ratios occur in this region in apo CBD$_{N1}$ (sample 1).
Figure 5.6. $T_1/T_2$ values for: (A) calcium-bound CBD$_{N1}$; (B) calcium-bound cellopentaose-bound CBD$_{N1}$. Open circles show data collected at 500 MHz, filled circles show data collected at 600 MHz. In (C) calcium-free $T_1/T_2$ values are shown as filled circles and calcium-bound values are shown as open circles. Arrows indicate the putative calcium-binding residues.
Here residues I4, G7, E14 and V17 all have noticeably higher than average $T_1/T_2$ ratios. In the calcium-bound form of CBD$_{N1}$ the $T_1/T_2$ ratios in this region are significantly lower, but still contain residues with higher than average values (T8 and V17). L139 in sample 1 also has a higher than average $T_1/T_2$ ratio. For samples 2 and 3 residues T8, V17, V34, G44, N81, T87 and G130 all have higher than average $T_1/T_2$ ratios. With cellopentaose binding, the values of these ratios, with the exception of T87 for which data was not obtainable in samples 4 and 5, all decrease. Only T8, V17 and V34 remaining noticeably higher than average in the cellopentaose bound state. The $T_1/T_2$ ratio of V48 is slightly higher than average in sample 4 and 5 but its value was not significantly changed from what it was in samples 2 and 3.

Model-free analysis of backbone $^{15}$N relaxation data

As described in the introduction to this chapter, the relaxation parameters ( $T_1$, $T_2$ and $^{1}$H-$^{15}$N NOE) of each residue were fit using the spectral density functions given in equations 5.6 and 5.9. An exchange parameter, $R_{ex}$, was also included in some models as outlined in table 5.2. The optimum spectral density models were chosen as described in the experimental section. Table 5.6 summarises the number of residues that fall within each class of model for each of samples 1 to 5. Also included is the number of residues for which no model was satisfactorily fit.

<table>
<thead>
<tr>
<th>model</th>
<th>sample$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1 (S$^2$)</td>
<td>1</td>
</tr>
<tr>
<td>2 (S$^2$, $\tau_e$)</td>
<td>10</td>
</tr>
<tr>
<td>3 (S$^2$, $R_{ex}$)</td>
<td>7</td>
</tr>
<tr>
<td>4 (S$^2$, $\tau_e$, $R_{ex}$)</td>
<td>19</td>
</tr>
<tr>
<td>5 (S$^2$, S$^T$,$\tau_s$)</td>
<td>47</td>
</tr>
<tr>
<td>none</td>
<td>4</td>
</tr>
</tbody>
</table>

$^a$ See table 5.1 for sample descriptions.
Figure 5.7 summarises the values of the fitting parameters used in the model-free analysis of free and cellopentaose-bound CBD$_{N1}$ at both 500 MHz and 600 MHz. Model-free data for the calcium and sugar-free form of CBD$_{N1}$ (sample 1) are shown in figure 5.8, with the data of sample 2 included for comparison. The average value of the square of the order parameter ($S^2$) over all residues for sample 1 is 0.79, for sample 2 this is 0.81, for sample 3 it is 0.82, sample 4, 0.80 and for sample 5 it is 0.79. The overall trends in the order parameter are similar in all forms of the protein. While there is variation in the $S^2$ value for each sample, the only region of the protein where the $S^2$ values are consistently low is at the C-terminus, residues 148-152.

A $\tau_e$ parameter was needed to fit the data throughout much of the protein in each sample. For all samples, longer than average $\tau_e$ values are found at the N-terminal region up to residue 15 and residues G44 to V48 in samples 1 to 3. Longer $\tau_e$ values for the N-terminal residues are found in the calcium free sample than in the calcium-loaded sample. The most significant difference between $\tau_e$ values in sugar-free and cellopentaose-bound forms of CBD$_{N1}$ is the reduction in the $\tau_e$ values for residues G44 to V48 in the sugar bound form.

Many of the residues in CBD$_{N1}$ need an $R_{ex}$ term to correctly model the experimental data. For all samples, $R_{ex}$ terms tend to occur at similar locations in the structure although with different values. For sample 1, the $R_{ex}$ terms are highest in the N-terminal region up to residue V17, and at residues L139 and C140. Obtaining data for sample 1 in the N-terminal region was hampered by fact that peaks in this region are in general very weak due to line-broadening. This line-broadening is less evident in the calcium-loaded sample. Line-broadening is indicative of the presence of motion on the chemical exchange ($R_{ex}$) time scale (millisecond). For samples 2 and 3 the $R_{ex}$ terms are largest at residues V17, V34, G44 and T87. The $R_{ex}$ term for these residues, and in general for all residues, is lower in samples 4 and 5.
Figure 5.7. Model-free parameters for: (left) calcium-bound (samples 2 and 3); (right) calcium-bound cellopentaose-bound (samples 4 and 5) CBD$_{N1}$. Open circles show data collected at 500 MHz, filled circles show data collected at 600 MHz. Bars indicate error values that are greater than the radius of the circles. The secondary structure of CBD$_{N1}$ is shown at the top of the diagram. Open and closed boxes indicate $\beta$-strands on sheet A (the oligosaccharide binding face) and sheet B, respectively.
Figure 5.8. Model-free parameters for calcium-free CBD$_{N1}$ (open circles; sample 1) and calcium-bound CBD$_{N1}$ (closed circles; sample 2). Bars indicate error values that are greater than the radius of the circles. The secondary structure of CBD$_{N1}$ is shown at the top of the diagram, the $\beta$-strands are labelled in this diagram.
Ultracentrifuge and viscosity measurements

Ultracentrifuge sedimentation velocity runs were performed on 0.35 mM samples of free and cellopentaose-bound forms of CBD\textsubscript{N1} using identical sample conditions. Both samples displayed single, symmetrical Schlieren peaks. Assuming an average partial specific volume of 0.73 cm\textsuperscript{3} g\textsuperscript{-1} for the bound and free forms of the protein, the calculated S\textsubscript{20,\textomega} values are 1.60 for the free sample and 1.70 for the cellopentaose-bound form of CBD\textsubscript{N1}. This slight increase in the S\textsubscript{20,\textomega} value is just beyond the range of expected experimental error. Based on sedimentation equilibrium measurements, cellopentaose-bound CBD\textsubscript{N1} is a monomeric protein with an apparent molecular mass of 15.1 kDa. This compares to a value of 14.1 kDa found previously for unbound CBD\textsubscript{N1} in identical conditions (chapter 2).

The viscosities of samples of both free and cellopentaose-bound CBD\textsubscript{N1} were measured at 1.9 mM and 1.8 mM, respectively, and also at three dilution levels. Also measured was the viscosity of the two buffers, with and without cellopentaose, used for the dilutions. This data is shown in figure 5.9. Fitting this data to a straight line results in slopes of 0.0471 cSt mM\textsuperscript{-1} and 0.0459 cSt mM\textsuperscript{-1}, and y-intercepts of 0.762 cSt and 0.759 cSt for cellopentaose-bound and unbound CBD\textsubscript{N1}, respectively.

Resonance assignment of calcium-loaded CBD\textsubscript{N1}

Previously, resonance assignments of cellotetraose-bound and calcium-bound CBD\textsubscript{N1} were obtained (chapter 3; Johnson et al 1996b). However, at the time, it was not know that calcium was bound by the protein. By following the titration of CBD\textsubscript{N1} with the calcium contaminated cellotetraose in reverse, to the level where CBD\textsubscript{N1} was approximately 20\% bound, then following a separate titration of apo-CBD\textsubscript{N1} with calcium, it was possible to assign unambiguously most of the resonances in a \textsuperscript{1}H-\textsuperscript{15}N HSQC spectrum of calcium-bound sugar-free CBD\textsubscript{N1}. To resolve the ambiguous assignments, HNCACB and CBCA(CO)NH experiments were run on the calcium-bound cellopentaose-free \textsuperscript{13}C/\textsuperscript{15}N/\textsuperscript{2}H(40\%) CBD\textsubscript{N1} sample. The triple-resonance data obtained from this 1.9 mM fractionally deuterated sample were superb, an example being shown in figure
Figure 5.9. (top) Values of viscosity plotted against protein concentration. Data collected in the presence of cellopentaose are shown by squares and a long dashed line. Data collected in the absence of oligosaccharide is shown by circles and a solid line. (bottom) Values of the overall rotational correlation time ($\tau_m$) (diamonds; short dashed lines) of samples 6 (0.3 mM), 8 (5.3 mM) and 9 (1.6 mM), and values of viscosity plotted against protein concentration.
3.3. The assignment process was straightforward. A labelled $^1$H-$^{15}$N HSQC spectrum showing these assignments is shown in figure 5.10.

The resonance assignment process was aided by the development of a new experiment that provides $^{15}$N-$^N^N$ correlations for residues that immediately follow methyl-containing amino acids (Muhandiram et al., 1997). This triple-resonance experiment requires an $^{15}$N, $^{13}$C and fractionally deuterated protein sample. The correlations shown by this experiment provide a means of helping to confirm the protein sequence location while carrying out resonance assignments in much the same way that a specifically-labelled protein sample does. A spectrum generated by this experiment is shown in figure 5.11. Three of the 69 methyl containing residues precede a proline and therefore do not give rise to $^{15}$N-$^N$ correlation peaks in this figure. All 66 expected peaks are observed in this spectrum, although the correlations for Asn50, Ser69, and Thr96 are weak.

Once the backbone $^{15}$N amide resonances were determined, the assignments of the methyl groups of CBD$_{N1}$ in the calcium-bound sugar-free state were obtained. H(CCO)NH-TOCSY and (H)C(CO)NH-TOCSY experiments, together with a constant time $^1$H-$^{13}$C HSQC on a sample of 10% fractionally $^{13}$C-labelled CBD$_{N1}$ were used to obtain these resonance assignments (figure 5.12). An example of the data obtained from the H(CCO)NH-TOCSY and (H)C(CO)NH-TOCSY experiments is shown in figure 3.4. Following the titration of the 10% fractionally $^{13}$C-labelled CBD$_{N1}$ sample with cellopentaose (Figure 5.13) provided a means of obtaining the methyl group assignments of the sugar-bound CBD$_{N1}$ sample.

*Methyl deuteron relaxation and analysis*

$T_1(I_2C_2D_2), T_1(I_2C_2)$ and $T_1p(I_2C_2D_y)$ values were measured to determine deuteron $T_1(D)$ and $T_1p(D)$ values for methyl groups in free and cellopentaose-bound CBD$_{N1}$, according to equations 5.11 and 5.12. Figure 5.14 shows a portion of the spectra used to determine the $T_1p(I_2C_2D_2)$ rate of calcium-bound cellopentaose-free CBD$_{N1}$. Examples for the decay curves for the $T_1(I_2C_2D_2), T_1(I_2C_2)$ and $T_1p(I_2C_2D_y)$ magnetization are provided in figure 5.15. Data were obtained for 50 methyl groups in the case of unbound CBD$_{N1}$, and 49 methyl groups for
Figure 5.10. 1H-15N HSQC spectrum of calcium-loaded cellobiose-free CBDN1. Aliased peaks are marked with an asterisk and are due to arginine side-chain nuclei.
Figure 5.11. (A) Spectrum illustrating the 15N-HN correlations of residues that immediately follow methyl containing amino acids. The peak labelled * is of opposite phase to all other peaks and arises from a CHD group of Glu151. The peak labelled ** is due to an impurity. (B) 15N-HN HSQC spectrum of CBDN1 recorded and processed in the same manner as the spectrum in (A).
Figure 5.12. Methyl region of a 1H-13C HSQC spectrum of 13C/15N and fractionally deuterated CBDN1 calicum-loaded and cellooligosaccharide-free.
Figure 5.13. Titration of calcium-loaded CBDN1 with cellopentaose monitored by 1H-13C HSQC spectra. Show is the region of the spectrum where the observed peaks are due to methyl groups. Only residues whose methyls experience a change in chemical shift are labelled.
Figure 5.14. Portion of the $^1$H-$^{13}$C correlation spectra recorded to measure the decay rates of $^{13}$C$_2$D$_2$ magnetization of methyls in calcium-bound and oligosaccharide-free CBDN1. Spectra are recorded with delay values of (A) 0.05 ms, (B) 8.6 ms, (C) 25.1 ms and (D) 50.0 ms.
Figure 5.15. Examples of the decay of $I_zC_zD_y$, $I_zC_zD_z$ and $I_zC_z$ magnetization. (top) $I_zC_zD_y$ magnetization decay is shown for L32$^{52}$ (cross), L14$^{52}$ (diamond), V72$^l$ (circle) and V36$^r$ (square). (middle) $I_zC_zD_z$ magnetization decay is shown for I54$^{51}$ (circle), V36$^r$ (diamond), L142$^{51}$ (cross) and V72$^l$ (square). (bottom) $I_zC_z$ magnetization decay is shown for A94 (diamond), L32$^{52}$ (circle), V72$^r$ (square) and A108 (cross).
cellopentaose-bound CBD\textsubscript{N1}. Data for 47 methyl groups were obtained in both the free and bound state. CBD\textsubscript{N1} contains 99 methyl groups. Data were not obtained for many of the methyl groups in the protein due to severe spectral overlap in the central part of the methyl region of the $^1$H-$^{13}$C HSQC spectrum, figure 5.12.

Figure 5.16 provides the $T_1(D)$ and $T_{1p}(D)$ values determined for both the free and cellopentaose-bound CBD\textsubscript{N1}. From the analysis of these data using equation 5.16, the order parameter $S_{2ax}$ was determined (figure 5.17). This provides information on the nanosecond-picosecond mobility of the methyl groups. As noted by Kay \textit{et al.} (1996) $S_{2ax}$ values display a much greater range of values than do $^{15}$N $S^2$ values. Excluding the methyls of the completely disordered C-terminal residue L152, values of $S_{2ax}$ range from a low of 0.209 for L32$^{\beta}$ in unbound CBD\textsubscript{N1} to a high of 0.911 for L95$^{\beta}$, also in the unbound form of CBD\textsubscript{N1}.

It can be expected that the length of the side-chain could affect $S_{2ax}$ values for different residue types (Kay \textit{et al.}, 1996). To correct for this positional dependency, average $S_{2ax}$ values were calculated for each methyl type (Ala C$^\beta$, Thr C$^\gamma$, Val C$^\gamma$, Ile C$^{\gamma 2}$, Ile C$^8$, and Leu C$^8$; CBD\textsubscript{N1} contains no methionine residues), and is denoted by $<S_{2ax}>$. The $S_{2ax}$ values for the methyls of the N and C-terminal residues A1 and L152 were not included in this calculation. Both the free and cellopentaose-bound values were included when calculating $<S_{2ax}>$. These data are presented in table 5.7. The difference between $S_{2ax}$ and the appropriate $<S_{2ax}>$ was then determined, and is shown in figure 5.17.

<table>
<thead>
<tr>
<th>methyl type</th>
<th>$&lt;S_{2ax}&gt;$</th>
<th>n$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala C$^\beta$</td>
<td>0.803</td>
<td>19</td>
</tr>
<tr>
<td>Thr C$^\gamma$</td>
<td>0.767</td>
<td>13</td>
</tr>
<tr>
<td>Val C$^\gamma$</td>
<td>0.603</td>
<td>16</td>
</tr>
<tr>
<td>Ile C$^{\gamma 2}$</td>
<td>0.743</td>
<td>6</td>
</tr>
<tr>
<td>Ile C$^8$</td>
<td>0.650</td>
<td>6</td>
</tr>
<tr>
<td>Leu C$^8$</td>
<td>0.546</td>
<td>33</td>
</tr>
</tbody>
</table>

$^a$ number of values used to calculate $<S_{2ax}>$. 

Table 5.7. Average $S_{2ax}$ values calculated for each methyl type found in CBD\textsubscript{N1}. 

Chapter 5-Dynamic Analysis of Ligand Binding

166
Figure 5.16. $T_1(D)$ (top) and $T_{1p}(D)$ (bottom) values for methyl groups in calcium-loaded CBD$_{N1}$ bound (closed circles) and not bound (open circles) to cellopentaose.
Figure 5.17. Order parameters squared, $S^{2}_{\text{axis}}$, describing the amplitudes of the motions of the methyl averaging axis for methyl groups in (A) calcium-loaded CBD\textsubscript{N1} and (C) the calcium-loaded and cellopentaose-bound CBD\textsubscript{N1} complex. Differences in $S^{2}_{\text{axis}}$ from the average of the square of the order parameters for each methyl type $\langle S^{2}_{\text{axis}} \rangle$, for methyls of (B) calcium-loaded CBD\textsubscript{N1} and (D) the calcium-loaded and cellopentaose-bound CBD\textsubscript{N1} complex. (E) Differences between $S^{2}_{\text{axis}}$ values of methyls in the cellopentaose-bound and free states of calcium-loaded CBD\textsubscript{N1}. Error bars are indicated for each value.
The patterns of $S^2_{\text{axis}} - <S^2_{\text{axis}}>$ values are very similar between the bound and free forms of CBD$_{N1}$. Aside from the N and C-terminal residues A1 and L152, the methyl groups that show the lowest $S^2_{\text{axis}} - <S^2_{\text{axis}}>$ values are L32 and L141. The side-chains of these residue occur across from each other in β-sheet B, and are directly adjacent to the disulphide bond connecting C33 and C140. This is very surprising considering these residues are completely buried in the hydrophobic core of the protein. The methyl groups of V34, which also lies next to the disulphide, is also relatively disordered, but not to as great an extent as is the case for L32 and L141. Other methyl groups that have low $S^2_{\text{axis}} - <S^2_{\text{axis}}>$ values are V17, L25, T70, V72 and A118. Residues corresponding to the most ordered methyl groups in CBD$_{N1}$, V36, L49, L62, V74, V88, V89, L95 and L129, are all buried in the core of the protein.

Discussion

Comments on the model-free analysis of $^{15}$N relaxation

A disproportionately high number of residues in CBD$_{N1}$ needed a two time scale spectral density model to fit the experimentally determined relaxation data (table 5.6). This is apparent from the fact that few residues, aside from the C-terminal residues S148 to L152, had low $T_1/T_2$ ratios (figure 5.6) indicative of motions on two time scales (Clore et al., 1990b). This frequent occurrence of residues fit by the two time scale model correlates with the strict error analysis for choosing models and the low error in the $T_1$ and $T_2$ values determined for CBD$_{N1}$. The average error in $T_1$ for samples 1 to 5 is 1.1%. For the $T_2$ values the average error is 1.0% and for the $^1$H-$^{15}$N NOE values it is 2.0%. This error range is much lower than reported in other studies (Farrow et al., 1994). When the data were fit with a minimum error in $T_1$ and $T_2$ of 3%, and a minimum error of 5% in the $^1$H-$^{15}$N NOE values the two time scale model was need much less often to fit the experimental data (table 5.8).
Table 5.8. Comparison of the spectral density models used to fit $T_1$, $T_2$ and $^{1}H^{-15}N$ NOE data for sample 5 with the original and introduced minimum errors.

<table>
<thead>
<tr>
<th>model</th>
<th>original errors</th>
<th>minimum errors$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ($S^2$)</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>2 ($S^2$, $\tau_e$)</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td>3 ($S^2$, $R_{ex}$)</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>4 ($S^2$, $\tau_e$, $R_{ex}$)</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>5 ($S^2$, $S_{fl}^2$, $\tau_s$)</td>
<td>29</td>
<td>7</td>
</tr>
<tr>
<td>none</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

$^a$ Minimum errors of 3% for $T_1$ and $T_2$ values and 5% in $^{1}H^{-15}N$ NOE values were used.

Although the different models used do change $S^2$ values, this does not change the trends observed in the data. Residues that have relatively low, or high, values of $S^2$ when originally fit still have relatively low, or high, values when increased error values are introduced into the data. Also, residues with $R_{ex}$ terms seem relatively unaffected by the introduction of minimum error values.

This dependence of model selection on the error range in the relaxation data does indicate that caution must be used in the interpretation of model-free analysis of relaxation data. In the discussion that follows this was kept in mind. The acquisition of data at two magnetic field strengths was especially useful in judging the reliability of trends in the data.

Global trends

As shown by the uniformity in $T_1$, $T_2$ and $^{1}H^{-15}N$ NOE data (figures 5.3 and 5.4) and the derived $S^2$ values (figures 5.6 and 5.7), the structure of CBD$_{N1}$ is well ordered throughout the protein. The four residues at the C-terminus of CBD$_{N1}$ are disordered, but no region within the protein is similarly unstructured. In native CenC these residues form the N-terminus of CBD$_{N2}$, thus the observed mobility may be due to the use of a protein fragment. In general, none of the residues with $T_1$, $T_2$ or $^{1}H^{-15}N$ NOE values that significantly differ from average lie in regions of secondary structure.
The only exceptions to this are the residues in \( \beta \)-strands B2 and B5 near the disulphide bond between C33 and C140. Conformational flexibility of the region near the disulphide is also reflected by low \( S^2_{\text{axis}} - <S^2_{\text{axis}} > \) values for the methyl groups of L32 and L141 in both bound and free \( \text{CBD}_{\text{N1}} \). It is therefore apparent that this region is motionally disordered in all samples, and that this disorder is not significantly affected by calcium or cellopentaose-binding. This motion might be related to conformational restrictions resulting from the disulphide linking the two \( \beta \)-strands together. Alternately, disulphide isomerization could lead to conformational flexibility, though this effect tends to occur on a slower time scale. Another possibility is that the sequence of \( \text{CBD}_{\text{N1}} \) used in all the structural and dynamic studies contains a mutation in this region. In native \( \text{CBD}_{\text{N1}} \), and all other family IV CBDs, residue 139 is a phenylalanine, whereas in \( \text{CBD}_{\text{N1}} \) used in this study, it is a leucine. This mutation was introduced during the initial subcloning of pTugN1n from pTZ18R-JC2. At first it was thought that this amino acid change was a result of incorrect sequencing of the original \( \text{CBD}_{\text{N1}} \) construct. However, upon sequencing \( \text{CBD}_{\text{N1N2}} \), it was revealed that the position corresponding to L139 in \( \text{CBD}_{\text{N1}} \) is a phenylalanine in \( \text{CBD}_{\text{N1N2}} \). It is therefore difficult to attribute the disorder observed in this region to any one effect when the native protein sequence in this region is not present. There might be subtle packing effects in this region of the core of the protein that are disrupted with the replacement of a phenylalanine by a leucine.

It is useful to compare the backbone rms deviation of the ensemble of \( \text{CBD}_{\text{N1}} \) structures with the relaxation data. The best parameter to compare the backbone rmsd to is the steady-state \( ^1\text{H}-^{15}\text{N} \) NOE measurement. Smaller, or less positive, \( ^1\text{H}-^{15}\text{N} \) NOE values are indicative of internal motion on the nanosecond time scale (Kay et al, 1989), as may be expected in a disordered region of the structure. Figure 5.18 shows a plot of the backbone rms deviation and the \( ^1\text{H}-^{15}\text{N} \) NOE values observed in sample 1 to 5 of \( \text{CBD}_{\text{N1}} \). This figure reveals that, aside from the N- and C-termini, the region with the largest rmsd, the loop area around residue 120, corresponds with a protein sequence of relatively low \( ^1\text{H}-^{15}\text{N} \) NOE. This indicates that the disorder observed in this region, and at the termini, in structure calculation results from true mobility in the protein.
**Figure 5.18.** Comparison between the rms deviation (rmsd) for the ensemble of 25 structures, aligned against the average structure (from figure 3.14), and the heteronuclear NOE values for CBD<sub>N1</sub> samples 1 to 5. High rmsd values indicate regions of the ensemble of structures with high variability. The lower the heteronuclear NOE values are, the more conformationally flexible the respective <sup>1</sup>H-<sup>15</sup>N bond vectors are on the nanosecond-picosecond time scale.
The loop region near residue 83, which also exhibits a high rmsd, does not show low $^1$H-$^{15}$N NOE values. The high rmsd here probably arises from it lying next to the disordered loop region near residue 120. As this hairpin loop lies at the edge of $\beta$-sheet A, the only cross-strand NOEs to structurally define this region are to the disordered region near residue 120. The high rmsd here thus reflects its proximity to a disordered region and a lack of NOEs to ordered regions of the protein, and not motion on the nanosecond time scale. The slightly low $^1$H-$^{15}$N NOE values found in the loop region around residue 100 are not reflected in a correspondingly high rmsd. This is probably due to this area being a short loop in comparison to that near residue 120. The well ordered $\beta$-strands at both ends of the residue 100 loop most likely fix the position of this region.

A general trend observed in this chapter was non-average relaxation parameters in the N-terminal region ($T_1/T_2$ ratios, $^1$H-$^{15}$N NOE and $R_{ex}$ values). This reflects disorder on both the millisecond and nanosecond time scale in this region and likely is the cause of the high rmsd values observed for residues located at the N-terminal.

**Effect of calcium binding**

It was previously shown (chapter 4) that CBD$_{N1}$ tightly bound calcium ions at a site lying on the face of CBD$_{N1}$ not involved in oligosaccharide binding. The putative atoms that ligate calcium are the backbone carbonyls T8, G30, and D142 and a side-chain oxygen of D142. The role of calcium is to stabilise the folded structure of CBD$_{N1}$ by tying together these three disparate regions of the protein. It is within this calcium-binding region that the most significant differences between the relaxation data of CBD$_{N1}$ in the calcium-free and calcium-bound form (samples 1 and 2) are found (figures 5.5, 5.6 and 5.8).

Upon calcium-binding, there is a reduction in the $T_1/T_2$ ratio in the N-terminal section of CBD$_{N1}$ (figure 5.6). This is reflected in a decrease in the value of the $R_{ex}$ term in this region in the model-free analysis (figure 5.8). This indicates that motion in this region on the millisecond time scale is reduced with calcium-binding. This effect was qualitatively recognized earlier by the
presence of line-broadened amide resonances in the calcium-free form. These resonances sharpen concurrently with calcium-binding.

Changes in $R_{EX}$ must be considered with caution as different processes may contribute to this term. Besides mobility of the protein on the millisecond time scale, in principle, the $R_{EX}$ terms could arise from chemical exchange from small amounts of calcium-bound CBD$_{N1}$ being present in solution according to the equilibrium below.

$$\text{CBD}_{N1} + \text{Ca}^{2+} \rightleftharpoons \text{CBD}_{N1}\text{Ca}^{2+}$$

Following the argument of Akke et al., (1993) the rate of exchange ($r_{ex}$) for calcium binding is given by:

$$r_{ex} = k_{off} + k_{on} [\text{Ca}^{2+}] = k_{off} + \beta k_{off}$$

where $\beta = [\text{CBD}_{N1}\cdot \text{Ca}] / [\text{CBD}_{N1}]$; $[\text{Ca}^{2+}]$, $[\text{CBD}_{N1}\cdot \text{Ca}]$ and $[\text{CBD}_{N1}]$ are the concentrations of calcium, CBD$_{N1}$ bound to calcium and calcium-free CBD$_{N1}$, respectively. As the relaxation experiments on the calcium-free form of CBD$_{N1}$ were collected in the presence of EDTA, whose affinity for calcium is much greater than that of CBD$_{N1}$, it is expected that $\beta$ is very low. This results in $r_{ex}$ effectively being equal to $k_{off}$. From chapter 4, $k_{off}$ is estimated to be 450 s$^{-1}$. For this rate to influence the $R_{EX}$ term, it must be comparable to the inverse of the time between the refocusing pulse and the formation of the spin-echo during the CPMG sequence used in the pulse sequence to measure $T_2$. This has a value of 2000 s$^{-1}$ in this study. As $r_{ex}$ for calcium binding is significantly lower than 2000 s$^{-1}$ chemical exchange between calcium-free CBD$_{N1}$ and a small amount of calcium-bound CBD$_{N1}$ will not contribute to the $R_{EX}$ terms. This strongly implies that the $R_{EX}$ terms observed for resonances from amides whose chemical shifts are perturbed by calcium-binding in sample 1, are due to exchange between distinct conformations of the protein.

Calcium-binding has a less dramatic effect on the relaxation properties of the other parts of the calcium-binding site. There is a decrease in the $T_1/T_2$ ratio and a decrease in $R_{EX}$ values of residues L139 and L141 with calcium-binding, again indicating a reduction in motion on the millisecond time scale with binding. But the change observed in this region is not nearly as
dramatic as similar changes at the N-terminus. The region around G30 also shows little change in its relaxation properties with calcium binding.

The observation that the largest changes in motions occur at the N-terminal region is most likely due to this segment not being part of any secondary structure element. The two other regions that bind calcium are either a short loop that connects two β-strands (G30), or a β-bulge (D142). These regions, though also not part of a secondary structure element, both occur as short two-residue segments between secondary structure elements. Thus both are already relatively structurally fixed compared to the N-terminal region.

The fact that the calcium-binding site is relatively disordered in the apo state seems contradictory when the previously determined on rate for calcium association is considered (chapter 4). This on-rate ($k_{on}$) was found to be $(5 \pm 2) \times 10^7$ M$^{-1}$ s$^{-1}$, which is only two orders of magnitude from being diffusion controlled. This suggests that the calcium-binding site is a structured, preformed entity with the necessary ligands correctly orientated, ready to quickly bind a calcium ion.

Although the protein segment around the calcium-ligating residue T8 is mobile on the millisecond time scale, it is fairly well-ordered on the picosecond time scale. The $S^2$ values, representing motion on this time scale, in this region in the apo form are not significantly lower than in other regions of the protein indicating this region is not a random coil (figure 5.8). They are also not significantly altered upon calcium-binding. This suggests that protein motions on these two time scales are not linked. The rapid on-rate most likely reflects the ordered picosecond time scale motion, and is not affected by the disordered motion observed on the millisecond time scale. Additionally, the magnitude of the disorder among the native-like states necessary to produce the observed decreases in $R_{ex}$ and $T_1/T_2$ values is unknown, and may only be slight.

It is possible to compare these results with those from a similar $^{15}$N relaxation study of the calcium-binding protein calbindin D$_{9k}$, apo, cadmium and calcium bound (Kördel et al., 1992; Akke et al., 1993). Here, a large increase in $S^2$ values at binding site II upon either calcium or cadmium ligation was observed. This reflects a large decrease in mobility on the picosecond time
scale of the backbone amides in this region. A similar effect was not observed for CBD$_{N1}$ as the binding site is already well ordered on this time scale in the apo form. Binding site I of calbindin D$_{9k}$ is well ordered in the apo form, consequently there was little effect on the order parameters of amides in this region with calcium binding. Also found in for calbindin D$_{9k}$ was significant conformational exchange terms ($R_{ex}$) in the apo and (Cd$^{2+}$)$_1$ state, with no significant conformational exchange terms found for the fully calcium-loaded protein (Ca$^{2+}$)$_2$. This is similar to what was found for CBD$_{N1}$.

*Effect of cellooligosaccharide binding on the dynamics of CBD$_{N1}$*

*i) Backbone $^{15}$N relaxation data*

There are limited changes in the dynamics of CBD$_{N1}$ with cellopentaose-binding as measured by backbone $^{15}$N relaxation. The oligosaccharide-binding site of CBD$_{N1}$ encompasses one of the $\beta$-sheets of the protein. In the sugar-free form of CBD$_{N1}$, the individual strands of this $\beta$-sheet are generally well ordered. They remain similarly well ordered when bound to cellopentaose. This is reflected in the uniformity of the $T_1$, $T_2$ and heteronuclear NOE values (figure 5.4) of these $\beta$-strand residues. The few changes in $^{15}$N relaxation data that do occur with cellopentaose-binding are observed at the very ends of the $\beta$-strands at the start of the loop regions of CBD$_{N1}$. In particular the $T_1/T_2$ ratios for V17, G44, N81 and G130 drop significantly. These residues are either the first or last residues of $\beta$-strands forming the binding face of CBD$_{N1}$. This indicates a decrease in mobility of these residues on the millisecond time scale. The reduction in mobility is similarly shown by a decrease in $R_{ex}$ terms of these residues (figure 5.7).

A similar consideration of the possibility of chemical exchange and not protein motion producing $R_{ex}$ terms, as was done in the section on calcium-binding, is not possible as the off rate of cellooligosaccharide association with CBD$_{N1}$ is unknown. It is known that sugar-binding is in the fast exchange limit of the NMR time scale. This is often much greater than 2000 s$^{-1}$, the rate
shown in the previous section for chemical exchange, not motional process to result in the observation of $R_{ex}$ terms.

On the nanosecond-picosecond time scale, as reflected in heteronuclear NOE and $S^2$ values, there is little change in CBD$_{N1}$ with celloctaose-binding. This probably reflects the fact that this protein is already well ordered on this time scale in the sugar-free form.

There is a small, but consistent, drop in $\tau_m$ values with oligosaccharide binding (table 5.5). This drop was observed in three separate samples of CBD$_{N1}$ (samples 2 and 3; samples 3 and 4; samples 9, 10 and 11). A drop in $\tau_m$ was not observed between samples 6 and 7. This implies that the bound form of the protein tumbles slightly faster in solution than the unbound form. This occurs even though this complex is 829 Da, or 5%, heavier when bound to celloctaose. The ultracentrifuge results showing a slight increase in $S_{20,\omega}$ of CBD$_{N1}$ when celloctaose-bound supports this observation of faster tumbling, resulting from a slightly more compact molecule when celloctaose-bound.

Together, the observed decrease in motion on the millisecond time scale of some residues, and a general decrease in $\tau_m$ values upon oligosaccharide binding indicates CBD$_{N1}$ is a less dynamic protein when bound. Whether this dynamic behaviour is a necessary part of oligosaccharide recognition and binding, or simply a result of steric restrictions imposed on the mobility of residues nearby the bound sugar is hard to determine.

A number of previous backbone $^{15}$N relaxation studies have observed a similar reduction in protein mobility with ligand binding. Fesik and co-workers observed a decrease in mobility of several residues on the picosecond time scale upon the binding of a tyrosine-phosphorylated peptide by a phosphotyrosine binding domain (Olejniczak et al., 1997). Hodsdon and Cistola (1997) observed decreased mobility on both the picosecond and millisecond time scales at one end of the binding site of intestinal fatty acid-binding protein upon forming a complex with palmitate. As previously mentioned the picosecond mobility of one of the calcium binding sites of calbindin D$_{9k}$ was dramatically reduced upon ion binding (Kördel et al., 1992; Akke et al., 1993). Conversely, an increase in motion on a picosecond time scale has also been observed upon ligand
binding (Yu et al., 1996). Similar findings as presented here for CBD\textsubscript{N1} were obtained by Farrow et al. (1994), where picosecond time scale disorder was not decreased upon ligand binding, but in the model-free analysis the experimental relaxation data for the complexed protein was fit with fewer exchange terms.

It is clear from the different results shown by these studies that a general correlation between protein mobility and ligand-binding is not found. Instead, other protein-specific considerations such as binding affinity and ligand specificity should be taken into consideration. A protein such as CBD\textsubscript{N1} that is not highly specific to a single substrate (Tomme et al., 1996a) may need to be flexible in the unbound state in order to correctly orient the side-chains involved in binding so that productive interactions can take place with a variety of substrates. When this motion is restricted in the ligand-bound state, as seen by a drop in R\textsubscript{ex} values for a small subset of residues on the binding face, an entropic penalty occurs. Proteins that are highly specific to a single ligand may not have to be as flexible and could be fixed in space and not entail this entropic penalty.

\textit{ii) Methyl containing deuteron relaxation}

While backbone \textsuperscript{15}N relaxation techniques are useful in studying protein-ligand interactions, a more intimate view of the binding event is obtained by studying the relaxation properties of the side-chains directly involved. The method used here for studying the dynamics of side-chain groups was methyl containing \textsuperscript{2}H relaxation. The work presented in this chapter is only the third reported application of this technique, the second studying ligand binding.

CBD\textsubscript{N1} is well suited to this method as it contains a number of solvent-exposed methyl containing side-chains in its binding face (figure 5.19). As a result of spectral overlap in the \textsuperscript{1}H-\textsuperscript{13}C HSQC spectrum of CBD\textsubscript{N1} (figure 5.12), it was not possible to obtain data for every methyl group in the binding face. However, a representative selection of data from methyls across the binding face was obtained.
Figure 5.19. Structure of CBD$_{N1}$ with the methyl containing residues present in the binding face shown in green. This figure was made using the programmes Molscript (Kraulis, 1991) and Raster3D (Merritt & Murphy, 1994).
Analysis of these data is hampered by the absence of a structure of the protein-oligosaccharide complex. Though the exact position of each methyl in relation to the substrate is unknown, the structure of CBD$_{N1}$ was determined from data collected on the cellotetraose-bound form of the protein. Also, intermolecular NOEs between the protein and ligand were detected for the methyl resonances of V17$^{11}$, V48$^{11}$, V48$^{12}$, L77$^{82}$ and A126$^{13}$ indicating close proximity, less than 5 Å, of these resonances and the bound oligosaccharide.

In the unbound form, of the methyls present in the binding face, only that of V17$^{12}$ is significantly more disordered than methyls of similar type (figure 5.17). Both methyls of L77 have typical $S_{2ax}^{2axis}$ values. However, the estimation of the relative disorder of leucine methyls in CBD$_{N1}$ are heavily biased by the large disorder found in L32 and L141. It is fair to say that the methyls of L77 are significantly disordered compared to most leucine methyls, other than L32 and L141. The methyl groups of T87 and A83 are of typical mobility. Data is unobtainable for the methyls of residues V48 and L126.

Upon cellopentaose-binding there is little change in the $S_{2ax}^{2axis}$ values of any of the binding-face methyls. Instead, the greatest change in $S_{2ax}^{2axis}$ occurs for T58$^{12}$ and L95$^{81}$ which becomes more and less ordered, respectively, upon ligand binding. The reasons for the changes observed for these groups is not known.

Surprisingly, the methyl groups V17$^{12}$, L77$^{82}$ and L77$^{81}$ remain disordered in the ligand-bound form, even though they must be completely buried by cellopentaose. The environment in the bound state may be similar to that found in the core of the protein. The disorder of these methyls when bound to a ligand could be related to similar mobility of the bound oligosaccharide within the binding groove. "Frame-shifting" of bound ligands was suggested by the chemical shift perturbation data produced by the binding of cellooligosaccharides (chapter 2; Johnson et al., 1996a). Similar changes in chemical shifts were observed when CBD$_{N1}$ bound cellotriose, cellotetraose, cellopentaose and cellohexaose. This occurs even though cellotriose and cellotetraose are shorter than the length of the binding cleft. This indicates that either these ligands bind tightly in multiple orientations in the cleft, or are mobile when bound.
As mentioned earlier, in the previous study using this technique (Kay et al., 1996) the methyl groups in the phosphotyrosine binding pocket of a SH2 domain were found to have their motions restricted upon binding. Methyl groups in other regions of the binding site that make contacts with the ligand remained disordered in both the bound and free state. The region where the methyls remain disordered when in contact with the peptide corresponds to the region of the peptide with relaxed binding specificity.

Similarly, in the case of \(\text{CBD}_{N1}\) where none of the methyls located in the binding cleft are highly ordered, either in the bound or free state there is also a lack of binding specificity. As shown by Tomme et al. (1996b) \(\text{CBD}_{N1}\) binds a variety of oligosaccharides. These include either \(\beta-1,4\) linked or mixed \(\beta-1,4\) and \(\beta-1,3\) glucose polymers, as well as the chemically modified forms of cellulose hydroxyethylcellulose, hydroxypropylmethylcellulose and carboxymethyl-cellulose. \(\text{CBD}_{N1}\) also binds chitin, chitosan, and sephadex. Also, as will be shown in chapter 6, \(\text{CBD}_{N1}\) binds cellooligosaccharides in multiple, not a single, orientations.

Motion on the millisecond time scale of residues in the binding face of \(\text{CBD}_{N1}\) was previously suggested by the observation of broad and weak \(H^\delta\) and \(H^\varepsilon\) signals from Y19 and Y85 (Johnson et al., 1996a). This was also shown by the absence of, or very weak signals of, the asparagine and glutamine groups present in the binding face, and not those in other regions of the molecule. When variants of \(\text{CBD}_{N1}\) with either Y19 or Y85 changed to alanine a large drop in affinity was observed, as well as the presence of strong signals from all Asn and Gln groups indicating the absence of motion on the millisecond time scale (\textit{vide infra}; chapter 6).

Together, the results of \(\text{CBD}_{N1}\) and those of the SH2 domain suggest that conformational flexibility of proteins coincides with areas of low protein-ligand specificity. This flexibility might be required for the recognition of a wide variety of different ligands in a common binding site by these proteins. Also, binding-site flexibility may not necessarily be reduced upon ligand-binding.
Changes in $\tau_m$ with viscosity

In the course of this study it was noticed that CBD$_{N1}$ samples experienced an increase in overall correlation time ($\tau_m$) with increased protein concentration. It is expected that an increase in protein concentration would result in a coincident increase in the viscosity of the protein solution. It was formulated by Debye that $\tau_m$ relates to viscosity by:

$$\tau_m = \frac{4\pi a^3 \eta_s}{3kT}$$

where $a$ is the effective radius of the solute, $T$ is the temperature, $k$ is the Boltzmann constant, and $\eta_s$ is the microviscosity of the environment of the solute.

As is shown in figure 5.9 there is a good correlation between viscosity and protein concentration. This graph also shows that the solution containing cellopentaose, and thus the bound form of CBD$_{N1}$, is slightly more viscous than that of the unbound sample. The finding that values of $\tau_m$ generally decrease when bound to cellopentaose (table 5.5), and in solution with a large excess of unbound sugar, appears to conflict with this observed increase in viscosity. To achieve this drop in $\tau_m$ the radius of the protein must decrease when cellopentaose-bound. This observation correlates well with the observed decrease in disorder of several loop-residues on the millisecond time scale upon ligand binding.

If the relationship between viscosity and protein concentration is assumed to be linear up to 5.3 mM the viscosity of sample 8 (table 5.1; table 5.5) would be 1.002 cSt. This reflects a 30% and 20% increase in the calculated viscosity of samples 6 and 9. This compares well with the measured 24% and 20% increase in $\tau_m$ values for sample 8 compared to samples 6 and 9, respectively. This increase in $\tau_m$ values with protein concentration, plotted together with the increase in viscosity is shown in figure 5.9. This finding emphasizes the need to keep sample conditions as similar as possible when comparing the results of different relaxation experiments.
Chapter 6

Structural Analysis of Ligand Binding

Abstract

The binding of TEMPO-labelled cellotriose and cellotetraose by CBD$_{N1}$ was studied to provide information about the possible orientations of oligosaccharides when bound. These two ligands are bound by CBD$_{N1}$ slightly tighter than the unmodified sugars. The TEMPO moiety contains a free electron which increases the relaxation rates, $R_1$ and $R_2$, of nearby nuclei. For both ligands, backbone $^1$H$_N \ R_1$ and $R_2$ rates were measured with the TEMPO group oxidized and reduced. From the difference in $R_1$ and $R_2$ values, the overall correlation time ($\tau_c$) and distance between the free electron and $^1$H$_N$ nuclei were determined. It was found that these spin-labelled ligands are bound in multiple orientations by CBD$_{N1}$. One set of orientations has the free electron near residue alanine 18, while in the other family of orientations, the free electron is close to glycine 86. These residues lie at opposite sides of the binding cleft. The binding to cellotetraose, and cellopentaose by two CBD$_{N1}$ variants, where tyrosine 19 and tyrosine 85 are changed to alanines, was studied. In both mutant proteins, binding affinity for both oligosaccharides is greatly reduced. A binding model is presented in which the flexibility of side-chains involved in hydrogen bonding formation to oligosaccharides is required. This flexibility allows for the binding of the same ligand in different orientations.
Introduction

Background

The structure of CBDN1 was calculated (chapter 3) using data collected in the presence of a large excess of cellotetraose (~40 fold). At that time, it was not known that calcium was bound by CBDN1, and therefore an excess of cellotetraose, with contaminating metal ion, was required to detect all backbone amide resonances (chapters 2 and 4). Intermolecular NOEs between the protein and cellotetraose were observed (chapter 3). However, as a large excess of oligosaccharide was used, and the association of CBDN1 with cellotetraose is in fast exchange on the NMR chemical-shift time scale, the chemical shifts of the sugar are heavily weighted to the unbound values. Since cellotetraose is composed entirely of β-(1,4)-linked glucose sugars, its spectrum is highly degenerate and not easy to assign. Thus it was not possible to assign these NOEs to specific ligand-protein interactions, as required to determine the structure of the oligosaccharide-protein complex.

The aim of this chapter is answer the question of how CBDN1 interacts with its ligands. Specifically, does the protein bind oligosaccharides in one or multiple conformations? Is the bound sugar orientated in one or two directions within the cleft? These questions are addressed in this chapter using nitroxide spin-labelled cellooligosaccharides to define the orientation of the sugar in the binding cleft, and using mutagenesis to probe the roles of two tyrosine side-chains in binding.

Theory on the use of spin labels

Structural insight into how CBDN1 interacts with oligosaccharides was gained by studying the binding of spin-labelled cellooligosaccharides. In this study, the spin-label used was a paramagnetic nitroxide free radical in the form of a 2,2,6,6-tetramethylpyrrolidine-1-oxyl group (I) (TEMPO).
The effect of the unpaired electron of the free radical is to increase the relaxation rates of the resonances of nearby nuclei in CBD$_N$ in a distance-dependent manner. Resonances closest to the spin-label will have their relaxation rates altered the most. This effect was studied in two ways. First, qualitatively, resonances closest to the spin label will decrease in intensity and eventually disappear upon the binding of a spin-labelled molecule. This provides immediate information on the orientation(s) of ligand binding in terms of which nuclei are closest to the free electron. Second, by measuring the relaxation rates ($R_1$, $R_2$) of the backbone amide H$_N$ nuclei, distances between the unpaired electron and the proton can be determined quantitatively.

The nitroxide group contains an unpaired electron which, in general terms, interacts with protons via dipolar coupling. Due to the size of the electron magnetic dipolar moment, this interaction extends to 25 Å and provides long range distance information. In contrast, NOE interactions between pairs of protons are limited to distances of less than ~5 Å. The magnetic interaction of an unpaired electron and a proton is described by the modified Solomon-Bloembergen equations (Solomon & Bloembergen, 1956). The enhancement of the proton’s spin-lattice ($\Delta R_1$) and spin-spin ($\Delta R_2$) relaxation rates are given by the equations (after Kosen, 1989 and Gillespie & Shortle, 1997):

$$\Delta R_1 = \Delta \left( \frac{1}{T_1} \right) = \frac{2K}{r^6} \left( \frac{3\tau_c}{1 + \omega_H^2 \tau_c^2} \right)$$

$$\Delta R_2 = \Delta \left( \frac{1}{T_2} \right) = \frac{K}{r^6} \left( 4\tau_c + \frac{3\tau_c}{(1 + w_H^2 \tau_c^2)} \right)$$
where \( K \) is the constant \( 1.23 \times 10^{-32} \text{ cm}^6 \text{s}^{-2} \) for a nitroxide radical, \( r \) is the distance between the electron and proton, \( \tau_c \) is the correlation time for the electron-proton vector and \( \omega_H \) is the Larmor frequency of the proton. \( \Delta(R) \) and \( \Delta(R_2) \) are the difference between the spin-lattice and spin-spin relaxation rates measured with the spin-label reduced (diamagnetic) and oxidised (paramagnetic), respectively. These equations are based on the assumptions that the vector between the electron and proton is free to undergo isotropic rotational diffusion, and that the distance \( r \) is constant.

The correlation time \( \tau_c \) is the sum of the contributions from the relaxation of the electron and motions of the electron-proton vector:

\[
1 / \tau_c = 1 / \tau_S + 1 / \tau_R
\]  

(6.3)

\( \tau_S \) is the longitudinal relaxation time of the free electron and \( \tau_R \) is the effective rotational correlation time of the vector, which is dependent on the motional characteristics of the protein. Typically for nitroxide radicals, \( \tau_S \) is longer than \( 10^{-7} \) s. Since \( \tau_R \) is in the range of \( 10^{-8} \) to \( 10^{-9} \) s, \( \tau_c \) is essentially equal to \( \tau_R \).

With \( \tau_c \) in the range of \( 10^{-8} \) to \( 10^{-9} \) s (chapter 5) and the NMR experiments carried out at 500 MHz \( (\omega_H / 2\pi) \) the term \( \tau_c \omega_H \) is greater than one. This enables \( \tau_c \) to be estimated from the ratio of \( \Delta R_2 \) by \( \Delta R_1 \) by rearranging equations 6.1 and 6.2 to obtain:

\[
\tau_c = \left( 6 \frac{\Delta R_2}{\Delta R_1} - 7 \right)^{1/2} \frac{4\omega_H^2}{\tau_c}
\]  

(6.4)

With a value of \( \tau_c \) and a measurement of either \( \Delta R_2 \) or \( \Delta R_1 \), the distance between the proton and electron, \( r \), can be determined by rearranging equations 6.1 and 6.2.

\[
\text{r}^6 = \frac{2K}{\Delta R_1} \left( \frac{3\tau_c}{3\tau_c^2 + \omega_H^2\tau_c^2} \right)
\]

(6.5)

\[
\text{r}^6 = \frac{K}{\Delta R_2} \left( \frac{4\tau_c + 3\tau_c^2}{3\tau_c^2 + \omega_H^2\tau_c^2} \right)
\]

(6.6)
Experimental methods

Spin labelled cellobiose saccharides

The spin-labelled compounds (2) and (3) used in this study studied were cellotriose and cellotetraose that contain the nitroxide group (1) attached at carbon 1 of the glucose located at the reducing end of the sugar chain. The nitroxide moiety (1) is approximately the same size and shape of a glucose ring. These compounds were synthesized by Lloyd McKenzie in the laboratory of Dr. Steve Withers, Department of Chemistry UBC, using the glycosynthase methodology developed in that laboratory. Full details of the synthesis and characterization of these compounds can be found in McKenzie (1997).

Preparation of \(\text{CBD}_{N1}\) samples and titration of nitroxide-labelled cellobiose saccharides

\(^{15}\text{N}\)-labelled \(\text{CBD}_{N1}\) used in titrations with nitroxide-labelled cellobiose saccharides was obtained by recycling the protein samples from the \(^{15}\text{N}\) relaxation experiments. As mentioned in chapter 5, oligosaccharide-bound \(\text{CBD}_{N1}\) can be recovered by unfolding the protein in 6 M urea in a microsep concentrator (Filtron). Upon centrifugation, the previously bound oligosaccharide passes through the membrane. The urea was removed by extensive exchange with a 50 mM sodium chloride, 50 mM sodium acetate (d3), 0.02% sodium azide, pH 6.1 buffer. Prior to the titration of the nitroxide-labelled cellobiose saccharides, the protein samples were saturated with
calcium by the addition of a 10 fold molar excess of calcium chloride dissolved in the same buffer as used with the protein.

For the nitroxide-labelled cellotriose (2), 2.8 mg was dissolved in 20 μL of the identical buffer as used for the protein, and added in 1, 3, 6 and 10 μL aliquots to the 0.32 mM protein sample. The final ligand-to-protein ratio was 22:1. In the case of the nitroxide-labelled cellotetraose (3), eight additions of a sugar solution were added to a 0.65 mM CBDN1 sample. The estimated final ligand-to-protein ratio was 2:1. Each point of these two titrations was monitored by the acquisition of a $^1\text{H}$-$^{15}\text{N}$ HSQC spectrum, collected with spectral widths of 6500 and 1450 Hz and with 1024 x 96 complex points in the $^1\text{H}$ and $^{15}\text{N}$ dimensions, respectively. All spectra were acquired on a Varian Unity 500 MHz spectrometer at 35 °C. Equilibrium association binding constants for these two ligands, based on their chemical shift perturbation upon ligand binding were obtained using the previously described methods (chapter 2 and chapter 4; Johnson et al., 1996a).

**NMR spectroscopy of nitroxide-labelled cellobiosaccharide-CBDN1 complexes**

Upon completion of the titration of CBDN1 with nitroxide-labelled cellobiosaccharides, a non-sensitivity enhanced $^1\text{H}$-$^{15}\text{N}$ HSQC spectrum to estimate $^1\text{H}$ N T$_2$ values was obtained as 1024 x 96 complex points with spectral widths of 6500 and 1450 in the $^1\text{H}$ and $^{15}\text{N}$ dimensions, respectively. A series of spectra to measure $^1\text{H}$ N T$_1$ values were then obtained using a sensitivity enhanced $^1\text{H}$-$^{15}\text{N}$ HSQC sequence (Kay et al., 1992) as a read-out of an inversion-recovery sequence (Carr & Purcell, 1954). For $^1\text{H}$ N T$_1$ measurements of the nitroxide-labelled cellotetraose, six spectra with delays of T= 0, 0.1, 0.2, 0.4, 0.8 and 2 s were recorded. The T=0.1 spectrum was measured twice in order to help estimate experimental error. For the nitroxide-labelled cellotriose, the $^1\text{H}$ N T$_1$ series had delay values of T = 0, 0.05, 0.1, 0.2, 0.4, 0.8 and 2 s. All the $^1\text{H}$ N T$_1$ spectra were acquired as 1024 x 96 complex points with spectral widths of 6500 and 1450 in the $^1\text{H}$ and $^{15}\text{N}$ dimensions, respectively. Selective water flip back pulses
were incorporated into both these $^1$H-$^1$5N HSQC sequences to ensure minimum perturbation of the water magnetization (Grzesiek & Bax, 1993; Zhang et al., 1994).

Upon completion of data collection with $\text{CBD}_{\text{N1}}$ bound to the paramagnetic cellobiose, the nitroxide functionality was reduced to the hydroxylamine form by the addition of two molar equivalents of solid L-ascorbic acid (Sigma). This reduction changes the nitroxide functionality on the bound spin-labelled cellobiose from a paramagnetic to a diamagnetic species. The pH* of the sample was adjusted back to 6.1 and a non-sensitivity enhanced $^1$H-$^1$5N HSQC spectrum and a $^1$H $T_1$ series with the same parameters as used with the oxidized sample, were collected.

Data analysis and calculation of $\Delta(R_1)$ and $\Delta(R_2)$

All NMR spectra were analysed using a combination of Felix v2.3, Felix95 (Biosym Technologies) and NMRPipe (Delaglio et al., 1995). When monitoring the binding of spin-labelled compounds with $\text{CBD}_{\text{N1}}$ by the decrease in intensity of peaks in the $^1$H-$^1$5N HSQC, the only window function applied to data was line broadening. When processing spectra to be used for peak volume measurement the data were processed with Lorentzian-to-Gaussian apodization. Relative peak volumes, absolute peak volumes and half-height linewidths were obtained using the nlinLS functionality in NMRPipe.

$^1$HN $T_1$ values were obtained by using relative peak volumes and the delay values (T) as input for the programme Imquick (Dr. Neil Farrow, University of Toronto). Imquick fits the relative peak volumes to a function of the form $I(T) = I(0) \exp(-T/T_1)$, where $I(T)$ is the relative peak volume at time $T$ and $I(0)$ is the intensity at time $T=0$. Errors in the measured relaxation rates are estimated using Monte Carlo procedures as described in Farrow et al. (1994). Once the per-residue $T_1$ values were obtained the reciprocal of the difference between $T_1$ values collected with oxidised and reduced ligands ($\Delta R_1$) was determined.
Two methods were used to measure per-residue values of $\Delta R_2$ from the non-sensitivity enhanced HSQC spectra. In the first, peak volumes were used to determine $\Delta R_2$. The ratio of a peak volume in the oxidised ($V_{\text{ox}}$) versus the reduced ($V_{\text{red}}$) form is:

$$\frac{V_{\text{ox}}}{V_{\text{red}}} = \frac{e^{-t/T_{2\text{ox}}}}{e^{-t/T_{2\text{red}}}}$$

(6.7)

where $t$ is the total time during the HSQC pulse sequence when $\text{HN}$ magnetization is in the transverse plane. By measuring peak volumes, not heights, the effect of differential $T_2$ relaxation during the detection period can be neglected. For the pulse sequence used in this study, the value of $t$ is 10.1 ms. This equation can be rearranged to:

$$V_{\text{ox} \cdot e^{-t/T_{2\text{ox}}}} = e^{-tAR_2}$$

(6.8)

Further rearranging gives:

$$\Delta R_2 = \Delta \left( \frac{1}{T_2} \right) = \left( \frac{1}{t} \right) \ln \left( \frac{V_{\text{red}}}{V_{\text{ox}}} \right)$$

(6.9)

Values of $\Delta(1/T_2)$ were also obtained using the proton half-height linewidths of the peaks. Since

$$R_2 = \frac{1}{T_2} = \pi LW$$

(6.10)

where LW is the half-height linewidth. It is easy to show that $\Delta(1/T_2)$ can be obtained from:

$$\Delta R_2 = \Delta \left( \frac{1}{T_2} \right) = \pi (LW_{\text{ox}} - LW_{\text{red}})$$

(6.11)

where $LW_{\text{ox}}$ and $LW_{\text{red}}$ are the half-height linewidths of the oxidised and reduced peaks, respectively.

_Titration of mutant CBD$_{N1}$ protein samples monitored by NMR_

The binding of mutant CBD$_{N1}$ protein samples to cellotetraose and cellopentaose was studied by NMR spectroscopy. In collaboration with Jeff Kormos and Peter Tomme (Dept. Microbiology, UBC), two CBD$_{N1}$ variants were studied as part of a larger investigation of the role played by various residues in oligosaccharide-binding. The CBD$_{N1}$ samples studied each had one
of the two solvent-exposed tyrosine residues (Tyr19 and Tyr85) changed to alanine. These CBD$_{N1}$ variants are subsequently referred to as Y19A and Y85A. $^{15}$N-labelled mutant CBD$_{N1}$ samples were prepared by Jeff Kormos (Dept. Microbiology, UBC).

Association binding constants were determined from the titration of $^{15}$N-labelled protein samples monitored by $^1$H-$^{15}$N HSQC spectroscopy. For each of the proteins studied, a single sample was used. This sample was split in half, with cellotetraose added to one half and cellopentaose added to the other. This ensures that the protein samples studied with each of the two cellooligosaccharides are identical. The concentration of Y19A CBD$_{N1}$ was 0.49 mM. Four aliquots of cellotetraose and five of cellopentaose was added. For the Y85A sample, a 0.58 mM sample was used, with six additions of cellotetraose and seven of cellopentaose added. For both sample a buffer of 50 mm potassium phosphate pH* 7.0 was used. This buffer is identical to that used in other binding studies performed using these mutant proteins (Jeff Kormos, pers. comm.).

$^1$H-$^{15}$N HSQC spectra on CBD$_{N1}$ variants were collected at 30 °C using the enhanced-sensitivity pulsed field gradient method of Kay et al. (1992) and Muhandiram and Kay (1994). Selective water flip-back pules were incorporated to minimize the perturbation of the bulk water (Grzesiek & Bax, 1993; Zhang et al., 1994). Equilibrium association binding constants were determined from the chemical shift perturbation of resonances in $^1$H-$^{15}$N HSQC spectra collected at each titration point. The data were analysed and $K_a$ values for these protein variants were obtained in a similar fashion to that previously outlined (chapter 2; Johnson et al., 1996b).

Results

**Binding of the nitroxide-labelled cellooligosaccharides to CBD$_{N1}$**

Association binding constants for the two nitroxide-labelled ligands were obtained using the methods previously described (chapter 2) with one minor modification as follows. In the case of the cellotetraose titration, very poor fits of the experimental data points to the curve calculated on
the basis of the regressed $K_a$ value were obtained. Upon the introduction of a scaling factor for the ligand concentration, which, along with the binding constant ($K_a$) and bound chemical shift ($\delta_b$), was allowed to vary, good fits of the experimental data to the calculated curve were obtained. The average value for this scaling factor was 1.8, indicating that more oligosaccharide was added than first thought. This scaling factor was not needed for the nitroxide-labelled cellotriose as the weight of this ligand used to prepare the ligand solution appeared to be accurate.

The $K_a$ values determined for the nitroxide-labelled cellooligosaccharides are summarised in table 6.1. For the sake of comparison, also included are the values previously determined for the relevant non-nitroxide cellooligosaccharide.

<table>
<thead>
<tr>
<th>Ligand</th>
<th>$K_a$ (M$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nitroxide-labelled cellotriose (2)</td>
<td>690 ± 262</td>
</tr>
<tr>
<td>cellotriose&lt;sup&gt;b&lt;/sup&gt;</td>
<td>180 ± 60</td>
</tr>
<tr>
<td>nitroxide-labelled cellotetraose (3)</td>
<td>(5000)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>cellotetraose&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4200 ± 720</td>
</tr>
</tbody>
</table>

<sup>a</sup> Data collected at 35 °C and pH 6.1 in 50 mM sodium chloride, 50 mM sodium acetate ($d^3$), 0.02% sodium azide and 10% D$_2$O / 90% H$_2$O. The reported $K_a$ values are the average of those determined from $^1$H and $^{15}$N chemical shift perturbations of G7, G15, V34, G44, T65, Q80, N81, T87, G130. The error range is one standard deviation.

<sup>b</sup> Data presented here are those determined in chapter 2 (Johnson <i>et al.</i>, 1996a).

<sup>c</sup> Due to uncertainty in the concentration of the ligand solution only an estimate for $K_a$ is given here.

**CBD$_{N1}$ binds nitroxide-labelled cellooligosaccharides in two distinct orientations**

The immediately observable effect of CBD$_{N1}$ binding the spin-labelled oligosaccharides is the decrease in intensity of the amide resonances closest to the unpaired electron. This decrease in intensity is readily observable for a number of peaks in the $^1$H-$^{15}$N HSQC spectra collected during the titration of CBD$_{N1}$ with both the nitroxide-labelled cellotriose and cellotetraose. Figure 6.1
Figure 6.1. 1H-15N HSQC spectra of CBDN1 bound to TEMPO-labelled cellotetraose. With the spin-label oxidized the free electron attenuates the signals of nearby nuclei. When the spin label is reduced this effect on the relaxation rates is no longer present. Resonances from residues that have their signal reduced the most are labelled.
shows the $^{1}H-^{15}N$ HSQC spectra of CBD$_{N1}$ in the presence of oxidized and reduced nitroxide-labelled cellotetraose. It is clear that a number of resonances present in the reduced form are absent in the oxidised form. These missing resonances include V17, A18, G20, Q42$^e$, G82, Y85 and T87. In this chapter resonances are said to “disappear” when their intensities fall below an arbitrary, but very low contour level.

The resonances that disappear can be classified into two groups. One for residues that lie on β-strands A1 and A2 of the binding face and another for those on strands A3, A4 and the loop connecting these strands (Figure 6.2). That two regions, on opposite ends of the binding face and separated by over 22 Å, are the most effected by binding the nitroxide-labelled celooligosaccharides clearly indicates that these ligands are bound by CBD$_{N1}$ in two distinct orientations. Based on which resonances disappear, one orientation has the nitroxide located near A18 H$_N$, the other with the nitroxide flipped end-over-end near and lying G86 H$_N$.

Figure 6.2. Schematic view of CBD$_{N1}$ with residues located in the binding face labelled. The TEMPO-labelled celooligosaccharides bind across the strands that make up this β-sheet. Residues whose backbone H$_N$ disappear with TEMPO-labelled cellotetraose-binding are underlined.
As shown by the ultracentrifuge results in chapter 2 and 5 there is no evidence for intermolecular association of CBDN1 either free or cellopentaose-bound. This rules out the possibility of a mechanism where the protein binds the spin-labelled cellobiosaccharides in a single orientation, but associates with itself end-to-end which would give rise to the relaxation rates of residues at one of the ends of the molecule being affected indirectly.

During the titration of CBDN1 with the nitroxyde-labelled cellobiosaccharides, the resonances that disappear lose their intensities at different concentrations of added ligand. At partial saturation, when the contribution of the nitroxyde to relaxation is lower, only protons closest to the nitroxyde electron will disappear. As saturation increases, the effect of the spin label increases. Protons at a greater distance from the spin-label will now have their relaxation rates attenuated proportionally more. The stage at which the resonances disappear with nitroxyde-labelled cellotetraose (3) binding is shown in Table 6.2. For nitroxyde-labelled cellotriose (2) binding these data are shown in Table 6.3.

### Table 6.2. Disappearance of resonances in $^{1}$H-$^{15}$N HSQC spectra induced by TEMPO-labelled cellotetraose (3) binding.$^{a}$

<table>
<thead>
<tr>
<th>Ligand-to-protein ratio needed to cause disappearance of resonance</th>
<th>Resonance that disappears</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4 : 1</td>
<td>A18</td>
</tr>
<tr>
<td>0.88 : 1</td>
<td>V17, Y19, Q42, G86</td>
</tr>
<tr>
<td>1.15 : 1</td>
<td>G20</td>
</tr>
<tr>
<td>1.42 : 1</td>
<td>Y85</td>
</tr>
<tr>
<td>1.75 : 1</td>
<td>T87$^{b}$</td>
</tr>
</tbody>
</table>

$^{a}$ Data collected at 35 °C and pH 6.1 in 50 mM sodium chloride, 50 mM sodium acetate (d3), 0.02% sodium azide and 10% D$_2$O / 90% H$_2$O.

$^{b}$ indicates resonance disappears upon binding the nitroxyde-labelled cellotetraose but not with nitroxyde-labelled cellotriose binding.
Table 6.3. Disappearance of resonances in $^1\text{H}-^{15}\text{N}$ HSQC spectra induced by TEMPO-labelled cellotriose (2) binding $^a$

<table>
<thead>
<tr>
<th>Ligand-to-protein ratio needed to cause disappearance of resonance</th>
<th>Resonance that disappears</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.09 : 1</td>
<td>V17, A18, Y19, G20</td>
</tr>
<tr>
<td>4.35 : 1</td>
<td>V48$^b$, N50$^b$</td>
</tr>
<tr>
<td>10.89 : 1</td>
<td>T21$^b$, A41$^b$, Q42$^e$, Y85, G86</td>
</tr>
<tr>
<td>21.77 : 1</td>
<td>G51$^b$, G82$^b$</td>
</tr>
<tr>
<td>undetermined$^c$</td>
<td>Y43$^b$</td>
</tr>
</tbody>
</table>

$^a$ Data collected at 35 °C and pH 6.1 in 50 mM sodium chloride, 50 mM sodium acetate (d3), 0.02% sodium azide and 10% D$_2$O / 90% H$_2$O.

$^b$ indicates resonance disappears upon binding the nitroxide-labelled cellotriose but not with nitroxide-labelled cellotetraose binding.

$^c$ unable to determine addition after which resonance disappears due to spectral overlap.

For both TEMPO-labelled ligands, resonances on strand A1 disappear at lower ligand-to-protein ratios than resonances in residues on strand A4 (figure 6.2). This may indicate that at sub-saturating concentrations one set of orientations is populated first. However, without knowledge of the electron-proton distances, this could result from resonances on strand A1 being closer to the free electron than resonances on strand A4, with both orientations populated equally.

The binding of the cellotriose spin-label is very weak. This is reflected in the low binding constant of $690 \pm 262 \text{ M}^{-1}$ found for the association of this ligand with CBD$_{N1}$. As a result, the concentrations of spin-label used in this titration were much higher than for the cellotetraose spin-label, and a significant amount of non-specific binding is observed. Non-specific binding is exemplified by the disappearance of residues that lie away from the binding site, or lie next to residues that appear unaffected by the spin-label. Examples of this non-specific binding include the disappearance of W16 H$^{E1}$ after 3 additions of ligand, and the disappearance of L146 H$^{N}$ after the second addition. The amides of V144 and A145 do not disappear.
Measurement of $\Delta R_1$ and $\Delta R_2$ values

For both the nitroxide-labelled cellotriose and nitroxide-labelled cellotetraose complexes the differences in $^1$H$^N$ R$_1$ and R$_2$ values between oxidized and reduced spin label were determined. R$_2$ values were measured two ways, one using peak volumes and another using half-height $^1$H$^N$ line widths in non-sensitivity enhanced $^1$H-$^15$N HSQC spectra. The peak volume method monitors the effect of the nitroxide before data collection. The line width measure reflects the effect of the nitroxide during the observation period. The $^1$H$^N$ R$_1$ and R$_2$ data are shown in figures 6.3 and 6.4 for the interaction of CBD$_{N1}$ with nitroxide-labelled cellotriose and cellotetraose, respectively. This analysis was limited to resonances that did not disappear entirely with the protein bound to oxidised spin-label. Thus, data are not obtainable for the resonances most affected by binding of the spin-labelled oligosaccharides.

From figure 6.4 it is immediately apparent that three regions of the CBD$_{N1}$ structure have their relaxation rates affected most by the binding of the cellotetraose spin-label. These are residues G15 to T21 on strand A1, residues Y43 to N51 on strand A2 and residues Q80 to D90 encompassing parts of strands A3 and A4 and the loop region connecting these two strands. As an indication of which $\Delta R_1$, $\Delta R_2$ and $\Delta$W values are significant, average values for these parameters were calculated. The average value of $\Delta R_1$ was found to be 0.29 s$^{-1}$ with a standard deviation of 0.38 s$^{-1}$, for $\Delta R_2$ the average is 8.6 s$^{-1}$ and the standard deviation is 46 s$^{-1}$, for $\Delta$W the average is 8 Hz with a standard deviation of 32 Hz. Residues that show a change in any of these values one standard deviation greater than average, or whose resonance is decreased in intensity such that $\Delta R_1$, $\Delta R_2$ or $\Delta$W could not be measured are shown in red in figure 6.5.

In the case of the nitroxide-labelled cellotriose (figure 6.3), four general areas are most affected by the binding of the spin label. These areas are the same as for the cellotetraose spin label with the addition of residues I125 to F127 on strand A5. Also, the region from residues Y43 to N51 on strand A2 seems to be more uniformly perturbed by this spin label than was the case for the cellotetraose spin label. The effects of non-specific binding are also evident, and are reflected by large values of either $\Delta R_2$ or $\Delta$W for residues V34, A35, R101 and D143 to V145.
Figure 6.3. Values of $\Delta R_1$, $\Delta R_2$ and $\Delta LW$ resulting from the binding of TEMPO-labelled cellotriose by CBD$_{N1}$. Four distinct regions are affected the most: on strand A1, centered around residue A18; across the length of strand A2; on strands A3 and A4 and the loop that connects them, residues Q80 to D90; and on strand A5, centered around A126. Non-specific binding results in high values of these parameters for residues Q101 and T138 to CHO. The only explanation for all these regions being affected is the binding of the ligand in multiple orientations.
Figure 6.4. Values of $\Delta R_1$, $\Delta R_2$, and $\Delta LW$ resulting from the binding of TEMPO-labelled cellotetraose by CBD$_{N1}$. Three distinct regions are affected the most: on strand A1 centered around residue A18; across the length of strand A2; on strands A3 and A4 and the loop that connects them, residues Q80 to D90. The only explanation for all these regions being affected is the binding of the ligand in multiple orientations.
Figure 6.5. $C^\alpha$ worm diagram of CBD$_{N1}$ with residues that experience the largest values of $\Delta R_1$, $\Delta R_2$ or $\Delta LW$ with TEMPO-labelled cellotetraose binding coloured red. Also coloured are residues whose amide resonances disappear with binding of this spin-labelled ligand. This view looks straight down onto the binding face of the molecule, the $\beta$-strands that make upface are labelled. Selected residues are labelled in yellow.
Calculation of $\tau_c$ values and electron-proton distances

As described by equations in the introduction to this chapter, once values of $\Delta R_1$ and $\Delta R_2$ are known it is facile to determine per-residue $\tau_c$ and $r$ values. The results of the $\tau_c$ calculation for CBD$_{N1}$ with the binding of the nitroxide-labelled cellotetraose (3) are summarised in figure 6.6. Due to the effects of non-specific binding a similar analysis of the data for cellotriose spin-label (2) was not done. It was possible to calculate 75 individual $\tau_c$ values for CBD$_{N1}$, the values ranged from $1.8 \times 10^{-10}$ s to $2.3 \times 10^{-8}$ s, the mean being $2.5 \times 10^{-9}$ s with a standard deviation of $2.7 \times 10^{-9}$ s. This compares to a value of the rotational correlation time of $7.4 \times 10^{-9}$ s determined by $^{15}$N relaxation methods (Chapter 5).

Using this mean value of $\tau_c$, electron-proton distances were determined from equations 6.5 and 6.6. As the distances have an $r^6$ dependence on $\tau_c$ they are very insensitive to the differences in $\tau_c$ observed in figure 6.6. Distances were calculated using both the volume and line-width methods of determining $\Delta R_2$. These three calculations of the distance are shown in figure 6.7. These graphs show similar trends to those seen in figure 6.4. That is, the regions of the protein that exhibit the shortest distance to the nitroxide are found on strands A1, A3 and A4. This is not at all surprising as these distances are calculated using the $\Delta R_1$, $\Delta R_2$ and $\Delta LW$ values shown in figure 6.4.

The distance values calculated using $\Delta R_1$ values include the low values of 10.7 Å (V17), 11.7 Å (Y19), 11.4 Å (G20), 11.8 Å (N50), 11.5 Å (G86) and 11.4 Å (T87). From $\Delta R_2$ values determined using peak volumes include the low values 9.2 Å (V17), 9.5 Å (Y19), 9.7 Å (G20), 10.5 Å (Y43), 10.9 Å (V48), 11.6 Å (G82), 11.4 Å (A83), 9.1 Å (G86) and 9.4 Å (T87). Finally using $\Delta R_2$ values determined using half-height peak line-widths the shortest distances are 11.1 Å (V17), 10.6 Å (G20), 11.4 Å (V48), 10.0 Å (N50), 11.6 Å (Q80), 11.1 Å (A83), 10.6 Å (G86), 10.1 Å (T87) and 11.0 Å (D90). Based on an estimated error of 10% in the measurement of peak volumes and line-widths, the error in $r$ calculated on $\Delta R_2$ values is likely to be at least ± 1.6 Å. The errors for $r$ determined by $\Delta R_1$ values are expected to be similar, on the order of ± 2 Å.
Figure 6.6. Values of the correlation time for the electron-proton vector ($\tau_c$) calculated from values of $\Delta R_1$ and $\Delta R_2$ for TEMPO-labelled cellotetraose-binding.
Figure 6.7. Values of the electron-proton distance (r) calculated from values of (A) \( \Delta R_1 \), (B) \( \Delta R_2 \), and (C) \( \Delta L_W \). Errors in r are estimated to be at least 2Å.
Cellulosic binding of CBD<sub>N1</sub> variants

The <sup>1</sup>H-<sup>15</sup>N HSQC spectra of the CBD<sub>N1</sub> variants Y19A and Y85A (Figure 6.8) are similar to that of wild type CBD<sub>N1</sub> for a large majority of residues. Differences are noted for resonances from nuclei near the site of mutation. Although not assigned, the overall similarity indicates that there are no gross structural changes associated with either of these mutations, such as the unfolding of part of the protein structure.

Binding constants for the association of mutant CBD<sub>N1</sub> proteins with cellotetraose and cellopentaose were determined from the chemical shift perturbation of the amide H<sub>N</sub> and <sup>15</sup>N nuclei as monitored during the titration by <sup>1</sup>H-<sup>15</sup>N HSQC spectra (Figure 6.9). For each titration, 10 to 16 different residues, corresponding to K<sub>a</sub> values from 15-31 different H<sub>N</sub> and <sup>15</sup>N nuclei, were averaged used to determine an overall K<sub>a</sub> value for that ligand protein pair. As the spectra of the CBD<sub>N1</sub> mutants did not exactly correspond to that of wild-type CBD<sub>N1</sub>, the amide resonances used to determine the K<sub>a</sub> values were not assigned to specific residues in the protein.

Both these CBD<sub>N1</sub> variants bind cellotriose and cellotetraose with approximately equal affinity (Table 6.4). This binding though, is much weaker than binding by wild type CBD<sub>N1</sub> to these ligands. The K<sub>a</sub> values for the mutant CBD<sub>N1</sub> proteins range from 0.6% to 1.4% of the corresponding wild-type values.

<table>
<thead>
<tr>
<th>Table 6.4. Association constants (K&lt;sub&gt;a&lt;/sub&gt;) of CBD&lt;sub&gt;N1&lt;/sub&gt; variants for two cellooligosaccharides.&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cellotetraose (M&lt;sup&gt;-1&lt;/sup&gt;)</strong></td>
</tr>
<tr>
<td>Y19A</td>
</tr>
<tr>
<td>Y85A</td>
</tr>
<tr>
<td>wild type&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Data collected at 30 °C and pH 7 in 50 mM potassium phosphate and 10% D<sub>2</sub>O/90% H<sub>2</sub>O. The reported K<sub>a</sub> values are the average of those determined from <sup>1</sup>H and <sup>15</sup>N chemical shift perturbations of 10 to 16 peaks. The error range is one standard deviation.

<sup>b</sup>Data presented here is that determined in chapter 2 (Johnson et al., 1996a).
Figure 6.8 1H-15N HSQC spectra of wild-type and two CBDN1 variants. The resonances corresponding the sites of mutation are indicated in the wild-type spectrum. Side-chain NH2 groups are connected by solid lines. All the side-chain NH2 resonances are present in the two variants, in wild-type CBDN1 the NH2 resonances of side-chains present in the binding face are not observed.
Figure 6.9. Binding of cellopentaose by the CBDN1 variant Y85A. Shown is a portion of 8 overlayed 1H-15N HSQC spectra in the presence of increasing amounts of cellopentaose. The arrows indicate which way the 1H-15N peaks shift with added sugar. Present in these spectra are all the NH2 resonances. The NH2 resonances of residues which lie in the binding face (N50, N81, Q124 and Q128) are not observed in spectra of wild-type CBDN1. These resonances also exhibit the largest changes in chemical shift with ligand binding. The assignments given are tentative.
Discussion

**Binding affinity of the nitroxide-labelled cellobioisaccharides**

CBD\(_{\text{NI}}\) binds the nitroxide-labelled cellobioisaccharides slightly tighter than the corresponding cellobioisaccharides (table 6.1). This indicates that the incorporation of the nitroxide label does not adversely effect the binding of these ligands by CBD\(_{\text{NI}}\), but rather slightly enhances the binding affinity. This can be explained by favorable hydrophobic and van der Waals interactions between the TEMPO group and residues on the protein surface. The TEMPO group and a glucose ring both present a relatively similar shape, both having a hydrophobic area that can interact with the protein.

One way in which the nitroxide moiety (1) differs from a glucose ring is in its lack of hydroxyl groups. It is thought that the hydroxyls on the glucose rings form hydrogen bonds with hydrophilic residues on the protein, providing the primary driving force for sugar binding by CBD\(_{\text{NI}}\) (Tomme et al., 1996a). The lack of hydroxyls on (1) probably explains why the nitroxide-labelled cellotetraose does not interact with CBD\(_{\text{NI}}\) as tightly as cellopentaose. And similarly why the nitroxide-labelled cellotriose is not bound as tightly as cellotetraose by CBD\(_{\text{NI}}\). These possible explanations for the increase in affinity of the TEMPO-labelled oligosaccharides suggest that the TEMPO group lies within the cleft when bound.

**Measurement of the relaxation parameters \(\Delta R_1\) and \(\Delta R_2\) and calculation of \(\tau_c\) and \(r\) values**

From the measured difference in the relaxation rates of H\(_N\) nuclei of CBD\(_{\text{NI}}\) with the cellotetraose spin-label oxidised and reduced, per-residue \(\tau_c\) values were obtained (figure 6.6). The average value obtained is 2.5 ns with a standard deviation of 2.7 ns. Though lower, this value is similar to the value of 7.4 ns determined for the overall rotational correlation time as determined by \(^{15}\text{N}\) relaxation experiments (chapter 5). The reason why the value is lower than that previously determined is not known, but may stem from presence of multiple binding orientations or motion of the TEMPO group when bound.
Using the values of $\Delta R_1$, $\Delta R_2$ and $\tau_c$, distances between the free electron and $H^N$ nuclei were determined (figure 6.7). These distances echo the trend in the data of the presence of two binding orientations, with the $H^N$ resonances of A18 and G86 being the two nuclei closest to the free electron. Both these nuclei are approximately 9 Å to 10 Å from the free electron. This is probably an underestimate of the true distance. This results from there being at least two binding orientations of the spin labelled cellulooligosaccharide. Upon saturation, approximately half the molecules have the spin label near these $H^N$ nuclei so the full effect of the free electron is not felt by the nucleus. Also, one assumption of equations 6.1 and 6.2 is that the distance $r$ is fixed. For CBD$_{N1}$ binding the spin-labelled compounds in two orientations, obviously $r$ is not fixed. With the free and bound ligands being in fast exchange on the NMR time scale this results in the measured $T_1$ and $T_2$ values of the oxidized having significant contributions from the alternate orientation, as well as staggered binding modes of the same orientation.

Binding of the spin-labelled cellotriose results in attenuation of the signal from resonances in all the strands of the binding face of CBD$_{N1}$ (figure 6.3). This indicates that this ligand is bound in orientations with the TEMPO lying in the middle of the binding face near residues A126 and V48, in addition to orientations near A18 and G86. Figure 6.10 shows these possible orientations. One of, and possibly both, orientations Ib and IIa exist to give rise to the attenuation of the signals from A126 and V48. Orientations Ia and IIb might exist and affect the relaxation rates of the resonances of A18 and G86, though, depending on the mobility of the ligand while bound, orientations Ib and IIa could produce the same effect. The TEMPO-labelled cellotetraose could also bind in each of these orientations, but as a result of its increased length, attenuation of signals in the middle strands of the binding face is less pronounced (figure 6.4).

At the start of this study it was hoped the distances obtained would be able to be used to help determine a structure of the bound oligosaccharide. But with the discovery that CBD$_{N1}$ binds these spin-labelled ligands in at least two orientations an accurate structure calculation is not feasible. There is no single structure to calculate, and the restraints derived from this study are only to one atom of a long sugar chain. It is hoped that the synthesis of cellopentaose with
**Figure 6.10.** Possible orientations of TEMPO-labelled cellotriose in the binding cleft of CBD_{NI}. The binding face is drawn as a box with the relative positions of Y19 and Y85 indicated. The ligand is drawn as a narrow rectangle with the TEMPO group indicated by shading. The length of the binding face, cellotriose and TEMPO group are drawn to scale. Evidence for the orientations Ib and Iia exists, while the existence of Ia and Iib is also likely. In addition to these binding modes it is possible for each orientation to exist in orientations rotated 180° about the lengthwise axis (see figure 6.12).
selectively incorporated $^{13}$C-labelled glucose rings will allow structures of the different orientations of cellopentaose bound to CBD$_{N1}$ to be determined.

However, insights into the nature of the complex can be gained from a rough model of CBD$_{N1}$ with a cellopentaose molecule manually positioned in the binding cleft (figure 6.11). This model shows the cellopentaose as a straight chain, but it is likely twisted when bound allowing the rings of Tyr19 and Tyr85 to stack flat against the pyranose rings of the sugar. Distances between H$_N$ atoms on the protein, and nuclei on the ligand where the free electron on the TEMPO might be are approximately 6 Å. This indicates that electron-H$_N$ distances in figure 6.7 are about 3 Å too long. Figure 6.11 clearly shows the length of cellopentaose is very similar to the length of the binding face. This corresponds with binding affinity being maximal for cellopentaose, longer chains bind with the same affinity (chapter 2; Johnson et al., 1996a; Tomme et al., 1996a).

**Ligand binding by CBD$_{N1}$ mutants**

Previous work had demonstrated the involvement of the side-chains of the tyrosine residues Y85 and Y19 of CBD$_{N1}$ in ligand binding. Difference ultraviolet absorbance spectroscopy showed that the environments of one or more tyrosine residues are perturbed by the addition of cellooligosaccharides (Johnson et al., 1996a). Similarly, the chemical shifts of both the aromatic H$_{5}$ and H$_{6}$ protons of these tyrosines are perturbed by cellotetraose binding (figure 2.11; Johnson et al., 1996a), and both these residues show NOEs to cellotetraose from their aromatic protons (Chapter 3; Johnson et al., 1996b). Finally, mutation of either Y85 or Y19 to alanine drastically reduces the affinity of CBD$_{N1}$ to phosphoric-acid swollen Avicel (Jeff Kormos, personal communication). Carbohydrate binding sites in proteins often contain aromatic residues that closely interact with the ligand (Vyas, 1991). In particular, cellulose-CBD interactions are often mediated by the interaction of a number of aromatic residues with cellulose (Chapter 1).

The binding of cellotetraose and cellopentaose to the Y19A and Y85A CBD$_{N1}$ mutant proteins were initially studied to see if these cellooligosaccharides were bound in a single orientation by CBD$_{N1}$, and if they are immobilized while bound. In the structure of CBD$_{N1}$
Figure 6.11. Two views of a model of $\text{CBD}_{N1}$ bound to cellopentaose. These views show only one of the possible orientations in which $\text{CBD}_{N1}$ binds cellopentaose.
presented in chapter 3, cellotetraose is not long enough to interact simultaneously with both Y85 and Y19. It was thought that if cellotetraose and cellopentaose were bound in one orientation, and the ligand is immobile when bound, it would be expected that the mutation of the tyrosine that interacts with cellotetraose would result in a large decrease in affinity for that ligand. The mutation of the tyrosine that does not interact with cellotetraose would result in no change in the affinity of CBD\textsubscript{N1} for cellotetraose. Conversely, for cellopentaose, which is expected to interact simultaneously with both tyrosines, the removal of either tyrosine should reduce the affinity of CBD\textsubscript{N1} for this ligand, as observed. However, recent structure calculations of CBD\textsubscript{N1} complexed with cellotetraose (Dr. E. Brun, personal comm.) show that it is possible for a cellotetraose molecule to simultaneously interact with both Tyr19 and Tyr 85. This results from the binding cleft of CBD\textsubscript{N1} becoming narrower in the structure of the cellotetraose-complex. No XPLOR energy penalty is incurred by this narrowing as very few distance restraints are found in this region of the protein to hold it in any specific conformation.

The fact that the mutation of either tyrosine drastically reduces the binding for cellotetraose and cellopentaose (table 6.4) proves that both tyrosines are important for ligand binding. Cellotetraose must interact with both Y19 and Y85 while bound to maintain tight binding, in both mutant proteins binding affinity is 1-2\% of wild-type affinity. This large drop in binding strength is consistent with the structure of the cellotetraose-CBD\textsubscript{N1} complex. If cellotetraose could not simultaneously interact with both tyrosines, as suggested by the sugar-free structure, and is bound in two orientations, as shown by the TEMPO-labelled cellotetraose results, its affinity for the two mutant proteins should be half that of wild-type.

Changes in the line-width of binding face residues in the CBD\textsubscript{N1} mutants

As mentioned in Chapter 3, a peculiar feature noticed during the resonance assignment of CBD\textsubscript{N1} was that the side-chain NH\textsubscript{2} resonances of the glutamine and asparagine residues, whose side-chains point into the binding groove, are severely line-broadened. This results in a large decrease in intensity of these resonances, probably resulting from intermediate exchange on the
NMR time scale (millisecond) between different conformations. This motion of binding face residues is not affected by cellobioigosaccharide binding to any observable extent, as judged by similar line-broadening in both free and ligand-bound spectra.

As these hydrophilic residues probably closely interact with oligosaccharides, a special effort was made to assign them. Most of these resonances were detected and assigned but they appear only as very weak signals. This decrease in intensity of resonances present in the binding face is not limited to \( \text{NH}_2 \)-containing residues but was also observed for both the \( \text{H}^8 \) and \( \text{H}^e \) resonances for Y19 and Y85 (chapter 2). In chapter 5 it was also observed that the side-chain methyls of V17 and L77 are disordered on the picosecond time scale, both when bound to cellopentaose and unbound.

It was therefore very surprising to observe these resonances in the \( ^1\text{H}-^{15}\text{N} \) HSQC spectra of both Y85A and Y19A CBD\( _{NI} \) (figures 6.8 and 6.9). As would be expected by their proposed close association with oligosaccharides, most likely through hydrogen-bond formation, these residue show a large change in chemical shift upon cellobioigosaccharide binding (figure 6.9). Indeed, these resonances show the greatest change in chemical shift with ligand binding of any resonances in both mutant CBD\( _{NI} \) proteins studied.

The appearance of sharp, strong peaks for all the \( \text{NH}_2 \) groups in the binding face indicates a change in the motion that originally gave rise to the linebroadening. As no "extra" set of resonances was observed in the \( ^1\text{H}-^{15}\text{N} \) HSQC, these resonances in the two mutant proteins probably have moved into the fast exchange limit on the NMR time scale. It is possible however, that the motion has shifted to the slow exchange limit, with only one conformation populated. Since the motion in wild-type protein is altered by the mutation of either Y19 or Y85, the motion must arise from the interaction of a specific residue composition of the binding face.

Upon mutation of either tyrosine 85 or 19 to an alanine, the motion that gives rise to the linebroadening is eliminated. Though some of the binding face glutamines and asparagines are close to one of these tyrosines, it seems surprising that a single mutation can alter the appearance of resonances far removed from the mutation site. It is possible the motions of the binding face
residues are linked together by a network of hydrogen-bonds. The removal of the tyrosyl hydroxyl group would then disrupt any hydrogen bonds it normally forms. It therefore seems the elimination of one member of this hydrogen bonding network, Y19 or Y85, can disrupt the entire network resulting in the elimination of the line-broadening. In the absence of detailed structural information on these mutant proteins, it is not possible to suggest other reasons for the observed effect.

Similar line-broadening was observed for a region of the protein Ras(1-171)GMPPNP (Ito et al., 1997) and was termed regional polysterism. This polysterism in Ras(1-171)GMPPNP was significantly reduced upon the binding of its downstream target, the Ras-binding domain of c-Raf-1. As discussed in chapter 5, I introduced the idea that the residues on the binding-face of CBD\textsubscript{NI} are disordered to enable this protein to recognize and bind a variety of different substrates. Ito et al. (1997) suggests a similar reason for the polysterism found in Ras(1-171)GMPPNP as this protein also recognizes a variety of target groups.

Implications for the association of CBD\textsubscript{NI} with cellooligosaccharides

From the observed patterns of how the nitroxide-labelled cellooligosaccharides (2) and (3) attenuate the relaxation rates of the amide proton nuclei of CBD\textsubscript{NI}, it is clear that these ligands are bound in two orientations (figures 6.1, 6.3 and 6.4). The nitroxide group can lie either near $\beta$-strand A1 close to Ala18, or strand A4 near residue Gly86 (figures 6.2 and 6.5). Mobility of bound ligand is suggested in the data from the cellooligosaccharide-binding studies presented in chapter 2. All the ligands studied, ranging in length from cellotriose to cellobiohexaose, produced similar patterns of perturbation of amide chemical shift of the protein. This indicates that these ligands are either mobile while bound, or there is a common structural change occurring with the binding of each cellooligosaccharide. Of course it is also possible for cellotetraose to bind in two orientations and also be mobile while bound. Figure 6.12 shows four possible orientations of cellotetraose.
Figure 6.12. Four possible orientations of cellotetraose. The reducing end of the oligosaccharide is marked R1. Oxygens are coloured red. The (a) and (b) orientations are flipped 180° about the lengthwise axis. The abundance of oxygens in cellotetraose means this oligosaccharide offers similar hydrogen bonding possibilities in all of these orientations.
It is possible that the attachment of the TEMPO moiety (1) alters the cellobiose to prevent this ligand from being bound in a single orientation. This would be the case if an interaction between the protein and the anomeric hydroxyl, which is not present in (2) or (3), is especially important for CBD binding to its ligands. This seems unlikely as all the orientations shown in figure 6.12 have oxygens at the end of the sugar chains for CBD to hydrogen bond with. However in all these orientations the hydroxyls are equatorial, if the anomeric hydroxyl is needed to be in the α anomer to be bound, introduction of the TEMPO could effect binding. The anomeric configuration of the bound sugar is unknown.

Also, the attachment of (1) probably does not affect its binding orientation as the nitroxide-labelled sugars studied are bound by CBD slightly tighter than the unmodified cellobiose (table 6.1). This indicates the introduction of the nitroxide group (1) at the end of a cellobiose does not interfere with binding, probably reflecting the fact that (1) is approximately the same size and shape as a glucose ring. The methyl groups present on (1) should produce some steric hindrance to binding by CBD. Despite this, the ligands (2) and (3) are bound tighter than the unmodified cellobiose.

As hydrogen bond formation between the oligosaccharide and the protein is the driving force for binding (Tomme et al., 1996a), the location of the oxygen atoms in four orientations of cellobiose (figure 6.12) should be considered. It is clear that there are many hydrogen bonding possibilities in all four orientations. Also, superficially at least, there a strong resemblance between the four conformations, especially between the pairs 1a, 2b and 1b, 2a (figure 6.12).

It should be noted here that, in the absence of a detailed structure of CBD bound to a cellobiose, figure 6.12 is only a representation showing the orientation of a great many possible conformations. It is likely that when bound, the C5-C6 bond in the sugar can occur in any of the possible rotamers, the angle of the glycosidic bond may change, and there is no guarantee that the sugar adopts the chair conformation as shown in figure 6.12.

For a mechanism where the sugar is mobile in the cleft while bound, hydrogen bonds must be continuously broken and formed. Similarly, if oligosaccharides are bound by CBD in two
orientations, different hydrogen bonding combinations must be possible. The high density of
hydroxyl groups on cellotetraose would make this possible. From figure 6.12 it is clear that these
four orientations offer numerous, and similar, hydrogen bonding possibilities.

It is likely that not all the oxygens in cellotetraose are involved in hydrogen bonding at the
same time. The association of cellotetraose with CBD$_N$1 involves a $\Delta G^\circ$ of -4.9 kcal mol$^{-1}$ at 35
°C (Tomme et al., 1996a). There are 14 hydroxyl groups and 21 oxygen atoms in total present in
cellotetraose. If the sole contribution to binding is assumed to be hydrogen bonding involving
only the hydroxyls, each would contribute -0.35 kcal mol$^{-1}$. Inclusion of hydrogen bonds to the
ether oxygens would reduce this further. From thermodynamic experiments involving proteins
binding deoxy-sugars, it has been shown that the presence of some hydroxyls of a sugar are not
necessary for binding, or even hinder binding, while others are essential. For three hydroxyls,
binding of the respective mono-deoxy sugar by the lectin GS-IV is eliminated (Lemieux, 1996).
This indicates that the removal of a single hydroxyl group reduces binding by as much as -6 kcal
mol$^{-1}$. For those hydroxyls whose elimination results in a measurable, but not complete, decrease
in affinity it is estimated that the elimination of an individual hydroxyl involves a $\Delta(\Delta G^\circ)$ of -2 kcal
mol$^{-1}$ (Quiocio, 1993) to -0.5 kcal mol$^{-1}$ (Lemieux, 1996). Using these measurements CBD$_N$1
likely forms 2.5 to 10 hydrogen bonds to cellotetraose. It is worth noting that these differences in
free energy are not due directly to the formation of a single hydrogen bond. Instead, they are the
differences between a state where water molecules hydrogen bond to the free protein and ligand,
that are then replaced by hydrogen bonds between the protein and ligand.

As noted earlier (chapter 3, Johnson et al., 1996b) the binding face of CBD$_N$1 contains a
number of hydrophilic residues capable of hydrogen bond formation. The work of Jeff Kormos
(personal communication) has shown that the substitution of any of Y19, N50, R75, N81, Y85,
Q124 or Q128 to alanine dramatically reduces binding affinity. Together, these residues have a
potential to form significantly more than the expected number of hydrogen bonds.

For a ligand that is bound in two orientations within the binding groove of CBD$_N$1, this
excess capacity of hydrogen bond formation might be required to produce the different hydrogen
bonding combinations necessary for productive binding. Similarly, for a ligand whose length is less than that of the cleft, cellotriose and cellotetraose, the observed mobility of the ligand may be facilitated by the presence of many hydrogen bond forming residues, with only a fraction of them being necessary to form for the ligand to be bound.

It was seen earlier, the methyls of V17 and L77 are motionally disordered on the picosecond time scale (chapter 5). Disorder of binding face residues on the millisecond time scale was discussed earlier in this chapter. In both cases motional disorder of the side-chains is observed in both the free and oligosaccharide bound forms. It is possible that this motion is required for the constant reorientation of side-chains to correctly position the necessary groups to form hydrogen bonds to the ligands. This enables CBD$_{NI}$ to bind different ligands as well as different orientations of the same ligand.

As mentioned in chapter 1, CBD$_{Cex}$ is mobile on the surface of cellulose while it is bound (Creagh et al., 1996; Jervis et al., 1997). In order to facilitate this mobility, it is likely that CBD$_{Cex}$ also binds cellulose in multiple orientations. As CBD$_{Cex}$ binds cellohexaose weakly, this hypothesis could be tested with experiments analogous to those discussed in this chapter using a spin-labelled cellohexaose molecule.

The work presented in this chapter provides the first structural evidence for the manner in which a CBD binds its ligand. These results indicate that CBD$_{NI}$ does not bind cellulose in one specific orientation. This function is consistent with CBD$_{NI}$ acting as an anchor point for the catalytic domain in the native CenC cellulase. In contrast to CBD$_{NI}$, cellulases bind and degrade cellulose in a single orientation. A biological role of feeding strands of cellulose into the active site of the catalytic domain would require the CBD to bind cellulose strands in one orientation to supply the catalytic domain with correctly oriented substrate. It is possible that the strand of sugar CBD$_{NI}$ binds is fed directly into the catalytic domain, but this would mean that one of the binding orientations of the CBD is enzymatically unproductive, resulting in a less efficient enzyme than if it bound in a single orientation.
When discussing CBD_{N1} it is important to keep in mind that in the intact enzyme CenC, CBD_{N1} is found in tandem with CBD_{N2} as CBD_{N1N2}. When binding experiments with the spin-labelled cellooligosaccharides were performed on CBD_{N2} it was determined that this protein also binds ligands in multiple orientations (E. Brun, personal communication). It was also determined that CBD_{N1N2} does not bind soluble glucans in a cooperative manner (Tomme et al., 1996a). However, it remains to be established whether the affect of one of the domains on the other is to influence the binding orientation, or specificity. If it is found that the two domains are independent of each other it is hard to imagine an efficient mechanism where the CenC CBDs feed the catalytic domain with substrate. Instead I believe the function of CBD_{N1N2} is to act as an anchor point on cellulose for the enzyme. This would be especially true if the linker region between the CBDs and the catalytic domain is flexible, enabling it to rotate to be correctly orientated with strands of cellulose which it can then degrade.
## Appendix

Table A1. Assignment of the $^1$H, $^{13}$C, and $^{15}$N NMR spectra of CBD$_N$1 in the presence of cellotetraose (35 °C, pH $^*$ 5.90).

<table>
<thead>
<tr>
<th>residue</th>
<th>$^{15}$N (H$^N$)</th>
<th>C$\alpha$</th>
<th>$^{13}$C$\alpha$ (H$^{\alpha}$)</th>
<th>$^{13}$C$\beta$ (H$^{\beta}$)$^a$</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>118.07 (8.39)</td>
<td>172.89</td>
<td>52.13 (4.19)</td>
<td>19.56 (1.56)</td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>112.30 (8.12)</td>
<td>174.84</td>
<td>43.80 (3.90, 4.62)</td>
<td>29.48 (2.23)</td>
<td>γ 28.12 (2.15, 2.35) δ 51.65 (4.05, 4.34) γ$^2$ 18.78 (0.95) γ$^1$ 26.49 (1.19, 1.34) δ$^1$ 15.04 (0.91)</td>
</tr>
<tr>
<td>P3</td>
<td>119.00 (8.05)</td>
<td>174.83</td>
<td>60.52 (4.51)</td>
<td>34.92 (2.86, 2.97)$^a$</td>
<td>δ 131.5 (7.38) e 131.74 (6.86) ξ 128.44 (5.86)</td>
</tr>
<tr>
<td>I4</td>
<td>110.62 (7.70)</td>
<td>176.55</td>
<td>60.75 (4.47)</td>
<td>70.17 (4.47)</td>
<td>γ$^2$ (1.18)</td>
</tr>
<tr>
<td>T8</td>
<td>118.47 (7.83)</td>
<td>177.12</td>
<td>56.24 (4.43)</td>
<td>40.96 (2.70, 2.58)$^a$</td>
<td>γ 179.94</td>
</tr>
<tr>
<td>F9</td>
<td>119.10 (8.05)</td>
<td>174.83</td>
<td>60.52 (4.51)</td>
<td>34.92 (2.86, 2.97)$^a$</td>
<td>δ 131.5 (7.38) e 131.74 (6.86) ξ 128.44 (5.86)</td>
</tr>
<tr>
<td>D10</td>
<td>118.07 (8.12)</td>
<td>174.84</td>
<td>43.80 (3.90, 4.62)</td>
<td>29.48 (2.23)</td>
<td>γ 36.28 (2.46) δ 183.2</td>
</tr>
<tr>
<td>D11</td>
<td>118.76 (8.81)</td>
<td>175.76</td>
<td>53.03 (4.81)</td>
<td>41.08 (2.72, 2.65)$^a$</td>
<td>γ 180.84</td>
</tr>
<tr>
<td>G12</td>
<td>110.62 (7.70)</td>
<td>176.55</td>
<td>60.75 (4.47)</td>
<td>70.17 (4.47)</td>
<td>γ$^2$ (1.18)</td>
</tr>
<tr>
<td>P13</td>
<td>117.23 (8.75)</td>
<td>176.13</td>
<td>57.14 (3.90)</td>
<td>27.81 (2.61)</td>
<td>γ 36.25 (2.36, 2.58) δ 183.8</td>
</tr>
<tr>
<td>E14</td>
<td>106.97 (8.19)</td>
<td>175.05</td>
<td>45.52 (3.79, 4.09)</td>
<td>41.08 (2.72, 2.65)$^a$</td>
<td>γ 180.84</td>
</tr>
<tr>
<td>G15</td>
<td>125.76 (8.33)</td>
<td>176.03</td>
<td>59.53 (4.46)</td>
<td>29.12 (3.37, 2.96)$^a$</td>
<td>δ$^1$ 129.3 (8.02) e$^1$ 132.79 (11.04) ζ$^2$ 113.52 (7.22) ξ$^3$ 121.0 (6.95) e$^3$ 121.04 (7.39) η$^2$ 122.5 (6.98)</td>
</tr>
<tr>
<td>W16</td>
<td>116.62 (9.23)</td>
<td>173.61</td>
<td>59.97 (4.69)</td>
<td>36.13 (2.26)</td>
<td>γ$^1$ 21.71 (1.04) γ$^2$ 19.89 (0.97)</td>
</tr>
<tr>
<td>V17</td>
<td>119.18 (8.86)</td>
<td>174.36</td>
<td>56.07 (5.13)</td>
<td>41.38 (3.09, 2.95)$^a$</td>
<td>δ 183.7 (6.81) e 117.3 (6.70)</td>
</tr>
<tr>
<td>A18</td>
<td>124.30 (8.56)</td>
<td>176.42</td>
<td>49.87 (5.65)</td>
<td>23.74 (1.43)</td>
<td>γ$^1$ 21.71 (1.04) γ$^2$ 19.89 (0.97)</td>
</tr>
<tr>
<td>Y19</td>
<td>111.62 (9.01)</td>
<td>173.77</td>
<td>45.83 (3.82, 4.72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appendix</td>
<td>221</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T21 108.29 (7.07)</td>
<td>60.73 (4.09)</td>
<td>69.85 (4.51)</td>
<td>$\gamma^2$ 22.85 (0.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D22 123.07 (8.65)</td>
<td>53.09 (4.74)</td>
<td>40.04 (2.08, 2.71)*</td>
<td>$\gamma$ 179.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G23 113.60 (7.70)</td>
<td>44.04 (3.89, 4.26)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P24 177.71</td>
<td>62.37 (4.52)</td>
<td>32.85 (2.15, 2.33)</td>
<td>$\gamma 26.78 (1.93, 2.06) \delta 49.29 (3.59, 3.64)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L25 120.89 (8.22)</td>
<td>55.84 (4.16)</td>
<td>42.45 (1.47, 1.55)</td>
<td>$\gamma 26.31 (1.81) \delta^1 25.55 (0.86) \delta^2 23.64 (1.57)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D26 124.10 (8.42)</td>
<td>53.19 (5.03)</td>
<td>44.49 (2.62)</td>
<td>$\gamma 180.9$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T27 117.65 (8.86)</td>
<td>61.34 (5.51)</td>
<td>68.21 (4.69)</td>
<td>$\gamma^2 21.49 (0.82)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S28 123.57 (8.75)</td>
<td>62.84 (4.12)</td>
<td>(3.95)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T29 113.12 (8.46)</td>
<td>62.27 (4.46)</td>
<td>70.44 (4.46)</td>
<td>$\gamma^2 21.21 (1.18)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G30 112.01 (8.23)</td>
<td>45.30 (3.75, 4.28)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A31 124.90 (7.37)</td>
<td>49.93 (4.39)</td>
<td>21.7 (1.04)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L32 124.52 (8.25)</td>
<td>54.81 (3.86)</td>
<td>40.16 (-0.63, 0.68)*</td>
<td>$\gamma 27.2 (0.73) \delta^1 26.10 (0.27) \delta^2 25.30 (0.14)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C33 126.39 (8.60)</td>
<td>55.00 (6.13)</td>
<td>47.35 (2.96, 2.75)*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V34 116.72 (9.02)</td>
<td>58.88 (4.39)</td>
<td>35.87 (1.71)</td>
<td>$\gamma^1 22.36 (0.57)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A35 127.51 (8.28)</td>
<td>50.60 (4.42)</td>
<td>19.23 (1.34)</td>
<td>$\gamma^2 18.81 (0.41)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V36 129.50 (9.11)</td>
<td>59.54 (3.82)</td>
<td>32.00 (0.80)</td>
<td>$\gamma^1 20.79 (0.80)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P37 176.69</td>
<td>62.36 (4.43)</td>
<td>32.42 (1.91, 2.24)</td>
<td>$\gamma 27.29 (1.91) \delta 51.87 (3.55, 3.82)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A38 123.83 (8.55)</td>
<td>169.78</td>
<td>52.59 (3.91)</td>
<td>18.77 (1.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G39 109.67 (8.40)</td>
<td>175.05</td>
<td>46.79 (3.83)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S40 114.71 (7.43)</td>
<td>173.07</td>
<td>58.99 (4.06)</td>
<td>65.23 (4.12, 3.84)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A41 124.79 (8.65)</td>
<td>177.82</td>
<td>50.09 (4.44)</td>
<td>21.48 (1.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q42 124.10 (8.42)</td>
<td>176.51</td>
<td>57.31 (3.25)</td>
<td>27.87 (1.48, 1.35)* $\gamma 32.33 (0.59, 1.46) \delta 179.5 \epsilon^2 112.31 (6.41, 6.80)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y43 122.08 (9.41)</td>
<td>177.82</td>
<td>59.97 (4.07)</td>
<td>36.38 (3.27)</td>
<td>$\delta 132.6 (6.97)$</td>
<td></td>
</tr>
<tr>
<td>G44 110.34 (8.56)</td>
<td>172.94</td>
<td>49.47 (3.80, 4.30)</td>
<td>$\epsilon 118.3 (6.74)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V45 116.63 (7.11)</td>
<td>175.25</td>
<td>59.33 (4.54)</td>
<td>35.27 (1.58)</td>
<td>$\gamma^1 21.84 (0.77)$</td>
<td></td>
</tr>
<tr>
<td>G46 117.65 (7.83)</td>
<td>179.66</td>
<td>46.21 (3.68, 4.65)</td>
<td>$\gamma^2 21.77 (0.95)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appendix</td>
<td>222</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V47</td>
<td>120.81 (8.60)</td>
<td>175.26</td>
<td>61.48 (5.13)</td>
<td>35.09 (1.97)</td>
<td>( \gamma^1 22.23 (1.00) )</td>
</tr>
<tr>
<td>V48</td>
<td>122.63 (9.59)</td>
<td>173.26</td>
<td>59.28 (5.58)</td>
<td>36.96 (2.01)</td>
<td>( \gamma^1 22.70 (0.95) )</td>
</tr>
<tr>
<td>L49</td>
<td>128.41 (8.66)</td>
<td>175.02</td>
<td>52.95 (4.09)</td>
<td>41.12 (-1.51, 0.64)</td>
<td>( \gamma 26.61 (0.70) )</td>
</tr>
<tr>
<td>N50</td>
<td>124.84 (8.38)</td>
<td>175.88</td>
<td>52.92 (5.02)</td>
<td>40.27 (2.47, 2.75)*</td>
<td>( \gamma 173.46 )</td>
</tr>
<tr>
<td>G51</td>
<td>111.04 (8.29)</td>
<td>173.44</td>
<td>46.65 (4.03, 4.17)</td>
<td>( \gamma^1, \gamma^2 21.09 (0.62) )</td>
<td></td>
</tr>
<tr>
<td>V52</td>
<td>125.18 (7.54)</td>
<td>172.93</td>
<td>61.30 (3.76)</td>
<td>33.60 (1.45)</td>
<td>( \gamma^1 25.48 (0.95, 1.45) )</td>
</tr>
<tr>
<td>A53</td>
<td>128.04 (7.89)</td>
<td>177.22</td>
<td>51.32 (4.27)</td>
<td>19.57 (1.36)</td>
<td>( \gamma 35.21 (0.99, 1.99) )</td>
</tr>
<tr>
<td>T58</td>
<td>128.06 (8.84)</td>
<td>175.24</td>
<td>54.61 (4.60)</td>
<td>31.50 (1.96)</td>
<td>( \gamma 35.74 (2.23) )</td>
</tr>
<tr>
<td>E55</td>
<td>127.82 (8.18)</td>
<td>177.42</td>
<td>57.97 (3.40)</td>
<td>28.92 (1.64)</td>
<td>( \gamma 34.95 (1.99) )</td>
</tr>
<tr>
<td>T59</td>
<td>125.48 (9.09)</td>
<td>174.35</td>
<td>44.87 (4.39, 4.32)</td>
<td>( \gamma^2 21.21 (0.56) )</td>
<td></td>
</tr>
<tr>
<td>V57</td>
<td>122.00 (8.08)</td>
<td>172.39</td>
<td>63.38 (4.16)</td>
<td>69.41 (4.39)</td>
<td>( \gamma^2 21.66 (0.94) )</td>
</tr>
<tr>
<td>E56</td>
<td>128.06 (8.84)</td>
<td>174.96</td>
<td>56.58 (4.98)</td>
<td>43.68 (1.41, 2.47)</td>
<td>( \delta 133.0 (6.61) )</td>
</tr>
<tr>
<td>G57</td>
<td>115.48 (9.09)</td>
<td>174.11</td>
<td>61.86 (5.24)</td>
<td>71.55 (3.86)</td>
<td>( \gamma^2 22.23 (1.08) )</td>
</tr>
<tr>
<td>T61</td>
<td>116.08 (8.84)</td>
<td>174.09</td>
<td>53.79 (5.38)</td>
<td>44.76 (1.42, 2.11)*</td>
<td>( \gamma 28.41 (1.80) )</td>
</tr>
<tr>
<td>L62</td>
<td>132.71 (9.58)</td>
<td>175.05</td>
<td>53.94 (5.73)</td>
<td>34.56 (1.84)</td>
<td>( \gamma 28.18 (1.70) )</td>
</tr>
<tr>
<td>R63</td>
<td>127.76 (9.31)</td>
<td>174.56</td>
<td>58.03 (4.90)</td>
<td>40.13 (3.65, 2.71)*</td>
<td>( \delta 132.8 (6.63) )</td>
</tr>
<tr>
<td>T64</td>
<td>115.18 (8.18)</td>
<td>172.56</td>
<td>53.94 (5.73)</td>
<td>34.56 (1.84)</td>
<td>( \gamma 28.18 (1.70) )</td>
</tr>
<tr>
<td>A65</td>
<td>118.00 (8.86)</td>
<td>173.83</td>
<td>61.02 (5.37)</td>
<td>70.02 (3.86)</td>
<td>( \gamma^2 21.86 (1.02) )</td>
</tr>
<tr>
<td>A66</td>
<td>128.89 (8.54)</td>
<td>176.91</td>
<td>50.46 (5.63)</td>
<td>24.61 (1.47)</td>
<td>( \gamma^2 21.61 (0.97) )</td>
</tr>
<tr>
<td>T67</td>
<td>117.76 (8.73)</td>
<td>172.01</td>
<td>61.76 (4.52)</td>
<td>71.58 (3.78)</td>
<td>( \gamma^2 18.17 (0.53) )</td>
</tr>
<tr>
<td>A68</td>
<td>129.41 (8.44)</td>
<td>177.18</td>
<td>50.21 (5.75)</td>
<td>25.65 (1.46)</td>
<td>( \gamma 180.3 )</td>
</tr>
<tr>
<td>S69</td>
<td>113.12 (8.46)</td>
<td>173.71</td>
<td>59.74 (4.27)</td>
<td>63.74 (3.95)</td>
<td>( \gamma^2 20.44 (0.15) )</td>
</tr>
<tr>
<td>T70</td>
<td>118.3 (7.35)</td>
<td>170.06</td>
<td>59.44 (4.36)</td>
<td>70.22 (3.94)</td>
<td>( \gamma^1 17.80 (-0.45) )</td>
</tr>
<tr>
<td>D71</td>
<td>122.74 (7.40)</td>
<td>175.55</td>
<td>54.42 (4.97)</td>
<td>39.86 (2.55)</td>
<td>( \gamma^2 20.44 (0.15) )</td>
</tr>
<tr>
<td>V72</td>
<td>125.73 (7.82)</td>
<td>171.87</td>
<td>60.24 (4.16)</td>
<td>34.68 (1.17)</td>
<td>( \gamma^2 20.44 (0.15) )</td>
</tr>
<tr>
<td>Code</td>
<td>Value1</td>
<td>Value2</td>
<td>Value3</td>
<td>Value4</td>
<td>Value5</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>T73</td>
<td>125.36</td>
<td>173.50</td>
<td>61.46</td>
<td>4.92</td>
<td>70.12</td>
</tr>
<tr>
<td>V74</td>
<td>115.64</td>
<td>173.17</td>
<td>57.82</td>
<td>3.50</td>
<td>31.48</td>
</tr>
<tr>
<td>R75</td>
<td>121.90</td>
<td>174.01</td>
<td>54.99</td>
<td>5.47</td>
<td>34.31</td>
</tr>
<tr>
<td>A76</td>
<td>131.15</td>
<td>176.16</td>
<td>50.41</td>
<td>5.15</td>
<td>20.73</td>
</tr>
<tr>
<td>L77</td>
<td>125.20</td>
<td>174.48</td>
<td>53.47</td>
<td>4.96</td>
<td>45.80</td>
</tr>
<tr>
<td>V78</td>
<td>120.86</td>
<td>176.66</td>
<td>61.53</td>
<td>5.06</td>
<td>32.97</td>
</tr>
<tr>
<td>G79</td>
<td>116.70</td>
<td>171.66</td>
<td>47.40</td>
<td>4.10</td>
<td></td>
</tr>
<tr>
<td>Q80</td>
<td>117.22</td>
<td>177.93</td>
<td>55.61</td>
<td>3.43</td>
<td>31.78</td>
</tr>
<tr>
<td>N81</td>
<td>124.48</td>
<td>174.24</td>
<td>51.56</td>
<td>4.76</td>
<td>36.26</td>
</tr>
<tr>
<td>G82</td>
<td>110.35</td>
<td>170.64</td>
<td>43.91</td>
<td>3.28</td>
<td>3.79</td>
</tr>
<tr>
<td>A83</td>
<td>124.45</td>
<td>175.73</td>
<td>51.21</td>
<td>3.34</td>
<td>16.89</td>
</tr>
<tr>
<td>P84</td>
<td>175.73</td>
<td>64.08</td>
<td>4.39</td>
<td></td>
<td>32.95</td>
</tr>
<tr>
<td>Y85</td>
<td>124.62</td>
<td>176.09</td>
<td>60.36</td>
<td>4.02</td>
<td>34.94</td>
</tr>
<tr>
<td>G86</td>
<td>109.38</td>
<td>172.30</td>
<td>45.55</td>
<td>3.67</td>
<td>4.09</td>
</tr>
<tr>
<td>T87</td>
<td>107.21</td>
<td>175.28</td>
<td>60.55</td>
<td>5.28</td>
<td>72.88</td>
</tr>
<tr>
<td>V88</td>
<td>110.32</td>
<td>174.96</td>
<td>60.46</td>
<td>4.54</td>
<td>33.26</td>
</tr>
<tr>
<td>L89</td>
<td>124.53</td>
<td>174.41</td>
<td>55.06</td>
<td>4.10</td>
<td>44.02</td>
</tr>
<tr>
<td>D90</td>
<td>129.12</td>
<td>175.30</td>
<td>54.61</td>
<td>4.84</td>
<td>41.60</td>
</tr>
<tr>
<td>T91</td>
<td>116.69</td>
<td>173.91</td>
<td>59.87</td>
<td>4.88</td>
<td>73.89</td>
</tr>
<tr>
<td>S92</td>
<td>117.91</td>
<td>175.30</td>
<td>54.14</td>
<td>5.79</td>
<td>64.68</td>
</tr>
<tr>
<td>P93</td>
<td>175.25</td>
<td>62.82</td>
<td>4.30</td>
<td></td>
<td>32.63</td>
</tr>
<tr>
<td>A94</td>
<td>125.73</td>
<td>177.29</td>
<td>51.52</td>
<td>4.64</td>
<td>18.83</td>
</tr>
<tr>
<td>L95</td>
<td>126.35</td>
<td>176.31</td>
<td>53.36</td>
<td>4.75</td>
<td>45.24</td>
</tr>
<tr>
<td>T96</td>
<td>115.65</td>
<td>174.08</td>
<td>60.14</td>
<td>4.82</td>
<td>71.52</td>
</tr>
<tr>
<td>S97</td>
<td>115.89</td>
<td>175.09</td>
<td>59.46</td>
<td>4.79</td>
<td>63.25</td>
</tr>
</tbody>
</table>
Appendix

<table>
<thead>
<tr>
<th>Code</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>E98</td>
<td>124.46</td>
<td>53.13 (4.75)</td>
<td>30.74 (1.81, 1.96)*</td>
<td>γ 35.80 (2.22) δ 183.9</td>
</tr>
<tr>
<td>P99</td>
<td>176.90</td>
<td>63.36 (4.33)</td>
<td>31.63 (1.72, 1.94)</td>
<td>γ 27.50 (1.97, 2.09)</td>
</tr>
<tr>
<td>R100</td>
<td>125.51 (8.06)</td>
<td>55.05 (4.55)</td>
<td>33.31 (1.81, 1.99)*</td>
<td>δ 50.47 (3.74, 3.84)</td>
</tr>
<tr>
<td>Q101</td>
<td>127.12 (8.50)</td>
<td>55.51 (4.69)</td>
<td>29.00 (2.08, 1.89)*</td>
<td>γ 27.84 (1.51) δ 43.82</td>
</tr>
<tr>
<td>V102</td>
<td>132.90 (9.26)</td>
<td>62.41 (3.92)</td>
<td>32.91 (0.55)</td>
<td>γ 34.47 (2.27) δ 180.9</td>
</tr>
<tr>
<td>T103</td>
<td>122.41 (8.12)</td>
<td>60.87 (5.15)</td>
<td>71.24 (3.93)</td>
<td>γ 21.89 (0.84)</td>
</tr>
<tr>
<td>E104</td>
<td>127.02 (9.38)</td>
<td>54.53 (4.97)</td>
<td>33.29 (1.86, 2.19)*</td>
<td>γ 20.58 (0.63)</td>
</tr>
<tr>
<td>T105</td>
<td>123.29 (8.86)</td>
<td>61.68 (5.75)</td>
<td>70.85 (3.87)</td>
<td>γ 21.78 (1.13)</td>
</tr>
<tr>
<td>F106</td>
<td>125.20 (9.34)</td>
<td>55.82 (5.11)</td>
<td>41.54 (3.03, 3.25)*</td>
<td>δ (6.96) ε 128.2 (6.72)</td>
</tr>
<tr>
<td>T107</td>
<td>121.41 (9.14)</td>
<td>61.89 (4.81)</td>
<td>69.38 (3.90)</td>
<td>γ 21.48 (1.00)</td>
</tr>
<tr>
<td>A108</td>
<td>132.06 (8.83)</td>
<td>54.38 (4.34)</td>
<td>20.07 (1.86)</td>
<td>γ 35.80 (2.22) δ 183.9</td>
</tr>
<tr>
<td>S109</td>
<td>119.19 (8.14)</td>
<td>57.66 (4.54)</td>
<td>63.90 (4.30, 3.99)*</td>
<td>γ 22.06 (1.30)</td>
</tr>
<tr>
<td>A110</td>
<td>124.36 (7.18)</td>
<td>51.05 (4.28)</td>
<td>23.00 (1.06)</td>
<td>γ 131.54 (6.97)</td>
</tr>
<tr>
<td>T111</td>
<td>115.08 (7.80)</td>
<td>60.26 (4.74)</td>
<td>71.32 (3.90)</td>
<td>ε 116.8 (6.29)</td>
</tr>
<tr>
<td>Y112</td>
<td>129.18 (9.44)</td>
<td>50.81 (5.66)</td>
<td>39.50 (2.60, 3.22)*</td>
<td>γ 26.78 (2.39, 2.47) δ (4.10)</td>
</tr>
<tr>
<td>P113</td>
<td>174.93</td>
<td>61.72 (4.73)</td>
<td>33.22 (2.25, 2.37)</td>
<td>γ 26.78 (2.39, 2.47) δ (4.10)</td>
</tr>
<tr>
<td>A114</td>
<td>119.13 (8.41)</td>
<td>53.20 (3.85)</td>
<td>19.68 (1.41)</td>
<td>γ 21.26 (1.20)</td>
</tr>
<tr>
<td>T115</td>
<td>111.01 (7.2)</td>
<td>57.68 (4.71)</td>
<td>70.57 (4.06)</td>
<td>γ 27.05 (2.01, )</td>
</tr>
<tr>
<td>P116</td>
<td>176.63</td>
<td>63.27 (4.51)</td>
<td>32.49 (2.45)</td>
<td>δ 50.61 (3.53, 3.82)</td>
</tr>
<tr>
<td>A117</td>
<td>124.93 (8.18)</td>
<td>177.76</td>
<td>50.14 (4.60)</td>
<td>22.23 (1.88, 1.52)</td>
</tr>
<tr>
<td>A118</td>
<td>126.55 (8.56)</td>
<td>177.98</td>
<td>54.67 (3.93)</td>
<td>17.74 (1.34)</td>
</tr>
<tr>
<td>D119</td>
<td>117.59 (8.41)</td>
<td>175.30</td>
<td>54.32 (4.01)</td>
<td>40.33 (2.77)</td>
</tr>
<tr>
<td>D120</td>
<td>120.10 (7.59)</td>
<td>52.17 (5.07)</td>
<td>41.93 (2.81, 2.22)*</td>
<td>γ 181.3</td>
</tr>
<tr>
<td>P121</td>
<td>175.09</td>
<td>62.59 (4.18)</td>
<td>32.38 (1.22, 1.87)</td>
<td>γ 26.78 (1.45, 1.56)</td>
</tr>
<tr>
<td>E122</td>
<td>121.34 (9.30)</td>
<td>175.46</td>
<td>54.35 (4.38)</td>
<td>29.81 (1.26, 1.93)</td>
</tr>
<tr>
<td>G123</td>
<td>107.86 (7.69)</td>
<td>174.50</td>
<td>45.06 (4.06, 4.25)</td>
<td>γ 35.82 (2.18, 2.37) δ 182.0</td>
</tr>
<tr>
<td>Q124</td>
<td>120.68 (9.27)</td>
<td>174.24</td>
<td>53.57 (5.61)</td>
<td>32.52 (1.66)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>γ 33.20 (1.67, 2.11)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>δ 179.4 ε 2 113.66 (8.42)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>121.96 (8.20)</td>
<td>174.45</td>
<td>60.73 (4.64)</td>
<td>40.03 (1.57)</td>
</tr>
<tr>
<td>A126</td>
<td>127.97 (8.93)</td>
<td>174.62</td>
<td>51.06 (5.34)</td>
<td>22.56 (1.12)</td>
</tr>
<tr>
<td>F127</td>
<td>122.26 (9.13)</td>
<td>175.77</td>
<td>56.91 (5.02)</td>
<td>39.88 (3.12, 2.51)*</td>
</tr>
<tr>
<td>Q128</td>
<td>123.94 (9.39)</td>
<td>175.77</td>
<td>55.85 (4.82)</td>
<td>29.61 (2.42, 2.12)*</td>
</tr>
<tr>
<td>L129</td>
<td>119.56 (8.07)</td>
<td>177.14</td>
<td>54.57 (4.98)</td>
<td>43.70 (1.10, 1.99)</td>
</tr>
<tr>
<td>G130</td>
<td>108.65 (8.00)</td>
<td>174.63</td>
<td>44.45 (3.72, 4.13)</td>
<td></td>
</tr>
<tr>
<td>G131</td>
<td>111.26 (8.12)</td>
<td>173.89</td>
<td>45.88 (3.34, 3.52)</td>
<td></td>
</tr>
<tr>
<td>F132</td>
<td>120.32 (6.98)</td>
<td>175.45</td>
<td>59.66 (4.51)</td>
<td>40.87 (2.34, 3.45)*</td>
</tr>
<tr>
<td>S133</td>
<td>113.37 (7.80)</td>
<td>175.65</td>
<td>56.38 (4.57)</td>
<td>64.72 (3.45, 3.52)*</td>
</tr>
<tr>
<td>A134</td>
<td>134.02 (9.12)</td>
<td>178.72</td>
<td>54.59 (4.32)</td>
<td>18.83 (1.47)</td>
</tr>
<tr>
<td>D135</td>
<td>120.51 (7.99)</td>
<td>175.06</td>
<td>53.71 (4.75)</td>
<td>42.71 (2.46, 2.87)*</td>
</tr>
<tr>
<td>A136</td>
<td>122.68 (8.25)</td>
<td>178.05</td>
<td>52.07 (4.58)</td>
<td>19.81 (1.48)</td>
</tr>
<tr>
<td>W137</td>
<td>120.37 (8.19)</td>
<td>172.86</td>
<td>56.34 (4.63)</td>
<td>29.55 (3.23)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T138</td>
<td>115.16 (9.07)</td>
<td>173.12</td>
<td>61.21 (4.93)</td>
<td>72.40 (3.88)</td>
</tr>
<tr>
<td>L139</td>
<td>129.72 (8.37)</td>
<td>175.61</td>
<td>53.25 (5.19)</td>
<td>44.63 (1.34, 1.59)*</td>
</tr>
<tr>
<td>C140</td>
<td>123.07 (8.72)</td>
<td>173.15</td>
<td>55.68 (5.63)</td>
<td>47.53 (2.72, 2.41)*</td>
</tr>
<tr>
<td>L141</td>
<td>127.07 (9.06)</td>
<td>174.36</td>
<td>52.74 (5.05)</td>
<td>47.06 (0.90, 1.03)</td>
</tr>
<tr>
<td>D142</td>
<td>120.31 (9.01)</td>
<td>173.37</td>
<td>51.43 (5.18)</td>
<td>44.56 (2.39, 2.51)*</td>
</tr>
<tr>
<td>D143</td>
<td>124.06 (9.03)</td>
<td>176.23</td>
<td>55.72 (4.28)</td>
<td>39.66 (2.60)</td>
</tr>
<tr>
<td>V144</td>
<td>120.40 (8.53)</td>
<td>176.62</td>
<td>60.77 (5.28)</td>
<td>33.40 (2.06)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A145</td>
<td>129.71 (9.51)</td>
<td>174.35</td>
<td>52.60 (4.38)</td>
<td>23.54 (1.44)</td>
</tr>
<tr>
<td>L146</td>
<td>122.49 (7.61)</td>
<td>174.71</td>
<td>54.44 (5.41)</td>
<td>44.49 (1.25, 2.02)*</td>
</tr>
<tr>
<td>D147</td>
<td>128.06 (8.84)</td>
<td>175.58</td>
<td>53.27 (5.19)</td>
<td>44.62 (2.54, 2.79)*</td>
</tr>
<tr>
<td>S148</td>
<td>116.39 (8.66)</td>
<td>174.35</td>
<td>56.94 (4.65)</td>
<td>65.30 (3.57)</td>
</tr>
</tbody>
</table>

* Denotes significance at the 0.05 level.
<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>E149</td>
<td>123.65 (8.20)</td>
<td>176.05</td>
<td>55.97 (4.28)</td>
<td>30.02 (1.78, 1.96)*</td>
<td>γ 35.53 (2.13) δ 183.2</td>
</tr>
<tr>
<td>V150</td>
<td>122.76 (8.03)</td>
<td>175.71</td>
<td>62.08 (4.08)</td>
<td>33.13 (2.01)</td>
<td>γ¹ 21.26 (0.88)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>γ² 21.26 (0.88)</td>
</tr>
<tr>
<td>E151</td>
<td>126.98 (8.32)</td>
<td>175.20</td>
<td>56.28 (4.29)</td>
<td>30.12 (1.90, 2.04)*</td>
<td>γ 35.99 (2.26) δ 180.8</td>
</tr>
<tr>
<td>L152</td>
<td>131.64 (7.83)</td>
<td>56.75 (4.17)</td>
<td>43.55 (1.53)</td>
<td>27.33 (1.53) δ (0.92)</td>
<td></td>
</tr>
</tbody>
</table>

*asterisk denotes stereospecifically assigned β protons, reported as H²β2 and H³β3, respectively. Otherwise all other methylene protons are listed by ascending chemical shift. Leucine and valine residues for which the methyls are not stereospecifically assigned are reported as δ and γ, respectively.
Bibliography


Bibliography


Hahn, M., Planas, A., Querol, E. & Heinemann, U. (1995c) Crystal structure of Bacillus licheniformis 1,3,1,4-β-D-glucan 4-glucanohydrolase at 1.8 Å resolution. FEBS Letters 374, 221-224.


Bibliography


Yamazaki, T., Foreman-Kay, J.D. & Kay, L. (1993) Two-dimensional NMR experiments for correlating \textsuperscript{13}C\textbeta\ and \textsuperscript{1}H\&\textgreek{e} chemical shifts of aromatic residues in \textsuperscript{13}C-labelled proteins via scalar couplings. *J. Am. Chem. Soc.* **115**, 11054-11055.

Yang, D. & Kay, L.E. (1996) The effects of cross correlation and cross relaxation on the measurement of deuterium T\textsubscript{1} and T\textsubscript{1p} relaxation times in \textsuperscript{13}CH\textsubscript{2}D spin systems. *J. Mag. Res. Ser. B.* **110**, 213-218.


