CRYSTAL STRUCTURE AND CATALYTIC MECHANISM
OF THE CATALYTIC DOMAIN OF THE \( \beta-1,4 \)-GLYCANASE CEX

by

André White

A thesis submitted in conformity
with the requirements for the degree of Doctor of Philosophy
Graduate Department of Medical Biophysics, University of Toronto

© Copyright by André White 1997
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.
Crystal Structure and Catalytic Mechanism of the Catalytic Domain of the β-1,4-Glycanase Cex.
Doctor of Philosophy, André White, department of Medical Biophysics, University of Toronto.

Cellulose-degrading microorganisms secrete a number of glycosyl hydrolases to degrade plant cell walls, the most abundant and renewable source of biomass. The bifunctional xylanase/cellulase (β-glycanase), Cex, secreted by the bacterium Cellulomonas fimii is an example of glycosyl hydrolases (family 10). This enzyme releases a disaccharide unit following the hydrolysis of its polysaccharide substrate. By analogy with lysozyme, the catalytic mechanism would proceed through an oxocarbonium ion-like species as the catalytic intermediate. However, recent evidence supports the existence of a covalent intermediate in the Cex catalytic mechanism; for example, cleavage of 2",4"-dinitrophenyl-2-deoxy-2-fluorocellobioside (2FDNPC) by Cex produces a covalent 2-fluorocellobiosyl-enzyme complex. A covalent 2-deoxy-glycosyl-enzyme intermediate has also been speculated to occur in the hydration of glycals by glycosyl hydrolases. In the present work, structural details of the catalytic domain of Cex (cex-cd), unliganded and with covalently complexed disaccharides, give insights into its folding motif, its active site structure and its mechanism of catalysis.

This thesis reports the high resolution (1.8 Å) crystal structure of cex-cd which is determined by the classical method of isomorphous replacement using heavy atoms. The main architecture of cex-cd consists of an elliptic (β/α)-barrel with eight β-strands forming the core. The active site is almost exclusively composed of conserved residues, suggesting a similar substrate recognition and catalytic mechanism among enzymes of family 10 of glycosyl hydrolases.

The structure of cex-cd determined in the presence of 2FDNPC reveals a fluorocellobioside molecule at the carboxy-terminal end of the (β/α)-barrel, which delimits the active site of the enzyme. This ligand is attached to the nucleophile Glu233 by a covalent link of α-configuration, which is consistent with a double displacement catalytic mechanism. In addition, a short exposure of cex-cd to the glycal cellobial leads to a 2-deoxy-cellobiosyl-enzyme covalent complex. From the structure of these complexes, seven structural characteristics of the covalent glycosyl-enzyme catalytic intermediate are identified. These studies help to understand catalysis by retaining β-1,4-glycosyl hydrolases and may be useful in the future for rational design of inhibitors.
ACKNOWLEDGMENT

I wish to express my sincere gratitude to my supervisor Dr. David R. Rose for his careful guidance, his extended patience, his invariable availability, and most of all his encouragement throughout my doctoral studies. Thanks David.

I extend my appreciation to my advisory committee, Dr. Victor Ling and Dr. Cheryl Arrowsmith, both of whom provided me with sound advice and valuable recommendations, and to my Ph.D. defense committee members who carefully scrutinized my work. Special thanks to my external examiners, Dr. P. Lynne Howell and Dr. P. Andrew Karplus, whose advice are of great value to me.

I'm indebted to our collaborators, in particular Dr. Steve Withers and his group who illuminated for me the complex catalytic mechanism of glycosyl hydrolases and who contributed with inhibitors rich in catalytic significance. I also thank Dr. Tony Warren, Dr. Neil Gilkes and Dr. Doug Kilburn for generous protein supply and helpful discussions.

I wish to extend my appreciation to my great labmates who provide a friendly and scientifically satisfying environment. I reserve a very special thank to Dr. Sona Vasudevan. Thanks Sona for your friendship and for your precious hints on ... life. Many thanks also Sona for our numerous scientifically rewarding discussions, which I value a lot. In addition, I want to thank Kathy Johns who untangled for me the mysteries of crystal soaking and handling, Sundeep Shahi, Leora Hoop, and Natalie Wong for their contribution to my projects, and Dr. Flip Hoedemaeker, Valerie Monem, and Dr. Alex Korn for incomparable help to and discussion with a challenged human being.

I gratefully acknowledge the Protein Engineering Network of Centres of Excellence (PENCE) for its financial support, and both the Fonds FCAR pour la formation de Chercheurs et l'Aide à la Recherche du Québec and the University of Toronto for doctorate studenships.
En mémoire de mon père Yvon
un homme simple aux nombreux talents
# TABLE OF CONTENT

## Chapter 1. Introduction

A) Rationale and thesis overview ................................................................. 2  
  Thesis outline ......................................................................................... 5  
B) Glycosyl hydrolases sequences and folding motifs .................................. 8  
  Catalytic domain of glycosyl hydrolases .............................................. 8  
  Classification by folding motif ............................................................. 8  
  Classification by substrate specificity .................................................. 10  
C) Catalytic mechanism of retaining β-glycosidases .................................... 13  
  Mechanism of lysozyme ....................................................................... 13  
  Nature of the intermediate: a paradigm ............................................... 18  
  Active site residue labeling ................................................................. 19  
  ALPH and saccharide distortion .......................................................... 22  
  Revised catalytic mechanism of β-glycosidases .................................. 25  
References ............................................................................................... 28  

## Chapter 2. Crystal structure of the Catalytic domain of the β-1,4-glycanase Cex from Cellulomonas fimi

Abstract .................................................................................................... 38  
Introduction ............................................................................................. 39  
Material and methods ............................................................................. 42  
  Crystallization and data collection ....................................................... 42  
  Isomorphous replacements .................................................................. 42  
  Structural analysis and model building ............................................... 43  
  Structure refinement ........................................................................... 43  
Results and discussion ........................................................................... 45  
  Multiple isomorphous replacement ..................................................... 46  
  Refinement and quality assessment of the current cex-cd molecular model . 47  
  Structural features ............................................................................. 48  
  Folding motif ....................................................................................... 51  
  Active site ............................................................................................ 53  
References ............................................................................................... 60
Chapter 3. Crystallographic observation of a covalent catalytic intermediate in a \( \beta \)-glycosidase

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>66</td>
</tr>
<tr>
<td>Introduction</td>
<td>67</td>
</tr>
<tr>
<td>Material and method.</td>
<td>68</td>
</tr>
<tr>
<td>Formaion of the complex</td>
<td>71</td>
</tr>
<tr>
<td>Structural analysis and refinement</td>
<td>71</td>
</tr>
<tr>
<td>Results and discussion</td>
<td>73</td>
</tr>
<tr>
<td>Steric structure of the glycosyl-enzyme attachment</td>
<td>76</td>
</tr>
<tr>
<td>Relevance of the substituent at the substrate C-2</td>
<td>82</td>
</tr>
<tr>
<td>Structural basis of the specificity by cex-cd</td>
<td>83</td>
</tr>
<tr>
<td>Implications</td>
<td>85</td>
</tr>
<tr>
<td>References</td>
<td>86</td>
</tr>
</tbody>
</table>

Chapter 4. Hydration of a glycal: The glycosyl-enzyme intermediate adopts a "full chair" conformation in the \( \beta \)-glycanase cex-cd

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>91</td>
</tr>
<tr>
<td>Introduction</td>
<td>92</td>
</tr>
<tr>
<td>Material and method.</td>
<td>93</td>
</tr>
<tr>
<td>Protein crystal preparation</td>
<td>96</td>
</tr>
<tr>
<td>Cellobial-cex-cd complex formation and data collection</td>
<td>96</td>
</tr>
<tr>
<td>Structure solution and refinement</td>
<td>97</td>
</tr>
<tr>
<td>Results and discussion</td>
<td>98</td>
</tr>
<tr>
<td>Conclusion</td>
<td>111</td>
</tr>
<tr>
<td>References</td>
<td>112</td>
</tr>
</tbody>
</table>

Chapter 5. General discussion, future directions, and concluding remarks

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>General discussion</td>
<td>115</td>
</tr>
<tr>
<td>Quality assessment of cex-cd model</td>
<td>116</td>
</tr>
<tr>
<td>Cex-cd folding motif</td>
<td>116</td>
</tr>
<tr>
<td>Structural characteristics of the catalytic intermediate</td>
<td>118</td>
</tr>
<tr>
<td>Similarity with haloalkane dehalogenase</td>
<td>119</td>
</tr>
<tr>
<td>Future directions</td>
<td>122</td>
</tr>
<tr>
<td>Concluding remarks</td>
<td>124</td>
</tr>
<tr>
<td>References</td>
<td>125</td>
</tr>
<tr>
<td>References</td>
<td>126</td>
</tr>
<tr>
<td>Appendix A</td>
<td>129</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Glucosyl hydrolases nomenclature</td>
<td>11</td>
</tr>
<tr>
<td>2.1</td>
<td>Diffraction data collection and MIRAS phasing</td>
<td>45</td>
</tr>
<tr>
<td>2.2</td>
<td>Heavy atom locations</td>
<td>46</td>
</tr>
<tr>
<td>2.3</td>
<td>MIRAS phasing</td>
<td>47</td>
</tr>
<tr>
<td>2.4</td>
<td>Crystallographic refinement statistics</td>
<td>48</td>
</tr>
<tr>
<td>3.1</td>
<td>Data collection and refinement statistics</td>
<td>74</td>
</tr>
<tr>
<td>4.1</td>
<td>Data collection statistics</td>
<td>98</td>
</tr>
<tr>
<td>4.2</td>
<td>Model refinement statistics</td>
<td>99</td>
</tr>
<tr>
<td>4.3</td>
<td>Hydrogen bonds between 2-deoxy-cellobioside and cex-cd</td>
<td>107</td>
</tr>
<tr>
<td>4.4</td>
<td>van der Waals contacts between 2-deoxy-cellobioside and cex-cd</td>
<td>108</td>
</tr>
<tr>
<td>A.1</td>
<td>Molecules proportion of the outer ( n ) layers of unit cells of the crystal</td>
<td>131</td>
</tr>
<tr>
<td>Figure Reference</td>
<td>Figure Title</td>
<td>Page</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1.1</td>
<td>Nomenclature for a molecule of β-D-glucopyranoside</td>
<td>12</td>
</tr>
<tr>
<td>1.2</td>
<td>Oxocarbonium ion-like transition state</td>
<td>15</td>
</tr>
<tr>
<td>1.3</td>
<td>Lysozyme catalytic mechanism</td>
<td>16</td>
</tr>
<tr>
<td>1.4</td>
<td>β-D-glucopyranose &quot;boat&quot; conformation</td>
<td>23</td>
</tr>
<tr>
<td>1.5</td>
<td>Presumed saccharide conformation through catalysis</td>
<td>25</td>
</tr>
<tr>
<td>1.6</td>
<td>Proposed catalytic mechanism for retaining β-1,4-glycosyl hydrolases</td>
<td>27</td>
</tr>
<tr>
<td>2.1</td>
<td>Cex gene product</td>
<td>39</td>
</tr>
<tr>
<td>2.2</td>
<td>Mobility of the residues and agreement of the model</td>
<td>49</td>
</tr>
<tr>
<td>2.3</td>
<td>Main chain (φ,ψ) angles analysis (Ramachandran Plot)</td>
<td>50</td>
</tr>
<tr>
<td>2.4</td>
<td>The fold of cex-cd, which forms an β/α-barrel</td>
<td>52</td>
</tr>
<tr>
<td>2.5</td>
<td>Active site of cex-cd</td>
<td>54</td>
</tr>
<tr>
<td>2.6</td>
<td>Glycosyl hydrolases family 10 consensus sequence</td>
<td>57</td>
</tr>
<tr>
<td>2.7</td>
<td>Position of the water molecules</td>
<td>58</td>
</tr>
<tr>
<td>2.8</td>
<td>A cluster of five water molecules which interacts with the loop 273-281</td>
<td>59</td>
</tr>
<tr>
<td>3.1</td>
<td>Cleavage of 2&quot;,4&quot;-dinitrophenyl-2-fluoro-2-deoxy-β-cellobioside (2F-DNPC)</td>
<td>70</td>
</tr>
<tr>
<td>3.2</td>
<td>Mobility of the residues and agreement of the model</td>
<td>75</td>
</tr>
<tr>
<td>3.3</td>
<td>Electron density map of the fluorocellobioside attached to Glu233</td>
<td>76</td>
</tr>
<tr>
<td>3.4</td>
<td>Fluorocellobioside binding region in cex-cd</td>
<td>77</td>
</tr>
<tr>
<td>3.5</td>
<td>19F-NMR spectra of cex-cd with 2-fluoro-disaccharides</td>
<td>79</td>
</tr>
<tr>
<td>3.6</td>
<td>Fluorocellobioside interactions with cex-cd</td>
<td>80</td>
</tr>
<tr>
<td>4.1</td>
<td>Cellobial chemical structure</td>
<td>94</td>
</tr>
<tr>
<td>4.2</td>
<td>Catalytic mechanism of hydration of glycals by β-glycosyl hydrolases</td>
<td>95</td>
</tr>
<tr>
<td>4.3</td>
<td>R-factor analysis as a function of resolution</td>
<td>100</td>
</tr>
<tr>
<td>4.4</td>
<td>Mobility of the residues and agreement of the model</td>
<td>101</td>
</tr>
<tr>
<td>4.5</td>
<td>Stereoview of the poorly resolved amino acids</td>
<td>102</td>
</tr>
<tr>
<td>4.6</td>
<td>Difference Fourier electron density omit map of the disaccharide</td>
<td>103</td>
</tr>
<tr>
<td>4.7</td>
<td>Surface charge distribution of cex-cd complexed to 2-deoxy-cellobioside</td>
<td>105</td>
</tr>
<tr>
<td>4.8</td>
<td>Active site of cex-cd with the bound ligand</td>
<td>106</td>
</tr>
<tr>
<td>4.9</td>
<td>Comparison of two covalent glycosyl-cex-cd complexes</td>
<td>109</td>
</tr>
<tr>
<td>4.10</td>
<td>Comparison of a 2-deoxy-α-cellobiosyl-Glu233</td>
<td>110</td>
</tr>
<tr>
<td>5.1</td>
<td>Structural characteristics of the catalytic intermediate of the retaining β-1,4-glycanase cex-cd</td>
<td>120</td>
</tr>
<tr>
<td>A.1</td>
<td>Geometry of a typical crystal of cex-cd</td>
<td>129</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>1D-NMR</td>
<td>One dimensional NMR</td>
<td></td>
</tr>
<tr>
<td>2F-DNPC</td>
<td>2&quot;,4&quot;-dinitrophenyl 2-fluoro-2-deoxy-cellobioside</td>
<td></td>
</tr>
<tr>
<td>2F-GMF</td>
<td>2-deoxy-2-fluoro-4-O-([β-D-glucopyranosyl]-β-D-mannopyranosyl fluoride</td>
<td></td>
</tr>
<tr>
<td>3D</td>
<td>Three-dimensional</td>
<td></td>
</tr>
<tr>
<td>19F-NMR</td>
<td>One dimensional fluorine-19 NMR</td>
<td></td>
</tr>
<tr>
<td>Ac</td>
<td>Acetate</td>
<td></td>
</tr>
<tr>
<td>ALPH</td>
<td>Anti-periplanar lone pair hypothesis</td>
<td></td>
</tr>
<tr>
<td>CBD</td>
<td>Cellulose-binding domain</td>
<td></td>
</tr>
<tr>
<td>CBD_{Cex}</td>
<td>CBD of Cex</td>
<td></td>
</tr>
<tr>
<td>CBHII</td>
<td>Cellobiohydrolase II from <em>Trichoderma reesei</em></td>
<td></td>
</tr>
<tr>
<td>Cex</td>
<td>β-1,4-glycanase (cellulase/xylanase) from the bacterium <em>Cellulomonas fimi</em></td>
<td></td>
</tr>
<tr>
<td>Cex-cd</td>
<td>Catalytic domain of Cex</td>
<td></td>
</tr>
<tr>
<td>COOH</td>
<td>Carboxyl-terminal</td>
<td></td>
</tr>
<tr>
<td>DhlA</td>
<td>Haloalkane dehalogenase</td>
<td></td>
</tr>
<tr>
<td>HEWL</td>
<td>Hen-egg white lysozyme</td>
<td></td>
</tr>
<tr>
<td>kDa</td>
<td>Kilo Dalton</td>
<td></td>
</tr>
<tr>
<td>KIE</td>
<td>Kinetic isotope effects</td>
<td></td>
</tr>
<tr>
<td>MIR</td>
<td>Multiple isomorphous replacement</td>
<td></td>
</tr>
<tr>
<td>MIRAS</td>
<td>Multiple isomorphous replacement with anomalous scattering signal</td>
<td></td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetyl-glucosamine</td>
<td></td>
</tr>
<tr>
<td>MurNAc</td>
<td>N-acetyl-muramic acid</td>
<td></td>
</tr>
<tr>
<td>NH_{2}</td>
<td>Amino-terminal</td>
<td></td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance spectroscopy</td>
<td></td>
</tr>
<tr>
<td>r.m.s.</td>
<td>Root mean square</td>
<td></td>
</tr>
<tr>
<td>TIM</td>
<td>Triose phosphate isomerase</td>
<td></td>
</tr>
<tr>
<td>TMLA</td>
<td>Trimethyl lead acetate</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>Electronegative atom</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 1

Introduction
A) RATIONALE AND THESIS OVERVIEW

An enzyme which cleaves a saccharide linkage with corresponding incorporation of a water molecule is defined as a glycosyl hydrolase. Glycosyl hydrolases are important, widely distributed metabolic and catabolic enzymes found in animals, plants and microorganisms. Mammals and plants produce O- and N-glycosyl hydrolases for the breakdown of larger carbohydrates. This process serves to incorporate the saccharides in maturing proteins, to derive energy, or to recycle the sugar units. Bacteria and fungi also secrete glycosyl hydrolases to derive both energy and metabolic components from plant biomass. One step toward a better understanding of the biology of glycosyl hydrolases consists of the understanding of their biochemical properties such as their folding motif and catalytic mechanism. Recently, there has been a great interest in using glycosyl hydrolases from microorganisms as industrial tools. A closer look at the composition of wood may help to understand the diversity of enzymes needed by microorganisms to extract monosaccharides from plants.

Wood is mainly composed of cellulose fibrils surrounded by hemicellulose, which in turn crosslinks to a generous amount of lignin; the chromophoric lignin links to hemicellulose at xylosyl units. Cellulose is a polymer of β-1,4-D-glucose; many chains of cellulose assemble to form micro-crystalline cellulose fibers of type Ia, Ib or II. Because of the high similarity of their chemistry, the relative composition of these types of chain within the cellulose fibers is poorly determined.

The high degree of compaction of micro-crystalline cellulose inherently limits its access by degradative enzymes. Because of its crystalline nature and its high degree of polymerization, cellulose is most efficiently broken down to glucose by using a mixture of enzymes. For instance, bacteria and fungi secrete three kinds of cellulases; together, endo-glucanases, exo-glucanases and β-glucosidases have been shown to yield β-D-glucose out of crystalline cellulose. Some bacteria secrete cellulosomes, which are organized multi-enzyme clusters, as a biological strategy to extract
monosaccharides from wood. These cellulosomes are believed to hydrolyze the complex plant cell walls quite efficiently, sometimes with synergism, and may provide a growth advantage for these cellulositic microorganisms (Béguin and Aubert, 1994).

The term hemicellulose, proposed about a century ago (Schulze, 1891), refers to the easily hydrolyzable part of plant cell walls. Found in lignified tissues of higher plants, the complex polysaccharides of hemicellulose bind non-covalently to cellulose both in the cell wall and intercellularly (Ferreira-Filho, 1994). The composition of hemicelluloses varies between plant species and may contain the following saccharides: L-arabinose, D-galacturonic acid, D-galactose, D-glucuronic acid, D-glucose, 4-O-methyl-D-glucoronic acid, D-xylose, O-acetyl groups, and ferulolate or coumarate esters via linked L-arabinose units to the polysaccharide backbone (Woodward, 1984; Puls and Poutanen, 1989). Xyloglucans, heteroxylans, and xylans are the most abundant components of hemicelluloses (Dey and Brinson, 1984). Imbedded between cellulose microfibrils, xylan molecules adsorb to cellulose fibers by van der Waals interactions and strong hydrogen bonds (Northcote, 1972; Ferreira-Filho, 1994). The hemicelluloses account for 30 to 40% of the total plant cell carbohydrates and are important structural components in lignocellulosic tissues (Williams, 1989; Zimmermann, 1989).

A current problem in the pulp and paper industries is the use of large quantities of both sodium hydroxide and chlorine-based bleaching agents in the extraction of cellulose fibers from both the hemicelluloses and the chromophoric lignin. Such chemicals pollute the environment and their use leads to undesirable side products. Moreover, the use of mechanical breakage, such as the Kraft pulping method, breaks cellulose fibers which lowers product quality. An alternative approach to pulping with chemicals is the use of enzymes, such as xylanases and acetylesterases, in an attempt to facilitate bleaching, that is to extract the chromophoric lignin from pulp (Grabski and Jeffries, 1991). The acetyl esterases remove acetyl groups from xylose units in polysaccharides, which gives access to the hemicellulose backbone for the xylanases, as was shown on extracts of birch wood (Poutanen and Puls, 1989). Hydrolysis at xylosyl sites
fragments the hemicellulose and thus allows the lignin to detach from its hemicellulose crosslinked counterpart. So the use of xylanases has great industrial interest.

Also of industrial interest is the production of monosaccharides from plant biomass, as a renewable source of energy. Derived from plants, recyclable paper constitutes a great source of glucose in the form of cellulose. Cellulases (β-1,4-glucanases) may be used to break down these fibers to glucose or cellobiose molecules. Biotechnological advances now permit the use of engineered bacteria to convert monosaccharides such as glucose or maltose to ethanol (Burchhardt and Ingram. 1992); so, plant biomass is a great source of renewable energy.

Further studies of glycosyl hydrolases are needed for a better understanding of their basis of catalysis. Studies with lysozyme revealed a catalytic mechanism which may be extended to other glycosyl hydrolases of its class. Because the products released by hen-egg white lysozyme retain their anomeric configuration β, this enzyme is said to be "retaining". The groundbreaking elucidation of the lysozyme three-dimensional structure in combination with molecular modeling studies suggested an ionized residual saccharide as the reactive species in the enzyme: inherently, such a species is presumed to be of sufficient lifetime to be an intermediate step in catalysis. Later studies such as kinetic isotope effects firmly established the involvement of such an ionic species in catalysis by lysozyme and other β-glycosyl hydrolases. However, studies with 2-fluorosaccharides suggest the existence of a covalent glycosyl-enzyme as the catalytic intermediate. Combining these two views led to a unified catalytic mechanism hypothesis in which saccharide cleavage proceeds with the formation and hydrolysis of a covalent glycosyl-enzyme species, via ionic transition states (Sinnott, 1990; Withers, 1995). Although more and more evidence suggests that the nature of the intermediate is a covalent species, this species had not been observed. Kinetic methods, chemical methods, and 1D-NMR spectroscopy do not give details on the spatial arrangement of functional chemical groups at the active centre of the enzyme, and thus of the biological context for catalysis. Much more detail about the spatial representation of both an enzyme and its complexed form with disaccharides may be provided with the technique of X-ray diffraction crystallography. X-ray diffraction data obtained at high resolution gives intricate structural information such as: the
architecture of the enzyme and its secondary structure elements; the conformation and precise location of its individual amino acids; the mapping and the conformation of the enzyme's active site; the residues located in the vicinity of the catalytic amino acids which may be involved in secondary catalytic assistance; the features defining the enzyme's specificity; the geometry of the catalytic intermediate; and the spatial distribution of the residues found in the vicinity of the site of saccharide attachment. These results may then provide a better understanding of the underlying catalytic mechanism of retaining \( \beta-1,4 \)-glycosyl hydrolases.

This thesis focuses on detailed three-dimensional analyses of one bacterial cellulase-xylanase, cex-cd, which is being extensively studied (Bray et al., 1996) in the context of both its possible industrial application and its relevance to the understanding of the process of catalysis for its class of enzymes. At the start of this work, the three-dimensional structure of bacterial and fungal glycosyl hydrolases revealed a remarkable variety of different folding motifs for enzymes of similar function (Davies and Henrissat, 1995). This observation suggested that the catalytic domain of cex may show yet another folding motif for microbial \( \beta \)-glycosyl hydrolases. In addition, chemical trapping using fluorocelllobiosides with cex-cd suggested that a covalent glycosyl-enzyme species may be obtained in the crystalline environment and structurally characterized. An attempt to verify these two hypotheses is presented here, with an emphasis on the understanding of both cex-cd specificity and its catalytic mechanism.

Thesis Outline

The remainder of Chapter One covers the evolution of our understanding of the mechanism of lysozyme. In particular there are two controversial aspects. Firstly, it has long been believed that the catalytic intermediate of this enzyme and its homologues is an ionic species, that is oxocarbonium ion-like. However, evidence has accumulated to suggest that the catalytic intermediate is a covalent species. Secondly, it is only recently that experimental evidence strongly supports a conformational change of the binding saccharide at the catalytic centre of the enzyme into a "boat"
conformation. These two controversial aspects are then combined with the initial mechanism of lysozyme into a revised catalytic mechanism for retaining β-1,4-glycosyl hydrolases.

Chapter Two reports the high resolution crystal structure of the catalytic domain of the β-glycanase Cex (cex-cd). This structure was determined by the classical method of isomorphous replacement with heavy atoms. The structure of the enzyme reveals a (β/α)₈-barrel folding motif, which was not previously observed for microorganisms' β-1,4-glycosyl hydrolases. The structure verifies that both the acid/base catalyst Glu127 and the nucleophile Glu233 are appropriately situated in a crevice, at the carboxy-terminal end of the (β/α)₈-barrel motif of the enzyme; this crevice is inferred to be the active site of the enzyme. In addition, further analysis of this active site suggests that the triad Asp235-His205-Glu233 may be important for catalysis. A modified version of this Chapter appeared in *Biochemistry* (White et al., 1994).

Chapter Three presents the crystal structure of a covalent fluorocellobioside-cex-cd complex at a resolution of 1.8 Å. This complex represents the catalytic intermediate which was previously proposed to occur but not actually observed. Interactions between the disaccharide and the enzyme's active site are analyzed in the context of the enzyme's better specificity for xylobioside substrate than cellobioside. This conformation also reveals a possible hydrogen bond between the hydroxyl group at position C2 of the substrate and the carbonyl oxygen of the esterified nucleophilic side chain of Glu233. Moreover, because such a complex is formed in the crystal, cex-cd is thus shown to be active in the crystalline environment; inherently, this enzymatic competence also validates the biological relevance of the enzyme conformation observed in cex-cd crystals. Implications of such a complex help in the understanding of the structural aspects of the covalent catalytic intermediate, which may be extended to all the retaining β-1,4-glycosyl hydrolases. A modified version of this Chapter has been published in *Nature Structural Biology* (White et al., 1996).

Chapter Four reports the crystal structure of a complex between cellobial and cex-cd. Glycals such as cellobial have been recognized to be hydrated by glycosyl hydrolases, a process believed to involve a covalent glycosyl-enzyme intermediate. The covalent 2-deoxy-cellobiosyl-
The ces-cd complex obtained is the first such example for a glycosyl hydrolase, to our knowledge. These results strengthen the conclusions in Chapter Three and give insights into the stereochemical course of hydration, which has been the subject of debate for the last fifteen years. These results are being prepared for publication.

The final Chapter Five includes the validation of the structures obtained, a justification of the biological relevance of the folding motif observed, seven structural characteristics of the covalent catalytic glycosyl-ces-cd intermediate, a previously unexpected similarity of catalytic machinery with the enzyme haloalkane dehalogenase, and future directions.
B) Glycosyl Hydrolases Sequences & Folding Motifs

The products of genes encoding glycosyl hydrolases frequently consist of a modular arrangement of domains. By definition, these enzymes include a catalytically active domain which hydrolyzes specific saccharide linkages. In addition, glycosyl hydrolases may contain additional functional domains which either bind to oligosaccharides or are homologous to fibronectin type III (FnIII). Although homologous sequences to these latter FnIII are occasionally observed in glycosyl hydrolases, saccharide-binding domains are more widespread.

Cellulose-binding domains (CBD) are examples of saccharide-binding domains. CBDs are commonly found as globular moieties of about 10 to 15 kDa and have been shown to bind either to cellulose or starch. Multidimensional NMR studies reveal that CBDs fold to compact domains rich in β-strand secondary structures, which present aromatic residues to the saccharide-binding region (Xu et al., 1995). The functional significance of CBDs is either to promote hydration of crystalline cellulose or to increase the processivity of their associated catalytic domains. Probed by scanning microscopy, recombinant CBD modules of the β-1,4-glycanase CenA from Cellulomonas fimi have been shown to hydrate cotton's microcrystalline cellulose leading to the disruption of the surface of cellulose and to the release of polysaccharide fragments from cotton fibres (Din et al., 1991); these smaller fragments are more accessible to further degradation by glycosyl hydrolases. Due to their ability to bind saccharides, CBDs may also increase the attachment time spent by glycosyl hydrolases on their substrate, leading to an increased processivity. Thus, CBDs add efficiency to glycosyl hydrolases.

Catalytic Domain of Glycosyl Hydrolases

Classification by folding motif

The catalytic domains of glycosyl hydrolases have been classified into 45 families according to hydrophobic cluster analysis (Gaboriaud et al., 1987; Lemesle-Varloot et al., 1990) and amino acid sequence alignments (Henrissat, 1991; Henrissat and Bairoch, 1993). As an example, the sequence alignment of family 10 reveals 36 highly conserved residues which are
landmarks for this family. These conserved residues are not all contiguous, which makes it more difficult to understand their biological significance. Given the rapid increase of the number of glycosyl hydrolase sequences, Henrissat and co-workers proposed a classification based on additional structural properties such as their protein architecture and their reaction products; these properties give clues to evolutionary relationships between these enzymes (Henrissat and Romeu, 1995). The elucidation of the folding motif of these enzymes and the sequence analysis with new genes rigorously test this classification into families. In support of the above classification is the similarity between the folding motif revealed by X-ray diffraction results for different enzymes of a common family. For example, the known structures of the family 10 xylanases all fold to a (β/α)8-barrel (Derewenda et al., 1994; Harris et al., 1994; White et al., 1994; Dominguez et al., 1995) whereas the functionally-related xylanases of family 11 form a twin pleated β-sheet (Campbell et al., 1993), which is a distinction predicted by Henrissat's classification.

Recently, similarity was identified between families by three means, suggesting grouping into 'superfamilies'. Sequence analysis revealed that some new sequences seem distantly related to more than one family. Using hydrophobic cluster analysis, genes of the families 11 and 12 were proposed to be structurally related (Törrönen et al., 1993). Sequence alignment and phylogenetic tree analyses revealed similarity between sequences of families 27 and 36 (Dagnall et al., 1995). Detailed X-ray structural analyses revealed a topologically similar folding motif for some families: 7 and 16 into a lectin-type folding motif (Divne et al., 1994); 19, 22, 23 and 24 into a lysozyme-architecture (Holm and Sander, 1994); 33 and 34 with a β-propeller architecture (Crennell et al., 1993). The largest superfamily includes enzymes from families 1, 2, 5, 10 and 17 which all fold into a (β/α)8-barrel, share three conserved residues, and hydrolyze their substrate with retention of anomeric configuration (Henrissat and Romeu, 1995; Jenkins et al., 1995). In this 4/7 superfamily, the two key catalytic carboxylates are exposed on β-strands 4 and 7. Enzymes of families 30, 35 39, and 42 were proposed to share the properties of the 4/7 superfamily, although their detailed structure is yet to be revealed (Henrissat and Romeu, 1995; Jenkins et al., 1995). All these similarities suggest common folding motif ancestors (Henrissat et al., 1995; Jenkins et al., 1995).
The concept of 'superfamily' may arise from a number of sources: a limited number of different architectures for these enzymes, a higher occurrence of some protein architectures than others for proteins, and possibly a demonstration that some protein folding motifs crystallize more readily than others. It has been previously suggested that the number of protein folding motifs is limited to only a few thousands (Chothia, 1992) and that only nine 'superfolds' account for about 30% of the known protein architectures (Orengo et al., 1994). For the 22 families of glycosyl hydrolases for which a three-dimensional structure is known, only ten distinct folding motifs are observed (Davies and Henrissat, 1995) of which the most commonly found is the 'classic' TIM-like (β/α)$_8$-barrel. Whether this observation expresses a common ancestor for (β/α)$_8$-barrel enzymes (Farber and Petsko, 1990; Farber, 1993), a more energetically favorable architecture, or a more readily crystallizable motif is still an open question.

Classification by substrate specificity

The classification of glycosyl hydrolases into families also suggests a conservation of substrate specificity. Substrate specificity may be further subdivided to three levels. First, glycosyl hydrolases are either exo, that is specific for the release of either mono- or disaccharides from the end of sugar chains, endo, that is with a random site of cleavage along a polysaccharide chain, or of the class of β-glucosidases which hydrolyze disaccharides. Family 6 of glycosyl hydrolases contains an exception in which members do not all have the same substrate specificity, such as the cellobiohydrolase CbhII from *Trichoderma reesei* and the endo-glucanase E2 from *Thermomonospora fusca*. From the crystal structure of CbhII (Rouvinen et al., 1990), it has been suggested that two surface loops restrict the active site's accessibility to the substrate, thus forming a "tunnel"; cellulose chains would venture into such a tunnel with its non-reducing-end first and be hydrolyzed as it migrates along. Sequence alignment reveals that these two loops are not present in E2 (Meinke et al., 1995). The crystal structure of E2 (Spezio et al., 1993) exhibits an accessible active site with which this enzyme binds at random locations along a polysaccharide. To analyze the exo/endo specificity, the group of Dr. A. Warren made a mutant (CbhΔΔC) of *Cellulomonas*
Jimi celllobiohydrolase A (CbhA) in which the corresponding loop (residues 373-387), presumed to define the exo activity (Meinke et al., 1995), was deleted. As expected, CbhAΔC exhibits an enhanced endoglucanase activity, as shown by a higher rate of fluidity increase in soluble O-(carboxymethyl) cellulose, with the fluidity index advancing from 13.3 \(\varphi_{sp}\text{-ml-mmol}^{-1}\) (typical of an exoglucanase activity) to 44.5 \(\varphi_{sp}\text{-ml-mmol}^{-1}\).

A second substrate specificity level refers to the carbohydrate composition, which may be of great diversity. For example, there are over \(10^{12}\) possible isomers of a reducing hexasaccharide (Laine, 1994). Although cellulose, \((\beta-1,4-D\text{-glucose})_n\), and chitin, \((\beta-1,4-D\text{-GlcNAc})_n\), constitute by far the most common polysaccharides in nature, many other sugars are used by living organisms in a myriad of biological functions, including storage, structure, and protection from proteolysis. Selective hydrolysis is thus represented by different families of glycosyl hydrolases.

Thirdly, glycosyl hydrolases may be classified on the basis of both their linkage-specificity and catalytic outcome. These enzymes are specific for either \(\alpha\) or \(\beta\) substrate linkages and may yield either retention or inversion of this configuration (Table 1.1). Retaining enzymes release

<table>
<thead>
<tr>
<th>Table 1.1. Glucosyl hydrolases nomenclature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substrate (a)</td>
</tr>
<tr>
<td>(\alpha)-glucoside</td>
</tr>
<tr>
<td>(\alpha)-glucoside</td>
</tr>
<tr>
<td>(\beta)-glucoside</td>
</tr>
<tr>
<td>(\beta)-glucoside</td>
</tr>
</tbody>
</table>

\(a\) The enol ether \(O\)-link which binds two saccharide units can have either an \(\alpha\) or \(\beta\) anomeric configuration defined respectively as an axial or equatorial bond of the saccharide developing a reducing-end.

product with the same anomeric configuration at the C1 of the newly formed reducing-end as the original substrate saccharide-saccharide link, whereas inverting glycosyl hydrolases yield products
with different configuration (Figures 1.1). For example, the retaining β-1,4-glycosyl hydrolase mammalian lysozyme releases product with 99.7% of conservation of the β-anomer (Dahlquist et al., 1969a). However, by the action of water, these reducing sugars undergo natural mutarotation at the anomeric centre and at equilibrium yield both enantiomers. Thus, the anomeric configuration of the product refers to the "initial product" obtained from the enzyme's action.

a)

\[
\text{re face}
\]

\[
\text{si face}
\]

b)

\[
\text{retaining enzyme}
\]
\[
\text{inverting enzyme}
\]

\[
\beta\text{-anomer}
\]
\[
\alpha\text{-anomer}
\]

**Figure 1.1 Nomenclature for a molecule of β-D-glucopyranoside.** a) Atomic and spatial identification for a β-glucopyranosyl unit. The equatorial hydroxyl at atom C1 defines the β configuration. The face \textit{si} refers to the side of the saccharide which presents C1, C3 and C5 coplanar, whereas \textit{re} is the opposite saccharide's side. b) Outcome of retaining and inverting β-1,4-glycosyl hydrolases. The difference between the two anomers refers to the position of the substituent (bold) at the carbon C1. Additional saccharides are represented by R' and R''.
C) **Catalytic Mechanism of Retaining β-Glycosidases**

The remarkable achievement of the three-dimensional structure determination of lysozyme (Blake *et al.*, 1965; Imoto *et al.*, 1972), the first enzyme structure revealed by X-ray diffraction, opened a window of opportunity to a better understanding of catalysis by an enzyme. Since the 1960s, the catalytic mechanism proposed for this glycosyl hydrolase served as a prototype for the understanding of other similar enzymes. However, there are two aspects of this mechanism which may need to be revisited, namely the *nature of the intermediate* and the *conformation of the binding saccharide*. This mechanism will first be introduced, with an emphasis on its underlying experimental evidence. Apparently conflicting results will then be presented, questioning the nature of the intermediate, and supporting evidence for a covalent glycosyl-enzyme species will be reviewed. The third part gives an overview of labeling agents, of which fluorosaccharides have been used to study the enzyme's catalytic process. The fourth part deals with saccharide's conformational dynamics, which correlate with the small molecule theory of stereoelectronic effects. Finally, a revised catalytic mechanism which may more appropriately explain evidence reported for lysozyme and the other glycosyl hydrolases will be drawn.

**Mechanism of Lysozyme**

Lysozyme (*N*-acetylmuramidase glucanohydrolase, E.C.3.2.1.17) hydrolyzes the *O*-glycosyl link between *N*-acetyl-**D**-glucosamine (GlcNAc) and *N*-acetyl-**D**-muramic (MurNac) acid saccharides. These saccharides are found in the peptidoglycan part of the cell wall's polysaccharides in bacteria. Chitin (GlcNAc)ₙ in insects and in crustaceous outer shells has also been recognized as a substrate for lysozyme. Prior to the elucidation of the three-dimensional structure of lysozyme, little was known about its catalytic mechanism. Subsequent studies on its mechanism of action have been slower than for other enzymes, such as the proteases, because of the difficulty of synthesis of oligosaccharide substrates and inhibitors.

The crystal structure of lysozyme reveals a long cleft running on the surface of the enzyme (Johnson and Phillips, 1965). Although carried on at a resolution of 6 Å, this study with
Saccharide inhibitors revealed the location of lysozyme's active site and some key catalytic residues. Further analyses of this cleft were reported for a non-productive complex with chitotriose, \((\text{GlcNAc})_3\), which occupies three binding subsites, denoted A, B, and C but does not fill the cleft completely (Phillips, 1966; Blake et al., 1967; Phillips, 1967). Structural studies of a complex with longer saccharides were difficult for two reasons, because lysozyme cleaves such longer sugars, and because one part of the enzyme's active site is partially blocked by another molecule in the crystal, due to molecular packing. In an attempt to describe further saccharide binding, Phillips and colleagues constructed a model with two trisaccharides in the active site of lysozyme; a duplicate of observed trisaccharide \((\text{GlcNAc})_3\) is moved into the vacant moiety of the enzyme's cleft (subsites D, E, and F), assuming that these two trisaccharides adopt a similar conformation. Given the substrate specificity and the results of this modeling, cleavage was deduced to occur at the saccharide linkage located between subsites D and E. More importantly, because of the steric interactions of the hydroxymethyl C6-OH group of the sugar at subsite D with Trp108, the main chain carbonyl of Asp52, and the acetamido group of the sugar residue at subsite C, the pyranose ring D was inferred to be distorted away from the energetically favorable "chair" to a "sofa" (half-chair) conformation (Phillips, 1967). A half-chair saccharide conformation had previously been proposed for an oxocarbenium ion in the transition state structure of acid hydrolysis of polysaccharides (Edward, 1955). Such an oxocarbenium ion-like species (Figure 1.2) was thus proposed to be involved in the lysozyme catalytic mechanism principally on the basis of \(\alpha\)-secondary deuterium kinetic isotope effects (Dahlquist et al., 1968) and positive binding energy \((\Delta\Delta G)\) calculations which favor a distorted rather than a chair conformation at subsite D (Chipman and Sharon, 1969). These two experiments were repeated for a number of glycosyl hydrolases and found to give consistent results with respect to lysozyme (Sinnott, 1990). The involvement of an oxocarbenium ion-like species as the transition state in catalysis by lysozyme is widely accepted, but the lack of experimental evidence at this point precludes a firm statement for the structure of the catalytic intermediate.
Figure 1.2. Oxocarbonium ion-like transition state. R groups represent substituents. Because of the sp² hybridization state at C1 of this residual saccharide and the likelihood of the participation of a lone pair of electrons of O5 in resonance with the electron-compromised C1, the cyclic atoms C5, O5, C1, and C2 are inferred to be coplanar. This residual saccharide is thus presumed to conform to a half-chair conformation as shown.

The catalytic mechanism for the retaining β-glycosyl hydrolase lysozyme from hen egg-white (HEWL), as originally proposed (Phillips, 1967; Imoto et al., 1972), is a general acid catalysis (Figure 1.3) which is summarized as follows:

1. The active form of HEWL is Asp52 deprotonated and Glu35 protonated;

   Kinetic analysis monitoring $k_{cat}/K_M$ as a function of pH and titration studies shows that catalysis by lysozyme is dependent on pH, such that there is an acid with abnormally high $pK_a$ 6.1-6.3 identified as Glu35 and a base of $pK_a$ ~4, which is Asp52 (Donovan et al., 1960; Pollock et al., 1967; Parsons and Raftery, 1972b; Parsons and Raftery, 1972a); these values were recently supported by two-dimensional $^1$H NMR spectroscopy (Bartik et al., 1994). Specific esterification of Asp52 by triethoxonium fluoroborate (Parsons and Raftery, 1970; Kuramitsu et al., 1977) and chemical modification of the carboxyl groups of lysozyme with carbodiimide both agree with the ionization states of Glu35 and Asp52 (Lin and Koshland, 1969). Site directed mutagenesis of either Glu35 or Asp52 abrogates lysozyme's activity (Malcolm et al., 1989). The crystal structure of lysozyme supports a higher $pK_a$ for Glu35 given its apolar environment in the proximity of Trp108 (Inoue et al., 1992). The ionic species of Asp52 and Glu35 are also strongly supported by neutron diffraction at 1.4 Å resolution, for which a proton is clearly resolved on Glu35 whereas Asp52 is deprotonated (Mason et al., 1984a; Mason et al., 1984b).
**Figure 1.3. Lysozyme catalytic mechanism.** An oxocarbenium ion-like species is formed and stabilized by counterions Glu35 and Asp52; then, a water molecule attacks this positively charged intermediate which leads to retention of anomeric configuration. The binding saccharide is presumed to be distorted toward the oxocarbenium ion-like "sofa" conformation.

2- Binding of a substrate to the active site of lysozyme imposes distortion of the saccharide unit at subsite D.

(a) Steric hindrance deduced from model building (Phillips, 1967), binding energy (ΔΔG) calculations (Chipman and Sharon, 1969) and binding of the transition state analog δ-lactone derived from GlcNAc₄ (Ford et al., 1974) all suggest a more favourable binding of a distorted saccharide, such as a "sofa" (half-chair), over a chair conformation. Because lactone inhibitors compete efficiently with the lysozyme substrate, as represented by $K_i = 8 \times 10^{-8}$ M, compared to $K_M$ of $10^{-5}$ M for the substrate GlcNAc₄, lysozyme displays a preferential binding for a trigonal geometry at C1 (Ford et al., 1974). Two lysozyme crystal structures predict a sofa conformation for the sugar at subsite D, namely hen egg-white lysozyme co-crystallized with the cell wall trisaccharide product 2-acetamido-2-deoxy-D-muramic acid - β(1,4)-2-acetamido-2-deoxy-D-glucosyl - β(1,4)-2-acetamido-2-deoxy-muramic acid (MurNAc-GlcNAc-MurNAc) (Strynadka and James, 1991) and the covalent adduct formation of the cell wall glycopeptide with a mutant Thr26→Glu of phage T4 lysozyme (Kuroki et al., 1993).
(b) Through hydrogen bonds, Glu35 would donate a proton to the linking β oxygen between the sugars occupying the subsites D and E, which leads to the weakening of the glycosidic bond C1–O.

3– Formation of an oxocarbenium ion-like species as a transition state.

(a) The glycosidic bond C1–O between subsites D and E is cleaved, leading to the release of the aglycone, that is the product with a newly exposed non-reducing-end from subsites E and F, and the development of a positive charge on the saccharide in subsite D.

(b) Simultaneously, this positively charged saccharide in subsite D becomes stabilized by adopting a full planar conformation at C5–O5–C1–C2 (the 'sofa' conformation), in which O5 and C1 are trigonal by virtue of a partial double bond formed by the delocalization of a lone pair electron of O5. The negatively charged Asp52 is presumably the main stabilizing factor as a counter-ion for this oxocarbenium ion-like species, which was also postulated to be a catalytic intermediate (Vernon, 1967). Theoretical calculations indicate that the charged Asp52 side chain would be sufficient for stabilizing the unstable transition state (Warshel and Levitt, 1976). The involvement of an oxocarbenium ion-like species is strongly supported by $^{18}\text{O}$ leaving group isotope effects (Rosenberg and Kirsch, 1981) and by α-secondary deuterium kinetic isotope effects (KIE) (Dahlquist et al., 1969b; Smith et al., 1973; Sinnott, 1987; Sinnott, 1990); for this latter measurement, a value greater than 1.00 indicates the hybridization state $sp^2$ of the adjacent atom, that is C-1. This hybridization is characteristic of a carbonium ion, which itself becomes stabilized by a resonance form with O5, thus forming an oxocarbenium ion-like residual saccharide.

(c) This ionic species is presumed to have a lifetime long enough to allow both the release of the first product and the diffusion of a water molecule into the vicinity of the reactive carbon C1, prior to the formation of covalent acylal (glycosylated-enzyme) intermediate (arguments against this conjecture are presented below).

4– A water molecule inserts at the re face (see Figure 1.1a) of the reactive oxocarbenium ion-like species and in the vicinity of Glu35, which abstracts a proton from that water which then attacks
the residual saccharide at the carbon C1. The stereochemistry of this nucleophilic addition by hen egg-white lysozyme leads to more than 99.7% of retention of the saccharide’s anomeric configuration (Dahlquist et al., 1969a).

**NATURE OF THE INTERMEDIATE: A PARADIGM**

As early as 1953, Koshland postulated that 'retaining' glycosyl hydrolases form a covalent glycosyl-enzyme intermediate as a catalytic event (Koshland, 1953). The formation of such a covalent intermediate supposes a double displacement type of mechanism. On the basis of similarity to small molecule reactions, Koshland proposed a tandem series of nucleophilic substitutions of order 2 ($S_N2$). An aglycone is first released with concomitant glycosyl-enzyme covalent intermediate formation by an $S_N2$ mechanism, which is then displaced by a water molecule through a second $S_N2$ event leading to the retention of the configuration of the saccharide.

Calculations based on theoretical expectation led the group of Jencks to show that a non-bonded oxocarbenium ion-like species has a lifetime in the order of $10^{-12}$ sec to $10^{-10}$ sec (Bennet and Sinnott, 1986; Amyes and Jencks, 1989), far too small to account for the turnover of (GlcNAc-MurNAc)$_3$ by lysozyme at $k_{cat} = 0.5$ sec$^{-1}$ (Rupley, 1967). As an alternative, the residual saccharide may react with the negatively charged side chain of Asp52, which is located in the vicinity of saccharide’s C1, thus forming a covalent glycosyl-enzyme complex. Such a complex is presumed chemically more stable (Sinnott, 1990).

Because of their active site structure similarity, their equivalent catalytic residues and environments, the similarity of their substrates and their product outcome, it is reasonable to assume that retaining β-glycosyl hydrolases, including lysozyme, proceed through a similar catalytic mechanism. Supporting evidence for the involvement of a covalent glycosyl-enzyme intermediate comes from glycosidases other than lysozyme, but the knowledge learned from β-glycosidases may extend to lysozyme.

Analysis of kinetics at sub-zero temperatures suggests the accumulation of a saccharide-enzyme intermediate in the β-galactosidase from *Escherichia coli* K12 (Fink and Angelides, 1975).
At a temperature below -10°C this enzyme shows both a much slower turnover of o-nitrophenyl-β-D-galactoside than at higher temperatures and a "burst" of o-nitrophenyl lasting a few seconds, which is linear with enzyme's concentration. Although kinetic rate constants were not measured at sub-zero temperatures, the presence of o-nitrophenyl bursts stoichiometric with enzyme concentration suggest the presence of a galactosyl-enzyme intermediate. It is unknown whether the temperature only slowed down the rate of catalysis or affected the mechanistic route of catalysis by this β-galactosidase.

Although kinetic isotope effects (KIEs) strongly suggest the involvement of an oxocarbenium ion-like transition state, further KIE studies reveal that this ionic species is involved at two occasions during catalysis. Depending on the chemical nature of the aglycone of synthetic substrates, the rate-limiting step will be either the formation of the intermediate or its hydrolysis. Using a variety of aglycone leaving groups coupled to a specific saccharide moiety, KIEs were observed for the glycosylation and the deglycosylation of β-glycosyl hydrolases, such as for Agrobacterium β-glucosidase (Kempton and Withers, 1992) and Cellulomonas fimii β-glycanase cex-cd (Tull and Withers, 1994). These results suggest that, for these two enzymes and consequently for homologous ones, there are two steps involved in the catalytic mechanism, which are separated by a tetrahedral species of the residual saccharide. This conclusion suggests the existence of a covalent glycosyl-enzyme intermediate, a tetrahedral species.

**ACTIVE SITE RESIDUE LABELING**

In an attempt to identify active site residues of β-glycosyl hydrolases, such as the nucleophile forming a tetrahedral species with the residual substrate, saccharide-derived crosslinking agents are often used. Crosslinking glycosides bind covalently to glycosyl hydrolases. A crosslinking agent contains two parts, that is an activable reactive group which forms a covalent attachment to the enzyme, and a joined saccharide moiety which provides specificity for binding.

Epoxide-based crosslinking agents offer the promise of localizing the enzyme's active site but may not be informative about the enzyme's catalytic mechanism for three reasons. Firstly, the
site of attachment may not necessarily be at the catalytic region of the active site. It has often been observed that glycosyl hydrolases present more than one binding subsite to an incoming ligand, such as lysozyme which binds 6 units of saccharide. For example, the inhibitor \((\text{GlcNAc})_2\) binds preferentially one-saccharide-subsite away from the active site catalytic residues Asp52 and Glu35 in partridge lysozyme (Turner and Howell, 1995). It is thus conceivable that a \((\text{GlcNAc})_n\)-based crosslinking agent may bind distal from the catalytic region of the active site and thus may label an otherwise unexpected enzyme residue. Indeed, in the crystal structure of human lysozyme complexed with affinity labeled 2',3'-epoxypropyl \(\beta\)-(GlcNAc)_2, both binding subsites B and C are occupied but with a different pattern of hydrogen bonding interactions with the enzyme's active site as compared with the natural substrate (Muraki et al., 1996). Therefore, enzyme labeling using epoxide compounds may perturb the mode of binding of its saccharide component. Secondly, the site of attachment may not be of predictable functionality. For example, it has been suggested that conduritol B binds covalently to \(Escherichia\ coli\ \beta\)-galactosidase's catalytic nucleophile, thus identified as Glu461 (Herrchen and Legler, 1984). However, this glutamate residue turned out to be the acid/base catalyst, whereas the nucleophile is now recognized to be Glu537 as shown with fluorosaccharide trapping agents (Gebler et al., 1992). In another example, three homologous crosslinking agents bind covalently to either Glu86 or Glu177 in the active site of the endo-1,4-xylanase II from \(Trichoderma reesei\) depending on their size, as shown by the high resolution crystal structure of their complex with this enzyme (Havukainen et al., 1996). In this example, an ideally specific labeling agent would bind to only one residue, which is not the case. These two examples clearly indicate a reduced specificity for the enzyme's site of attachment when epoxide crosslinking agents are used. Thirdly, the aglycone moiety of affinity crosslinking chemicals necessarily requires additional atoms which will mediate the attachment between the residual glycoside and the targeted amino acid. Given the restricted space in the active site of an enzyme, it is desirable not to insert unnecessary atoms which may occlude the active site pocket and thus impair its structural interpretation. There are at least two reported examples of detailed three-dimensional structures in which the covalent glycosyl-enzyme species may not be
representative of a catalytic condition: the *Bacillus* β-1,3-1,4-glycanase covalently complexed with 3,4-epoxybutyl β-D-celllobioside (Keitel et al., 1993), and the *endo*-1,4-xylanase II from *Trichoderma reesei* covalently bound with epoxyalkyl xylosides (Havukainen et al., 1996). In these two examples, a chain of non-saccharide atoms populates the enzyme's active site, thus complicating possible understanding of the enzyme's catalytic mechanism.

In contrast to epoxides, fluorosaccharides are minimally modified saccharides which specifically bind to glycosyl hydrolases. Upon reaction with a retaining β-glycosyl hydrolase, a fluorosaccharide forms a reversible covalent species specifically with the enzyme's nucleophile (Withers et al., 1987). Compared to its natural substrate homologue, the fluorosaccharide used in this work bears a hydroxyl to fluorine substitution at the C2 carbon, that is next to the anomeric carbon. This substitution is isosteric and thus may not impair the binding ability of fluorosaccharides by steric hindrance. Additional atoms are not needed between the saccharide's C-1 and the enzyme's catalytic nucleophile. Moreover, because the enzyme processes a fluorosaccharide in a similar manner as an unsubstituted substrate, the resulting attachment stereochemistry is more likely to be catalytically significant. The slow hydrolysis of the covalent glycosyl-enzyme species also suggests that this complex is catalytically competent, which was shown both for the *Agrobacterium* β-glucosidase (Withers et al., 1990; Street et al., 1992) and for the *Cellulomonas fimii* β-1,4-glycanase cex-cd (Tull and Withers, 1994). In addition, these results show that the reactivation is accelerated by the reverse reaction of transglycosylation in the presence of a suitable saccharide acceptor.

With the use of specific fluorosaccharides, the catalytic intermediate has been isolated and characterized for a number of enzymes. The intermediate is a covalent attachment of the residual saccharide to a carboxylate (Asp or Glu) which acts as a catalytic nucleophile. Using radiolabelled 2-deoxy-2-fluoroglycosides, the catalytic nucleophile has been identified for the 'retaining' β-1,4-glycosidases *Agrobacterium faecalis* β-glucosidase (Withers et al., 1990), *Escherichia coli* β-galactosidase (Gebler et al., 1992), human glucocerebrosidase (Miao et al., 1994), *Cellulomonas fimii* β-1,4-glycanase (cellulase/xylanase) Cex (Tull et al., 1991) and *Clostridium thermocellum*
endo-glucanase CelC (Wang et al., 1993). A recent technique for detection of glycopeptides consists of the identification of the glycosylated peptide from a hydrolysate by its modified mass using electrospray mass spectroscopy for detection (Tull et al., 1994). This technique alleviates the use of radioactive compounds and, combined with the isolation of the labeled peptide, permits its solid phase amino acid sequencing (Withers, 1995); this procedure has been used with the β-1,4-glycanase cex-cd, the Clostridium thermocellum endoglucanase C, and the Agrobacterium faecalis β-glucosidase (Tull et al., 1994). The fluorosaccharides also directly identify the enzyme's active site nucleophile, which is important for a better understanding of its catalytic mechanism.

**ALPH AND SACCHARIDE DISTORTION**

P. Deslongchamps (1983) presented the Antiperiplanar Lone Pair Hypothesis (ALPH) as a basis to explain organic chemical reactions. This hypothesis predicts both the chemical course and the outcome for the elimination reaction, which proceeds more efficiently for a group which is located both anti- and coplanar to a lone pair of electrons of an adjacent atom. This hypothesis has received substantial experimental support and may predict both the transition state geometry and the chemical outcome of chemical reactions (Deslongchamps, 1993).

Extension of ALPH to biological processes led Post and Karplus (PK) to propose an alternative catalytic mechanism for lysozyme. On the basis of molecular dynamics, stereochemical and stereoelectronic considerations, an opening of the sugar ring would be involved as the route of catalysis by HEWL (Post and Karplus, 1986). Catalysis proposed by PK inherently proceed without positive charge buildup at the anomeric centre, which does not correlate with experimental evidence such as kinetic isotope effects (KIEs) (Sinnott, 1993); α-secondary KIEs and direct KIEs for example are not explained by PK's mechanism. Moreover, PK's mechanism predicts a linearized saccharide linked to the enzyme's nucleophile as the catalytic intermediate, whereas the saccharide observed at this catalytic step is a glycopyranosyl group, which is cyclic (White et al., 1996). This experimental evidence questions the involvement of the opening process. For these reasons, the mechanism proposed by Post and Karplus will not be considered further here.
For β-glycosyl hydrolases such as lysozyme, Deslongchamps firmly stated that the saccharide at the cleavage site must first adopt a boat conformation (Figure 1.4) so to fulfill stereoelectronic requirements (Deslongchamps, 1983; Deslongchamps, 1993). Such a boat conformation may be involved in catalysis.

A bacterial enterotoxin is an example of a saccharide-binding protein which constrains a distortion onto its ligand. The 2.1 Å resolution crystal structure of the heat-labile enterotoxin from *Escherichia coli* complexed to the tumor marker disaccharide D-Gal-β-1,3-GalNAc reveals one sugar unit as a chair and the other as a boat conformation (van den Akker et al., 1996). The geometric rearrangement of the N-acetylgalactosyl residue allows the attached T-antigen protein moiety to be directed toward the outside of the binding pocket of this bacterial toxin.

**Figure 1.4. β-D-glucopyranose boat conformation.** Although a chair glucose has its hydroxyl groups equatorial (nearly horizontal) at positions C1, C2, C3 and C4, the higher energy boat form presents its C2 hydroxyl substituent, its β oxygen at C1 and one of the lone pairs of electrons (shaded area) at O5 all axial (vertical). These two latter chemical constituents are now positioned anti-periplanar; such an alignment is not possible in a chair conformation for a β anomer.

Crystallographic evidence also suggest that the saccharide at position -1 may distort upon binding to retaining β-glycosyl hydrolases. The unusual boat form is observed for sialic acid (N-acetylneuraminic acid) in the influenza virus neuraminidase (Varghese et al., 1992). The carbohydrate distortion in this enzyme may be attributed to the carboxylic substituent at C1 which forms stabilizing electrostatic interactions with a triarginyl cluster (residues 118, 292, and 371) and advantageous packing with Tyr106. A third example of the involvement of a boat conformation in binding to a protein concerns the substrate di-N-acetylglucosamine (di-GlcNAc) binding to a
Serratia marcescens chitobiase (Tews et al., 1996). The crystal structure of this bacterial β-glycosyl hydrolase reveals a (β/α)8-barrel which predictably binds its substrate at the carboxyl-terminal end of the barrel. A stable complex of chitobiase with di-GlcNAc has been obtained with a short ten minute incubation followed by flash-freezing at 100 K. The N-acetyl group of the second GlcNAc sugar is believed to participate in catalysis as the reaction's nucleophile and to act on the C1 involved in a boat conformation. In addition to suggesting such an anchimeric assistance by the substituent at C2, this crystal structure refined to 1.9 Å resolution shows a non-hydrolyzed disaccharide at positions1 +1, -1, in which the sugar unit at position -1 forms a boat. A fourth example has recently been presented at a meeting (Davies, 1996). Davies presented the crystal structure of the retaining β-glycosyl hydrolase EGI in complex with a thio substituted non-hydrolyzable tetrasaccharide inhibitor. The sugar unit at the position -1 is constrained to adopt a boat form. Details of this complex have not yet been published. These examples suggest that β-glycosyl hydrolases may impose a significant deformation, a boat, to their incoming saccharide ligand or to their product at the binding subsite position -1.

Saccharide distortion involving the anomeric centre has a catalytic significance. The boat conformation of the incoming saccharide at position -1 correlates with the stereoelectronic requirements predicted by ALPH, such that the leaving oxygen of the aglycone is now found antiperiplanar to a lone pair of electrons of the cyclic oxygen O5 (Figure 1.4). The progression of this strained conformation to a chair form of the intermediate must proceed through a planar arrangement. This inference correlates well with the wealth of data suggesting the involvement of an oxocarbenium ion-like species, i.e. a sofa conformation as previously described. Hence during catalysis by retaining β-glycosyl hydrolases, the substantial dynamics of the anomeric centre of the saccharide at position -1 may be simplified as the pattern shown in Figure 1.5.

---

1 Binding subsites are numbered according to their position relative to the site of cleavage; this cleavage is defined to occur between subsites -1 and +1. Additional saccharide units which bind at the non-reducing end of subsite -1 are identified with negative numbers. Conversely, positive numbers are assigned to saccharide units which bind to the reducing end side of the substrate.
REVISED CATALYTIC MECHANISM OF $\beta$-GLYCOSIDASES

Whether the catalytic intermediate is a long lived oxocarbenium ion-like residual saccharide or a covalent glycosyl-enzyme species has led to a long lasting debate. This question on the nature of the intermediate had a turning point through combining both the oxocarbenium ion-like and covalent glycosyl-enzyme species as separate steps of a united catalytic mechanism (Sinnott, 1987). Hydrolysis by retaining $\beta$-glycosyl hydrolases, which includes lysozyme, would proceed through the formation and hydrolysis of a covalent glycosyl-enzyme intermediate via oxocarbenium ion-like transition states. This process redefines the involvement of the two essential key carboxylates/carboxylic acids, as one carboxyl group is assigned to an acid/base catalyst function (Glu35 in HEWL) whereas the other is assigned to a catalytic nucleophile function (Asp52 in HEWL). Because it has not received direct evidence, the nature of the intermediate is still speculative at this point. Taken together, evidence suggests the catalytic mechanism presented in Figure 1.6 with the following steps:

1- **Substrate binding:**

   (a) The active form of the enzyme contains a deprotonated nucleophile (Asp52 in HEWL) and a protonated acid/base catalyst (Glu35 in HEWL); The substrate may occupy 3 to 8 binding subsites of which the saccharide unit at position -1 adopts a "twisted-boat" conformation;

   (b) Through a direct hydrogen bond, the acid/base catalyst (Glu35 in HEWL) donates a proton to the linking oxygen of the scissile bond, which leads to the weakening of the glycosidic bond C1–O;
2– Bond cleavage and first transition state:

(a) The saccharide at the anomeric centre at the side of the scissile glycosidic bond (subsite D in HEWL) becomes distorted from its chair conformation towards a planar shape (the so-called "sofa" conformation);

(b) The glycosidic bond C1–O is cleaved, which leads both to the release of the aglycone (GlcNAc–GlcNAc from subsites E and F in HEWL) and to the development of a positive charge on the saccharide's newly reducing end (subsite D in HEWL);

(c) Concomitantly this positively charged saccharide becomes stabilized by adopting a full planar conformation at C5–O5–C1–C2 (the "sofa" conformation), in which a partial double bond is formed by the contribution of a lone pair of electron of O5; this oxocarbenium ion-like species is believed to be a transition state of short lifetime;

3 Intermediate:

The residual charged saccharide collapses to form a covalent glycosyl-enzyme intermediate of inverted anomic configuration, in which the saccharide adopts a "full chair" conformation; these two species (intermediate and transition state) are in equilibrium, favouring the more stable covalent form, but being sufficiently unstable to be reactivated to the next transition state;

4– Second transition state:

An incoming water molecule inserts at the re face of the reactive oxocarbenium ion-like and in the vicinity of the acid/base catalyst; the covalent glycosyl-enzyme intermediate reenters a transition state conformation;

5– Product formation:

The acid/base catalyst abstracts a proton from that water molecule concomitant to the nucleophilic addition to the reactive saccharide at the equatorial position of C1; the stereochemistry of the addition of the hydroxyl group leads to the overall retention of the saccharide's anomic configuration.
Figure 1.6. Proposed catalytic mechanism for retaining $\beta$-1,4-glycosyl hydrolases. A covalent $\alpha$-glycosyl-enzyme intermediate is formed and then hydrolyzed via oxocarbonium ion-like transition states. The binding saccharide conformation is presumed to undergo conformational dynamics, for stereoelectronic considerations; this saccharide binds as a twisted-boat (1), deforms to a half-chair at the first transition state (2), relaxes to a chair at the covalent intermediate (3), deforms again to a half-chair at the second transition state (4), and yields a twisted-boat conformation as the product (5).
REFERENCES


Bartik, K., Redfield, C. and Dobson, C. M. (1994) Measurement of the individual pK\textsubscript{a} values of acidic residues of hen and turkey lysozymes by two-dimensional \textsuperscript{1}H NMR. *Biophys. J.* **66**, 1180-1184.


Foundation for Biotechnical and Industrial Fermentation Research, Espoo, Finland) pp. 63-72.


Chapter 2

Crystal Structure of the Catalytic Domain of
the β-1,4-glycanase Cex from Cellulomonas fimi

This chapter describes the high resolution (1.8 Å) crystal structure of a bacterial enzyme. This structure is determined using the classical method of multiple isomorphous replacement with heavy atoms. At the submission of this work for publication, there was no other example of homologous protein structure available in the literature. This chapter is adapted from a paper which appeared under the same title [White, A., Withers, S. G., Gilkes, N. R., & Rose, D. R. (1994) Biochemistry 33:12546-12552]. Reprinted in part with permission from Biochemistry (1994) 33:12546-12552, ©1994 American Chemical Society.
ABSTRACT

\[ \beta-1,4\text{-glycanases, principally cellulases and xylanases, are responsible for the hydrolysis of plant biomass. The bifunctional \( \beta-1,4\text{-xylanase/glucanase} \) Cex from the bacterium \textit{Cellulomonas fimi}, one of a large family of cellulases/xylanases, depolymerizes polysaccharides and releases a disaccharide unit from the substrate non-reducing end. Hydrolysis occurs with net retention of the anomeric configuration of the sugar through a double-displacement mechanism involving a covalent glycosyl-enzyme intermediate. The active site nucleophile, \text{Glu}233, has been unambiguously identified by trapping of such an intermediate [Tull et al. (1991) \textit{J. Biol. Chem.} 266, 15621-15625] and the acid-base catalyst, \text{Glu}127, by detailed kinetic analysis of mutants [MacLeod et al. (1994) \textit{Biochemistry} 33, 6371-6376]. However, little is known about the enzyme's overall folding motif and its active site architecture. We report here the high resolution crystal structure of the catalytic domain of Cex. The atomic structure refinement results in a model that includes 2400 protein atoms and 171 water molecules, with an \( R \)-factor of 0.201 for data extending to 1.8 Å resolution. The protein forms a \( \beta/\alpha\)-barrel which consists of eight parallel \( \beta \)-strands, which is a novel folding motif pattern for a microbial \( \beta \)-glycanase. The active site, inferred from the location of \text{Glu}233, \text{Glu} 127, and of other conserved residues, is an open cleft on the carboxy-terminal end of the \( \beta/\alpha\)-barrel. An extensive hydrogen-bonding network stabilizes the ionization states of the key residues; in particular, the \text{Asp}235-\text{His}205-\text{Glu}233 hydrogen-bonding network may play a role in modulating the ionization state of \text{Glu}233 and in controlling local charge balance during the reaction. \]
**INTRODUCTION**

Cellulases (β-1,4-glucanases) and β-1,4-xylanases are examples of a broad range of glycoside hydrolyzing enzymes produced by microorganisms that make use of plant biomass as a source of carbon and energy. They are usually modular enzymes composed of two or more discrete structural and functional units. Typically, in enzymes derived from both bacteria and fungi, a catalytic domain is joined to a substrate-binding domain by a short linker sequence; additional domains are sometimes involved (Gilkes et al., 1991b). The catalytic domains of over 150 known β-1,4-glucanases and β-1,4-xylanases can be classified into a few distinct families on the basis of amino acid sequence similarities (Henrissat and Bairoch, 1993). Comparative analysis suggests that, despite their apparent diversity, all these enzymes arose by mutation and shuffling of relatively few progenitor modules (Henrissat and Bairoch, 1993).

*Cellulomonas fimi* is a gram positive cellulolytic bacterium that synthesizes at least six β-1,4-glucanases (Xyl A, Xyl B, Xyl C, Xyl D, Cbh I and Cbh II) and four β-1,4-xylanases (Cex and Xyl A) (Meinke et al., 1994; Shen et al., 1994). One of these, Cex, also known as XynB (Hazlewood and Gilbert, 1993), is a 47.1 kDa enzyme comprised of an NH₂-terminal catalytic domain (cex-cd) and a COOH-terminal cellulose-binding domain (CBDcex) joined by a glycosylated linker polypeptide (residues 315-334) containing 20 alternating proline and threonine residues (O’Neill et al., 1986) (Figure 2.1). In addition, a 42 amino acid-long signal peptide is released through the process of secretion of the mature Cex protein (O’Neill et al., 1986).

![Figure 2.1. Cex gene product. Disulfide bridges are represented by S—-S.](image-url)
On the basis of heterologous expression of each of the domains of Cex in the bacterium *Streptomyces lividans*, it has been observed that the linker region is the primary site of glycosylation with 4 to 12 units of D-mannose and α-D-galactoside (Langsford *et al.*, 1987; MacLeod *et al.*, 1992; Ong *et al.*, 1994). Furthermore, this glycosylation is believed to protect Cex from proteolysis and increases this enzyme's affinity for bacterial microcrystalline cellulose (Ong *et al.*, 1994). Small-angle X-ray scattering analysis indicates that the structure of Cex resembles a tadpole in which the catalytic domain corresponds to the globular head and the cellulose-binding domain to the extended tail (Gilkes, N.R. *et al.*, unpublished data). Cex was first recognized as a β-1,4-glucanase, and analysis of carboxymethyl cellulose hydrolysis showed that its action was predominantly exohydrolytic (Gilkes *et al.*, 1984). Subsequent analysis revealed that its catalytic efficiency on p-nitrophenyl xylobioside was about 50-fold higher than on the cellobioside analog (Gilkes *et al.*, 1991a). This preference is a common feature of all the enzymes currently known in the family 10 (formerly family F) of glycosyl hydrolases (Gilkes *et al.*, 1991b; Henrissat and Bairoch, 1993). The specificity of cex-cd for both β-1,4-xylosidic and glucosidic bonds is of interest not only mechanistically but also because of the importance of xylanases in various biotechnological processes. Some applications, such as enzymatic prebleaching of Kraft pulps for paper manufacture, require xylanase enzymes that are free of β-1,4-glucanase activity, while others would benefit from enzymes with dual specificities (Wong and Saddler, 1992). In addition to providing an atomic basis for its catalytic mechanism, the determination of the structure of cex-cd should yield clues as to how the specificity for each substrate is controlled; this would ultimately allow the engineering of such enzymes to change their substrate preference.

Kinetic studies on cex-cd have provided good evidence for a double-displacement catalytic mechanism in which a glycosyl-enzyme intermediate is formed via the attack of a nucleophilic side-chain and is then hydrolyzed to release the product with the same anomeric configuration as the substrate (Sinnott, 1990). Both the formation of the intermediate (glycosylation) and its hydrolysis (deglycosylation) require an acid/base catalytic assistance and both proceed via oxocarbonium ion-
like transition states (Withers et al., 1986; Tull and Withers, 1994). The nucleophilic residue was identified as Glu233 by trapping of the glycosyl-enzyme intermediate using a 2-deoxy-2-fluorocellobioside substrate analog, followed by proteolysis of the complex, and sequencing of the labeled peptide (Tull et al., 1991). The intermediate that was stably trapped in these experiments was shown to be catalytically competent, as the addition of sugar acceptors such as cellobiose resulted in turnover via transglycosylation, the reverse reaction of the enzyme (Tull et al., 1991). Site-directed mutagenesis at Glu233 has yielded mutants with kinetic properties fully consistent with this role (MacLeod, 1994). Mutations at Glu127, followed by detailed kinetic analysis, confirmed its role as an acid/base catalyst (MacLeod et al., 1994). Determination of the structure of the Cex catalytic domain at high resolution provides an overall picture of the folding motif of the enzyme and details of the intricate features of the active site, including the positions of Glu233, Glu127, and other residues involved in substrate binding and enzyme catalysis. This provides a molecular basis for future mutagenesis studies, and will add to our understanding of the mechanism of this family of retaining β-glucanases.

Three-dimensional structures of known bacterial and fungal β-glucanases and xylanases show a remarkable variety of folding motifs. The Trichoderma reesei Cbh II exo-glucanase forms a parallel seven stranded α/β-barrel (Rouvinen et al., 1990), which is also seen in the thermophilic endocellulase E2 catalytic domain from Thermomonospora fusca (Spezio et al., 1993); the Humicola insolens endoglucanase V folds into a six stranded β-barrel (Davies et al., 1993; Davies et al., 1995; Davies et al., 1996); the endoglucanase CelD from Clostridium thermocellum forms an (α/α)₆-barrel (Juy et al., 1992); the Bacillus subtilis and Trichoderma harzianum 20 kDa xylanases fold into a sandwich of two β-pleated sheets (Campbell et al., 1993) (D. R. Rose and R. Campbell, in preparation), a folding motif similar to that of the hybrid Bacillus 1,3-1,4-β-glucanase (Keitel et al., 1993). Until the acceptance of this chapter for publication (White et al., 1994), there was no reported crystal structure of a family 10 β-1,4-glycanase, though two other crystallization reports had been available (Pickersgill et al., 1993; Souchon et al., 1994). Here we report the high resolution crystal structure of the catalytic domain of the β-1,4-glycanase Cex.
MATERIALS AND METHODS

CRYSTALLIZATION AND DATA COLLECTION

The cex-cd was obtained by limited proteolysis with papain of recombinant Cex protein (Gilkes et al., 1988) and then crystallized as previously described (Bedarkar et al., 1992). These crystals typically reach a size of 0.7 mm x 0.7 mm x 0.5 mm. Two useful heavy atom derivatives were obtained by soaking protein crystals in 15% (w/v) PEG 4000, 100 mM sodium acetate pH 4.6, and either 20 mM trimethyl lead acetate (TMLA) for 30 hr or 0.5 mM (COOCH₃)₂Hg for 3 hr followed by a 21 hr backsoak. Each data set was measured from a single crystal at room temperature on a Xuong-Hamlin area detector connected to a Rigaku RU-200 rotating anode X-ray generator operated at 40 kV and 150 mA. Oscillation frames of 0.1° on ω were measured, from which data up to 1.8 Å resolution were integrated and then reduced using the software by Howard and Neilsen (Howard et al., 1985). Bijvoet pairs for both derivative data sets were measured in order to obtain anomalous scattering data.

ISOMORPHOUS REPLACEMENTS

The positions of the heavy atoms were determined from difference Patterson maps and subsequently refined using the software package PHASES (Furey and Swaminathan, 1990); the single site for the lead derivative was solved first and then used along with its anomalous scattering signal to calculate phases, in order to locate the positions of the mercury atoms by difference Fourier synthesis. The crystallographic tetragonal space group ambiguity was resolved by observing the lowest residual obtained for the combined refinement of both heavy atoms, in either P4₁₂₂ or P4₃₂₁₂. The crystals belong to the space group P4₁₂₁₂. The heavy atom positions of both derivatives were refined together and their phases combined at 2.3 Å resolution using the package PHASES. An automated solvent flattening procedure (Wang, 1985) as implemented in the program PHASES was carried out to improve the combined protein phases. The heavy atom positions were then refined again, this time against the phases obtained following the solvent
flattening algorithm. Protein phases to 2.3 Å resolution were determined and used to generate an electron density map.

**STRUCTURAL ANALYSIS AND MODEL BUILDING**

A primary tracing of the protein was obtained using the program BONES, as implemented in the program MAPMAN (Kleywegt and Jones, 1994). This trace was edited and used as a guide to build a polymer of Cα into the electron density map. All the model building and graphical rebuilding was carried out on a Silicon Graphics Indigo, using the program O version 5.9.1 (Jones et al., 1990; Jones et al., 1991). A molecular model of residues 3 to 312 of cex-cd was then interpreted from the electron density map, using the polymer of Cα as a guide and assisted by a library of substructures (Jones and Thirup, 1986). Whereas residues 1 and 2 were located at a later stage, the carboxyl-terminal region beyond Phe312 was not clearly defined. Because the catalytic domain was obtained by proteolysis of the mature form of the enzyme, we cannot predict with certainty the carboxyl-terminal content.

**STRUCTURE REFINEMENT**

All the protein structure refinements were performed using the program X-PLOR version 3.1 (Brünger et al., 1987), in which 10% of the diffraction data were set aside to compute an $R_{\text{free}}$ so as to follow the refinement progress objectively (Brünger, 1992). The topology and parameter values used as ideal geometry are coded in the files topcshdx.cho and paramcshdx.pro, as previously compiled (Engh and Huber, 1991). The initial protein atomic model was first assigned a uniform isotropic temperature factor of 20 Å$^2$ and its stereochemistry was regularized with 40 cycles of energy minimization by conjugated gradient against all 15 to 2.3 Å data with $F > 2\sigma$ as follows. This model was then refined by simulated annealing using the slow-cool protocol with molecular dynamics in which the structure was slowly cooled from 3000 to 300 K in steps of 25 K. At each temperature, 50 cycles of molecular dynamics and energy minimization were performed. An additional positional refinement of 120 cycles was applied, followed by
temperature factor refinement of first an overall and then individual atomic $B$-values. Through this process, a tight stereochemistry was maintained.

An electron density map was computed using the $2mF_{\text{obs}} - DF_{\text{calc}}$ coefficients\(^2\) and phases combined from the model and the experimental MIR, as implemented in the program SIGMAA (Read, 1986). Incorrectly interpreted regions were located by a "residue real space correlation" (Jones et al., 1991) and further corrected with the use of simulated-annealing omit maps ($F_{\text{calc}}$, $\phi_{\text{calc}}$ computed with the omission of less than 10% of the residues). A model rebuilding was done using this electron density map, followed by 120 cycles of positional refinement at 300 K.

Peaks above 3σ in a SIGMAA-weighted $F_{\sigma} - F_c$ map were located using the command PEAKPICK, as implemented in the program MAPMAN (Kleywegt and Jones, 1994). As a first approximation, these peaks were assigned as water molecules. The region of the electron density map for each water molecule was visually examined. Each water molecule, assigned as an oxygen with $sp^3$ hybridization, had to meet the following criteria: acceptable hydrogen bond geometry (angles) with respect to its neighboring non-hydrogen atoms; within hydrogen bond distance of 2.4-3.5 Å to at least one non-hydrogen protein atom or the oxygen of another water molecule; electron density peak $\geq 3\sigma$ in a SIGMAA-weighted $F_{\sigma} - F_c$ map; temperature factor $B \leq 60$ Å\(^2\) after refinement.

Data to 1.8 Å resolution were then included in rounds of positional refinement and model rebuilding. The tight stereochemical constraints were progressively released through these rounds of model rebuilding and refinement. The stereochemistry of the model was verified using the software package PROCHECK (Laskowski et al., 1993). The secondary structure elements were assigned by the criteria of Kabsch and Sander (Kabsch and Sander, 1983), as implemented in PROCHECK. The final model atomic coordinates have been deposited with Brookhaven Protein Data Bank (accession number 2EXO).

\(^2\) The weight $m$ is the figure of merit, whereas $D$ is a coefficient dependent on an estimate of the coordinate errors.
RESULTS AND DISCUSSION

Between 89 and 94\% of the predicted reflections were measured for the three data sets reported in Table 2.1. On average, four intensity measurements contribute to each reflection. Similar unit cell parameters suggest that the incorporation of the heavy atoms used does not perturb the molecular structure of the protein. Hence, these derivatives are assumed isomorphous.

Table 2.1. Diffraction Data Collection and MIRAS phasing

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>(CH$_3$)$_3$PbAc$^a$</th>
<th>HgAc$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a, b$ (Å)$^b$</td>
<td>88.15</td>
<td>88.42</td>
<td>88.21</td>
</tr>
<tr>
<td>$c$ (Å)</td>
<td>81.11</td>
<td>81.27</td>
<td>81.16</td>
</tr>
<tr>
<td>Unit cell volume (Å$^3$)</td>
<td>630259</td>
<td>635377</td>
<td>631506</td>
</tr>
<tr>
<td>Resolution limit (Å)</td>
<td>1.8</td>
<td>2.2</td>
<td>2.3</td>
</tr>
<tr>
<td>No. of observations</td>
<td>128888</td>
<td>151941</td>
<td>200753</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>29059</td>
<td>31654$^d$</td>
<td>24920$^d$</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.4</td>
<td>4.8</td>
<td>8.1</td>
</tr>
<tr>
<td>Data completeness (%) $</td>
<td>94</td>
<td>93</td>
<td>89</td>
</tr>
</tbody>
</table>

$R_{sym}$ (%)$^c$

|                         | 6.1    | 6.0                  | 4.2      |

$^a$ Ac means acetate. $^b$ The unit cell angles $\alpha$, $\beta$, and $\gamma$ are held fixed to 90\(^\circ\), which is inherent to a tetragonal system. $^c R_{sym} = \sum |l_i - <l>| / \sum <l>$, where $l_i$ is the scaled intensity of the $i^{th}$ measurement and $<l>$ is the mean intensity for that reflection. $^d$ The Bijvoet pairs are considered independent reflections for both heavy atom derivative data sets.
MULTIPLE ISOMORPHOUS REPLACEMENT

A single lead site was located from the Harker section \( \langle \frac{1}{2}, v, w \rangle \) of the difference Patterson of the TMLA derivative with the use of the Fourier coefficients \( |F_{isomorphous} - F_{native}|^2 \). The location of the heavy atom is consistent with the Harker section \( \langle \frac{1}{2}, v, w \rangle \) of the difference anomalous scattering signal, with the use of the Fourier coefficients \( |F_+ - F_-|^2 \). The heavy atom position was refined for both isomorphous and anomalous scattering data and results in a combined figure of merit of 0.67 at 3.0 Å resolution. An automated solvent flattening filtering improved the figure of merit to 0.85 for the protein phases; a solvent content of 40% was used, compared to 46% as estimated from the unit cell. These phases were then used to locate the mercury heavy atom sites by difference Fourier synthesis, which identified three major sites (Table 2.2). A combined positional refinement of the lead and mercury heavy atoms yielded an overall combined figure of merit of 0.71 (Table 2.3), which was further improved by solvent flattening. The new phases were then used to re-refine the position of the heavy atoms, followed by solvent flattening. The final figure of merit is 0.88 for the protein phases within the resolution range 15 to 2.3 Å.

Table 2.2. Heavy atom locations

<table>
<thead>
<tr>
<th>Type</th>
<th>Site</th>
<th>x (Å)</th>
<th>y (Å)</th>
<th>z (Å)</th>
<th>q</th>
<th>B (Å²)</th>
<th>Close protein atoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pb</td>
<td>1</td>
<td>50.70</td>
<td>18.98</td>
<td>6.93</td>
<td>1.02</td>
<td>20.0(^a)</td>
<td>Glu76 O(^{\varepsilon1,\varepsilon2}), Asp14 O(^{\delta1,\delta2})</td>
</tr>
<tr>
<td>Hg</td>
<td>1</td>
<td>11.12</td>
<td>75.44</td>
<td>7.95</td>
<td>1.02</td>
<td>39.22</td>
<td>His85 N(^{\varepsilon}), Asp138 O(^{\delta2}, O)</td>
</tr>
<tr>
<td>Hg</td>
<td>2</td>
<td>18.91</td>
<td>67.56</td>
<td>1.23</td>
<td>1.00</td>
<td>13.45</td>
<td>His80 N(^{\varepsilon}), Glu233 O(^{\varepsilon1,\varepsilon2})</td>
</tr>
<tr>
<td>Hg</td>
<td>3</td>
<td>21.27</td>
<td>67.26</td>
<td>3.63</td>
<td>0.95</td>
<td>14.44</td>
<td>His205 N(^{\varepsilon}), Glu233 O(^{\varepsilon1})</td>
</tr>
</tbody>
</table>

\(^a\) The thermal factor B of the lead derivative was not refined but the occupancy was.
Table 2.3. **MIRAS phasing**

<table>
<thead>
<tr>
<th>Heavy atom</th>
<th>Conc'n (mM)</th>
<th>Soaking (h)</th>
<th>No. sites</th>
<th>( R_{\text{der}}^a ) (%)</th>
<th>( R_{\text{Cullis}}^b ) (%)</th>
<th>Phasing</th>
<th>Phasing power ( ^c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(CH(_3)(_3))PbAc</td>
<td>15</td>
<td>30</td>
<td>1</td>
<td>8.6</td>
<td>60</td>
<td>1.70</td>
<td>1.42</td>
</tr>
<tr>
<td>HgAc(_2)</td>
<td>0.5</td>
<td>3</td>
<td>3</td>
<td>5.0</td>
<td>61</td>
<td>1.89</td>
<td>1.53</td>
</tr>
</tbody>
</table>

\( ^a R_{\text{der}} = \sum |F_{\text{PH}} - F_{\text{PI}}| / \sum (F_{\text{P}} + F_{\text{PH}}) \), where \( F_{\text{P}} \) and \( F_{\text{PH}} \) are the native and derivative structure factor amplitudes respectively. \( ^b R_{\text{Cullis}} = \sum (|F_{\text{PH}}| - |F_{\text{PI}}|) / \sum |F_{\text{PH}}| + |F_{\text{PI}}| \) for centric reflections. \( ^c \) Phasing power is the average of \( <F_{\text{P}}>_\varepsilon \) values, where \( <F_{\text{P}}>_\varepsilon \) is the mean amplitude of the heavy atom structure factor of Miller indices \( \mathbf{h} \) for the data within a resolution shell, and \( \varepsilon \) is the r.m.s. lack of closure error.

**REFINEMENT AND QUALITY ASSESSMENT OF THE CURRENT CEX-CD MOLECULAR MODEL**

The refinement by simulated annealing of 2400 non-hydrogen atoms, *i.e.* 312 residues, of the native cex-cd, improved the \( R \)-factor from 38.4 to 24.2\% at 2.3 Å resolution. The resolution was increased to 1.8 Å and 171 water molecules were added to the model, followed by rounds of model building and refinement by X-PLOR. The current value of the \( R \)-factor is 20.1\% and the \( R_{\text{free}} \) is 28.2\%, for the resolution range 8 to 1.8 Å and \( F > 2\sigma \) (Table 2.4). For the current model, the root-mean-square deviations from ideality are 0.017 Å for bond distances, 2.0° for bond angles, and 1.8° for "improper" angles. The average temperature factors are 22 Å\(^2\) for main chains, 23 Å\(^2\) for side chains and 38 Å\(^2\) for the water molecules. The "residue real space correlation"\(^3\) for the model has an average value per residue of 0.88 (\( \sigma = 0.03 \)), with the lowest values between 0.75 and 0.80 for Glu6, Asn96, Ser98, Lys188, and Thr245. Regions of low correlation also show large \( B \)-values indicating that they are part of flexible regions of the structure (Figure 2.2). Analysis of the stereochemistry of the model with PROCHECK (Laskowski *et al.*, 1993) shows concordance with other protein structures determined at comparable resolution. For example, among the non-glycine and non-proline residues, 252 (91\%) have their \( \phi, \psi \) angles in the

---

\(^3\) The real-space-correlation coefficient is defined as the correlation of the model-derived \( F_C \) map and the \( 2F_C - F_C \) electron density maps, computed on a residue to residue basis. (Jones *et al.*, 1991)
"most favoured" regions of the Ramachandran plot, whereas the remaining 24 (9%) are in the "allowed" regions (Figure 2.3).

Table 2.4. **Crystallographic Refinement Statistics**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. non-hydrogen atoms</td>
<td>2400 (residues 1 to 312)</td>
</tr>
<tr>
<td>No. water molecules</td>
<td>171</td>
</tr>
<tr>
<td>R.m.s. deviation, bond length (Å)</td>
<td>0.017</td>
</tr>
<tr>
<td>R.m.s. deviation, bond angle (deg)</td>
<td>2.01</td>
</tr>
<tr>
<td>R.m.s. deviation, improper angles (deg)</td>
<td>1.84</td>
</tr>
<tr>
<td>R.m.s. deviation, dihedral angles (deg)</td>
<td>24.1</td>
</tr>
</tbody>
</table>

Resolution (Å)  
8.0 to 2.83  2.83 to 2.26  2.26 to 1.98  1.98 to 1.80

Reflections used for refinement:

| No. of reflections (F>2σ) | 6664  6028  5491  4115 |
| R-factor (shell)          | 0.175  0.226  0.232  0.238 |
| R-factor (cumulative)     | 0.175  0.190  0.197  0.201 |

Reflections set aside from refinement:

| No. of reflections (F>2σ) | 755  691  566  453 |
| R_free (shell)            | 0.245  0.316  0.315  0.353 |
| R_free (cumulative)       | 0.245  0.266  0.274  0.282 |

\[ a \text{ R-factor} = \frac{\Sigma |F_{\text{obs}}| - |F_{\text{calc}}|}{\Sigma |F_{\text{obs}}|} \text{ where } |F_{\text{obs}}| \text{ and } |F_{\text{calc}}| \text{ are the observed and calculated structure factor amplitudes respectively.} \]

**STRUCTURAL FEATURES**

The catalytic domain of Cex is a globular protein rich in secondary structure. There are three single-turn α-helices, and ten longer ones. Alternating with the α-helices are ten β-strands, of which eight are parallel and form the elliptical central core of a (β/α)-barrel (Figure 2.4). Among the secondary elements, the long β-strand β7 makes an excursion out of the barrel to form
Figure 2.2. Mobility of the residues and agreement of the model, as evaluated with the main chain real-space correlation (continuous line) and temperature factor ($B$) profiles (dashed line). The real-space correlation (Jones et al., 1991) for the main chain of each residue of the complexed cex-cd model was evaluated against $2F_o-F_c$ electron density maps, in which phases were derived either from the final model.

...a short antiparallel $\beta$-sheet with both $\beta$6b and $\beta$9. The amino- and carboxy-terminal $\alpha$-helices, $\alpha$0 and $\alpha$9, are adjacent and parallel to each other. Disulfide bridges were predicted to join Cys167 with Cys199 (CC1) and Cys261 with Cys267 (CC2) (Gilkes et al., 1991a). Both disulfide bridges are now confirmed; CC1 joins the strands $\beta$5 and $\beta$6a, while CC2 constrains the residues 264-266 to form a turn connecting $\alpha$7 to $\beta$8.
Figure 2.3. Main chain ($\phi, \psi$) angles analysis (Ramachandran Plot). Glycines (filled triangles) and non-glycine (filled squares) residues are shown, except for the prolines. The denser the shading, more favorable a conformation of the main chain of the corresponding amino acid. This preferential distribution is based on 118 structures of resolution of at least 2 Å and $R$-factor no greater that 20% as implemented in the algorithm PROCHECK (Laskowski et al., 1993).

There are two cis-peptides in the three-dimensional structure of cex-cd. The peptide bond Thr240-Pro241 is located in a turn between strand $\beta7$ and helix $\alpha7$ and adopts a cis conformation;
the cis-Pro241 is stabilized by Lys246 Nε, which sits within hydrogen bond distance of the backbone oxygens of Arg239 and Pro241. The interpretation of this region is consistent with a simulated-annealing omit map, in which residues 239 to 245 were omitted. Although a cis-peptide is energetically less favorable for residues other than proline, the peptide bond His80-Thr81 also adopts the cis conformation; this electron density map interpretation was confirmed by a simulated-annealing omit-map in which the residue 81 and its neighboring atoms within 6.5 Å distance were left out while the remaining of the model was subjected to simulated annealing, as proposed by the group of Brünger (Hodel et al., 1992). Because the carbonyl oxygen of His80 would otherwise clash with the carbonyl of Asp123, it is strained and its angles ψ and ω (peptide bond rotation angle) are rotated by about 180° thus producing a cis-peptide. The His80 backbone O position is now stabilized by one water molecule in a favorable hydrogen bond position. The backbone of residues 80 to 82 and 123 form a "special β-bulge" of type "SP3" (Chan et al., 1993), a conformation which incorporates the unusual orientation of the backbone of cis-Thr81. Although infrequently observed (Stewart et al., 1990; MacArthur and Thornton, 1991), non-proline cis peptides are often associated with specific functions such as ligand binding or positioning of catalytic residues (Herzberg and Moult, 1991). So it is possible that the steric strain on the peptide bond His80-Thr81 confers a preferable local environment for catalysis and/or substrate binding.

FOLDING MOTIF

The catalytic domain of Cex folds into a (β/α)8-barrel that contains an elliptical core of eight parallel β-strands (Figure 2.4). First observed in triose phosphate isomerase (TIM) (Banner et al., 1975), the β/α-barrel is a common folding pattern for enzymes with a wide variety of different activities; these enzymes have been postulated to be related by divergent evolution (Farber and Petsko, 1990). The folding motif of the family 10 enzyme cex-cd is distinct from the reported crystal structures of enzymes of other families of bacterial and fungal β-glucanases and xylanases. The cex-cd architecture differs from family 6 (previously family B) enzymes CbhII and E2cd (Rouvinen et al., 1990; Spezio et al., 1993), the most closely related microbial β-glucanase
Figure 2.4. The folding motif of cex-cd, which forms an $\beta/\alpha$-barrel, as drawn by the program MOLSCRIPT (Kraulis, 1991). a) View along the core of the $\beta$-barrel. Residue numbers refer to the acid/base catalyst Glu127 and the nucleophile Glu233, whereas N and C identify the amino- and carboxy-termini respectively. The secondary structures locations are $\alpha0$ (4-10), $\beta1$ (14-19), $\alpha1$ (27-36), $\beta2$ (39-42), $\alpha2a$ (48-51), $\alpha2b$ (61-73), $\beta3$ (76-83), $\alpha3a$ (90-93), $\alpha3b$ (97-114), $\beta4$ (121-126), $\alpha4a$ (140-145), $\alpha4b$ (149-160), $\beta5$ (165-170), $\alpha5$ (178-193), $\beta6a$ (199-202), $\beta6b$ (205-207), $\alpha6$ (215-223), $\beta7$ (228-239), $\alpha7$ (244-262), $\beta8$ (267-272), $\alpha8$ (282-285), $\beta9$ (289-290) and $\alpha9$ (304-311). For clarity, the $\beta$-strands are not identified. b) A side view of cex-cd.

topologies reported to date, for the following two reasons: all the $\beta$-strands of the cex-cd barrel are linked from one to the next by at least one $\alpha$-helix, whereas neither CbhII nor E2$_{cd}$ includes an $\alpha$-helix between $\beta VI$, $\beta VII$, and $\beta VIII$, and the cex-cd $\alpha/\beta$-barrel core includes an eighth parallel $\beta$-strand, unlike those of CbhII and E2$_{cd}$. In addition, the cex-cd active site crevice appears to be more accessible in comparison to the tunnel-like CbhII cleft, even though both enzymes exhibit exo activity. Moreover, because of its eight-stranded $\beta/\alpha$-barrel and its ten $\alpha$-helices, the cex-cd is unlikely to be related to the single-helix six-stranded $\beta$-barrel of EngV (Davies et al., 1993), which is a member of family 45 (previously family K). The structure of the family 10 cex-cd more closely
resembles the plant (1-3)- and (1-3,1-4)-β-glycanases reported by Varghese et al. (Varghese et al., 1994) as well as α-amylase (Luo et al., 1994), all of which adopt a (β/α)8-barrel architecture.

**ACTIVE SITE**

The active site location is inferred from previously identified catalytic residues Glu233 and Glu127 and other conserved residues. It is a significant deep crevice which spans over 30 Å on the protein surface. This crevice is situated at the carboxyl-terminal end of the β/α-barrel. Interestingly, this end of the barrel is the preferential location for the active site of more than 20 enzymes (Farber and Petsko, 1990; Brädén, 1991; Farber, 1993).

The catalytic residues nucleophile Glu233 and acid/base catalyst Glu127 are suitably disposed within the crevice of the active site, with their carboxyl groups facing together at a separation of 5.5 Å. This separation is similar to those seen in other "retaining" β-glycanases whose structures have been solved: 5.4 Å between Glu78 and Glu172 for the *Bacillus circulans* xylanase (Campbell et al., 1993) and 5.5 Å between Asp52 and Glu35 of hen egg white lysozyme (Imoto et al., 1972). Such a separation is presumably optimal for the efficient formation of a glycosyl-enzyme intermediate on Glu233 of cex-cd while at the same time allowing Glu127 to protonate the departing aglycone in a concerted manner. A greater separation between active site carboxyl groups is observed for the inverting β-glycosidase E2, an endocellulase from *Thermomonospora fusca* (Spezio et al., 1993), in which 9.5 Å separates the acid catalyst Asp117 and Glu265; this distance is consistent with the different mechanism in which a water molecule together with the substrate must be located between the two carboxyl groups.

A hydrogen-bonding network involving the catalytic residues is found at the active site, and is likely important in determining their ionization behavior during catalysis (Figure 2.5). The nucleophile, Glu233, probably exists in a deprotonated form in the free enzyme to facilitate attack at the anomeric centre of the saccharide substrate. Glu233 is within hydrogen bond distance of

---

4 Because the carboxylate groups of the two catalytic residues are the actual active site catalysts, the distance is measured between the oxygen atoms of these carboxylates.
Figure 2.5. Active site of cex-cd. (a) Stereoview displayed with the program O of a representative section of the electron density map $2mF_{o} - DF_{o}$ contoured at 2.5σ. The view focuses on the active site of cex-cd. (b) Stereo diagram of the environment of the catalytic residues Glu127 and Glu233 in the active site of cex-cd (subset of a), as drawn by the program SETOR (Evans, 1993). For clarity, only selected side chains are shown along with their respective Cα (circles). Probable hydrogen bonds (dashed lines) and their lengths (Å) are indicated.
both His205 and Asn169. His205 in turn interacts with Asp235, another conserved residue. Interestingly, in only one of the twenty or more enzymes from this family, ORF4 from *Caldocellum saccharolyticum* (Luthi *et al.*, 1990), this aspartic acid residue is not conserved; it is replaced by a threonine. However, a compensating replacement of the (otherwise) conserved His205 by an asparagine maintains charge equivalence. As this trio of residues (Glu233-His205-Asp235) is highly conserved within the family, we postulate that it plays an important role in maintaining the ionization state of the nucleophile.

By contrast, the acid catalyst Glu127 must be in its protonated state for the first step in catalysis. This ionization state of Glu127 may be facilitated by interactions with Asn126, Gln203, Trp84 and a water molecule (Figure 2.5b). Formation of the glycosyl-enzyme intermediate on Glu233 will affect the ionization behavior of Glu127 by decreasing the local charge density and possibly by altering the hydrogen-bonding network. As a consequence, Glu127 should be in an appropriate state for its role as a base in the second step of the mechanism.

A fourth conserved carboxylic amino acid located near the active site is Asp123 (Figure 2.5a). This residue is hydrogen-bonded to the active site conserved residue His80 which itself may be involved in substrate recognition. MacLeod noted that mutants in which Asp123 is replaced by Ala have considerably reduced catalytic activity (MacLeod *et al.*, 1996), suggesting either some direct role of Asp123 in catalysis or some important role in modulating other interactions. A direct role in catalysis is unlikely, because the Asp123 side chain is not exposed in the active site. From the structure, it appears that replacement of this residue by Ala compromises the ability of His80 to interact with the substrate. Interestingly, much of the lost activity could be restored to the mutant by addition of exogenous azide. The structure would suggest that this restoration is due to azide binding at the site of the missing side chain and satisfying local charge requirements, thereby enabling His80 to fulfill its role better. Similar results have been obtained with mutants of trypsin, which could be re-activated by addition of exogenous acetate (Perona *et al.*, 1994).

Inspection of sequence alignments of the proteins within this family identifies a number of insertions that are present in some members. As might be expected, all these insertions are found
to be on the outer surface of the protein, mostly located in loop regions, and are thus easily accommodated. Other critical residues for binding of the polysaccharide chain substrate can be inferred on the basis of their location in the structure and their conservation on sequence alignment. Figure 2.6 illustrates the location of the 27 conserved residues from the sequence alignment of the 20 known members of the family 10 (Gilkes et al., 1991b). Although they are not contiguous in sequence, the side chain of these conserved residues group together in space. They distribute in the following categories: (a) A cluster of five hydrophobic residues form part of a hydrophobic pocket embedded between the α-helices α3b, α4a, α4b, the β-strands β3, β4, β5, and the loop joining α5 to β6a; (b) A core of 18 amino acids which includes both key catalytic carboxylates Glu127 and Glu233 is located at the C-terminal end of the β/α-barrel; (c) Among the remaining scattered conserved residues, Pro89 joins a loop to the α-helix α3a, Gly194 immediately follows the α-helix α5, whereas Asp277 forms a salt bridge to Lys302. Most of these residues are located at or near the active site. This observation suggests a highly similar catalytic mechanism and substrate recognition for these enzymes.

One particularly useful piece of information in trying to identify the polysaccharide binding site and to assess the orientation of the chain within that site, is the recent work on mutants of xylanase A from Streptomyces lividans (Moreau et al., 1994), an enzyme from the same sequence-related family. On the basis of kinetic studies with a mutant in which the conserved residue Asn173 was replaced by Asp, the authors concluded that this residue was involved in important hydrogen-bonding interactions with a xylose residue three units away from the cleavage site, in the direction of the reducing terminus. The structure of cex-cd is in complete agreement with this proposal, since this residue (Asn172 in Cex) is located some 11.3 Å away from the nucleophilic residue Glu 233. Such a distance is consistent with the placement of at least two complete sugar residues between these two amino acid residues, with Asn172 interacting with the third. This observation also suggests that when the scissile bond is located between Glu127 and 233, the reducing end of the polysaccharide is oriented towards Asn172, and thus the non-reducing disaccharide moiety extends towards Gln87, another completely conserved residue.
Figure 2.6. Glycosyl hydrolases family 10 consensus sequence. (a) Shown is the amino acid sequence of cex-cd (line 1), the consensus sequence (cons., line 2), the location of these conserved residues (line 3; \(1\)=active site, \(2\)=hydrophobic cluster, \(3\)=the others), and the secondary structure elements (2nd str., line 3; \(\alpha\)=\(\alpha\)-helix, \(\beta\)=\(\beta\)-strand, \(\cdot\cdot\cdot\)=coiled) (Kabsch and Sander, 1983). The consensus refers to residues that are either identical in all, or all but one, known sequences of the family 10 proteins. (b) Projection in stereo, from the program O, of the C\(\alpha\) of cex-cd with side chains of residues (bold) conserved (consensus) in the family 10.
The active site cleft of cex-cd also contains several tryptophans (84, 273, 281); these three residues extend their side chains into the vicinity of Glu127 and Glu233. The tryptophans are likely to be involved in the stacking of either cellulose or xylan in the active site, but Trp273 and Trp281 may also have a structural role. These two tryptophans delimit the extent of the loop which joins the strand β8 to the helix α8. Analysis of the water molecule distribution (Figure 2.7) reveals that the backbone of this loop is tightly stabilized on the surface of the protein by a well ordered, internal cluster of five water molecules (Figure 2.8). These are within tight hydrogen-bonding distance to several backbone and side-chain atoms within the loop. Furthermore, the middle of this loop is pinned to the rest of the structure by a tight salt bridge between Asp277 and Lys302, two residues that are completely conserved in all members of the family. This interaction further points to an important role for this loop, though the nature of this role has not yet been defined.
Figure 2.8. A cluster of five water molecules which interacts with the loop 273-281. This stereo diagram was generated with SETOR (Evans, 1993). Water molecules are represented by a plus sign associated with a un-lettered number. For clarity, the backbone is shown with circles and the side chains of only the residues that are within 3.5 Å of the cluster of water molecules are shown.

In summary, the folding motif of the catalytic domain of Cex is a (β/α)8-barrel. The active site is now proposed to be located in a crevice in which the nucleophile Glu233 and the acid/base catalyst Glu127 expose their side chains. This crevice is composed of conserved residues, which suggest a conserved mode of substrate binding and possibly of catalytic mechanism for the enzymes with such a conserved sequence.
References


Chapter 3

Crystallographic Observation of a Covalent Catalytic Intermediate in a β-Glycosidase

This chapter describes the crystal structure of a covalent glycosyl-enzyme complex with cex-cd. This species represents the intermediate involved in the catalysis by retaining β-1,4-glycosyl hydrolases; this event was not previously structurally characterized and now provides clues about the catalysis by retaining β-glycosyl hydrolases. The inhibitor synthesis [Tull, D. et al. (1991) J. Biol. Chem. 266:15621-15625], the one dimensional NMR and the kinetic analyses were carried out by Drs. L. P. McIntosh, D. Tull, and S. G. Withers. I also express my thanks to K. Johns for her valuable help with crystal soaking experiments. This chapter is adapted from a manuscript published [White, A., Tull, D., Johns, K., Withers, S. G., Rose, D. R. (1996) Nature Struct. Biol. 3:149-155]. Reprinted with permission from Nature Structural Biology (1996) 3:149-155, ©1996 Nature America Inc.
The three-dimensional structure of a catalytically competent glycosyl-enzyme intermediate of a retaining β-1,4-glycanase has been determined at a resolution of 1.8 Å by X-ray crystallography. A fluorinated slow substrate forms an α-d-glycopyranosyl linkage to one of the two invariant carboxylates, Glu233, as supported in solution by 19F-NMR studies. The resulting ester linkage is coplanar with the cyclic oxygen of the proximal saccharide and is inferred to form a strong hydrogen bond with the 2-hydroxyl of that saccharide unit in natural substrates. The active site architecture of this covalent intermediate gives insights into both the classical double-displacement catalytic mechanism and the basis for the enzyme specificity.
INTRODUCTION

The nature of the catalytic intermediate of retaining β-glycosidases\(^5\) has been a matter of controversy for many years. Koshland first suggested a double-displacement mechanism for this class of enzymes wherein a covalent glycosyl-enzyme intermediate is formed upon cleavage of the glycosidic bond (Koshland, 1953). The transition states for the formation and hydrolysis of that intermediate are known to have substantial oxocarbenium ion character (Sinnott, 1990). On the basis of pioneering crystallographic studies of hen egg-white lysozyme, Phillips and co-workers suggested a variation of this mechanism in which the intermediate is an ion pair of significant lifetime (Phillips, 1967). Support for this ion-pair intermediate has been obtained from a number of studies, ranging from the early determination of the structure of a tri-(N-acetylglucosaminyl lactone)-lysozyme complex (Ford \textit{et al.}, 1974) to more recent refined high-resolution crystal structures with bound oligosaccharides (Strynadka and James, 1991; Song \textit{et al.}, 1994; Turner and Howell, 1995). The structures of these complexes with non-covalently bound glycosides, however, do not explicitly eliminate the possibility of a covalent catalytic intermediate for β-glycosidases.

Evidence from β-glycosidases other than lysozyme has generally favoured a covalent intermediate (Sinnott, 1990; McCarter and Withers, 1994).

1 Kinetic isotope effects\(^6\) measured for each of the two catalytic steps, formation and hydrolysis of the intermediate, are inconsistent with an ion pair intermediate (Sinnott, 1978). Such \(\alpha\)-secondary deuterium kinetic isotope effects provide a measure of the hybridization of the anomeric carbon C-1 of the substrate. Existence of an ion-pair intermediate would require an

\(^5\) Retaining glycosidases hydrolyze a glycosidic O-linkage with net retention of anomeric configuration at the carbon C-1 of the product which has a newly reducing end. For example, retaining β-glycosidases release products with β-anomeric configuration.

\(^6\) \(\alpha\)-secondary deuterium kinetic isotope effects, obtained by comparing the respective rates \(k_{cat}\) of hydrolysis of \([H]\) and \([D]\)-substitution at the atom C1 of the substrate, is indicative of the hybridization of this anomeric carbon C1 for the transition state of the step in question. A positive effect \((k_{cat}/k_0 > 1.0)\) would indicate an increased \(sp^2\) character. Resonance forms for this C1 involve the delocalization of a lone pair of electrons from O5 which itself also adopts a \(sp^2\) character. A necessary implication of the partial double bond between C1 and O5 is the planarity of the ring involving the atoms C5, O5, C1, and C2.
inverse isotope effect \( (k_{H}/k_{D} < 1) \) for the second catalytic step. However, the observed ratio \( k_{H}/k_{D} = 1.11 \) for this catalytic step is consistent only with a tetrahedral geometry at the C-1 of the intermediate (Sinnott, 1990; Tull and Withers, 1994).

(2) Additional evidence for the nature of the catalytic intermediate has come from results with 2-deoxy-2-fluoroglycosides (Withers et al., 1988; Street et al., 1989; Tull et al., 1991; Street et al., 1992; McCarter et al., 1993; Tull et al., 1994; Tull and Withers, 1994; Withers and Aebersold, 1995) by which a covalent intermediate species accumulates (Figure 3.1), a process known as "trapping". A good leaving group, such as 2,4-dinitrophenol, is added as the aglycone moiety which increase the rate of the first step in catalysis, with relatively reactive leaving groups. These minimally modified substrates are designed to undergo the first step of catalysis, the formation of a glycosyl-enzyme intermediate, but to hydrolyze very slowly, up to \( 10^6-10^7 \)-fold slower than the parent sugar (Withers and Aebersold, 1995). This reduced rate has two sources; one is inductive destabilization of the positively charged oxocarbenium ion-like transition states by the electronegative fluorine, the other is the loss of crucial transition state hydrogen bonding interactions (Roeser and Legler, 1981; McCarter et al., 1992). Such a hypothetical interaction may involve C-2 hydroxyl as a hydrogen bond donor to the protein, which cannot be fulfilled by a fluorine atom. Combined together, these effects provide a rationale for a decrease in transition state stabilization by more than \( 30 \text{ kJ mol}^{-1} \) as a result of substitution on C-2. (Roeser and Legler, 1981; Wolfenden and Kati, 1991; McCarter et al., 1992) The net result is the accumulation of the covalent intermediate for an extended period of time, as observed both by electrospray ionization mass spectrometry and by \(^{19}\text{F-NMR} \) spectroscopy (Withers et al., 1988; Withers and Street, 1988; Withers et al., 1990; Tull et al., 1991; Tull and Withers, 1994; Withers and Aebersold, 1995).

(3) The covalent intermediate is then turned over at a slower rate, either by hydrolysis or transglycosylation, indicating that this complex is a catalytically competent species (Withers et al., 1990; Street et al., 1992; Tull and Withers, 1994).

Taken together, the above results strongly suggest that trapped fluoroglycosides are representative of the natural glycosyl-enzyme intermediate for retaining \( \beta-1,4-\)glycosidases. One
such example is the 2",4"-dinitrophenyl 2-deoxy-2-fluoro-cellobioside (2-FDNPC, Figure 3.1). Moreover, when combined with electrospray mass spectroscopy, these 2-fluoroglycosides are now straightforward tools for the identification of the catalytic nucleophile of retaining β-glycosidases (Withers and Aebersold, 1995). This compound 2F-DNPC has been successful with cex-cd which hydrolyzes either xylan or cellulose with retention of anomeric configuration (Gilkes et al., 1984; Gilkes et al., 1991a; Tull et al., 1991; Tull and Withers, 1994).

Figure 3.1. Cleavage of 2",4"-dinitrophenyl-2-fluoro-2-deoxy-β-cellobioside (2FDNPC). Cex-cd cleaves 2F-DNPC, releases dinitrophenol, and forms a covalent fluorocellobiosyl-enzyme species with Glu233. In this process, the catalytic nucleophile Glu233 acts in a concerted manner with the acid/base catalyst Glu127 (Sinnott, 1990; MacLeod et al., 1994).

In an attempt to characterize the conformation of such a glycosyl-enzyme species, and to provide insights into interactions at the active site that are crucial to substrate specificity and catalysis, we have determined the crystal structure of a trapped fluorocellobiosyl-enzyme complex for the catalytic domain of the xylanase/glucanase Cex (cex-cd) from Cellulomonas fimi.
MATERIALS AND METHODS

FORMATION OF THE COMPLEX

The catalytic domain of recombinant β-glycanase Cex (cex-cd) has been crystallized as previously described (Bedarkar et al., 1992). Crystals of native cex-cd were soaked with various concentrations of 2",4"-dinitrophenyl-2-fluoro-2-deoxy-β-cellobioside (2F-DNPC) in a 10 μL drop of the reservoir solution. The drop was suspended over 1 mL of 15% polyethyleneglycol 4000 (PEG4000) and 100 mM acetate (pH 4.6). Although this reservoir solution contains the same buffer as the crystallization conditions, the precipitating agent is now doubled, that is 15% PEG4000, to avoid dissolution of the protein crystal.

Soaked cex-cd crystals in 0.5, 1 or 2 mM 2F-DNPC for either 3, 11 or 24 hours were analyzed with X-rays on a San Diego Multiwire Systems area detector. X-ray diffraction data were collected by 90 sec oscillation frames at an angular speed of 0.1°/frame around ω over 48 hrs. The X-ray-sensitive detector plates were centered at 2θo=-16° and 2θ1=35° so to extend the resolution to 1.8 Å and to include a sufficient overlap of data between detectors, in anticipation of the scaling of the data. The observed intensities were reduced as previously described for the native cex-cd (White et al., 1994). The data reported here were measured on a single crystal. An equivalent crystal form was also obtained by crystallization of the complexed enzyme (data not shown).

STRUCTURAL ANALYSIS AND REFINEMENT

The position of the cex-cd model (White et al., 1994) was optimized by rigid-body refinement and then refined with X-PLOR (version 3.1) (Brünger et al., 1987). The starting model contains the 312 amino-terminal residues of cex-cd, in which Glu 233 was modified to an alanine, in order to assess the position of its side chain. Simulated annealing protocol was applied to the model, in which the temperature was decreased from 3000 to 300 K by steps of 25 K, followed by standard positional and individual isotropic B-value refinements. The refinement progress was monitored with the free R value (Brünger, 1992) for 7% of the reflections set aside prior to the refinement. Model work and analysis was done using the program O (Jones et al., 1990; Jones et
In order to test the model further, an extensive analysis using "simulated-annealing omit-maps" (Hodel et al., 1992) was carried out for the whole polypeptide chain, in which segments of 10 to 13 residues were omitted prior to rounds of simulated annealing with the temperature decreasing from 1000 to 300 K.

The side chain of residue Gln87 is found disordered and treated as an alanine in the model. After several rounds of model building and positional refinement, water molecules were added. Each peak above 3σ in a |F₀|-|F₂| map was visually examined and assigned as a water molecule if it met the following criteria: the peak centroid was within hydrogen bond distance (2.5 to 3.5 Å) to at least one non-hydrogen atom of either the protein, the ligand or another water molecule; acceptable hydrogen bond angles; temperature factor \( B < 70 \text{ Å}^2 \); consistency with a peak on 2|F₀|-|F₂| and 3|F₀|-2|F₂| electron density maps. No bulk solvent correction was used with the model.

Following refinement of the resulting model, the disaccharide was incorporated into the model and refined with an occupancy of 0.75. A modified version of the files "param3_mod.cho" and "toph3.cho" (X-PLOR) were used for the disaccharide restraints and applied as follows: bond lengths of the ligand were restrained with the weights and distances suggested with the exception of the covalent linkage which was restrained to an ideal value of 1.43 Å with the same weight as the rest of the saccharide; restrained weights on angles involving the ligand carbons C-1 and C-2 were reduced to 10% of the weight applied for the remaining of the disaccharide; dihedral angles were released for angles involving the ligand's carbons C-1 and C-2 whereas full dihedral restraints were applied to the rest of the disaccharide. The resulting refined model contains 2396 non-hydrogen protein atoms, 22 disaccharide atoms and 189 water molecules and has good agreement with geometrical parameters for well determined structures as verified using PROCHECK (Laskowski et al., 1993). The atomic coordinates are being deposited with Protein Data Bank (entry code 1EXP).
RESULTS AND DISCUSSION

A careful analysis of the diffraction pattern revealed that many crystals examined were not useful. One criteria of rejection was the misindexing of about 10% of the data. On the basis of this observation, many soaked crystals were discarded. We found that a soak of a cex-cd crystal in 0.5 mM inhibitor for 11 hrs followed by immediate mounting of the crystal in a capillary were appropriate conditions. This crystal belongs to space group $P4_12_12$ with unit cell dimensions $a=b=88.17 \text{ Å}$ and $c=81.10 \text{ Å}$. X-ray diffraction data to 1.8 Å were collected, which gives a total of 139457 intensity measurements which were reduced to 29464 unique reflections, representing 99% of the predicted data (Table 3.1).

Diffusion of the slow substrate 2"4"-dinitrophenyl-2-deoxy-2-fluoro-β-cellobioside (2F-DNPC) into crystals of cex-cd leads to the appearance of an intense yellow colour within the crystal, which indicates the release of the chromophore dinitrophenol. Inherently, cex-cd cleaves the slow substrate 2F-DNPC (Figure 3.1) in the conditions of the soaking. The final model contains 2396 non-hydrogen protein atoms, 22 saccharide atoms and 189 water molecules. The $R$-factor for this model is 0.208 (free $R$ value 0.283) with reasonable stereochemistry (Table 3.1). A plot of $R$-factor against resolution (Luzzati, 1952) or the SIGMAA method (Read, 1986) gives an upper estimate of coordinate error, assuming model errors to be solely responsible for the difference between $|F_{\text{obs}}|-|F_{\text{calc}}|$. This estimated average coordinate error is 0.25 Å. The main chain of the model makes a reasonable fit with the electron density map as shown in Figure 3.2, with a mean real-space correlation of 0.93 against the final electron density map. Low values of correlation agree with corresponding high $B$-factors, suggesting larger mobility in these regions. Conversely, high correlation values, such as for the active site residues, correspond to low $B$-values.
Table 3.1. Data Collection and Refinement Statistics

<table>
<thead>
<tr>
<th>Resolution (Å)</th>
<th>8 - 1.8</th>
<th>1.95 - 1.8&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>139457</td>
<td>16567</td>
</tr>
<tr>
<td>Unique reflections (% completeness)</td>
<td>29464 (99)</td>
<td>5644 (97)</td>
</tr>
<tr>
<td>$R_{sym}$ (%)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4.1</td>
<td>28.4</td>
</tr>
<tr>
<td>Signal to noise ($\langle l/\sigma \rangle$)</td>
<td>19.7</td>
<td>2.1</td>
</tr>
</tbody>
</table>

**Refinement:**

| No. reflections ($|F|>2\sigma$) (% completeness) | 24597 (83) | 4155 (71) |
|-----------------------------------------------|------------|-----------|
| $R_{factor}$ (%)<sup>c</sup> | 20.8       | 25.0      |
| No. reflections for free $R$ value (% completeness)<sup>e</sup> | 1866 (6)    | 291 (5)   |
| free $R$ value (%) | 28.3       | 34.0      |
| Non-hydrogen protein atoms | 2396       |
| main-chains $<B$-factor ($\AA^2$) | 24.2       |
| side-chains $<B$-factor ($\AA^2$) | 25.9       |
| Identified water molecules | 189        |
| $<B$-factor ($\AA^2$) | 41.9       |
| Fluorocellulobioside atoms | 22         |
| $<B$-factor ($\AA^2$) | 38.2       |
| Rms deviation from ideality<sup>d</sup>: |         |
| bond lengths ($\AA$) | 0.014      |
| bond angles (°) | 2.0        |
| improper angles (°) | 1.5        |
| dihedrals (°) | 22.4       |

Estimated average error in coordinates, Luzzati method (Luzzati, 1952) ($\AA$) | 0.25 |

<sup>a</sup> Highest resolution shell of data.
<sup>b</sup> $R_{sym} = \sum l_i - \langle b \rangle / \sum l_i$, where $l_i$ is the scaled intensity of the $i^{th}$ measurement and $\langle b \rangle$ is the mean intensity for that reflection.
<sup>c</sup> $R_{factor} = \sum ||F_{calc}|-|F_{obs}||/\sum |F_{obs}|$ where $F_{calc}$ and $F_{obs}$ are the calculated and observed structure factor amplitudes, respectively.
<sup>d</sup> Standard stereochemical parameters (Engh and Huber, 1991) were used in the refinement.
<sup>e</sup> These represent 7% of the 26463 reflections with $|F|>2\sigma$. 
Figure 3.2. Mobility of the residues and agreement of the model, as evaluated with the main chain real-space correlation and temperature factor (B) profiles (dotted lines). The real-space correlation (Jones et al., 1991) for the main chain of each residue of the complexed cex-cd model was evaluated against $2|F_o| - |F_c|$ electron density maps, in which phases were derived either from the final model (thick lines) or from simulated-annealing omit maps (thin lines).

A significant element of electron density is found to be an extension of Glu233 (Figure 3.3). This entity has the size and shape of the anticipated fluorocelllobioside, and was assigned accordingly. The carbohydrate-binding cleft of homologous xylanases is composed of several subsites (Derewenda et al., 1994a; Harris et al., 1994; Dominguez et al., 1995; Harris et al., 1996). In cex-cd, the disaccharide occupies two binding subsites in a deep crevice at the carboxyl-
terminal end of the \((\beta/\alpha)_8\)-barrel motif (Figure 3.4). This C-terminal region mediates the interactions with the disaccharide by presenting 18 conserved residues into the enzyme's active site, a binding environment that provides a rationale for the substrate specificity of cex-cd.

\[\text{Figure 3.3. Electron density map (3|F_o|-2|F_c|) of the fluorocellobioside attached to Glu233. a) This map (contour level of 1\sigma) was generated using phases of the model cex-cd before the incorporation of the disaccharide. The disaccharide was later included with the model and refined. This final model overlays the electron density map shown. b) Corresponding schematic drawing of the disaccharide attachment to the nucleophile Glu233.}\]

**STEREOCHEMISTRY OF THE GLYCOSYL-ENZYME ATTACHMENT**

On the basis of a double-displacement catalytic mechanism, it is expected that the covalent intermediate of a retaining \(\beta\)-glycosidase adopts an \(\alpha\) anomeric configuration with a catalytic nucleophile. This catalytic nucleophile for cex-cd has been previously identified as Glu233 by labeling with the 2-fluorocellobioside and sequencing of the purified glycopeptide (Tull et al., 1991: Tull et al., 1994). In the crystal structure, the C-1 of the proximal saccharide is now
confirmed to be tetrahedral and bonded to the conserved residue Glu233 $O^\varepsilon_1$ by a covalent linkage of 1.43 Å, in the $\alpha$ anomeric configuration (Figure 3.3). There is no significant element of electron density corresponding to the $\beta$ configuration. The carbon C-1 of the attached saccharide is found syn to the ester group of Glu233, a preferential location for a nucleophilic attack by a carboxylic acid (Gandour, 1981). In addition, this ester linkage is found to be coplanar with the cyclic O-5 of the attached saccharide (Figure 3.3), which may have stereoelectronic implications.

**Figure 3.4. Fluorocellobioside binding region in cex-cd.** The ligand (red for oxygens, yellow for carbons and magenta for fluorine atom) binds to the nucleophile Glu233 (orange) in the active site of cex-cd ($\alpha$-helices in blue and $\beta$-strands in green) by a covalent linkage (light blue). The acid/base catalyst Glu127 (orange) is located 5.5 Å from the nucleophile in this retaining enzyme. This figure was prepared with SETOR (Evans, 1993).
The covalent attachment of the glycosyl adduct to cex-cd in the crystal structure is supported in solution by $^{19}\text{F-NMR}$ spectra (Figure 3.5). It has been shown in a closely related system (Withers and Street, 1988) that this technique can be used to distinguish $\alpha$ and $\beta$ anomeric configurations of a covalently attached fluorosaccharide. The broad resonance at $\delta - 195.5$ ppm in the spectrum of cex-cd with 2F-DNPC (Figure 3.5a) arises from the covalent fluorocellobiosyl-enzyme complex, the large linewidth indicating the covalent attachment. Formation of the intermediate results in a change of chemical shift of less than three ppm relative to 2F-DNPC, which is insufficient to draw conclusions about the stereochemistry of the glycosyl-enzyme linkage. However, $^{19}\text{F-NMR}$ chemical shifts of $\alpha$- and $\beta$-2-fluoromannosides differ by approximately 20 ppm, sufficient for the reliable identification of the anomeric configuration of the covalent species (Withers and Street, 1988). The corresponding 2-deoxy-2-fluoro-4-O-{$\beta$-D-glucopyranosyl}-$\beta$-D-mannopyranosyl fluoride (2F-GMF) with an axial fluorine at position C-2 was therefore synthesized and found to be a time-dependent inactivator of cex-cd ($k_i = 0.069\ \text{min}^{-1}$, $K_i = 160$ mM, measured as previously (Sinnott, 1990; Tull and Withers, 1994)) and thus, to form a covalent glycosyl-enzyme intermediate. The large dissociation constant of this inhibitor necessitated use of a large excess of 2F-GMF in order to achieve saturation. Therefore, as is seen in Figure 3.5b, the only resonances apparent in this region of the spectrum are those from F-2 of the excess 2F-GMF at $\delta - 224.3$ ppm and from F-2 of the product of spontaneous hydrolysis at $\delta - 205.4$ ppm. The chemical shift and coupling constants of this latter species confirm it as the more stable $\alpha$-anomer, but unfortunately this resonance obscures the one due to the covalently linked glycosyl-enzyme intermediate. However, upon dialysis of the sample the $^{19}\text{F-NMR}$ spectrum reveals a resonance at $\delta - 205.6$ ppm due to the covalent glycosyl-enzyme (Figure 3.5c). This chemical shift is consistent only with an $\alpha$-anomeric linkage.
Figure 3.5. $^{19}$F-NMR spectra of cex-cd with 2-fluoro-disaccharides. a) cex-cd (0.2 mM) plus 2F-DNPC (1.0 mM). The peaks at $\delta$ -195.4 ppm with a broader line width was assigned to the covalent intermediate. The peaks at $\delta$ -198.1 ppm and -198.4 ppm are unliganded 2F-DNPC and hydrolyzed products respectively. b) Cex-cd (0.2 mM) plus 2F-GMF (30 mM). c) Same as in b, after dialysis. $^{19}$F-NMR were acquired essentially as described previously (Withers and Street, 1988). This figure was prepared with SETOR (Evans, 1993). The inhibitor synthesis (Tull et al., 1991), the one dimensional NMR and the kinetic analyses were carried out by Drs. L.P. McIntosh, D. Tull and S. G. Withers.
Another conserved acidic residue, Glu127, has previously been identified as the acid/base catalyst on the basis of kinetic analysis of mutants (MacLeod et al., 1994). This residue was also affinity-labelled with N-bromoacetyl cellobiosylamine (Black et al., 1993) as an important residue, but only recently identified (Tull et al., 1996). Glu127 forms a hydrogen bond with both Trp84 and a water molecule, and is located 4.5 Å away from the Cl of the proximal saccharide in the complex. It's distance of 5.5 Å from the nucleophile O\textsuperscript{E1} agrees with observed distances in retaining β-1,4-glycosyl hydrolases (McCarter and Withers, 1994; White et al., 1994). The location of the acid/base catalyst Glu127 is consistent with its role of protonating the glycosidic bond during formation of the glycosyl-enzyme and of deprotonating an incoming water molecule during the hydrolysis of the complex.

The nucleophilicity of Glu233 is likely to be influenced by its immediate environment. Glu233 side chain is parallel to both side chains Gln203 and Trp273 and found between them (Figure 3.6). This hydrophobic environment may favour a lower pK\textsubscript{a} for Glu233, which in turn may increase the nucleophilicity of this residue. In both the complexed (Figure 3.6) and the uncomplexed cex-cd, Glu233 accepts hydrogen bonds from Asn169 and His205. This imidazole ring in turn is properly located to hydrogen bond to Asp235. In the cellobiosyl-enzyme complex, the dyad Asp235 - His205 is likely to be important in assisting the hydrolysis of the glycosyl-enzyme complex by stabilizing the released carboxylate. Likewise, this dyad could serve to maintain the ionization of the catalytic nucleophile, stabilizing its deprotonated state in the free enzyme.

**Figure 3.6. Fluorocellobioside interactions with cex-cd.** a, The two invariant carboxylates Glu127 and Glu233 are shown in red and the other conserved residues of cex-cd (Gilkes et al., 1991b; White et al., 1994) are shown in green. The fluorocellobioside (same color scheme as in Figure 3.4) is covalently bonded (blue) to the nucleophile Glu233 in an α-configuration. b, Schematic diagram of the possible hydrogen bond network (dashed lines with length in Å indicated) of cex-cd involved in the recognition of the fluorocellobioside.
Relevance of the Substituent at the Substrate C-2

The fluorine substituent F-2 of the cellobioside substrate is involved in specific interactions with the enzyme. The fluorine F-2 accepts a hydrogen bond from Asn126 N$^{\text{82}}$ and is close (3.1 Å) to the nucleophile Glu233 O$^{\text{52}}$ (Figure 3.6). A replacement of the equivalent asparagine (Asn139) to either Ala, Gln or Asp in the homologous CelC leads to a substantial decrease (greater than 100-fold) in catalysis (Navas and Béguin, 1992), suggesting an important role for this residue. Moreover, this amino acid is conserved in more than 150 glycosyl hydrolases, that is in families 1, 2, 5, 10, 17, 30, 35, 39, and 42 (Domínguez et al., 1996). The various substrates recognized by these enzymes all present an equatorial OH-2 group (Domínguez et al., 1996), although families 5 and 10 have some residual activity with mannan as well. In cex-cd, interaction of Asn126 with the equatorial fluorine substituent of the proximal saccharide may provide clues about conserved means of specificity by these glycosyl hydrolases.

Studies on cex-cd and related retaining β-glycosidases have shown that interactions with the hydroxyl group OH-2 of the natural substrate contribute 30 to 40 kJ mol$^{-1}$ to the stabilization of the transition states for glycosyl transfer (Roeser and Legler, 1981; Wolfenden and Kati, 1991; McCarter et al., 1992). Because deoxygenation inductively stabilizes the oxocarbenium ion-like transition states, thus inherently increasing the reaction rate, the observed decreases in reaction rate must be due to the removal of important transition state binding interactions. The value of 30 to 40 kJ mol$^{-1}$ therefore represents a minimum estimate of the contribution of these interactions. The source of such strong interactions is therefore of considerable interest. While interactions with Asn126 likely contribute to this stabilization it seems probable that the strongest interaction would be one with O$^{\text{52}}$ of Glu233. Such a hydrogen bond with a charged acceptor could provide the large interaction energies observed, approaching those of the low-barrier hydrogen bonds observed in other systems (Cleland and Kreevoy, 1994; Frey et al., 1994; Tobin et al., 1995). This type of ligand C-2 interaction to nucleophile may bring clues to the possibility of substrate-assisted catalysis. However, the structure of a good transition state analogue complex will be needed to provide further insights.
An additional interaction is found to involve a C–H···F–C type of hydrogen bond. The fluorine atom of the disaccharide is located in the same plane as the imidazole C^e1 of His80. In addition, these atoms are 3.0 Å apart, which taken together are the characteristics of a typical hydrogen bond. The involvement of a methylene group (C–H) in either a hydrogen bond or a weakly polar interaction is recognized to occur in proteins (Burley and Petsko, 1988; Derewenda et al., 1994b). In the latter report, serine proteases were compared to X-ray and neutron diffraction results of small organic molecules and later, this study was extended to a statistical analysis of 13 high resolution protein structures (Derewenda et al., 1995). The C–H···X–C bond^7 stereochemistry reported is typical of commonly attributed hydrogen bonds in proteins (Ippolito et al., 1990). Our results show that the C–H···F–C interaction between His80 C^e1 and the ligand is similar to the C–H···X–C bond refered to above, which raises the possibility that this interaction ligand to protein may provide a biological rationale for one element of substrate specificity by cex-cd.

**STRUCTURAL BASIS OF THE SPECIFICITY BY CEX-CD**

Cex-cd hydrolyzes xylan 40-fold more efficiently than cellulose (Tull and Withers, 1994). Xylan is a β-1,4 linked polymer of D-xylose, a saccharide unit similar to glucose but lacking the hydroxymethyl group on the carbon C-5. To accommodate the C-5-hydroxymethyl group of the distal and proximal saccharides respectively, Gln87 is now disordered and Trp281 is rotated by 31° around its C^δ-C^γ axis as compared to the uncomplexed structure. These less favorable rearrangements would be unnecessary in a xylobiosyl complex, which is consistent with cex-cd's preference for xylobiose over cellobiose at this site.

Cex-cd has been reported to prefer cleaving a cellobiose unit rather than glucose from the non-reducing end of cellulose. This preference is also reflected in the greater activity with aryl cellobioside substrates than arylglucosides (Tull et al., 1991; Tull and Withers, 1994). In the

---

^7 X is an electronegative atom, in most cases an oxygen. This type of hydrogen bond was originally proposed thirty seven years ago (Pauling, 1960), and later reviewed for small organic molecules on the basis of crystallographic structures (Taylor and Kennard, 1982). The type of interaction involving a C–H hydrogen bond donor group has been largely ignored for macromolecule accounts until recently (Burley and Petsko, 1988).
crystal structure. upon binding of fluorocellobioside, cex-cd does not undergo significant conformational changes, as shown by a rms deviation of 0.2 Å for the Cα's between the complexed and free enzyme. Thus, a minimal rearrangement takes place upon complex formation.

A closer analysis of the carbohydrate binding reveals an important contribution by the loop of residues 43 to 47, which joins the β-strand β2 to the α-helix α2a (White et al., 1994). Among these amino acids 43 to 47, three are conserved in the family 10 of glycosyl hydrolases (see Figure 2.6). One of these residues, the side chain of Lys47, shows a significant decrease in its temperature factor, from 23.6 to 15.8 Å², which may be explained by its interactions with the ligand. The proximal saccharide makes hydrogen bonds with Lys47, His80, and Asn126 whereas the distal saccharide significantly contributes through hydrogen bonding with residues Glu43, Asn44, Lys47, and Trp273 (Figure 3.6). This strong interaction for the distal saccharide correlates with the extensive six hydrogen bonds observed for N-acetylglucosamine bound to lysozyme at the equivalent ligand-binding subsite "C" (Strynadka and James, 1991). Thus, in cex-cd the loop 43-47 may play a dominant role in the preference of the enzyme for the release of disaccharides rather than monosaccharides and may explain the preferential recognition of a disaccharide by cex-cd over a monosaccharide.

Binding of a ligand in the active site of an enzyme does not necessarily ensure that it binds at the immediate vicinity of the catalytic residues. Such an example is partridge lysozyme, for which the saccharide (GlcNAc)₃ occupies the subsites A, B, and C, whereas the β-1,4 linkage hydrolysis takes place between subsites D and E (Turner and Howell, 1995). Thus, the ligand binds at more than 6 Å from the catalytic residues Glu35 and Asp52. In contrast, the use of a fluorocellobioside as a ligand, which leads to a covalent bond with the catalytic nucleophile of the enzyme, unambiguously identifies two substrate-binding subsites in the active site of cex-cd.

Among the water molecules located in the structure, one forms a hydrogen bond to C-4-OH of the distal saccharide, and another one is found within hydrogen bonding distance to both the acid/base catalyst Glu127 and the active site residue Gln203. Although it is 4.0 Å away from the sugar C-1, this latter water molecule is the most likely candidate to be the nucleophile in the second
step of catalysis. Whether or not this designated water actually performs the nucleophilic attack on C-1 or forms a hydrogen bond with an incoming nucleophilic water molecule still remains an open question.

IMPLICATIONS

The structures of complexes of other affinity-labeled glycosidases have been reported previously. Examples include adducts with glycosyl epoxide affinity labels of lysozyme (Moult et al., 1973) and a β-1,3-1,4-glucanase (Keitel et al., 1993). However, these complexes are not structural analogs of a covalent intermediate since they are not catalytically competent and there is no direct linkage between the enzyme's catalytic nucleophilic side chain and C-1 of the saccharide. In addition, the structure of a T4-lysozyme mutant complexed with a substrate covalently attached at the site of mutation has been determined (Kuroki et al., 1993), but that complex is not the natural intermediate in catalysis. The use of fluoroglycosides as mechanism-based slow substrates for structural studies does address directly the stereochemistry of the disaccharide attached to the enzyme's catalytic nucleophile, in addition to allowing a description of a β-1,4-glycosidase in the context of a covalent intermediate. As the glycosyl-enzyme complex is generated in the crystal, the activity of the uncomplexed cex-cd enzyme in the crystalline state is confirmed. Moreover the covalent complex presented here unambiguously identifies two saccharide-binding subsites of the enzyme and their specific carbohydrate-enzyme interactions. In particular, the proximity of the fluorine atom F-2 to Glu233 Oe2 suggests a rationale for the importance of the hydroxyl group at this position. In addition, this structure provides useful insights into the role of specific interactions in determining the biotechnologically important specificity for xylan versus cellulose.
REFERENCES


Cellulomonas fimi exoglycanase as the residue modified by N-bromoacetyl

without radiolabels: tandem mass spectrometric localization of the tagged active-site

study of the exo-β-1,4-glycanase from Cellulomonas fimi. Biochemistry 33, 6363-6370.

(1991) Glutamic acid 274 is the nucleophile in the active site of a "retaining" exoglucanase
from Cellulomonas fimi. J. Biol. Chem. 266, 15621-15625.

without tri-N-acetylchitotriose inhibitor at 1.9 Å resolution. Protein Sci. 4, 442-449.

domain of the β-1,4-glycanase Cex from Cellulomonas fimi. Biochemistry 33, 12546-
12552.

Withers, S. G. and Aebersold, R. (1995) Approaches to labeling and identification of active site

class of specific mechanism-based glycosidase inhibitors. J. Biol. Chem. 263, 7929-7932.

Withers, S. G. and Street, I. P. (1988) Identification of a covalent α-D-glucopyranosyl enzyme

(1990) Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile

Chapter 4

Hydration of a Glycal: The Glycosyl-Enzyme Intermediate
Adopts a "Full Chair" Conformation in the β-Glycanase Cex-cd.

This chapter describes the crystal structure of a covalent 2-deoxy-glycosyl-enzyme complex with cex-cd. This species is obtained by esterification of the double bond of cellulobial which is a glycal. Such a covalent species has been postulated to occur in the hydration of glycals by retaining β-glycosyl hydrolases, but was not previously observed. This chapter describes its structural features [White, A., Withers, S. G., and Rose, D. R. (1996) to be submitted].
ABSTRACT

It is well established from biochemical and structural studies that a saccharide deformation takes place during catalysis by glycosyl hydrolases. In the context of two catalytic carboxylates in the active site of most of the retaining β-1,4-glycosidases, the hydrolysis is believed to proceed by a double displacement catalytic mechanism through a covalent intermediate with oxocarbenium transition states. Our previous crystallographic studies reveal that a covalent α-glycosyl-enzyme catalytic intermediate can be accommodated in the confined space of the active site of the enzyme cex-cd (Chapter 3). Moreover, such an intermediate has previously been proposed as a transient event through the hydration of glycals (1,2-unsaturated saccharides) such as cellobial. We present the crystal structure of cex-cd determined in the presence of cellobial.

Cex-cd is the catalytic domain of a bacterial xylanase/cellulase and is a representative enzyme of the family 10 of glycosyl hydrolases. We previously solved the crystal structure of the unliganded form of cex-cd by classical heavy atom methods (Chapter 2). This model is now being used to determine the crystal structure of cex-cd with cellobial. Our results suggest that cex-cd converts this ligand, and so forms a covalent 2-deoxy-α-cellobiosyl-enzyme species which represents the catalytic intermediate of hydration of cellobial. Compared to the fluoro-cellobiosyl covalent complex (Chapter 3), the 2-deoxy-cellobiosyl residual disaccharide occupies the same subsites and makes a similar network of interactions with the enzyme. These two disaccharides each adopt a "full chair" conformation which differs from the planar arrangement of the transition states. Comparison of the structure of these liganded forms of cex-cd provides information on the catalytic mechanism of retaining β-glycosidases.
**INTRODUCTION**

β-1.4-glycanases are classes of glycosyl hydrolases specific for D-xylose and D-glucose-contained saccharides. In the context of two catalytic carboxylates in the active site of most retaining β-1.4-glycosyl hydrolases, substrate hydrolysis is believed to proceed by a double displacement catalytic mechanism through a covalent catalytic intermediate with oxocarbenium ion-like transition states (Sinnott, 1990; Withers, 1995). Using fluorosaccharides, the trapping of a covalent glycosyl-enzyme complex allowed the examination of the catalytic intermediate. With the β-glycanase cex-cd, such a catalytic covalent species was characterized kinetically (Tull et al., 1991; Tull and Withers, 1994), chemically (Tull et al., 1996), and structurally with both one dimensional NMR analysis and high resolution X-ray crystallography (White et al., 1996).

Our previous crystallographic studies using a fluorocellobioside slow substrate with the β-glycanase cex-cd provided a first look at the postulated catalytic intermediate (White et al., 1996). A covalent glycosyl-enzyme may be accommodated in the confined space of the active site of cex-cd, and confirms the α anomic configuration of its link to Glu233. This link is found at the syn position of the nucleophile Glu233 and nearly coplanar with the cyclic oxygen of the bond saccharide. The resulting ester moiety of the nucleophile Glu233 is inferred to form a hydrogen bond with the hydroxyl substituent at C2 of the residual fluorosaccharide. In this structure the attached saccharide adopts a "full" chair conformation which differs from the postulated planar arrangement of the transition states.

Glycals, such as cellobial (Figure 4.1), are glycosyl hydrolase inhibitors and are slowly hydrated by their host enzyme, yielding 2-deoxy-glycosides as the products through the addition of either water or alcohol to the double bond of glycals (Lalégerie et al., 1982). Trapping by denaturation allows the isolation of glycosylated peptides suggesting that a covalent glycosyl-enzyme intermediate is involved in this hydration, as shown for the hydration of D-glycals by *E. coli lacZ* β-galactosidase (Wentworth and Wolfenden, 1974; Viratelle and Yon, 1980) and of β-glucosidase A3 of *A. wentii* (Legler et al., 1979). This intermediate would take the form of a
covalent 2-deoxy-glycosyl-enzyme, which is presumed identical to the intermediate involved with the hydrolysis of \( p \)-nitrophenyl 2-deoxy-\( \beta \)-D-glucopyranosides (Roeser and Legler, 1981). The structural characterization of such a covalent intermediate from a \( D \)-glycal is difficult. For example, the high resolution crystal structure of the retaining \( \alpha \)-glycosyl hydrolase soybean \( \beta \)-amylase bathed with \( D \)-maltal for 40 min shows the accumulation of the hydrated product 2-deoxy-\( \alpha \)-maltose (Mikami et al., 1994). To our knowledge, there is no other report of a complex of a glycal with a glycosyl hydrolase.

**Figure 4.1. Cellbrial chemical structure.**
The proximal saccharide ring is unsaturated at C1-C2 which confers a trigonal geometry for these carbons. The inherent planar geometry makes this compound somewhat structurally similar to the transition states.

The stereochemistry of the hydration of glycals has been analyzed for \( D \)-glucal with sweet almond \( \beta \)-glucosidase (Hehre et al., 1977), for the addition of glycerol on \( D \)-galactal with \( lacZ \) \( \beta \)-galactosidase (Lehmann and Zieger, 1977), and for a number of other glycosyl hydrolases. On the basis of deuterium labeling experiments with these \( \beta \)-glycosyl hydrolases, a concerted cis-addition mechanism through a 6-member ring was proposed (Hehre et al., 1977; Lehmann and Zieger, 1977) that involves a glycosyl cation transition state (Legler et al., 1980). This interpretation means that a proton is added from the \( si \) face (bottom) of the saccharide, that is from the same side as the attack from the enzyme’s catalytic nucleophile; a covalent intermediate with the anomeric configuration \( \alpha \) is thus formed. So, the catalytic mechanism of hydration by a \( \beta \)-glycosyl hydrolase proceeds as shown on figure 4.2.

In this catalytic mechanism, the nucleophile needs to be protonated, whereas the acid/base catalyst should be ionized (Figure 4.2). At acidic pH both the nucleophile and the acid/base catalyst would be protonated, which suggests that the condition at step IV may not be fulfilled and
Figure 4.2. Catalytic mechanism of hydration of glycals by $\beta$-glycosyl hydrolases.

thus would lead to the accumulation of the covalent intermediate, the step III. In an attempt to test this hypothesis, we report the crystal structure of the $\beta$-glycanase cex-cd determined at pH 4.6 in the presence of cellobial (Figure 4.1).

Our structure indicates a covalent 2-deoxy-$\alpha$-cellobiosyl-cex-cd (compound III) as a predominant species which allows its structural examination. Our results show a "full chair" conformation for the 2-deoxy-$\alpha$-cellobiosyl moiety. The inhibitory species cellobial is a covalent adduct to the nucleophile Glu233.
MATERIALS AND METHODS

PROTEIN CRYSTAL PREPARATION

The catalytic domain cex-cd obtained from recombinant Cex was generously provided by the team of Dr. R. A. J. Warren (University of British Columbia, Vancouver). This protein is crystallized by vapour diffusion against a 1 ml reservoir containing 100 mM sodium acetate buffer pH 4.6 with 7.5 to 8.5% deionized polyethylene glycol 4000: drops of 2 μl of cex-cd at 67 mg·ml⁻¹ were mixed with an equal volume of the reservoir solution and suspended over this reservoir solution in a sealed container as previously described (Bedarkar et al., 1992). Useful large crystals of 0.3 to 0.8 mm on each edge appeared, typically within a week at room temperature. In some cases, seeding with small crystals leads to a useful larger crystal within a day. Some of the crystals appeared only after a few months, such that a membrane of dehydrated polyethylene glycol formed at the surface of the drop.

CELLOBIAL-CEX-CD COMPLEX FORMATION AND DATA COLLECTION

Large monocrystals were transferred to 10 μL drops of a reservoir solution containing 100 mM sodium acetate buffer pH 4.6 and 15% deionized polyethylene glycol 4000. A 56.8 mM solution of cellobial is prepared with 100μL of the reservoir solution. To the above drops, 1.76 μL and 2.14 μL of the cellobial solution are added, which gives 10 mM final; each of these additions is followed by 4 hours incubation at room temperature. The crystals were mounted in quartz capillaries which were then sealed and mounted on a goniometer head. X-ray diffraction data were measured using a San Diego Multiwire Detector equipped with two bidimensional area detectors and a Rigaku RU200 copper rotating anode operated at 40 kV and 150 mA. Soaking of cex-cd with cellobial induces less than 0.2% change to the P4₁2₁₂ unit cell parameters, which are now a=b=88.266 Å and c=81.264 Å. The intensity of 95939 X-ray Bragg reflections to a resolution of 2 Å were measured through ω scans of 0.1° frames and reduced to 19913 unique reflections.
(Table 4.1) using the SDMS software (Howard et al., 1985). Merging of the symmetry-equivalent measurements yielded an agreement value $R_{sym}$ of 8.2%.

**STRUCTURE SOLUTION AND REFINEMENT**

The crystal structure of the unliganded cex-cd (White et al., 1994) was used to solve its complexed form with celllobial. Refinement is carried out with the program X-PLOR (Brünger et al., 1987; Brünger, 1993). Following the exclusion of the water molecules, the position of the cex-cd model as a rigid entity was optimized through 50 cycles of rigid-body-refinement using data from 8 to 2.2 Å resolution. The slow-cool protocol for refinement with simulated annealing combined with molecular dynamics was applied, in which the temperature was decreased from 3000 to 300 K in steps of 25 K with 25 ps of molecular dynamics at each step. Following this procedure, cycles of standard positional refinement in alternance with model rebuilding using O (Jones et al., 1990; Jones et al., 1991) allowed for the identification of water molecules. In this process, strong peaks were located using SIGMAA-weighted (Read, 1986) $F_{obs}-F_{calc}$ difference electron density maps generated with the CCP4 package (Collaborative Computational Project, 1994), analyzed using a new software and visually examined for consistency on a SIGMAA-weighted $2F_{obs}-F_{calc}$ electron density map using O. Because it is disordered, Gln87 has been treated as an alanine throughout the refinement.

A celllobial molecule was positioned in a SIGMAA-weighted (Read, 1986) $F_{obs}-F_{calc}$ electron density map and incorporated with the protein model for refinement using X-PLOR. Ideal geometry for the saccharides was derived from available dictionaries for X-PLOR and supplemented by analogy to similar compounds obtained from Cambridge Structural Database. For instance, the geometry of the D-glucal was built on the basis of the glucal 3,4,6-tri-O-acetyl-1,2-dideoxy-D-hex-1-enopyranose (Krajewski et al., 1979), whereas the 2-deoxy-α-glucosyl moiety was built using α-D-glucose stereochemistry. Energy components were attributed with analogy to the available dictionaries for X-PLOR. The current refined model contains 2396 non-hydrogen protein atoms, 122 water molecules and 21 non-hydrogen ligand atoms at full occupancy.
RESULTS AND DISCUSSION

Diffraction data to a maximum resolution of 2 Å were measured and cover 90% of the possible reflections (Table 4.1). On average, there are nearly five observations for each reflection; this value decreases to about three at the highest resolution shell of data, that is at 2.15-2.0 Å resolution. Although the X-ray signal is weak for this last resolution shell, the corresponding reflections are kept for the structure analysis.

<table>
<thead>
<tr>
<th>Table 4.1. Data Collection Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>Redundancy</td>
</tr>
<tr>
<td>$R_{sym}$</td>
</tr>
<tr>
<td>Unique reflections (% completeness)</td>
</tr>
<tr>
<td>$&lt;I/\sigma I&gt;$</td>
</tr>
<tr>
<td>Reflections with $I/\sigma I$ (% completeness)</td>
</tr>
</tbody>
</table>

The crystal structure of cex-cd with a disaccharide derived from cellobial has been refined to a resolution of 2 Å. The model agrees with the diffraction data with $R$-factor of 0.19, free-$R$ value of 0.30, rms deviation from ideal geometry of 0.013 Å on bond length and 1.7° on bond angles (Table 4.2). An estimated average coordinates error $\langle |\Delta r| \rangle$ may be estimated by the method of Luzzati (1952). This method is based on an analysis of the $R$-factor varying with the resolution and is assumed to explain the residual information left in a difference Fourier ($F_{obs}$-$F_{calc}$) solely by the model's positional error $\langle |\Delta r| \rangle$. However, it has been noted that the use of free-$R$ value for this analysis may more adequately represent the estimated average coordinate error (Kleyweg: and
Brünger, 1996). This for the current cex-cd model is shown on Figure 4.3. Compared to ideal predictions, the average positional error is evaluated to be 0.35 Å. A two-line weighting scheme (w) is then applied to \( |F_{\text{obs}}| \), according to the method of Dave Smith (Smith, 1997), such that:

\[
\begin{align*}
\text{if } \sin \theta / \lambda & \leq 0.174 \text{ then } w = 1/[A + B(\sin \theta / \lambda - 1/6)] \\
\text{if } \sin \theta / \lambda & \leq 0.174 \text{ then } w = 1/[C + D(\sin \theta / \lambda - 1/6)]
\end{align*}
\]

where the coefficients A, B, C, D are determined using a plot \(<F_{\text{obs}}-F_{\text{calc}}>\) against \(\sin \theta / \lambda\). Such a

---

**Table 4.2. Model Refinement Statistics**

<table>
<thead>
<tr>
<th></th>
<th>without weighting</th>
<th>with weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>High resolution</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>8 - 2.0</td>
<td>2.15 - 2.0</td>
</tr>
<tr>
<td>Reflections used for refinement ((</td>
<td>\mathbf{F}</td>
<td>&gt;2\sigma))</td>
</tr>
<tr>
<td>(R) factor</td>
<td>0.179</td>
<td>0.27</td>
</tr>
<tr>
<td>Reflections used for free-(R) value</td>
<td>1115</td>
<td>160</td>
</tr>
<tr>
<td>Free-(R) value</td>
<td>0.265</td>
<td>0.34</td>
</tr>
<tr>
<td>Data/parameter ratio</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>Non-H protein atoms</td>
<td>2396</td>
<td></td>
</tr>
<tr>
<td>Main chains (B)-value (Å(^2))</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Side chains (B)-value (Å(^2))</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Located water molecules ((B)-value, Å(^2))(^a)</td>
<td>122 (34)</td>
<td></td>
</tr>
<tr>
<td>Non-H ligand atoms ((B)-value, Å(^2))(^a)</td>
<td>21 (33)</td>
<td></td>
</tr>
<tr>
<td>R.m.s. deviation from ideality(^b):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond length (Å)</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>bond angle (deg.)</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>improper angles (deg.)</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>dihedral angles (deg.)</td>
<td>23.6</td>
<td></td>
</tr>
<tr>
<td>Estimated (\langle \Delta r \rangle) (Å) using (R)-factor</td>
<td>0.25-0.30</td>
<td></td>
</tr>
<tr>
<td>Estimated (\langle \Delta r \rangle) (Å) using free-(R)</td>
<td>0.35</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Full occupancy is assumed for the disaccharide.  
\(^b\) Engh and Huber (1991) parameters were used during refinement.
Figure 4.3. *R*-factor analysis as a function of resolution. The *R*-factor (all $|F_{\text{obs}}| > 0$) was calculated either before (○•○) or after (●•●) applying a weighting scheme on the data as a function of resolution; corresponding free-*R* values are also shown for either before (△•△) or after applying the same weighting scheme (●•●). Predicted values of *R*-factor are traced (----) for different estimates of coordinate error ($\langle|\Delta r|\rangle$) as previously suggested (Luzzati, 1952).

scheme assigns an increasing weight to the data with increasing resolution. As a result, the *R*-factor at high resolution reduces, but the overall free-*R* value increases and thus, the estimated positional error increases too to 0.35-0.40 Å (Figure 4.3). Because of the increased free-*R* value, this weighting scheme is not further used here. Refinement statistics are presented in Table 4.2.
Figure 4.4. Mobility of the residues and agreement of the model, as evaluated with the main chain real-space correlation (continuous line) and temperature factor ($B$) profiles (dashed line). The real-space correlation (Jones et al., 1991) for the main chain of each residue of the complexed cex-cd model was evaluated against $2|F_o|-|F_c|$ electron density maps, in which phases were derived from the final model.

Because the $R$-factor is an indicator of the overall agreement of the model, it is relatively insensitive to localized errors. For example, the region of a loop may be inadequately explained by atoms which may still give a low value of the overall $R$-factor. As an attempt to assess the agreement of the model with the electron density results, a correlation (real-space-fit) is computed between the experimental and its corresponding calculated electron density elements at a residue level (Jones et al., 1991). A perfect match results in a correlation value of unity, whereas in the worst scenario this value may reach negative values. Figure 4.4 presents the correlation obtained
Figure 4.5. Stereoview of the poorly resolved amino acids. The side chain of the sixteen residues poorly defined, which are found in three loops, are shown in bold on the Ca tracing of cex-cd.

for the 312 residues of the model when compared to the final \((2F_{\text{obs}}-F_{\text{calc}})\) electron density map, given an overall temperature factor \(B\) of 18.3 Å\(^2\). The correlation ranges from 63% to 95% with an average value of 88%. Because of its mobility, the reliability of Gln87 can not be evaluated. Regions of high agreement include the residues involved in interaction with the ligand, whereas the lowest correlation (≤80%) is noted for residues 94 to 98 which join the secondary structure elements α3a to α3b, 239 to 245 which join β7 to α7, and 284 to 287 which ends α8. These three regions correspond to loops found at the carboxyl-terminal end of the \((\beta/\alpha)_8\)-barrel of cex-cd (Figure 4.5), which may explain both their greater mobility, as expressed by high temperature factor values \(B\) (Figure 4.4), and their poorer definition in the electron density map (Figure 4.4).

The root-mean-square deviation of the Ca atoms between this structure and native cex-cd is 0.19 Å, which suggests that the ligand does not disturb significantly the conformation of cex-cd.
Figure 4.6. Difference Fourier electron density omit map of the disaccharide. a) On this stereoview, the contour level is 3.5σ. Both the disaccharide and Glu233 were omitted from the Fourier coefficient synthesis. b) Schematic diagram of the attached disaccharide seen in (a). c) Same as in (a), but rotated by 90° along the vertical axis.
The one region of highest deviation is the loop which joins the secondary structure elements α8 to β9: this region is located at the surface of the protein and have a higher mobility (Figure 4.4). Among the residues of this loop, Trp281 is the most significantly displaced and rotated. This change may allows to accomodate the substrate, as observed in the fluorocelllobiosyl-enzyme complex (Chapter 3).

Difference Fourier synthesis shows residual positive electron density at the active site of cex-cd (Figure 4.6). This residual object clearly represents two units of saccharide. Because cex-cd hydrates cellobial (S. G. Withers, personal communication), an eight hour soaking time may result in up to three different compounds, that is I, III, and V shown Figure 4.2: species II and IV are transition states and will not be observed experimentally. The absence of a significant element of electron density at the β position of the anomeric carbon indicates that the observed ligand can not be the hydrated reaction product 2-deoxy-celllobioside (compound V). So, this element of electron density may represent either compound I or III. Interpretation of this element of electron density as a cellobial molecule (compound I) followed by refinement leads to a close proximity (distance of 2 Å or less) between the proximal saccharide’s C1 and Glu233 O\textsuperscript{E1}, which is clearly unfavourable. It is thus reasonable to assume that the predominant species observed in the electron density map is not compound I, but the compound III, that is the covalent 2-deoxy-α-cellobiosyl-enzyme. Additional evidence for this species includes the connectivity observed in the electron density map between C1 and Glu233 O\textsuperscript{E1} and the weak signal at the C2 position (Figure 4.6). This latter observation may indicate a more flexible CH\textsubscript{2} group as opposed to the rigid double bond of cellobial. Taken together, this evidence suggests that the predominant disaccharide species observed is the covalent 2-deoxy-α-cellobiosyl-cex-cd complex (compound III).

The 2-deoxy-cellobioside binds in an extended crevice at the carboxy-terminal end of the (β/α)\textsubscript{8}-barrel motif of cex-cd (Figure 4.7) as anticipated (White et al., 1996). This crevice extends over 30 Å, which indicates that more glucose or xylose units may be accomodated in the binding region of cex-cd. This observation correlates with the carbohydrate-binding site of homologous
Figure 4.7. Surface charge distribution of cex-cd complexed to 2-deoxy-cellobioside. On this stereo-image, acidic regions are in red and basic regions in blue. The disaccharide is shown covalently linked to Glu233. The free carboxylate is the acid/base catalyst Glu127. The extended crevice displays predominantly negative charges and may accommodate more than two units of saccharide.

xylanases (Harris et al., 1994; Dominguez et al., 1995). The inspection of the charge distribution at the surface of cex-cd (Figure 4.7) indicates that this enzyme presents an acidic binding region to its ligand such that the side chains Glu127, Glu43, Asp170 and Asp235 are exposed in this crevice. Of these, Glu127 participates in catalysis as an acid/base catalyst. The disaccharide linked on Glu233 prevents the ionization of this amino acid: at the release of this adduct, Glu233 will act
Figure 4.8. **Active site of cex-cd with the bound ligand.** Atoms of the adduct 2-deoxy-\(\alpha\)-cellobiosyl-Glu233 are yellow for carbons and red for oxygens. The acid/base catalyst Glu127 is indicated in orange, the other conserved residues of cex-cd are green, and the water molecules are shown as blue spheres. Hydrogen bonds appear as dotted blue lines. Upon binding of the ligand to cex-cd, Gln87 (shown as an alanine) becomes disordered due to steric hindrance with C6-OH of the distal saccharide, and the tryptophanyl side chain of 281 is moved away by a rotation of 30° aound C1β-C1γ to accommodate the C6-OH of the proximal saccharide. The water molecule which forms a hydrogen bond with Glu127 is the only candidate as the nucleophile for the second displacement within the catalytic mechanism.

as a counterion for the stabilization of the positively charged transition state.

The recognition of the 2-deoxy-cellobioside by cex-cd involves hydrogen bonds and van der Waals interactions (Figure 4.8). Analysis of this recognition reveals that ten amino acids are involved in the disaccharide recognition, namely Glu43, Asn44, Lys47, His80, Trp84, Gln203, His205, Glu233, Trp273, and Trp281. These residues form six specific hydrogen bonds (Table 4.3) and 73 van der Waals interactions (Table 4.4) with the ligand. It is well recognized that
aromatic side chains can pack against saccharide rings, which is a common mode of ligand recognition in carbohydrate-binding proteins (Quiocio, 1986; Quiocio, 1988). Among the aromatic side chains in cex-cd, Trp273 is such an example of packing with the proximal saccharide of the ligand.

Binding interactions of 2-deoxy-cellobioside are similar to these previously observed for a 2-fluoro-cellobioside complex to cex-cd (Figure 4.9). As expected, the active site conformation of cex-cd is highly conserved upon binding of either disaccharide (Figure 4.9). The one difference between the chemistry of the two disaccharides concerns the substituent at the C2 position of the proximal saccharide. In the 2-fluoro-cellobioside complex to cex-cd, a fluorine atom accepts a hydrogen bond from Asn126 N^δ2. Because of the absence of substituent at the C2 position in 2-deoxy-cellobioside, its complex with cex-cd does not involve interaction with Asn126.

Table 4.3. Hydrogen bonds between 2-deoxy-cellobioside and cex-cd

<table>
<thead>
<tr>
<th>Hydrogen bond partner^a</th>
<th>Distal saccharide</th>
<th>Proximal saccharide</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O2</td>
<td>O3</td>
</tr>
<tr>
<td></td>
<td>Trp273 N^e1 (2.6), Glu43 O^e (2.8)</td>
<td>Asn44 N^δ2 (3.2), Wat497 (3.1)</td>
</tr>
<tr>
<td></td>
<td>O4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wat493 (3.3), Wat495 (2.9), Wat508 (3.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>O3</td>
<td>O6</td>
</tr>
<tr>
<td></td>
<td>His80 N^e (2.7), Lys47 N^e (2.7)</td>
<td>Wat363 (2.5)</td>
</tr>
</tbody>
</table>

^a Distance is shown in brackets. "Wat" stands for a water molecule.

It has been reported that the catalytic intermediate of retaining β-1,4-glycosyl hydrolases forms a nearly half-chair conformation. The crystal structure of a Thr26→Glu mutant of T4-lysozyme (Kuroki et al., 1993) and the structure of hen-egg white lysozyme with the product MurNac-GlcNAc-MurNac (Strynadka and James, 1991) both reveal a distorted sugar toward
### Table 4.4. van der Waals contacts between 2-deoxy-cellobioside and cex-cd

<table>
<thead>
<tr>
<th>Distal saccharide:</th>
<th>Proximal saccharide:</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Gln203 (C^5, O^e1), Glu233 (C^7, C^5, O^e2), His205 N^e2, Wat580</td>
</tr>
<tr>
<td>C2</td>
<td>His80 (N^e2, C^c1), Glu233 (C^5, O^e2, O^c1), Wat580</td>
</tr>
<tr>
<td>C3</td>
<td>Lys47 N^5, Trp273 C^c2, Glu233 O^c1, Trp273 C^c^2, His80 (C^c1, N^e2)</td>
</tr>
<tr>
<td>O2</td>
<td>Trp84 (C^c^3, C^n^2), His80 (C^d^2, C^c^1), Lys47 C^e</td>
</tr>
<tr>
<td>C4</td>
<td>Trp273 N^c^1, Glu233 O^c^1, Wat363</td>
</tr>
<tr>
<td>O4</td>
<td>Lys47 N^5, Trp273 (C^c^2, C^c^2, N^c^1)</td>
</tr>
<tr>
<td>C5</td>
<td>Trp273 N^c^1, His205 N^e^2, Glu233 O^c^1, Wat363</td>
</tr>
<tr>
<td>O5</td>
<td>His205 (C^c^1, C^d^2, N^e^2), Gln203 O^c^1, Glu233 (C^5, O^e^1)</td>
</tr>
<tr>
<td>C6</td>
<td>Trp281 (C^7, C^d^2, C^c^2, C^c^3, C^c^2, C^c^3), Wat363</td>
</tr>
<tr>
<td>O6</td>
<td>Trp281 (C^d^2, C^c^3, C^c^3), Wat580</td>
</tr>
</tbody>
</table>
Figure 4.9. Comparison of two covalent glycosyl-cex-cd complexes. The enzyme’s side chains of the 2-deoxy-α-cellobiosyl-cex-cd complex are shown in green, except for Glu127 and Glu233 which are colored orange. Overlaid in blue are the side chains of the 2-fluoro-2-deoxy-α-cellobiosyl-cex-cd complex (White et al., 1996). These two disaccharides differ by their substituent at the C2 position of their proximal saccharide: either a hydrogen or a fluorine atom. For clarity, the disaccharides are not shown.

transition state planar conformation. The inspection of the geometry of the 2-deoxy-cellobioside in cex-cd shows that both saccharide units form a “full chair” conformation (Figures 4.8 and 4.10): the proximal 2-deoxy-glucose ring resembles both the 2-fluoro-2-deoxy-glucose attached to Glu233 and a β-D-glucose more than a cellobial molecule.
Figure 4.10. Comparison of a 2-deoxy-\(\alpha\)-cellobiosyl-Glu233 (blue) to a) 2-fluoro-2-deoxy-\(\alpha\)-cellobiosyl-Glu233 complex previously reported, b) cellobial molecule, and c) \(\beta\)-D-glucose. Atoms of the overlayed molecules shown in stereo are colored red for oxygens and yellow for carbons.
CONCLUSION

The crystal structure of cex-cd with cellobial has been determined to a resolution of 2 Å with an $R$-factor of 0.19 and good stereochemistry for the model. In this complex, one disaccharide binds proximal to both the nucleophile Glu233 and the acid/base catalyst Glu127 of cex-cd. Cex-cd converts the bound cellobial molecule to its 2-deoxy-α-cellobiosyl homologue, which remains covalently bound to the nucleophile Glu233 with inverted, or α, anomeric configuration. The proximal unit of saccharide of the ligand forms a "full chair" conformation, which is similar to a relaxed glucose molecule. This 2-deoxy-cellobiosyl moiety binds to the same location as the 2-fluoro-cellobiosyl to cex-cd and makes a similar network of interactions to the enzyme.
REFERENCES


Chap. 5

General Discussion
Future Directions
Concluding Remarks
GENERAL DISCUSSION

The crystal structure of the catalytic domain of the β-1,4-glycanase cex and its covalent complexes with disaccharides now provide clues on both the folding motif of this enzyme and the mechanism of retaining β-glycosyl hydrolases. This discussion contains four sections. Firstly, the validity of the X-ray diffraction results is assessed with quantitative indicators. The validity of many of the structural and functional conclusions in this thesis depends on the high quality and accuracy of the crystal structure determination. Secondly, because cex-cd is catalytically active in the crystalline environment, its structure obtained by X-ray diffraction is thus biologically relevant. Thirdly, seven structural characteristics of the catalytic intermediate are presented. These characteristics may form the basis for the understanding of the broad range of retaining β-glycosyl hydrolases. Fourthly, a previously unexpected structural similarity with the enzyme haloalkane dehalogenase leads to clues about the anomeric configuration outcome for catalysis by retaining β-glycosyl hydrolases. The argument is based on the restricted accessibility of a water molecule in the vicinity of the catalytic nucleophile.

QUALITY ASSESSMENT OF CEX-CD MODEL

In order to draw conclusions about biological implications of the three-dimensional structures obtained, the quality of the X-ray diffraction results should be assessed for both the model stereochemistry and the crystallographic interpretation.

Firstly, the crystallographic indicators are reasonable. The R-factors are in the order of 19 to 20%, which are expected values for comparable structures at similar resolution of 2.0 to 1.8 Å (Laskowski et al., 1993). However, it is well recognized that a low R-factor may still be obtained with a structure containing significant errors (Kleywegt and Jones, 1995). That is, a low R-factor indicates little residual crystallographic information non-interpreted by a given model, whatever the model is. It is thus imperative to consider proper controls through the model refinement. One such control is the free R value, in which a set of 5 to 10% of randomly chosen X-ray diffraction
data is set aside as a control of refinement (Brünger, 1992). The monitoring of this test set of data provides an unbiased means to follow the normal progression of the refinement, that is the improvement in optimizing the model against the crystallographic data. Overinterpretation of the data for example would lead to a higher free R value. For cex-cd, this value reaches 28 to 30%, which is an acceptable value (Kleywegt and Brünger, 1996).

Secondly, the interpretation of the electron density has been extensively analyzed for the 2-fluorocellobioside complex with simulated annealing omit maps for all the 312 residues of cex-cd. Groups of 10 to 13 residues were omitted from the model prior to a round of simulated annealing in which the temperature was decreased from 1000 to 300 K, and then followed by 250 cycles of energy minimization. The excluded segments of amino acids were then re-analyzed with an electron density map generated from the simulated-anneal-refined residual model. In effect, this process significantly reduces the bias in the electron density map calculation, due to model inclusion with the refinement (Hodel et al., 1992). The interpretation of this stretch of amino acids in the newly generated electron density map was then quantified by residue-based real-space correlation (Jones et al., 1991). High correlation values obtained with this indicator validates the model of cex-cd (see Chapter 3).

Thirdly, the resulting protein model has to make chemical sense. It is thus reasonable to restrict the divergence of the model’s stereochemistry close to ideal values. Divergence values are about 1.5 to 2° on bond angles and 0.011 to 0.017 Å on bond lengths for cex-cd. These values are within acceptable ranges when compared to other structures of good quality (Laskowski et al., 1993). An additional parameter of stereochemistry refers to the analysis of the main chain dihedral angles (φ,ψ), known as the Ramachandran plot (Ramakrishnan and Ramachandran, 1965). Such an analysis reveals that the angles of the main chain of cex-cd all appear in "allowed" or "most favourable" regions on a Ramachandran plot.

Taken together, good values obtained for the above validators support the reliability of the three-dimensional structure of cex-cd.
**Cex-cd Folding Motif**

The X-ray diffraction analysis has established that the catalytic domain of cex folds into a (β/α)₈-barrel. The high resolution (1.8 Å) crystal structure of cex-cd permits the positioning and the orienting of its 312 N-terminal residues. This inherently reveals the secondary structure elements and the tertiary structure of the enzyme. This folding motif is now supported by crystal structures of homologous Family F (family 10) enzymes. These homologs are the high resolution structure of both *Pseudomonas fluorescens* xylanase A (Harris et al., 1994; Harris et al., 1996) and *Clostridium thermocellum* xylanase XynZ (Dominguez et al., 1995), and the 2.6 Å resolution structure of *Streptomyces lividans* xylanase A (Derewenda et al., 1994).

Because a residual saccharide is observed upon adding 2F-DNPC to cex-cd crystals, it strongly suggests that this enzyme is catalytically active in the crystalline environment, at least for the first step of catalysis. The catalytic intermediate, a 2-fluoro-2-deoxy-cellobioside residual saccharide, is clearly present in the active site of cex-cd as shown in the crystal structure of this complex. This observation implies the cleavage of 2F-DNPC to 2-fluoro-2-deoxy-cellobioside. Moreover, because the residual saccharide covalently complexes with the enzyme, it suggests that the cleavage of 2F-DNPC must have occurred. The cleavage does not take place only with proteins of the surface of the crystal, but involves molecules packed in the crystal. The occupancy of the covalent ligand in the crystal structure is clearly above 0.5, which means the ratio of bond-ligand to protein involves at least half of the enzyme molecules. As a comparison, the calculated amount of enzyme molecules covering the surface layer of a crystal of cex-cd represents 1:25000 of the crystal content (see Appendix A), which cannot account for the protein-ligand occupancy of more than 0.5. Therefore, cex-cd molecules found inside the crystal must have been involved in the formation of a 2-fluoro-2-deoxy-cellobioside complex, and thus, cex-cd is catalytically active in the crystalline environment. Therefore, the crystal structure obtained for cex-cd is biologically relevant.
STRUCTURAL CHARACTERISTICS OF THE CATALYTIC INTERMEDIATE

Until the present results, structural characterization of the catalytic intermediate of β-1,4-glycosyl hydrolases was restricted due to the absence of a detailed three-dimensional structure. Combining the results from both the 2-deoxy-cellobiosyl-cex-cd and 2-fluoro-2-deoxy-cellobiosyl-cex-cd complexes leads to the following features of the catalytic intermediate (Figure 5.1):

1. The intermediate is a covalent glycosyl-enzyme species with inverted configuration α at the anomeric carbon C1 of the proximal saccharide. This configuration was expected on the basis of α-secondary deuterium kinetic isotope effects for both the glycosylation and deglycosylation steps, and also on the basis of trapping of the intermediate with fluoro-saccharides followed by analysis with one-dimensional 19F-NMR studies. The crystallographic results obtained with cex-cd now confirm these previous inferences.

2. The proximal saccharide is cyclic. This observation clearly challenges the previously proposed mechanism by the group of M. Karplus (Post and Karplus, 1986). In this latter mechanism based on stereoelectronic considerations, the retaining β-glycosyl hydrolase lysozyme would hydrolyze its substrate through an endocyclic opening of the binding saccharide. This mechanism of Post and Karplus does not explain established experimental evidence, as explained by Sinnott in an authoritative critique (Sinnott, 1993). For example, kinetic isotope effects, which are typical of an ionic transition state, can not be explained by the mechanism of Post and Karplus. However, a cyclic residual saccharide, as observed in the crystal structures of cex-cd, is consistent with the kinetic isotope effect results.

3. The proximal saccharide adopts a "full chair" conformation. Both 2-fluoro-cellobioside and 2-deoxy-cellobioside complexes illustrate this conformation. Moreover, both these complexes with cex-cd are catalytically competent, that is the covalent complex does turn over by releasing the reaction product and the free enzyme; this validates their conformation as true catalytic intermediates.
Figure 5.1. Structural characteristics of the catalytic intermediate of the retaining β-1,4-glycanase cex-cd. The features may be conserved in the other retaining β-1,4-glycosyl hydrolases which make use of two carboxylates as the key catalytic residues. The hatched planes are residues (Trp273, Gln203) which form a close stacking with the nucleophile.

This "full chair" conformation is in contrast to previous models. Because the proximal saccharide at the transition states bears a "sofa" geometry which is planar at C5-O5-C1-C2, it has long been thought that the catalytic intermediate approaches this distortion. For instance, crystallographic results of both the analogous lysozyme with the reaction product MurNAc-GlcNAc-MurNAc trisaccharide (Kelly et al., 1979; Strynadka and James, 1991), and the Thr26 to Glu mutant of T4-lysozyme with a covalent substrate (Kuroki et al., 1993) suggest a conformation somewhat between the "sofa" and the "chair" forms for the catalytic intermediate. Because of their inherent non-natural modifications, these systems may not be representative of the natural catalytic intermediate, whereas the complexes with cex-cd are. In other reports, the "full chair" conformation was observed for epoxide labeling agent forming a "dead-end" attachment in the active site of lysozyme (Moult et al., 1973) and of β-1,3-1,4-glucanase (Keitel et al., 1993). However, because of the additional carbon atoms between the saccharide and the enzyme's nucleophile, the geometry of these glycosyl-enzyme complexes may not be representative of the catalytic intermediate.
4. The proximal saccharide binds at the \textit{syn} position of the (now) esterified side chain of the nucleophile Glu233 of cex-cd. By similarity to small molecules, this position is a preferred location to the \textit{anti} in chemical reactions (Gandour, 1981).

5. The orientation of Glu233 side chain is such that it is found nearly coplanar with the cyclic oxygen O5 of its attached saccharide (Figure 5.1). As a consequence, one lone pair of electrons of O5 is inherently coplanar with the side chain of Glu233. It is not known what is the influence of this coplanarity, although it may have long range stereoelectronic implications.

6. The other key catalytic residue, the acid/base catalyst Glu127, is located 4.5 Å from the saccharide C1 and 5.5 Å from the nucleophile Glu233. This latter correlates with other retaining β-glycosyl hydrolases whose crystal structures are known (McCarter and Withers, 1994; White \textit{et al.}, 1994). The distance of 4.5 Å is presumed to allow the insertion of a water molecule between the acid/base catalyst and C1, which will displace the residual saccharide; the product will thus be formed with retention of configuration at C1. The side chain of both the acid/base catalyst and the nucleophile will be located on the same side of the polysaccharide substrate, as shown for the Endocellulase E1 from \textit{Acidothermus cellulolyticus} in complex with cellotetraose (Sakon \textit{et al.}, 1996). Hence, in addition to the distance, the geometrical disposition of the acid/base catalyst relative to the nucleophile is likely to be of great importance.

7. The hydroxyl substituent at C2 of the natural saccharide substrate is inferred to donate a hydrogen bond to the carbonyl oxygen of the ester group of the enzyme's nucleophile Glu233 and is observed to accept a hydrogen bond from the amido group of a conserved asparagine residue. Previous kinetic data using a variety of substrates have shown that this substituent at the C2 position contributes at least 8 kcal/mol of binding energy to the transition state stabilization (Roeser and Legler, 1981; Wolfenden and Kati, 1991; McCarter \textit{et al.}, 1992). Inspection of the crystal structure of cex-cd covalently complexed with a fluorocellobioside as a catalytic intermediate reveals a close interaction (3.1 Å) with good angle values between the fluorine substituent at the C2 position and the nucleophile Glu233 O\text{E2}. Thus, a similarly positioned hydroxyl group at C2 is
inferred to have all the characteristics of a hydrogen bond; this hydrogen bond provides a good rationale for explaining the high binding energy attributed to the C2 substituent. It may be a low barrier short hydrogen bond of the type described elsewhere (Cleland and Kreevoy, 1994; Frey et al., 1994; Tobin et al., 1995).

SIMILARITY WITH HALOALKANE DEHALOGENASE

The catalytic machinery of cex-cd shows some resemblance to the active site of haloalkane dehalogenase (DhIA), an unrelated enzyme. Although these enzymes differ in sequence, architecture, and substrate, they both host a similar catalytic triad "carboxylate-histidine-carboxylate" at their active centre and they both hydrolyze their respective substrates by a double displacement through a covalent catalytic intermediate. Comparison of their active sites, that is the local environments of their catalytic nucleophiles, leads to a better understanding of catalysis by cex-cd.

DhIA catalyzes the nucleophilic substitution of a chlorine atom for a hydroxyl group. Using both an Asp to Asn mutant and mass spectrometric analysis of $^{18}O$ incorporation from $H_2^{18}O$ (Pries et al., 1994) it has previously been shown that Asp124 is the key catalytic nucleophile in the active site of Xanthobacter autotrophicus GJ10 DhIA. This nucleophile has also been shown to form a covalent attachment to the residual substrate upon mutation of His289 to Gln (Pries et al., 1995). The catalytic triad Asp124-His289-Asp260 of DhIA (Franken et al., 1991; Verschueren et al., 1993) is structurally similar to the triad Glu233-His205-Asp235 of cex-cd. Because these triads are within the immediate environment of the catalytic nucleophile of both enzymes, these triads are presumed to be functionally significant.

Both DhIA and cex-cd make use of a nucleophilic water to displace a covalent residual substrate-nucleophile complex; this represents the second displacement within their respective catalytic mechanism. A closer analysis shows that the mode of hydrolysis of the covalent intermediate of DhIA is different for the displacement of the residual substrate than for its analog in cex-cd. In DhIA, a nucleophilic water molecule attacks at the carbon of the ester group, not at the substrate linking carbon, as demonstrated with catalysis in $^{18}O$ labeled water (Pries et al., 1994).
Moreover, the location of that nucleophilic water has been assigned in DhlA by X-ray crystallography and shown to be in a favourable binding environment (Verschueren et al., 1993). This nucleophilic attack on the carbon of the ester leads to the substitution of an atom of the nucleophilic residue. In other words, the enzyme DhlA does not maintain its integrity through catalysis, because a water molecule substitutes an oxygen atom of this enzyme which leads to a labeled protein. This mode of nucleophilic attack by a water in DhlA is fundamentally different from the one of cex-cd: hydrolysis of the catalytic intermediate of cex-cd involves an oxocarbonium ion-like transition state followed by a nucleophilic attack of a water on the residual substrate, not on a residue of the enzyme, which ultimately leads to the integrity of cex-cd through catalysis.

Steric hindrance may explain the difference in hydrolysis of cex-cd versus DhlA. Examination of the environment of cex-cd’s nucleophile Glu233 shows that this residue is embedded between the side chains of Trp281 and Gln203. A necessary consequence is the restricted access for a water to insert at the vicinity of the carbon of the esterified nucleophile at the catalytic intermediate step. Another water forms a hydrogen bond with Glu233, but it is located in plane with this esterified Glu233 in the anti position. This latter water molecule may not be a nucleophile for the esterified Glu233 because of its relative position; a water is more favourably reactive toward an ester if positioned normal to the plane of the ester, not anti. Thus, this restricted accessibility of a water to the correct position near the nucleophile Glu233 of cex-cd may be one explanation that the attack of a water on the covalent glycosyl-Glu233 catalytic intermediate is performed at the residual substrate, as opposed to the more chemically reactive ester carbon.

Furthermore, the above steric hindrance in cex-cd may explain why a double displacement catalytic mechanism leads to the retention of the saccharide anomeric configuration. Indeed the observed preference of nucleophilic attack of a water onto the residual saccharide C1, as opposed to the esterified Glu233, implies a second inversion of configuration through catalysis; the first inversion is the formation of the covalent intermediate by the formation of an α link with the nucleophile Glu233, given a substrate of β configuration. Thus, the anomeric configuration of the C1 of the saccharide undergoes changes of the form β → α → β. This mode differs from the
following hypothetical situation: If a nucleophilic water attacks the electrophilic centre of the ester linkage at its carbon, the residual saccharide will be released without a second inversion of configuration, thus following a mode $\beta \rightarrow \alpha \rightarrow \alpha$ which is an overall inversion of configuration; however, this is not observed for cex-cd. Experiments clearly demonstrate that the hydrolysis of saccharides by cex-cd yields an overall retention of the configuration $\beta$ of the saccharide by a double displacement. To ensure this retention of configuration, the more labile carbon of the esterified Glu233 must be inaccessible to an incoming nucleophilic water molecule; both Trp281 and Gln203 side chains contribute to this required steric hindrance. Thus, for cex-cd, a double displacement explains the retention of configuration through its catalytic mechanism.

**Future Directions**

Knowledge of the detailed three-dimensional structure of cex-cd with and without ligand may help in protein engineering of this enzyme. More specifically, the side chains of residues Gln87 and Trp281 are the only sites of significant movement upon binding of the cellobiosides. A closer analysis reveals that these side chains are moved to accommodate the C6-OH groups of both units of saccharide. These 2 carboxymethyl groups define precisely the difference between xylan and cellulose, and thus provide clues about preferential binding of xylobiosyl compared to cellobiosyl compounds (activity ratio 50:1). So, it is reasonable to speculate that point mutation at these sites (87 and/or 281) to shorter side chains may reduce constraints imposed by carboxymethyl groups in binding. In doing so a mutant cex-cd may display a reduced specificity, and bind both xylan and cellulose. Such a bifunctional specificity is a test of this fundamental hypothesis and may have industrial applications in the breakdown of complex polysaccharide structures, such as hemicelluloses in the pulp and paper industries, to their saccharide components.

The active site conformation of cex-cd may help to understand critical conserved elements which define catalysis by enzymes of its family. Means commonly used to extract conserved features between proteins include sequence alignment as a first step. Conserved or identical residues may be recognized, but little information about their functionality can be obtained by this
means alone. Alignment of secondary structure elements and folding motif comparison may help to provide a context for the conserved amino acids. As more and more crystal structures of glycosyl hydrolases become available, their comparison may help to identify the local environment of catalytic residues. Placed in broader perspective, the comparison of the detailed three-dimensional structures of glycosyl hydrolases may help to define specific conserved characteristics of their active sites. Such characteristics may either be conserved residues, conserved side chain functionality such as a carboxylate, hydrogen bond ability of similarly placed side chains or hydrophobic local properties. In this context the crystal structure of a defined catalytic event, such as the covalent catalytic intermediate of cex-cd, provides a framework on which to define a functional map of the active site of retaining β-glycosyl hydrolases. Clearly, amino acid sequence alignment alone is insufficient for such a prospect.

Active site comparison of homologous enzymes may further help in areas of rational drug design. For example, the structure-based design of 4-guanidino-neu5Ac2en has been shown to be effective in the 10^{-11} to 10^{-10} M range as an inhibitor of influenza virus neuraminidase (Colman, 1995); neuraminidase is a β-1,4-glycosyl hydrolase necessary for the virion infectivity. In the context of rational drug design, the three-dimensional structure of cex-cd and its complex to disaccharides may provide clues in the rational design of catalytic intermediate mimicking compounds for β-1,4-glycosyl hydrolases.

**Concluding Remarks**

The crystal structures of cex-cd and its complexes with disaccharides have now been described. Special attention to quantitative crystallographic validators brings confidence to the qualitative interpretation of the enzyme folding motif and mechanism. The structure of the covalent catalytic intermediate with an emphasis on previously unavailable geometrical considerations gives insights into specific structural characteristics of such an event in catalysis. Taken together, these results open a window of opportunities for the exploration of both the cex-cd structure and the catalytic intermediate of this class of enzymes.
REFERENCES


Appendix A

This appendix focusses on the evaluation of the ratio surface to volume for a crystal of cex-cd. To simplify the calculations, this ratio is represented in terms of amount of molecules and a single shape of crystal is considered (Figure A.1).

Figure A.1. Geometry of a typical crystal of cex-cd.

1. SURFACE AREA OF THE CRYSTAL

Along the following faces of the crystal, the surface $S$ is:

- faces 010: $S_{010} = a_1 \cdot a_5$
- faces 001: $S_{001} = a_1 \cdot a_2$
- faces 101: $S_{101} = \frac{a_2}{4}\sqrt{4a_3^2 + (a_5 - a_4)^2}$
- faces 110: $S_{110} = \frac{a_2 + a_4}{4}\sqrt{4a_3^2 + a_2^2}$

So, the total surface of the crystal:

$$S_{\text{crystal}} = 2S_{010} + 2S_{001} + 4S_{101} + 4S_{110}$$

$$\therefore S_{\text{crystal}} = 2\left(a_1(a_5 + a_2) + a_2\sqrt{a_3^2 + \left(\frac{a_5 - a_4}{2}\right)^2} + a_5 + a_4\sqrt{a_3^2 + \frac{a_2^2}{4}}\right)$$
Now, let \( d \) be the thickness of one layer of protein molecules, approximating for \( n \) layers of protein molecules on the surface of the crystal, and assuming that \( n \) is small, gives the number of molecules \( S_{\text{molecules}} \) on the surface of the crystal:

\[
S_{\text{molecules}} = n d S_{\text{crystal}}
\]

2. VOLUME OF THE CRYSTAL

To simplify this calculation, the crystal is fragmented in three parts which have the following volumes:

- central block: \( V_1 = a_1a_2a_5 \)
- square-base pyramid: \( V_2 = \frac{a_2a_3(a_5 - a_4)}{3} \)
- side’s central triangle: \( V_3 = \frac{a_2a_4a_3}{2} \)

So, the total volume of the crystal is:

\[
V_{\text{crystal}} = V_1 + 2V_2 + 2V_3
\]

\[
= a_1a_2a_5 + 2\frac{a_2a_3(a_5 - a_4)}{3} + 2\frac{a_2a_4a_3}{2}
\]

\[
\therefore V_{\text{crystal}} = a_2\left(a_1a_5 + \frac{a_4}{3}(2a_5 + a_4)\right)
\]

3. NUMERICAL CALCULATION

The dimensions of a typical crystal of cex-cd are \( a_1=0.5 \) mm, \( a_2=0.5 \) mm, \( a_3=0.15 \) mm, \( a_4=0.1 \) mm and \( a_5=0.6 \) mm and the radius \( (d) \) of the protein is about 40 Å. The surface to volume ratio is thus given by:

\[
\frac{nds_{\text{crystal}}}{V_{\text{crystal}}} = \frac{n \cdot 7.2 \times 10^{-6} \text{ mm}^3}{0.183 \text{ mm}^3} = n \cdot 3.9 \times 10^{-5}
\]

So, the surface of the crystal represents the molecule proportion shown in Table A.1. This means that the outer layer of protein molecules on the crystal surface includes 1:25000 of the molecules of
the crystal. Similarly, if we consider that the outer three hundred layers of molecules delineates the surface of the crystal, this outer moiety represents only about 1:85 of the total molecules of the crystal.

Table A.1. Molecules proportion of the outer \( n \) layers of molecules of the crystal.

<table>
<thead>
<tr>
<th>( n ) layers</th>
<th>( \frac{nS_{\text{crystal}}}{V_{\text{crystal}}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( 3.9 \times 10^{-5} )</td>
</tr>
<tr>
<td>50</td>
<td>0.0020</td>
</tr>
<tr>
<td>300</td>
<td>0.012</td>
</tr>
</tbody>
</table>