CRYSTALLOGRAPHIC STUDIES OF CARBOHYDRATE MODIFYING PROTEINS: MECHANISMS OF ACTION IN POLYSACCHARIDE BINDING AND CATALYSIS

by

Valerie Notenboom

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 Valerie Notenboom 2000
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.

0-612-49935-9
Crystallographic Studies of Carbohydrate Modifying Proteins: Mechanisms of Action in Polysaccharide Binding and Catalysis

Doctor of Philosophy 2000
Valerie Notenboom

Department of Medical Biophysics
University of Toronto

Abstract

Enzymatic degradation of cellulose and hemicellulose abundant in plant biomass is achieved by the activity of diverse hydrolytic enzyme systems secreted by microbial organisms. Such enzymes can be found incorporated within a cellulosome or expressed in soluble form. Subtle differences in specificities among the enzymes complement each other to degrade the diverse polymeric substrates efficiently. These enzymes frequently embody a modular structure, consisting of one or more substrate binding domains and a catalytic domain of the glycosyl hydrolase enzyme class.

This thesis explores the three-dimensional structures and mechanisms of action in examples of each of these two types of modules, a cellulose binding domain expressed by the thermostable Thermotoga maritima and a β-1,4-glycanase from Cellulomonas fimi with xylanase and cellulase activity.

The first part of this manuscript describes the X-ray crystal structure of the 17 kDa family 9 binding module from T. maritima (CBM9-2) to 1.9 Å resolution, as determined by multiple-wavelength anomalous dispersion using selenium-substituted methionine residues as the source for anomalous scattering. The structure reveals a β-sandwich fold stabilized by three calcium ions, and suggests a novel mode of substrate binding through a sandwich-like binding pocket formed by two tryptophan side chains. The crystal structures of CBM9-2 in complex with several ligands allows for
interpretation of the mechanism of binding in this module, which was determined to interact specifically with reducing ends.

The remainder of the thesis addresses mechanistic aspects of the C. fimii family 10 catalytic domain, Cex. Cex hydrolyzes xylan, and to a lesser extent cellulose, with retention of β-configuration at the anomeric centre of the cleaved O-glycosidic bond. This implies a 2-step SN1-like catalytic mechanism involving a covalent enzyme-substrate intermediate flanked by energetically unfavourable transition states with significant oxocarbenium cationic character. Through extensive analysis by X-ray crystallography as well as kinetic studies with mutants, a detailed mechanistic picture of Cex has emerged describing unique enzyme characteristics such as substrate specificity and transition state stabilization. Furthermore, insights gained throughout the project have led to the design of strong mechanism-based inhibitors, whose crystal structures bound to the enzyme in turn have been characterized and described.
# Table of Contents

**Introduction and Thesis Overview**................................................................. 1

**Chapter 1** Crystal Structures of the Family 9 Carbohydrate-Binding Module from *Thermotoga maritima* Xylanase 10A in Native and Ligand Bound Forms................................................................. 21

- **Abstract**
- **Introduction**
- **Materials and Methods**
  - **Materials**
  - **Bacterial strains and plasmid**
  - **Media and growth conditions**
  - **Production and purification of wild type CBM9-2**
  - **Preparation of reduced cellulose and barley β-glucan**
  - **Production and purification of seleno-methionine containing CBM9-2**
  - **Crystallization of native and Semet CBM-9.2**
  - **Data collection and processing**
  - **Determination of selenium atom positions and MAD phasing**
  - **Model building and refinement**

- **Results and Discussion**
  - **Structure of CBM9-2**
  - **Structure of CBM9-2 in complex with cellobiose**
  - **Structure of CBM9-2 with glucose**
  - **Exclusive binding of CBM9-2 to the reducing ends of polysaccharides**

- **References**

**Chapter 2** Exploring the Cellulose/Xylan Specificity of the β-1,4-Glycosidase Cex from *Cellulomonas fimic* through Crystallography and Mutation.......................................................................................................................... 50
### Chapter 3 Insights into Transition State Stabilization of the \(\beta-1,4-\) Glycosidase Cex by Covalent Intermediate Accumulation

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>79</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>80</td>
</tr>
<tr>
<td>- Chemicals, media components and enzymes</td>
<td>82</td>
</tr>
<tr>
<td>- Bacterial strains and plasmids</td>
<td>83</td>
</tr>
<tr>
<td>- Construction of mutants</td>
<td>83</td>
</tr>
<tr>
<td>- Production of mutant proteins</td>
<td>83</td>
</tr>
<tr>
<td>- Characterization of mutants</td>
<td>84</td>
</tr>
<tr>
<td>- Kinetic studies</td>
<td>84</td>
</tr>
<tr>
<td>- Crystallization and data collection of Cex-DM</td>
<td>85</td>
</tr>
<tr>
<td>- Phase approximation</td>
<td>85</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>88</td>
</tr>
<tr>
<td>- Purification and structural characterization</td>
<td>88</td>
</tr>
<tr>
<td>- pH dependence</td>
<td>88</td>
</tr>
<tr>
<td>- Kinetic analysis</td>
<td>89</td>
</tr>
<tr>
<td>- Crystal structure of Cex-cd E127A/H205N</td>
<td>96</td>
</tr>
<tr>
<td>References</td>
<td>103</td>
</tr>
</tbody>
</table>
Chapter 6  Tautomerization of an Isofagomine-Derived Glyconolactam in the Active Site of Cex: Implications for the Role of Asn126 in Substrate

Binding........................................................................................................................................ 171

Abstract.......................................................................................................................................... 172

Introduction...................................................................................................................................... 173

Materials and Methods..................................................................................................................... 174

• Crystal preparation, data collection and structure determination............................................. 174

Results and Discussion.................................................................................................................... 176

• Structure of wtCex with the isofagomine lactam..................................................................... 176

• Inhibition of the isofagomine lactam in Cex N126A............................................................... 179

References........................................................................................................................................ 181

Concluding Remarks and Future Directions.................................................................................. 184

Appendix A  Nomenclature for saccharides and saccharide conformations............................... 190

Appendix B  A special case of MAD: pseudo-MIR using MAD and native datasets................. 194
List of Figures

1.1-a Ribbon diagram of the crystal structure of cellobiose bound CBM9-2 35
1.1-b Topology map of the three-dimensional fold of CBM9-2 36
1.2 Calcium-binding interactions in three calcium binding sites of CBM9-2 37
1.3 2IFol-IFcl difference density for cellobiose bound to CBM9-2 39
1.4 Schematic representation of cellobiose and glucose interactions with CBM9-2 40
1.5 2IFol-IFcl difference density for glucose bound to CBM9-2 42
1.6 Stereo representation of a superposition of cellobiose and glucose in the binding site of CBM9-2 42

2.1 Enzyme inactivation by cleavage of 2",4"-dinitrophenyl 2-fluoro-2-deoxy-β-xylobiose 59
2.2 Active site interactions of the covalent 2F-xylobiosyl-Cex-cd structure 62
2.3-a Stereo diagram of 2IFol-IIFcl electron density in 2F-xylobiosyl-Cex-cd 64
2.3-b Stereo representation of a superposition of uncomplexed Cex-cd, 2F-cellobiosyl-Cex-cd and 2F-xylobiosyl-Cex-cd 65
2.4 Plot of log(kcat/Km) for Gln87Met mutant vs. wild type enzyme for β-glycosides 71

3.1 Proposed catalytic mechanism of Cex 80
3.2 Dependence of kcat/Km upon pH for wtCex and mutants for PNPC hydrolysis 89
3.3-a Stereo representation of 2IFol-IIFcl electron density in Cex H205N/E127A with cellobiose 95
3.3-b Stereo representation of a superposition of the covalent structures of Cex H205N/E127A with cellobiose and wtCex with 2F-cellobiose 95
3.4-a -1 subsite interactions in the covalent intermediate structure of wtCex with 2F-xylobiose 99
3.4-b -1 subsite interactions of the fam 5 A. cellulolyticus E1 with cellotetraose 99
3.5 Proposed role of H-bonding between 2’OH and Glu233 during catalysis 102
4.1 Active site cartoon of Cex in covalent complex with 2F-xylobiose
4.2-a Schematic representation of mutagenesis approach for Asn126Ala
4.2-b Schematic representation of mutagenesis approach for Asn169Ala
4.3 Protein expression levels of wtCex and Cex Asn126Ala and Asn169Ala
4.4 Structures of the β-glycosides DNPC and PNPC
4.5 Reaction mechanism of Cex, with rate constant expressions
4.6-a Stereo representation of the active site of N169A
4.6-b Stereo representation of the active site of N126A

5.1-a Reaction mechanism of Cex,
5.1-b Reaction mechanism of Bcx
5.2 Xylobiose derived inhibitor series against Cex
5.3 1Fol-1Fc1 density of four inhibitors (a-d) in Cex
5.4-a Anti- and syn protonation trajectories in β-glycosidases
5.4-b Proposed anti- and syn-protonation of lactam oxime in β-glycosidases
5.5 Stereo diagram of a superposition of Cex in complex with imidazole, and 2F-xylobiose
5.6 Interactions of Tyr69 with the covalent intermediate in Bcx

6.1 Potential tautomeration of isofagomine lactam in enzyme bound form
6.2 2!Fol-1Fc1 electron density around the active site of Cex in complex with isofagomine lactam
List of Tables

1.1-a Synchrotron MAD data collection statistics of CBM9-2
1.1-b 'Home' rotating anode data collection statistics of CBM9-2 complexes
1.2 Anomalous, dispersive and isomorphous differences between datasets
1.3 Structure refinement and model quality statistics

2.1 Data collection statistics for 2F-xylobiosyl-Cex-cd
2.2 Structure & refinement statistics
2.3 Comparison of ideal and measured angles
2.4 Kinetic parameters of β-glycosides in Cex and Cex-Gln87Met

3.1-a Data collection statistics for Cex H205N/E127A with cellobiose
3.1-b Structure and refinement statistics
3.2 Michaelis-Menten parameters for hydrolysis of aryl cellobiosides by Cex and mutants

4.1 Data collection statistics for Cex N126A and N169A
4.2 Structure refinement and model quality statistics of N126A and N169A
4.3 Michaelis-Menten parameters for wtCex, N126A and N169A

5.1 Data collection statistics for Cex in complex with 4 inhibitors
5.2 Structure refinement and model quality statistics of 4 inhibitor complexes

6.1 Data collection statistics for Cex in complex with isofagomine lactam
6.2 Structure refinement and model quality statistics
6.3 Kinetic parameters and for Cex and Cex N126A against five xylobiosederived iminosugars
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>Ångström, $10^{-10}$ m</td>
</tr>
<tr>
<td>Bcx</td>
<td><em>Bacillus circulans</em> xylanase</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>CBD</td>
<td>Cellulose-binding domain</td>
</tr>
<tr>
<td>CBM</td>
<td>Carbohydrate-binding module</td>
</tr>
<tr>
<td>Cex</td>
<td><em>Cellulomonas fimis</em> exoglycanase</td>
</tr>
<tr>
<td>Cex-cd</td>
<td>Cex-catalytic domain</td>
</tr>
<tr>
<td>CMC</td>
<td>Carboxymethyl cellulose</td>
</tr>
<tr>
<td>DNPC</td>
<td>2&quot;,4&quot;-dinitrophenyl cellobiose</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediamine tetra-acetic acid</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-thiogalactoside</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase pairs</td>
</tr>
<tr>
<td>kcat</td>
<td>catalytic rate constant [s$^{-1}$]</td>
</tr>
<tr>
<td>kDa</td>
<td>kiloDalton</td>
</tr>
<tr>
<td>$K_i$</td>
<td>inhibition constant</td>
</tr>
<tr>
<td>Km</td>
<td>Michaelis constant</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani growth medium</td>
</tr>
<tr>
<td>MAD</td>
<td>Multiple wavelength Anomalous Dispersion</td>
</tr>
<tr>
<td>MIR</td>
<td>Multiple Isomorphous Replacement</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>PhC</td>
<td>Phenyl-β-D-cellobioside</td>
</tr>
<tr>
<td>PhX$_2$</td>
<td>Phenyl-β-D-xylobioside</td>
</tr>
<tr>
<td>PNPC</td>
<td>para-nitrophenyl cellobiose</td>
</tr>
<tr>
<td>rmsd</td>
<td>root mean square deviation</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SeMet</td>
<td>Selenium-substituted methionine</td>
</tr>
<tr>
<td>TB</td>
<td>Terrific broth</td>
</tr>
<tr>
<td>TYP</td>
<td>Tryptone, Yeast extract, Phosphate growth medium</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra-violet</td>
</tr>
</tbody>
</table>
Formulas for Reliability Indices

Abbreviations:

R \hspace{1em} \text{residual – usually expressed as percentage, or fraction of 1}

I_{hkl} \hspace{1em} \text{Intensity of a reflection at (h,k,l)}

F_{obs} \hspace{1em} \text{structure factor amplitude of an observed reflection}

F_{calc} \hspace{1em} \text{theoretical or ideal structure factor amplitude of a reflection based on model}

\Delta F^\pm \hspace{1em} \text{structure factor amplitude difference between Friedel mates (Bijvoet difference)}

• R-factor for comparing symmetry related reflection intensities within dataset

\[ R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_{hkl} - \overline{I}_{hkl}|}{\sum_{hkl} \sum_i \overline{I}_{hkl}} \]

• R-factor describing Bijvoet differences used in isomorphous replacement methods, describing the contribution of an anomalous scatterer to the diffraction

\[ R_{\text{anom}} = \sqrt{\frac{\sum_{hkl} (|\Delta F_{\text{obs}}^\pm| - |\Delta F_{\text{calc}}^\pm|)^2}{\sum_{hkl} (|\Delta F_{\text{obs}}^\pm|)^2}} \]

• Crystallographic R-factor as an indicator for the correctness of a structural model

\[ R_{\text{cryst}} = \frac{\sum_{hkl} |F_{\text{obs}} - k|F_{\text{calc}}|}{\sum_{hkl} |F_{\text{obs}}|} \]
Cross-validation R-factor, based on a subset of T reflections not used in refinement

\[ R_{\text{free}} = \frac{\sum_{\mathcal{R} \notin \mathcal{T}} |F_{\text{obs}}| - k|F_{\text{calc}}|}{\sum_{\mathcal{R} \notin \mathcal{T}} |F_{\text{obs}}|} \]
Introduction
and
Thesis Overview
The enzymatic degradation of plant cell wall polymers arises from a complex and productive collaboration between a plethora of enzymes and ancillary modules expressed by polysaccharolytic microorganisms. The most dramatic examples are found in many cellulolytic clostridia such as *Clostridium thermocellum* and *Clostridium cellulovorans*, which express an interdigitated array of proteins on their cell surface, termed the cellulosome. The cellulosome consists of a multi domain scaffolding protein containing a cellulose binding domain, a cell-anchoring domain and many smaller domains called cohesins. Cohesins serve as docking sites for an impressive array of cellulases and, in some cases, xylanases and chitinases, each specializing in a slightly different mode of attack on the substrate to synergistically and efficiently degrade cellulose and hemicellulose (Beguin & Aubert, 1994; Din *et al*., 1994; Fernandes *et al*., 1999; Garcia-Campayo & Beguin, 1997; Shoham *et al*., 1999).

![Cellulosome Diagram](image)

Adapted from: (Shoham *et al*., 1999)
Cellulose is a rigid and chemically inert substrate, made up of microfibrils of β-1,4 linked glucose polymers, which assemble into a highly ordered, compact crystalline solid with a staircase organization (Blackwell, 1982; Sarko, 1986). Plant cell walls mainly consist of crystalline cellulose surrounded by hemicellulose, a matrix of such polymers as xylan and galactan. Instead of cellulosome scaffolds, many bacteria and fungi excrete an equally complex variety of soluble versions of cellulolytic proteins. Especially in bacteria, these often consist of modular proteins, comprising a catalytic domain and one or more ancillary domains. The most common type of ancillary domain is the carbohydrate-binding domain (hereafter called CBM, for carbohydrate-binding module), some of which were previously called cellulose-binding domains (Tomme et al., 1995). In many cases the presence of such domains enables or enhances the hydrolytic activity of an enzyme (Gal et al., 1997; Hall et al., 1995; Irwin et al., 1998). It is thought to do so by directing the enzyme onto the substrate and thereby increasing the local concentration of the enzyme (Bolam et al., 1998). There is some evidence for non-hydrolytic disruption of crystalline cellulose by binding domains, which would make the polymers more accessible to the enzyme (Din et al., 1991).

Based on sequence similarities CBMs have been classified into 21 distinct families, which reflect their ligand specificity as well as their overall mode of binding (Coutinho & Henrissat, 1999b). Structural studies of CBMs have allowed some insights into the mechanisms of substrate binding, although a crystal structure of a CBM in complex with cellulose or xylan has remained elusive, for two main reasons. Firstly, many cellulose binding CBMs have been solved by NMR methods, which does not easily allow for resolving multiple carbohydrate chemical shifts, although some information can still be derived from carefully designed experiments (Weimar et al., 2000). Secondly, the insoluble nature of the ligand cellulose or xylan often precludes structural studies by NMR or X-ray crystallography. The binding sites of the CBMs for which structures have been determined, as predicted by inference from the structure, NMR titrations or site-
directed mutagenesis, are of two general types: flat surfaces comprising predominantly aromatic residues that are thought to interact with crystalline cellulosic surfaces, and extended shallow grooves that presumably bind individual polymer chains. Other types of CBMs include those that bind soluble fractions of polymers, perhaps to capture partially hydrolyzed and released fragments of cellulose, that can be degraded into even smaller sugar units by catalytic domains.

As mentioned earlier, CBMs as expressed by bacteria and fungi are often part of a modular gene product comprising at least one catalytic domain with a specificity to hydrolyze glycosidic bonds. Modules on a single polypeptide chain do not necessarily have complementary ligand specificities. Often, xylanases are expressed in conjunction with cellulose-binding modules, perhaps reflecting the nature of the natural substrate: cellulose is often found in the presence of xylan in hemicellulose, and the cellulose binder would localize the enzyme near its substrate. In turn, some xylanases also have some degree of cellulase activity.

Catalytic domains are termed glycosyl hydrolases, or glycosidases, and have been classified into over 81 families to date, based on sequence similarities as revealed by hydrophobic cluster analysis (HCA) (Davies & Henrissat, 1995; Henrissat, 1991; Henrissat, 1998). Unlike IUB nomenclature, which is based solely on substrate specificity, HCA finds its strength in the direct relationship between amino acid sequence homologies and folding similarities. It therefore reflects not only structural features of these enzymes, but also evolutionary relationships and mechanistic approaches. Combined, enzymes within a family share conserved catalytic sites and hence substrate specificity. Inter-family relationships between enzymes are sometimes revealed based on structural information or more sophisticated sequence comparisons, and these enzymes can be grouped into clans (Coutinho & Henrissat, 1999a; Coutinho & Henrissat, 1999b).

Glycosidases fall into two general mechanistic classes, depending on whether the glycosidic bond is hydrolyzed with net inversion or net retention of anomeric
configuration. Both classes have active sites containing a pair of carboxylic acids, usually glutamates. In the case of the inverting glycosidases the mechanism involves a direct displacement at the anomeric center by a base-activated water molecule. Reaction proceeds via an oxocarbenium ion-like transition state, with the two carboxylic acids suitably positioned approximately 10 Å apart, to function as general acid and general base catalysts. Retaining glycosidases use a double-displacement mechanism in which a covalent glycosyl-enzyme intermediate is formed and hydrolyzed via oxocarbenium ion-like transition states. The two carboxylic acid side chains are closer together in these enzymes such that one functions as a general acid catalyst, while the other functions as a nucleophile, directly attacking at the anomeric center to form a covalent glycosyl-enzyme intermediate with expulsion of the aglycone leaving group. In a second step, water attacks at the anomeric center with general base catalytic assistance from the same carboxyl group that originally donated a proton (i.e. the ‘acid/base catalyst’), thereby releasing the product sugar with net retention of anomeric configuration (see below).

(Davies & Henrissat, 1995; Davies et al., 1998b; Ly & Withers, 1999; Sinnott, 1990; Zechel & Withers, 2000).

![Proposed β-retaining glycosidase mechanism](image-url)
Cellulases and xylanases can be found in at least 13 glycosyl hydrolase families. The difficulty to disrupt and hydrolyze a substrate such as cellulose is appropriately reflected in the variety of modes of attack employed by these cellulases. Exo-acting β-glucanases attack a chain from the polymer termini, bearing specificity for the reducing end or the non-reducing end of the chain, whereas endo-acting β-glucanases hydrolyze the glycosidic bond internally. β-Glucosidases specifically hydrolyze released disaccharides to yield single glucose units. A sub-category of cellulases are those that remain fixed onto the cellulose chain and sequentially hydrolyze disaccharides: the so-called processive cellulases.

Best representing the cellulases are perhaps those that can be grouped into clan GH-A, based on sequence similarity. Enzymes within this clan have a common (α/β)₆ barrel fold, and are often referred to as 4/7 enzymes, because their respective active site glutamatess are located on strands 4 and 7 (Jenkins et al., 1995; Sakon et al., 1996). A wealth of information has been accumulating on these enzymes in recent years, accelerated by the fact that many structures have been solved by X-ray crystallography which has led to very detailed insights into the workings of these enzymes, and has prompted fascinating mechanistic studies. A brief synopsis of the most relevant findings in this area is outlined below.

Much of the early work on β-retaining glycosidases focused on the identification of the two active site carboxylic acids serving as the nucleophile and the acid/base catalyst, whose existence are implied by the mechanism proposed by Koshland (Koshland, 1953), and in analogy to the proposed lysozyme mechanism of action (Blake et al., 1967; Phillips, 1967). Initially, epoxide-based crosslinking agents were used to locate nucleophilic carboxylates. This method often provided ambiguous results due to the promiscuous chemical nature of these reactive compounds (e.g. Havukainen et al., 1996). The use of the mechanism-based 2-deoxy-2-fluoro sugar analogs specifically trapped the covalent intermediate of the reaction, and by sequencing the purified
glycopeptides isolated from proteolytic digests the nucleophile could be identified (Tull et al., 1991; Withers & Aebersold, 1995). Site-directed mutagenesis coupled to detailed kinetic analysis led to the identification of acid/base catalysts. Upon the advent of powerful sequence alignments using HCA, these approaches became mostly redundant since the active site residues could then be accurately assigned to enzymes within families or clans. Further delineation of the glycosidase mechanism was permitted through identification of rate-limiting steps, as well as investigation of the transition-state structure for each step using detailed kinetic analyses. Such mechanistic insights, derived from pre-steady- and steady-state kinetics, Brønsted relationships, kinetic isotope effect measurements, inactivation experiments, and pH studies, provided further supporting evidence for the double-displacement mechanism of retaining β-glycosidases (Bray et al., 1997; MacLeod et al., 1994; MacLeod et al., 1996; Tull & Withers, 1994).

The nature of the β-retaining reaction intermediate was first observed crystallographically in the family 10 β-1,4-glycanase Cex from the soil bacterium Cellulomonas fimii with 2-deoxy-2-fluoro cellobiose covalently bound to the nucleophile (White et al., 1996). The structure showed the saccharide to be in the relaxed 'C, "chair" conformation, covalently linked to the nucleophile through an α-glycosidic bond from the anomeric carbon C1, in agreement with the proposed reaction mechanism. In order to arrive at this intermediate the substrate first has to go through a high-energy transition state, proposed to be of oxo-carbenium ion-like character and, subsequently, in a 'H3 "half chair" conformation (for nomenclature, see Appendix A). Initial attempts at trapping the pre-transition state structure, the Michaelis complex, in order to attain its crystal structure, involved the use of non-hydrolyzable thiol-linked sugars in the family 7 Fusarium oxysporum endoglucanase I (Sulzenbacher et al., 1996). The substrate was found to span the point of enzymatic cleavage (-1/+1) and adopted a distorted 'S3 skew boat conformation. This arrangement results in the upward, pseudo-axial orientation of the C-O bond being cleaved as predicted by stereoelectronic requirements on the
transition state. More recently, the family 5 cellulase Cel5A from *Bacillus agaradhaerens* was used in the elegant crystallographic studies of several “snapshots” along the enzymatic pathway (Davies *et al.*, 1998a), including the product complex, the \( 1S_3 \) skew boat Michaelis complex prior to leaving group departure (figure), and the relaxed \( 4C_1 \) “chair” covalent intermediate.

![Michaelis complex in \( 1S_3 \) skew boat conformation](image)

Detailed understanding of the glycosidase mechanism is crucial in the design and improvement of glycosidase inhibitors. Especially subtle differences in areas such as substrate specificity and transition state structure could prove essential in optimizing inhibitive powers of lead compounds targeting a specific glycosidase function. Glycosidase inhibitors have already been effective as anti-viral agents in the treatment of influenza and HIV infection (Ratner *et al.*, 1991; Taylor, 1996), the treatment of diabetes (Martin & Montgomery, 1996) and as antimetastatic agents (Pili *et al.*, 1995), and they are expected to represent a much broader therapeutic realm in the near future if the ubiquity of glycosilation in nature is an indication. In turn, analysis of crystal structures of glycosidases in complex with such inhibitors can provide great insight into the interpretation of the mechanistic behaviour of these enzymes, and as such this iterative process becomes mutually productive.
The study of glycosidase inhibitors has traditionally focused on two main types of compounds: the basic, nitrogen containing piperidines, such as the naturally occurring deoxynojirimycin and castanospermine (shown below), and the compounds derived from glucono-1,5-lactone, originally observed to be a strong β-glycosidase inhibitor by Ezaki (Ezaki, 1940).

![Chemical structures](image)

deoxynojirimycin  castanospermine  transition state  gluconolactone  tautomer

The obvious comparison to the proposed transition state structure of glycosidases, the positively charged oxocarbenium ion-like \( \text{H}^+ \) half chair, argues for the importance of tight binding enzyme inhibitors to mimic transition state structure in shape and charge, i.e. the flattened chair conformation and the positive charge build-up on the anomeric carbon and cyclic oxygen. However, the relative importance of these features in glycosidase inhibition can seem ambiguous, since the inhibitory profile of these compounds varies widely among many of the glycosidases tested, especially among β-retaining glycosidases. Based on the observation that protonation of the inhibitor at its exocyclic heteroatom by the enzyme can be important for its effectiveness, β-retaining glycosidases are further classified as "syn" or "anti" protonators (Heightman & Vasella, 1999). Such classification depends upon the trajectory of the proton transfer event relative to the C1-O5 bond, which is dictated by the relative position of the acid/base catalyst to the incoming substrate. To illustrate, many members of the lactone-derived inhibitor class, for example the imidazoles, do not contain a syn lone pair, and are generally only good inhibitors against anti-protonators. A schematic representation of syn vs. anti protonation is shown below:
Recently, a surprising new mechanistic approach as been unveiled in two family 11 xylanases based on their covalent 2F-xylobiose enzyme intermediate crystal structures (Sabini et al., 1999; Sidhu et al., 1999). These structures show the catalytic intermediate to be a distorted $2_{5}B$ boat, which suggests a mechanistic itinerary involving $4C_{1} - 2H_{3} - \tilde{3}S_{0} - 2_{5}B - 4C_{1}$ conformations. The transition state structure, though not a “classic” $4H_{3}$ half chair still retains its stereoelectronic requirements by forming a planar organization around C2-C1-O5-C5 and positions the anomeric carbon optimally for nucleophilic attack. Not surprisingly, many of the inhibitors tested so far have been ineffective against family 11 xylanases, probably due to their inherently different approach to transition state stabilization. A new generation of specifically designed inhibitors would be of great value to study this mechanism more extensively.

In light of the above described developments in the field of carbohydrate binding proteins and glycosidase mechanisms in recent years, I will present the work that I completed on those topics as a graduate student. Below is a summary of the prepared thesis.
Thesis Overview

The first chapter in this thesis describes the 1.9 Å crystal structure, determined by the MAD method (see appendix B), of the family 9 carbohydrate binding module from *Thermotoga maritima* xylanase 10A (CBM9-2) in native and ligand bound forms. These structures reveal several unique and surprising features in carbohydrate and cellulose binding. Two tryptophan residues form a sandwich around the ligand, which is held tightly into place by multiple polar interactions surrounding the reducing end of the sugar. Glucose binds in a different fashion than cellobiose, but it maintains the overall mode of binding through hydrogen bonding and dual tryptophan stacking. This is a first opportunity to analyze protein-ligand interactions of a cellulose-binder at the atomic level. The structures also led to the hypothesis that CBM9-2 binds exclusively to the reducing ends of cellulose, and not to the crystalline surface of this substrate as anticipated from the binding experiments.

The remainder of the work described in this thesis revolves around the catalytic module of the family 10 xylanase Cex from *Cellulomonas fimi*. Cex has been well characterized structurally and kinetically, both in this laboratory and in that of Drs. S. G. Withers and R. A. J. Warren at UBC. In light of this, Cex was used as a crystallographic template to probe many aspects of the retaining β-glycosidase reaction mechanism.

Cex, an endoxylanase, preferentially hydrolyzes xylan but also displays significant hydrolytic activity against carboxymethylcellulose. More specifically, Cex hydrolyzes xylosides approximately 40-fold more efficiently than it does glucosides (Gilkes *et al.*, 1991). Chapter 2 provides insight into the structural nature of its preference for xylan, by analyzing the crystal structure of the Cex catalytic domain trapped as its covalent 2-deoxy-2-fluoroxylobiosyl-enzyme intermediate. Together with the crystal structure of unliganded Cex (White *et al.*, 1994) and the previously determined crystal structure of the covalent 2-deoxy-2-fluorocellobiosyl-Cex intermediate (White *et al.*, 1996), this structure provides a convincing rationale for the
observed substrate specificity of this enzyme. Two active site residues, Gln87 and Trp281, are found to sterically hinder the binding of glucosides and must rearrange to accommodate these substrates. Such rearrangements are not necessary for the binding of xylobiosides, and therefore should result in more favourable binding of xylan versus cellulose.

Cex hydrolyzes its substrates via a double-displacement mechanism involving formation and hydrolysis of a covalent glycosyl-enzyme intermediate flanked by oxocarbenium ion-like transition states, as proposed by Koshland (Koshland, 1953). Interactions at the 2-position of the proximal saccharide are extremely important to transition state stabilization in a number of glycosidases, contributing at least 40 kJ.mol\(^{-1}\) in the case of Cex (Namchuk & Withers, 1995). Unfortunately, since the 2-position is substituted by fluorine in the trapped complexes it has been impossible to identify unequivocally the interactions involved at this position. Chapter 3 directly addresses the nature of the interactions between the enzyme and the 2-hydroxyl of the glycone in the bond being hydrolyzed, by presenting the structure of a covalent enzyme-substrate intermediate in the His205N/E127A double mutant of Cex with an unsubstituted cellobiosyl moiety. The structure suggests that the major source of this stabilization may be a strong hydrogen bond between the 2-hydroxyl and the side chain carbonyl oxygen of the nucleophilic carboxylate. This interaction could provide enormous transition state stabilization and may prove to be widespread in glycosidase catalysis.

Chapter 4 probes the roles of two conserved secondary catalytic residues, Asn126 and Asn169 in the active site of Cex, by kinetic and crystallographic analysis of their respective alanine mutants. The kinetic profile of the Asn126Ala mutant indicates that the rate-limiting step of the reaction shifts from deglycosylation to glycosylation, whereas the Asn169Ala mutation results in minimal effect on catalysis. The implications of these results for the role of Asn126 and Asn169 in catalysis are discussed in this chapter.
Considerable efforts have been expended in recent years in the design, synthesis and testing of glycosidase inhibitors, with the hope of not only learning more about the active site structures and mechanisms of these enzymes, but also of possibly generating new therapeutic agents. Applications of glycosidase inhibition can be found in areas such as treatments for influenza and non-insulin dependent diabetes (Martin & Montgomery, 1996; Taylor, 1996). A great need exists for structural information to validate theories of inhibitor design and suggest improvements. Chapter 5 describes the crystal structures of four xylobiose derived nitrogen-containing inhibitors in complex with Cex. These inhibitors address key features of the reaction mechanism, such as syn vs. anti protonation trajectories of the departing sugar during glycosylation, and boat vs. chair transition state and intermediate itineraries of the reaction mechanism. The differences in reaction mechanisms between Cex and Bcx, a family 11 endoxylanase, are discussed in light of these results.

Based on the results with the inhibitors discussed in chapter 5, a xylobiose-derived lactam, related to xylobiosylisofagomine, was prepared and assayed against Cex. In Chapter 6 the crystal structure of this inhibitor-enzyme complex is described. The resulting structure is completely consistent with the inhibitor binding as the iminol tautomer. This observation is confirmed through kinetic analysis of this inhibitor against the Asn126Ala mutant of Cex, which suggests the iminol tautomer interacts with the Asn126 residue in a fashion similar to the 2-hydroxy containing inhibitors.
References


Chapter 1

Crystal Structures of the Family 9 Carbohydrate-Binding Module from
Thermotoga maritima Xylanase 10A in Native and Ligand Bound Forms

This chapter has been reformatted from the original manuscript as prepared for
submission to the journal Biochemistry (ACS publications): Valerie Notenboom,
Alisdair B. Boraston, R. Antony J. Warren, Douglas G. Kilburn, and David R. Rose
(2000) Crystal structures of the family 9 carbohydrate binding module from Thermotoga
maritima xylanase 10A in native and ligand bound forms.
Abstract

The C-terminal module of the thermostable *Thermotoga maritima* xylanase 10A (CBM9-2) is a carbohydrate-binding module belonging to family 9. CBM9-2 binds to amorphous and crystalline cellulose, a range of soluble di- and monosaccharides as well as to cello- and xylo oligomers of different degrees of polymerization (Boraston *et al.*, in preparation). The crystal structure of CBM9-2 was determined by the multiwavelength anomalous dispersion method to 1.9 Å resolution. CBM9-2 assumes a β-sandwich fold and contains three metal binding sites. The bound metal atoms, which are most likely calcium cations, are in an octahedral coordination. The crystal structures of CBM9-2 in complex with glucose and cellobiose were also determined to identify the sugar-binding site and provide insight into the structural basis for sugar binding by CBM9-2. The sugar-binding site is a solvent exposed slot sufficient in depth, width and length to accommodate a disaccharide. Two tryptophan residues are stacked together on the surface of the protein forming the sugar-binding site. From the complex structures with glucose and cellobiose it was inferred that CBM9-2 binds exclusively to the reducing end of mono-, di-, and oligosaccharides with an intricate hydrogen bonding network involving mainly charged residues, as well as stacking interactions by Trp175 and Trp71. The specificity of CBM9-2 for the reducing ends of polysaccharides was confirmed when it was unable to bind reduced forms of cellulose and β-glucan. The binding interactions are limited to disaccharides as was expected from calorimetric data. Comparison of the glucose and cellobiose complexes revealed surprising differences in binding of these two substrates by CBM9-2. Cellobiose was found to bind in a distinct orientation from glucose, while still maintaining optimal stacking and electrostatic interactions with the reducing end sugar.
Introduction

Microbial degradation of cellulose and hemicellulose found in plant biomass is achieved by the activity of diverse polysaccharolytic enzyme systems (Beguin & Aubert, 1994; Tomme et al., 1995). The majority of polysaccharidases produced by cellulolytic organisms have modular structures comprising a catalytic module and one or more ancillary modules (Gilkes et al., 1991). The most common type of ancillary module is the carbohydrate-binding module (CBM), some of which were previously called cellulose-binding domains (Tomme et al., 1995). In many cases the presence of the CBM enhances the hydrolytic activity of an enzyme. It is thought to do so by increasing the concentration of the enzyme on the substrate (Bolam et al., 1998), though there is some evidence for non-hydrolytic disruption of cellulose by a CBM (Din et al., 1991).

Structural studies of CBMs have allowed insights into the mechanisms of substrate binding. The binding sites of the CBMs for which structures have been determined are of two general types: flat surfaces comprising predominantly aromatic residues and extended shallow grooves.

The family 1 CBMs from Trichoderma reesei (Kraulis et al., 1989; Mattinen et al., 1998), the family 2a CBM from xylanase 10A of Cellulomonas fimi (Xu et al., 1995), the family 3 CBM from CipC of Clostridium thermocellum (Tormo et al., 1996), and the family 5 CBM from cellulase 5A of Erwinia crysanthemi (Brun et al., 1997) are all β-sheet polypeptides. The binding faces of CBMs from these families, identified by site directed mutagenesis (Din et al., 1994; Goldstein & Doi, 1994; Linder et al., 1995; Nagy et al., 1998; Simpson & Barras, 1999), are flat faces and composed mainly of tryptophan and tryosine residues. The ligand of these CBMs, crystalline cellulose, is a regular array of cellulose chains with a staircase organization (Blackwell, 1982; Sarko, 1986). The binding faces of these examples of family 1, 2a, 3, and 5 CBMs seem well suited to bind the flat surfaces of cellulose crystals. In the case of CBM2a, thermodynamic data have suggested that dehydration of the aromatic binding face provides an entropic driving
force for binding (Creagh *et al.*, 1996). Beyond this, other protein-cellulose interactions are speculative at best.

The tandem N-terminal family 4 CBMs (CBM4-1 and CBM4-2) from cellulase 9B of *C. fimi* (Johnson *et al.*, 1996) and the internal family 2b CBM (CBM2b-1) from xylanase 11b of *C. fimi* (Simpson *et al.*, 1999) bind soluble β-1,4-glucan and β-1,4-xylan chains, respectively. The binding sites of these CBMs, identified by NMR spectroscopy, are shallow grooves sufficient in width and length to accommodate individual glycan chains up to 5 sugar units in length (but no smaller than 3). As with most carbohydrate-binding proteins, the involvement of aromatic residues in ligand binding is implicated; however, in contrast to CBM2a, thermodynamic studies of CBM4-1 and CBM2a-1 implied a predominance of polar and van der Waals interactions with the ligand (Simpson *et al.*, 1999; Tomme *et al.*, 1996). The C-terminal family 9 CBM (CBM9-2) from xylanase 10A of *Thermotoga maritima* is an extremely thermostable module (Wassenberg *et al.*, 1997) with the unique ability to bind crystalline cellulose, amorphous cellulose, and disaccharides tightly (Boraston *et al.*, 2000). The thermostability and unprecedented capacity to bind all cellulose allomorphs, in addition to many types of small soluble sugars, make it an interesting candidate for structural studies. Here the crystal structure of CBM9-2 in native and ligand bound forms to 1.9Å resolution is described. For the first time we can describe the specific intermolecular contacts that mediate the binding of a CBM with cellulose.
Materials and Methods

Materials.

All chemicals and buffer reagents used, were purchased from Sigma (St. Louis, MO), unless otherwise specified. Cellotetraose was from Megazyme (Dublin, Ireland, UK).

Bacterial strains and plasmid.

*Thermotoga maritima* CBM9-2 (Boraston et al., 2000) was produced in *Escherichia coli* strains JM101, or the BL21 derived B834(DE3) for selenomethionine incorporation (Novagen, Madison, WI). The plasmid used was pET28a (Novagen, Milwaukee, WI).

Media and growth conditions.

*E.coli* cultures were grown routinely in TB or TYP medium unless otherwise stated (Sambrook et al., 1989) at 37 °C, or at 30 °C for production of protein. The medium contained 50 mg/L kanamycin of culture for strains carrying pET28 plasmids.

Production and purification of wild type CBM9-2.

60 L of TYP medium, inoculated with *E. coli* BL21(DE3)/ pETCBM9-2, was grown at 30 °C to OD₆₀₀nm of ~0.6 upon which Isopropyl β-D-thiogalactoside (IPTG) was added to 0.3 mM for production of the polypeptide CBM9-2 and incubation continued overnight. The cells were harvested by centrifugation at 15,000 rpm (40,000 x g) and 4 °C for 15 min, and re-suspended in 1 L of 50 mM potassium phosphate buffer, pH 7.0 (KP). Cell lysis was accomplished by three times passing through a French™ press. Unbroken cells and debris were removed by centrifugation as above. A 300 mL sample of the cleared cell extract was loaded at 1 mL/min onto a 200 mL column (packed bed volume) of beaded cellulose (Sigma). The column was washed with 5 column volumes of 50 mM potassium phosphate, pH 7.0, containing 500 mM NaCl, followed by a 5 column wash with 50 mM potassium phosphate, pH 7.0. Washing steps were performed at a flow rate of 3 mL/min. Bound CBM9-2 was desorbed by passing a solution of 1 M glucose (in 50 mM potassium phosphate, pH 7.0) through the column at a flow rate of 3 mL/min. 10 mL fractions of the glucose elution were collected and screened by SDS-PAGE. Fractions containing pure
polypeptide of the expected size (21 kDa) were pooled. The polypeptide was extensively buffer exchanged and concentrated into 50 mM potassium phosphate, pH 7.0, by ultrafiltration using a 1K Ultrafiltron membrane (Filtron Technology Corp., Northborough, MA). The concentrated solution was stored at 4 °C. Final yields of pure CBM9-2 ranged from 80-100 mg per 2 L culture.

*Preparation of reduced cellulose and barley β-glucan.*

1 g of Avicel™ (PH 101, Little Island, County Cork, Ireland) was suspended in 15 mL of 100 mM NH₄OH (pH 10.8). 150 mg of NaBH₄ was added to this, covered, and incubated for 6 hours with stirring at room temperature. The cellulose was pelleted by centrifugation at 11000 x g and washed with 40 mL of distilled water. This process was repeated four times. The final cellulose pellet was resuspended in 10 mL of distilled water. A control sample to which NaBH₄ was not added was prepared in parallel. The final concentration of these samples was determined by the dry weights of duplicate 1 mL samples. 100 mg of barely β-glucan (Sigma, St. Louis, MO) was dissolved in 5 mL of 100 mM NH₄OH (pH 10.8). 50 mg of NaBH₄ was added to this, covered, and incubated for 6 hours with stirring at room temperature. The sample was dialyzed into distilled water for 4 days using a 1000 Da cutoff membrane (Spectra/Por®, Spectrum Laboratories, Rancho Dominguez, Ca). A control sample to which NaBH₄ was not added was prepared in parallel. The final concentration of these samples was determined by the phenol sulfuric acid assay for total sugar (Chaplin, 1986) using glucose standards.

*Production and purification of Seleno-Methionine containing CBM9-2.*

The pET28a plasmid containing the CBM9-2 construct was transformed into competent *E. coli* B834(DE3) methionine auxotrophs by heat shock. A 15 mL culture was grown on enriched minimal medium (2× M9 salts supplemented with all regular amino acids to 40 mg/L, vitamins riboflavin, niacinamide, pyridoxine monohydrochloride and thiamine to 1 mg/L, 0.4% glucose, 25 mg/L FeSO₄ and 2 mM MgSO₄) overnight at 37 °C with 50 mg/L kanamycin. 5 mL of this culture was added to 1 L medium as above, with
methionine substituted by seleno-L-methionine at 100 mg/L, and allowed to grow to an OD₆₀₀ of 0.7. If the doubling time was found to be longer than 2 hrs, then an additional 1 mL of a 65:35 Selenomet:Met culture in stationary phase was added. IPTG was added to 0.3 mM to induce protein production and cells harvested after 5 hours by centrifugation. CBM9-2 was purified essentially as described above and stored at 30 mg/mL at 4 °C. The final selenomethionine protein yield from this 1 L culture was approximately 80 mg.

**Crystallization of Native and SeMet CBM-9.2.**

Tetragonal crystals were grown using the vapour diffusion technique from hanging and sitting drops in ~15% PEG8000 and 0.2 mM sodium acetate, pH 4.5. Long needles extending beyond the crystallization drop generally appeared immediately upon mixing the protein and reservoir solutions if either the protein concentration or the PEG concentration was too high. Temperature or pH did not prove to be factors in nucleation rates at the ranges screened (4-37 °C, and pH 3-9). Optimized precipitant and protein concentration combinations did produce 3-dimensional crystals that appeared after a week or more at room temperatures, with dimensions reaching 0.2 x 0.2 x 0.5-1.5 mm. The crystals are of space group P4₁2₁2 with cell dimensions a=b=56.8 Å and c=123.1 Å, allowing for 1 molecule per asymmetric unit with a solvent content of ~43 %. The SeMet substituted protein crystallized in identical conditions as wild type protein. Crystals of CBM-9.2 complexed with glucose and cellobiose were produced by supplementing the reservoir solution with 50 - 100 mM substrate and adjusting precipitant concentrations.

**Data collection and Processing.**

Prior to data collection crystals were bathed in artificial mother liquor with 30% glycerol added as cryoprotectant and quickly frozen in a stream of nitrogen gas at 100 K in a rayon fiber loop. A native data set was collected in-house as well as data from a SeMet crystal, on a Mar345 image plate using Osmic mirror focussed Cu-Kα X-rays, generated from a rotating anode operating at 100mA x 50kV. 90 oscillations of 1° were collected
in 5 min exposures. All data were indexed, integrated and scaled with the Denzo/Scalepack package (Otwinowski & Minor, 1997). Two peaks could be visualized on a difference Patterson map calculated in the program PHASES at this point (Furey & Swaminathan, 1990), which suggested significant incorporation of selenium atoms in the protein crystals (not shown). Subsequently a three-wavelength MAD data set was collected at the Advanced Photon Source (APS) at Argonne National Laboratories at the BioCARS 14-BM-D, on a 1K CCD area detector at 105 mm crystal to detector distance. The absorption edge for selenium was detected by scanning through the theoretical X-ray absorption edge. From this three wavelengths were chosen, one at the inflection point or edge of the absorption profile (f-min, $\lambda = 0.9799$ Å), one at its peak (f'-max, $\lambda = 0.9797$ Å) and one reference set at a remote wavelength ($\lambda = 0.9537$ Å). Data sets were collected to 1.8 Å resolution, ensuring Bijvoet pairs were measured close together in time to avoid unnecessary error due to possible crystal deterioration. Exposure times ranged from 5-10 s for 1" oscillations, depending on the flux of the beam at the time of the experiment. The program package Denzo/Scalepack was used to refine experimental parameters and process data (Otwinowski & Minor, 1997). In addition, glucose and cellobiose complexes were crystallized and data collected in-house as described above. Experimental statistics are summarized in Tables 1.1a & 1.1b.
### Table 1.1a. Data collection statistics from synchrotron radiation

<table>
<thead>
<tr>
<th></th>
<th>Peak</th>
<th>Edge</th>
<th>Remote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength</td>
<td>0.9797 Å</td>
<td>0.9799 Å</td>
<td>0.9537 Å</td>
</tr>
<tr>
<td>Resolution</td>
<td>1.9 Å</td>
<td>1.8 Å</td>
<td>1.8 Å</td>
</tr>
<tr>
<td>Observations</td>
<td>318505</td>
<td>305530</td>
<td>318615</td>
</tr>
<tr>
<td>Unique refl.</td>
<td>16435</td>
<td>19085</td>
<td>19135</td>
</tr>
<tr>
<td>Completeness</td>
<td>99.8 %</td>
<td>99.0 %</td>
<td>99.5 %</td>
</tr>
<tr>
<td>I/σ</td>
<td>10.2</td>
<td>9.3</td>
<td>10.4</td>
</tr>
<tr>
<td>R$_{merge}$</td>
<td>6.9 %</td>
<td>7.2 %</td>
<td>5.8 %</td>
</tr>
</tbody>
</table>

### Table 1.1b. Data collection statistics from rotating anode

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>SeMet</th>
<th>Glucose</th>
<th>(Glu)$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength</td>
<td>1.54 Å</td>
<td>1.54 Å</td>
<td>1.54 Å</td>
<td>1.54 Å</td>
</tr>
<tr>
<td>Resolution</td>
<td>1.9 Å</td>
<td>1.9 Å</td>
<td>1.8 Å</td>
<td>1.9 Å</td>
</tr>
<tr>
<td>Observations</td>
<td>90745</td>
<td>119264</td>
<td>158924</td>
<td>272755</td>
</tr>
<tr>
<td>Unique refl.</td>
<td>15478</td>
<td>14908</td>
<td>19725</td>
<td>17513</td>
</tr>
<tr>
<td>Completeness</td>
<td>97.7 %</td>
<td>95.2 %</td>
<td>96.2 %</td>
<td>99.4 %</td>
</tr>
<tr>
<td>I/σ</td>
<td>8.8</td>
<td>16.0</td>
<td>10.7</td>
<td>9.5</td>
</tr>
<tr>
<td>R$_{merge}$</td>
<td>6.5 %</td>
<td>4.3 %</td>
<td>6.2 %</td>
<td>7.7 %</td>
</tr>
</tbody>
</table>

$$R_{merge} = \frac{\sum \sum |I_{hkl} - \overline{I}_{hkl}|}{\sum \sum |I_{hkl}|}$$
Determination of selenium atom positions and phasing

Initially the program Solve was used in an attempt to locate directly the 3 expected selenium atoms from the MAD dataset alone (Terwilliger & Berendzen, 1999). Both possible enantiomorphic space groups were tested but none of the sites were found. Using the home data sets collected from native and SeMet crystals, a combined MAD/MIR routine was attempted by treating the home data set pair as an additional isomorphous heavy atom system (see also Appendix B). The R-values for the anomalous (diagonals), dispersive and isomorphous differences are listed in Table 1.2.

<table>
<thead>
<tr>
<th></th>
<th>Edge</th>
<th>Peak</th>
<th>Remote</th>
<th>Native (1.54 Å)</th>
<th>SeMet (1.54 Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edge</td>
<td>0.067</td>
<td>0.021</td>
<td>0.027</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Bijvoet)</td>
<td>(dispersive)</td>
<td>(dispersive)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td></td>
<td>0.055</td>
<td>0.037</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Bijvoet)</td>
<td>(dispersive)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remote</td>
<td></td>
<td></td>
<td>0.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Bijvoet)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Native</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
<td>0.129</td>
</tr>
<tr>
<td>(1.54 Å)</td>
<td></td>
<td></td>
<td></td>
<td>(isomorphous)</td>
<td></td>
</tr>
</tbody>
</table>

The dispersive differences within the MAD data set are within the error range of data measurement as expressed by the Bijvoet differences of centric reflections (not shown) and should not be expected to yield a clear solution. However, in combination with the additional ‘heavy atom’ data set it was hoped that the isomorphous differences between S and Se would be significant enough to find initial peaks from which the MAD data could be interpreted. Using this approach Solve found 2 selenium sites with occupancies of 0.63 and 0.38, the 3rd expected site located at the N-terminal remained elusive and was presumed disordered. The phases calculated from these two positions and intensities to
1.8 Å from the home selenomethionine dataset resulted in the modest average figure of merit of 0.42. The corresponding electron density was still uninterpretable at the individual residue level, but distinct protein and solvent regions could be identified suggesting the phase estimates were meaningful. Density modification using DM in the program suite CCP4 (Collaborative Computational Project, 1994; Cowtan & Zhang, 1999), implementing histogram matching and solvent flattening, was used to improve the map quality. The solvent content value proved a significant factor in terms of map improvement by DM, changes of as little as 1% resulted in drastic differences in map appearance. The resulting map based on 43% solvent content was of sufficient quality to be completely traced, except for some loop regions. Interestingly, no density beyond C-β was seen in this experimental map for each of the SeMet residues, both of which are located on the surface of the protein and assume multiple conformations.

Model building and Refinement.

The initial solvent flattened map was manually traced using the graphical interface in O (Jones et al., 1990). Most regions in the electron density were reasonably well defined, with distinct carbonyl oxygens and recognizable side chains in most cases, except for the two methionine residues and some disordered solvent exposed residues. Three loop regions seemed poorly defined upon first inspection and were omitted from the polypeptide trace. Closer inspection, however, revealed that these loops included unanticipated octahedrally coordinated metal-binding sites and were, in fact, well ordered. 188 residues were fitted, as well as three metal ions, which were initially treated as water molecules because their identity was unknown. This model underwent rigid body refinement in the program CNS (Brünger et al., 1998), implementing bulk solvent correction and a maximum likelihood refinement target using structure factor amplitudes, against the 1.9Å wild type data collected in-house with a random 10% of reflections set aside for cross-validation purposes (Brünger et al., 1998; Kleywegt & Brünger, 1996). Visual inspection of 2IFol-IFcl (1σ) and IFol-IFcl (+3σ, -3σ) electron density with
calculated model phases allowed for some manual adjustments of poorly fitted regions. The three metal ions were assumed to be Ca\(^{2+}\) cations at this stage, based on the octahedral geometry of the coordination and the size of the difference peaks (~35\(\sigma\)), as well as the precedence of Ca\(^{2+}\) ions present in thermally stable proteins in general (Hwang et al., 1997) and CBMs in particular (Daniel et al., 1996; Ding et al., 1999; Tormo et al., 1996). The model was then subjected to simulated annealing starting from 3000K in steps of 25K, followed by positional refinement and grouped thermal parameter refinement. Several cycles of manual water building was performed at this point. Upon convergence of free-R, individual B-factors were refined. Structures of the cellobiose and glucose complexed proteins were determined by molecular replacement of the unliganded model (without water molecules) against the diffraction data of the complexed crystals. The substrate was located and built into 3\(\sigma\) |Fol-IFcl difference density and these coordinates included in the model. This model then was refined as described above, except simulated annealing was commenced at 1500K.
Results and Discussion

Structure of CBM9-2.

The native protein model consists of 1484 non-hydrogen atoms, 3 calcium atoms and 154 ordered water molecules. Structure refinement statistics are summarized in Table 1.3.

Table 1.3. Structure refinement and model quality statistics

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>Glucose</th>
<th>Cellobose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum resolution used</td>
<td>1.90 Å</td>
<td>1.90 Å</td>
<td>1.90 Å</td>
</tr>
<tr>
<td>Cell (a=b, c in Å, SG P4_2 2)</td>
<td>56.56, 122.57</td>
<td>56.76, 123.10</td>
<td>56.67, 122.97</td>
</tr>
<tr>
<td>R-factor, all data (%)</td>
<td>20.1</td>
<td>19.9</td>
<td>20.7</td>
</tr>
<tr>
<td>R-free, all data (%)</td>
<td>23.2</td>
<td>22.1</td>
<td>23.6</td>
</tr>
<tr>
<td>No. protein atoms</td>
<td>1484</td>
<td>1484</td>
<td>1484</td>
</tr>
<tr>
<td>No. substrate atoms</td>
<td>N/A</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>No. solvent atoms</td>
<td>154</td>
<td>161</td>
<td>146</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond lengths (Å)</td>
<td>0.005</td>
<td>0.005</td>
<td>0.006</td>
</tr>
<tr>
<td>bond angles (°)</td>
<td>1.38</td>
<td>1.39</td>
<td>1.42</td>
</tr>
<tr>
<td>dihedral angles (°)</td>
<td>24.1</td>
<td>23.2</td>
<td>22.7</td>
</tr>
<tr>
<td>Average B-factor (Å²)</td>
<td>14.01</td>
<td>11.23</td>
<td>11.04</td>
</tr>
<tr>
<td>Ave. B-factor substr. (Å²)</td>
<td>N/A</td>
<td>14.27</td>
<td>14.03</td>
</tr>
</tbody>
</table>

\[
R_{cryst} = \frac{\sum_{hkl} ||F_{obs}|| - k ||F_{calc}||}{\sum_{hkl} ||F_{obs}||} \\
R_{free} = \frac{\sum_{hkl \neq T} ||F_{obs}|| - k ||F_{calc}||}{\sum_{hkl \neq T} ||F_{obs}||}
\]
A ribbon diagram of the crystal structure (of the substrate bound form) and a topology representation are shown in figures 1.1a and b. The 3D structure consists of two twisted anti-parallel β-sheets each made up of five β-strands (β2-β3-β4-β9-β8 and β7-β6-β5-β10-β11) to form a β-sandwich fold, flanked by an additional two stranded parallel β-sheet (β1-β12) comprising the N- and C-termini. A short α-helix starts off a loop connecting β1 and β2, which is stabilized by a Ca\(^{2+}\) ion. Two other Ca\(^{2+}\) ions are also found at the periphery of the protein, one stabilizing a loop connecting strands β5 and β6, and one sandwiched between β3 and β4. This last site also forms contacts with β10 and β11, thus connecting the two β-sheets. The coordination of this and the two other Ca\(^{2+}\) sites are depicted in Figure 1.2a-c, all exhibiting close to perfect octahedral geometry involving oxygens originating from a combination of main-chain carbonyls, waters and side-chain carbonyls. Several side-chain carbonyls (Glu130 in a), Asp93 and 94 in b) and Asp74 and Asp60 in c) appear to form interactions with the calcium ion through their deprotonated resonance form, with the negative charge distributed over both oxygens. Substrate binding affinities of CBM9-2 are not dependent on Ca\(^{2+}\) or Mg\(^{2+}\) concentration (Boraston et al., 2000). It has been demonstrated that CBM9-2 binds cellobiose effectively at temperatures to at least 70 °C. Since none of the calcium sites is in direct contact with the binding site, the observed calcium binding sites most likely serve a structural role, stabilizing the three-dimensional organization of the polypeptide thereby contributing to the thermostability of the protein.
Figure 1.1a  Ribbon diagram of CBM9-2 showing the overall fold of the molecule. Three octahedral metal binding sites, presumed calcium, are drawn in magenta. Two Trp residues, shown in yellow, bind a cellobiose moiety in a sandwich-like fashion. The reducing end of the sugar points inward into the protein.
Figure 1.1b  Topology map depicting the secondary structure organization of CBM9-2, with blue arrows representing β-strands and the yellow box an α-helix. Three calcium ions are shown in magenta, around the major contributing binding elements. Two antiparallel β-sheets of five strands each form the bulk of the molecule, with a parallel sheet formed by the N- and C-termini on the outside.
Figure 1.2  Three calcium binding sites in CBM9-2. Glu130 (a), Asp94 and Asp93 (b), Asp74 and Asp60 (c) bind calcium in a bifurcated fashion, or through a delocalized charge on the residue.
CBM9-2 has been shown to bind amorphous and crystalline cellulose, as well as a range of small soluble sugars including mono- and disaccharides and cello- and xylooligosaccharides. Cellobiose bound with the highest affinity of the substrates tested, and cellooligosaccharides with a degree of polymerization greater than two did not bind with higher affinity, suggesting a disaccharide might fully occupy the binding region (Boraston et al., 2000). Two tryptophan residues, Trp175 and Trp71, were found in parallel orientation at the base of the second twisted β-sheet in such that both residues could interact with a carbohydrate placed between these residues. The stacking of aromatic amino acid sidechains against the apolar surfaces of carbohydrates is common in protein-carbohydrate interactions. For this reason, and the prevalence of planar polar amino acids at the base of the cleft formed by these tryptophans, this was predicted to constitute the binding region of CBM9-2.

*Structure of CBM9-2 in complex with cellobiose.*

The native protein model was used in rigid body refinement against data of the cellobiose-CBM9-2 complex to optimize the orientation of the model in the unit cell. Difference electron density from |F₀| - |F₁| and 2|F₀| - |F₁| (contoured at 3σ and 1σ respectively) clearly revealed a single cellobiose molecule sandwiched between Trp71 and Trp175 as anticipated (Figure 1.3). A cellobiose molecule was easily fitted into the density in a single orientation, with the reducing end of the sugar pointed into the protein. An extensive network of hydrogen bonding interactions involving all hydroxyl groups and the cyclic oxygen is formed around the reducing end (proximal) sugar. Of particular interest are interactions formed with the OH₁ and OH₂ groups of the reducing sugar. Binding experiments have shown that substrates with modified functional groups at these positions (such as 1-O-methyl-glucopyranosides, GlcNAC or mannose, Al Boraston, UBC, personal communication) are unable to compete with cellulose for binding to CBM9-2, suggesting interactions at these positions are extremely important. Indeed, the anomeric hydroxyl group (OH₁) forms three hydrogen bonds with Asn172, Arg161 and
Gln151: OH2 also forms two interactions, with Glu77 and an ordered water molecule, which is traced back to the cyclic nitrogen of Trp175. A schematic representation of interactions in the binding site is shown in Figure 1.4a. The tryptophan residues provide planar hydrophobic stacking interactions spanning the whole disaccharide, centering around the glycosidic bond and seemingly involving both sugar rings equally. The non-reducing or distal saccharide receives fewer hydrogen bonds from the protein than the proximal sugar, the OH6 hydroxyl forms a hydrogen bond with Arg98.

**Figure 1.3** 2Fo-Fc electron density contoured at 1σ, using phases from the protein model only. A cellobiose molecule (left) was easily fitted into the density with one clear orientation. Two Trp residues (right, stereo) stack the sugar in a ‘sandwich’.
Figures 1.4 a and b  Schematic representations of cellobiose (top) and glucose (bottom) binding.
Structure of CBM9-2 with glucose.

CBM9-2 is unique among carbohydrate binding modules for binding monosaccharides with relatively high affinity. Calorimetric analysis of binding indicated that glucose may undergo a different mechanism of binding compared to the disaccharides tested, thus it was of interest to investigate the crystal structure of CBM9-2 in complex with glucose. The structure was acquired as described above. Again, difference electron density using the unbound structure as a starting model clearly revealed a glucose moiety bound in the binding pocket of the protein, allowing for straight-forward building of the sugar into the structure (Figure 1.5). Surprisingly, glucose is bound in the pocket in a markedly different orientation than cellobiose. Although the reducing end of the sugar remains oriented inward, none of the specific hydrogen bonding interactions are maintained, but are replaced by a set of new interactions (Figure 1.4b). Roughly, the glucose molecule, compared to the proximal ring of cellobiose, has rotated 180° around the axis defined by C1-C4, and an additional 60° in the plane of the ring. This not only maintains proper stacking interactions with the Trp ‘sandwich’, but allows for a crude overlap of functional groups in space, and therefore almost equal number of hydrogen bonds are formed with the protein as seen with cellobiose (Figure 1.6). Interestingly, OH1 and OH2 in glucose overlap almost perfectly with OH2 and OH1 of cellobiose, respectively, supporting the importance of these two hydroxyl groups as discussed in the previous paragraph, but not in the conserved fashion that was anticipated.
Figure 1.5  2Fo-Fc electron density contoured at 1σ, using phases from the protein model only. A glucose molecule (left) was easily fitted into the density with one clear orientation.

Figure 1.6  Stereo representation of the binding site interactions in CBM9-2 with cellobiose (yellow) and glucose (green). Only the hydrogen bonding interactions with cellobiose are shown, for clarity (see figure 1.4b for glucose interactions).
CBM9-2 binds exclusively to the reducing ends of polysaccharides.
The structures of CBM9-2 complexed with small soluble saccharides showed the significance of the reducing ends of these sugars to binding. This importance of this mechanism in the binding of CBM9-2 to polysaccharides was confirmed by Altering the reducing ends of cellulose and barley β-glucan by reduction of the hemi-acetal linkage at C-1 with sodium borohydride, thus opening the sugar ring. CBM9-2 did not bind to reduced Avicel™ (data not shown). The affinity and capacity of CBM2a (formerly CBDcα), which binds crystalline and amorphous cellulose, for reduced Avicel were unchanged (not shown) suggesting that the gross structure of the cellulose was not significantly perturbed by the reducing agent. CBM9-2 did not bind reduced barley β-glucan; however, CBM17, which can bind internally on the glucan polymer (A. Boraston, unpublished results), bound as well to reduced barley β-glucan as it did to unmodified barley β-glucan. These results demonstrate the specificity of CBM9-2 for the reducing ends of cellulose and β(1→3)(1→4)-glucans. Therefore, it is probable that CBM9-2 binds these polysaccharides by the same mechanism that it binds the small soluble sugars. CBM9-2 is the only currently characterized CBM demonstrated to be specific for the reducing ends of polysaccharides. This may provide a non-destructive method to determine the accessibility of reducing ends in cellulosic substrates.
References


Chapter 2

Exploring the Cellulose/Xylan Specificity of the β-1,4-Glycosidase Cex from *Cellulomonas fimi* through Crystallography and Mutation

This chapter has been reformatted from original publication: Valerie Notenboom, Camelia Birsan, R. Antony J. Warren, Stephen G. Withers, and David R. Rose (1998) Exploring the cellulose/xylan specificity of the β-1,4-glycanase Cex from *Cellulomonas fimi* through crystallography and mutation *Biochemistry*, 37, 4751-4758.

Copyright © 1998 American Chemical Society
Abstract
The retaining β-1,4-glycanase Cex from Cellulomonas fimi, a family 10 glycosyl hydrolase, hydrolyzes xylan 40-fold more efficiently than cellulose. To gain insight into the nature of its preference for xylan, we determined the crystal structure of the Cex catalytic domain (Cex-cd) trapped as its covalent 2-deoxy-2-fluoroxyllobiosyl-enzyme intermediate to 1.9 Å resolution. Together with the crystal structure of unliganded Cex-cd [White, A., et al. (1994) Biochemistry 33, 12546-12552] and the previously determined crystal structure of the covalent 2-deoxy-2-fluorocellobiosyl-Cex-cd intermediate [White, A., et al. (1996) Nat. Struct. Biol. 3, 149-154], this structure provides a convincing rationale for the observed substrate specificity in Cex. Two active site residues, Gln87 and Trp281, are found to sterically hinder the binding of glucosides and must rearrange to accommodate these substrates. Such rearrangements are not necessary for the binding of xylobiosides. The importance of this observation was tested by examining the catalytic behavior of the enzyme with Gln87 mutated to Met. This mutation had no measurable effect on substrate affinity or turnover number relative to the wild type enzyme, indicating that the Met side chain could accommodate the glucoside moiety as effectively as the wild type Gln residue.
Introduction

The β-1,4-glycanase Cex from the soil bacterium *Cellulomonas fimii* is capable of hydrolyzing both cellulose and xylan as well as a range of soluble aryl glycosides based upon glucose and xylose. The gene encoding Cex was cloned, expressed in *Escherichia coli*, and subsequently sequenced (O'Neill et al., 1986). When produced in *C. fimii*, Cex is glycosylated, but the recombinant form is not. Glycosylation has no apparent effect on catalytic activity; its main function appears to be protection against proteolysis (Langsford et al., 1987). Cex is a 47 kDa protein comprising an N-terminal catalytic domain (35 kDa) and a C-terminal cellulose binding domain (12 kDa) held together by a linker peptide rich in proline and threonine residues. The domains in Cex retain their respective catalytic and cellulose-binding properties when separated by limited proteolysis (Gilkes et al., 1989; Gilkes et al., 1988).

Cex is a retaining β-glycanase (Withers et al., 1986). Hence, hydrolysis of substrates occurs with retention of anomic configuration, by a double-displacement mechanism involving formation and hydrolysis of a covalent glycosyl-enzyme intermediate via oxocarbenium ion-like transition states, as proposed by Koshland (Koshland, 1953). This mechanism implies the existence of two catalytic residues in the active site of Cex. One, the nucleophile, is present in the ionized form and stabilizes an oxocarbenium ion transition state, subsequently forming a glycosyl-enzyme intermediate; the other acts as the acid-base catalyst, protonating the glycosidic oxygen of the scissile bond. The nucleophilic residue was identified as Glu233 by trapping a covalent 2-deoxy-2-fluoroglucosyl-enzyme intermediate and then sequencing the purified glycopeptide isolated from proteolytic digests (Tull et al., 1991). The assignment was later confirmed by mutation-kinetic analysis (MacLeod et al., 1996). Site-directed mutation of Glu127 yielded mutants with kinetic parameters fully consistent with the role of acid-base catalyst for this residue (MacLeod et al., 1994). A detailed kinetic study of Cex (Tull & Withers, 1994) allowed delineation of the mechanism through identification of rate-
limiting steps, as well as investigation of the transition-state structure for each step. Such mechanistic insights, derived from pre-steady- and steady-state kinetics, Brønsted relationships, kinetic isotope effect measurements, inactivation experiments, and pH studies, provide further supporting evidence for the double-displacement mechanism of Cex and other retaining β–glycosidases.

The determination of the X-ray crystal structure of the Cex catalytic domain (Cex-cd) (White et al., 1994) provides a structural basis for explaining the wealth of biochemical information that is available for Cex and related glycosidases. Cex-cd folds into an eight-stranded parallel (α/β)₈-barrel, with an open cleft at the carboxyl-terminal end, proposed to be the active site. The two catalytic glutamates, Glu233 and Glu127, are located on either side of the cleft, being separated by a distance of 5.5 Å, consistent with a retaining mechanism. 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 aglycone in a concerted manner (White et al., 1994). Diffusion of the mechanism-based inactivator 2,4-dinitrophenyl 2-deoxy-2-fluorocellobiose into crystals of Cex-cd led to the formation of a covalent 2-deoxy-2-fluorocellobiosyl-enzyme intermediate. The structure of this glycosyl-enzyme complex was determined by X-ray crystallography (White et al., 1996) and, together with the structural information on the catalytic domain alone, provided more insight into interactions at the active site that are crucial to the catalytic mechanism and substrate preference of Cex.

On the basis of amino acid sequence alignments, Cex has been assigned to family 10 of glycosyl hydrolases (Henrissat & Bairoch, 1993). Family 10 members are primarily categorized as endolytic β–1,4-xylanases, although modest activity against cellobiosides has been reported for several of them (Grepinet et al., 1988; Haas et al., 1992; Luthi et al., 1990; Shareck et al., 1991). In addition, low activity against carboxymethylcellulose (CMC) has been reported in some instances (Luthi et al., 1990). Clearly, glycosyl hydrolases of family 10 are largely xylanolytic but can to a varying degree hydrolyze
cellulose as well, in contrast to the low-molecular weight family 11 enzymes which exclusively hydrolyze xylan. In the case of Cex, the $k_{cat}/K_m$ values for the hydrolysis of aryl xylobiosides are 30-100 times higher than those for the corresponding aryl cellobiosides (Gilkes et al., 1991). Since xylan is a $\beta$-1,4-linked polymer of D-xylose, a saccharide unit similar to glucose but lacking the hydroxymethyl group on C-5, it seems likely that the presence of this group is somehow inhibitory to catalysis. Comparison of the complexed and uncomplexed Cex-cd structures reveals that upon formation of the 2-deoxy-2-fluorocellobiosyl-enzyme intermediate, the side chains of Gln87 and Trp281 reorient from their positions in the native enzyme to accommodate the C-5 hydroxymethyl groups of the distal and proximal glucosyl units, respectively (White et al., 1996).

The aim of the study reported here is to assess whether this reorientation may contribute to the substrate specificity of Cex. The structure of the xylobiosyl-enzyme intermediate addresses whether this reorientation is also required in binding that substrate and whether there are other structural explanations for the specificity. An understanding of the atomic basis for substrate specificity could be achieved by attempting to modify the Cex active site through site-directed mutation to improve its specificity for xylan degradation.

**Materials and Methods**

**General Methods.**

Aryl glycoside substrates were synthesized as described previously (Kempton & Withers, 1992; Tull & Withers, 1994; Ziser et al., 1995) or, where available, were purchased from Sigma Chemical Co. All buffer reagents, unless otherwise specified, were also obtained from Sigma. Cex and CexQ87M were purified as described previously (O'Neill et al., 1986). The cellulose binding domain was removed from Cex by papain digestion, and the Cex-cd was purified as described previously (Ziser et al., 1995).

**Crystallization and Inhibitor Soak..**
The crystallization conditions for wild type Cex-cd are well established (Bedarkar et al., 1992). Crystals appear overnight and generally grow to a maximum size of $1 \times 0.5 \times 0.5$ mm$^3$ within 3 days. Prior to addition of the substrate, the crystals were transferred to a larger volume of artificial mother liquor with a higher polyethylene glycol concentration (from 10 to 15%) in which they were found to be most stable. 2,4-Dinitrophenyl 2-deoxy-2-fluoro-β-xylobioside (2F-DNPX$_2$) was added as a solid to the mother liquor surrounding the crystal prior to data collection. A soaking time of 48 h proved to be optimal in terms of both merging R-factors and substrate occupancy.

**Data Collection and Processing.**

X-rays for diffraction data were generated from a CuKα target on a Rigaku RU200 generator operated at 6 kW (40 kV and 150 mA). Intensities were measured on a San Diego Multiwire Systems area detector, and data were indexed and reduced with the complementary software (Howard, 1985). Refinement was performed by X-Plor version 3.851 (Brünger, 1996). Seven percent of the data (1898 reflections) were set aside as an objective test set to monitor the refinement process and were not used during refinement except for map calculations where all data were included (Kleywegt & Brünger, 1996). The electron density map calculated using phases from only the protein model (unliganded Cex-cd, PDB entry code 2EXO) after 40 cycles of rigid body minimization, followed by simulated annealing from 3000 to 300 K, showed unambiguously the 2-deoxy-2-fluoroxylobioside moiety covalently bound to the catalytic nucleophile Glu233. Subsequent refinement involved several cycles of simulated annealing and positional refinement routines, after each of which electron density was calculated and carefully inspected. Ordered water molecules were included in the structure after every cycle of refinement based on 2lFol-lFcl maps contoured at a map standard deviation of $\sigma$ and lFol-lFcl density contoured at $3\sigma$. The disaccharide was not included in the refinement process until the last cycle, in an effort to avoid biasing the electron density surrounding the sugar moiety. X-ray scattering factors for the fluorine atom were added. Standard X-plor
carbohydrate parameter and topology files (X-plo r param3.cho and toph3.cho, respectively) were adjusted to describe the substrate accurately, i.e., including the fluorine atom on the proximal xylose moiety and the covalent \( \alpha \)-link between the nucleophile Glu233 and the substrate. Only half of the suggested energies were used for the sugar restraints, to allow for some flexibility during refinement and to increase the weight on experimental parameters. Dihedral and improper restraints were not imposed on the disaccharide. Individual B-factor refinement was applied after the R-free value showed no more convergence during energy minimization routines.

*Generation of Cex Gln87Met.*

The mutation Gln87Met was made in vitro by overlap extension (Higuchi, 1990), using pUC12-1.1cex(PTIS) (O'Neill *et al.*, 1986). Synthetic oligonucleotides used for mutagenesis were prepared by the UBC Nucleic Acid and Protein Sequencing Unit (NAPS) with an Applied Biosystems 380A DNA synthesizer. The following primers were used, with underlining indicating where nucleotides were changed: GTA TGG CAC TCG MG CTG CCC GAC TGG and CCA GTC GGG CAG CAT CGA GTC CCA TAC. *E. coli* JM101 was transformed with the mutated plasmid. Transformants were selected on ampicillin, and the identity of the plasmid was checked therein by restriction digestion and sequencing of that part of the cex gene containing the mutation. The 500-base pair BamHI-NotI fragment from an appropriate clone was used to replace the corresponding fragment in wild type pUC12-1.1cex(PTIS). After transformation of *E. coli* JM101 and selection on ampicillin, the plasmid was sequenced between the BamHI and NotI sites to ensure that only the desired mutation was present. The plasmid, designated pCexQ187M, was used to transform various *E. coli* strains which were then tested for production of the mutant protein. *E. coli* Topp5 gave the best yield (data not shown) and was used for enzyme production.

*Kinetic Analysis of the Gln87Met Mutant.*
The enzymic hydrolysis rates were measured by spectrophotometric detection of phenol release or by measuring the amount of reducing sugars liberated during hydrolysis of sugar polymers. Michaelis-Menten parameters for the enzymatic hydrolysis of all aryl glycoside substrates, with the exception of PhC and PhX₂, were determined using a continuous assay, by recording changes in absorbance using a UV/Vis Pye-Unicam 8700 spectrophotometer equipped with a temperature-controlled circulating water bath. Solutions of the appropriate substrate concentrations in 50 mM citrate or phosphate buffer (pH 5.5 or 7.0, respectively) with 1 mg of BSA mL⁻¹ at 25 or 37 °C were preincubated within the spectrophotometer until thermally equilibrated; reactions were initiated by the addition of enzyme. Hydrolysis was continuously monitored at a wavelength at which there was a convenient absorbance difference between the initial glycoside and the phenol product as previously reported (Ziser et al., 1995). The concentration of enzyme added was selected so that less than 10% of the total substrate was converted to product during analysis. Substrate concentrations employed ranged from 0.2 to 5 times the Km value ultimately determined, wherever possible. Hydrolysis rates for the substrates PhC and PhX₂ were determined at 37 °C using a stopped assay. Substrate at different concentrations in 50 mM phosphate buffer, with 1 mg of BSA mL⁻¹ (pH 7.0) (190 μL), was prewarmed at 37 °C and the reaction initiated by the addition of 10 μL of enzyme. After an appropriate time, 0.6 mL of 2.0 M Na₃PO₄ (pH 12.1) was added to stop the reaction. The absorbance of the released phenolate at 288 nm was determined immediately, corrected for the spontaneous hydrolysis of substrate and the background absorbance of the enzyme (Δ(= 2.17 mM⁻¹ cm⁻¹ for phenol at pH 12.1), and used to calculate rates. The program Grafit (Leatherbarrow, 1990) was used to perform nonlinear regression analysis in the determination of Km and Vmax.

CM-cellulase and xylanase activities were measured by the p-hydroxybenzoic acid hydrazide (HBAH) method (Miller, 1960) which quantitates the production of reducing sugar. Initial rates of CM-cellulose hydrolysis were determined by incubating 100 μL of
an enzyme solution with 900 μL of 2% CM-cellulose in 50 mM citrate buffer at pH 6.8 and 30 °C. Aliquots of 100 μL were removed at 0, 2, 4, 6, 8, and 10 min intervals and added to 1 mL of 5 mM NaOH in glass test tubes. After all samples were taken, 1 mL of HBAH reagent was added to each tube and the samples were heated in a boiling water bath for 12 min. When tubes had cooled to room temperature, A420nm was measured and the amount of reducing sugar released was determined using a glucose standard curve generated under the same conditions. The activity is expressed as inverse seconds or nanomoles per milliliter of reducing sugars released per second per nanomoles per milliliter of enzyme. To ensure the validity of the assays for enzyme activity, concentrations of enzyme were chosen so that the production of reducing sugar in the assay fell within the linear response range of the standard curve. Appropriate controls for background reducing sugar and/or for the presence of chemical reagents that might interfere with the assay were always included with the assay procedures. Substrate for the xylanase assays was prepared by dissolving 6-10 g of birchwood xylan in 100-200 mL of distilled water at room temperature for 3-4 h with continuous stirring. Insoluble material was removed by centrifugation at 16000g in a GSA rotor for 20 min. The clear supernatant was lyophilized, and the dry, water-soluble product was used as substrate at a concentration of 0.2%; the release of reducing sugars was determined as described above.

Results and Discussion


Due to the very poor solubility of the slow substrate 2F-DNPX₂ at the concentration being sought (0.5 mM), soaking the Cex-cd crystal in a solution containing a significant amount of solid material proved to be necessary. The intent was that the substrate would partition into the crystal via the small amount in solution. Successful soaking was indicated by the appearance of yellow 2,4-dinitrophenolate in the solution overnight as a consequence of the reaction shown in Figure 2.1, in which the covalent 2-deoxy-2-
fluoroxylobiosyl-enzyme intermediate is formed. An additional indication of reaction occurring was the fact that the crystal surface appeared to be cracked a few hours after the addition of 2F-DNPX₂, but these cracks annealed over time.

**Figure 2.1**

X-ray diffraction data were collected to 1.9 Å resolution at room temperature. A single crystal was sufficient for a complete data set, because of the resilience and high symmetry of the tetragonal crystals (P4₁2₁2). After three cycles of data reduction, the indexing parameters and merging residuals converged to yield the statistics shown in Table 2.1. The unit cell dimensions of the complexed crystal differed only slightly from those of the native (Δa = Δb = 0.05 Å, Δc = 0.28 Å). Rigid body refinement indicated a slight rotation of the molecule in the unit cell compared to the packing of native Cex-cd. The structure and refinement statistics are listed in Table 2.2. The current model contains 2399 non-hydrogen protein atoms, 18 non-hydrogen saccharide atoms, and 123 ordered water molecules. Data between 10 and 1.9 Å resolution and with an F/σ of >2 refined to a crystallographic R-factor of 0.21 and an R-free of 0.26. Of the non-glycine residues, 91% have φ and ψ angles located in the most favored regions, 8% fall in the additional allowed regions, and the remaining residue Thr81 assumes the unusual cis conformation,
as described previously (Tull & Withers, 1994).

**Table 2.1 Data collection statistics**

<table>
<thead>
<tr>
<th>Resolution shell</th>
<th>to 1.9Å</th>
<th>2.05 – 1.90 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>197626</td>
<td>25539</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>25857</td>
<td>5050</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.94</td>
<td>99.90</td>
</tr>
<tr>
<td>I/σ</td>
<td>16.19</td>
<td>3.58</td>
</tr>
<tr>
<td>R_{merge}</td>
<td>8.03</td>
<td>27.03</td>
</tr>
</tbody>
</table>

$$R_{merge} = \frac{\sum_{\mathbf{hkl}} \sum_{i} |I_{hkl} - \overline{I}_{hkl}|}{\sum_{\mathbf{hkl}} \sum_{i} I_{hkl}}$$
### Table 2.2: Structure and Refinement Statistics

<table>
<thead>
<tr>
<th>Resolution shell</th>
<th>10 – 1.9 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell (a=b, c in Å)</td>
<td>88.18, 81.29</td>
</tr>
<tr>
<td>R-factor (%)</td>
<td>21.2</td>
</tr>
<tr>
<td>R-free (%)</td>
<td>26.0</td>
</tr>
<tr>
<td>No. protein atoms</td>
<td>2399</td>
</tr>
<tr>
<td>No. substrate atoms</td>
<td>18</td>
</tr>
<tr>
<td>No. solvent atoms</td>
<td>123</td>
</tr>
<tr>
<td>R.m.s. deviation s</td>
<td></td>
</tr>
<tr>
<td>bond lengths (Å)</td>
<td>0.005</td>
</tr>
<tr>
<td>bond angles (°)</td>
<td>1.23</td>
</tr>
<tr>
<td>dihedral angles (°)</td>
<td>22.8</td>
</tr>
<tr>
<td>Average B-factor (Å²)</td>
<td>13.0</td>
</tr>
<tr>
<td>Ave. B-factor substr. (Å²)</td>
<td>6.64</td>
</tr>
<tr>
<td>Ave. B-factor solvent (Å²)</td>
<td>23.20</td>
</tr>
</tbody>
</table>

\[
R_{crys} = \frac{\sum_{hkl} |F_{obs}| - k|F_{calc}|}{\sum_{hkl} |F_{obs}|}, \quad R_{free}^{\text{free}} = \frac{\sum_{hkl \in T} |F_{obs}| - k|F_{calc}|}{\sum_{hkl \in T} |F_{obs}|}
\]
The active site of Cex trapped as its 2-deoxy-2-fluoroxyllobiosyl-enzyme intermediate is shown in Figure 2.2, indicating the relevant interactions of the enzyme with the sugar moiety.

**Figure 2.2** Schematic representation of interactions in the active site region of the Cex-cd 2F-xylobioside complex. Distances are measured in Ångstroms. The distance between 2'F and the carbonyl oxygen of E233 is 2.6Å (not shown). The boxes below the diagram roughly indicate the area of the binding subsite. The -3 subsite is shown in dashed lines because it is not occupied by substrate.

The covalent nature of the glycosyl-enzyme species is clearly seen in Figure 2.3a, the contiguous electron density between the carboxyl side chain of Glu233 and the anomeric carbon of the sugar moiety being quite evident. The electron density displayed is that determined for the 2-deoxy-2-fluoroxyllobiosyl-Cex-cd structure described in this paper. The glycosyl-enzyme bond between C-1 of the proximal sugar and the nucleophilic oxygen of Glu233 is 1.45 Å in length. Table 2.3 lists a selection of bond angles of the final refined conformation of the proximal (covalently linked) 2-deoxy-2-fluoroxyllose unit, as compared to suggested parameters during refinement. The fluorine substituent at
C-2 of the proximal saccharide is seen to be only slightly removed from its expected position in the relaxed $\mathbf{C_1}$ conformation. The dihedral angle 3'-OH-C-3-C-2-2'-F measures 58° as opposed to 66° for 3'-OH-C-3-C-2-2'-OH in the distal sugar. This slight dihedral "flattening" is also seen in the crystal structure of the cellobiosyl-Cex-cd covalent intermediate, in which the proximal 2'-hydroxyl group is retained, formed by a catalytically deficient mutant of the enzyme (Notenboom et al., 1998). Therefore, this is likely to be a characteristic of the covalent reaction intermediate rather than a consequence of the fluorine substitution. Interestingly, even larger distortion was observed in the 2-deoxy-2-fluorocellobiosyl-Cex-cd structure, with a dihedral angle of 21° being found (White et al., 1996). One consequence of, or perhaps reason for, this distortion is the apparent presence of a hydrogen bond between the fluorine at C-2 of the 2-fluorocellobiosyl moiety and the side chain amide nitrogen of Asn126, a highly conserved residue. No such strong interaction is seen in the structure of the 2-deoxy-2-fluoroxylobiosyl-Cex-cd, the separation of 3.3 Å being too great. A water molecule is found positioned between the acid-base catalyst Glu127 (2.9 Å) and C-1 (3.6 Å) of the covalently linked proximal xylose moiety in the latter structure, in an ideal position for attacking the anomeric carbon.
Figure 2.3a  Stereo diagram of 2|Fol|-|Fc| electron density for the 2F-xylobiosyl-Cex-cd structure, contoured at 1.2σ, superposed on representative atoms of the final refined structure. The density was calculated from all data to 1.9Å resolution and the final protein and water coordinates. The figure was drawn using the program Setor (Evans, 1993).
Figure 2.3b  Stereo representation of a superposition of uncomplexed Cex-cd (White et al. 1994, pdb entry code 2EXO) shown in orange, 2F-cellobiosyl-Cex-cd (White et al. 1996, pdb entry code 1EXP) shown in green, and 2F-xylobiosyl-Cex-cd shown in bulky white, with oxygen atoms shown in red, nitrogens in blue and the fluorine atom in yellow. Trp281, flanking the proximal sugar, does not need to reorient to accommodate the hydroxymethylene group as seen in the 2F-cellobiosyl-enzyme complex (a 31° rotation around its $C_\beta$-$C_\gamma$ bond). Glu87 is disordered in the 2F-cellobiosyl-enzyme structure, but is stabilised in the 2F-xylobiosyl-enzyme complex through a water molecule and Ser86, which also adopts a different conformation in this interaction. The figure was drawn in Setor (Evans, 1993).
Table 2.3: Comparison of ideal and measured angles

<table>
<thead>
<tr>
<th>Angle (deg)</th>
<th>Imposed</th>
<th>observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1-C2-C3</td>
<td>110.4</td>
<td>112.0</td>
</tr>
<tr>
<td>C3-C2-F</td>
<td>109.7</td>
<td>112.6</td>
</tr>
<tr>
<td>C5-O5-C1</td>
<td>112.0</td>
<td>109.1</td>
</tr>
<tr>
<td>C1-C2-F</td>
<td>109.7</td>
<td>111.7</td>
</tr>
<tr>
<td>C1-Oε1-CE</td>
<td>112.0</td>
<td>116.3</td>
</tr>
<tr>
<td>Oε1-C1-C2</td>
<td>109.7</td>
<td>111.3</td>
</tr>
<tr>
<td>Oε1-CE-Oε2</td>
<td>123.0</td>
<td>122.5</td>
</tr>
<tr>
<td>Oε1-C1-O5</td>
<td>112.1</td>
<td>109.7</td>
</tr>
</tbody>
</table>

Apart from the slight changes in conformation of the sugar moiety, the significant differences in structure between the free Cex-cd, 2-deoxy-2-fluorocellubiosyl-Cex-cd, and 2-deoxy-2-fluoroxylbiosyl-Cex-cd are those seen within the enzyme active site, specifically in the positions of the side chains of Gin87 and Trp281. These changes are shown in Figure 2.3b which presents a superposition of the sugar moieties in the two covalent intermediates and of the positions of selected side chains for all three proteins.
As can be seen, the indole side chain of Trp281 is displaced significantly from its position in the free enzyme upon formation of the 2-deoxy-2-fluorocellobiosyl intermediate, whereas no such change is seen upon formation of the 2-deoxy-2-fluoroxyllobiosyl-enzyme intermediate. These conformational changes appear to be caused by repulsive steric interactions between the hydroxymethyl group at C-5 of the glucose of the proximal sugar and the tryptophan side chain. Such interactions do not occur when a xylose residue is present at this position. Similarly, there appear also to be repulsive, steric interactions between the hydroxymethyl group of the distal sugar and Gln87. In the native enzyme, the side chain of Gln87 is somewhat disordered, with side chain B-factors refined to an average of 47.5 Å²; apparently, it is forced outward in the 2-deoxy-2-fluorocellobiosyl-Cex-cd intermediate. By contrast, the side chain of Gln87 is well defined in the 2-deoxy-2-fluoroxyllobiosyl-Cex-cd intermediate (average side chain B-factors of 11.8 Å²) in a position previously occupied by the C-5 hydroxymethyl side chain of glucose; it appears to form a hydrogen bond with the endocyclic oxygen atom (O-5) of the distal sugar. Furthermore, Ser86 assumes a different side chain conformation in the 2-deoxy-2-fluoroxyllobiosyl-Cex-cd intermediate, compared to that in both crystal structures described previously, now forming a hydrogen bond to a water molecule not seen in the 2-deoxy-2-fluorocellobiosyl-Cex-cd intermediate. In turn, this water molecule forms a fairly close interaction (2.9 Å) with Gln87, stabilizing its side chain even more (Figure 2.3a,b). When the structures are superimposed on other family 10 crystal structures found in the literature, the only differences in the active sites can be found at positions equivalent to Ser86 and Gln87 of Cex. For example, in XynZ from *Cellulomonas thermocellum* ((Domínguez et al., 1996); PDB entry code 1XYZ), an Asn corresponds to Ser86 of Cex; and in XylA from *Pseudomonas fluorescens* ssp. cellulosa ((Harris et al., 1994); PDB entry code 1CLX), a Pro and a Tyr correspond to Ser86 and Gln87, respectively, of Cex. It is possible that these positions, though located in the active site, are not as sensitive to evolutionary pressure as the strictly conserved residues.
with respect to their contribution to the catalytic mechanism. This could therefore result in a more promiscuous enzyme that recognizes both xylosides and glucosides. It should be noted that differences in interactions during binding of the polymers xylan and cellulose at the more remote subsites could play a significant role in substrate specificity as well. In addition, the cellulose binding domain at the C-terminal end of Cex possibly affects hydrolysis rates of cellulose polymers. However, it appears that the specificity for xylan over cellulose substrate arises, at least in part, from repulsive steric interactions with Gln87 and Trp281 in the cellobiosyl-Cex-cd complex, as well as from the stabilization of Gln87 in the xylobiosyl-Cex-cd complex. Attempts to engineer an improved xylanase/cellulase ratio should logically focus upon these two residues initially. Our focus was directed in the first instance upon Gln87 because Trp281 cannot be replaced with a bulkier amino acid. It may be possible in the future to prevent displacement of Trp281 from the active site by introducing a bulkier residue behind Trp281, in the position occupied in the wild type enzyme by phenylalanine.

Our initial intention was to generate mutants in which Gln87 was replaced by Met, His, and Tyr. Met was chosen as a residue of slightly greater dimensions than Gln, and one with which hydrogen bonding possibilities are limited since it cannot donate a hydrogen bond. His was chosen as an amino acid with a bulkier side chain, but one which retains most of the hydrogen bonding possibilities, while Tyr is a larger residue with some polar character. Unfortunately, all attempts to obtain the His and Tyr mutants were unsuccessful; workable quantities of either protein could not be produced. Fortunately, sufficient quantities of the Met mutant were readily generated.

**Catalytic Properties of the Gln87Met Mutant.**

The Gln87Met mutant behaved in a manner virtually identical to that of the wild type enzyme during purification; both enzymes were desorbed from the cellulose affinity column by approximately the same volume of eluant, and both ran as a single bands with an Mr of 47 kDa on SDS-PAGE (>95% purity by inspection). The yield of purified
The enzyme ranged from 15 to 25 mg per liter of liquid culture. Mass spectrometric analysis confirmed the molecular mass of the mutant protein as 47 120 Da, exactly as predicted. Correct folding was indicated by the CD spectrum and the thermal stability of the mutant enzyme being virtually identical to those of the wild type enzyme. Values of $k_{cat}$ and $K_m$ for the mutant and wild type enzymes were determined on a range of glucoside, celllobioside, xyloside, and xylobioside substrates (Table 2.4). Substitution of Gln87 by Met in the binding site of Cex did not alter the $K_m$ and $k_{cat}$ values for the two series of substrates significantly. This is best expressed in the form of the linear free energy relationship in which values of $\log(k_{cat}/K_m)$, the specificity constant, for each substrate with the Gln87Met mutant, are plotted against the same parameter for the wild type enzyme (Figure 2.4). The excellent correlation coefficient ($\rho = 0.974$) and the slope of unity ($r = 1.03$) confirm that the two enzymes have essentially identical specificities. The small variations observed are most likely due to perturbations in the local environment within the protein. Thus, in contrast to what had been hoped, the Gln87Met mutant did not show a decreased affinity for cello-oligosaccharide substrates. Presumably, the methionine residue is still sufficiently flexible to undergo a rearrangement similar to that observed for Gln87 upon binding the gluco substrates in the active site. It is possible that the hydrogen bonding interaction between Gln87 and the endocyclic oxygen of the distal sugar either does not play a very important role in specificity or is compensated for in some way. It is likely, therefore, that the binding interactions with the C-5 hydroxymethylene groups in the active site are more complex than initially thought, and that a single point mutation in the catalytic domain of Cex may be insufficient to produce the desired changes in the specificity of the enzyme. It is significant in this context that the Gln87Ser mutant of Cex, constructed in a manner similar to that of the Gln87Met mutant, behaves kinetically like the latter with the same substrates (A. MacLeod, personal communication).
Table 2.4  Kinetic parameters for the hydrolysis of β-glycosides by Cex and the Gln87Met mutant at 37 °C, pH 7.0.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Enzyme</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat}/K_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNPG</td>
<td>Cex wt</td>
<td>0.024</td>
<td>8.3</td>
<td>2.9 x 10$^{-3}$</td>
</tr>
<tr>
<td></td>
<td>Gln87Met</td>
<td>0.024</td>
<td>7.4</td>
<td>3.2 x 10$^{-3}$</td>
</tr>
<tr>
<td>3,4-DNPG</td>
<td>Cex wt</td>
<td>2.9</td>
<td>6.5</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>Gln87Met</td>
<td>1.2</td>
<td>6.0</td>
<td>0.20</td>
</tr>
<tr>
<td>PNPC</td>
<td>Cex wt</td>
<td>15.8</td>
<td>0.60</td>
<td>26.3</td>
</tr>
<tr>
<td></td>
<td>Gln87Met</td>
<td>9.2</td>
<td>0.34</td>
<td>27.1</td>
</tr>
<tr>
<td>3,4-DNPC</td>
<td>Cex wt</td>
<td>9.7</td>
<td>0.14</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Gln87Met</td>
<td>8.3</td>
<td>0.065</td>
<td>128</td>
</tr>
<tr>
<td>PhC</td>
<td>Cex wt</td>
<td>-</td>
<td>-</td>
<td>1.0 x 10$^{-4}$</td>
</tr>
<tr>
<td></td>
<td>Gln87Met</td>
<td>-</td>
<td>-</td>
<td>2.3 x 10$^{-4}$</td>
</tr>
<tr>
<td>PNPX</td>
<td>Cex wt</td>
<td>2.6</td>
<td>20</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Gln87Met</td>
<td>1.2</td>
<td>67</td>
<td>0.018</td>
</tr>
<tr>
<td>3,4-DNPX</td>
<td>Cex wt</td>
<td>22</td>
<td>7.9</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>Gln87Met</td>
<td>9.0</td>
<td>7.2</td>
<td>1.2</td>
</tr>
<tr>
<td>PNPX$_2$</td>
<td>Cex wt</td>
<td>39.8</td>
<td>0.018</td>
<td>2200</td>
</tr>
<tr>
<td></td>
<td>Gln87Met</td>
<td>57.9</td>
<td>0.12</td>
<td>482</td>
</tr>
<tr>
<td>3,4-DNPX$_2$</td>
<td>Cex wt</td>
<td>22</td>
<td>0.012</td>
<td>1840</td>
</tr>
<tr>
<td></td>
<td>Gln87Met</td>
<td>38.2</td>
<td>0.038</td>
<td>1005</td>
</tr>
<tr>
<td>PhX$_2$</td>
<td>Cex wt</td>
<td>10.4</td>
<td>8.6</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>Gln87Met</td>
<td>18.7</td>
<td>21.2</td>
<td>0.9</td>
</tr>
<tr>
<td>2% CMC</td>
<td>Cex wt</td>
<td>0.052</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Gln87Met</td>
<td>0.025</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1% Xylan</td>
<td>Cex wt</td>
<td>0.47</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Gln87Met</td>
<td>0.40</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Abbreviations: PNP – para-nitrophenol; DNP – dinitrophenol; X – xylose; X$_2$ – xylobiose; C – cellobiose; G – glucose; CMC – carboxymethyl cellulose; Ph – phenol
Figure 2.4  Plot of log(kcat/Km) for hydrolysis of a series of aryl glycosides by the Gln87Met mutant of Cex versus the same parameter for the wild type enzyme.
In conclusion, the crystallographic analysis of the Cex-cd-2F-xylobiosyl covalent intermediate has provided a possible structural rationale for the preference of Cex for xylan over cellulose. The C-5 substituent on the glucose rings causes the displacement of two side chains relative to the unliganded enzyme, Gln87 and Trp281, a displacement that is not necessary for the binding of xylobiose. An additional component may be the reduced strain of the sugar ring in the xylobiosyl-enzyme covalent intermediate, as compared to that of the cellobiosyl complex. A first attempt to test these conclusions with a Gln87 to Met mutant did not succeed in altering the cellulose/xylan specificity. A Met side chain presumably retains enough flexibility to accommodate a glucose moiety as well as xylose in the -2 subsite. Alternative strategies could include the incorporation of a bulkier side chain in this position or the substitution by a smaller side chain, to test for the opposite result, an improvement in cellulase activity. As the Gln side chain is disordered in both the wild type and cellobiose intermediate, but is observed to be well ordered in the xylobiose complex, there may be an entropic cost to binding that is actually greater for xylan binding than for cellulose binding. Therefore, the effects of various side chains at position 87 may be revealing in terms of the entropic versus the enthalpic contributions to binding. Structure-based studies along these lines are underway.
References


Chapter 3

Insights into Transition State Stabilization of the β-1,4-Glycosidase Cex by Covalent Intermediate Accumulation.

This chapter has been reformatted from original publication: Valerie Notenboom, Camelia Birsan, Mark Nitz, David R. Rose, R. Antony J. Warren, and Stephen G. Withers (1998) Insights into transition state stabilization of the β-1,4-glycosidase Cex by covalent intermediate accumulation Nature Structural Biology, 5, 812-818.

Copyright © 1998 Nature Publications
Abstract

The catalytic mechanism of ‘retaining’ β-glycosidases has been the subject of considerable interest and debate for many years. The visualization of a covalent glycosyl enzyme intermediate by X-ray crystallography was first accomplished with a saccharide substrate substituted with fluorine at its 2-position (White et al., 1996). The structure implicated major roles for residue His205 and for the 2-hydroxyl position of the proximal saccharide in binding and catalysis. Here we have studied the kinetic behaviour of various His205 mutants. One of these mutants, a double mutant H205N/E127A, has been used to stabilize a covalent glycosyl-enzyme intermediate involving an unsubstituted sugar, permitting crystallographic analysis of the interactions between its 2-hydroxyl group and the enzyme.
**Introduction**

Cellulases and xylanases are key enzymes involved in the degradation of β-glucan biomass, acting through hydrolysis of the β-glycosidic linkages. The amino acid sequences are known for well over 200 such enzymes and are grouped into 64 families of related amino acid sequences; 13 of the families contain only β-1,4-glycanases (Henrissat, 1991; Henrissat & Bairoch, 1993; Henrissat & Bairoch, 1996; Tomme et al., 1995). All members of a family exhibit the same catalytic mechanism (Gebler et al., 1992): the stereochemistry at the anomeric center of the bond being cleaved is either inverted or retained as a consequence of hydrolysis. Inverting glycosidases appear to use a mechanism in which a general acid/base catalyzed direct displacement occurs at the anomeric center via an oxocarbenium ion-like transition state. Retaining glycosidases, however, use a double-displacement mechanism (Figure 3.1) in which a covalent glycosyl-enzyme intermediate is formed and hydrolyzed in a general acid/base catalyzed process via oxocarbenium ion-like transition states, or possibly via oxocarbenium ion intermediates (McCarter & Withers, 1994; Sinnott, 1990).

**Figure 3.1**
The β-1,4-glycanase Cex from *Cellulomonas fimi* is a 47.1 kDa enzyme comprising an N-terminal catalytic domain and a C-terminal cellulose binding domain connected by a proline-threonine rich linker (O'Neill *et al.*, 1986). The catalytic domain (Cex-cd) is a member of family 10 of the glycoside hydrolases (Gilkes *et al.*, 1991; MacLeod *et al.*, 1994). Like some other members of family 10, Cex hydrolyses xylan and, to a lesser extent, carboxymethylcellulose, as well as a range of soluble aryl xylosides, xylobiosides, glucosides and celllobiosides (Gilkes *et al.*, 1991; Tull & Withers, 1994). Cex is a retaining enzyme (O'Neill *et al.*, 1986), whose catalytic nucleophile is Glu233, identified by trapping of a covalent 2-deoxy-2-fluoro-glycosyl-enzyme intermediate (Tull & Withers, 1994; Tull *et al.*, 1991). The acid/base catalyst was identified as Glu127 by detailed kinetic analysis (MacLeod *et al.*, 1994; MacLeod *et al.*, 1996). The X-ray crystal structure of the enzyme (White *et al.*, 1994) identified the amino acid residues forming the environments around these key carboxylates and likely, therefore, to play important roles in catalysis. His205 is of particular interest, being hydrogen-bonded to both the nucleophile, Glu233, and to Asp235, a trio of residues highly conserved within family 10 enzymes. This paper describes the detailed kinetic analysis of His205 mutants of Cex, whose properties suggested an approach to determining the three-dimensional structure of a true glycosyl-enzyme covalent intermediate formed during catalysis by a retaining glycosidase. The original formulation of this mechanism proposed a covalent intermediate (Koshland, 1953) whereas an oxocarbenium ion/carboxylate ion pair was proposed later for the mechanism of a retaining lysozyme (Phillips, 1967). However, a substantial body of evidence for enzymes other than lysozyme now favours the covalent species. Of particular importance in this regard are the α-secondary deuterium kinetic isotope effects observed with several retaining β-glycosidases. If the principal intermediate formed is an ion pair, then isotope effects for the deglycosylation step should be inverse (k_H/k_D <1). However, measured isotope effects for this step have invariably been greater than 1, ranging from k_H/k_D = 1.25 for *E. coli* β-galactosidase
(Sinnott & Souchard, 1973) to $k_H/k_D = 1.11$ for both an *Agrobacterium* sp. β-glucosidase (Kempton & Withers, 1992) and for Cex (Tull & Withers, 1994), strongly supporting a covalent (sp³) intermediate flanked by sp² hybridized transition states. Direct evidence for the covalent nature of this intermediate comes from the detection by $^{19}$F-NMR (Withers et al., 1988) and electrospray mass spectrometry (Withers & Aebersold, 1995) of stable intermediates formed from 2-deoxy-2-fluoro glycosides. Recently, the three-dimensional structures of such intermediates formed by Cex have been determined crystallographically (Notenboom et al., 1998; White et al., 1996). Interactions at the 2-position are extremely important to transition state stabilization in a number of glycosidases, contributing at least 10 kcal.mol⁻¹ in the case of Cex and similar amounts in a range of other glycosidases (Namchuk & Withers, 1995). Unfortunately, since the 2-position is substituted by fluorine in the trapped complexes it has been impossible to identify unequivocally the interactions involved at this position. Here we directly address the nature of the interactions between the enzyme and the 2-hydroxyl of the glycone in the bond being hydrolyzed, by presenting the structure of a covalent cellobiosyl-enzyme intermediate formed between the His205N/E127A double mutant of Cex-cd (Cex-DM) and an unsubstituted cellobiosyl moiety.

**Materials and Methods**

*Chemicals, media components and enzymes*

All chemicals used were of analytical or HPLC grade; they were obtained from Sigma or BDH. All media components were from Difco. Restriction endonucleases, DNA polymerase, ligase and nucleotides were from Pharmacia or New England Biolabs. Sequenase Version 2.0 was from United States Biochemical. Buffers and other solutions were prepared as described previously (Sambrook et al., 1989).
**Bacterial strains and plasmids**

*Escherichia coli* strain JM101 (Yanisch-Perron et al., 1985) was used as the host for genetic manipulations; *E. coli* strain Topp5 (Stratagene, La Jolla, CA) was used for protein production. Plasmids pUC12-1.1cex (PTIS) and pUC12-1.1cex E127A (PTIS) were described previously (MacLeod et al., 1994; O'Neill et al., 1986).

**Construction of mutants**

The CexH205N and CexH205A mutants were made by overlap-extension PCR using pUC12-1.1cex (PTIS) as template (Higuchi, 1990). The primers were designed so that the final PCR products contained the desired mutation flanked by BamHI and NotI sites to allow the corresponding sequence in wild-type pUC12-1.1cex (PTIS) to be replaced with the mutated fragment, and a silent mutation that introduced a BstXI site for preliminary screening of ampicillin-resistant transformants. The presence of only the desired mutations in each construct was confirmed by sequencing the 500 bp BamHI-NotI fragment from each mutant, using the modified dideoxy chain termination method (Tabor & Richardson, 1987) with the following modifications: in the primer extension reaction, T7 DNA polymerase (Sequenase) was used, the reaction temperature was increased to 43°C, and 7-deazaGTP was substituted for dGTP. The CexE127A/H205N double mutant was constructed by recombining appropriate restriction fragments: the 3293 bp BamHI-MluI fragment from pUC12-1.1cex E127A (PTIS) carrying the E127A mutation and the 1224 bp fragment from pUC12-1.1cex H205N (PTIS) carrying the H205N mutation. The plasmids in ampicillin-resistant transformants were screened for correct size, restriction patterns and loss of the BstXI site. The presence of only the desired mutations was confirmed by DNA sequencing.

**Production of mutant proteins**

*E. coli* Topp5 was transformed with the plasmids encoding the mutant proteins.
The mutant proteins were produced and purified by affinity chromatography (MacLeod et al., 1994). The Cex catalytic domain was obtained by proteolytic removal of the CBD using papain (Gilkes et al., 1991). Purity was assessed by SDS-PAGE.

Characterization of mutants

The molecular weights of purified Cex and Cex mutants were determined by Ion Spray Mass Spectrometry performed on a PE SCIEX API 300 ion spray LC/MS system. CD spectra were recorded on a JASCO J-720 spectropolarimeter controlled with J-700 software. Spectra were obtained at a protein concentration of 0.3 mg.mL⁻¹ in 5 mM phosphate buffer, pH 7.0, 25 °C using a 100 μL silicon quartz cell with a 1 mm path length. Spectra were recorded four times from 190 to 300 nm at a scan rate of 50 nm/min, using a 2.0 s response and a sensitivity of 10 mdeg. The thermal stabilities of the proteins were assessed by monitoring changes in circular dichroism at 220 nm while ramping up the temperature.

Kinetic studies

Kinetic studies (Table 3.1) were performed in 50 mM citrate or phosphate buffer, pH 5.5 or 7.0 respectively, 1 mg.mL⁻¹ BSA at 25 °C or 37 °C, as described previously (MacLeod et al., 1996; Tull & Withers, 1994). Inactivation studies were performed using 1.5 mM 2F-DNPC. The dependence of kcat/Km upon pH (Figure 3.2) for Cex wild-type and the H205 mutants was determined by progress curve analysis at low substrate concentration using a PNPC concentration of 0.2 × Km in the appropriate buffer containing 1 mg/mL BSA and 145 mM NaCl at 25°C. The pH of each reaction mixture was measured after completion of reaction to ensure that the pH had not fluctuated during the reaction. Measurement of the pre-steady state burst was carried out by adding to 2,4-DNPC (190 μl, 40 μM) in 50 mM phosphate buffer, 0.1% BSA, pH 7.0 at 25 °C, 10 μl enzyme (2.6 μM His205Ala or 2.3 μM His205Asn) and the reaction followed by UV/visible spectrophotometry at 400 nm. Extrapolation of the linear portion of the burst phase and steady state phase back to time zero determined the size of the burst.
Crystallization and Data Collection of Cex-DM

Crystals of the covalent Cex-DM cellobiosyl intermediate were grown by vapour diffusion from hanging drops in the same space group (P4₁2₁2) as wild type Cex-cd in the same conditions as wild type enzyme (Bedarkar et al., 1992). A complete data set was obtained at room temperature from a single crystal mounted in a 1.5 mm glass capillary tube. The crystal was exposed to 6kW (40kV × 150mA) Copper Kα X-rays (λ=1.54 Å), from a Rigaku RU200 rotating anode. Diffraction data were collected on a San Diego Multiwire Detector System to 1.73 Å resolution from 60s exposures on 0.1° oscillation frames. Data collection statistics are listed in Table 3.2.

Phase Approximation

7% of reflections were selected and set aside prior to any structure refinement routines, in order to monitor the R-free factor (Kleywegt & Brünger, 1996). These reflections were included during map calculations. The coordinates for the catalytic domain of wild type Cex (2EXO) were used as an initial model to position the molecule in the asymmetric. X-Plor version 3.851 (Brünger, 1996) was used for refinement. Initially a 40-cycle rigid body minimization search was performed using data from 8.0 Å to 2.2 Å. Data from 15.0Å to 2.0Å were then used for simulated annealing from 3000K to 300K in 25° intervals, followed by positional refinement. Prior to this routine the point mutated residues were changed to Ala in the model, in order to confirm the location and nature of the mutations. Electron density maps with coefficients 2IFol-IFcl and IFol-IFc₁, were subsequently calculated using phases from this model and all data to 1.8 Å, displayed at an average map contour level of 1.0σ and 3.0/-3.0σ, respectively. Residues that were found to have adopted conformations other than those seen in wild type Cex were then changed to alanine, and this model also underwent simulated annealing in an effort to remove any remaining bias. Waters were included to the model based on the appearance of spherical density in 2IFol-IFcl and IFol-IFc₁ electron density maps, and when the suggested position was in an appropriate orientation to donate or accept a hydrogen bond.
When the R-free of the model ceased to converge (at 32%), a celllobiosyl molecule was built into the 2|Fol-lFcl electron density with the phases calculated from the protein and water atoms. Parameter and topology files (param3.cho and toph3.cho, resp.) were adjusted to reflect the geometry and chemistry of the covalent α-glycosidic bond to Glu233, but dihedral and improper angle restraints for the substrate were left unimposed. Individual B-factors were determined after incorporation of water molecules and the substrate into the model. No significant change in the R-free value was observed when anisotropic B-factor scaling was applied. Statistics for the refined structure are listed in Table 3.1a and 3.1b.

Table 3.1a Data collection statistics

<table>
<thead>
<tr>
<th>Resolution shell</th>
<th>to 1.74Å</th>
<th>1.87 - 1.74 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>255846</td>
<td>23362</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>32211</td>
<td>5297</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.10</td>
<td>80.40</td>
</tr>
<tr>
<td>I/σ</td>
<td>19.01</td>
<td>2.01</td>
</tr>
<tr>
<td>R_{merge}</td>
<td>6.81</td>
<td>36.11</td>
</tr>
</tbody>
</table>

\[
R_{merge} = \frac{\sum \sum |I_{hkl} - \overline{I}_{hkl}|}{\sum \sum \overline{I}_{hkl}}
\]
### Table 3.1b: Structure and Refinement Statistics

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution shell</td>
<td>15 – 1.80 Å</td>
</tr>
<tr>
<td>Cell (a=b, c in Å)</td>
<td>88.41, 80.67</td>
</tr>
<tr>
<td>R-factor, all data (%)</td>
<td>22.7</td>
</tr>
<tr>
<td>R-free, all data (%)</td>
<td>28.5</td>
</tr>
<tr>
<td>No. protein atoms</td>
<td>2390</td>
</tr>
<tr>
<td>No. substrate atoms</td>
<td>22</td>
</tr>
<tr>
<td>No. solvent atoms</td>
<td>131</td>
</tr>
<tr>
<td>R.m.s. deviation s</td>
<td></td>
</tr>
<tr>
<td>bond lengths (Å)</td>
<td>0.011</td>
</tr>
<tr>
<td>bond angles (°)</td>
<td>1.58</td>
</tr>
<tr>
<td>dihedral angles (°)</td>
<td>23.25</td>
</tr>
<tr>
<td>Average B-factor (Å²)</td>
<td>20.33</td>
</tr>
<tr>
<td>Ave. B-factor substr. (Å²)</td>
<td>34.48</td>
</tr>
<tr>
<td>Ave. B-factor solvent (Å²)</td>
<td>34.53</td>
</tr>
</tbody>
</table>

\[
R_{crys} = \frac{\sum_{hkl} |F_{obs}| - k|F_{calc}|}{\sum_{hkl}|F_{obs}|} \quad R_{free} = \frac{\sum_{hkl<} |F_{obs}| - k|F_{calc}|}{\sum_{hkl<}|F_{obs}|}
\]
Results and Discussion

Purification and Structural Characterization

Purification of the two single mutants of Cex (His205Asn and His205Ala) proceeded with essentially no differences in behaviour from that of the wild type enzyme, final protein yields ranging from 0.5 to 1.0 mg/liter of culture. Electrospray mass spectrometric analysis of the purified proteins confirmed the expected mass changes relative to wild type Cex, as follows: wild type Cex, 47118 (+5); His205Ala, 47058 (+3); His205Asn, 47099 (+4). Circular dichroism spectra (data not shown) of the mutants were essentially identical to those of the wild type enzyme, indicating that no major changes in structure had occurred. The mutants were, however, significantly less thermally stable than the wild type enzyme, as revealed through measurement of melting temperatures, monitored by CD. Consequently, kinetic investigations were performed at 25 °C rather than 37 °C.

pH Dependence

The pH dependence of kcat/Km for hydrolysis of PNPC by each enzyme species is shown in Figure 3.2. The wild type enzyme shows a bell-shaped dependence upon pH which fits well to a model involving two essential ionizing groups of pKa 5.0 and 7.3, consistent with the previously determined pH-dependence for hydrolysis of 2,4-DNPC. Since the pH-dependence of kcat/Km reflects ionizations in the free enzyme, these two groups are assigned as the nucleophile and acid/base catalyst respectively, as has been directly demonstrated for the mechanistically similar Bacillus circulans xylanase (McIntosh et al., 1996). Quite different behaviour was observed for the two mutants, a weak pH dependence being observed for His205Asn, while an apparent single, non-essential ionization was seen for His205Ala. Such large changes in ionization behavior are difficult to rationalize in detail, but are consistent with the intimate involvement of His205 in hydrogen bonding directly to the catalytic nucleophile, Glu233, and possibly to the acid/base catalyst, Glu127, which is only 3.3 Å away. Presumably upon removal of
His205, the pKa of Glu233 moves outside the range studied resulting in the removal of the 'acid' ionization of His205Ala. In any case, as a consequence of these different pH dependences subsequent kinetic studies were performed at both pH 7.0 (the optimum pH for wild type Cex and the pH at which previous studies had been performed (MacLeod et al., 1994; MacLeod et al., 1996; Tull & Withers, 1994) and pH 5.5 (the optimum pH for His205Ala).

Figure 3.2 Dependence of $k_{cat}/K_m$ upon pH for hydrolysis of PNPC by wild type Cex and mutants

**Kinetic analysis**

Kinetic parameters for the hydrolysis of three aryl glycoside substrates of different aglycone pKa (2,4-DNP, 3.96; 3,4-DNP, 5.96; PNP, 7.18) were determined at both pH 7.0 and 5.5 and these results are presented in Table 3.2. Values of $k_{cat}$ for the three
substrates with each individual mutant are essentially identical, suggesting that the second step (deglycosylation) which is common to each, is rate-determining. This interpretation is further borne out by the very low Km values observed for the substrates with good leaving groups. In such cases the glycosylation step is very fast compared to deglycosylation, with the result being significant accumulation of the glycosyl-enzyme intermediate, and a lowering of Km values (Fersht, 1985; Tull & Withers, 1994). Further evidence for rate-limiting deglycosylation was obtained through the observation of a full-sized burst of release of one equivalent of aglycone per equivalent of enzyme in a pre-steady state kinetic analysis (data not shown). Accumulation of the celllobiosyl-enzyme intermediate was also observed by means of electrospray mass spectrometry, the masses of each mutant incubated with 2,4-DNPC increasing by an amount (325 Da) that corresponds to the addition of a celllobiosyl moiety. Thus the measured mass of the His205Ala mutant in the presence of 2,4-DNPC was 47376 (± 4), while that of the His205Asn mutant was 47418 (± 5). The large differences in Km values and the observation of a full-sized burst also clearly indicate that the catalytic activity observed is a true activity of the mutant and does not arise from a small contaminant of wild type Cex in a sample of completely inactive mutant. Additional confirmation was obtained by monitoring the inactivation of the His205Asn mutant in the presence of the mechanism-based inactivator 2,4-dinitrophenyl-2-deoxy-2-fluoro-β-celllobioside. Complete inactivation was observed in each case according to pseudo-first-order kinetics, with apparent first order rate constants for inactivation of $k_{obs} = 0.0164 \text{ min}^{-1}$ (wild type) and $k_{obs} = 0.0013 \text{ min}^{-1}$ (His205Asn). The complete inactivation according to pseudo-first-order kinetics, and the ten-fold slower inactivation of the mutant are consistent only with a pure mutant of low activity. Inspection of $k_{cat}$ values in Table 3.2 reveals that the deglycosylation step is slowed some 4000-5000 fold as a consequence of the His205Ala mutation, and a slightly larger amount (8000-9000 fold) for the substitution with Asn. This indicates a very important
role for His205 in the deglycosylation step, likely involving stabilization of the carboxylate of Glu233 as it departs from the sugar anomeric centre. Similar rate reductions were observed upon mutation of Tyr298 to Phe in Agrobacterium sp. β-glucosidase (Gebler et al., 1995). This residue also hydrogen bonds to the catalytic nucleophile (Glu358) and an equivalent role in stabilizing the position of, and charge development on, the carboxylate was proposed.
Table 3.2 Michaelis-Menten parameters for the hydrolysis of aryl cellobiosides by Cex and mutants at pH 7.0 and 5.5

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>pH</th>
<th>Substrate</th>
<th>$k_{cat}$ (s⁻¹)</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat}/K_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.0</td>
<td>2,4-DNPC</td>
<td>3.9</td>
<td>0.12</td>
<td>32.1</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td></td>
<td>3.9</td>
<td>0.082</td>
<td>48</td>
</tr>
<tr>
<td>Cex-wt</td>
<td>7.0</td>
<td>3,4-DNPC</td>
<td>3.8</td>
<td>0.17</td>
<td>22.2</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td></td>
<td>5.9</td>
<td>0.19</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>PNPC</td>
<td>3.5</td>
<td>0.56</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td></td>
<td>4.1</td>
<td>0.70</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>2,4-DNPC</td>
<td>$8.2 \times 10^4$</td>
<td>$0.007 \pm 0.002$</td>
<td>$1.2 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td></td>
<td>$1.0 \times 10^{-3}$</td>
<td>$0.0007 \pm 0.0003$</td>
<td>$1.4 \pm 0.6$</td>
</tr>
<tr>
<td>H205A</td>
<td>7.0</td>
<td>3,4-DNPC</td>
<td>$8.6 \times 10^4$</td>
<td>0.032</td>
<td>$2.7 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td></td>
<td>$1.1 \times 10^{-3}$</td>
<td>0.023</td>
<td>$4.8 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>PNPC</td>
<td>$9.1 \times 10^4$</td>
<td>0.82</td>
<td>$1.1 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td></td>
<td>$1.2 \times 10^{-3}$</td>
<td>0.16</td>
<td>$7.5 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>2,4-DNPC</td>
<td>$4.9 \times 10^4$</td>
<td>0.03</td>
<td>$1.6 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td></td>
<td>$5.2 \times 10^4$</td>
<td>0.008</td>
<td>$7.4 \times 10^{-2}$</td>
</tr>
<tr>
<td>H205N</td>
<td>7.0</td>
<td>3,4-DNPC</td>
<td>$5.1 \times 10^4$</td>
<td>0.083</td>
<td>$6.1 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td></td>
<td>$6.6 \times 10^4$</td>
<td>0.024</td>
<td>$2.8 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>PNPC</td>
<td>$4.6 \times 10^4$</td>
<td>0.67</td>
<td>$6.9 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td></td>
<td>$4.7 \times 10^4$</td>
<td>0.19</td>
<td>$2.5 \times 10^{-3}$</td>
</tr>
</tbody>
</table>
The very significant effects of the mutations of His205 on the glycosylation step are reflected in the kcat/Km values which range from approximately 300-9000 fold lower than those measured for the wild type enzyme with the same substrates. Some insight into the transition state structure for this step is obtained by construction of Brønsted relationships correlating kcat/Km values measured at both pH 7.0 and pH5.5 with leaving group pKa values (not shown). The slopes of such plots (βp values) reflect the degree of negative charge development on the glycosidic oxygen at the transition state, and the following values were obtained: Cex-wt, -0.2; His205Ala, -0.6; His205Asn, -0.4. The larger βp values obtained upon removal of the imidazole side chain, particularly with the Ala mutant, indicate that His205 plays a significant role in stabilization of the negative charge on the phenolic oxygen at the transition state. This is either effected directly via a through-space field effect, or indirectly through correct placement of the acid catalyst via the intricate hydrogen bonding network in which His205 is intimately involved. Removal of this charge and/or disruption of the network would alter the charge distribution at the transition state.

The major consequence of mutating His205 could therefore be the incorrect placement of the acid/base catalyst Glu127 since the kinetic consequences observed here are very similar to those observed upon mutation of Glu127 (MacLeod et al., 1996). However, the results with a double mutant in which both Glu127 and His205 have been mutated argue strongly against this interpretation. If the sole role of His205 is correct placement of the acid/base catalyst, the double mutant should be no less active than either single mutant. On the other hand, if the roles are quite different the effects should be approximately additive. The latter was clearly found to be the case since addition of 2,4-DNPC to the His205Asn/Glu127Ala double mutant resulted in the formation of a very stable cellobiosyl-enzyme intermediate. This is indicated by the stoichiometric release of one equivalent of 2,4-DNP (data not shown) and by the expected increase in mass of the mutant from 47041 (± 4) to 47365 (± 4). This intermediate is extremely stable, no
significant steady state turnover of 2,4-DNPC being observed even upon lengthy incubation. These findings are consistent with the 8500 fold rate reduction expected for the His205Asn mutation coupled with the 300-fold reduction expected for the Glu127Ala change: a total of 2.5 \times 10^6 fold. It is therefore impossible to obtain meaningful steady state kinetic parameters with this mutant, but is nonetheless very clear that the effect of the two mutations is additive. These results also indicate the double mutant to be an excellent candidate for crystallization and structure determination of a stable cellobiosyl-enzyme intermediate.

Determination of the 3-dimensional structure of this intermediate complex first requires the removal of the cellulose binding domain by cleavage with papain since the intact enzyme has never been crystallized. Unfortunately papain digestion under the conditions used for the wild type enzyme resulted also in degradation of the catalytic domain, presumably because the double mutation substantially destabilized its structure. This obstacle was surmounted by first reacting the enzyme with 2,4-DNPC, thereby forming the cellobiosyl-enzyme intermediate which might be expected to be more stable than the free enzyme, by analogy with findings on the 2-deoxy-2-fluoroglucosyl-enzyme complex of Agrobacterium sp. \( \beta \)-glucosidase (Street et al., 1992). Indeed, papain digestion of this species yielded the intact cellobiosyl complex of the catalytic domain, as evidenced by polyacrylamide gel electrophoresis and mass spectrometry. The mass of the trapped catalytic domain was 35070. Upon incubation overnight with 0.5M sodium azide, the mass reduced to 34753, explained by the loss of the cellobiosyl moiety, whereas in the absence of azide, no mass change was observed. This is consistent with the azide activation of the Glu127Ala single mutant (MacLeod et al., 1994; MacLeod et al., 1996) which arises from reaction of the intermediate with azide to form cellobiosyl azide, thereby indicating that the domain is still catalytically competent. The trapped catalytic domain was purified by gel filtration and crystallized directly in that form.
Figure 3.3  

(a) Stereo representation of initial 2|Fo|−|Fc| density contoured at a map standard deviation level of 1.0, superimposed on the refined Cex-DM cellobiosyl structure. Continuous electron density between the substrate and the catalytic nucleophile suggests a covalent bond. The distance between 2′OH of the covalently linked proximal sugar and the carbonyl oxygen of the nucleophile is 2.37 Å.  

(b) Stereo diagram of a superposition of the Cex-DM cellobiosyl active site (white) and wild type Cex trapped as the 2-deoxy-2-fluoro-cellobiosyl-enzyme (White et al., 1996) (green). The nucleophile Glu233 has assumed a different orientation in the double mutant structure, resulting in a shift in the substrate position as well. Glu233 is still positioned optimally between residues Gln203 and Trp273, as seen in the wild type structure. Hydrogen bonds are shown in dotted lines. The conserved hydrogen bonding triad Glu233-His205-Asp235 does not form in the double mutant structure, but a new network is formed. This figure was generated with Setor (Evans, 1993).
Crystal Structure of Cex-cd E127A/H205N

Several structures of covalent glycosyl-enzyme complexes have been reported recently (Burmeister et al., 1997; Notenboom et al., 1998; White et al., 1996). However, in all these examples the structures reported have been 2-deoxy-2-fluoroglycosyl-enzyme species. In contrast, here we report the crystallographic observation of a covalent glycosyl-enzyme intermediate complex using a natural (unsubstituted) substrate. A cellobiose molecule is found covalently bound to the nucleophile Glu233, and its proximal sugar ring adopts an almost perfectly relaxed $^4C_1$ conformation (Figure 3.3a), the only perturbation being that the dihedral angle describing the proximal 3'OH-C3-C2-2'OH is found to be 57\(^\circ\) (67\(^\circ\) in the distal sugar). This is in good agreement with the geometry seen in other complexes (Burmeister et al., 1997; Notenboom et al., 1998), with the exception of the 2F-cellobiosyl complex with wild type Cex (White et al., 1996).

The orientation of the nucleophile Glu233 was found to be altered from that in the wild type enzyme. Difference electron density (IFol-IFcl maps, from which this and other surrounding residues were omitted in the density calculation) suggested a rotation of 180\(^\circ\) around Chi-1, essentially translating the carboxyl group of the residue by approximately 1\(\AA\), but keeping it in the same plane as in the wild type structure, and leaving Glu233 positioned between the residues Trp273 and Gln203 as seen in wild type Cex and in other family 10 xylanases (Dominguez et al., 1995; Harris et al., 1994; White et al., 1994). This reorientation of the nucleophile is presumably a result of the His205Asn mutation disrupting the highly conserved hydrogen bonding triad Glu233-His205-Asp235 in Family 10 glycoside hydrolases, which may be responsible for proper ionization and/or the orientation of the nucleophile. While other members of the 4/7 superfamily have a conserved tyrosine interacting with the nucleophilic Glu, the family 10 enzymes C. thermocellum Xylanase XynZ (Dominguez et al., 1995), P. fluorescens XylanaseA (Harris et al., 1994), and S. lividans Xylanase A (Derewenda et al., 1994) all form the same His-Asp interaction around the nucleophile. As the nucleophile Glu233 cannot form
a hydrogen bond with Asn205 in the Cex-DM crystal structure, a new hydrogen bonding triad (Gln203-Asn205-Asp235) is now formed in which Glu233 has been replaced by Gln203 (Figure 3.3b). In crystal structures of wild type Cex-cd Gln203 donates a hydrogen bond to the acid/base catalyst Glu127 through its NH group. Since Glu127 is mutated to Ala in this enzyme Gln203 is free to form a hydrogen bond with Asn205, leading to a less than ideal positioning of the nucleophile due to its exclusion from the hydrogen bonding triad. In addition, the pH profiles described earlier show that the His205Asn mutation alters the ionization behaviour of the nucleophile during catalysis, rendering it sub-optimal for nucleophilic attack in the glycosylation step. The covalent enzyme-substrate intermediate was able to accumulate however, because an excellent leaving group such as DNP allows the glycosylation of the enzyme to proceed.

Due to the repositioning of the nucleophile as described above, the bound disaccharide has been shifted approximately 1Å with respect to the two previously reported wild type fluoroglycosyl-Cex complexes (Notenboom et al., 1998; White et al., 1996). Despite the shift in positioning, the two glucose moieties maintain most of their hydrogen bonding interactions with the active site residues surrounding the -1 and -2 substrate binding sites. However, one key interaction seen in previous studies of Cex is missing. The proximal sugar C-2 hydroxyl group fails to form a hydrogen bond with Asn126, a strictly conserved residue in the 4/7 glycoside hydrolase superfamily (Dominguez et al., 1995). In wild type Cex Asn126 interacts relatively weakly with the proximal 2'F group of the substrate at a distance of 3.3Å, and forms a close hydrogen bond with Asn169 (2.9Å). Asn169, in turn, is found to form a hydrogen bond with the side-chain carbonyl of the nucleophile Glu233. Interestingly, in the Cex-DM complex the Asn126-Asn169 interaction is maintained at 2.9Å, whereas the 2'OH-Asn126 distance has increased to 4.4Å. Although Asn169 is conserved in Family 10 glycoside hydrolases, this residue is not conserved in other members of the 4/7 superfamily, in contrast to Asn126. Instead, in these enzymes an arginine residue (strictly conserved in all 4/7 families, except family
10) is oriented into the active site from a different part of the sequence (eg. Arg62 from Family 5 A. cellulolyticus EI (Sakon et al., 1996) and Arg95 from Family 1 S. alba myrosinase (Burmeister et al., 1997)) such that this residue forms similar interactions to those of Asn169 with the nucleophile and Asn126 (Figure 3.4a+b) in Cex.
Figure 3.4  Comparison of -1 subsite residues of a) wild type family 10 C. fimi Cex in covalent complex with the inhibitor 2-Fluoro-2-deoxy-xylobiose (Notenboom et al., 1998) and b) wild type family 5 A. cellulolyticus E1 in complex with cellotetraose spanning the active site (Sakon et al., 1996), shown in stereo representation (Evans, 1993). Asn169 from strand5 in Cex and Arg62 from strand2 in E1 each donate hydrogen bonds to both the nucleophilic Glu and to a conserved Asn (126 and 161, resp.) in both enzymes. This Asn in turn donates a weak hydrogen bond (3.3Å) to the substituent at the saccharide 2' position. See also Dominguez et al., 1995.
An unusually short distance of 2.37Å is measured between the C-2 hydroxyl group and the carbonyl oxygen of the nucleophilic carboxylate group of Glu233. Such a close interaction was not seen in the fluoro-substituted complexes, perhaps due to the electronegative nature of the substituting atom which renders it unable to donate a hydrogen bond to the side chain carbonyl of the nucleophile. This is a very short hydrogen bond, and hydrogen bonding distances of less than 2.5Å have been described previously as possibly being of the low-barrier type (Cleland & Kreevoy, 1994; Frey et al., 1994; Tobin et al., 1995). More specifically, low barrier hydrogen bonds between substrate and enzyme are postulated to form when the distance between the two hetero atoms sharing the hydrogen becomes shorter than 2.5Å, and their pKa values have reached similar magnitudes (e.g. during catalysis). Energies of such short hydrogen bonds have been estimated to lie between 10 to 20 kcal.mol$^{-1}$ and they are thought to play an important role in transition state stabilization during enzymatic catalysis (Cleland & Kreevoy, 1994; Gerlt et al., 1997). Considerable controversy has surrounded the explanation of these short, strong hydrogen bonding interactions (Guthrie, 1996; Shan & Herschlag, 1996a; Shan & Herschlag, 1996b; Warshel & Papazyan, 1996) but there is no doubt that very strong interactions can develop at single sites in the transition state (Wolfenden & Kati, 1991). Further, it is clear that very strong interactions in the range of 8 to 10 kcal.mol$^{-1}$ are present at the sugar 2-position in the transition states of a number of retaining β-glycosidases, including Cex (McCarter et al., 1992; Namchuk & Withers, 1995; White et al., 1996; Wolfenden & Kati, 1991). While some of this stabilizing interaction at the 2-position may result from hydrogen bonding to Asn126, the fact that this hydrogen bond disappears in the complex with the double mutant suggests that it may not be the major contributor. It seems more probable that the major source of this stabilization is a strong hydrogen bond with the side chain carbonyl oxygen of the nucleophilic carboxylate. The interaction observed within this structure is not, of course, that present at the transition state. Further, it may be modified somewhat as a
consequence of the mutations necessary to stabilize the structure. However, the similarity of the covalent intermediate structures, stabilized in quite different ways, and also the fact that Glu233 must form a covalent bond at the adjacent, anomeric centre would indicate that this important interaction must be present in the natural enzyme/substrate system. It seems probable that this interaction becomes optimized geometrically as the transition state is reached both by virtue of the flattening of the sugar ring and through the approaching carboxylate as the covalent bond is formed, as shown in Figure 3.5. In addition this hydrogen bond will be optimized electronically at the transition state since the development of positive charge on the sugar ring at C-1 and O-5 must transiently increase the acidity of the 2-hydroxyl. In these ways the interaction at the 2-position can provide enormous transition state stabilization. This mechanism of transition state stabilization may prove to be widespread in glycosidase catalysis. However it cannot be a universal phenomenon since neither α-glucosidases nor β-mannosidases can achieve the correct geometry for this interaction.
Figure 3.5  The proposed role of hydrogen bonding at OH-2 (-1 subsite) showing the Glu233 carboxylate interaction with the saccharide OH-2 in transition state stabilization.


Chapter 4

Kinetic and Crystallographic Analysis of the Cex Active Site Mutants
Asn126Ala and Asn169Ala
Abstract
The β-1,4-glycanase Cex from *Cellulomonas fimi* belongs to family 10 of glycosyl hydrolases, which comprises mainly endo-xylanases with a variable degree of activity against cellulose amongst its members. The roles of two conserved residues, Asn126 and Asn169 in the active site of Cex were probed by kinetic and crystallographic analysis of their respective alanine mutants. The mutant proteins were assayed for activity against p-nitrophenyl (PNPC) and 2", 4"-dinitrophenyl cellobiose (DNPC). Elimination of the Asn126 functionality resulted in a mutant that displayed reduced activity particularly against the compound with the poorer leaving group, PNPC, where $k_{cat}$ was reduced 200 fold. An additional consequence of this mutation was the dramatic increase of $K_m$ against PNPC but not DNPC, which led to the conclusion that the rate-limiting step of the reaction had been shifted from deglycosylation to glycosylation. The Asn169Ala mutation bore minimal effect on catalytic activity, as well as on $K_m$ values against both substrates tested. The crystal structures of both N126A and N169A were determined at 2.1Å and 2.0Å resolution respectively. In the case of N169A a water molecule was found in place of the void generated by the mutation. No significant structural reorganization had taken place in the active site of N169A. In N126A Trp84 rotates inward into the binding pocket and possibly obstructs binding of incoming substrates somewhat, although this effect is not expected to be significant due to the dependence of $k_{cat}/K_m$ on leaving group ability of the substrate. The implications of these results on reaction mechanics are discussed in this chapter.
Introduction

The β-1,4-glycanase Cex from Cellulomonas fimi belongs to family 10 of glycosyl hydrolases, which comprises mainly endo-xylanases with a variable degree of activity against cellulose amongst its members (Gilkes et al., 1991; MacLeod et al., 1994). Family 10 enzymes fall into the 4/7 superfamily of glycosyl hydrolases, members of which are related by sequence conservation and structural similarity. All 4/7 superfamily enzymes have the (α/β)_8 fold, and the two catalytic carboxylic acids lie on β-strands 4 and 7 (Jenkins et al., 1995). A large number of active site residues are strictly conserved within this superfamily, suggesting important roles in catalysis. Figure 4.1 depicts an active site cartoon of Cex, in which most of the residues that are strictly conserved are shown in the context of the covalent 2F-xylobiose structure (Notenboom et al., 1998b).

![Active site cartoon of Cex in complex with 2F-xylobiose, showing hydrogen bonding interactions of conserved residues with the substrate.](image-url)
Of interest are the interactions surrounding the nucleophile Glu233 and the 2'OH group of the proximal saccharide. As discussed in the previous chapter, interactions at the 2-position are extremely important to transition state stabilization in a number of glycosidases, contributing at least 40 kJ mol⁻¹ in the case of Cex and similar amounts in a range of other glycosidases (Namchuk & Withers, 1995). Unfortunately, the 2-position is substituted by fluorine in the trapped complex, which leaves interpretation of the interactions involved at this position ambiguous. In the 2F-xylobiose structure Asn126 comes within hydrogen bonding distance of the fluorine atom. However, in the double mutant H205N/E127A structure with accumulated covalent cellobiose (Chapter 3 and (Notenboom et al., 1998a)) this interaction is lost and a short hydrogen bond is found between Oe2 of the nucleophile Glu233 and 2'OH. Glu233 and Asn126 in turn accept hydrogen bonds from Asn169, which may play a role in positioning of these side chains. It appears that an intricate and dynamic network of interactions surrounds the 2' position during catalysis, possibly providing transition state stabilizing forces and allowing for specific substrate recognition.

In order to gain further insight into the roles of the residues surrounding the 2' position, the active site mutants of Cex, Asn126Ala and Asn169Ala, were generated and analyzed kinetically.

Materials and Methods

Materials.

Restriction endonucleases were from New England Biolabs (Mississauga, ON, Canada). T4-DNA ligase was from Gibco-BRL (Burlington, ON, Canada). PWO DNA polymerase and the Expand™ High fidelity PCR system were from Boehringer Mannheim (Laval, PQ, Canada). Isopropyl β-D-thiogalactoside (IPTG) and CF-1 cellulose were from Sigma (St. Louis, MO, USA).

Bacterial strains and plasmids.
*Escherichia coli* strain BL21 was used as the host strain for most genetic manipulations, Novo Blue cells (Stratagene, La Jolla, CA) for cloning and JM101 for protein expression (Sambrook *et al.*, 1989). The DNA plasmid used was pUC12-1.1_ex(PT1S) (O'Neill *et al.*, 1986).

**Media and growth conditions.**

*E. coli* cultures were grown in TYP medium, containing the following (per litre): 16 g tryptone, 16 g yeast extract, 5 g NaCl, 2.5 g K₂HPO₄. Plasmid containing *E. coli* cultures were grown in TYP medium supplemented with Kanamycin at 50 μg/ml at 30° in 6L shake flasks at 225 rpm for production of protein. Bacterial growth was monitored spectrophotometrically at A600nm.

**DNA manipulations.**

Ligations and restriction endonuclease digestions were performed as recommended by manufacturers and suppliers. DNA fragments were purified from agarose gels with Qiaquick Spin PCR purification columns (Qiagen, Valencia, CA, USA). Oligonucleotide primers were synthesized by the UBC Nucleic Acid and Protein Service (NAPS) with an Applied Biosystems Model 380A DNA synthesizer. DNA was sequenced by the NAPS unit, using AmpliTaq dye termination cycle sequencing protocol and an Applied Biosystems Model 377 sequencer. Plasmid DNA was extracted with Qiagen Miniprep system (Qiagen, Valencia, CA, USA). PCR was run using a PTC-100 Programmable Thermal Controller (MJ Research Inc., Watertown, MA, USA). 100 μl PCR reactions contained 50 ng plasmid template, 100 pmols of each primer, 1μl PWO polymerase, 400μM dNTPs and 5% DMSO in 1× PWO buffer (10 mM Tris-HCl, pH 8.85, 25 mM KCl, 5mM (NH₄)₂SO₄ and 2 mM MgSO₄). PCR cycling was set up as follows: 1. 60 s, 95°C, 2. 30 s, 95°C, 3. 30 s, 65°C, 4. 60 s, 72°C, 5. repeat from step 2 for 25 cycles with the temperature decreasing by 0.5° with each cycle, 6. 30 s, 95°C, 7. 30 s, 55°C, 8. 60 s, 72°C, 9. repeat from step 6 for 25 cycles, 10. 120 s, 72°C, 11. 4°C.

**Site-directed mutagenesis.**
Site-directed mutagenesis was performed by PCR using the 2 primer (N126A) or the mega-primer (N169A) extension method, generating the Cex mutants N126A and N169A, as outlined in Fig4.2 a+b. The expression plasmid pUC12-1.1cex(PTIS) was used as the DNA template (Gilkes et al., 1991; MacLeod et al., 1994). The following primers were synthesized, mutations are depicted enlarged in bold and restriction sites underlined and specified in brackets.

**n126a**

5' GTC CTG CCG GCC GCC GCC GTC GGC GAA **AGC TTC**

GGC GAC GAC GTC 3' (Hind III)

**n169a**

5' CTC GAC GTT GTA GTC **GGC** GAT GCA **AAG CTT** GGC CGT

3' (Hind III)

**BamH1**

5' CTC GCC CGG **GGA TCC** TAG GA 3' (Bam HI)

**Apa1**

5' CCG GTG CCC **GGG CCC** GTC GG 3' (Apa1)

The mutation N126A was generated by extending the fragment flanked by the primers n126a and BamH1 on the pUC12-1.1cex(PTIS) template. The PCR product was isolated and purified from agarose gel and digested with the restriction endonucleases Not I and BamH I and was ligated into pUC12-1.1cex(PTIS) with the Not I/BamH I region excised. The N169 site on pUC12-1.1cex(PTIS) was not flanked by appropriate unique restriction sites and hence a mega primer containing the desired mutation was generated by PCR amplifying the DNA flanked by the primers n169a and BamH1. This fragment was purified from a 1.2 % agarose gel. All of the product was then used in a subsequent PCR reaction as a primer with Apa1 on the pUC12-1.1cex(PTIS) template. The resulting ~1500bp were visualized and purified from a 1.0 % agarose gel. The DNA containing the N169A mutation was digested with BamH1 and BssHII. This product was then ligated
into the appropriately digested plasmid as described above. The ligation mixtures were cloned into competent NovoBlue cells using the heat shock transformation protocol as recommended by the suppliers. Both mutations were designed to add a second Hind III restriction site to the plasmid by silent mutation. This allowed for transformants to be screened by restriction site analysis of purified double stranded plasmid DNA using Hind III restriction endonuclease, revealing a twice cut plasmid for each correct mutation. The mutations were confirmed by DNA sequencing.
Figure 4.2a+b (opposite page) Introduction of the Ala mutations into the pUC12-1.1cex plasmid using PCR.
1. Extend fragments
2. Isolate product
3. Combine fragment, primer and template
4. Extend fragment
5. Isolate product
6. Digest with Not I and BssH II
7. Ligate into pUC-12-1.1cex
Protein expression and purification.

The mutant and wild type plasmids were transformed into *E. coli* JM101 cells by electroporation. 2L cultures were allowed to grow in Kan supplemented TYP medium to OD$_{600nm}$ of ~1.5 at which point IPTG was added to induce protein expression. Cells were harvested 3-5 hours after induction by centrifugation at 5,000 x g for 15 minutes. The cell pellets were resuspended in 40 ml of 50 mM phosphate buffer, pH 7.0. From the resuspension, the equivalent of 1 ml of cells was boiled in SDS loading buffer and analyzed by SDS-PAGE for protein expression levels. Cell lysis was accomplished by passing through a French™ pressure cell three times. PMSF and EDTA were added to 1 mM and the ruptured cell fragments removed by centrifugation at 15,000 x g for 45 min at 4 °C. The remaining cell extract was bound to 50 g CF-1 cellulose swollen in 1 L phosphate buffer and rocked overnight at 4 °C. The cellulose was poured into a glass column (l=1 m, d=5 cm) and washed with 2 L phosphate buffer in 1 M NaCl by gravity flow. The column was then attached to a GradiFrac™ fraction collector (Pharmacia, Uppsala, Sweden) and allowed to equilibrate with phosphate buffer at a flow rate of 3 ml/min. Bound protein was eluted with distilled water at the same flow rate, monitoring OD$_{280nm}$ of the eluate at all times. Peak fractions were pooled, filtered through 0.22 µM pores and concentrated to ~2 mg/mL using Amicon PM10 membranes. All protein samples were analyzed for purity on SDS-PAGE before further experiments. For preparation of the crystallizable (catalytic) fragment of Cex, typically 20 ml of each of the full length protein samples (2 mg/mL) was digested with 400 µg immobilized papain overnight at 37 °C in 100mM Tris-HCl, pH 7.5, 2 mM EDTA, 0.02% azide and 6µl β-mercaptoethanol. After completion of digestion the samples were dialyzed against 10 mM Tris-HCl, pH 7.5 and loaded onto a HighTrap Q column (Pharmacia) attached to a BioRad fraction collector. The column was washed at 1 mL/min with loading buffer for 10 column volumes allowing the cleaved 12 kD CBD$_{Cex}$ domain to pass through without binding. The remaining catalytic fragment was eluted at 1 mL/min by increasing
concentration gradient to 300 mM NaCl in Tris-HCl buffer, pH 7.5. Peak fractions were pooled and concentrated to 45 mg/mL in 10 mM Tris-HCl, pH 7.5.

Crystallization, data collection and structure determination of Cex N126A and N169A.

Tetragonal crystals (P4₁₂₂) were grown from the N126A and N169A mutant protein samples by the hanging drop vapour diffusion method from 20% PEG4000 and 0.1 M Na acetate, pH 4.6, as described previously (Bedarkar et al., 1992). Crystals were sequentially transferred into cryo protectant solutions of artificial mother liquor supplemented with increasing amounts of glycerol, to a final glycerol concentration of 30%. Crystals were then scooped into a rayon fiber loop and placed in a stream of N₂ gas at 100K. In order to reduce mosaic spread the crystals, once frozen, were brought back to room temperature for ~20 seconds and then quickly refrozen to allow for re-annealing.

All data were collected in-house on a Mar345 image plate using Osmic mirror focussed Cu-Kα X-rays, generated from a rotating anode operating at 100mA × 50kV. Oscillations of 1° were collected in 60-120s exposures, depending on the size of the crystal. All data were indexed, integrated and scaled in the Denzo/Scalepack program suite (Otwinowski & Minor, 1997). Data collection statistics are listed in Table 4.1

<table>
<thead>
<tr>
<th></th>
<th>N126A</th>
<th>N169A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Resolution</td>
<td>2.1 Å</td>
<td>2.0 Å</td>
</tr>
<tr>
<td>Number of observations</td>
<td>261228</td>
<td>247731</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>17833</td>
<td>17870</td>
</tr>
<tr>
<td>Completeness</td>
<td>96.0 %</td>
<td>93.6 %</td>
</tr>
<tr>
<td>I/σ</td>
<td>11.1</td>
<td>9.6</td>
</tr>
<tr>
<td>R_merge</td>
<td>6.7 %</td>
<td>7.5 %</td>
</tr>
</tbody>
</table>

\[
R_{\text{merge}} = \frac{\sum_{hkl} \sum_{i} |I_{hkl} - \overline{I}_{hkl}|}{\sum_{hkl} \sum_{i} I_{hkl}}
\]
At the start of refinement 10% of reflections were set aside for cross-validation purposes (Kleywegt & Brünger, 1996). Upon freezing the unit cell axes decrease in length and the Cex crystals become essentially non-isomorphous with those mounted at room temperature. The exact position of the molecule in the asymmetric unit was easily found by rotation and translation searches in X-Plor (Brünger, 1996) using the wt-Cex coordinates (PDB code 2EXO) with the appropriate coordinate deletions reflecting the mutations. All subsequent refinement was performed in the CNS program package (pre-release version 0.5) (Brünger et al., 1998). After low-resolution anisotropic bulk-solvent corrections the molecular replacement model was subjected to a simulated annealing routine starting at 3000K and a stepwise decrease of 25K, with maximum likelihood refinement methods using structure factor amplitudes. From this model 2IFol-IFcl and lFol-lFcl electron density maps were calculated and examined at σ levels of 1.0 and 3.0 respectively. At this point several rounds of water building were performed. Finally, individual B-factors were refined. Final structure quality statistics are summarized in Table 4.2.
Table 4.2 Structure refinement and model quality statistics

<table>
<thead>
<tr>
<th></th>
<th>N126A</th>
<th>N169A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell (a=b, c in Å)</td>
<td>85.3, 79.1</td>
<td>85.6, 79.3</td>
</tr>
<tr>
<td>R-factor, all data (%)</td>
<td>21.6</td>
<td>20.9</td>
</tr>
<tr>
<td>R-free, all data (%)</td>
<td>24.5</td>
<td>24.1</td>
</tr>
<tr>
<td>Rms deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.006</td>
<td>0.007</td>
</tr>
<tr>
<td>Bond angles (Å)</td>
<td>1.34</td>
<td>1.41</td>
</tr>
</tbody>
</table>

\[
R_{\text{cryst}} = \frac{\sum_{hkl} |F_{\text{obs}}| - k |F_{\text{calc}}|}{\sum_{hkl} |F_{\text{obs}}|}, R_{\text{free}} = \frac{\sum_{hkl \notin T} |F_{\text{obs}}| - k |F_{\text{calc}}|}{\sum_{hkl \notin T} |F_{\text{obs}}|}
\]

Kinetic analysis of mutants.

Michaelis-Menten parameters for N126A and N169A were determined for DNPC and PNPC using a continuous assay, recording absorbency changes at \( \lambda = 400 \) nm with a UV/Vis Pye-Unicam 8700 spectrophotometer, operating at 37 °C. Substrate solutions in 50 mM phosphate buffer, pH 7.0 and 1 mg/mL BSA were pre-incubated at 37 °C, and the reaction initiated by injection of enzyme. Accurate values were obtained for \( V_{\text{max}} \) using a range of substrate concentrations between 0.2 and 5 times the estimated \( K_{\text{m}} \) values. Non-linear regression analysis was used to determine exact values for \( V_{\text{max}} \) and \( K_{\text{m}} \) within the program GraFit (Leatherbarrow, 1990).
Results and Discussion

Production of mutant protein.

Mutations were introduced into the pUC12-1.1cex (PTIS) plasmid as described above. Screening for transformants containing the desired mutation was performed by digestion of the extracted plasmid with Hind III. In the case of N126A two bands were observed upon agarose gel separation of 1.7 kb and 2.8 kb, the N169A plasmid also digested into two bands of similar size, as expected. Sequence analysis confirmed the mutations. The mutants were expressed and purified as described. Both mutants behaved essentially identical to wild type protein during expression and purification procedures. Since purification is largely based on functional affinity chromatography by specific binding to cellulose these results suggested that the mutations had not introduced a significant alteration in the fold of the protein. Expression levels of the mutants were comparable to that of wild type protein (Figure 4.3), yielding on average 50 mg of purified protein from a 2 L culture grown in TYP medium using the above described protocol.

![Figure 4.3 SDS-PAGE separation of 47.1 kD wild type Cex and mutants. After 3 hours of induction with IPTG the equivalent of 1 ml of cells were boiled in SDS loading buffer. A molecular weight marker is loaded in the left lane. Cracked appearance due to inappropriate drying conditions.](image-url)
Kinetic analysis of mutants.

Values for $K_m$, $k_{cat}$ and $k_{cat}/K_m$ were determined for two aryl-cellobiosides 4"-nitrophenyl-β-D-cellobioside (PNPC) and 2",4"-dinitrophenyl-β-D-cellobioside (DNPC) with both mutants.

![Chemical structures of PNPC and DNPC](image)

**Figure 4.4** The β-glycosides DNPC (top) and PNPC (bottom)

The aglycone leaving groups of these substrates have markedly different pKa values (7.18 and 3.96, respectively). As a result in the first (glycosylation) step of hydrolysis little acid assistance is required for leaving group release of DNPC whereas PNPC is aided by accepting a proton from the acid/base catalyst. If the generated mutations affect the protonation machinery of the enzyme, these compounds should best reveal such changes. Steady-state kinetics were performed at pH 7.0 at 37°C in the presence of 1 mg/mL BSA, the results are listed below in Table 4.3.
## Table 4.3 Michaelis Menten parameters for wtCex, N126A and N169A

<table>
<thead>
<tr>
<th></th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat}/K_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt Cex</td>
<td>DNPC 6.98</td>
<td>0.060</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>PNPC 11.2</td>
<td>0.530</td>
<td>21.3</td>
</tr>
<tr>
<td>N126A</td>
<td>DNPC 0.539</td>
<td>0.040</td>
<td>13.5</td>
</tr>
<tr>
<td></td>
<td>PNPC 0.052</td>
<td>11.75</td>
<td>0.0044</td>
</tr>
<tr>
<td>N169A</td>
<td>DNPC 1.79</td>
<td>0.040</td>
<td>44.9</td>
</tr>
<tr>
<td></td>
<td>PNPC 3.87</td>
<td>0.592</td>
<td>6.54</td>
</tr>
</tbody>
</table>
Cex hydrolyzes its substrates in a two-step mechanism involving a covalent glycosyl-
enzyme intermediate:

\[
\begin{align*}
E+S & \xrightleftharpoons[k_1]{k_{-1}} E.S \\
& \xrightarrow{k_2} E-S \\
& \xrightarrow{k_3} E+P
\end{align*}
\]

**Figure 4.5** Reaction mechanism of Cex, expressed in rate constants

From the mechanism, \( K_m \) and \( k_{cat} \) can be expressed in terms of individual rate constants:

\[
\begin{align*}
k_{cat} &= \frac{k_2k_3}{k_2 + k_3} \\
K_m &= \frac{k_3(k_{-1} + k_2)}{k_1(k_2 + k_3)} \\
k_{cat/K_m} &= \frac{k_1k_2}{k_{-1} + k_2}
\end{align*}
\]

When the deglycosylation (\( k_3 \)) step is rate-limiting to catalysis, such as in wild type Cex, \( k_3 \ll k_2 \) and the expression for \( K_m \) reduces to \( k_3(k_{-1} + k_2)/k_2k_1 \), resulting in accumulation of the glycosyl-enzyme intermediate. This will cause the enzyme to become saturated with substrate at lower concentrations than the maximum \( k_2 \) rate, and an apparent low \( K_m \) is observed. Alternatively, in the event where the glycosylation step (\( k_2 \)) becomes rate limiting, the expression for \( K_m \) simplifies to \( K_m = k_{-1}/k_1 \), since \( k_2 \ll k_1 \). This value can effectively be interpreted as the dissociation constant \( K_4 \) of the substrate. In wild type Cex the \( k_{cat} \) values are not dependent on leaving group pKa, indicating that
glycosylation proceeds much faster than deglycosylation, the rate-limiting step. The N169A mutation results in reduced catalytic efficiency, but does not appear to alter the rate-limiting step of the reaction. The reduced catalytic rates can be the result of the loss of hydrogen bonding interactions of Asn169, which interacts with the Oe2 carbonyl oxygen of the nucleophile and the carbonyl oxygen of Asn126 in wild type enzyme. By contrast, the above data clearly shows different behavior in the N126A mutant. An approximately 10-fold decrease in $k_{cat}$ of PNPC is observed compared to DNPC, as well as a 300-fold increase in $K_m$. The dependence on the aglycone leaving group on $k_{cat}$ as well as the drastic increase in $K_m$ for PNPC suggests that in the N126A mutant the rate-limiting step has become $k_2$, or glycosylation. The observed effect of this mutation can arise in several ways. In wild type Cex, the Asn126 residue is found to form a hydrogen bond with the 2F atom of the proximal sugar in the covalent 2F-xylobiose and 2F-cellobiose intermediate structures (Notenboom et al., 1998b; White et al., 1996). This residue is thus assumed to constitute an important hydrogen bonding interaction during hydrolysis. It is possible that the absence of this interaction has resulted in poorer binding or recognition of the substrate to form the Michaelis complex. Alternatively, since Asn126 is in close proximity, and connected through a hydrogen bonding network via Gln203 to the acid/base catalyst Glu127, it is plausible that the mutation to alanine has resulted in a charge disruption affecting Glu127, which could greatly compromise its protonation efficiency.

Interestingly, the N126A mutant of the family 10 xylanase XYLA from *Pseudomonas fluorescens* resulted in an apparently contrasting kinetic profile (Charnock et al., 1997). Modification of Asn-126 to alanine caused a significant decrease in the catalytic activity of XYLA, and in the case of DNPC, in a large decrease in the $K_m$. However, in native XYLA the rate-limiting step in catalysis is glycosylation, whereas in wtCex deglycosylation is rate-limiting. These results are thus consistent with the conserved residue Asn126 affecting the catalytic machinery in such a way that the acid/base catalyst
does not perform optimally. In Cex, glycosylation may become rate-limiting upon the N126A mutation because Glu127 cannot provide proper acid assistance to the leaving group, which is especially apparent in PNPC hydrolysis. Similarly, in Xyla, DNPC hydrolysis does not require a proton to its leaving group, but, apparently, base catalysis through the abstraction of a proton from water cannot be achieved, and the deglycosylation step becomes rate-limiting. In order to gain further insight into the effects of these mutations, the structures of both mutants were determined by X-ray crystallography.

*Crystal structures of Cex N126A and N169A.*

The structures of CexN126A and CexN169A were determined as described in the materials and methods section. No significant changes in the overall fold of the protein were observed in either structure. At the site of Ala169 in the N169A mutant a water molecule is found occupying the void created by the mutation (Fig 4.6a). This water molecule makes a contact with the (non-nucleophilic) Oe2 oxygen of Glu233 as seen in the asparagine residue in wild type Cex. Presumably this new interaction could substitute to a certain degree the one lost through the mutation, thereby obscuring the full effects of losing the Asn169 functionality. However, if the function of Asn169 is to modulate the pKa of its neighbours, then of course a water molecule cannot completely substitute for the loss of an acidic residue.

As a result of the mutation Trp84, which lies 'behind' Asn126, collapses partially into the generated space by a rotation of $-30^\circ$ around Chi-1 (Fig 4.6b). As a result, an incoming substrate might be partially obstructed with this active site arrangement. Indeed, with Trp84 in the position as found in the mutant enzyme in unbound form, a contact of less than 2 Å would be formed with a substrate (as modeled from wild-type covalent and inhibitor structures reported in this manuscript). This may explain in part the observed rate decrease of this mutant. However, the Km of DNPC hydrolysis by N126A was not altered, but was found significantly increased with PNPC. This suggests that the
potential binding obstruction caused by the Trp84 movement cannot be solely responsible for the observed rate changes and increase in Km. It is more likely that the effects of the mutation are of electrostatic nature, including a possible pKa change of the neighbouring acid/base catalyst Glu127 and the loss of an important hydrogen bonding interaction of Asn126 with 2’OH, which may result in improper positioning of the substrate upon binding. Figures 4.6 a and b show the active sites of the mutants as revealed by their crystal structures. The structure of 2F-xylobiose in Cex is superimposed in both cases for comparison.

Figure 4.6 a (top) and b (bottom) Stereo representations of a) CexN126A active site in green and b) CexN169A active site in yellow, superimposed on wtCex-2Fxylolbiose (white), for comparison.
In summary, it appears that the role of Asn126 is at least bifunctional in catalysis. It is clearly involved in modulating the acid/base catalyst activity as illustrated by results above, but the interaction between the amino group and the 2'OH group in the proximal sugar during substrate binding, transition state and/or intermediate structure stabilization cannot be discounted. The role of Asn126 in substrate binding through 2'OH will be further discussed in the final chapter of this thesis (Chapter 6), in which inhibitor studies in wild type and Cex-Asn126Ala shed light on this interaction.
References


Jenkins, J., Lo Leggio, L., Harris, G. & Pickersgill, R. (1995). \(\beta\)-Glucosidase, \(\beta\)-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold \(\beta/\alpha\) architecture and with two
conserved glutamates near the carboxy-terminal ends of β-strands four and seven.


Chapter 5

Structural Studies of Xylobiose-Derived Aza-Sugars Complexed with Cex from *Cellulomonas fimi* Provide Detailed Insight into the Basis of Inhibition.

This chapter has been reformatted from original publication as submitted to the journal *Biochemistry (ACS publications)*: Valerie Notenboom, Spencer J. Williams, Roland Hoos, Stephen G. Withers, and David R. Rose (2000) Structural studies of xylobiose-derived aza-sugars with Cex from *Cellulomonas fimi* provide detailed insight into the mechanics of inhibition.
Abstract
The family 10 xylanase Cex from *Cellulomonas fimic* was used as a crystallographic template to study the nature of inhibition by four aza-sugars that were designed to probe different aspects of transition state structure. The crystal structures were determined to better than 2Å resolution of Cex in complex with xylobiose-derived isofagomine, deoxynojirimycin, lactam oxime and imidazole. These compounds were found to provide excellent inhibition against Cex, with $K_i$ values ranging from 4800 to 150 nM. By contrast, these compounds did not inhibit the family 11 xylanase Bcx from *Bacillus circulans* to a considerable extent. Although Cex and Bcx both catalyze the configuration retaining hydrolysis of xylan, their reaction mechanisms may differ significantly in transition state and intermediate structure, as suggested by their respective covalent 2-deoxy-2-fluoroxylobiosyl intermediate structures [Notenboom, V. *et al.* (1996) *Biochemistry* 37, 4751-4758; Sidhu, G. *et al.* (1999) *Biochemistry* 38, 5346-5354] The crystallographic observations obtained from the Cex inhibitor complexes are used to explain the relatively poor inhibition of these compounds against Bcx.
Introduction

Considerable efforts have been expended in recent years in the design, synthesis and testing of glycosidase inhibitors, with the hope of not only learning more about the active site structures and mechanisms of these interesting enzymes, but also of possibly generating new therapeutic agents. Indeed, some successes have been achieved in that regard with the development of specific inhibitors of viral neuraminidases (Taylor, 1996) and of intestinal \( \alpha \)-glucosidases (Martin & Montgomery, 1996) that are marketed now as treatments for influenza and non insulin dependent diabetes mellitus, respectively. Much of the effort towards the synthesis of reversible, non-covalent inhibitors has gone into the synthesis of nitrogen-containing sugar analogues, of which there are several general classes. Designs for these have been inspired both by the structures of naturally occurring inhibitors and by mechanistic considerations, as has been summarized in several reviews (Berecibar et al., 1999; Bols, 1998; Ganem, 1996; Heightman & Vasella, 1999; Hughes & Rudge, 1994; Legler, 1990; Stütz, 1999).

Glycosidases fall into two general mechanistic classes, depending on whether the glycosidic bond is hydrolyzed with net inversion or net retention of anomeric configuration. Both enzyme classes have active sites containing a pair of carboxylic acids. In the case of the inverting glycosidases the mechanism involves a direct displacement at the anomeric center by a base activated water molecule. Reaction proceeds via an oxocarbenium ion-like transition state, with the two carboxylic acids suitably positioned approximately 10 Å apart, to function as general acid and general base catalysts. Retaining glycosidases use a double-displacement mechanism in which a covalent glycosyl-enzyme intermediate is formed and hydrolyzed via oxocarbenium ion-like transition states. The two carboxylic acid side chains are closer together in these enzymes such that one functions as a general acid catalyst, while the other functions as a nucleophile, directly attacking at the anomeric center to form a covalent acylal intermediate with expulsion of the aglycone leaving group. In a second step water attacks
at the anomeric center with general base catalytic assistance from the same carboxyl group that originally donated a proton, thereby releasing the product sugar with net retention of anomeric configuration (Davies & Henrissat, 1995; Davies et al., 1998b; Ly & Withers, 1999; Sinnott, 1990; Zechel & Withers, 2000).

Sequence information is available on several thousand glycosidases, and based on sequence similarities these have been classified into at least 76 families (Henrissat & Bairoch, 1996). Three-dimensional structural information is now also available for large numbers of glycosidases, with members of some 30 families having been subjected to such analysis at the time of writing. These structures, along with conjectures on the basis of sequence similarities, have revealed that some of the sequence-related families can be grouped into clans on the basis of similarities in three-dimensional folds and locations of the key catalytic carboxylic acids (Davies & Henrissat, 1995). Structural information is also available on several glycosyl-enzyme intermediates that have been trapped through the use of fluorosugar analogues. This has provided valuable insights into the identities, locations and potential roles of a number of other essential active site residues (Davies et al., 1998a; Notenboom et al., 1998a; Notenboom et al., 1998b; Uitdehaag et al., 1999; White et al., 1996).

A widely studied class of inhibitors has been that of the 5-amino-5-deoxy-sugar analogues, epitomized by the naturally occurring D-gluco analogues nojirimycin and L-deoxynojirimycin (Stütz, 1999). These are proposed to bind to the enzyme active site in a protonated form such that the positively charged nitrogen interacts favorably with the negatively charged carboxyl groups at the active site. They generally function as better inhibitors of α-glycosidases than of β-glycosidases and have frequently been suggested to mimic either the protonated form of the substrate or the transition state itself, hence their tight binding. A closely related class of inhibitors, the isofagomines, have an endocyclic nitrogen located at the position corresponding to the anomeric center, thereby possibly mimicking the charge generated at the anomeric carbon in the transition state (Bols,
Interestingly, these prove to have the complementary inhibitory profile to that of the nojirimycins, inhibiting \( \beta \)-glycosidases much better than \( \alpha \)-glycosidases, leading to suggestions of inherent differences in transition state structure between the two enzyme classes (Bols, 1998; Zechel & Withers, 2000).

Another somewhat different class of nitrogen containing inhibitors is that in which the anomeric center is \( sp^2 \)-hybridised, thereby providing better stereochemical resemblance to the transition state. The progenitor of this class of inhibitors is gluconolactone, first discovered as an inhibitor of \( \beta \)-glucosidases by Ezaki (Ezaki, 1940). However, these glyconolactones have not proved to be generally useful due to their instability. Nitrogen-containing versions of this class of compounds include gluconolactam and derivatives in which a nitrogen is additionally incorporated at the 'glycosidic' position, such as the amidines, amidrazones, the lactam oximes and the bicyclic tetrazoles and imidazoles (Ganem, 1996; Heightman & Vasella, 1999). Such compounds have a better claim to being transition state analogues than the nojirimycin type of compounds since they more closely reflect the partially planar nature of the sugar ring and the build-up of positive charge on C1 and the ring oxygen. Further, the presence of an exocyclic heteroatom at C1 should better mimic the departing aglycone and provide a site for interaction of the active site acid catalyst. Such considerations, however, depend upon the acid catalyst being located appropriately in the active site for interaction with the available lone pair of the exocyclic heteroatom. This has led to the further classification of glycosidases as either anti- or syn protonators, depending upon the trajectory of the proton transfer event relative to the C1-O5 bond (Heightman & Vasella, 1999). Since many members of this inhibitor class, for example the imidazoles, do not contain a syn lone pair then it might be expected that these inhibitors would be specific for anti protonators. However this hypothesis has only been tested in one case, that being one in which other factors confused the interpretation (Vonhoff et al., 1999).
Despite the considerable efforts put into the design, synthesis and testing of glycosidase inhibitors, there have been remarkably few crystal structures determined of complexes of inhibitors with enzymes. This is quite surprising at first glance, given the potential importance of the information to be gleaned from such structures, both in terms of understanding why they work so well and also in providing ideas for improvements in design. Part of the explanation for this has been that testing of such inhibitors has generally been carried out on glycosidases for which no structural information is available, in particular mammalian glycosidases. Part is a consequence of a lack of coordination between synthetic chemists and X-ray crystallographers. Thus, the only structure available of a complex between a deoxynojirimycin-type inhibitor and an enzyme is that with the inverting family 15 $\alpha$-glucosidase glucoamylase (Harris et al., 1993). Very recently, the structure of the natural product castanospermine with a family 5 $\beta$-glucosidase has been published, the only such structure available for $\beta$-glycosidases (Cutfield et al., 1999). Likewise the only structure available for a complex with a nitrogen containing inhibitor of the sp$^2$-hybridized class is that recently published on the cellobiose-derived imidazole bound to the cellulase Cel5A from Bacillus agaradhaerens (Varrot et al., 1999). A great need exists for such structural information to validate theories of inhibitor design and suggest improvements. Excellent candidates for such a study are the family 10 xylanase Cex from Cellulomonas fimi and the family 11 xylanase Bcx from Bacillus circulans. High resolution crystal structures are available for both enzymes and for the structures of their glycosyl-enzyme intermediates (Notenboom et al., 1998b; Sidhu et al., 1999; Wakarchuk et al., 1994; White et al., 1996; White et al., 1994). Further, the two enzymes differ in their protonation trajectory, the family 10 (Clan GH-A) Cex being an anti-protonator and the family 11 (Clan GH-B) Bcx being a syn-protonator. In addition, on the basis of the structures of their 2-deoxy-2-fluoroxylolobiosyl-enzyme intermediates it has been surmised that the two enzymes likely proceed via two different transition state conformations, Cex proceeding via a $^4H_3$ half-
chair and Bcx via a $^{25}B$ boat (Figures 5.1a and b) (Notenboom et al., 1998b; Sidhu et al., 1999; White et al., 1996)

**Figure 5.1a** Proposed catalytic mechanism of the family 10 glycosidase Cex
To this end we have recently reported the synthesis of a series of four nitrogen-containing inhibitors in the xylobiose series, two of the $sp^3$-hybridised class and two of the $sp^2$-hybridized class and we have reported inhibition studies with both Cex and Bcx (Figure 5.2) (Williams et al., 1999). Interestingly, excellent inhibition (micromolar to nanomolar) was seen for all inhibitors with Cex, but in the case of Bcx they functioned as only very modest (millimolar) inhibitors. In this manuscript we describe the structural analysis of the complexes formed by these four inhibitors with the Cex catalytic domain and we use these observations to explain why these inhibitors function relatively poorly with Bcx.

Figure 5.2  Xylobiose derived inhibitor series against Cex, clockwise from top left: Imidazole, Lactam oxime, Deoxyojirimycin and Isofagomine
Materials and Methods

Crystal handling.

Tetragonal crystals (P4_1,2,2) were grown by hanging drop vapour diffusion method from wt-Cex over-expressed in E. coli, which was papain digested and purified as described previously (Bedarkar et al., 1992; Gilkes et al., 1988). Inhibitors were synthesized as described (Williams et al., 2000). For soaking experiments inhibitor concentration was typically brought to approximately 40 mM concentration in artificial mother liquor (20% PEG4000, 100 mM sodium acetate, pH 4.6). 2 μl of inhibitor solution was then added to a crystal in a ~4 μl hanging drop and allowed to equilibrate. Soaking times ranged from 2 to 48 hours prior to data collection, with no apparent difference in final occupancy values.

Data collection and Processing.

Inhibitor soaked crystals were sequentially transferred into cryosolutions of artificial mother liquor supplemented with increasing amounts of glycerol, to a final glycerol concentration of 30%. Crystals were then scooped into a rayon fiber loop and placed in a stream of N₂ gas at 100K. In order to reduce mosaic spread the crystals, once frozen, were brought back to room temperature for ~20 seconds and then quickly refrozen to allow for re-annealing. All data were collected in-house on a Mar345 image plate using Osmic mirror focussed Cu-Kα X-rays, generated from a rotating anode operating at 100mA × 50kV. Oscillations of 1° were collected in 60-120s exposures, depending on the size of the crystal. All data were indexed, integrated and scaled in the Denzo/Scalepack program package (Otwinowski & Minor, 1997). Data collection statistics are listed in Table 5.1.

Phasing.

10% of reflections were set aside for cross-validation purposes. Upon freezing the unit cell axes decrease in length and the Cex crystals become essentially non-isomorphous with those mounted at room temperature. The exact position of the molecule in the
asymmetric unit was easily found using rotation and translation searches in X-plor (Brünger, 1996) using the wtCex coordinates (PDB code 2EXO). All subsequent refinement was performed in the CNS program package (version 0.5) (Brünger et al., 1998). After low-resolution anisotropic bulk-solvent corrections the molecular replacement model was subjected to a simulated annealing routine starting at 3000K and a stepwise decrease of 25K, with maximum likelihood refinement methods using structure factor amplitudes. From this model 2lFol-lFcl and lFol-lFcl electron density maps were calculated and examined at σ levels of 1.0 and 3.0 respectively. In all 4 structures difference (lFol-lFcl) density unambiguously allowed for locating and interpreting the inhibitor in the active site. At this point several rounds of water building were performed prior to incorporation of ligand coordinates in an effort to avoid biasing the density. Upon free R-factor convergence the ligands were built in manually, and a positional refinement was performed using Xplo2D (Kleywegt, 1997) generated topology and parameter files augmented to reflect bond lengths and angles of the compounds in solution (Williams et al., 1999), with dihedral and torsional restraints removed for the proximal (-1) sugar. Finally, individual B-factors were refined, as well as occupancy levels for the inhibitors, which yielded full incorporation of the compounds in all 4 structures. Final structure quality statistics are summarized in Table 5.2.
## Table 5.1 Data Collection Statistics

<table>
<thead>
<tr>
<th></th>
<th>Nolirimycin</th>
<th>Isofagomine</th>
<th>Imidazole</th>
<th>Lactam oxime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum resolution</td>
<td>1.82 Å</td>
<td>1.95 Å</td>
<td>1.90 Å</td>
<td>1.72 Å</td>
</tr>
<tr>
<td>measured observations</td>
<td>519114</td>
<td>385031</td>
<td>415786</td>
<td>424842</td>
</tr>
<tr>
<td>unique reflections</td>
<td>27874</td>
<td>20562</td>
<td>24224</td>
<td>33537</td>
</tr>
<tr>
<td>U/σ</td>
<td>14.7</td>
<td>11.1</td>
<td>13.6</td>
<td>14.0</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.6 (97.4)</td>
<td>93.9 (85.6)</td>
<td>97.2 (83.8)</td>
<td>94.9 (89.7)</td>
</tr>
<tr>
<td>1Rmerge (%)</td>
<td>4.3 (24.6)</td>
<td>7.5 (31.1)</td>
<td>4.3 (25.6)</td>
<td>4.5 (23.5)</td>
</tr>
</tbody>
</table>

\[
R_{merge} = \frac{\sum_{hkl} \sum_{i} |I_{hkl} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_{i} \langle I_{hkl} \rangle}
\]
Table 5.2  Structure refinement and model quality statistics

<table>
<thead>
<tr>
<th></th>
<th>Deoxy-nojirimycin</th>
<th>Isofagomine</th>
<th>Imidazole</th>
<th>Lactam oxime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell (a=b,c in Å)</td>
<td>87.04, 80.27</td>
<td>85.31, 79.10</td>
<td>85.34, 78.60</td>
<td>87.25, 80.36</td>
</tr>
<tr>
<td>'R-factor, all data (%)</td>
<td>20.2</td>
<td>18.7</td>
<td>18.8</td>
<td>20.2</td>
</tr>
<tr>
<td>'R-free, all data (%)</td>
<td>23.0</td>
<td>22.4</td>
<td>21.9</td>
<td>21.7</td>
</tr>
<tr>
<td>No. protein atoms</td>
<td>2399</td>
<td>2399</td>
<td>2399</td>
<td>2399</td>
</tr>
<tr>
<td>No. substrate atoms</td>
<td>18</td>
<td>17</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>No. solvent molecules</td>
<td>269</td>
<td>232</td>
<td>249</td>
<td>271</td>
</tr>
</tbody>
</table>

R.m.s deviation from ideal

- Bond lengths (Å) | 0.006 | 0.009 | 0.007 | 0.006 |
- Bond angles (°)  | 1.4   | 1.5   | 1.4   | 1.4   |
- Dihedrals (°)    | 22.3  | 22.2  | 22.4  | 22.6  |
- Improper (°)      | 0.85  | 0.74  | 0.82  | 0.79  |

Average B-factor (Å²) | 14.80 | 23.7  | 14.20 | 13.80 |
Average B-factor substr. (Å²) | 13.59 | 17.82 | 12.03 | 11.79 |

\[
R_{e|x}^{\text{cryst}} = \frac{\sum_{hkl} |F_{obs}|-k|F_{calc}|}{\sum_{hkl} |F_{obs}|}, \quad R_{e|x}^{\text{free}} = \frac{\sum_{hkl\in T} |F_{obs}|-k|F_{calc}|}{\sum_{hkl\in T} |F_{obs}|}
\]
Results

**Xylobiose-derived Isofagomine**

The xylobiose-derived isofagomine was found to inhibit Cex with a $K_i$ of 130 nM (Williams et al., 1999). The crystals of the complex with Cex-cd diffracted to 1.95 Å resolution, the structure was solved to an R-factor of 18.7%, and the $R_{free}$ converged to 22.4%. No significant changes in the positions of protein atoms in this complex compared to that of the wild type enzyme were observed upon binding of the inhibitor. lFol-lFcl difference density (Figure 5.3a) unambiguously showed the disaccharide occupying the -1 and -2 binding sub-sites as expected, and the sugar could readily be built into the density. The distal (-2) xylosyl residue makes contacts as seen in the covalent 2-deoxy-2-fluoro-xylobiosyl Cex complex (Notenboom et al., 1998b). The proximal sugar adopts a relaxed $^4C_1$ chair conformation, but despite its excellent inhibition, appears to make surprisingly few contacts with the enzyme. Interactions of the 2-hydroxyl of the substrate with the enzyme are thought to play a crucial role in stabilization of the transition state, however, this molecule lacks a hydroxyl group at the 2-position and, correspondingly, no important contacts with the 2-position and either Asn126 or Glu233 are observed. The major interaction of the inhibitor with the enzyme is between the ‘anomeric’ nitrogen and the Oe1 oxygen of the nucleophile, at a distance of 2.6 Å. Presumably, this interaction constitutes a strong hydrogen bond, with the nucleophile bearing a negative charge and the basic nitrogen bearing a significant positive charge. The only other significant interactions with the inhibitor in the −1 sub-site is found between the 3-hydroxyl and His80 and Lys47, as seen in the 2-deoxy-2-fluoro-xylobiosyl Cex complex.

**Xylobiose-derived deoxynojirimycin**

Xylobiose-derived deoxynojirimycin soaked Cex-cd crystals diffracted to 1.82 Å and the data were refined to R and $R_{free}$ factors of 20.2 and 23.0% respectively, see Table 5.2. Excellent difference density allowed for relatively easy interpretation of the position of
the substrate. The disaccharide was found to occupy the -2 and -1 subsites as anticipated, with the distal xylose moiety making very similar contacts with the enzyme as that seen for the 2-deoxy-2-fluoro-xylobiosyl Cex complex (Figure 5.3b). The deoxynojirimycin ring, like the isofagomine, assumes a \( ^4C_1 \) conformation. His80 and Lys47 again form hydrogen bonds with the 3-hydroxyl group. The 2-hydroxyl group forms a relatively short hydrogen bond with the Oe2 oxygen from Glu233 (2.8 Å) and the amine of Asn126 (2.7 Å). Surprisingly, no protein interactions are found with the endocyclic nitrogen of the inhibitor. Instead, two water molecules are positioned to make hydrogen bonds with this atom, one of which forms a hydrogen bond back to the protein through Gln203.

**Xylobose-derived lactam oxime**

The complex of the lactam oxime with Cex-cd was solved to a resolution of 1.72 Å and refined to R and \( R_{\text{free}} \) of 20.2, and 21.7% respectively. Difference density contoured around the active site revealed clearly the location of the inhibitor and allowed the compound to be incorporated into the model in the -2 and -1 subsites (Figure 5.3c). The distal xylosyl residue was found to form the same interactions with the enzyme as previously described. The lactam oxime moiety assumes planar geometry around the C1 carbon, as dictated by the exocyclic double bond off this atom. The geometry of this double bond is clearly (Z), as predicted based on comparisons of the \(^{13}\)C NMR chemical shift with the corresponding per-acetate (Williams et al., 1999). The proximal ring is seen to bind in a \( ^4E \) conformation. The ring nitrogen forms a strong intramolecular hydrogen bond with the exocyclic hydroxyl ‘arm’, at a distance of 2.5 Å, residing in the same plane as that described by N1, C1 and the exocyclic nitrogen. As seen before, the 3-hydroxyl is found to interact with Lys47 and His80, and the 2-hydroxyl forms hydrogen bonds with Asn126 and Oe2 from Glu233, the nucleophile. The exocyclic nitrogen forms a close hydrogen bonding contact with the acid/base catalyst (2.6 Å) as well as with Gln203 (2.6 Å). Gln203 also interacts with the exocyclic hydroxyl of the ligand.
**Xylobiose-derived imidazole**

The imidazole complex was solved to 1.90 Å resolution, and refined to a crystallographic R-factor of 18.8% and a corresponding R$_{free}$ of 21.9%. The compound was easily built into difference density, occupying the -1 and -2 binding subites as expected (Figure 5.3d). The distal xylosyl ring forms interactions as described above. The proximal ring assumes a $^4E$ conformation similar to that of the lactam oxime. The imidazole ring is observed to be highly planar around C1. Similar interactions are found as with the lactam oxime, its exocyclic nitrogen interacts with the acid/base catalyst (Glu127) as well as Gln203.

All four of the above-described structures show minimal change in Cα-trace (rmsd with wild type ranges from 0.3-0.4 Å) as well as no significant perturbation in the orientation of the active site residues.

---

**Figure 5.3 a-d (next page)**

Stereo representation of 2Fol-lFcl density (left) contoured at 1σ, phased with protein coordinates after rigid body refinement with native Cex for each of the xylobiose-derived inhibitors: a) Isofagomine b) Deoxynojirimycin c) Lactam oxime and d) Imidazole. The model coordinates for the compounds are superimposed for clarity, together with the nucleophile Glu233 and the acid/base catalyst Glu127. Schematic representations of relevant active site interactions with the proximal (-1) sugar and Cex are shown on the right. Only in the case of xylobiosyl isofagomine (2a) the interactions around the distal sugar are drawn, these remain the same for the other complexes but are omitted for clarity.
Discussion

Basic, hydroxylated azasugars, over many years, have been widely demonstrated to act as powerful glycosidase inhibitors. The obvious similarity of such compounds to certain aspects of the transition state has been noted and five principal factors have been suggested as making important contributions to the transition state character of an inhibitor. These may be listed as charge distribution, trigonal 'anomeric' centre, half-chair conformation, proper relative configuration and the ability to be directionally protonated. Our results confirm that all four inhibitors have the correct relative configuration. The four inhibitors satisfy the four remaining requirements to varying degrees and will be discussed individually.

The deoxyoctimycin and the isofagomine xylobiose-derived analogs

Since the discovery of the potent inhibitory effects of deoxyoctimycin toward glycosidases in the early 1970's, this class of compound has engendered a vast amount of interest concerning their synthesis and the evaluation of their biological properties (Hughes & Rudge, 1994; Ichikawa et al., 1998; Stütz, 1999). Nojirimycin derived inhibitors have been widely suggested to be powerful glycosidase inhibitors because their basic endocyclic nitrogen, when protonated, resembles the positively charged transition states of the enzyme catalyzed hydrolysis reaction. These compounds are generally better inhibitors of α-glycosidases than of β-glycosidases, although the xylobiose-derived deoxyoctimycin is a good inhibitor of Cex with a $K_i$ value of 4.8 μM. It has been suggested that nojirimycin-like compounds may be effective against β-glycosidases through interaction of the endocyclic nitrogen with the acid-base catalyst. In the crystal structure determined here, no specific protein-substrate interactions are observed involving the endocyclic nitrogen. Instead, it binds two water molecules, one in an axial direction above the nitrogen and one placed equatorially from the nitrogen. Only the equatorial water molecule may be traced back to the protein through a hydrogen bond with Gln203. The nearest oxygen of the acid/base catalyst Glu127 is 4.0 Å distant from
the endocyclic nitrogen. Thus, at least in Cex, this compound seems to derive its inhibitory effect predominantly from specific interactions with the 2- and 3-OH groups and from interactions of the distal saccharide with the protein, and not through the formation of a strong hydrogen bond with the nitrogen atom. Though there is no formation of a strong hydrogen bond with the nitrogen atom, electrostatic interactions in the proximal binding site presumably do contribute to its inhibition. It should be kept in mind that while the deoxynojirimycin analog is the weakest of the four studied inhibitors in Cex, it still binds Cex tighter than it does its product, xylobiose.

Of some relevance here is the X-ray structure of deoxynojirimycin bound to glucoamylase, an inverting α-glycanase, determined by Harris et al. (Harris et al., 1993).

This enzyme catalyzes the hydrolysis of α-1,4-linked glucans and folds in an (α/α)_6 barrel. In this enzyme, the two catalytically essential carboxyl groups are spaced approximately 9.5 Å apart to allow for the interspacing of the nucleophilic water molecule. Consequently, only a weak interaction between the ring nitrogen and the two carboxyl groups was seen, the closest being 3.9 Å. Despite this weak interaction, deoxynojirimycin is a strong inhibitor of glucoamylase with a $K_i$ value of 96 μM. Notably, this structure showed the inhibitor bound in a $4C_1$ conformation similar to that observed here.

Recently, the crystal structure of castanospermine bound to Exg from Candida albicans was determined (Cutfield et al., 1999). Exg is a family 5 enzyme that catalyzes the hydrolysis of β-1,3-linked glucans through a retaining mechanism and folds in a (β/α)_6 barrel similar to that of Cex. The bicyclic castanospermine molecule may be considered a derivative of deoxynojirimycin modified by a bridging ethylene group between C6 and the ring endocyclic nitrogen. The X-ray structure of castanospermine alone shows that the piperidine ring of this molecule prefers a $4C_1$ conformation. However, in the complex with Exg, castanospermine was shown to bind in a $1S_3$ twist-boat conformation allowing for a relatively close contact (2.69 Å) between the ring nitrogen and the nucleophile
residue, leading to the suggestion that this molecule was binding in a conformation approaching the $^4H_3$ conformation of the transition state, and thus was binding as a transition state analogue. Despite Exg belonging to the same clan as Cex, the castanospermine structure is markedly different from the inhibitor structures observed in Cex.

Thus, it appears that deoxynojirimycins may bind to different enzymes in quite different conformations and thus the degree of transition state resemblance may differ from one glycosidase to the next. Also, the interactions of enzyme residues with the nitrogen atom of the inhibitor may vary on a case-by-case basis even with enzymes of the same clan.

The second structure determined here was of the isofagomine bound to Cex. In this class of inhibitor the anomeric carbon is substituted with an sp$^3$-hybridized basic nitrogen atom and the endocyclic oxygen is replaced with a methylene group. The xylobiosyl isofagomine is an excellent inhibitor of Cex, the best seen for this enzyme thus far, with a $K_i$ value of 130 nM (Williams et al., 1999). The strong affinity of this compound for Cex is attributed to the formation of a strong hydrogen bond between the nucleophile Glu233 and the (presumably) positively charged nitrogen of this compound. Although its inhibition is extremely good, the isofagomine is an unlikely candidate for a transition state analogue of this enzyme owing to the fact that the ring conformation is that of a relaxed $4C_1$ chair, as opposed to the $^4H_3$ half chair conformation expected of the transition state. Perhaps the tight binding exhibited by this compound is somewhat fortuitous; indeed, there is a striking similarity between the isofagomine Cex complex and that of the 2-deoxy-2-fluoro-xylobiosyl Cex complex suggesting that this compound may bind with analogy to the intermediate glycosyl enzyme.

A further difference between this complex and the proposed transition state results from the lack of a 2-hydroxyl group present in the substrate. Interactions of the enzyme with the 2-hydroxyl group of the substrate has been suggested to provide upwards of 40 kJ mol$^{-1}$ of stabilization of the transition state of the enzyme catalyzed reaction (Namchuk &
Withers, 1995). In the case of Cex, X-ray crystal structures of trapped glycosyl enzyme species have implicated N126 and E233 as the main residues that interact with this position (White et al., 1996; Notenboom et al., 1998). Owing to the lack of a 2-hydroxyl group, there are no strong interactions with either N126 or E233 and the 2-position. Introduction of a 2-hydroxyl group into an isofagomine would take advantage of these interactions, however, such a molecule is unlikely to be stable and would dehydrate in common with other hemiaminals. Recently, Williams et al. reported the synthesis of an isofagomine-derived lactam, which was a surprisingly powerful inhibitor of Cex ($K_i$ 340 nM) (Williams et al., 2000). Mutation of the N126 residue to alanine severely compromised the affinity of this molecule for this mutant ($K_i$ = 2.0 mM) by an amount that corresponds to a change in free energy of $-5.4$ kcal mol$^{-1}$. This figure was very similar to that seen for other 2-hydroxylated inhibitors and for $p$-nitrophenyl $\beta$-cellobioside. In the X-ray structure of this molecule with Cex, the piperidine was seen to bind in a remarkably planar $E_5$ conformation (see also Chapter 6 of this manuscript). On the basis of these results Williams and coworkers concluded that the high affinity of this molecule resulted from a tautomerization of the lactam to an iminol, which would provide a basic nitrogen at the anomeric position, a hydroxyl group at the 2-position, and an sp$^2$-hybridized anomeric centre. Protonation of the iminol would allow a strong electrostatic interaction between the iminium ion and the carboxylate of the nucleophile. Importantly, the restoration of a hydroxyl group at C2 could take advantage of the large amount of binding energy used for stabilization of the transition state.

*The imidazole and the lactam oxime*

Another class of inhibitors investigated here is that in which the anomeric carbon is sp$^2$-hybridized, the ring oxygen being substituted by a nitrogen atom and having an exocyclic double bonded nitrogen off Cl. Besides being suggested to resemble the proposed $4H_3$ conformation of the Cex transition state, these inhibitors have a charge distribution resembling that of the hypothetical transition state. In this class of compounds,
protonation of the 'glycosidic' nitrogen may resemble protonation of the departing aglycon at the transition state. More specifically, these compounds have been suggested as being able to distinguish between syn- and anti-protonation by glycosidases (Figure 5.4a and b). The orientation of the lone pair on the 'glycosidic' nitrogen of such an inhibitor allows only for anti-protonation by the acid/base catalyst relative to the C1-O5 bond and may preclude interaction coming from a syn protonation trajectory.

**Figure 5.4a** Schematic representation of the anti- (left) and syn- (right) protonation trajectories in retaining β-glycosidases relative to the H1-C1-O5 plane of the sugar

**Figure 5.4b** Schematic representation of possible interactions of the compound lactam oxime in anti (left) and syn (right) protonating glycosidases.
In Cex, an anti-protonator, the xylobiose-derived imidazole and lactam oxime, were excellent inhibitors with $K_i$ values of 150 and 370 nM, respectively. Analysis of the inhibitor-Cex complex crystal structures indeed showed strong hydrogen bonds (2.6 Å) between the exocyclic nitrogen and the acid/base catalyst Glu127. Similarly, in a recently published complex of a cellobiose-derived imidazole with Cel5A, also a clan GH-A glycosyl hydrolase and anti-protonator, equivalent interactions were found (Varrot et al., 1999). The high resolution of that structure (0.97 Å) allowed assignments of individual protons, locating a shared proton between the acid/base catalyst and the 'glycosidic' nitrogen. However, in both of these cases, the transition state of the reaction is highly likely to be formed from the Michaelis complex which possesses a $1S_2$ skew boat conformation, resulting in the upward, pseudo-axial orientation of the C-O bond being cleaved and thus protonic assistance is delivered in the anti-direction but from above the plane of the sugar ring (Davies et al., 1998a; Sulzenbacher et al., 1996). This is in line with observations in the Cel5A crystal structures of substrate and inhibitor complexes, in which the protonated glycosidic oxygen is displaced significantly as the reaction progresses along the reaction coordinate with very little movement of the acid/base catalyst, while maintaining hydrogen bonding distance throughout (Davies et al., 1998a).

As Davies and co-workers suggest, the protonation machinery of these enzymes seem to be able to sustain a large margin of positional variation and still maintain efficient protonic assistance, a likely reason for the ubiquity of carboxylates as acid catalysts in glycosyl hydrolases. This would suggest the imidazole and lactam oxime may not be entirely representative of the interactions formed at the transition state, since their exocyclic nitrogens are planar off the anomeric carbon and clearly below the predicted pseudo-axial orientation expected of the glycosidic oxygen. One possible outcome of the sub-optimal in-plane protonation of the 'glycosidic' nitrogens of the imidazole and the lactam oxime could be a slight upward displacement of the position of the sugar ring of these inhibitors compared to that expected for the transition state (Figure 5.5). Moreover,
the inability for the nucleophile to approach C1 to within 'bonding' distance would tend to favor this slight upward displacement.

Figure 5.5
Stereo representation of the superposition of Cex in complex in covalent complex with 2F-xylobiose in yellow (PDB code 2xyl) and with bound xylobiosyl imidazole. The imidazole moiety is located slightly upward from the expected transition state orientation.
Although molecules such as the lactam oxime and imidazole have been suggested to be good mimics of the \( ^4H_3 \) conformation of the transition state of the enzyme-catalyzed reaction, the structures determined herein show that these two compounds bind in a \( ^4E \) conformation, in agreement with the structure determined of the cellobiose-derived imidazole with Cel5A from \( \textit{B. agaradhaerens} \) (Varrot et al., 1999). That these compounds bind to Cex in a \( ^4E \) conformation is surprising and may be a result of an upward displacement of the sugar ring compared to its location in the transition state as suggested above. Raising the sugar ring in this fashion while maintaining the position of C-4 and O-4 through the invariant positioning of the –2 sugar may prevent deformation of the ring into a \( ^4H_3 \) conformation. Relaxation of the ring into a \( ^4H_3 \) conformation could cause a subtle change in the position of the 2-hydroxyl that would shorten and possibly strengthen the interaction of the 2-OH with the carbonyl oxygen of the nucleophile. However, in these complexes, the nucleophile is only able to come within non-bonding distance of Cl and thus the sugar ring fails to bind in the precise manner as that of the transition state. It is of interest here to review the structures that have been observed for similar sp\(^2\)-hybridized molecules. The structure of the d-gluco lactam oxime in D\(_2\)O solution was concluded to be a \( ^4C_1 \) chair based on the \(^1\)H NMR spectrum. On the other hand, the X-ray crystal structures of d-gluconolactam and the d-gluco lactam oxime showed these molecules to be in conformations close to that of a \( ^4H_3 \) half-chair. Xylose-derived isofagomine lactam was also shown to adopt a \( ^4H_3 \) conformation in a single crystal X-ray determination. However, the corresponding disaccharide complexed with Cex was shown to bind in a \( E_3 \) conformation (Williams et al., 2000). Thus, the preferred conformations of these molecules in the solid state appears to give very little indication of the conformations adopted when bound to an enzyme, where specific interactions may tip the delicate balance of one conformation in favor of another.

Some ambiguity exists in the location of the double bond in molecules such as the lactam oxime. For the neutral molecule, convincing evidence has been presented to show that
the double bond is exo-cyclic. However, upon protonation of the molecule, the double bond is expected to become delocalized over the N5-C1-N1 system. The resolution of the structure determined here does not allow for an accurate measurement of the bond-lengths. However, the distances observed here are consistent with such a structure.

*Implications for the catalytic mechanism*

Detailed structural information has been gathered over the last decade on glycosidases. Probably the best characterized enzyme thus far in terms of substrate/inhibitor complexes is the retaining family 5 cellulase from *B. agaradhaerans* for which X-ray structures have been obtained for all of the stable species along the reaction coordinate, namely the free enzyme, the Michaelis complex, the glycosyl enzyme intermediate and the product complex (Davies et al., 1998a). These data are highly supportive of the mechanism proceeding through a 4C1 - 1S - 4H - 4C itinerary. Cex, with a similar (β/α)8 barrel fold, belongs to the same clan as Cel5A (GH-A) and is believed to proceed through a similar itinerary.

What seems clear here is that none of these inhibitors are binding in exactly the way expected of the transition state, a point not particularly surprising given the wide range of structures and of course the limitations of chemical synthesis. However, to a large extent, these molecules interact with active site residues in an identical manner, particularly about O3 and O4. The greatest differences between these structures lie in interactions of nitrogen atoms of the inhibitors with the catalytic machinery of the enzyme.

Unfortunately, none of these compounds can truly mimic the hypothetical transition state in its entirety, particularly the partial positive charge which develops on O5, C1 and O1, the presence of an aglycon, and the trigonal bipyramidal geometry about C1 with elongated apical bonds. Remarkable insight has been gained into such a structure through the recently determined structure of a 10-C-5 hypervalent carbon compound which may be considered a stable transition state analogue of the bimolecular nucleophilic substitution reaction (Akiba et al., 1999). Stabilization of a dioxacarbenium
ion was achieved by two methoxy groups mounted on an anthracene scaffold. In this structure, the apical carbon oxygen bonds were shown to be approximately 2.45 Å in length, and molecular modeling suggested that true partial bonds were formed between the central carbon atom and its apical neighbours. Such a molecule is perhaps not a true analogue of the transition state expected for glycoside hydrolysis as the two apical oxygen atoms in this molecule were observed to be sp²-hybridized, presumably to allow for back-bonding from the aromatic π-system into the unoccupied p-orbitals. Nonetheless, this compound represents the first successful preparation of a molecule possessing a five coordinate carbon with elongated partial bonds off carbon.

Possible reasons for the poor inhibition of Bcx

While these four inhibitors are potent inhibitors of Cex, a family 10 xylanase, comparable inhibition was not achieved for another enzyme, the family 11 xylanase Bcx from Bacillus circulans (Williams et al., 2000). However, it should be noted, that inhibition by these compounds in Bcx was still considerably better than that observed for xylobiose, thus these compounds do provide some additional inhibitory effect in Bcx as well. Bcx processes the natural substrate, xylan, at a similar rate as Cex, however, in contrast to Cex, Bcx has no activity towards cellobiose and carboxymethyl cellulose. There are substantial differences in the two enzymes' structures, Bcx having a β-jelly roll fold while Cex folds as a (β/α)₆ barrel.

Recently, the structure of the 2-deoxy-2-fluoro-xylobiosyl enzyme formed on Bcx was determined (Sidhu et al., 1999) (and later the xylanase from B. agaradhaerens (Sabini et al., 1999)), and the 2-deoxy-2-fluoro-xylosyl moiety was seen to be in a 2.5B conformation, unlike the 4C₁ conformation observed on Cex. These observations lead to the suggestion by Sidhu and co-workers that these two enzymes differ in their mechanism, Cex proceeding through a 'classic' 4C₁ - 1S₃ - 4H₃ - 4C₁ itinerary (Figure 5.1a), while a 4C₁ - 2H₁ - 2S₀ - 2.5B - 4C₁ itinerary was suggested for Bcx (Figure 5.1b) (Sidhu et al., 1999). According to this proposal, the transition state for the reaction
catalyzed by Bcx was suggested to be a \(^{2,5}\)B conformation. The geometry observed in the intermediate structure keeps the C2-C1-O5-C5 atoms in the same plane, thereby obeying the stereochemical prerequisite of the oxocarbenium ion-like transition state, while facilitating the subsequent nucleophilic attack during deglycosylation. The distortion required for this type of mechanism likely exclude glucose-derived substrates that possess a bulky hydroxymethyl substituent, in line with the observation that xylans are the exclusive substrate for family 11 enzymes. A further difference between these two enzymes is the location of the catalytic acid residue. Bcx has the acid/base residue located 'syn' to the C1-O5 bond of the substrate while, as mentioned above, the acid/base in Cex is positioned \(\text{anti}\) to the C1-O5 bond.

The detailed structural information gathered in the course of these studies with Cex allows for an analysis of the possible reasons for the poor affinity observed with these inhibitors with Bcx. Interestingly, inhibition constants for the xylobiose-derived deoxynojirimycin and isofagomine against Bcx were modest, with \(K_i\) values of 1.5 and 1.1 mM, respectively (Williams et al., 1999). In this context, it is indeed surprising that Bcx was inhibited by neither the deoxynojirimycin nor isofagomine compounds to an appreciable extent, since these compounds should be relatively amenable to distortion into the \(^{2,5}\)B conformation of the transition state. However, even in cyclohexane, there is a 6.4 kcal mol\(^{-1}\) free energy difference between a chair and a boat conformation that could be responsible for the poor affinity. Manual docking of these compounds into the Bcx active sites in their relaxed \(4C_1\) conformations (overlaying the 2-deoxy-2-fluoro-xylobiosyl Bcx complex on the C4-C3-C2-C1 atoms of the proximal sugar and the nucleophile) revealed no obvious clashes between the enzyme and these compounds, other than a possible close contact of Tyr69 against C6 of the isofagomine or N5 of the deoxynojirimycin. One possible explanation for the lack of inhibition by these compounds in Bcx could lie in the absence of an endocyclic oxygen atom in their ring structure. In the X-ray structure of the 2-deoxy-2-fluoro-xylobiosyl Bcx complex
mentioned above, the phenolic hydroxyl of Tyr69 was seen to be in close proximity to the endocyclic oxygen of the inhibitor. Whether this interaction was in the form of a hydrogen bond between these two groups could not be determined, however, it was suggested instead that the close proximity of Tyr69 to the endocyclic oxygen destabilizes the ground state relative to the transition state and thus reduces the activation energy of the reaction. A role for this residue in catalysis was suggested wherein a bifurcated hydrogen bond is formed from Tyr69 to both the endocyclic oxygen, O5, and the carbonyl oxygen of the nucleophile, as shown in Figure 5.6

![Figure 5.6](image_url)  
**Figure 5.6** Interactions around the distorted covalent intermediate in Bcx (from: (Sidhu et al., 1999))

Upon breakdown of this intermediate, desymmetryzation of the bifurcated hydrogen bond would assist in the departure of the enzyme carboxylate, thus inducing a partial negative charge on the phenolic oxygen that would serve to stabilize partial positive charge development on the endocyclic oxygen as the reaction approaches the second transition state. The geometry of such an interaction is only optimized in a $25^\circ B$ conformation, since
the oxygen lone pair is pointed in the wrong direction in the $^4C_1$ conformation. Substitution of Tyr69 with Phe leads to an inactive mutant, supportive of a critical role in catalysis for this residue. Quite possibly, interactions of Tyr69 with the isofagomine may be particularly unfavorable, as it cannot accept a hydrogen bond from Tyr69 at this position, as the endocyclic oxygen atom of the substrate has been substituted by a methylene group. In the case of the deoxynojirimycin derivative, positioning of C1 of the inhibitor to within non-bonding distance of the nucleophile could well cause a steric clash between N5 and Tyr69. In both cases these interactions would explain why good inhibition by these two inhibitors is not seen in Bcx. In addition, the lack of a departing group in these non-covalent inhibitors may also play a role in the lack of observed inhibition. Protein interactions at the +1 binding site with the aglycon may provide an additional driving force for distortion of the 2-deoxy-2-fluoro-xylosyl moiety into the observed distorted geometry, which is then maintained by specific stabilizing interactions in the active site. A product complex of an inactive nucleophile mutant of Bcx, with Glu172 modified to Cys, complexed with xylo tetraose has been determined and that shows xylobiose bound in the $-1$ and $-2$ subsites. In this complex, the $-1$ sugar bound is bound in a more conventional $^4C_1$ conformation, with an increase in the O5-Tyr69 separation. It may be that both the deoxynojirimycin and the isofagomine bind to Bcx in a $^4C_1$ conformation and that their affinity for this enzyme more closely resembles that of the product than of the transition state.

Bcx is not inhibited significantly by the xylobiosyl imidazole or lactam oxime molecules. Most probably, the poor affinity of these two compounds for Bcx lies in the spatial positioning of the acid/base residue. Bcx belongs to family 11 of the glycosyl hydrolase classification scheme and is thus a syn-protonator. Consequently, in this enzyme, the non-bonding nitrogen orbitals of the imidazole and lactam oxime are directed away from the acid-base catalyst.
A further possibility for the poor affinity of these compounds for Bcx could be for reasons similar to those discussed above for the deoxynojirimycin and isofagomine, namely, possible steric clashes with the inhibitors and other active site residues. These two compounds should be able to deform to a \( 25B \) conformation with little difficulty. However, such a conformation would have to go through the high energy \( 2SO \) conformation, which the enzyme might not be able to accomplish with these inhibitors, possibly, in part, because they are lacking +1 subsite sugars. Moreover, if these inhibitors bind to Bcx in a \( 25B \) conformation, significant clashes are likely to occur between the phenolic hydroxyl of Tyr69 and the oxime or imidazole moieties. Conversely, if these compounds bind in a \( 4H \), or a \( 4E \) conformation, a close contact between the oxime or imidazole moieties and Val37 of Bcx would result.

Summary

Inhibitors designed as structural and mechanistic probes, in conjunction with X-ray crystallographic studies, can be powerful tools to deconvolute subtleties in reaction mechanisms of enzymes as well as mechanistic differences between families of enzymes, and allows for a more comprehensive interpretation of kinetic data. In this paper we describe the crystal structures of a family 10 xylanase in the context of four different inhibitors, designed to address key features of the reaction mechanism. Besides resembling the transition state of family 10 mechanisms in anomeric configuration and charge, the lactone-inspired xylobiosyl imidazole and lactam oxime specifically interact with the acid catalyst of the enzyme as expected. This confirms the classification of these enzymes as anti-protonators and is consistent with assignment of this type of inhibitor as 'transition state analogs'. However, the binding of these compounds in an envelope conformation rather than a half-chair is an interesting point of difference from the transition state. Less clear is the interpretation of the structures of the sp\(^{3}\)-hybridized inhibitors, the xylobiosyl isofagomine and deoxynojirimycin. The isofagomine is perhaps the furtherest from being a true transition state analog of the family 10
glycosidases, its $4C_1$ geometry and lack of interactions with almost all of the active site residues excepting the nucleophile suggests it may be a fortuitous binder rather than a transition state mimic. The deoxynojirimycin on the other hand lacks strong interactions of the protein with the endocyclic nitrogen; rather it interacts with two water molecules, but maintains good interactions at the 2-position.

Perhaps the most striking feature of these compounds is the lack of inhibition they exhibit against Bcx. To comprehend the differences between these two enzymes fully, comparative crystal structures of Bcx in complex with non-covalent inhibitors will be crucial. The current repertoire of inhibitors does not reflect the family 11 mechanism, and thus the details of distinct mechanistic features such as a syn-protonation trajectory and a $2.5B$ transition state conformation remain speculative. Ideally, X-ray structures of these four compounds with Bcx would cast light on the reasons for the low affinity of these compounds for Bcx. However, to this date, we have been singularly unsuccessful in obtaining such structures. Not surprisingly, the poor affinity of these compounds for Bcx appears to prevent high occupancy in the active site of this enzyme. As we continue to gain more insight into the details of enzyme catalyzed glycoside hydrolysis it will be interesting to see whether a new generation of inhibitors could be developed even more specifically geared towards the differences in these two enzymes.
References


Chapter 6

Tautomerization of an Isofagomine-Derived Glyconolactam in the Active Site of Cex: Implications for the Role of Asn126 in Substrate Binding


Copyright © 1998 American Chemical Society
Abstract

A xylobiose-derived lactam, related to xylobiosylisofagomine, was prepared and assayed against a retaining, family 10 xylanase from *Cellulomonas fimi*. This remarkably simple compound is a potent inhibitor of this enzyme ($K_i = 0.34 \mu M$). The crystal structure of the inhibitor-enzyme complex was determined and is completely consistent with the inhibitor binding as the iminol tautomer. The inhibitor binds in a rather unusual $E_3$ conformation although the X-ray structure for the corresponding monosaccharide lactam uncomplexed indicates this compound prefers a $4H_3$ conformation. The $K_i$ value determined for the inhibitor with the N126A mutant, relative to other 2-deoxy and 2-hydroxy inhibitors, is most consistent with this inhibitor binding as the iminol tautomer and interacting with the N126 residue in a fashion similar to the 2-hydroxy containing inhibitors.
**Introduction**

Retaining glycosidases catalyze the hydrolysis of glycosides with overall retention of anomeric stereochemistry, normally with the assistance of two carboxyl groups, one functioning as an acid/base and the other as a nucleophile (Sinnott, 1990). The reaction usually proceeds through a covalent glycosyl enzyme and the transition-states leading to and from this intermediate have been shown to possess considerable oxocarbenium-ion like character (Zechel & Withers, 1999). Many glycosidase inhibitors have been developed over the last 30 years with varying degrees of success. Particularly successful compounds include a number of sugar-shaped heterocycles containing a basic nitrogen at positions that correspond to C1, O5 and the glycosidic oxygen (Heightman & Vasella, 1999; Stütz, 1999). Of these, 1-N-iminosugars are particularly powerful inhibitors of retaining β-glycosidases. However, unlike the substrates for most glycosidases and most other iminosugar inhibitors, these 1-N-iminosugars lack a hydroxyl at C2 (Bols, 1998; Ichikawa et al., 1998). This is unfortunate since the interaction of active-site residues, particularly of the catalytic nucleophile, with OH2 has been suggested to provide upwards of 30-40 kJ mol⁻¹ to the stabilization of the transition-state of enzyme-catalyzed glycoside hydrolysis (Mega & Matsushima, 1983; Notenboom et al., 1998a; Roeser & Legler, 1981; White et al., 1996). The possibility therefore exists to develop much stronger inhibitors of this type if this interaction can be included. However, 1-N-iminosugars with a hydroxyl at C2 are not expected to be stable compounds, since they will presumably dehydrate in common with many hemiaminals. An attractive way of incorporating this interaction into a 1-N-iminosugar is provided in the compound isofagime lactam, whose crystal structure in the enzyme bound form is described here. By binding as the protonated iminol tautomer, this compound could provide a hydroxyl group at C2 and a positive charge at the anomeric position and in part reflect the planarity of the sugar ring at the transition state (Figure 6.1). Investigation of the mode of interaction of such a compound may cast light on some aspects of glycosidase
mechanism, in particular the role of OH2 in the transition-state of the enzyme catalyzed reaction.

![Structure of OH2 in the transition-state of the enzyme catalyzed reaction]

**Figure 6.1.** Potential tautomerisation of the isofagomine lactam in enzyme bound form

Basic hydroxylated iminosugars have been previously shown to be extremely effective inhibitors of many retaining glycosidases (Stütz, 1999). Most, if not all, of these inhibitors have been suggested to mimic various aspects of the transition-state of glycoside hydrolysis, namely the positive charge that develops on C1 and O5 and/or the planarity of the sugar ring, in particular atoms C5, O5, C1, and C2 (Heightman & Vasella, 1999). In Chapter 5 the crystal structures of xylobiose-derived versions of a number of representatives of the main classes of iminosugar inhibitors in complex with Cex were discussed; two compounds with sp² hybridized nitrogen atoms in place of the glycosidic oxygen, namely the imidazole and the lactam oxime, and two compounds with sp³ nitrogen atoms in place of C1 or O5, the deoxynojirimycin and the isofagomine (see also (Williams et al., 2000)). All of these compounds were effective inhibitors of Cex, with $K_i$ values ranging from 150 nM to 4800 nM (Williams et al., 2000). It was hoped that by installing a carbonyl into the isofagomine, affording the isofagomine lactam, we could provide a simple new class of non-basic inhibitors of glycosidases, and of Cex in particular.

**Materials and Methods**

*Crystal preparation, data collection and structure determination*

Tetragonal crystals (P4₁2₁2) were grown from the N126A and N169A mutant protein samples by hanging drop vapour diffusion method from 20% PEG4000 and 0.1 M Na
acetate, pH 4.6, as described previously (Bedarkar et al., 1992). Cex crystals were soaked in artificial mother liquor containing at least 40 mM of the xylobiose derived isofagomine lactam for several hours prior to data collection. A crystal was sequentially transferred into cryo protectant solutions of artificial mother liquor and placed in a stream of N₂ gas at 100K. Diffraction data were recorded in-house to 2.0Å resolution at 100 K on a Mar345 image plate using Osmic mirror focussed Cu-Kα X-rays, generated from a rotating anode operating at 100mA x 50kV. Oscillations of 1° were collected in 120s exposures. Recorded reflections were indexed, integrated and scaled in the Denzo/Scalepack program package (Otwinowski & Minor, 1997). The data set was 99% complete with R_sym of 5.7% and an average I/σ of 13.3, using 258208 observations for 21259 unique reflections, see Table 6.1.

<table>
<thead>
<tr>
<th>Table 6.1  Data collection statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cex-isofagomine lactam</strong></td>
</tr>
<tr>
<td>Maximum Resolution</td>
</tr>
<tr>
<td>Number of observations</td>
</tr>
<tr>
<td>Unique reflections</td>
</tr>
<tr>
<td>Completeness</td>
</tr>
<tr>
<td>I/σ</td>
</tr>
<tr>
<td>R_merge</td>
</tr>
</tbody>
</table>

\[
R_{merge} = \frac{\sum_{hkl} \sum_{i} |f_{hkl} - \overline{f}_{hkl}|}{\sum_{hkl} \sum_{i} f_{hkl}}
\]
The structure was refined using wt Cex as a starting model (PDB code 2EXO) with the CNS program package (Brünger et al., 1998) as described in Chapter 5, to an R-factor of 20.9% (R-free 25.0%). Structure refinement and model quality statistics are summarized in Table 6.2.

**Table 6.2 Structure refinement and model quality statistics**

<table>
<thead>
<tr>
<th>Cex-isofagomine lactam</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell (a=b, c in Å)</strong></td>
</tr>
<tr>
<td><strong>R-factor, all data (%)</strong></td>
</tr>
<tr>
<td><strong>R-free, all data (%)</strong></td>
</tr>
<tr>
<td><strong>No. protein atoms</strong></td>
</tr>
<tr>
<td><strong>No. solvent atoms</strong></td>
</tr>
<tr>
<td><strong>R.m.s. deviation s</strong></td>
</tr>
<tr>
<td>bond lengths (Å)</td>
</tr>
<tr>
<td>bond angles (°)</td>
</tr>
<tr>
<td>dihedral angles (°)</td>
</tr>
<tr>
<td><strong>Average B-factor (Å²)</strong></td>
</tr>
<tr>
<td><strong>Average B-factor substr. (Å²)</strong></td>
</tr>
</tbody>
</table>

$$R_{cryst} = \frac{\sum_{hkl} |F_{obs}| - k|F_{calc}|}{\sum_{hkl} |F_{obs}|}, \quad R_{free} = \frac{\sum_{hkl \notin T} |F_{obs}| - k|F_{calc}|}{\sum_{hkl \notin T} |F_{obs}|}$$

**Results and Discussion**

To understand the structural basis for the remarkable inhibition shown by the isofagomine lactam, the X-ray crystal structure of the complex with Cex was determined, to 2.0 Å resolution. This structure shows the inhibitor binding in the $-1$ and $-2$ subsites in
Figure 6.2  Stereo diagram of 2|Fo|-|Fc| electron density from the protein model, contoured at 1σ. The stick-model of the isofagomine lactam is superimposed on the density for clarity. The inhibitor appears to bind in a E5 conformation, which implies the tautomerization to the iminol form predominates in the enzyme bound form. The hydrogen bonding network involving Glu233 and Asn126 further supports this observation.
The conformation of the piperidine ring when bound in this complex is completely consistent with its binding as the iminol tautomer. The torsional angles measured about the C6-N1-C2-C3 and C6-N1-C2-O2 systems were 21.3° and 175.5°, with C6, N1, C2, C3 and O2 lying very nearly in a plane. The amide tautomer necessarily requires atoms N1, C2, C3 and O2 to be in plane with each other. However, upon tautomerisation to the iminol, there is an additional requirement that C6 lies in this same plane. The catalytic nucleophile, Glu233 lies directly below N1 at a distance of 2.98 Å. If the substrate does in fact bind in the form of the iminol, the lone pair on nitrogen is directed in the plane of the ring and is not capable of forming a strong hydrogen bond with the nucleophile. Nevertheless, if the iminol is protonated, a strong electrostatic interaction between the nucleophile and the iminium ion is possible. Also observed in the crystal structure is a relatively short contact of 2.59 Å between Oe2 of the nucleophile and the lactam oxygen. This short contact is reminiscent of that observed in the X-ray crystal structure of the cellobiosyl enzyme formed with the H205N/E127A Cex double mutant where a distance of 2.37 Å was observed between the carbonyl oxygen and OH2 (Notenboom et al., 1998a). A significant H-bonding interaction (2.97 Å) is also seen between the amide nitrogen of N126 and O2. N126 is a highly conserved residue in family 10 and in clan GH-A and, on the basis of crystal structures of the 2-deoxy-2-fluoro-cellobiosyl (White et al., 1996) and 2-deoxy-2-fluoro-xylobiosyl enzyme (Notenboom et al., 1998b), has been suggested as a residue that hydrogen bonds directly with OH2. Surprisingly, in the case of the cellobiosyl enzyme formed with the H205N/E127A Cex double mutant, the N126 residue fails to form a hydrogen bond with OH2 (see also Chapter 3). Interestingly, a single crystal X-ray analysis of the monosaccharide (xylose derived) lactam showed that the uncomplexed molecule is in a $^4H_3$ conformation similar to that expected for the transition state. While there is only a slight difference between the $E_3$ conformation of the isofagomine lactam-Cex complex and the $^4H_3$ conformation of the unbound native analogue, the fact that it does not bind to the enzyme in the same conformation as the
uncomplexed monosaccharide, and indeed the conformation expected of the transition state, indicates that the isofagomine lactam is not binding exactly as a transition state analogue. The observation of significant interactions between the nucleophile and both N1 and O2 is most consistent with the substrate binding in the protonated iminol form. Indeed, if the amide tautomer were bound to Cex, these interactions with the nucleophile would be strongly destabilizing.

*Inhibition of the isofagomine lactam in Cex N126A*

In light of the above results, it is of interest to determine the importance of interactions of O2 of isofagomine lactam with the N126 residue. The $K_i$ values of the inhibitor with the Cex N126A mutant were determined and compared with equivalent values for the inhibitors. Table 6.3 shows $K_i$ values for the xylobiose derived inhibitors with wild-type Cex and the N126A mutant, as well as values for the contributions of the interactions with N126 to inhibitor binding.

Table 6.3. Kinetic parameters for xylobiose-derived iminosugars and $p$-nitrophenyl $\beta$-cellobioside with wild-type Cex and the N126A mutant. (1= isofagomine lactam; 2= imidazole; 3= lactam oxime; 4= nojirimycin; 5= isofagomine; see also Chapter 5)

<table>
<thead>
<tr>
<th>$K_i$ (µM)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>PNP $\beta$-cellobioside</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild type</td>
<td>0.34</td>
<td>0.15*</td>
<td>0.37*</td>
<td>5.8*</td>
<td>0.13*</td>
<td>$k_{cat}/K_m = 26.3 \text{ s}^{-1}\text{mM}^{-1}$</td>
</tr>
<tr>
<td>N126A</td>
<td>2000</td>
<td>200</td>
<td>1300</td>
<td>4800</td>
<td>1.6</td>
<td>$k_{cat}/K_m = 0.0044 \text{ s}^{-1}\text{mM}^{-1}$</td>
</tr>
<tr>
<td>$\Delta \Delta G$ (kcal mol$^{-1}$)</td>
<td>-5.3</td>
<td>-4.4</td>
<td>-5.0</td>
<td>-4.1</td>
<td>-1.5</td>
<td>$\Delta \Delta G^\circ = 5.4 \text{ kcal mol}^{-1}$</td>
</tr>
</tbody>
</table>

* Data taken from Williams et al. 2000.

As can be seen, for compounds that have a hydroxyl at the 2 position, binding to the N126A mutant is considerably weaker than to wild type enzyme, as expected due to the loss of important hydrogen bonding in the N126A mutant. By contrast, binding of
isofagomine to the mutant is compromised to a much lesser extent, consistent with the absence of an interaction at that position. Significantly, the consequence of the mutation of N126 upon binding of the isofagomine lactam is consistent with that observed for inhibitors that contain a 2-hydroxyl group. This strongly implies that there are similar interactions in the two cases and thus provides additional evidence that the isofagomine lactam binds in its iminol form. It is also interesting to compare these results with the effect of mutating N126 upon catalysis, as reflected in $k_{cat}/K_m$ values. The loss in transition state stabilization observed ($\Delta \Delta G^f = 5.4$ kcal mol$^{-1}$) is very similar to that seen for binding of these inhibitors, implying that these inhibitors are, at least in part, mimicking the reaction transition state. However, a more detailed analysis will be required to properly probe this behavior. The isofagomine lactam may therefore represents an example of a possible new class of potent glycosidase inhibitors. Work is continuing to explore the generality of this approach.

* NB: The kinetic analysis described in this paragraph was performed by Dr. S. J. Williams and J. Wicki, in the laboratory of Dr. S. G. Withers, Dept. of Chemistry, UBC, Vancouver, as part of the P.E.N.C.E. Theme F collaborative effort.
References


Concluding Remarks
and
Future Directions
**T. maritima xylanase 10A**

The structure of the family 9T. maritima carbohydrate binding module CBM9-2 represents the first three-dimensional structure of a cellulose-interacting protein complexed with its substrate. The two Trp residues that are stacking the ligand do not appear to dictate an absolute conformation to the binding, however the orientation of the plane of the ring is well-defined. Since glucose in ring conformation contains all equatorial hydroxyl groups, the two faces of the ring are not distinguishable from each other, and neither face is particularly hydrophobic. Therefore, the Trp residues likely provide stabilizing van der Waals interactions orienting the plane of the ligand into the binding site, and preclude sugars from binding that contain axial hydroxyl groups on their rings. The electrostatic interactions in the binding pocket of CBM9-2 render the protein specific for the reducing end of glucose moieties. This is in apparent contrast to known carbohydrate binders (e.g. in the lectin family), in which a single Trp residue is often found in the binding site that specifically recognizes the most hydrophobic face of the ring structure, thereby providing an extra element in substrate specificity.

Preliminary results of the crystal complexes of CBM9-2 with maltose (Glu-\(\alpha\)-1\(\rightarrow\)4-Glu), lactose (Gal-\(\beta\)-1\(\rightarrow\)4-Glu), cellotetraose and xylopentaose reveal the binding interactions for the reducing end pyranosides of these compounds to be as seen in the CBM9-2 glucose complex. In the case of the oligosaccharides cellotetraose and xylopentaose, only density for the first two sugars could be seen, the remaining units are presumed disordered. This further confirms the specificity of CBM9-2 to be for reducing end disaccharides in particular.

The unique orientation of the cellobiose moiety is not easy to comprehend in the context of the other ligand complexes, which all bind in the same fashion as glucose. A data set of CBM9-2 with cellobiose was recently collected to 1.15 Å resolution, which should allow for the refinement of this complex to great accuracy, followed by careful analysis of the binding interactions. It is hoped that docking studies and energy
calculations for the binding of these complexes will reveal a convincing rationale for the observed binding modes.

The identity of the coordinated metal ions can be unambiguously determined by either atomic absorption techniques or fluorescence scanning at the X-ray absorption edge for calcium and other possible candidates such as magnesium and cadmium.

It will be of interest to determine the binding specificities and crystal structure of the second family 9 cellulose-binding module of the \textit{T. maritima} xylanase 10A gene product, CBM9-1. It is not known whether this module contains any metal ligands, but the sequence alignment of the two domains does point to conservation in these regions. The construct encoding this domain has been cloned, expressed and purified, and small crystals have been grown in similar conditions as CBM9-2. Perhaps some insight will be gained in terms of complementary binding profiles of these two ancillary domains. Though alignment of the two proteins reveals sequence identity for 26\% of the residues and sequence homology of 43\%, the two Trp residues in the binding site were surprisingly not conserved.

Clearly, structural aspects of cellulose binding is still largely a dark area, and much remains to be discovered before a more complete picture can be drawn of the collaboration between substrate binding modules and their catalytic counterparts.

\textbf{Cex, Cex mechanism and Implications for Inhibitor Design}

The catalytic domain of Cex has been extensively studied both crystallographically and kinetically. Great insights have been gained since the discovery of the cex gene in 1988 (Gilkes \textit{et al.}, 1988) and the solution of its crystal structure in 1994 (White \textit{et al.}, 1994). These efforts then led to the first crystallographic observation of the catalytic covalent enzyme-substrate intermediate, which strongly supported the proposed existence of this species along the reaction coordinate of retaining glycosidases (White \textit{et al.}, 1996).
The work cited in this thesis describes further details of the reaction mechanism as applied to Cex and family 10 glycosyl hydrolases in particular, and, to some extent, retaining glycosidases in general. In combination with studies on other enzymes of the clan GH-A a general mechanistic model for the glycosylation step in these enzymes might look as follows:

A crystal structure of an active site spanning complex in wtCex, possibly revealing the Michaelis complex, is still elusive and would be of great interest to pursue. However, in the mean time, an extrapolation may be warranted based on the crystallographic observation of this species in Cel5A (family 5) and endoglucanase I (family 7) (Davies et al., 1998; Sulzenbacher et al., 1996).
Inhibitor design aided by structural analysis of such compounds in enzyme bound form (so-called ‘rational drug-design’) has the advantage of being an iterative process from which information can be used to improve on the potency of these inhibitors as well as further our knowledge of the mechanism of the reaction being inhibited. In the case of Cex, inhibitors were designed to mimic the proposed transition state in shape and charge, in addition to retained stereochemical characteristics of the natural substrate (xylan) to optimize ligand-protein interactions. This approach proved fruitful, such that nanomolar inhibition was found in all but one of the xylobiose-derived inhibitors against Cex. Surprisingly, these inhibitors were not as potent against the family 11 xylanase, Bcx. The next generation of inhibitors might benefit from incorporating aspects of the Bcx mechanism, which assumes a boat conformation as a covalent catalytic intermediate. Since the crystal structure of this enzyme has been solved, it should be possible to follow a deductive path similar to Cex in inhibitor design.

Hopefully, some of the wisdom gained by the work described in this theses, as well as the work cited, will lead to a better understanding of glycosidase mechanism and glycosidase inhibition.
References


Appendix A

Nomenclature for Cyclic Pyranoside Conformations
**Boat conformations**  The reference plane is defined by two sides of the boat, with its variants both above or below the defined plane, e.g. $^{25}\text{B}$ and $B_{0.3}$.

Skew conformations  Skew conformations are defined by a reference plane containing three adjacent atoms and a non-adjacent atom. The reference plane (of which there are two possible) is chosen such that the lowest possible atom number is exo-planar. Skew boats form mid-way between two related boat conformations, e.g. $^{14}\text{B} \rightarrow ^{1}\text{S}_5 \rightarrow B_{2.5}$.

Half-chair conformations  Four adjacent co-planar atoms form a half-chair conformation, with the two exo-planar variants designated below and above the plane.
Six-membered sugar rings assume many discrete or transitory conformations in solution or during chemical reactions and catalysis. The following is a synopsis focusing on the nomenclature to describe and distinguish the most common occurrences of pyranoside ring forms. The representations are of idealized rings and are mostly intended to be descriptive.

The approximate conformation of a six-membered carbohydrate ring is depicted by an italic, capital letter describing the overall ring shape (the 'descriptor'), annotated by the variants of that ring in numbers, designating the atoms that lie outside of the defined plane for that conformation. The variants that lie 'above' the plane are written left of the descriptor in superscript, and the variants 'below' the plane are written in subscript on the right side of the descriptor, e.g. $^4C_1$.

The following descriptors apply to six-membered rings: $C$ – for chair (preferred conformation); $B$ – for boat; $S$ – for skew; $H$ – for half-chair; $E$ – for envelope and $I$ – for twist conformations.

**Chair conformations** A plane is defined by two parallel ring sides, such that the lowest number carbon atom is exo-planar. For pyranosides there are two possibilities, namely $^1C_4$ and $^4C_1$.
**Envelope conformations** Envelopes are rare in six-membered rings, and form when five adjacent atoms become co-planar. The remaining atom is defined either above or below the plane.
Appendix B

A special case of MAD: pseudo-MIR using MAD and native datasets
Anomalous scattering occurs when the energy of and X-ray approaches the frequency of oscillations in a bound electronic orbital. This results in a resonance condition perturbing the normal (or Thomson) scattering of an element. Mathematically, the scattering factor is described to gain an anomalous component, which is, by definition, dependent on the wavelength of the incoming beam, and can be subdivided into a real ($f'$) and an imaginary ($if''$) component:

$$f = f_0 + f' (\lambda) + if'' (\lambda)$$

An absorption profile of anomalous scattering as a function of X-ray wavelength then shows that $f'$ reaches its peak value at the point where $f'$ is at its inflection point. This wavelength is termed the absorption edge of an anomalously scattering element. Similarly, $f''$ is at its inflection point when $f'$ reaches a minimum:

Anomalous scattering factor spectrum for selenium around its absorption edge as a function of X-ray energy expressed in wavelength (Å). The imaginary component $f''$ is drawn in the upper curve, and the real component $f'$ in the lower curve. From: Hendrickson, 1991.
The Multi-wavelength Anomalous Dispersion (MAD) technique of X-ray structure determination takes advantage of the fact that, at and around the X-ray absorption edge of an anomalous scatterer, there are measurable differences of the real and imaginary components of the scattering factor, from which phase information can be obtained. Generally, datasets are measured at the absorption edge (f'{-} max, or "peak"), at its inflection point (f'-min, or "edge") and at one or two wavelengths remote from the previous two energies ("remote").

Traditionally, phase information has been extracted using the least squares method implemented in the program MADSYS (Hendrickson et al., 1988). In this method, phase differences between the anomalously scattering atoms and all atoms in the structure are estimated by solving for the locations of the anomalous scatterers from a set of equations based on the theoretical contributions of the real and imaginary components from the heavy atoms (Hendrickson, 1991).

Another, faster approach has been developed to derive phase information from a MAD experiment by approximating this data to the mathematically simpler isomorphous replacement problem, as is implemented in the program package Solve (Terwilliger & Berendzen, 1999). To do this, two assumptions have to be made and are reasonable for most straightforward MAD experiments:

1. At different wavelengths the anomalous scatterering contribution varies in magnitude, but not in phase. This implies the anomalous scatterers in a crystal all have the same scattering factor.

2. The contribution to the structure factor from anomalous scatterers is small compared to all atoms combined.

In a MAD diffraction experiment one typically measures the difference between structure factor amplitudes for Bijvoet pairs of reflections at three or four discrete wavelengths (ΔF_{th} = |F^{+}(λ_{ab}) - F^{-}(λ_{ab})|), which arises from the imaginary (f') contribution to the scattering, and the dispersive differences of structure factors between wavelengths.
dependent on the real component \( f' \) (\( \Delta F_{\Delta} = |F_i| - |F_j| \)). The Bijvoet and dispersive differences contain complementary information since their phases are orthogonal to each other, and thus, if measured carefully, should suffice to solve for the exact locations of the anomalous scatterers in the unit cell.

In order to convert MAD data to an isomorphous replacement problem, the measured variables are expressed in terms that approximate those in the standard SIR (and SAS) equations (for details see (Blundell & Johnson, 1976)). The derivation for this approximation has been described in detail in the literature and is summarized below (Terwilliger, 1994; Terwilliger, 1997):

In essence, an expression is sought relating the experimental multi-wavelength data to the parameters \( F_o, F_{\text{TOT}} \) and \( \Delta_{\text{ISO}} \) at a random standardized wavelength \( \lambda_o \), which correspond roughly to the variables \( F_p, F_{\text{PH}} \) and \( F_H \) in the classic SIR case (1).

\[
F_{PH} = F_P + F_H
\]

Firstly, the total contribution from the scatterers in the unit cell at a wavelength \( \lambda_i \) can be expressed as the sum of the contributions of the non-anomalous scattering of the anomalous scatterers (i.e. including the dispersive differences) and the structure factor from all other atoms.

\[
\overline{F}(\lambda_i) = |F_0 + F_H(\lambda_i)|
\]

From assumption #2 (2) is approximated to

\[
F_{\text{tot}}(\lambda_i) = F_0 + F_H(\lambda_i)\cos(\phi_0 - \phi_H(\lambda_i))
\]
Assumption #1 allows for conversion of the dispersive differences between datasets by appropriately adjusting the structure factor magnitudes, since the phases remain constant with varying wavelengths

\[ F_{n}(\lambda_i) = \frac{F_{n}(\lambda_0)f_0 + f'(\lambda_0)}{f_0 + f'(\lambda_0)} \]

Similarly, the anomalous differences between wavelengths are related as

\[ \Delta_{ANO}(\lambda_i) = \Delta_{ANO}(\lambda_0)\frac{f''(\lambda_i)}{f''(\lambda_0)} \]

from which Bijvoet differences can be estimated at a standard wavelength $\lambda_0$, using measurements at any other wavelength.

From (3) and (4) expressions can be found for $\Delta_{ISO}$ at $\lambda_0$

\[ \Delta_{ISO}(\lambda_0) = F_{H} \times \cos(\phi_e - \phi_i) \]

\[ \Delta_{ISO}(\lambda_0) = [f_0 + f'(\lambda_0)] \frac{\overline{F}(\lambda_i) - \overline{F}(\lambda_i)}{\overline{F}(\lambda_i) - \overline{F}(\lambda_i)} \]

which can be used, with experimental data, to express $F_o$, $F_{TOT}$ at the standard wavelength $\lambda_0$

\[ F_o = \overline{F}(\lambda_i) - \Delta_{ISO}(\lambda_0) \frac{f_0 + f'(\lambda_0)}{f_0 + f'(\lambda_0)} \]

\[ \overline{F}(\lambda_0) = F_o + \Delta_{ISO}(\lambda_0) \]

A major advantage of this approach is that it is possible to simultaneously import several isomorphous datasets, from MAD and single-wavelength experiments, and homogenize them into a pseudo-MIRAS problem. This can be especially useful when using selenomethionine substituted protein crystals, since it should generally not be a problem to generate isomorphous crystals of both protein forms. By adding a single-wavelength
non-seleomethionyl dataset to the three- or four-wavelength MAD datasets from the selenomethionine crystals, the scattering difference between sulfur and selenium can be exploited and better initial phase approximations (and hence better experimental electron density maps) can be obtained (Ramakrishnan & Biou, 1997).

The assumptions formulated above have been tested empirically for the impact they might have on the resulting estimated structure factors with which phase approximations are made, using a hypothetical ideal MAD dataset (errors less than 0.01%) and one with 4% experimental error in the measured reflections (Terwilliger, 1994). It was found that the errors introduced by the approximations were considerably smaller than those resulting from experimental error, with the majority of approximations for $F_o$ close to their ideal counterparts as were estimates obtained for $\Delta_{iso}$ values. Perhaps ironically, in cases where the MAD signal is less pronounced, and estimations are less accurate, the Solve method appears to be best at deducing initial phase approximations, and would be the method of choice for solving a protein structure at present.
References


