Catalytic strategies of glycoside hydrolases

Luis Petersen
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd
Part of the Biological Engineering Commons, and the Chemical Engineering Commons

Recommended Citation
http://lib.dr.iastate.edu/etd/11144

This Dissertation is brought to you for free and open access by the Graduate College at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Catalytic strategies of glycoside hydrolases

by

Luis Petersen

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Chemical Engineering

Program of Study Committee:
Peter J. Reilly, Major Professor
Mark S. Gordon
Monica H. Lamm
Brent H. Shanks
Olga A. Zabotina

Iowa State University
Ames, Iowa
2010

Copyright © Luis Petersen, 2010. All rights reserved.
# Table of Contents

## Chapter 1: Overview

1

## Chapter 2: Literature Review

5

- Classification and nomenclature used for glycoside hydrolases 5
- Classification of glycoside hydrolases 5
- Nomenclature of glycoside hydrolases 5
- The mechanistic features of catalysis by glycoside hydrolases 7
  - General mechanisms of glycoside hydrolases 8
  - Syn vs anti protonators 9
  - Oxacarbenium ion-like transition state 9
  - Conformational itineraries in different GHs 11
  - Cremer–Pople parameters 12
  - Pyranosyl ring conformations identified in GH families 13
  - Why do glycoside hydrolases distort the sugar ring in subsite –1? 18
  - Non-covalent interactions 22
  - Unresolved mechanistic issues 23
- GH8 endoglucanase and GH43 β-xylosidase 24
  - Enzymatic deconstruction of the plant cell wall 24
- GH8 endoglucanase 26
- GH43 β-xylosidase 29
- Golgi GH38 α-mannosidase and the N–linked glycosylation pathway 30
Overview of computational methods

Current challenges in modeling enzymatic reactions

The size of the system

Time-scale problem in molecular dynamics

Multiple reaction coordinates

QM/MM metadynamics

QM region

Molecular mechanics region and classical molecular dynamics

Metadynamics

References

Chapter 3: The Mechanism of Cellulose Hydrolysis by Inverting GH8

Endoglucanases: A QM/MM Metadynamics Study

Abstract

Introduction

Computational methods

Classical molecular dynamics

QM/MM molecular dynamics

QM/MM metadynamics

Results and discussion

Classical and QM/MM molecular dynamics simulations

QM/MM metadynamics simulation
Chapter 4. Transition-State Stabilization and Dynamics in a Glycoside Hydrolase

Family 8 Endoglucanase

Abstract

Introduction

Computational methods

Molecular dynamics simulations

Results

Discussion

Conclusion

References

Chapter 5: Mechanism of Xylobiose Hydrolysis by GH43 β-Xylosidase

Abstract

Introduction

Computational methods

Classical molecular dynamics and system setup

QM/MM molecular dynamics

QM/MM metadynamics
Chapter 7: Molecular Mechanism of the Glycosylation Step Catalyzed by Golgi α-Mannosidase II

Abstract 130

Introduction 131

Computational methods 134

Classical molecular dynamics simulations 134

QM/MM molecular dynamics simulations 135

QM/MM metadynamics simulations 136

Results and discussion 138

Metadynamics simulation of GMII 138

The FES and the minimal energy reaction pathway 139

The role of zinc and the partial deprotonation of the O2’–H02’ bond 144

Conformational itinerary of the glycon mannosyl ring along the reaction path 145

References 148

Chapter 8: Conclusions and Future Directions 152

Acknowledgements 155
Chapter 1: Overview

Glycoside hydrolases (GHs) are enzymes that catalyze the hydrolysis of the glycosidic bond between two carbohydrate residues or a carbohydrate unit linked to a non-carbohydrate aglycon unit. They are ubiquitous in Nature. About 1% of the genome of any organism encodes for GHs,¹ and they are involved in several processes that are essential for life on earth. They participate, for example, in cellular recognition processes; carbohydrate moieties are covalently linked to proteins and lipids, where they interact with other molecules to mediate recognition in and among cells. Post-processing of the attached carbohydrates, via addition and trimming of individual carbohydrate units, is crucial for this process, and GHs and glycosyltransferases are employed in this step. In addition, the hydrolysis of carbohydrate polymers, such as cellulose and starch, into smaller sugar units provides some of the required building blocks to construct biopolymers and obtain energy in organisms.

GHs are also important in industrial settings. The variety of roles that they play in the human body has allowed researchers to identify some of these enzymes as drug targets for a number of diseases including cancer and influenza.²,³ In addition, GHs found in other organisms have been employed to break down starch, cellulose, and hemicellulose to produce fermentable sugars that can be converted into biofuels.⁴-⁶

Scientists have gathered a great deal of information about GHs. We know the primary sequences of thousands of GHs, as well as their evolutionary relationships.⁷ Three-dimensional structural information is available for hundreds of GHs, and two general hydrolysis mechanisms applicable to most GH have been discovered (see below). Nevertheless, many details on the hydrolysis mechanism and mode of action of specific GHs remain unclear. For
instance, the oxacarbenium ion nature of the transition state (TS) of the hydrolysis reaction that occurs in glycoside hydrolysis is now common knowledge, but the structural features of the TS, and the extent to which these features depend on the enzyme’s active site, are still under investigation. This information is crucial for GHs identified as drug targets, as it can guide the design of molecules that resemble the TS and test them as drug candidates. In addition, a detailed understanding of the hydrolysis mechanisms and the molecular interactions that make GHs such remarkable catalysts can provide important insight to develop efficient catalysts to use in industrial settings.

Because of the short life span of a reaction’s TS and other important unstable states along the reaction pathway, it is very difficult to use experimental techniques to obtain structural information on these states. Nevertheless, modern computational techniques can now fill this gap and shed light into this problem.

This dissertation presents the results of computational investigations that give insight into the mechanistic details of a cellulase, an enzyme that breaks the interior glycosidic bonds of cellulose chains, a β-xylosidase that breaks xylosyl residues from the end of xylosyl polymer chains present in hemicellulose, and a clinically relevant Golgi α-mannosidase that is part of the N-linked glycosylation pathway, and has been identified as a target against breast, skin, and colon cancer.

The next chapter presents the literature review. It mainly focuses on the current knowledge of GH mechanisms. An overview of the computational methods and the current challenges of modeling enzymatic reactions is also presented at the end of Chapter 2. Chapter 3 contains the results of the explicit quantum mechanics/molecular mechanics (QM/MM) metadynamics simulation of the enzymatic hydrolysis of cellulose in a GH8 endoglucanase.
This provides insight into the enzymatic hydrolysis mechanism and the conformational itinerary (see below) of this cellulase. In addition, the structure of the TS complexed with the enzyme was obtained from this work. Chapter 4 extends the work from Chapter 3 to provide a detailed analysis of the non-covalent interaction of GH8 endoglucanase and the role of active-site amino acid residues in stabilizing the reaction TS. We take TS and Michaelis complex active-site structures of the enzyme obtained from the QM/MM analysis of the reaction and use high accuracy ab initio methods to measure the energetic contributions of individual amino acids to stabilize the TS. Chapter 5 deals with a hemicellulase: β-xylosidase from family GH43. We use QM/MM metadynamics to model the xylobiose hydrolysis performed by the enzyme. In Chapter 6, the twisting of scissile glycosidic bonds, imposed by GHs from 14 different GH families, was investigated by constructing hybrid (non-integral) QM::MM maps that measure the energy as a function of glycosidic bond twisting for eight different substrates and inhibitors of GHs. The scissile bond dihedrals observed in many GH crystal structures were overlaid on the maps to obtain an estimate of the strain imposed by the enzyme on the substrate. Turning from plant-cell wall degrading GHs, Chapter 7 presents a QM/MM metadynamics investigation of an α-mannosidase from family GH38 to obtain the detailed molecular mechanism of the enzyme. Concluding remarks and future directions are discussed in Chapter 8.

References


Chapter 2: Literature review

Classification and nomenclature used for glycoside hydrolases

Classification of glycoside hydrolases

The enzyme classification system of the International Union of Biochemistry and Molecular Biology is based on substrate specificity, and the Enzyme Commission (EC) number corresponding to glycoside hydrolases (GHs) is EC 3.2.1.x. However, the structural relationship (if any) between specific enzymes is not revealed by this classification system. These relationships are important because enzyme catalytic mechanisms are better related to their three-dimensional (tertiary) structures than to their substrate specificities. Alternatively, the carbohydrate-active enzyme (CAZy) classification system for GHs (and other carbohydrate-active enzymes) is based on amino acid sequence comparisons.\textsuperscript{1,2} The enzyme’s primary sequence determines its tertiary structure. Therefore, GHs that have similar primary sequences (homologous enzymes) also share similar tertiary structures. The active site and catalytic mechanism of sequence-related enzymes are also very similar, which permits extrapolation of structural and mechanistic data between related GHs.

Based on this classification system, GHs are divided into >100 families.\textsuperscript{1} Each GH family contains enzymes with different specificities, but similar tertiary structures and mechanisms. The CAZy classification system also helps to reveal evolutionary relationships between enzymes and even between entire GH families. Families with similar tertiary structures (folds) are grouped into clans. There are 14 clans of GH families to date.\textsuperscript{1}

Nomenclature of glycoside hydrolases

The sugar nomenclature followed throughout this paper is summarized in Figure 2.1. This
shows a glucopyranosyl ring, but the nomenclature does not change for other sugars. The anomeric carbon atom is referred to as C1, and the numbering progresses so that the aldo- or keto-oxygen atom in the molecule has a high number.

![Figure 2.1. Nomenclature of a glycopyranosyl ring.](image)

Carbohydrate-active enzymes have a unique nomenclature. Many of these enzymes have very long binding sites capable of accommodating a number of sugar residues. The region of the enzyme that binds one residue of the chain is called a subsite. The numbering of subsites begins with the two sugar rings that share the glycosidic oxygen atom of the scissile bond. These are numbered –1 and +1, pointing toward the nonreducing and reducing ends of the chain, respectively. The scissile bond itself belongs to the ring in subsite –1. Toward the reducing end, subsite numbering increases (+1, +2, +3…etc.) and toward the nonreducing end, subsite numbers decrease (–1, –2, –3…etc.). Subsite –1 is the most important, as it holds the sugar ring that bears the scissile bond and contains the anomeric carbon atom that is the center of the hydrolysis reaction.

To refer to specific atoms that belong to pyranosyl rings in subsite –1, an apostrophe is added to the atom label shown in Figure 2.1. For example, the anomeric carbon atom is designated as C1’, the ring oxygen atom is O5’, and so on. The glycosidic oxygen atom of
the scissile bond belongs to the next subsite (subsite +1), and no apostrophe is required.

**The mechanistic features of catalysis by glycoside hydrolases**

The most general catalytic mechanisms for GHs were proposed decades ago. Since then, several revisions have been made and research has focused on important details about the action of GH enzymes. In this section we review the general mechanisms of most GHs, the experimental and computational evidence on the hydrolysis reaction transition state (TS), and the enzyme-imposed distortion of pyranosyl rings to facilitate catalysis. The section finishes with a summary of unresolved mechanistic issues that are currently under investigation.

**General mechanisms of glycoside hydrolases**

Most GHs follow one of two possible general mechanisms depicted in Figure 2.2 (but there are some rare exceptions). These are called the retaining and inverting mechanisms, based on the stereochemical outcome of the hydrolysis, either retention or inversion of the product’s anomic configuration. Both catalytic mechanisms employ two carboxylate groups, provided by either a glutamate or aspartate amino acid residue (Figure 2.2). One of the carboxylate groups is located underneath the anomic carbon atom, while the other one is positioned laterally above the plane of the sugar ring. The role of these carboxylate amino acid residues differs slightly for retainers and inverters. GHs that retain the product’s anomic configuration perform the hydrolysis reaction in two steps. In the first step (usually called the glycosylation step), one of the enzyme’s carboxylate groups, the nucleophile, forms a covalent bond with the anomic carbon atom of the pyranosyl unit in the active site, while a second carboxylate group, regularly known as the acid/base catalytic residue, donates a proton to the glycosidic oxygen atom of the leaving group (Figure 2.2A). The second step is
Figure 2.2A. GH retaining mechanism.

Figure 2.2B. GH inverting mechanism.
the cleavage of the enzyme–substrate intermediate. The carboxylate group that acted as proton donor in the first step now coordinates a water molecule and assists its nucleophilic attack to form the product. This is followed by product release, emptying the active site for the next catalytic cycle.

The mechanism of inverting GHs occurs in only one step (Figure 2.2B). One carboxylate group donates its proton to the leaving group’s glycosidic oxygen atom while the second carboxylate group assists the nucleophilic addition of a water molecule to the anomeric carbon atom. This carboxylate group collects the proton and the inversion reaction is completed.

**Syn vs anti protonators**

The wealth of structural information of GHs complexed with natural substrates and TS analogs reveals that the proton donor is positioned in a lateral position with respect to the glycosidic oxygen atom of the scissile bond to facilitate protonation. If the proton donor is located on the same side of the endocyclic oxygen atom (the ring oxygen atom), the enzyme is called a *syn* protonator. If the proton donor is positioned on the opposite side of the ring oxygen atom the enzyme is an *anti* protonator. This observation is useful for comparisons between GH active sites.

**Oxacarbenium ion-like transition state**

The TS of the hydrolysis reaction in glycopyranosides have been studied with substrate inhibitors, kinetic isotope effects (KIEs), and computational methods.

There is substantial experimental and theoretical evidence indicating that nucleophilic substitutions on glycosides are on the borderline between a highly dissociative (\(A_N D_N\)) mechanism and a (\(D_N^* A_N\)) mechanism. The \(A_N D_N\) mechanism consists of a SN2-type TS with
participation of both leaving group and nucleophile at the TS. In the $D_N^*A_N$ mechanism the
leaving group dissociates with no participation of the nucleophile, forming a very unstable
cation intermediate that is trapped by the nucleophile immediately after its formation.

TS structures are difficult to determine experimentally. This is because TS lifetimes are
very small ($\sim 10^{-12}$ s). Experimental information that has given the most insight into the
nature of the TS in GHs has come mainly from measurement of KIEs. In this technique, the
rate of reaction of the natural substrate is compared to an isotopically labeled substrate at a
specific atom close to the reaction center. A substitution of a particular atom of the substrate
by its heavier isotope leads to a lower vibrational frequency, which, in turn, leads to an
increased reaction rate (because of an increase in activation energy due to a lower zero-point
energy). A ratio between the rates of the natural substrate over the isotopically labeled sub-
strate gives the KIE. The measured KIE can be compared to values of known reactions to
determine the type of mechanisms and structural information on the TS.

Evidence from measuring KIEs shows that TSs in GHs (and glycoside substitutions in
general) have significant oxacarbenium ion character.\textsuperscript{8-12} The positive charge is partly deloc-
alized throughout the sugar ring, but the anomeric carbon and ring oxygen atoms bear most
of it. The C1–O5 bond significantly shortens and adopts double-bond character, as the ano-
meric center becomes sp2-hybridized. This imposes the requirement that at the TS, the C1,
C2, C5, and O5 atoms of the pyranosyl six-membered sugar ring must lie on a plane. Sugar
ring conformations (see below) that comply with this requirement are similar to $4H_3$, $4E$, $E_3$,
$3H_4$, $3E$, $4E$, $B_{2,5}$, and $2.5B$. Therefore, at the TS a pyranosyl ring must have one of these con-
formations (or close to one of them).

Some of the KIE studies on glycoside substitutions are done in solution. Nevertheless,
three studies of hydrolytic reactions inside a GH enzyme are worth mentioning. In a recent work on a retaining β-xylosidase, Withers and co-workers measured the α-deuterium-secondary KIE on the enzymatic hydrolysis reaction.\textsuperscript{11} Their results showed that the hydrolysis proceeds through a mechanism similar to an A\textsubscript{N}D\textsubscript{N}. The TS was described as oxocarbenium ion-like, with both the nucleophile and leaving group positioned substantially far away from the anomeric carbon atom, leading to a TS with highly cationic character.

Berti and coworkers\textsuperscript{12} measured primary and secondary KIE on two retaining GHs, an α- and a β-glucosidase. They found that the α-glucosidase performs hydrolysis through an D\textsubscript{N}*A\textsubscript{N} mechanism while the β-glucosidase uses an A\textsubscript{N}D\textsubscript{N}. The same reactions were performed under acid catalysis in solution and both reactions follow a D\textsubscript{N}*A\textsubscript{N} mechanism.

KIE studies on enzymatic hydrolysis of glycosides include that of two α-D-glucopyranosyl pyridinium salts with yeast α-glucosidase.\textsuperscript{13} Here again the measurement of the KIE indicates a highly dissociative A\textsubscript{N}D\textsubscript{N}-like mechanism. In addition, the TS of the enzymatic reaction was compared to that in solution. The main difference was that the nucleophile was slightly involved in the enzymatic reaction TS, while it is not in the reaction in solution. This suggests that enzymes not only stabilize the TS of the reaction they catalyze; they change it.\textsuperscript{9}

Further evidence of an oxocarbenium ion-like TS in GHs comes from inhibition studies. Binding affinity measurements of non-hydrolyzable TS analogs shows tighter binding to the active site of GHs when they are structurally and electronically similar to an oxocarbenium ion species.\textsuperscript{5,7}

**Conformational itineraries in different GHs**

As mentioned in the last section, the sp\textsubscript{2}-hydridized C1–O5 bond of the glycopyranosyl
rings at the TS imposes planarity on the C2–C1–O5–C5 ring atoms. This forces a distortion of the sugar ring, away from its most stable chair form. The ring distortion is not only associated with the TS but is also observed in equilibrium enzyme–substrate complex structures, demonstrating that the sugar ring bearing the scissile bond (the one bound in subsite –1) changes its conformation along the reaction coordinate. We refer to these conformational changes as the glycopyranosyl itinerary (or conformational itinerary).

Tertiary crystal structures of GHs complexed with substrate or inhibitors show pyranosyl rings in a variety of distorted conformations, suggesting that enzymes can follow different itineraries. In this section, the basis for the distortion and its importance will be discussed, together with the results of relevant studies.

The number of possible conformations adopted by six-membered rings is astonishing. Therefore, before proceeding it is important to introduce a framework for their classification.

**Cremer-Pople parameters**

Cremer and Pople showed that all the possible conformations of six-membered rings can be mapped onto a spherical representation that is based on three coordinates (\(Q\), \(\phi\), and \(\theta\)), analogous to spherical coordinates, that depend only on the positions of the ring atoms. In the spherical representation, the chair forms are located at the north (\(^1C_1\)) and south (\(^1C_4\)) poles. All the boat and skew–boat conformations are located along the equator, and all the half-chairs and envelopes are located between the poles and equator of the sphere.

A projection of the sphere in two dimensions, viewed from the north pole, is widely used. This representation is sometimes referred to as the Stoddart diagram (Figure 2.3).
Pyranosyl ring conformations identified in GH families

There is now overwhelming structural data that reveals equilibrium enzyme–substrate complexes where the sugar ring in subsite –1 is in a distorted conformation.\textsuperscript{6,7,14,16-19} Tables 1 and 2 show the data of ring distortion in $\beta$- and $\alpha$GHs, respectively, obtained from crystallographic and computational studies.

There are several tools to investigate the itineraries harnessed by specific GHs. One of the most used is x-ray data from crystallographic studies. Here, the challenge is to “trap” the enzyme–substrate complex by impairing the enzyme or the substrate to prevent catalysis.

Site-directed mutagenesis is commonly used to mutate one of the catalytic amino acid residues and remove the catalytic power of the enzyme. Then, the enzyme can be crystallized with its natural substrate in the active site, and the tertiary structure of the complex can be
obtained. Sometimes the interactions between the mutated amino acid and the substrate are important for the ring distortion, in which case one may obtain unreliable results.

The Michaelis complex column shows some of the distorted pyranosyl conformers, some of which have been obtained through x-ray crystallography (see below). For wild-type enzymes (or mutants that retain activity), the product’s ring conformation can also be obtained.

The covalent intermediate in the middle step of retaining GHs has also been observed with x-ray crystallography. This was done using a very interesting kinetic strategy that involved modification of both the substrate and the enzyme. The proton donor of the enzyme was mutated, while the substrate was a 2-deoxy-fluoropyranosyl compound bonded to a very good leaving group. The rationale behind these modifications was the following: The rate-determining step in retaining GHs is the formation of the covalent intermediate, the glycosylation step, while the second step to form the product (deglycosylation), occurs faster. To trap the covalent intermediate, it is required to make the glycosylation reaction fast, but to prevent deglycosylation from happening. The use of a very good leaving group causes the glycosylation reaction to occur without requiring a proton donor (which is also required for the second step). Furthermore, the oxocarbenium ion-like TS has an electron deficiency at the anomeric center, which becomes even more pronounced if one of the substituents of the pyranosyl ring is fluorine, the most electronegative element. The introduction of fluorine to the substrate, together with the removal of the acid/base catalytic amino acid, prevents the deglycosylation reaction from happening quickly, leaving enough time for its crystallization and x-ray three-dimensional structure to be determined.

If the ring distortion of both the Michaelis complex and the covalent intermediate (or product for inverters) of a specific GH are known, the conformation of the TS can be
inferred/hypothesized with help of the Cremer–Pople sphere (Figure 2.3).\textsuperscript{14,21}

Trapping substrates for their observation with x-ray crystallography has been done for several enzymes. Here we present the work of a selected few. Ducros and coworkers studied a retaining GH26 $\beta$-mannanase.\textsuperscript{22} They impaired the enzyme and trapped the Michaelis complex with a distorted $\beta^{\text{1}S_5}$ mannose conformation in subsite $-1$. The covalent intermediate showed an $\alpha^{-\text{O}S_2}$ conformation. This led the researchers to hypothesize a $B_{2,5}$ conformation of the TS. The $B_{2,5}$ conformation is flanked by both the $1S_5$ and $0S_2$ conformations in the Cremer–Pople sphere (Figure 2.3). Even further, they predicted that other $\beta$-mannanases would follow the same (or very similar) itineraries.

A partial confirmation of this prediction was obtained in their later work with a GH2 mannanase, where they crystallized the wild-type (non-modified) enzyme complexed with many inhibitors.\textsuperscript{7} The ones that showed better binding were in $B_{2,5}$ conformations, suggesting that the $B_{2,5}$ is indeed a likely TS conformation in both GH2 and GH26.

Similar examples of itineraries in other GH enzymes with different specificities have been published. Varrot and co-workers crystallized the Michaelis complex and covalent intermediate of a retaining GH5 endoglucanase.\textsuperscript{23} The Michaelis complex was seen in a distorted $\beta^{\text{1}S_3}$ conformation, while the covalent intermediate showed an undistorted $\alpha^{-4C_1}$ conformation. Again, the $^4H_3$ conformation is flanked in the Cremer–Pople sphere by those observed in the crystal structures ($^1S_3$ and $^4C_1$), which makes this conformation a likely candidate for the TS.

It can be seen in Table 2.1 that the $\beta^{\text{1}S_3}$ conformation is the most common Michaelis complex, and also that these complexes also have an $\alpha^{-4C_1}$ product or covalent intermediate.
Table 2.1. Selected X-ray crystal and computational structures of GHs complexed with distorted β-glycopyranosyl rings.

<table>
<thead>
<tr>
<th>Family</th>
<th>Enzyme action</th>
<th>Mechanism</th>
<th>Michaelis complex</th>
<th>TS</th>
<th>Product/CI.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>β-mannosidase</td>
<td>retaining</td>
<td>β⁻¹S₃</td>
<td>B₂S₅</td>
<td>α⁻⁰S₂</td>
</tr>
<tr>
<td>5</td>
<td>1,4-β-endoglucanase</td>
<td>retaining</td>
<td>β⁻¹S₃</td>
<td>—</td>
<td>α⁻²C₁</td>
</tr>
<tr>
<td>6</td>
<td>cellobiohydrolase</td>
<td>inverting</td>
<td>β⁻¹S₀</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>7</td>
<td>1,4-β-endoglucanase</td>
<td>retaining</td>
<td>β⁻¹S₃</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>8</td>
<td>1,4-β-endoglucanase</td>
<td>inverting</td>
<td>β⁻²S₂/²S₀</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>9</td>
<td>cellobiohydrolase</td>
<td>inverting</td>
<td>β⁻¹E</td>
<td>—</td>
<td>α⁻⁴C₁</td>
</tr>
<tr>
<td>10</td>
<td>xylanase</td>
<td>retaining</td>
<td>β⁻¹S₃</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>11</td>
<td>xylanase</td>
<td>retaining</td>
<td>—</td>
<td>α⁻²⁵B</td>
<td>—</td>
</tr>
<tr>
<td>12</td>
<td>1,4-β-glucanase</td>
<td>retaining</td>
<td>β⁻¹S₃</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>16</td>
<td>1,3,1,4-β-glucanase</td>
<td>retaining</td>
<td>β⁻¹S₃</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>18</td>
<td>chitinase</td>
<td>retaining</td>
<td>β⁻¹⁴₁⁴B/¹⁴S₃</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>20</td>
<td>chitobiase</td>
<td>retaining</td>
<td>β⁻¹S₃</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>26</td>
<td>β-mannanase</td>
<td>retaining</td>
<td>β⁻¹S₃</td>
<td>α⁻⁰S₂</td>
<td>α⁻⁰S₂</td>
</tr>
<tr>
<td>44</td>
<td>endoglucanase</td>
<td>retaining</td>
<td>β⁻¹S₃</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>74</td>
<td>oligoxyloglucan CBH</td>
<td>inverting</td>
<td>β⁻¹S₄⁴H₃</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>94</td>
<td>cellobiose phosphorylase</td>
<td>inverting</td>
<td>β⁻¹S₃</td>
<td>⁴H₃</td>
<td>α⁻⁴C₁</td>
</tr>
</tbody>
</table>

Table 2.2. Selected X-ray crystal and computational structures of GHs complexed with distorted α-glycopyranosyl rings.

<table>
<thead>
<tr>
<th>Family</th>
<th>Enzyme action</th>
<th>Mechanism</th>
<th>Ligand</th>
<th>TS</th>
<th>Product/CI.</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>cyclodextrin transferase</td>
<td>retaining</td>
<td>α⁻⁴C₁</td>
<td>—</td>
<td>β⁻¹C₁</td>
</tr>
<tr>
<td>15</td>
<td>glucoamylase</td>
<td>inverting</td>
<td>α⁻⁴C₁</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>33</td>
<td>sialidase</td>
<td>retaining</td>
<td>α⁻¹B₂₅</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>38</td>
<td>α-1,3 –1,6-mannosidase</td>
<td>retaining</td>
<td>α⁻¹H₅</td>
<td>—</td>
<td>α⁻¹S₅</td>
</tr>
<tr>
<td>47</td>
<td>α-1,2-mannosidase</td>
<td>inverting</td>
<td>α⁻¹S₁</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>77</td>
<td>amylomaltase</td>
<td>retaining</td>
<td>α⁻¹C₁</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

It is likely that all these GHs have a $^4H₃$-like TS conformation and therefore that the $β⁻¹S₃ → ^4H₃[TS] → α⁻⁴C₁$ itinerary is the most common in GHs.

X-ray crystallography is very useful tool to elucidate GH itineraries, but it is limited. In many cases the enzyme is mutated, and this can affect the rest of the active site and the distortion of sugar ring –1. Substrates are sometimes non-natural, which may also affect distortion. Most importantly, the TS is not a stable intermediate and cannot be observed by these techniques. The TS conformation has to be predicted based on equilibrium data, and
not by direct measurement. These limitations can be overcome with computational tech-
niques (although they have their own limitations), and they have been used to study a variety
of GHs.

Biarnés and co-workers\textsuperscript{24} have done first-principle QM/MM studies on a GH16 endogluc-
ucanase. The whole enzyme substrate and solvent (water) were explicitly included. They
showed that the most stable Michaelis complex conformation adopts a $\beta^{-1}S_3$ conformation in
subsite $-1$. The other stable conformation in the dynamics calculation was the $^4C_1$ chair form.
This is consistent with an itinerary similar to $\beta^{-1}S_3 \rightarrow ^4H_3[TS] \rightarrow \alpha^{-4}C_1$.

Other less expensive computational tools, such as classical molecular dynamics (MD)
and automated docking, have been used for this purpose. Fushinobu and coworkers carried
out both classical MD and automated docking simulations on a GH94 cellobiose phosphory-
lase.\textsuperscript{6} The crystallographic structure had only phosphate and glycerol in the active site. Auto-
mated docking was used to model a cellobiose molecule with its nonreducing-end sugar res-
didue distorted to different conformations, with their energies and positions within the active
site of all the different conformers measured. The cellobiose molecule with the nonreducing-
end $\beta^{-1}S_3$ sugar ring had the lowest binding energy, suggesting that the Michaelis complex
adopts this conformation. The enzyme with the $\beta^{-1}S_3$ distorted substrate was then submitted
to classical MD simulations, which proved that this conformation was indeed the most stable
when bound to the active site. The eight possible ring conformations (those with planar C2–
C1–O5–C5 atoms) of the oxocarbenium-ion TS were docked in the enzyme active site. The
$^4H_3$ conformer binds tighter that the rest, suggesting a $\beta^{-1}S_3 \rightarrow ^4H_3[TS] \rightarrow \alpha^{-4}C_1$ itinerary
similar to that of many other GHs.
The itinerary of a GH47 α-mannosidase was also analyzed with automated docking by Mulakala and coworkers.\textsuperscript{25} All possible mannosyl conformations were submitted to automated docking, and forces were calculated to obtain insight on the sugar ring’s dynamics. A non-standard $\alpha^{-1}C_4 \rightarrow \alpha^{-3}S_1 \rightarrow \beta^{-4}E[TS]$ itinerary was proposed.

Some relevant observations have been made with respect to the different conformations found inside the enzymes. First, GHs that act on $\beta$-anomeric glycopyranoses distort the sugar ring in such a way that the scissile glycosidic oxygen is positioned in an axial or pseudo-axial orientation. For enzymes that catalyze hydrolysis in $\alpha$-anomeric glycopyranoses, the most stable (in solution) chair conformation, $\alpha^{-4}C_1$, has its scissile glycosidic oxygen is positioned in an axial position. Nevertheless, distortion is sometimes observed for these enzymes (see Table 2.2).

**Why do glycoside hydrolases distort the sugar ring in subsite –1?**

There are several possible advantages of distorting the sugar ring at the center of the reaction. These include pre-activation of the the substrate for catalysis by 1) receiving assistance from stereoelectronic effects;\textsuperscript{26,27} 2) positioning the ring in a conformation that is closer to the TS conformation;\textsuperscript{28} 3) imposing strain on the substrate to reduce the reaction barrier; and 4) removing steric hinderance from the $C1$–$H$ atom at the time of the nucleophilic attack. These points are further explained below.

Part of the substrate pre-activation for catalysis comes from stereoelectronic effects in the pyranosyl ring.\textsuperscript{26,27} Stereoelectronic theory dictates that, to facilitate the formation of the TS, one of the ring oxygen atom’s lone pairs must be oriented antiperiplanar to the scissile glycosidic bond (Figure 2.4A). As the scissile bond elongates to form the TS, the anomeric carbon
atom becomes electron-deficient (positively charged) and electron donation from the ring oxygen atom becomes important. If the ring conformation complies with stereoelectronic requirements, the ring oxygen atom’s lone pair can easily share its electrons with the antibonding orbital of the anomic carbon atom and form the partial double C1–O5 bond observed at the TS. This is called the antiperiplanar lone pair hypothesis (ALPH). Figure 2.4 shows the electronic distribution as the glycosidic bond breaks.

![Figure 2.4](image_url)

**Figure 2.4.** Steroelectronic effects in glycopyranosyl rings. A) pre-TS $\beta^{-1}S_5$, showing the oxygen atom’s lone pair in an antiperiplanar position with respect to the leaving group. B) $\beta^H_1$-TS conformation showing the oxygen atom’s orbital forming a partial double bond with the anomic carbon atom.

The pyranosyl pre-TS conformations that comply with the stereoelectronic requirement are different for $\alpha$- and $\beta$-anomeric glycopyranoses. If the glycosidic bond is in a $\beta$-configuration, the ALPH-compliant conformations are $1S_5$, $4B$, $1S_3$, $B_{3,0}$, $2S_0$ and $1C_4$. The ALPH-compliant conformations for $\alpha$-anomeric glycopyranoses are the complete opposite, $5S_1$, $B_{1,4}$, $3S_1$, $\Omega B$, $O S_2$, and most importantly $4C_1$, viewed from the Cremer–Pople sphere.

The $4C_1$ chair is the most stable of all conformations for both $\alpha$- and $\beta$-anomers. The $\alpha$-anomers have the advantage that its most stable conformation is also ALPH-compliant and already pre-activated for the anomeric substitution. On the other hand, $\beta$-anomers undergo a
conformational flip, away from the most stable \(^4C_1\) chair form, before proceeding with the anomeric substitution.

Another advantage of the ring distortion in \(\beta\)-anomers is the removal of steric hinderance between the incoming nucleophile and the C1–H atom. When a \(\beta\)-glycopyranose is in its relaxed chair conformation, the glycosidic bond is positioned in an equatorial position, and the C1–H atom takes the axial position. The nucleophilic attack occurs from underneath the C1 atom, which is blocked by the C1–H atom in the \(^4C_1\) conformation. ALPH-compliant conformations also remove this steric impediment.

Biarnés \textit{et al.} did first-principle DFT calculations of a fully solvated, \(\beta\)-glucopyranose molecule.\(^{28}\) They sampled all possible ring conformations, measured their energy, and calculated their structural and electronic features. As expected, \(\beta^{-4}C_1\) was the lowest-energy conformer. Also, ALPH-compliant conformers were lower in energy than non-ALPH-compliant ones (except for \(\beta^{-4}C_1\)). The ALPH-compliant conformations resemble structurally and electronically the oxocarbenium ion-like TS. They showed a significant elongation of the glycosidic bond, shortening of the C1–O5 bond, and positive charge build-up at the anomeric center. These results suggest that distortion brings the pyranose ring closer to the TS, structurally and electronically.

Most of the proposed itineraries are consistent with the electrophilic migration mechanism originally proposed for lysozyme.\(^{20}\) In this mechanism, the anomeric substitution proceeds by migration of the electron-deficient (electrophilic) anomeric carbon atom toward the nucleophile. Both the leaving group and the nucleophile are fairly rigid in the enzyme active site, so the atomic displacements needed to break the glycosidic bond and form the product are mainly carried out by the sugar ring in subsite \(-1\), via conformational changes.\(^{21}\) This is
better explained with the example in Figure 2.5. Here one can see the reactant is in the common distorted conformation $\beta^{-1}S_3$, which positions the anomeric carbon atom above the average plane of the ring. The leaving group is fairly fixed and the elongation of the glycosidic bond occurs by a downward movement of the C1 carbon atom, reaching the $^4H_3$ TS. A further downward movement of the C1 atom “traps” the nucleophile, which is also fairly rigid.

\[ \text{Figure 2.5. Electrophilic migration mechanism. Average positions of the catalytic acid and base are shown in yellow. Example of conformations adopted by the glycopyranosyl ring along the reaction coordinate are shown. The reactant is in black, the TS is in blue, and the product (covalent intermediate) is in yellow. It can be seen that the anomeric carbon migrates toward the nucleophile to perform the substitution.} \]

In conclusion, ring distortion seems to be a clever strategy of GHs that could provide 1) electronic assistance to facilitate formation of the TS, via ALPH; 2) a closer structural and electronic resemblance to the TS by elongating the scissile bond, shortening the C1–O5 bond, and increasing the positive charge at the anomeric center; 3) a solution to remove the steric hinderance between the C1–H atom and the nucleophile in $\beta$-glycopyranoses; 4) a solution to the rigid nature of the leaving group and nucleophile in the enzyme active site by
forcing an electrophilic migration mechanism; and 5) imposed strain on the substrate that can lower the reaction energy barrier.

**Non-covalent interactions**

It is common to cite the two mechanistically important carboxylate groups employed by GHs as the only catalytic machinery of these enzymes. Nevertheless, non-covalent interactions play a crucial role in enzymatic catalysis. For the most part, enzymes achieve outstanding rate accelerations based on their ability to stabilize the TSs of the reactions that they catalyze. Detailed analysis of enzyme–substrate interactions that contribute to TS stabilization is required to understand enzyme catalytic strategies.

Non-covalent interactions are particularly important in GHs. The carbohydrate substrates that they hydrolyze have multiple hydroxyl groups that make hydrogen bonds with the amino acid residues in the enzyme active site. Some of these hydrogen bonds tend to strengthen at the TS, partially relieving the electron deficiency (i.e. stabilizing) of the oxocarbenium ion-like TS.19,32

Withers and coworkers experimentally measured the energetic contribution of the non-covalent interactions between the substrate’s hydroxyl groups and the active site of a β-glucosidase.33 A variety of substrates lacking specific hydroxyl groups were synthesized and kinetically tested with the enzyme. The difference in kinetic parameters between regular and modified substrates was used to calculate energetic contributions of non-covalent interactions. The modified substrates affected binding of the TS much more than that of the ground state. The interactions with the hydroxyl group at the 2-position (C2–OH) were the strongest, providing at least 18 kJ/mol of TS stabilization. Interactions with the rest of the hydroxyl
positions were weaker, but still significant. Hydrogen bonds with the 3- and 6-positions (C3–OH and C6–OH) contributed at least 9 kJ/mol. These results demonstrate the impact that non-covalent interactions can have in GH catalysis.

The great majority of GHs have a strong hydrogen-bond acceptor (usually with a carboxylate group) that interacts with the 2-hydroxyl group. The proximity of the C2–OH group to the anomeric center (the most electron-deficient region of the TS) makes it the best hydroxyl position to form strong hydrogen-bond contacts to relieve the positive charge build-up at the TS. Hydroxyl groups in other positions are also commonly make strong hydrogen bonds.

There are also other non-covalent interactions commonly used by GHs to help stabilize the TS. Amino acids that contain a hydrophobic platform frequently form stacking interactions with the sugar ring at the –1 position. It was suggested that these interactions also stabilize the developing positive charge of the oxocarbenium ion-like TS.

**Unresolved mechanistic issues**

The discovery of the two catalytically relevant carboxylate amino acid residues, almost ubiquitous in GHs (Figure 2.2), was made decades ago, in 1953. Nevertheless, a great variety of GH families, each with a different active site, carry out the retaining or inverting mechanisms in different ways. These variations between the mechanisms in different GH families are the focus of current research, including:

- The structural and electronic features of hydrolysis TSs and how they depend on the specific GH enzyme.
- The role of non-covalent interactions between the active sites of specific GHs and the substrates that they hydrolyze.
- Sugar ring distortion inside the enzyme’s active site and how it depends on the interactions with the enzyme, the general mechanism, and the intrinsic properties of the sugar. Also, the different conformations of the sugar ring along the reaction pathway are still undetermined for many GH families.
- Understanding the special features in the mechanisms of certain GH families with unusual active sites, such as the presence of a metal ion interacting with the substrate (e.g. members from family GH38), the absence of a catalytic base (e.g. members from GH48 and GH6) and even enzymes that may follow completely different mechanisms.

In this thesis, we focus on shedding light on some of these unresolved mechanistic issues for enzymes from families GH8, GH38, and GH43. In the next section we present some of the work done for these enzyme families that serves as the foundation to our work.

**GH8 endoglucanase and GH43 β-xylosidase**

The following section presents an introduction to two of the GHs studied in this work, GH8 endoglucanase and GH43 β-xylosidase, which are involved in the deconstruction of the plant cell wall. First, we describe the general process of hydrolysis of the plant cell wall and the general role of cellulases and hemicellulases.

**Enzymatic deconstruction of the plant cell wall**

The fermentation of small sugars obtained from biomass degradation is widely recognized as a sustainable source of renewable energy. Unfortunately, hydrolysis of the plant cell wall to smaller sugar moieties presents difficult challenges that limit our ability to competitively obtain biofuel from biomass. In its most common form, cellulose chains are buried deep inside the plant’s cell wall, tangled with hemicellulose chains and layers of
lignin that protect the whole structure from degradation.\textsuperscript{35,36} In Nature, a combination of cellulases and hemicellulases are used by bacteria and fungi to untangle and hydrolyze the cellulose chains. Understanding Nature’s solution to this problem is an important step to overcome the challenges of biomass deconstruction.

Organisms release cellulases and hemicellulases to the cell’s exterior, where they attach to the plant’s cell wall and degrade cellulose into smaller sugars. There are two main strategies used by cellulose-dependent organisms to release these enzymes. Some organisms release various cellulases and hemicellulases to diffuse freely in the extracellular environment. The oligosaccharide produced by the enzymes are then taken up by the cell and further hydrolyzed into smaller sugars by intracellular glycoside hydrolases. Other organisms release the enzymes with attached domains so that they form very large multi-enzymatic assemblies called cellulosomes\textsuperscript{37,38} that are attached to the cell surface.

The GHs secreted to the extracellular environment work synergistically to break the plant’s cell wall. GHs with three different active-site topologies, cleft, well, and tunnel, have been found in the enzyme cocktails released from the cell. A cleft-shaped active site can accommodate various sugar residues and hydrolyze glycosidic bonds in the interior of the carbohydrate chain. The reaction may be followed by product release and subsequent binding of a new carbohydrate chain, or the chain may move along the cleft and be cleaved again. If the active site is a tunnel, a long polysaccharide chain can be accommodated, but the chain has to enter the active site through the tunnel’s cavity; therefore these enzymes can only break off the end residues of the chain, and then pull the chain across the tunnel for the next reaction. The enzyme proceeds until the chain is completely degraded or it escapes from the tunnel. GHs that have wells as active sites are quite common. They also cleave the end resi-
dues of the chains, but after one cleavage, the chain must retract from the active site to allow product release and permit another chain to enter the cavity for the next reaction.

The different topologies allow enzymes to work in concert and to make the degradation process more efficient. In cellulases, for example, endoglucanases (EGs), which have cleft-shaped active sites, cleave cellulose chains at random points, releasing chain ends that can be used by cellobiohydrolases (CBHs) that contain tunnel-shaped active sites. CBHs bind long pieces of cellulose chains across the tunnel and degrade them in a processive fashion. They break two glucosyl moieties (cellobiose) from the chain end and, after product release, the chain is pulled inside the tunnel to relocate it in the right position for the next round of catalysis. This process repeats until the chain is completely degraded or the chain escapes from the tunnel. The cellobiose released is then taken up by β-glucosidases, which have a well-shaped active site and form the individual glucose units that the cell requires. Hemicellulases with cleft and well-shaped active sites also exist, but tunnels are specific for cellulases.

In this work, we focused on a specific endoglucanase from the cellulosome in *Clostridium thermocellum* (GH8 endoglucanase) and a hemicellulase from *Geobacillus stearothermophilus* T-6 that degrades the nonreducing-end xylosyl residues from oligosaccharides and xylan chains (GH43 β-xylosidase).

**GH8 endoglucanase**

As part of the cellulosome of *C. thermocellum*, the function of GH8 endoglucanase is to hydrolyze the interior β-1,4-glycosidic bonds present in cellulose chains. Its cleft-shaped active site (Figure 2.6), binds several glucosyl residues from the cellulose chain and performs
Figure 2.6. X-ray crystal structures (up) and active sites (down) of GH8 endoglucanase (with E95N mutation) and GH43 β-xylosidase (with D128A mutation).
hydrolysis using the inverting mechanism (Figure 2.2B). The crystal structure of GH8 endo-
glucanase has been solved for the wild-type enzyme and for its E95G mutant complexed with cellopentaose. Its catalytic amino acid residues are Glu95, the proton donor, and Asp278, the nucleophilic assistant, identified by their active-site locations.

The tertiary structures of GH8 members display \((\alpha/\alpha)_6\) folds, which contain six \(\alpha\)-helices forming an inner circle, surrounded by other six \(\alpha\)-helices in an outer circle (Figure 2.6). Members of GH8 are evolutionarily related to members of GH48, and they share a similar tertiary structure. Both GH8 and GH48 are part of the GH-M clan.

The cellulose chain that binds the active site of GH8 is fixed inside the cleft through stacking interaction at subsites \(-3, -2, +1, +2\). The only subsite forming extensive hydrogen bonding with the substrate is subsite \(-1\), where the reaction occurs. Within this subsite, the proton donor (Glu95) forms a hydrogen bond with the glycosidic oxygen atom of the scissile bond, and a lateral aspartate (Asp152) forms two hydrogen bonds with the C2′–OH and C3′–OH hydroxyl groups. In addition, the nucleophilic water molecule is located directly underneath the anomeric carbon atom (C1′) held in place through hydrogen bonds with the proton acceptor (general base) (Asp278) and a tyrosine residue (Tyr215).

Important insight into the role of the lateral aspartate (Asp152) and the tyrosine (Tyr215) interacting with the nucleophilic water molecule was obtained through mutagenesis experiments. Replacing Asp152 with alanine caused complete enzyme inactivation, and this mutant was even less active than the proton donor and proton acceptor mutants. Because of its interaction with the C2′–OH and C3′–OH hydroxyl groups, it was proposed that Asp152 stabilizes the reaction TS. The Tyr215 mutation to phenylalanine (only one hydroxyl group difference) was also deleterious for GH8 activity, so it was proposed that its hydrogen bond
with the nucleophilic water molecule is necessary for the correct water orientation during the nucleophilic attack. In addition, Collins and coworkers\textsuperscript{40} measured the melting temperature of the wild-type enzyme and its Y215A mutant to measure enzyme stability upon mutation. They found a significant melting temperature decrease in the Y215A mutant that led them to conclude that Tyr215 is important for GH8 stability.

Another important feature of the GH8 endoglucanase active site is the distortion that it forces on the sugar ring occupying subsite \textendash\textendash 1. The glucopyranosyl ring in the crystal structure is distorted toward a $2\,S_{2^1}^B$ conformation.\textsuperscript{16} A crystal structure of the D152A mutant shows no ring distortion (i.e. the ring was observed in the regular $4_1^1$ form), confirming the importance of the bidentate interaction of Asp152 toward the correct ring distortion.

\textbf{GH43 $\beta$-xylosidase}

Hemicellulose is a complex polymer made of xylans, glucans, mannans, and arabinans. Of all of these polymers, xylan is the most abundant in the plant cell wall, accounting for 30\textendash35\%. After several xylanases\textsuperscript{42} located in the extracellular environment break the plant cell wall xylan, the oligosaccharides are then transported into the cell. A \textit{Geobacillus stearothermophilus} T-6 GH43 $\beta$-xylosidase (also known as XynB3) is responsible for hydrolyzing $\beta$-1,4-glycosidic bonds from the nonreducing ends of xylooligosaccharides.\textsuperscript{43}

XynB3 removes one xylose at a time from the oligosaccharide chain, and its end product is a xylose monomer.\textsuperscript{43} This is consistent with a well-type active site. The tertiary structure of XynB3 was solved at 2.15-Å resolution, and it displays a five-bladed $\beta$-propeller fold (Figure 2.6).\textsuperscript{44} Its active site consists of a mainly hydrophobic pocket that binds the nonreducing end xylosyl residue that contains the scissile bond. The hydrophobic pocket is formed by Phe32,
Trp74, Phe127, and Phe506. Trp74 forms stacking interactions with the xylosyl residue in subsite –1 (Figure 2.6). In addition, there are three carboxylate groups forming hydrogen bonds with the hydroxyl groups (one of them, Asp128, is not shown in Figure 2.6, because it was mutated in the crystal structure) and an arginine residue that forms hydrogen bonds with the oxygen atom at the C4’–OH position.

There are very few interactions between the enzyme and the xylosyl residue in subsite +1, consistent with the fact that XynB3 is not very selective at this subsite. The only interactions observed at this subsite are between the C3–OH group and the proton donor, and between the C3–OH group and enzyme backbone. The two carboxylate groups shown in Figure 2.6, Asp15 and Glu187, are the nucleophilic assistant (or general base) and the proton donor, respectively. Their roles were first confirmed by mutagenesis studies, followed by activity measurements in a variety of substrates with a range of leaving-group pK_a.s. As the leaving group pK_a decreases (i.e. better leaving group), the enzyme activity should increase. If the proton donor is mutated, the activity of the enzyme against substrates with good leaving groups should not be greatly affected, and this amino acid (Glu187) can be identified. Interestingly, the mutation of the lateral Asp128 residue results in a completely inactive enzyme against all substrates. It was proposed that this residue helps to modulate the pK_a of the proton donor and to stabilize the reaction TS via hydrogen bonding with the substrate.

**Golgi GH38 α-mannosidase and the N–linked glycosylation pathway**

Protein glycosylation is a cellular process consisting of the addition and modification of carbohydrate structures onto a protein chain. This process is responsible for crucial cellular functions such as protein trafficking, quality control, and folding. The addition of carbohy-
drates to protein chains starts at the endoplasmic reticulum (ER), where the lipid molecule dolichol is covalently attached to the initial carbohydrate structure Glc3Man9GlcNAc2 that will be transferred to the nacent protein (Glc is glucose, Man is mannose, and GlcNAc is N-acetylglucosamine). The carbohydrate structure is attached to the amino groups of asparagine residues when they are located within the sequences Asn–X–Ser or Asn–X–Thr, X being any residue except proline. This is the beginning of the N–linked glycosylation pathway.

The carbohydrate structure is trimmed several times inside the ER. First, three glucosyl units are hydrolyzed from Glc3Man9GlcNAc2 to form Man9GlcNAc2 (Figure 2.7). Then three mannosyl residues are cleaved from the structure in a reaction catalyzed by ER α-mannosidase. The protein is transferred to the Glogi apparatus, where Golgi α-mannosidase I cleaves one more mannosyl residue to form the product Man5GlcNAc2. At this point, the first addition to the carbohydrate structure occurs. A GlcNAc residue is incorporated into the structure with a transfer reaction catalyzed by transferase I. Subsequently, Golgi α-mannosidase II cleaves two mannosyl residues from GlcNAcMan5GlcNAc2, ending the trimming steps in the N-glycosylation pathway. After this, several glycosyl transferases attach several GlcNAc, galactosyl, and sialic acid moieties to form complex-type oligosaccharides (Figure 2.7).

The complex-type oligosaccharides shown in Figure 2.7, specifically the tetra-branched structures, give rise to complicated structures that contain an elongated chain of β-1,6-mannosyl residues at one of the branches. These structures with β-1,6 branches can be found in increased concentrations in some malignant tumor cells, and this is correlated with their metastatic properties.
Inhibition of certain enzymes in the N-linked glycosylation pathway has been proposed as a means to reduce the amounts of β-1,6 branches.

One such target is Golgi α-mannosidase II (GMII), and its inhibition with swainsonine has shown potential in the treatment against breast, colon, and skin cancers.\textsuperscript{51,52} Unfortunately,
swainsonine also inhibits some lysosomal mannosidases, producing undesirable side effects.\textsuperscript{51,53,54}

GMIIs belong to GH38, together with other $\alpha$-mannosidases.\textsuperscript{1} There are tertiary structures available in GH38 that belong to GMI from \textit{D. melanogaster} and lysosomal $\alpha$-mannosidase from \textit{B. taurus}. GH38 members have a mannosidase fold that consist of a 1000-amino acid globular domain with an N-terminal $\alpha/\beta$-domain and a C terminal with all $\beta$-sheets.\textsuperscript{55} GH38 members hydrolyze mannosyl glycosidic bonds using a retaining mechanism, as observed by trapping a mannose substrate in a covalent intermediate state.\textsuperscript{56} With the same experiment, the identity of the nucleophilic amino acid residue, Asp204 (\textit{D. melanogaster} nomenclature), was revealed. These results were later confirmed by crystal structures of the covalent intermediate.\textsuperscript{57} Asp341 was then proposed as the general acid/base catalytic residue (proton donor) based on its proximity to the glycosidic oxygen atom of the scissile bond. Surprisingly, the active site in GH38 members contains a zinc ion that interacts directly with the nucleophile (Asp204) and two hydroxyl groups (C2$'–$OH and C3$'–$OH) of the substrate.\textsuperscript{55}

GH38 is the first known GH family whose members employ a catalytic Zn ion.\textsuperscript{19} Three other amino acid residues, His 90, Asp92, and His471, are also coordinated by the zinc ion to give a hexacoordinated ion–ligand complex.

The mannosyl residue bound to subsite $–1$ forms several hydrogen bonds with the protein’s active site. A lateral aspartate, Asp92, hydrogen-bonds with the C2$'–$OH group. Analogous interactions with the C2$'–$OH group are observed in many other GHs. In addition, a second aspartate, Asp472, forms hydrogen bonds with the C3$'–$OH and C4$'–$OH groups. These hydrogen bonds likely stabilize the positively charged oxacarbenium ion–like TS.
The presence of a positively charged Zn ion in the active site might seem counterintuitive, as the oxacarbenium ion-like TS is also positively charged and the enzyme’s active site environment must provide a stabilizing effect toward the TS to perform catalysis. Because of the interaction with the oxygen atoms of the substrate’s C2’–OH and C3’–OH groups, it was proposed that the zinc ion helps to induce distortion of the mannosyl ring. It was also suggested that the Zn ion contributes toward TS stabilization, but no explanation was given as to how this may happen.

**Figure 2.8.** X-ray crystal structure of GMII in complex with its natural substrate reveals the enzyme subsites that bind the M3 and G3 sugar moieties.

The GMII tertiary structure complexed with its natural substrate (GlcNAcMan₅GlcNAc) was recently obtained (Figure 2.8). The GlcNAc residue five units away from the scissile bond (G3 in Figure 2.8) is required for catalysis by GMII. The crystal structure with the natural substrate revealed an enzyme subsite, the anchor site, that recognizes the distant G3 sugar (green in Figure 2.8), which explains the selective action of the enzyme toward sub-
strates having the G3 residue. In addition, a second subsite, the holding site, binds the M3 sugar (purple in Figure 2.8). The presence of this subsite led the authors to propose that GMII cleaves the M5 residue prior to the M3 sugar, but that they use the same catalytic site.

Overview of computational methods

This section briefly describes the computational methods used in this research. The challenges of modeling enzymatic reactions are presented first to give an idea to those unfamiliar with molecular simulations and to provide a rationale for the computational methods chosen in this work. Simulations of enzymatic reactions are presented in Chapters 3, 5, and 7.

Current challenges in modeling enzymatic reactions

The size of the system

Chemical reactions, in general, need to be described by the electronic rearrangement that occurs as a molecule transforms from reactant to product, and this is done with electronic structure (i.e. quantum chemistry) calculations.

With currently available computer power, quantum chemistry calculations can only be performed for a small number of atoms (usually fewer than than 200), depending on the level of approximation and the type of calculation. A regular protein with 350 amino acid residues has about 6000 atoms. In addition, considering solvation of the protein by a water box, the atom count climbs up to ~40,000. Clearly, electronic structure calculations alone cannot deal with the enzymatic system if the whole enzyme and solvent are modeled.

Time-scale problem in molecular dynamics
If one is interested in studying the enzymatic system with MD to obtain the time evolution of the system (to obtain free energies, for example), the time scale of the simulations becomes an important problem.

In *ab initio* (or first principles) MD (i.e. MD with quantum chemistry methods), the time that can be simulated is on the order of picoseconds. If one is modeling a reacting system, such as an enzyme in its Michaelis complex state, the simulation needs to be as long as the reaction time ($1/k_{cat}$ for enzymes) to spontaneously observe the reaction. For instance, $1/k_{cat}$ for the fastest GHs is about 0.001 s, orders of magnitude higher than the time attainable in *ab initio* simulations. Even classical simulations that can model atomic motion without electronic rearrangements (i.e. no reactions) can only model nanoseconds of dynamics.

Together, the system size and its time scale make modeling of enzymatic reactions a difficult task for *ab initio* MD simulations. Fortunately, there are some techniques presented in this section that can be used to overcome these problems with present computational power.

**Multiple reaction coordinates**

In addition to the challenges discussed above, there are some reactions that present an extra complication: the presence of multiple reaction coordinates to describe the reaction pathway. For example, in GHs there are two bond breaking/making events that occur during the reaction (Figure 2.9). For this system, two reaction coordinates are required, one to describe the proton transfer between the proton donor and glycosidic oxygen atom, and a second one to model the nucleophilic attack and departure of the leaving group.

These two bond-making/breaking events occur simultaneously during the reaction and they need to be modeled together.
QM/MM metadynamics

As mentioned earlier, there are techniques that can be used to model these complicated systems. To overcome the problem of the system size, quantum chemistry methods can be combined with a computationally cheaper force field methodology to simplify the problem. The rationale here is that the electronic rearrangements throughout an enzymatic reaction occur only within the protein active site (usually less than 100 atoms), and quantum chemistry calculations are only needed for this part. The rest of the system and solvent can be treated with the force field. The time scale and multiple reaction coordinates problems can be solved simultaneously using the metadynamics technique, which induces the reaction.

Figure 2.9. The two reaction coordinates needed to describe the reaction in a GH enzyme.

QM region

MD for the enzyme active site was modeled with the Car–Parrinello (CP) scheme, using Density Functional Theory (DFT). The CP–MD approach is better illustrated by comparing it to traditional ab initio MD, also called Born–Oppenheimer molecular dynamics (BO–MD).
In BO–MD, an optimized wave function of the electronic system needs to be calculated to obtain the energy for a particular nuclear configuration. The energy gradient can then be used to calculate the forces on the nuclei. As nuclei are treated classically, Newton’s equations can be integrated (with a time step of ~1 fs) to obtain new positions for the nuclei. This process is repeated many times to obtain an MD trajectory. Energy calculations by optimizing the wave function are very time-consuming, and this step needs to be repeated many times during a BO–MD simulation. In CP–MD, the wave function optimization needs to be carried out only at the beginning of the simulation. This is accomplished by letting the wave function parameters evolve simultaneously with the nuclei along the MD, by including wave function parameters as fictitious masses in the dynamics. The price to pay for including the electronic degrees of freedom into the dynamics is that a smaller time step (~0.12 fs) is required to describe the nuclear motion. Nevertheless, CP–MD is computationally cheaper, which allows treatment of larger systems and longer simulation times.

**Molecular mechanics region and classical molecular dynamics**

In classical or force-field MD, atoms are the smallest units of the simulated system. Electrons are not explicitly accounted for. Nevertheless, the effect of electrons (i.e., bonding and nonbonding interactions) is implicitly accounted for by assigning properties to the atoms according to their types (e.g. element, hybridization…etc.).

The system energy depends on the atomic coordinates in the system and a set of parameters that are usually fitted to experimental data or calculations of higher accuracy. The set of atom types, parameters, and energy functions that describes the system is called a force field. The AMBER force field, for example uses the following energy function:
The energy function contains terms describing the energy required to displace a certain part of the system away from its “most stable” conformation. For example, the first term describes covalent bonds between two atoms. The parameter $K_r$, which depends on the two atom types bonded, multiplies the difference between the actual bond length ($r$) and the pre-specified average bond length for the atom types involved ($r_{eq}$). The other energy terms correspond to angle bending of two bonds, torsion of a bond, and non-covalent van der Waals and electrostatic interactions.

With the system energy, the forces can be calculated and the time evolution (dynamics) of the system can be explored using Newton’s second law. This method allows simulating large systems of up to $10^6$ atoms in the nanosecond time scale. The main disadvantage of classical simulation is that electronic rearrangements of the system cannot simply occur (like in BO–MD or CP–MD).

In this work, the Amber package was used to carry out classical MD simulations. The AMBER force field was used to describe protein systems, while the carbohydrates were modeled with GLYCAM$^{62}$ and water with TIP3P.$^{63}$

**Metadynamics**

Even with *ab initio* MD, where electronic calculations are included, the picosecond time scale attainable in the simulations is not enough to observe reactions with significant energy barriers. Metadynamics$^{59}$ is a recently developed technique that allows the simulation of
processes separated by significant energy barriers, such as chemical reactions.\textsuperscript{64} It requires the definition of a set of collective variables (CVs) that describe the transition of interest. During the simulation, small Gaussian hills are added to the potential energy of the system in CV space, to discourage the system from visiting the same regions of CV space previously visited in the simulation. As the metadynamics simulation proceeds, the small Gaussian hills accumulate until enough energy is added to force the system out of a minimum through the lowest energy pathway to the other side of the potential energy surface. This is illustrated in Figure 2.10, at the beginning of the simulation (T0 in Figure 2.10) the system behaves as in regular dynamics. As the simulation time proceeds, new Gaussians are added and new regions of the potential energy surface become available to the system until it can escape the minimum (at T4). At the end of the simulation (at T9), both minima are filled and the system can do the transition back and forth. The free energy of the system can be estimated by summing all the Gaussian hills required fill the wells.\textsuperscript{65} This methodology has been applied to a variety of problems in chemistry, biology, and material science.\textsuperscript{28,66-69}
Figure 2.10. Schematic of a metadynamics run. The system starts at T0 and Gaussian hills are added in CV space until the transition can occur through the lowest energy maximum (at T4). The simulation is stopped when both wells are filled (at T9).

References


(2) Coutinho, P.; Henrissat, B. *Carbohydrate-Active Enzymes Server* **1999**.


Chapter 3: Mechanism of Cellulose Hydrolysis by Inverting GH8

Endoglucanases: A QM/MM Metadynamics Study

Modified from a paper published in *The Journal of Physical Chemistry B*

Luis Petersen, Albert Ardèvol,¹ Carme Rovira,¹ and Peter J. Reilly

Abstract

A detailed understanding of the catalytic strategy of cellulases is key to finding alternative ways to hydrolyze cellulose to mono-, di-, and oligosaccharides. Endoglucanases from glycoside hydrolase family 8 (GH8) catalyze the hydrolysis of β-1,4-glycosidic bonds in cellulose by an inverting mechanism believed to involve a oxacarbenium ion-like transition state (TS) with a boat-type conformation of the glucosyl unit in subsite –1. In this work, hydrolysis by *Clostridium thermocellum* endo-1,4-glucanase A was computationally simulated with quantum mechanics/molecular mechanics metadynamics based on Density Functional Theory. Our calculations show that the glucosyl residue in subsite –1 in the Michaelis complex is in a distorted $^2S_O/2,5^B$ ring conformation, agreeing well with its crystal structure. In addition, our simulations capture the cationic oxacarbenium ion-like character of the TS, with a partially-formed double bond between the ring oxygen and C1’ carbon atoms. They also provide previously unknown structural information of important states along the reaction pathway. The simulations clearly show for the first time in GH8 members that the TS features a boat-type conformation of the glucosyl unit in subsite –1. The overall catalytic mechanism follows a $D_N^* A_N$-like mechanism and a $β-^2 S_O \rightarrow 2,5^B [TS] \rightarrow α-^5 S_1$ conformational

¹ Parc Científic Barcelona, Barcelona, Spain
itinerary along the reaction coordinate, consistent with the anti-periplanar lone pair hypothesis. Because of the structural similarities and sequence homology among all GH8 members, our results can be extended to all GH8 cellulases, xylanases, and other endoglucanases. In addition, we provide evidence supporting the role of Asp278 as the catalytic proton acceptor (general base) for GH-8a subfamily members.

Keywords: boat-type conformation, CPMD, Density Functional Theory, GH8 endoglucanase, metadynamics, transition-state itinerary

**Introduction**

Glycoside hydrolases (GHs) catalyze the hydrolysis of glycosidic bonds that join either two carbohydrate moieties or a carbohydrate glycon and a non-carbohydrate aglycon. At present GHs are divided into more than 100 families\(^1\) based on primary sequence data. Known tertiary structures exist in most families, many complexed with their substrates or inhibitors. Despite this wealth of structure–function information, further supported by mechanistic studies from kinetic isotope effects, computational simulations, and enzyme kinetics,\(^2\)\(^-\)\(^4\) there are still several mechanistic aspects of GHs that are not fully understood and that are being researched.\(^3\)

There is overwhelming evidence that GHs distort the glycon sugar ring bearing the scissile glycosidic bond and further change the ring conformation along the reaction coordinate.\(^5\)\(^,\)\(^6\) The conformational interconversions that occur in the active sites of specific GHs have been the subject of many recent studies.\(^7\)\(^-\)\(^14\) Nevertheless, direct evidence of conformational itineraries is still missing for many GH families. Furthermore, detailed structural
features of the enzymatic transition state (TS) in GHs are still fairly unknown. This issue has been difficult to address because of the extremely short life of an oxacarbenium ion-like TS, which makes it very difficult to obtain detailed structural insight with experimental methods. Most of the information that we have about TS structures in GHs comes from kinetic isotope effect measurements and structural information of GHs complexed with TS mimics.\textsuperscript{12, 15–16} Fortunately, reliable computational techniques can be exploited to explicitly model reactions inside GHs. To date, the reaction mechanism in GHs has been modeled only for GH22 lysozyme and GH2 β-galactosidase by first-principles approaches, providing evidence of a covalent glycosyl–enzyme intermediate for these retaining GHs.\textsuperscript{17–19}

Cellulases, found in many GH families, have received much attention for their connection with the biofuels industry. The slow degradation of cellulose, as well as that of other plant cell-wall components like hemicellulose and lignin, is one of the principal impediments to competitively produce fuel from biomass.\textsuperscript{20–22} Cellulase catalytic mechanisms have attracted interest because these enzymes have been optimized by many years of evolution to hydrolyze cellulose to mono-, di-, and oligosaccharides. A deep understanding of the catalytic strategy of these enzymes would be very valuable in pursuing viable alternatives to perform cellulose hydrolysis.

GH family 8 (GH8) enzymes catalyze the hydrolysis of β-1,4-glycosidic bonds present in cellulose, chitosan, and xylan. The GH8 active site is a cleft that allows substrate binding and subsequent bond cleavage at interior chain positions. GH8 belongs to clan M (GH–M) along with GH48, with their members having (α/α)_6 tertiary folds.\textsuperscript{1,23}

GH8 enzymes use the inverting mechanism (Figure 2.2). Two carboxylate amino acid residues, either aspartate or glutamate, are employed by most inverting GHs to catalyze
hydrolysis, one protonating the scissile glycosidic oxygen atom (O4) and the other coordinating the nucleophile, a water molecule, to assist its deprotonation to complete the reaction.\textsuperscript{24} The proton donor and putative proton acceptor (general base) were identified based on structural information as Glu95 and Asp278, respectively,\textsuperscript{25} in \textit{Clostridium thermocellum}. However, the identification of Asp278 as the proton acceptor has been controversial in light of recent kinetic studies that show that some GH8 members in which residues analogous to Asp278 has been mutated still retain significant activity.\textsuperscript{26,27} Also, a phylogenetic study showed that the putative proton acceptor is not strictly conserved in GH8; instead its location shifts within the active site.\textsuperscript{28}

Guerin and coworkers published a high-resolution crystal structure, PDB 1KWF, of a mutated (E95Q) GH8 endoglucanase (EG) from the \textit{C. thermocellum} cellulosome (Figure 3.1).\textsuperscript{25} A cellopentaose molecule is bound in a groove-shaped active site, characteristic of EGs, spanning subsites –3, –2, –1, +1, and +2. The active site forces the cellopentaose chain to kink, and it distorts the glucosyl residue in subsite –1 away from the ground-state $^4C_1$ conformation to a $^2S_O$/$^2B$ ring conformation, while the other four glucosyl residues remain in the relaxed $^4C_1$ conformation. The catalytic proton donor, Glu95, is \textit{anti}-positioned (i.e. pointing towards the O4 lone pair that is at the opposite side to the ring O5’ atom).\textsuperscript{29} The putative catalytic proton acceptor, Asp278, is located below the average plane of the glucosyl ring occupying subsite –1, hydrogen-binding the nucleophilic water molecule (Figure 3.1). Tyr215 forms a second hydrogen bond with this water molecule, fixing it in place for optimal catalysis.\textsuperscript{25,26} Asp152 forms a bidentate interaction with the C2’–OH and C3’–OH hydroxyl groups of the glucosyl residue in subsite –1.
Figure 3.1. Active-site structure of *C. thermocellum* GH8 EG (PDB entry 1KWF). Red spheres are water molecules. The Gln95 residue of the crystal structure has been back-mutated to Glu. For the sake of clarity, the protein is oriented with the negative glucosyl subsites (–1, –2, and –3) at the right-hand-side.

The finding of a $^2S_O/2.5B$ type of distortion in the Michaelis complex of GH8 EG, together with that of a stable $2.5B$ glycosyl-enzyme intermediate in GH11, led to the suggestion that these GHs operate via a $2.5B$-type TS. In this work, this prediction is tested by first-principles quantum mechanics/molecular mechanics (QM/MM) simulations of cellulose hydrolysis catalyzed by GH8 EGs. The simulations were meant to gain insight into the enzyme catalytic mechanism, the structural features of the reaction TS, and the interconversions of the glucosyl unit occupying subsite –1 along the reaction coordinate. Also, further evidence that supports the role of Asp278 as the proton acceptor is presented. To the best of our knowledge, the hydrolysis reaction of an inverting GH has not previously been investigated by first-principles methods.
Computational methods

Classical molecular dynamics

The structure of the GH8 EG–cellopentaose complex described above was submitted to classical molecular dynamics (MD) simulations with Amber9 software. The FF99SB Amber and GLYCAM06 force fields were used to model the protein and cellopentaose substrate, respectively. The E95Q mutation of the crystal structure was manually changed back to glutamic acid. The crystal structure used, PDB 1KWF, contains two conformations. The conformation that contains the cellopentaose substrate spanning subsites –3 to +2 was used for this study. The protonation states of ionizable amino acids were selected based on a visual inspection of their microenvironments. The protonation states of Glu95 and Asp278 were chosen based on the common knowledge of the enzymatic inverting mechanism, where the proton donor begins the reaction in a neutral state and the proton acceptor must be deprotonated (Figure 2.2). Moreover, this choice led to good agreement between the computed and experimental structure at the active-site region. Asp152 was modeled in its deprotonated state to correctly simulate the two hydrogen bonds with hydroxyl groups of the substrate in subsite –1 (C2′–OH and C3′–OH) observed in the crystal structure. The rest of the aspartate and glutamate residues were deprotonated. All seven histidine residues, none of which are close to the enzyme active site, were modeled in their neutral states. The water molecules observed in the crystal structure were retained, and additional water molecules were added to fully solvate the protein, creating a water box with a 10-Å cushion around the protein surface. The water molecules were modeled with the TIP3P force field. Seven sodium ions were added to neutralize the system. The MD simulation was carried out in several steps. First, the system was subjected to energy minimization with fixed protein and substrate, and then the
constraints were removed and the whole system was minimized. To gradually reach the desired temperature of 300 K, weak spatial constraints were initially added to the protein and substrate while the water molecules and sodium ions were allowed to move freely. Finally, all constraints were removed and the MD simulation was extended to 1 ns, when the system had reached equilibrium. The structure was analyzed with the VMD package. A snapshot of the equilibrated system was used as a starting point for further QM/MM simulations.

**QM/MM molecular dynamics**

The method developed by Laio et al. was used. This method combines the first-principles MD method of Car and Parrinello (CPMD) with a force-field MD methodology (i.e. QM/MM CPMD). In this approach, the system is partitioned into QM and MM regions. The dynamics of the atoms in the QM region depends on the electronic density, $\rho(r)$, computed with Density Functional Theory (DFT), whereas the dynamics of the atoms in the MM region is ruled by an empirical force field. The QM/MM interface is modeled by the use of link-atom pseudopotentials that saturate the QM region. The electrostatic interactions between the QM and MM regions are handled via a fully Hamiltonian coupling scheme, where the short-range electrostatic interactions between the QM and the MM regions were explicitly taken into account for all atoms. An appropriately modified Coulomb potential was used to ensure that no unphysical escape of the electronic density from the QM to the MM region occurs. The electrostatic interactions with the more distant MM atoms are treated via a multipole expansion. Bonded and van der Waals interactions between the QM and the MM regions are treated with the standard Amber force field. Long-range electrostatic interactions between MM atoms are described with P3M implementation, using a $64 \times 64 \times 64$ mesh.
An accurate description of energetic, dynamic, and structural features of biological systems has been previously obtained with this methodology, confirming its reliability (see refs. 40, 41 and references therein).

The QM region was confined in an isolated $17.3 \times 19.1 \times 13.3$ Å supercell that included 66 atoms, consisting of the side chains of Glu95 and Asp278, capped at the Cα with a link-atom pseudopotential, and the glucosyl residues in subsites –1 and +1 capped at the C1 of subsite –2 and C4 of subsite +2. The nucleophilic water molecule was also included. Kohn–Sham orbitals were expanded in a plane-wave basis set with a kinetic energy cutoff of 70 Ry. *Ab initio* pseudopotentials generated by the Troullier–Martins scheme\(^42\) were used. The Perdew–Burke–Ernzerhoff functional\(^43\) in the generalized gradient-corrected approximation of DFT was selected. This functional was chosen based on its reliability in describing hydrogen bonds\(^44\) and it is the one that we used in our previous QM/MM work on a retaining GH.\(^8\)

A constant temperature of 300 K was reached by coupling the system to a Nosé–Hoover thermostat\(^45\) of 3500 cm\(^{-1}\) frequency.

Structural optimizations were done using QM/MM MD with annealing of electronic velocities until the maximal component of the nuclear gradient was $10^{-4}$ au. The time step used for the annealing simulations was 0.072 fs and the fictitious mass of the electrons was 1000 au. For the MD simulation coupled with a thermostat, a time step of 0.12 fs and a fictitious electron mass of 700 au were used. The TS electronic structure was analyzed in terms of localized orbitals according to the Wannier functions methodology.\(^46\)

Structural optimizations to determine the identity of the catalytic proton acceptor were carried out in a larger $19.0 \times 21.7 \times 20.8$ Å QM box. This included the side chains of Glu95 and Asp278; Tyr215, which coordinates the nucleophilic water molecule; Asp152, which
interacts with the C2’–OH and C3’–OH hydroxyl groups; seven water molecules; and the glycon and aglycon. All the amino acids were capped with link-atom pseudopotentials at the Cα position.

**QM/MM metadynamics**

One of the main advantages of *ab initio* MD is the possibility of modeling chemical reactions. Nevertheless, present computational power allows for only picoseconds of *ab initio* QM/MM MD simulations of large systems such as proteins, not long enough to observe spontaneous reactions with high energy barriers. The metadynamics technique is a novel methodology that can be used to induce the reaction by selecting a set of collective variables that include the relevant modes of the TS of the simulated reaction. The collective variables can be any function of the nuclear coordinates of the system, as long as they are able to distinguish the different states of the system (i.e., reactants, TS, and products). In this technique, small repulsive Gaussian-like potentials are added to the regions of space that were already visited, preventing the system from exploring the same regions and thereby allowing the system to overcome significant energy barriers. The small repulsive potential terms are used to construct the free energy surface of the system, which can be estimated as the negative of the sum of the Gaussian potential terms. This method has recently been applied to a variety of problems in biophysics, chemistry, and material science (see e.g. references 50–54).

Two collective variables, CV1 and CV2, were chosen to model the hydrolysis of the β-1,4-glycosidic bond between two glucosyl residues; each of them was a difference of coordination numbers (CN’s). The CN between two atoms, $CN_{i,j}$, indicates whether a covalent bond exists (1 = bond, 0 = no bond). CV1 is the difference $CN_{C1',Owat} - CN_{C1',O4}$ (Figure 3.2).
Thus CV1 measures the formation of the α-glycosidic bond between the C1’ atom and the incoming water molecule’s oxygen atom, with the simultaneous rupture of the β-glycosidic bond. CV2 is the difference $CN_{HGlu95,O4} - CN_{HGlu95,OGlu95}$ (Figure 3.2). Thus CV2 measures the proton transfer between the catalytic proton donor and the O4 atom. The $CN$ between two atoms $i$ and $j$ is defined\(^\text{47}\) by

$$CN_{ij} = \frac{1 - \left(\frac{d_{ij}}{d_0}\right)^p}{1 - \left(\frac{d_{ij}}{d_0}\right)^{p+q}}$$

where $d_{ij}$ is the internuclear distance of the atoms involved, $d_0$ is the threshold distance for bonding, and $p$ and $q$ are exponents that determine the steepness of the decay of $CN_{ij}$ with respect to $d_{ij}$. The selected values were $d_0 = 3.03$ au, $p = 12$, and $q = 2$ for CV1 and $d_0 = 2.08$ au, $p = 16$, $q = 3.6$ for CV2. It is important to note that this choice of collective variables does not introduce any bias concerning the type of nucleophilic substitution mechanism. In other words, it does not dictate whether the mechanism is of $A_N D_N$ (fully concerted mechanism), $D_N^* A_N$ type (dissociative mechanism, i.e. departure of the leaving group prior to attack by the nucleophile) or $D_N + A_N$ (dissociative mechanism, with the presence of an intermediate carbocation).\(^\text{55,56}\)

**Figure 3.2.** Collective variables (CV1 in green and CV2 in red) used to model hydrolysis of the glycosidic bond catalyzed by *C. thermocellum* GH8 EG (see definition in the text).
The metadynamics simulations were done within the Car–Parrinello approach. In this scheme, the Car–Parrinello Lagrangian is extended by extra terms describing the fictitious dynamics of the collective variables. These additional fictitious particles are coupled through a harmonic potential to the value of the selected collective variables in the real system. The mass of this fictitious particle and the force constant of the coupling potential were tested to ensure that the coupled particle naturally follows the value of the associated collective variable in the real system. The selected mass values of the fictitious particles were 15 and 5 amu for CV1 and CV2, respectively, while those of the force constant were 1.5 and 1 au for CV1 and CV2. The height of the Gaussian terms was 1.5 kcal mol\(^{-1}\), which ensures sufficient accuracy for the reconstruction of the free energy surface. The width of the Gaussian terms was 0.05 Å, according to the oscillations of the selected collective variables observed in a free dynamics. A new Gaussian-like potential was added every 400 MD steps.

**Results and discussion**

**Classical and QM/MM molecular dynamics simulations**

A model of the fully solvated wild-type enzyme–substrate complex at its natural conditions (300 K and 1 atm) was obtained by performing a classical MD simulation. The overall structure after the simulation is very similar to the original crystal structure, with an overall RMSD of 0.60 Å. The length of the hydrogen bond between Glu95, the catalytic proton donor, and the glycosidic oxygen atom is 1.87 Å. Hydrogen bonds with Asp278, the putative proton acceptor, and Tyr215 fix the nucleophilic water molecule in place, and are 2.2 Å and 3.7 Å long, respectively. The oxygen atom of the nucleophilic water molecule is 3.79 Å below the C1’ atom. The glucosyl ring in subsite –1 remained in its distorted \(^2\!S_{O}/2.5\!B\) confor-
mation throughout the classical MD simulation. Also, the nucleophilic water molecule remained underneath the C1’ all through the classical MD run. The simulation was then continued by QM/MM for 1.5 ps; during this time no significant conformational change occurred.

**QM/MM metadynamics simulation**

The system behavior during the metadynamics simulation is displayed in Figure 3.3, which shows the variation of the main distances characterizing the reaction. First explored was the reactant state, with the proton donor in its neutral (protonated) state. During the first 7 ps of the simulation, the proton moved back and forth from Glu95 to the O4 atom several times (the black line). At ~7.2 ps, with the O4 atom protonated, the nucleophilic attack

![Figure 3.3](image)

**Figure 3.3.** Time evolution of some important distances along the metadynamics simulation. Glu95: proton donor; O4: glycosidic oxygen atom; Hwat: nearer proton of the nucleophilic water molecule; Asp278: putative proton acceptor; C1’: glycon anomeric oxygen atom. The forward reaction occurs at ~7.2 ps and the reverse reaction occurs at ~31 ps.
occurred. The β-glycosidic bond (the blue line) lengthened while the distance between the C1’ atom and the oxygen atom of the incoming water molecule (the red line) shortened to within bonding distance. Also, one of the hydrogen atoms of the water molecule transferred to Asp278 (the green line).

After the forward reaction occurred, the product state was explored for ~24 ps. At ~15 ps, Arg281 pulled the deprotonated Glu95 toward it. This interaction is probably a strategy of the enzyme to prevent the reverse reaction, by positioning the catalytic proton donor in an unproductive local minimum while it is deprotonated. Therefore, a constraint in the CP Lagrangian\textsuperscript{57} forcing the Arg281···Glu95 distance to increase was applied for < 1 ps, after which Glu95 returned to its original position. While exploring the product state, the activated proton moved between the O4 atom and Glu95 several times (the gray and black lines) until the reverse reaction occurred at ~31 ps. The simulation was continued until ~33 ps, when no further significant change in the free energy surface (described below) was observed.

The reaction free energy landscape and minimal-energy pathway

Figure 3.4 shows the free energy surface of the GH8 EG reaction. The axes in the figure represent the two reaction coordinates used (CV1 and CV2). The minimal-energy pathway from the reactants to the products traverses the reactant state, the protonated glycosidic oxygen atom, the TS, and the product state. When the system is in the reactant state the values of CV1 and CV2 are ~0.8 each. The values of CV1 and CV2 for the product state are ~ −0.8 each.

The reactant well is approximately 36 kcal mol\textsuperscript{−1} below the TS conformational free energy, which is the maximal energy point in the reaction pathway. When the Glu95 proton
is transferred to the β-glycosidic oxygen atom (O4), the free energy increases to about 12 kcal mol$^{-1}$ below the TS energy. Our calculations support cellulose hydrolysis as an exothermic reaction, as the products are ~12 kcal mol$^{-1}$ below the reactants.

Snapshots of representative states along the reaction pathway are depicted in Figure 3.5. Table 3.1 shows average distances, computed from all configurations falling into a small region around the stationary point of the free energy surface (Figure 3.4) for structures that correspond to each of its minima (reactants, protonation of the glycosidic oxygen atom, and products) and to the TS.

**Figure 3.4.** Free energy surface of cellopentaose hydrolysis by a GH8 EG. CV1 measures the rupture of the β-glycosidic bond with the simultaneous formation of the α-glycosidic bond. CV2 measures the proton donation between Glu95 and the oxygen atom of the scissile β-glycosidic bond. Contours are separated by 3 kcal mol$^{-1}$. 
The reactant conformation is shown in Figure 3.5A. Here, protonated Glu95 forms a hydrogen bond of 1.98 Å with the O4 atom (Table 3.1). Asp278 and Tyr215 form hydrogen bonds with the nucleophilic water molecule and position it 3.82 Å below the C1’ atom. The glycon ring in the reactant state (the Michaelis complex) is maintained in an intermediate $\beta^{2S_0\beta.5}B$ conformation, in good agreement with the experimental structure.\textsuperscript{25} This ring conformation preactivates the substrate for catalysis by moving the glycosidic bond toward an axial position, increasing the charge of the C1’ atom\textsuperscript{10} and removing steric conflict between the H1’ atom and the incoming nucleophilic water molecule.

The next step in the reaction pathway is proton transfer from Glu95 to the O4 atom (Figure 3.5B), thereby increasing the aglycon’s leaving group ability. Glycosidic oxygen protonation also increases the length of the $\beta$-glycosidic bond by 0.11 Å (Table 3.1) and increases the energy of the system, facilitating the reaction (Figure 3.4).
Table 3.1: Average Bond Distances and Their Standard Deviations at Different Stages of the Reaction Pathway (Å)

<table>
<thead>
<tr>
<th>bond</th>
<th>reactants</th>
<th>O4 protonated</th>
<th>TS</th>
<th>products</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1’–O4</td>
<td>1.51 ± 0.07</td>
<td>1.62 ± 0.07</td>
<td>2.45 ± 0.17</td>
<td>3.83 ± 0.29</td>
</tr>
<tr>
<td>C1’–OWat</td>
<td>3.82 ± 0.34</td>
<td>4.14 ± 0.19</td>
<td>3.16 ± 0.38</td>
<td>1.41 ± 0.04</td>
</tr>
<tr>
<td>O_{Glu95}–H_{Glu95}</td>
<td>1.02 ± 0.04</td>
<td>1.34 ± 0.05</td>
<td>1.47 ± 0.06</td>
<td>2.31 ± 0.38</td>
</tr>
<tr>
<td>H_{Glu95}–O4</td>
<td>1.98 ± 0.20</td>
<td>1.14 ± 0.04</td>
<td>1.08 ± 0.04</td>
<td>0.98 ± 0.03</td>
</tr>
<tr>
<td>O_{Asp278}–H_{Wat}</td>
<td>1.79 ± 0.26</td>
<td>1.77 ± 0.12</td>
<td>1.69 ± 0.11</td>
<td>1.01 ± 0.03</td>
</tr>
<tr>
<td>O_{Wat}–H_{Wat}</td>
<td>1.01 ± 0.02</td>
<td>1.01 ± 0.02</td>
<td>1.02 ± 0.03</td>
<td>2.01 ± 0.19</td>
</tr>
<tr>
<td>O_{Tyr215}–H_{2Wat}</td>
<td>1.91 ± 0.24</td>
<td>1.91 ± 0.25</td>
<td>2.03 ± 0.24</td>
<td>2.17 ± 0.33</td>
</tr>
<tr>
<td>C1’–O5’</td>
<td>1.42 ± 0.04</td>
<td>1.37 ± 0.02</td>
<td>1.29 ± 0.02</td>
<td>1.45 ± 0.03</td>
</tr>
</tbody>
</table>

After the glycosidic oxygen atom is protonated, the dissociation (Dₙ) step takes place. The β-glycosidic bond lengthens by a further 0.83 Å until it disengages sufficiently for the complex to reach the TS (Figure 3.5C). Simultaneously, the glycon becomes an oxacarbenium ion-like species, where the C1’–O5’ bond shortens by 0.13 Å compared to its reactant state (Table 3.1) and becomes sp²-hybridized, forcing the C2’, C1’, O5’, and C5’ ring atoms into a plane. Calculated Wannier orbitals show the formation of a partial double C1’–O5’ bond at the TS (Figure 3.6). Upon formation of the oxacarbenium ion-like TS, the C1’ atom moves closer to the oxygen atom of the nucleophilic water molecule, by changing the glucopyranosyl ring conformation from a $^2S_O/2.5B$ into a $^2.5B$ (Figure 3.5C).

At the TS, the glycosidic oxygen atom is 2.45 Å from the C1’ atom, while the nucleophilic water molecule is 3.16 Å from the C1’ atom (Table 3.1). The O_{water}–C1’–O4 angle is
~133 ± 8°. Calculation of RESP partial charges indicate a highly cationic oxacarbenium ion-like TS (Figure 3.6), bearing a substantial positive charge at the C1’ atom (0.72, compared to 0.50 and 0.44 in the reactants and products, respectively). One of the lone pairs of the ring oxygen atom helps to stabilize the positive charge buildup by forming a partial double bond. The nucleophilic water molecule is polarized by its interactions with the putative proton acceptor Asp278, with the O_{Asp278}⋯H_{Wat} distance decreasing by 0.1 Å from the reactants to the TS (Table 3.1). The interactions of the C2’–OH and C3’–OH groups of the glucosyl residue in subsite –1 with Asp152 (Figure 3.1) are likely to relieve the electron deficiency of the TS and provide a stabilizing effect.

In the following association (AN) step to form the product, the glycon changes from the $^{2,5}B$ TS to an $\alpha$-$^{5}S_{1}$ conformation by moving its C1’ atom further and simultaneously trapping the water molecule. In a concerted manner, a proton from the water molecule is transferred to Asp278 to complete the reaction (Figure 3.5D). The product well is very long and is characterized by three minima (Figure 3.4), two of which represent the aglycon protonated state (CV2 ≈ −0.4 and −0.9). Another minimum (CV2 ≈ 0.4), much less stable than the other ones, represents re-protonation of the proton donor.

In summary, our results show that the glucosyl residue occupying subsite –1 follows a $\beta$-$^{2}S_{0} \rightarrow ^{2,5}B [TS] \rightarrow \alpha$-$^{5}S_{1}$ conformational itinerary along the reaction coordinate, confirming previous predictions of a $^{2,5}B$ TS in GH8-catalyzed hydrolyses. A second and less rigorous analysis was done by automated docking to test the ability of the GH8 EG to bind possible TS conformations. The docking results, available as Supplementary Information, further support a $^{2,5}B$ transition state.
Stereoelectronic effects on the $\beta^2S_O \rightarrow ^{2.5}B [TS] \rightarrow \alpha^5S_1$-like itinerary

Stereoelectronic effects are generally considered, but not without opposing views$^{59}$ to play an important role in anomeric substitution reactions (such as hydrolysis) of pyranoses.$^{60}$ The ring oxygen atom facilitates the reaction by donating electrons to stabilize the positive charge that develops in the anomeric center at the TS. Positioning the oxygen lone pairs in an optimal orientation with respect to the C1’ atom facilitates this electron transfer. For pyranosyl rings with a $\beta$-glycosidic bond, stereoelectronic theory dictates that the ring must undergo a conformational change away from the most stable $^4C_1$ chair form into a skew–boat conformation that allows the overlap between the anti-periplanar lone pair of the ring oxygen atom and the antibonding orbital of the $\beta$-glycosidic bond. This requirement prepares the ring for the substitution reaction. As the $\beta$-glycosidic bond elongates, the anti-periplanar lone pair in the O5’ atom proceeds to form the partial C1’–O5’ double bond at the TS (Figure 3.6). This is known as the anti-periplanar lone pair hypothesis (ALPH).

The pyranosyl ring conformations observed in GH active sites usually comply with the ALPH requirement. In the GH8 EG, the itinerary of the glucosyl ring in subsite –1 along the reaction coordinate follows a pathway similar to a $\beta^2S_O \rightarrow ^{2.5}B [TS] \rightarrow \alpha^5S_1$. Figure 3.7 shows two views of the Cremer–Pople sphere$^{61}$ with the pyranosyl ring conformations visited throughout the metadynamics simulation. This conformational itinerary is consistent with ALPH, as the pre-TS $\beta^2S_O$ conformation allows the overlap of the anti-periplanar lone pair of the ring oxygen atom with the antibonding orbital of the $\beta$-glycosidic bond. Nevertheless, a $^{2.5}B$ conformation has a C5–H2 flagpole interaction and eclipsed substituents in the C3–C4 positions of the pyranosyl ring that increase the energy of this conformation.
Figure 3.6. Structural and electronic features of the oxacarbenium ion-like TS. RESP partial charges of the more relevant atoms are shown as black labels. Wannier orbitals are shown for the sp2-hybridized C1’–O5’ bond (yellow) and for the ring oxygen lone pair (gray). The distances between the C1’, O4, and O_water are shown in blue.

Figure 3.7. Conformational itinerary of the glycon glucosyl ring along the reaction coordinate. All ring conformations of the metadynamics simulation are mapped onto a Cremer–Pople sphere. Blue: reactant ring conformations visited before the TS; cyan: product ring conformations visited after the TS. The red dot indicates the TS ring conformation.
Figure 3.8. A: Superposition of reactant (black) and product (blue) states. The C1’ atom moves the longest distance to trap the nucleophilic water molecule. B: Glycon ring conformations of the reactant state (black), TS (yellow), and product state (blue).

Substrate distortion during catalysis in GHs has been often discussed in relation to the relative movement of the anomeric carbon atom with respect to the nucleophile.\textsuperscript{4,62,63} Specifically, the C1’ atom becomes electrophilic as the glycosidic bond breaks, and it migrates toward the nucleophile to trap it, forming the new bond. In the case of GH8 EG, the $\beta^2S_0$ conformation of the glucosyl ring in subsite –1 in the Michaelis complex has its C1’ atom placed higher relative to the $^{2,5}B$ [TS] conformation, which is the next step in the reaction coordinate (Figure 3.8B). Therefore, a simple C1’ movement from the $\beta^2S_0$ conformer reaches the $^{2,5}B$ conformation and places the C2’, C1’, O5’, and C5’ ring atoms in a plane, a requirement imposed by the oxacarbenium ion-like, sp2-hybridized, C1’–O5’ bond. The $\alpha^5S_1$ product conformation is also a mere C1’ movement down from the TS conformation.
Figure 3.8A shows a superimposition of the reactant and product states in the GH8 EG, demonstrating that the C1’ atom migrates furthest to trap the nucleophilic water molecule. Therefore, the $\beta^{-2}S_0 \rightarrow 2.5B\ [TS] \rightarrow \alpha^{-5}S_1$ itinerary is consistent with the so-called electrophilic migration mechanism.

**The GH8 endoglucanase proton acceptor**

As mentioned above, there has been some controversy about the identity of the GH proton acceptor. This is because the position of this residue in the active site is not conserved among GH8 members, which is unusual within GH families. This led to a subdivision of GH8 into three subfamilies (GH-8a, GH-8b, and GH-8c) depending on the location of the proton acceptor.28 Moreover, kinetic studies of GH8 members that have residues corresponding to Asp278 show significant retention of activity upon mutation of the putative catalytic base.26,27

The crystal structure of the GH8 EG has a water molecule underneath the C1’ atom in a good position to proceed with the substitution. Asp278 and Tyr215 form hydrogen bonds with the nucleophilic water molecule. The crystallographic model25 appears to demonstrate that the proton acceptor is Asp278, as most GHs employ a pair of carboxylate-containing amino acids as proton donor and proton acceptor. Although Tyr215 also hydrogen-bonds the nucleophilic water molecule, it seems unlikely that it could be the catalytic proton acceptor, given that its $pK_a$ would have to be considerably lower to assume this role (there is no positively charged residue in the vicinity of Tyr215 that could justify a drastic $pK_a$ change of this residue). Tyr215 is strictly conserved within GH8 and even further within clan GH-M. The mutation of this residue to phenylalanine leads to an almost complete loss of activity in GH8
members. Even though tyrosine can act as nucleophile in some retaining GHs, the role of Tyr215 in GH8 is apparently to fix the water molecule in a correct position for catalysis.

To confirm the identity of the proton acceptor in GH8 *C. thermocellum* EG, we calculated where the proton released from the nucleophilic water molecule after hydrolysis was likely to go, by performing a series of QM/MM geometric optimizations and energy calculations for three different situations: 1) after Asp278 takes the proton; 2) after Tyr215, which forms a second hydrogen bond with the nucleophilic water molecule, takes the proton, leading to the tyrosine hydroxyl group becoming positively charged; 3) after the solvent becomes the proton acceptor, forming an H$_3$O$^+$ cation. The GH8 active site is fairly solvent-exposed, with several water molecules underneath the substrate, but only one is coordinated by Asp278 and Try215 and is correctly positioned for catalysis. The rest of the water molecules could be responsible for stabilizing the proton.

The first case is the result from the metadynamics simulation. In the second case, the proton on Tyr215 rapidly transfers to Asp278 via the C1’–OH group. In the third case, the geometric optimization that starts with an H$_3$O$^+$ cation is unstable and the proton also migrates back to Asp278. Therefore our results support Asp278 as the proton acceptor of the GH8 *C. thermocellum* EG. However, the enzymatic mechanism and identity of the proton acceptor in the unusual situation where Asp278 is not present, such as in GH8 EG mutants or in GH48 members, remain unknown and awaits study.

**Conclusions**

Enzymatic hydrolysis of cellulose by GH8 *C. thermocellum* EG was explicitly modeled with DFT QM/MM methods. The reaction was activated by using the metadynamics
approach, and the free energy of the reaction with respect to two reaction coordinates was calculated. The Michaelis complex (i.e. the bound reactants) shows that the glucosyl unit occupying subsite –1 is in a distorted $^2S_0/^{2,5}B$ ring conformation, in good agreement with experiments. The minimal-energy pathway from reactants to products passes through a transient state characterized by protonation of the glycosidic oxygen atom (O4) and a substantial lengthening of the glycosidic bond. The glucosyl residue occupying subsite –1 takes a $^{2,5}B$ conformation at the TS, whose structural features are consistent with an oxacarbenium ion-like structure (i.e. an sp2-hybridized C1’–O5’ bond that forces planarity of the C2’, C1’, O5’, and C5’ ring atoms). This is the first time that a boat-type transition state has been found in GHs by first-principles calculations, confirming previous hypothesis of a $^{2,5}B$-type TS in GH8 enzymes. The conformation of the glucosyl residue occupying subsite –1 is $^5S_1$ in the product state. Therefore, our calculations predict that catalysis by GH8 members takes place via a $\beta-^2S_0 \rightarrow ^{2,5}B [TS] \rightarrow \alpha-^5S_1$ itinerary of the glycon ring along the reaction coordinate, consistent with the anti-periplanar lone pair hypothesis. The reaction path calculations, as well as additional QM/MM calculations on alternative scenarios (i.e. Tyr215 or solvent water as a catalytic proton acceptor), provide evidence that Asp278 is the catalytic proton acceptor of GH8 C. thermocellum EG. In view of the sequence homology and structural similarity among all GH-8a members, our results also apply to cellulases, xylanases, and other EGs belonging to subfamily GH-8a.

Acknowledgments

The authors thank the National Research Initiative of the U.S. Department of Agriculture Cooperative State Research, Education and Extension Service (grant number 2007-35504-
18252), the Generalitat de Catalunya (grant 2005SGR-00036), and the Ministerio de Ciencia e Innovación (MICINN) (grant FIS2008-03845) for supporting this work. A. A. acknowledges a FPU research fellowship from MCINN. We are also deeply grateful to Wim Nerinckx (University of Ghent) for his continued advice. We are very grateful to Pedro Alzari (Institut Pasteur) for discussions on the GH8 catalytic mechanism, to Mercedes Alfonso–Prieto for her help with CPMD, and to the Iowa State University High Performance Computation System for making its facilities available to us.

**References**


(25) Guerin, D. M.; Lascombe, M. B.; Costabel, M.; Souchon, H.; Lamzin, V.; Benuin, P.;
(28) Adachi, W.; Sakihama, Y.; Shimizu, S.; Fukazawa, T. S. T.; Suzuki, M.; Yatsunami,
(33) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels,
    014113.


(59) Sinnott, M. Biochem J. 1984, 224, 817-821


Chapter 4: Transition-State Stabilization and Dynamics in a Glycoside Hydrolase Family 8 Endoglucanase

Luis Petersen and Peter J. Reilly

Abstract

The active site of a glycoside hydrolase Family 8 inverting endoglucanase from Clostridium thermocellum binding the Michaelis complex and transition-state structures of cellotriose was computationally modeled using ab initio calculations. The contribution of individual active-site amino acids to lower the energy barrier of cellulose hydrolysis (with respect to the reactant state) was calculated at the MP2 (6-31G*) level of theory. The amino acid residues that contribute the most to lower the reaction barrier are Asp152, Asp278, Arg281, and Arg84. Many body contributions were important, and they significantly decreased transition-state stabilization, compared to the addition of all the individual contributions. In addition, classical molecular dynamics simulations were carried out with the complete and fully-solvated protein complexed with a cellopentaose substrate. Virtual mutations were simulated to measure the effect of specific active-site amino acids on ring puckering of the glucosyl residue in subsite -1. The E95A and D152A mutants showed the most significant changes in glycon ring puckering.

Introduction

Fermentable sugars obtained from cellulose hydrolysis are a promising source of bio-renewable fuel. Unfortunately, the 1,4-β-glycosidic bonds in cellulose are particularly stable, making cellulose degradation a difficult task. In addition, contributing to biomass recalcification...
trance is the intricate chemical structure of the plant cell wall embedding cellulose fibrils.  

Nevertheless, enzymatic cellulose degradation is ubiquitous in nature. Cellulases and hemicellulases are secreted by many organisms to untangle and release cellulose chains from plant cell walls and hydrolyze them to simple sugars. A detailed understanding of the enzymatic hydrolysis would be very valuable in finding efficient alternatives to degrade biomass.

Cellulases and glycoside hydrolases (GHs) in general are among the most proficient enzymes on earth, achieving rate enhancements up to $10^{11}$ over the uncatalyzed reaction in solution. This catalytic power is due largely to the enzyme’s ability to stabilize the transition state (TS) of the catalyzed reaction. Therefore, an understanding of cellulase catalytic strategies requires analysis of the interactions between the enzyme active site and the reaction TS. The short life of the TS makes it difficult to analyze by experimental tools. Fortunately, we can use accurate computer simulations to deal with this problem, as has been done for several enzymatic systems.

In GHs, non-covalent interactions are important for TS stabilization. Hydrogen bonds between the hydroxyl groups of the carbohydrate substrate and the enzyme active site were estimated to contribute more than 10 kcal/mol toward TS stabilization, especially by interactions with the C2’–OH substrate’s hydroxyl group.

In this work, we present quantum mechanics (QM) simulations of Clostridium thermocellum GH Family 8 (GH8) endoglucanase. *Ab initio* calculations were done to obtain insight into the TS stabilization provided by the enzyme’s active site amino acids. The energy of the reactant and TS were measured in a gas-phase reference reaction and in the presence of individual active-site amino acid residues. The difference in energy barrier upon addition of active-site residues provides insight into the energetic contribution of each active-site residue remaining.
to stabilize the reaction TS (or destabilize the Michaelis complex). In this way, we provide a quantitative analysis of the cellulase active site that can be useful for protein engineering, design of enzyme mimicry catalyst systems, and as a framework to compare the active sites of different GHs.

Several studies have been done on the *C. thermocellum* GH8 endoglucanase. Guerin and coworkers published the crystal structure 1KWF of the mutated (E95Q) enzyme, showing a cellopentaose molecule bound in a groove-shaped active site, characteristic of endoglucanases (Figure 2.6). The active site forces the cellopentaose chain to kink and distorts the glucosyl residue in subsite –1 away from the ground-state $^4C_1$ conformation into a $^2SO^{2,5}B$ ring conformation. In addition, enzymatic cellulose hydrolysis by a GH8 endoglucanase was explicitly modeled using QM/MM metadynamics simulations. The conformational interconversions of the glucosyl unit occupying subsite –1 were predicted to follow a $^\beta^{2}S_0 \rightarrow ^{2,5}B\ [TS] \rightarrow ^\alpha^{-5}S_1$-like itinerary along the reaction coordinate. The structural and electronic features of the enzymatic TS were obtained from this previous work.

GH8 endoglucanases hydrolyze cellulose chains by an inverting mechanism (Figure 2.2B). Inverting GHs usually use two carboxylate amino acid residues (Asp or Glu) to catalyze hydrolysis, one protonating the scissile-bond glycosidic oxygen atom and the other coordinating with the nucleophile, a water molecule, to assist its deprotonation to complete the reaction.

**Computational methods**

The structure of the GH8 endoglucanase TS complex used for the present work was obtained from a previous QM/MM metadynamics simulation (presented in Chapter 3) using a
QM/MM MD method combining first-principles Car–Parrinello MD with a force field.\textsuperscript{13,14} The metadynamics technique\textsuperscript{15,16} was used to induce the hydrolysis reaction.

For the present work, snapshots of the QM/MM metadynamics simulation, corresponding to the reactant and TS, were selected. Three structures were taken from each state (reactants and TS) and an average of the energetic contributions (see below) was taken. To make the system tractable with high-accuracy \textit{ab initio} simulations, the system was simplified by taking the side chains of the active-site amino acid residues Arg281, Asp152, Asp278, Glu95, Ser94, Trp132, Tyr215, Tyr372, the nucleophilic water molecule, and a water molecule that hydrogen-bonds the substrate and Ser94, from the QM/MM metadynamics snapshots of the reactant and TS (Figure 4.1). A cellotriose substrate spanning subsites \(-2, -1,\) and \(+1\) was also taken from the snapshots. The side chains of the active-site amino acid residues were capped with methyl groups, as shown in Figure 4.1.

To obtain insight into the catalytic strategy of GH8 endoglucanases, a differential TS stabilization (DTSS)\textsuperscript{8} was performed (Figure 4.2). DTSS refers to the stabilization measured with respect to the reactant state (Michelis complex). The energy barrier in the presence of an active site amino acid (\(\Delta E_{\text{enzyme}}\)) minus the energy barrier of the gas-phase reference reaction (\(\Delta E_{\text{gas}}\)) gives an estimate of the amino acid residue’s contribution to TS stabilization (or inhibition) (Figure 4.2). If the net effect of a specific residue is to lower the reaction barrier, the value of its contribution will be negative. Positive values indicate an inhibitory effect of the amino acid residue. Lowering of the energy barrier can arise from stabilizing the TS more than the Michaelis complex or by destabilizing the reactant state (Michaelis complex) more than the TS. The same calculation can be done for all active-site residues except the proton donor Glu95, because its chemical structure is different at the reactant state, where it is
Figure 4.1. System used for DTSS analysis. The snapshots were taken from previous QM/MM metadynamics simulations.

Figure 4.2. Differential TS stabilization (DTSS) analysis. The contribution of a specific amino acid residue to lower the reaction barrier is measured by $(\Delta E_{\text{enzyme}} - \Delta E_{\text{gas}})$. The bottom curve represents the energy on the presence of an active-site amino acid residue. The upper line is a gas reference reaction with no amino acid residues (except for Glu95 that is part of the reaction).
protonated, than at the TS, where it is deprotonated.

The single-point calculations were carried out at the correlated MP2/6-31G* level of theory, and the estimation for the effect of many-body interactions was carried out at the RHF/6-31G* level. All electronic structure calculations were done with GAMESS.17

**Molecular dynamics simulations**

The mutated (E95Q) crystal structure of GH8 EG complexed with cellopentaose was used to perform molecular dynamics (MD) simulations using Amber 9.18 Hydrogen atoms were added to the crystal structure using AmberTools and the point mutation E95Q was manually changed back to Glu95. The H++ interface was used to assign the correct amino acid residue protonation states.19 The proton donor, Glu95, was protonated during all MD simulations. All water molecules observed in the crystal structure were retained and additional water molecules were added to fully solvate the protein, creating a water box with a 10-Å cushion around the protein surface. The water molecules were modeled with the TIP3P force field.20 Amber force-field parameters for the protein,21 and GLYCAM22 was used to model the crystal-structure cellopentaose molecule. Virtual mutations were introduced with PyMOL.23 Seven Na+ ions were added to neutralize the system. The MD simulation was carried out in several steps. First, the system was subjected to energy minimization while fixing the protein and substrate, then the constraints were removed and the whole system was minimized. To reach the desired temperature of 300 K in a gradual manner, weak spatial constraints were initially added to the protein and substrate while the water molecules and Na+ ions were allowed to move freely. In the final simulation, all constraints were removed and the MD simulation was extended to 1 ns, when the system was equilibrated. The same
process was repeated for all simulations. Periodic boundary conditions were applied for the system. The time step used was 2 fs. The structural features of the system were analyzed with the VMD package.\textsuperscript{24}

Results

We conducted \textit{ab initio} calculations in the GH8 endoglucanase active site to gain insight into the contribution of specific active-site amino acid residues to lower the reaction energy barrier. Figure 4.3 shows the energetic contribution ($\Delta E_{\text{enzyme}} - \Delta E_{\text{gas}}$) of individual active-site amino acid residues toward TS stabilization with respect to the reactant state (Michaelis complex). All residues investigated are conserved within subfamily A of GH8, which includes all its cellulases, xylanases, and other endoglucanases.\textsuperscript{25}

The DTSS analysis indicates that Asp152 is the active-site amino acid residue that provides the largest contribution (–11.26 kcal/mol) toward TS stabilization of the GH8 endoglucanase (at the MP2/6-31G* level). The proton acceptor Asp278 also substantially lowers, by –9.87 kcal/mol, the energy barrier. Two arginine residues, Arg281 and Arg84, contribute by –8.59 and –6.07 kcal/mol, respectively. Both tyrosine residues, Tyr372 and Tyr215, do not play a crucial role in stabilizing the TS, contributing only –1.87 and –1.37 kcal/mol, respectively. Finally, two amino acid residues, Trp132 and Ser94, lightly inhibit TS formation. The proton donor cannot be analyzed quantitatively using DTSS as the rest of the amino acids, because the chemical structure of Glu95 is different at the enzyme pre-TS complex, when it is protonated, than in the enzyme TS complex, when it is deprotonated.
Many-body interactions turned out to be important, especially between Asp152 and Arg281, which form a hydrogen bond (Figure 4.1). When Asp152 and Arg281 are taken together in the calculation, their contribution to lower the energy barrier reduces to $-10.75$ kcal/mol (at the RHF/6-31G* level), compared to $-18.73$ kcal/mol, which is the sum of the individual contributions of both amino acids. In total, there is a $17.6$-kcal/mol difference between the sum of the individual contributions and the contribution of all the active-site amino acids taken together in the same calculation.

![Figure 4.3](image)

**Figure 4.3.** Contribution of specific amino acid residues to lower the energy barrier of hydrolysis in GH8 endoglucanase calculated at the MP2/6-31G*. Negative values represent lowering of the energy barrier. Positive values represent an increase on the energy barrier (inhibition effect).

Due to the importance of many-body interactions, the DTSS analysis performed above cannot be taken as a quantitative measure of the role of active-site amino acids residues to stabilize the reaction’s TS. Nevertheless, it is interesting to identify the types of interactions that lower the cellulose hydrolysis energy barrier. For instance, negatively charged residues
have the largest contributions to lower the reaction barrier. The cellulose hydrolysis TS has significant oxacarbenium-ion character and it is positively charged. Negatively charged residues relieve the electron deficiency of the TS, providing a stabilizing effect. Surprisingly, positively charged arginine amino acids also provide a significant contribution. This can occur, for example, when the amino acid destabilizes the reactant state more than the TS, or if the positive amino acid pulls electron charge from its neighboring amino acids to share with the substrate at the TS.

Classical MD simulations of wild-type GH8 endoglucanase and four virtual mutants (E95A, D278A, D152A, and Y215F) were performed to identify the amino acids responsible for distorting the Michaelis complex glucopyranosyl ring. During the simulations, the conformations of the pyranosyl ring were followed and mapped onto a Cremer-Pople sphere. Figure 4.4 shows a two-dimensional projection (viewed from the north pole) of the sphere for six different simulations, including two for the wild-type enzyme and one for each virtual mutant analyzed. Two wild-type simulations were necessary because the force field was unable to reproduce the hydrogen bond between the hydrogen atom between the proton donor (Glu95) and the glycosidic oxygen atom. Therefore, a distance constraint was placed to force this hydrogen bond. The ring conformations visited during both wild-type simulations (with and without constraint) are slightly different and therefore we show both. The mutants that show the most ring distortion are E95A and D152A. In both simulations, the $^2S_o/2.5B$ conformation turned into the most stable $^4C_1$ conformation. The rest of the virtual mutants did not change the ring conformation significantly during the simulation. Additional changes may be observed for longer simulation times.
Figure 4.4. Cremer-Pople parameters of the pyranosyl ring in subsite –1 for different simulations of virtual mutations.
Discussion

The hydrogen-bond interactions that Asp152 forms with the C2’–OH and C3’–OH hydroxyl groups are crucial for GH8 EG’s activity. A mutagenesis analysis of a GH8 endoglucanase\(^27\) and xylanase\(^28\) shows that the D152A mutation causes a complete loss of activity, having the most deleterious effect of all mutated GH8 active-site residues, including the proton donor and nucleophilic assistant. Asp152 is necessary for ring distortion, based on crystallographic evidence of the D152A mutant.\(^29\) Our simulations show that the non-covalent interactions of Asp152 with the glucosyl ring in subsite –1 are very important for GH8 endoglucanase catalysis. Asp152 provides a large contribution to stabilize the reaction TS. In addition, it is partially responsible for the distortion of the pyranosyl ring, as evidenced in our classical MD simulation of the D152A virtual mutant.

The experimental single-point mutations Y215F and Y215A also cause large activity decreases in GH8 endoglucanase and xylanase.\(^27,28\) It was suggested that Tyr215 is mainly important to hold the nucleophilic water molecule in place,\(^10,28\) and for structural purposes due to an observed decrease in melting temperature upon its mutation to phenylalanine.\(^28\) The DTSS analysis shows little contribution of Tyr215 toward lowering the energy barrier, which suggests a structural role for these residues, consistent with experimental predictions.

The DTSS analysis identifies Asp278, the proton acceptor, as an important amino acid to lower the hydrolysis energy barrier. This contrasts with experimental mutagenesis studies that show little change in activity upon mutation of Asp278 to alanine. However, these experiments were done with the soluble unnatural substrate carboxymethyl cellulose, which differs from the natural cellulosic substrate at the C6 position, a carboxymethyl group replacing the hydroxyl group (C6–OH) present in cellulose. The carboxylate group of Asp278 directly
interacts with the C6–OH at subsite +1 (see Figure 3.5); therefore, the effects of enzyme mutation on activity towards CMC are probably not the same as those on cellulose.

As mentioned above, the proton donor cannot be analyzed using DTSS. Nevertheless, a qualitative analysis of the interactions of the proton donor with the substrate and TS is still insightful. At the Michaelis complex, the proton donor interacts closely with the glycosidic oxygen atom of the scissile bond, which becomes protonated at the TS (Figure 4.1). Also, Glu95 forms a hydrogen bond with the C2–OH group at the +1 subsite. This interaction helps to twist the glycosidic bond, which accounts for ~3 kcal/mol (see chapter 6).

**Conclusion**

*Ab initio* calculations were done to measure the contribution of the important residues Ser94, Trp132, Asp152, Tyr215, Asp278, Arg281, and Tyr372 toward TS stabilization in GH8 endoglucanase. The active-site residues that provide the largest contributions toward TS stabilization (at the MP2/6-31G* level) are Asp152, the proton acceptor Asp 278, and two arginine residues, Arg281 and Arg84. Many-body effects, not present in the individual contribution calculations, turned out to be important. In addition, MD simulation of GH8 EG wild-type and E95A, D152A, Y215F, and D278A virtual mutants show that the proton donor Glu95 and lateral aspartate Asp152 significantly affect distortion of the glucosyl residue occupying subsite –1.

**References**


Chapter 5: Mechanism of Xylobiose Hydrolysis by GH43 β-Xylosidase

Ian J. Barker, Luis Petersen, and Peter J. Reilly

Abstract

Glycoside hydrolases cleave the glycosidic linkage between two carbohydrate moieties. They are among the most efficient enzymes currently known. β-Xylosidases from glycoside family 43 hydrolyze the nonreducing ends of xylooligomers using an inverting mechanism. Although the general mechanism and catalytic amino acid residues of β-xylosidases are known, the nature of the reaction’s transition state and the conformations adopted by the xylopyranosyl ring along the reaction pathway are still elusive. In this work, the xylobiose hydrolysis reaction catalyzed by XynB3, a β-xylosidase produced by Geobacillus stearothermophilus T-6, was explicitly modeled using first-principles quantum mechanics/molecular mechanics Car–Parrinello metadynamics. We present the reaction's free energy surface and its previously undetermined reaction pathway. The simulations also show that the reaction proceeds through a $^{2,5}B$-type transition state with significant oxacarbenium-ion character.

Introduction

Glycoside hydrolases (GHs) catalyze the hydrolysis of the glycosidic bond between two carbohydrate residues or a carbohydrate and a non-carbohydrate moiety. Spontaneous hydrolysis of this linkage has an estimated half-life of over 5 million years. GHs may accelerate this reaction by $10^{17}$ times the spontaneous rate, suggesting that they are among the most efficient catalysts studied. In general, GH active sites are highly specific for substrate binding.

---

1 Author contributions: I.J.B. and L.P. performed calculations; L.P. analyzed the present data; I.J.B., L.P., and P.J.R. wrote the paper.
and catalyze hydrolysis using a general acid and a general base. The spatial arrangement of these residues in each enzyme active site determines whether the product anomeric configuration is inverted or retained during hydrolysis (Figure 2.2).

The CAZy (Carbohydrate-Active enZYme) database classifies GHs and other carbohydrate-active enzymes into families by amino acid sequence, implying that all family members share a common ancestry and mechanism and a similar overall protein fold. Of the 100+ GH families found in CAZy, GH family 43 (GH43) members, many of which are β-xylosidases (EC 3.2.1.37), come almost exclusively from bacteria, fungi, and plants. *Geobacillus stearothermophilus* T-6 specifically produces XynB3, an inverting GH43 β-xylosidase with β-sandwich and five bladed β-propeller domains (Figure 5.1). Its active site is a pocket, closed off at one end by a single loop at the end of the β-sandwich domain, that binds one xylosyl residue at a time. The enzyme exhibits *exo* action, releasing single xylosyl units from the nonreducing ends of short xylooligomers.

![Figure 5.1](image.png)

**Figure 5.1.** Structure of XynB3 showing its β-sandwich domain (left) and a five-bladed β-propeller domain (right) that form a pocket-shaped active site.
Inverting GHs use a single-displacement mechanism in which one carboxylic acid acts as a general base/nucleophile and another acts as a general acid/proton donor, involving an oxacarbenium-ion transition state. In XynB3, Asp15 and Glu187 are the general base and general acid, respectively. Asp15 coordinates a water molecule that attacks the glycon C1’ atom, while Glu187 protonates the leaving group at the glycosidic oxygen atom.

Recently obtained XynB3 crystal structures of E187G and D128G mutants has xylobiose molecules in the active site. Each crystal structure is a tetrameric complex of XynB3 enzymes, all binding xylobiose molecules. Interestingly, the xylosyl rings bound to subsite –1 have different conformations, even for the same mutant. In the proton donor mutant (E187G), two xylosyl rings are in a $^4C_1$ chair conformation (the most stable conformation in solution) and the other two are in a $^2S_{O2,5}B$ conformation. The D128G mutant has one xylobiose molecule in a $^4C_1$ conformation, two in a $^2S_{O2,5}B$ conformation, and one $^0E$ conformation. Glycon ring distortion of the Michaelis complex away from the $^4C_1$ conformation is a common feature of $\beta$-glycosidases.

In this work, we model the hydrolysis of the $\beta$-1,4 glycosidic bond in xylobiose by GH43 $\beta$-xylosidase using first-principles quantum mechanics/molecular mechanics (QM/MM) Car–Parrinello metadynamics to obtain insight into the detailed mechanism of the enzyme. In addition, we provide new structural and dynamical information for the wild-type Michaelis complex, transition state, and product states. Furthermore, we followed the conformational itinerary of the glycon xylosyl ring along the hydrolysis reaction.

**Computational Methods**
Most of the computational methods are very similar to those presented in Chapter 3. Therefore, we focus on the important details that are specific for this simulation.

**Classical molecular dynamics and system setup**

The XynB3 GH43 β-xylosidase structure was subjected to classical molecular dynamics (MD) using the Amber9 software package. The FF99SB force field was employed to model the protein, while the GLYCAM06 force field was used to model the xylobiose substrate. The D128G mutation was corrected manually to simulate the wild-type active site using PyMOL. Glu187 was protonated and Asp15 was left unprotonated, based on prior knowledge of the inverting reaction mechanism. The protonated states of ionizable amino acids was selected based on their microenvironment. The remaining aspartate and glutamate residues were modeled in their unprotonated state, while His151, His210, His253, and His362 were protonated to a charge of +1. The other seven histidine residues were simulated in their neutral states. The original crystal-structure water molecules were retained and a water box of 10 Å from the protein surface was added and modeled with using the TIP3P force field. Eleven sodium ions were added to neutralize the system.

**QM/MM molecular dynamics**

The QM/MM implementation of Laio et al. within the Car–Parrinello MD scheme was used. The QM region was defined as a 14.9 x 18.1 x 19.5 Å box containing 55 atoms, and including the side chains of Asp15 (capped at the β-carbon atom) and Glu187 (capped at the γ-carbon atom). The nucleophilic water molecule was also included in this region. Kohn–Sham orbitals were expanded in a plane-wave basis set with an 80-Ry kinetic energy cutoff. Ab initio pseudopotentials were generated by the Troullier–Martins scheme. The Perdew–
Burke–Ernzerhoff functional\textsuperscript{20} was selected in the generalized gradient-corrected approxima-
tion of Density Functional Theory based on its reliability in describing hydrogen bonds.\textsuperscript{21}

The structure was optimized using QM/MM MD by annealing the electronic velocities until the maximal component of the nuclear gradient was $10^{-4}$ au. The annealing simulation time step was 0.072 fs and the fictitious electron mass was 1000 au. The MD simulation coupled to a thermostat had a time step of 0.12 fs and a fictitious electron mass of 700 au.

**QM/MM metadynamics**

For this system, a difference of coordination numbers ($CN$) was used as collective variables (CV). Two CVs were used to indicate reaction progress. CV1 is $CN_{C1',O4} - CN_{C1',Owat}$, a measure of glycosidic bond breakage and water nucleophilic attack. CV2 is $CN_{HGlu187,OGlu187} - CN_{O4,HGlu187}$, which indicates proton transfer between the proton donor and the glycosidic oxygen atom. $CN$ is defined as

$$CN_{ij} = \frac{1-\left(d_{ij}/d^0\right)^p}{1-\left(d_{ij}/d^0\right)^{p+q}}$$

In this case, $d^0 = 4.35$ au, $p = 12$, and $q = 3$ were chosen for CV1, and $d^0 = 2.44$ au, $p = 14$, and $q = 6$ were chosen for CV2. The fictitious masses of CV1 and CV2, required to implement the metadynamics technique in the Car–Parrinello MD scheme,\textsuperscript{22-24} were 15 and 5 amu, respectively, and the force constants of CV1 and CV2 were 1.5 and 1 au, respectively. Gaussian hills, 0.05 Å wide and 1.25 kcal mol\textsuperscript{-1} high, were added to the potential well every 400 MD steps. The heights of these were small enough to ensure that the calculated free-energy surface is accurate.
Results and discussion

The QM/MM metadynamics simulation

A total of 51 ps of metadynamics were required to model the enzymatic hydrolysis of xylobiose inside GH43 β-xylosidase and to construct the free energy surface (FES) (Figure 5.2). The glycosidic bond cleavage, with the simultaneous formation of a covalent bond between the nucleophilic water molecule and the anomeric carbon atom (C1’), occurred after 7 ps of metadynamics simulation. The rest of the 51 ps was spent exploring the product well, which is much deeper and longer than the reactant well.

Figure 5.2: Free energy surface of xylose hydrolysis by GH43 β-xylosidase. Contours are separated by 3 kcal/mol.
The free energy surface and GH43-catalyzed hydrolysis

The Gaussian hills added to induce the reaction were used to construct the FES (Figure 5.2). Here, the reactant, TS, and product states can be identified. The free energy barrier was estimated at 24 kcal/mol. While no experimental energy barrier value was found for this enzymatic reaction, the calculated barrier lies in the range of experimental values found for other GH enzymes. The reaction’s free energy (i.e. the difference between reactant and product free energies) was estimated at 30 kcal/mol.

All the structures from the metadynamics trajectory that correspond to each of the two minima (reactant and product) and the TS were selected to compute an average of important distances (Table 5.1). Representative structures for the reactant, TS, and product states are shown in Figure 5.3.

**Table 5.1: Average Bond Distances and Their Standard Deviations at Different Stages of the Reaction Pathway (Å)**

<table>
<thead>
<tr>
<th>bond</th>
<th>reactants</th>
<th>TS</th>
<th>products</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1’–O4</td>
<td>1.49 ± 0.04</td>
<td>2.62 ± 0.11</td>
<td>4.05 ± 0.32</td>
</tr>
<tr>
<td>C1’–OWat</td>
<td>4.12 ± 0.19</td>
<td>2.75 ± 0.06</td>
<td>1.45 ± 0.04</td>
</tr>
<tr>
<td>OGlu187–HGlu187</td>
<td>1.03 ± 0.03</td>
<td>1.56 ± 0.09</td>
<td>1.37 ± 0.13</td>
</tr>
<tr>
<td>HGlu187–O4</td>
<td>1.56 ± 0.06</td>
<td>1.03 ± 0.02</td>
<td>1.11 ± 0.07</td>
</tr>
<tr>
<td>OAsp15–HWat1</td>
<td>1.99 ± 0.27</td>
<td>1.83 ± 0.13</td>
<td>1.00 ± 0.03</td>
</tr>
<tr>
<td>NHis249–HWat2</td>
<td>1.87 ± 0.12</td>
<td>2.00 ± 0.02</td>
<td>1.99 ± 0.21</td>
</tr>
<tr>
<td>HArg288–OWat</td>
<td>3.07 ± 0.32</td>
<td>2.59 ± 0.06</td>
<td>2.72 ± 0.31</td>
</tr>
<tr>
<td>OWat–HWat1</td>
<td>0.99 ± 0.03</td>
<td>0.99 ± 0.02</td>
<td>2.13 ± 0.39</td>
</tr>
<tr>
<td>C1’–O5’</td>
<td>1.40 ± 0.02</td>
<td>1.28 ± 0.00</td>
<td>1.42 ± 0.03</td>
</tr>
</tbody>
</table>

The reaction starts with the xylobiose molecule bound to the enzyme’s active site, with its glycon pyranosyl ring distorted away from the most stable $^4C_1$ chair form (see below). The scissile glycosidic bond is 1.49 Å long at the reactant state. The nucleophilic water molecule is held in place, 4.12 Å underneath the anomeric carbon atom, by forming a 1.99-Å-long
hydrogen bond with the proton acceptor (Asp15) and two additional hydrogen bonds with active-site amino acid residues, one of 1.87 Å with His249 and the other of 3.07 Å with Arg288. The proton donor (Glu187) forms a 1.56-Å-long hydrogen bond with the scissile glycosidic oxygen atom (Figure 5.3).

![Figure 5.3: Structures for important states along the reaction pathway. Left panel: reactant, center panel: transition state, and right panel: product. The enzyme is shown in the background with yellow and cyan colors for β-sheets and loops, respectively.](image)

When the TS is reached, the glycosidic bond elongates substantially to 2.62 Å and the nucleophilic water molecule approaches within 2.75 Å of the anomeric carbon atom. The proton acceptor (Asp15) tightens its hydrogen bond with the nucleophilic water molecule from 1.99 Å at the reactant state to 1.83 Å at the TS. The C1’ atom becomes sp2-hybridized, forming a partial double bond with the ring oxygen atom (O5’), as evidenced by a shortening of the C1’-O5’ bond from 1.40 Å at the reactant state to 1.28 Å at the TS. This is consistent with a TS with significant oxacarbenium ion character.

The TS then collapses to the products, where the glycosidic oxygen atom (O4) is now 4.05 Å away from the C1’ atom, and the nucleophilic water molecule forms a covalent bond of 1.45 Å with the anomeric carbon atom, transferring the extra proton to the proton acceptor
(Asp15) to finish the reaction. The C1’ atom recovers sp3 hybridization with a C1’–O5’ distance of 1.42 Å.

The lateral aspartate Asp128 forms a bidentate interaction with the C2’–OH and C3’–OH hydroxyl groups throughout the reaction. An analogous interaction was observed experimentally and computationally in a GH8 endoglucanase.\textsuperscript{26,27} The interaction of a negatively charged amino acid residue with the hydroxyl groups, especially the C2’–OH group, is likely to help relieve the electron deficiency of the TS, thereby providing a stabilizing effect.

**The conformational itinerary of GH43 β-xylosidase**

As mentioned above, the crystal structures of GH43 β-xylosidase (even in their mutated states) show some bound ligands with distorted glycon conformations, instead of the $^4C_1$ form, which is the most stable in solution. In this work, we tracked the glycon xylosyl ring conformation along the metadynamics simulation to gain insight into the conformational itinerary of GH43 β-xylosidase. The conformations visited before and after the reaction were mapped onto a two-dimensional projection of the Cremer–Pople sphere\textsuperscript{28} viewed from the north pole (Figure 5.4).

The conformations visited before the reaction span a large portion of the sphere, from $^2S_0$ to $B_{1,4}$ conformations, suggesting that the active site does not hold the glycon ring very tightly. Interestingly, the mutated crystal structures also show a variety of active-site glycon conformations. The average xylosyl ring conformation at the TS has a structure very close to the $^{2,5}B$, represented by the middle sphere with a yellow star. After the TS, a smaller region of the Cremer–Pople sphere, between $^{2,5}B$ and $^5S_1$ conformations, was populated. Because of the large range of conformations visited throughout the trajectory, we focused our attention to
the small part where the forward reaction occurs. These conformations are shown in the middle sphere of Figure 5.4. The reaction occurs on the near vicinity of the $^{2.5}B$ conformer, with the reactants protruding into the $^2S_0$ region (as observed in the crystal structures) and the products going into the $^5S_1$ region.

![Figure 5.4: Conformational itinerary of the glycon xylosyl ring along the metadynamics simulation. The conformations visited before the TS are shown in blue on the left diagram. The conformations visited after the TS are shown in red on the right diagram. In the middle diagram, the average TS conformation is shown with a yellow star. The blue and red lines in the middle diagram represent the conformations visited before and after the reaction, respectively, for a small part of the trajectory where the reaction occurs (see text).](image)

To summarize, the GH43-catalyzed hydrolysis of xylobiose was modeled with first-principles Car–Parrinello QM/MM metadynamics. The FES of the reaction was constructed, allowing the identification the important states along the reaction. Previously unknown structural information for the wild-type Michaelis complex, TS, and product was obtained with the simulation. The active site can accommodate a range of xylosyl ring conformations, especially at the Michaelis complex.

**References**


Chapter 6: Twisting of glycosidic bonds by hydrolases

Modified from a paper published in Carbohydrate Research

Glenn P. Johnson¹, Luis Petersen, Alfred D. French¹, Peter J. Reilly

Abstract

Patterns of scissile bond twisting have been found in crystal structures of glycoside hydrolases (GHs) that are complexed with substrates and inhibitors. To estimate the increased potential energy in the substrates that results from this twisting, we have plotted torsion angles for the scissile bonds on hybrid Quantum Mechanics::Molecular Mechanics energy surfaces. Eight such maps were constructed, including one for α-maltose and three for different forms of methyl α-acarviosinide to provide energies for twisting of α-(1,4) glycosidic bonds. Maps were also made for β-thiocecclobiose and for three β-cellobiose conformers having different glycon ring shapes to model distortions of β-(1,4) glycosidic bonds. Different GH families twist scissile glycosidic bonds differently, increasing their potential energies from 0.5 to 9.5 kcal/mol. In general, the direction of twisting of the glycosidic bond away from the conformation of lowest intramolecular energy correlates with the position (syn or anti) of the proton donor with respect to the glycon’s ring oxygen atom. That correlation suggests that glycosidic bond distortion is important for the optimal orientation of one of the glycosidic oxygen lone pairs toward the enzyme’s proton donor.

¹ Southern Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, New Orleans, LA, USA

Introduction

It has been known for many years that the substrate carbohydrate residue immediately to the nonreducing side of the scissile glycosidic bond (in enzyme subsite −1) is distorted during the cleavage reaction catalyzed by many glycoside hydrolases (GHs).¹ Evidence of ring conformations of higher intramolecular energy than for the optimal ⁴C₁ shape has become overwhelming in recent years as crystal structures of GH–ligand complexes have been solved.²,³

A second distortion, that of the torsion angles for the two bonds connected to the glycosidic oxygen atom, has been discussed much less. One article was devoted to twisting distortions for all linkages in oligo- and polysaccharides bound by proteins.⁴ Most linkages had low-energy conformations, but some scissile linkages corresponded to high energies. Distortion of substrate torsion angles by GHs was also mentioned by Fujimoto et al.⁵

Several different cleavage functions could be enhanced by twisting glycosidic bonds. Distortion of the scissile bond can improve access of catalytic amino acid residues to the glycosidic oxygen atom and the C₁′ atom of the nonreducing-side carbohydrate residue. Changes in the electronic structure resulting from torsional changes could enhance reactivity.⁶ Also, an increase of $E_{\text{Intra}}$, the intramolecular energy above that for the lowest-energy conformation of the scissile bond torsion angles, could reduce the energy barrier to cleavage, although strain was ruled out as the major contribution toward lowering the reaction barrier. Electrostatic interactions are proposed to have a more important role.⁷

There is now a sufficient number of GH crystal structures with carbohydrates bound in active sites to address the question of the effect of GH–ligand binding on scissile bond
geometry. Our approach was to plot the available data for different \( \alpha \)- and \( \beta \)-glycosidic linkages involving nitrogen, oxygen, and sulfur [NOS] on maps of \( E_{\text{Intra}} \) values. Thus, we have plotted experimental values of \( \phi (O5'–C1'–[NOS]4–C4) \) and \( \psi (C1'–[NOS]4–C4–C5) \) (Fig. 6.1) on energy surfaces that are based on hybrid Quantum Mechanics::Molecular Mechanics (QM::MM) computations in an attempt to gain new insights on this question.

**Figure 6.1.** Structures of molecules used to construct energy contour maps. Ring conformations are denoted with labels. Superscripts and subscripts denote the atoms above or below the plane of the ring, respectively.

**Glycoside hydrolase families and mechanisms**

GHs have been classified into >100 families based on similarities in their primary structures. Tertiary structures, i.e., the overall arrangement of helices, strands, and loops, are expected to be quite similar within individual families. Crystal structures are known for
enzymes with a ligand bound in subsites −1 and +1 that belong to fourteen families that
cleave either α-1,4 or β-1,4 glycosidic bonds. These subsites contain the glycosyl residues to
the immediate nonreducing and reducing sides, the glycon and aglycon, respectively, of the
scissile glycosidic bond (Table 6.1). Members of five families hydrolyze α-glycosidic bonds,
while members of the other nine cleave β-glycosidic bonds.

Occasionally, different GH families appear to be distantly related to each other by having
similar tertiary structures and mechanisms, and therefore they are grouped into clans.8
Among the families listed in Table 6.1, GH Families 2 and 5 (GH2 and GH5) are part of Clan
GH-A, while GH13 and GH77 belong to Clan GH-H. Other families in Table 6.1 are either
the only representatives of their clans with known tertiary structures containing glucosyl or
acarbose-derived carbohydrate ligands in subsites −1 and +1, or they have not been classified
into clans.

GHs hydrolyze glycosidic bonds by two general mechanisms,9 with some exceptions. In
the inverting mechanism, a carboxyl group on an amino acid side-chain donates a proton to
the scissile glycosidic oxygen atom. Meanwhile, the oxygen atom of a water molecule
coordinated by a dissociated carboxyl group on a second amino acid side-chain forms a
partial bond with the C1′ atom of the carbohydrate ring in subsite −1. The complex proceeds
through the transition state with the bond between O5′ and C1′ atoms assuming partial double
bond character. That causes the C5′–O5′–C1′–C2′ atoms to form a plane. Finally, the glyco-
sidic bond is broken and a hydroxyl group from the coordinated water molecule replaces it,
inverting the configuration of the carbohydrate residue. The remaining proton from the water
molecule is accepted by the dissociated carboxyl group.
<table>
<thead>
<tr>
<th>GH family</th>
<th>GH clan</th>
<th>GH family members</th>
<th>No. of structures</th>
<th>Catalytic mechanism</th>
<th>Ligand glycon conformation</th>
<th>Ligand bond configuration</th>
<th>Position of proton donor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>A</td>
<td>β-Galactosidase</td>
<td>1</td>
<td>R</td>
<td>$^4C_1$</td>
<td>β-(1,4)</td>
<td>anti</td>
</tr>
<tr>
<td>3</td>
<td>–</td>
<td>β-Glucan glucohydrolase</td>
<td>2</td>
<td>R</td>
<td>$^4C_1$</td>
<td>β-(1,4)</td>
<td>anti</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>Endocellulase, EG</td>
<td>4</td>
<td>R</td>
<td>$^4C_1$</td>
<td>β-(1,4)</td>
<td>anti</td>
</tr>
<tr>
<td>6</td>
<td>–</td>
<td>CBH, EG</td>
<td>3</td>
<td>I</td>
<td>$^4C_1$, $^2S_o$</td>
<td>β-(1,4)</td>
<td>syn</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>CBH, EG</td>
<td>1</td>
<td>R</td>
<td>Distorted $^{1,4}B$</td>
<td>β-(1,4)</td>
<td>syn</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>EG</td>
<td>1</td>
<td>I</td>
<td>$^{2,5}B$</td>
<td>β-(1,4)</td>
<td>anti</td>
</tr>
<tr>
<td>9</td>
<td>–</td>
<td>CBH</td>
<td>1</td>
<td>I</td>
<td>$^4E$</td>
<td>β-(1,4)</td>
<td>syn</td>
</tr>
<tr>
<td>12</td>
<td>C</td>
<td>EG</td>
<td>2</td>
<td>R</td>
<td>$^1S_3$</td>
<td>β-(1,4)</td>
<td>syn</td>
</tr>
<tr>
<td>13</td>
<td>H</td>
<td>CGTase, α-amylase, neopullulanase, glucon 1,4-α-maltotetraosidase, amylosucrase, 4-α-glucanotransferase, maltogenic amylase</td>
<td>42</td>
<td>R</td>
<td>$^2H_3$, $^4C_1$</td>
<td>α-(1,4)</td>
<td>anti</td>
</tr>
<tr>
<td>14</td>
<td>–</td>
<td>β-Amylase</td>
<td>4</td>
<td>I</td>
<td>$^4C_1$</td>
<td>α-(1,4)</td>
<td>syn</td>
</tr>
<tr>
<td>15</td>
<td>L</td>
<td>Glucoamylase, glucodextranase</td>
<td>5</td>
<td>I</td>
<td>$^2H_3$, $^2E$</td>
<td>α-(1,4)</td>
<td>syn</td>
</tr>
<tr>
<td>44</td>
<td>–</td>
<td>EG</td>
<td>1</td>
<td>R</td>
<td>$^1S_3$</td>
<td>β-(1,4)</td>
<td>anti</td>
</tr>
<tr>
<td>57</td>
<td>–</td>
<td>4-α-Glucanotransferase</td>
<td>1</td>
<td>R</td>
<td>$^2H_3$</td>
<td>α-(1,4)</td>
<td>anti</td>
</tr>
<tr>
<td>77</td>
<td>H</td>
<td>4-α-Glucanotransferase</td>
<td>1</td>
<td>R</td>
<td>$^4C_1$</td>
<td>α-(1,4)</td>
<td>anti</td>
</tr>
</tbody>
</table>

Abbreviations: CBH: cellobiohydrolase; CGTase: cyclomaltodextrin glucanotransferase; EG: endo-1,4-β-glucanase; GH: glycoside hydrolase; I: inverting; R: retaining.
In the retaining mechanism, the protonated carboxyl group donates a proton to the glycosidic oxygen atom, and the dissociated carboxyl group forms a covalent bond with the C1’ atom. The bond between the C1’ atom and the glycosidic oxygen atom is broken and a hydroxyl group from the water molecule hydrates the C1’ atom, breaking the covalent bond with the carboxyl group. This double-displacement reaction retains the anomeric configuration of the glycon. Of the families listed in Table 6.1, five have inverting mechanisms and nine have retaining mechanisms.

1. The protonated carboxyl group is usually positioned laterally on either side of the O4 atom (the glycosidic oxygen atom). The different positions of the proton donor are labeled as syn or anti. Syn proton donors are on the same side as the O5’ atom (the glycon ring oxygen atom), while anti proton donors are on the opposite side (Fig. 6.2).

Figure 6.2. The different location of the proton donor in GHs. The proton donor is labeled as syn if it is on the same side of the glycon ring oxygen atom. If it is on the opposite side, it is labeled as anti. Dihedral twisting is different for syn (right) than for anti (left).
Methods

Gathering data on individual structures

Crystal structures of GHs with ligands spanning subsites −1 and +1 (the cleavage site) were gathered through the CAZy database. They were visualized with PyMOL, and their \( \phi \) and \( \psi \) values were determined. Publications associated with the crystal structures were consulted to confirm ligand placement in the active site and glycon conformation. When more than one catalytic domain occurred in the crystal structure, the torsion angles of their bound ligands were averaged.

The 68 GH structures (supplementary data) with ligands having \( \alpha \)- or \( \beta \)-(1,4) glycosidic bonds over the enzyme cleavage site and with glucosyl or acarbose-derived residues in subsite −1 were investigated further. Residues in subsite +1 were not so restricted, and all mutated enzyme forms were considered equally.

As mentioned earlier, many GHs, especially those acting on \( \beta \)-glycosidic bonds, preactivate the substrate for catalysis by distorting the sugar ring bearing the scissile bond away from the most stable \( ^4C_1 \) form. Crystal structures of GH–ligand complexes show several different glycon conformations (Table 6.1). Many nonoptimal ligand ring conformations are maintained by being part of putative transition-state analogs such as acarbose, a potent inhibitor of amylases and other enzymes that hydrolyze the \( \alpha \)-(1,4) glycosidic bonds of starch and its partial hydrolyzates. The nonreducing-end ring of acarbose is a \( ^2H_3 \) conformer maintained by a double bond between its C5′ and C7′ atoms, the latter replacing the O5′ atom normally found in pyranosyl rings. In addition, a protonated nitrogen atom replaces the
glycosidic oxygen atom and a methyl group replaces the methanol group at the C5 atom of the residue to the reducing side of the scissile bond. Often, GHs incubated with acarbose retain sufficient activity to rearrange it into various derivatives consisting of acarviosinose (the two non-glucosyl residues of acarbose) (Fig. 6.1) and different numbers of glucosyl residues to either side of this fragment.

Most 1,4-glucosyl-acting GHs with known tertiary structures bind six types of ligands (Fig. 6.1), leading to construction of eight hybrid QM::MM maps of $E_{\text{Intra}}$ at various $\phi/\psi$ values of the bound ligands: 1) a map for $\alpha$-maltose (Fig. 6.3, with energies taken from Johnson, et al.\textsuperscript{14}) for GH13, GH14, and GH77 enzymes that bind ligands with $\alpha$-(1,4) scissile bonds and glycon $^4C_1$ conformers; 2) three maps of the $R$, $S$, and positively-charged forms of methyl $\alpha$-acarviosinide (Fig. 6.4) for GH13, GH15, and GH57 members that bind ligands with $\alpha$-(1,4) scissile bonds and glycon $^2H_3$ conformers; 3) a map of $\beta$-$^4C_1$-cellobiose (Fig. 6.5) for GH2, GH5, GH6, and GH7 enzymes that bind ligands with $\beta$-(1,4) scissile bonds and glycon $^4C_1$ conformers; 4) a map of $\beta$-thiocellobiose (Fig. 6.6) for GH3, GH5, and GH6 members that bind ligands with $\beta$-(1,4) scissile bonds and glycon $^4C_1$ conformers linked by a sulfur atom; 5) a map of $\beta$-$^1S_3$-cellobiose (Fig. 6.7) for GH9, GH12 and GH44 enzymes that bind ligands with $\beta$-(1,4) scissile bonds and glycon $^1S_3$ conformers; and 6) a map of $\beta$-$^2S_0$-cellobiose (Fig. 6.8) for a GH8 member that binds ligands with $\beta$-(1,4) scissile bonds and glycon $^2S_0$ conformers.

**Energy determination**

A detailed description of the non-integral hybrid QM::MM method is provided elsewhere.\textsuperscript{15} This method, when used with an elevated dielectric constant, has successfully
predicted the conformations of numerous disaccharide linkages in small-molecule crystal structures, typically finding a majority within the 1 kcal/mol contour,\textsuperscript{14,16,17} and with no distortions greater than 4 kcal/mol.\textsuperscript{15} Furthermore, the distribution of energies follows a Boltzmann-like exponential decay curve, indicating that the method is predictive for condensed-phase systems.\textsuperscript{18} The elevated dielectric constant diminishes by 80% the otherwise dominating influence of intramolecular hydrogen bonds without the need to explicitly consider neighboring molecules. These hybrid maps are constructed from three different $\phi$ $\psi$ surfaces with energies calculated at 20° intervals of $\phi$ and $\psi$.

1. A QM relative energy map to describe the core part of the molecule that includes the sugar rings and glycosidic linkage (the analog). For $\alpha$-maltose and $\beta$-$^4C_1$-cellobiose, the analog consists of a tetrahydropyran (THP) dimer. For the methyl $\alpha$-acarviosinide, $\beta$-thiocellobiose, $\beta$-$^1S_3$-cellobiose, and $\beta$-$^2S_0$-cellobiose maps, the hydroxyl groups of the full disaccharide were replaced by hydrogen atoms, giving a dimer of methyl THP (methyl cyclohexene and methyl THP for the methyl $\alpha$-acarviosinide analog).

2. An MM relative energy map of the analog molecule as above.

3. An MM relative energy map of the full disaccharide to account for the interaction energy of the hydroxyl groups.

The hybrid map is constructed from the above relative energy maps by subtracting the MM analog map from the MM full disaccharide map and adding in the QM analog map. This effectively replaces the MM energy contribution from the core of the molecule with that of QM calculations. The non-integral hybrid method was advantageous in the present work because parameters for the MM force field were not fully developed for the thiocellobiose
and methyl α-acarviosinide molecules.

QM analog maps

The QM analog maps were constructed from QM calculations using GAMESS\textsuperscript{19} or Jaguar version 7.0\textsuperscript{20} software. The QM calculations, except for those for the β\textsuperscript{-4}C\textsubscript{1}-cellobiose analog, were done at the HF/6-31G* level of theory. β\textsuperscript{-4}C\textsubscript{1}-cellobiose analog calculations used B3LYP/6-311++G** theory, based on structures optimized at the B3LYP/6-31G* level. Constraints were used to maintain φ and ψ at each increment.

Additional steps were needed for some of the analog computations. For the β\textsuperscript{-1}S\textsubscript{3}- and β\textsuperscript{-2}S\textsubscript{0}-cellobiose analog calculations, three extra torsion constraints were placed on the non-reducing ring to maintain the desired conformation. Also, some of the β\textsuperscript{-1}S\textsubscript{3}- and β\textsuperscript{-2}S\textsubscript{0}-cellobiose analogs in high-energy regions of φ/ψ space required a constraint on the glycosidic angle to achieve convergence of the minimization. The values of those constrained angles were derived from the comparable MM calculation. Finally, some of the β\textsuperscript{-1}S\textsubscript{3}- and β\textsuperscript{-2}S\textsubscript{0}-cellobiose analog calculations required constraints on the reducing ring to maintain its 4\textsubscript{C1} conformation.

The other QM analog calculations that needed special handling were those for the methyl α-acarviosinide analog, specifically the R and S chiral configurations of the neutral molecules. The single hydrogen atom and the corresponding lone pair attached to the linkage nitrogen atom of the neutral molecules make the linkage nitrogen a chiral atom. However, a constraint was needed on the hydrogen atom to ensure sampling of both chiral configurations. This was accomplished by using the harmonic constraint capability of Jaguar and an improper dihedral angle as a proxy for the chirality. A harmonic constraint was placed on the
C4–C1′–N4–H improper dihedral angle of the molecules with \( R \) and \( S \) chiral configurations. The value of the constraint was \(-135^\circ\) for the \( R \) configuration and \(+135^\circ\) for the \( S \) configuration with a well half-width of 40°. Thus, the value of the improper dihedral was constrained to the value of the target ±40°. The force constant of the constraint was 10 kcal mol\(^{-1}\), which sufficed for most of the calculations at the respective \( \phi/\psi \) points. In some cases, the constraint was not strong enough and the improper dihedral would reverse sign (and chirality) by crossing the ±180° value. In these cases, a progressively higher constraint force constant was used until the constraint was held in the appropriate well or, in rare cases, the computation was aborted.

**MM analog maps**

The MM analog maps were constructed with the MM3(96) program.\(^{21}\) The dielectric constant was set to 1.5, the value used when parameterizing MM3 and recommended for isolated molecules. Variation of the dielectric constant has little effect on the analog maps but was chosen none-the-less for maximum compatibility with the QM calculations. As done during the QM calculations, the \( \phi \) and \( \psi \) values were held to their appropriate values via torsion constraints in the MM3 dihedral driver routines. Also like the QM analog calculations, some of the molecular systems required special handling. For the \( \beta^{-1}S_3 \) and \( \beta^{-2}S_0 \) cellobiose analog calculations, three extra torsion constraints were placed on the nonreducing ring to maintain it in the desired conformation. However, unlike the QM calculations, no additional constraints were needed on either the glycosidic angle or the reducing ring. Since MM3 does not support constraining improper dihedral angles and does not have harmonic constraint capability, a different technique was needed to constrain the hydrogen atom on the
linkage nitrogen of methyl \(\alpha\)-acarviosinide to insure complete sampling of the chiral configurations.

In lieu of using an improper torsion constraint, a constraint was placed on the \(\psi_H\) torsion angle (C5–C4–N4–H). Since this hydrogen atom is also the terminal atom of the improper torsion angle (C4–C1–N4–H), there is a direct relationship between these proper and improper torsion angles. Since the improper torsion angle can assume a range of values, depending on the values of \(\phi\) and \(\psi\), it is necessary to scan a range of the related \(\psi_H\) torsion angle. To do this, a set of values ranging from 120°–240° in 5° increments was subtracted from the \(\psi_H\) torsion angle after \(\psi\) was set. The terminal hydrogen atom attached to the nitrogen atom of the \(\psi_H\) torsion angle was set in a way that only the terminal hydrogen atom was moved. This effectively put the improper torsion angle in the range of \(-120\°\) to \(+120\°\), including \(\pm180\°\), thus sampling the entire chiral range. However, since a chiral assignment cannot be made on the \(\pm180\°\) configurations, those were not included on the maps. This range was not as broad as that of the QM harmonic constraint. However, the QM data showed that the effective limits of the improper torsion were \(-120\°\) to \(-175\°\) and \(120\°\) to \(175\°\) for \(R\) and \(S\) chirality, respectively. Each of the resulting \(\psi_H\) conformations was used to calculate the data points at each required \(\phi/\psi\) value. The lowest energy conformation at each chiral configuration for each \(\phi/\psi\) point was then used for the respective MM3 analog maps.

**MM disaccharide maps**

The MM disaccharide maps were constructed from calculations with MM3(96), except the hydrogen bonding parameters were from the MM3(92) version of the force field and the dielectric constant was set to 7.5. For the chiral methyl \(\alpha\)-acarviosinide maps, the \(\psi_H\) con-
straint used at each $\phi/\psi$ point of the disaccharide map was that of the respective MM analog map at the equivalent $\phi/\psi$ point. A set of starting conformations was generated, based on variations of the exo-cyclic group orientations, and each resulting conformation was considered at each $\phi/\psi$ point of the map. The lowest energy conformation at each $\phi/\psi$ point was used for the respective maps.

Energies for the $\alpha$-maltose map were based on a set of 58 conformations as described in previous work. A total of 2187 conformations for the methyl $\alpha$-acarviosinide molecules were generated by

1. Setting the O5–C1–O1–CH$_3$ torsion angle to 60°.
2. Systematically rotating the remaining 7 rotatable exo-cyclic groups to values of −60°, 180°, and 60°.

Cellobiose and thiocellobiose molecules possess more exo-cyclic groups than methyl $\alpha$-acarviosinide, so a conformational search procedure was employed for them. An outline of the procedure follows:

1. A structure was either obtained from crystal coordinates or sketched with Maestro and minimized with MacroModel version 9.5 using the OPLS-2005 force field.
2. A Monte Carlo multiple minimum conformational search protocol was set up as follows:
   a. All torsion angles, except those of the pyranosyl rings, were set as search variables.
   b. The rings were constrained with three alternating torsion constraints to hold the appropriate ring shape during minimization. The constraints for the $^1S_3$ and $^2S_0$ rings had force constants of 1000 kJ/mol, while the $^4C_1$ rings had a softer force constant of 100 kJ/mol.
c. 30,000 Monte Carlo steps were performed.

d. The number of variables altered in each step was 2–4.

e. The energy window (energy above global minimum at the time) for acceptance of a conformer was 50 kJ/mol (12 kcal/mol).

f. The least used structures were used as starting geometries for subsequent Monte Carlo steps.

g. The similarity criteria included distance checks of atoms as well as torsion angle differences involving polar hydrogen atoms. The distance threshold was 0.25 Å and the torsion threshold was set to 60°.

h. The structures generated at each Monte Carlo step that passed initial checks were minimized with OPLS-2005 for up to 5000 steps.

3. Step 2 was performed ten times, for a total of 300,000 Monte Carlo steps. This seemed to be sufficient to approach convergence of the search.

4. Each member of the resulting conformation set was transformed as follows:

   a. The glycosidic angle was set to 150°.18

   b. The $\phi$ and $\psi$ torsion angles were set to those of the global minimum of the set.

5. A redundant conformer elimination procedure was performed. This used heavy atoms and polar hydrogen atoms for testing the maximal distance threshold of 0.25 Å between conformations. This eliminates redundant conformations after normalizing the linkage variables in step 4 above.

   The numbers of conformations generated for $\beta$-cellobiose and $\beta$-thiocellobiose molecules with the above procedure were:

   - $\beta^{-4}C_{1}$-cellobiose: 1863
• $\beta^{-1}S_3$-cellobiose: 2871
• $\beta^{-2}S_O$-cellobiose: 3485
• $\beta$-thiocellobiose: 2277

These numbers of starting structures are far greater than in any of our previous calculations.

**Results and discussion**

**Energy contour maps**

Eight $\phi/\psi$ energy contour maps were computed using the hybrid QM::MM technique to measure the strain on the glycosidic bond upon twisting. Torsion angles obtained from Protein Data Bank crystal structures of GHs with bound ligands spanning the cleavage site were plotted on the corresponding $\phi/\psi$ energy contour maps to gain insight into the glycosidic bond distortion that occurs upon binding by the enzyme’s active site.

For $\alpha$-1,4 glycosidic bonds we constructed an $\alpha$-maltose and three methyl $\alpha$-acarviosinide energy contour maps (Figs. 6.3, 6.4). Maltose oligomers are the natural substrates of members of several GH families, while methyl $\alpha$-acarviosinide is a powerful inhibitor of maltose- and starch-degrading enzymes. The glycosidic nitrogen atom of methyl $\alpha$-acarviosinide, when singly protonated, may have either the $R$ or the $S$ chiral configuration. It can also be doubly protonated, making the molecule positively charged. X-ray crystal structure studies of proteins are not usually able to provide hydrogen atom positions, so the protonation state of methyl $\alpha$-acarviosinide could not be determined. Therefore, we plotted methyl $\alpha$-acarviosinide torsion angle measurements on all three maps (methyl $\alpha$-$R$-acarviosinide, methyl $\alpha$-$S$-acarviosinide, and methyl $\alpha$-acarviosinide-$H^+$).
Similarly, for β-1,4 glycosidic bonds we constructed maps of β-cellobiose and β-thiocellobiose with $^4C_1$ glycon conformations (Figs. 6.5, 6.6). As mentioned above, many cellulases (as well as other GHs) distort the glycon away from the most stable $^4C_1$ conformation. The most common distortion observed in cellulase crystal structures is to a $^1S_3$ (or nearby) conformation (Fig. 6.7). The $^2S_0$ (or nearby) conformation has been found in some inverting cellulases (Fig. 6.8).

**Figure 6.3.** $E_{\text{intra}}$ contour map of α-maltose with its two α-glucosyl residues in $^4C_1$ conformations, and with crystal-structure ligand torsion angles plotted on it. Green, GH13; yellow, GH14; purple, GH77.
**α-1,4 Glycosidic bonds**

The α-maltose energy contour map has a global minimum near $\phi = 100^\circ$, $\psi = -140^\circ$ (Fig. 6.3). The three methyl α-acarviosinide maps (methyl α-$R$-acarviosinide, methyl α-$S$-acarvio-

![Image of energy contour maps]

**Figure 6.4.** $E_{\text{intra}}$ contour map of methyl α-acarviosinide with its glycon in the $^2H_3$ conformation and its aglycon in the $^4C_1$ conformation, and with crystal-structure ligand torsion angles plotted on it. A) methyl α-$R$-acarviosinide; B) methyl α-$S$-acarviosinide; C) methyl α-acarviosinide-$H^+$. Green, GH13; cyan, GH15; red, GH57

sinide, and methyl α-acarviosinide-$H^+$) are qualitatively very similar to each other, the main
difference among them being the position of their global minima (Fig. 6.4). The methyl α-S-acarviosinide and methyl α-acarviosinide-H⁺ maps have global minima near $\phi = 70^\circ$, $\psi = -140^\circ$, while that in the methyl α-R-acarviosinide map is near $\phi = 130^\circ$, $\psi = -90^\circ$. The global minimum regions are smaller in all three methyl α-acarviosinide maps than in the α-

Figure 6.5. $E_{\text{intra}}$ contour map of β-cellobiose with its two β-glucosyl residues in $^4C_1$ conformations, and with crystal-structure ligand torsion angles plotted on it. Cyan, GH2; pink, GH5; red, GH6; yellow, GH7 maltose map, but a second low-energy region is present in all of the former. All four maps have low-energy valleys running along $50^\circ < \phi < 140^\circ$, with less deep ones in the range of
−180° < ψ < −100°. The methyl R- and S-α-acarviosinide maps are generally similar to the previously published MM3 map that combined the energies for both forms, but the maps for the protonated form have some important differences.25 Those MM3 maps were based on a much smaller number of starting geometries as well as a lower dielectric constant. The pure MM3 map for the protonated form has a global minimum at $\phi = 67^\circ$ and $\psi = 42^\circ$ (after

**Figure 6.6.** $E_{\text{intra}}$ contour map of β-thiocellobiose with its two β-glucosyl residues in $^4C_1$ conformations, and with crystal-structure ligand torsion angles plotted on it. Green, GH3; pink, GH5; red, GH6 converting from $\phi_H$ and $\psi_H$), whereas the global minimum on the hybrid map is near $\phi = 70^\circ$, $\psi = −140^\circ$. Although each global minimum corresponds to a secondary minimum on the
previous map, a very low-energy minimum on the hybrid map has no close counterpart on the published pure MM3 map. The observed crystal structure conformations are clustered around the global minimum on the hybrid map.

**β-1,4 Glycosidic bonds**

The β-1\(^1\)-cellobiose map has two low-energy valleys (Fig. 6.5). The first one runs along $-130^\circ < \phi < -50^\circ$, while the second one runs along $-160^\circ < \psi < -80^\circ$. The global minimum

---

**Figure 6.7.** $E_{\text{intra}}$ contour map of β-cellobiose with its glycon and aglycon β-glucosyl residues in $1S_3$ and $2C_1$ conformations, respectively, and with crystal-structure ligand torsion angles plotted on it. Cyan, GH9; red, GH12; green, GH44
is near $\phi = -60^\circ$, $\psi = -120^\circ$. Energy calculations on this new map differ from those on previous QM::MM3 cellobiose maps\textsuperscript{4,16,17} in that the number of starting geometries was greatly expanded and a dielectric constant of 7.5 was employed instead of 3.5 as in Refs. 16 and 17. Still, the features of these maps are quite similar.

The $\beta$-thiocellobiose map is similar to the $\beta$-$4^C_1$-cellobiose map, having the same valleys running along $\phi$ and $\psi$ (Fig. 6.6). However, the $\beta$-thiocellobiose map has two widely-separated minima of almost the same $E_{\text{Intra}}$ value. The first one lies near $\phi < -60^\circ$, $\psi < -120^\circ$, as in

![Figure 6.8. $E_{\text{Intra}}$ contour map of $\beta$-cellobiose with its glycon and aglycon $\beta$-glucosyl residues in $^{2}S_0$ and $^{4}C_1$ conformations, respectively, and with crystal-structure ligand torsion angles plotted on it. Green: GH8](image-url)
the $\beta^{-4}C_1$-cellobiose map, while the second one is near $\phi < -70^\circ$, $\psi < -290^\circ$.

Distortion of the glycon ring significantly reduces the allowed (low-energy) $\phi/\psi$ regions. The difference is most noticeable in the $\beta^{-1}S_3$-cellobiose energy contour map, where the low-energy region is contained only within $-190^\circ < \phi < -70^\circ$, $-200^\circ < \psi < -60^\circ$ (Fig. 6.7). The global minimum in the $\beta^{-1}S_3$-cellobiose map is shifted 40° down in both $\phi$ and $\psi$ directions compared to the $\beta^{-4}C_1$-cellobiose map, and the energy in other regions is much higher than that in the $\beta^{-4}C_1$-cellobiose map, with energy values >40 kcal/mol above the global minimum at the mountaintops. The main reason for the high energy values in the $\beta^{-1}S_3$-cellobiose map is the steric clash between the bulky C6-OH hydroxyl groups of both sugar rings, which occurs more often with the $^1S_3$ disaccharide because the aglycon ring is more axial to the glycon.

In the $\beta^{-2}S_0$ conformation the aglycon is pseudo-axial to the glycon. This causes C6-OH clashes in $\beta^{-2}S_0$-cellobiose to be less frequent than in $\beta^{-1}S_3$-cellobiose, and the energy range from global minimum to mountaintop is not as pronounced in $\beta^{-2}S_0$-cellobiose as in $\beta^{-1}S_3$-cellobiose (Fig. 6.8). The global minima in the $\beta^{-2}S_0$- and the $\beta^{-4}C_1$-cellobiose maps are in very similar positions (contained within $-100^\circ < \phi < -50^\circ$ and $-180^\circ < \psi < -100^\circ$).

**Ligand intramolecular energies and torsion angles**

Values of $\phi$ and $\psi$ from crystal structures are plotted on Figs. 6.3–6.8. GHs in different families that cleave $\alpha$- or $\beta$-1,4 bonds twist these bonds to different values of $\phi$ and $\psi$, corresponding to different $E_{\text{Intra}}$ values. Table 6.2 shows the average $E_{\text{Intra}}$ value for each GH family. Values of $E_{\text{Intra}}$ shown in parentheses in Table 6.2 are of crystal structures of wild-
type GHs or of those having slight mutations (e.g. Glu to Gln) binding the natural substrate. Interestingly, those enzymes with more radical mutations bind ligands with the same or lower $E_{\text{Intra}}$ values than the average $E_{\text{Intra}}$ of the family, suggesting that these mutations interfere with the enzyme’s ability to twist glycosidic bonds optimally for reaction.

GHs that contain a proton donor in the \textit{anti} position twist the scissile bond in the opposite direction (with respect to the lowest energy point on the map) than GHs with a \textit{syn}-positioned proton donor. In GHs that break $\alpha$-1,4-glycosidic bonds, the data point to a preference

\begin{table}[h]
\centering
\begin{tabular}{llc}
\hline
GH family & Map & Average $E_{\text{Intra}}$ in all structures, kcal/mol \\
\hline
2 & $\beta$-4$\text{C}_1$-Cellobiose & 1.2 \\
3 & $\beta$-Thiocellobiose & 2 \\
5 & $\beta$-4$\text{C}_1$-Cellobiose & 2 (4)$^a$ \\
5 & $\beta$-Thiocellobiose & 0.5 \\
6 & $\beta$-4$\text{C}_1$-Cellobiose & 5.2 \\
6 & $\beta$-Thiocellobiose & 1.8 \\
7 & $\beta$-4$\text{C}_1$-Cellobiose & 1.5 \\
8 & $\beta$-2S$_3$-Cellobiose & 2.5 (2.5) \\
9 & $\beta$-1S$_3$-Cellobiose & 1.5 (1.5) \\
12 & $\beta$-1S$_3$-Cellobiose & 1.9 (1.9) \\
13 & $\alpha$-Maltose & 4.3 (6.2) \\
13 & Methyl $\alpha$-acarviosinide & 3.6 \\
14 & $\alpha$-Maltose & 9.5 (9.7) \\
15 & Methyl $\alpha$-acarviosinide & 2.6 \\
44 & $\beta$-1S$_3$-Cellobiose & 2.1 (2.1) \\
57 & Methyl $\alpha$-acarviosinide & 2.2 \\
77 & $\alpha$-Maltose & 1.3 (1.3) \\
\hline
\end{tabular}
\caption{$E_{\text{Intra}}$ values for GH families containing crystal structures of enzymes with oligosaccharides bound in subsites $-1$ and $+1$}
\end{table}

$^a$Average $E_{\text{Intra}}$ of wild-type structures and those with slight mutations (e.g. Glu to Gln)

for ligands bound to GHs with an \textit{anti}-positioned proton donor that lies to the left (lower $\phi$ values) of the minimum (Figs. 6.3, 6.4). Ligands bound to GHs with a \textit{syn}-positioned proton donor lie to the right (higher $\phi$ values) of the minimum. For GHs that break $\beta$-1,4 glyco-
sidic bonds, the opposite is true. Data for GHs with anti-positioned proton donors are located to the right of the lowest energy point on the maps, while the left side of the minimum is populated by data from GHs containing a syn-positioned proton donor. This suggests that twisting of the glycosidic bond is necessary for the correct orientation of one of the glycosidic oxygen lone pairs toward the proton donor. The only exceptions are the three GH13 outliers with a mutated proton donor (see below) in the α-maltose map, one point in the methyl α-acarviosinide map for a GH57 member, a GH2 member in the β-cellobiose map, and a GH44 member in the β-1S₃-cellobiose map. However, the latter two are very close to the global minimum.

The majority of GH13 members and the one GH77 member, all part of Clan H, bind ligands of both ⁴C₁ (α-maltose) and ²H₃ (methyl α-acarviosinide) glycon conformations with torsion angles of 10° < φ < 60° and −170° < ψ < −130° (Figs. 6.3, 6.4). This is consistent with the fact that methyl α-acarviosinide is an excellent inhibitor of maltose-degrading enzymes. However, three GH13 structures have ligands with φ values from 130° to 170° and with ψ values between −110° and −90° (Fig. 6.3). These are the only structures in which the catalytic proton donor (Glu257) has been mutated to alanine, which probably caused their different locations on the φ/ψ map. A superimposition of GH13 members shows a significantly different sugar binding position in the three outliers. Average values of E_{Intra} for all GH13-bound maltose structures are greater than the values found at the global minimum by ≈ 4.3 kcal/mol (Fig. 6.3). Wild-type GH13 enzymes bind α-maltose oligomers with higher E_{Intra} values, an average of ≈ 6.2 kcal/mol. For GH13 members twisting the methyl α-acarviosinide glycosidic bond, the average value of E_{Intra} is ≈ 3.6 kcal/mol. The one GH77 mem-
ber available for analysis imposes strain with an $E_{\text{Intra}}$ value of $\approx 1.3$ kcal/mol.

All four GH14 members bind ligands with glycons in the $^4C_1$ conformation. Their torsion angles are grouped on the $\alpha$-maltose map far from those of GH13 and GH77 members, near $\phi = 140^\circ$, $\psi = -260^\circ$ (Fig. 6.3). GH14 members distort the glycosidic bond to $E_{\text{Intra}}$ values of $\approx 9.5$ kcal/mol on average, imposing the greatest scissile bond strain observed in GHs so far. Wild-type GH14 members distort the glycosidic bond to the slightly higher $E_{\text{Intra}}$ value of $\approx 9.7$ kcal/mol.

The five GH15 members and one GH57 member bind ligands with glycons having $^2H_3$ conformations, since methyl $\alpha$-acarviosinide groups are found in subsites $-1$ and $+1$. Their torsion angles are found around $\phi = 100^\circ$ and $\psi = -120^\circ$ to $-110^\circ$ (Fig. 6.4), distant from GH13, GH14, and GH77 members found in Fig. 6.3 and GH13 members found in Fig. 6.4. GH15 and GH57 enzymes bind methyl $\alpha$-acarviosinide and twist the glycosidic bond torsions to $E_{\text{Intra}}$ values of $\approx 2.6$ and $\approx 2.2$ kcal/mol, respectively.

Figs. 6.5–6.8 show the torsion angles of GHs that hydrolyze $\beta$-1,4 glycosidic bonds. Because oxygen and sulfur glycosidic atoms yield substantially different valence angles and bond lengths, comparisons between members of the same GH family (in this case, GH5 and GH6) having ligands with different glycosidic atoms should be made with caution.

On average, GHs that hydrolyze $\beta$-1,4-glycosidic bonds force their ligands to have torsion angles that yield lower $E_{\text{Intra}}$ values than those imposed by GHs that hydrolyze $\alpha$-1,4-glycosidic bonds.

GH5 members bind cello-oligosaccharides with an imposed average $E_{\text{Intra}}$ value of $\approx 2$ kcal/mol, with one of the GH5-bound ligands lying very close to the global minimum of the
β-C\(_1\)-cellobiose map and the other lying \(\approx 4\) kcal/mol above the minimum (Fig. 6.5). There are also two GH5 structures that bind thiocello-oligosaccharide ligands with an average \(E_{\text{Intra}}\) value of \(\approx 0.5\) kcal/mol (Fig. 6.6). For GH2, which is part of the same Clan A as GH5, there is only one available crystal structure for analysis, and it binds a lactose molecule. The only difference between cellobiose and lactose is the configuration of the hydroxyl group attached to the glycon C4 atom, and their \(\phi/\psi\) maps are very similar to each other.\(^{26}\) The available data point for GH2 would fall near the minimum of the \(\beta-\)\(C\(_1\)-cellobiose map, at \(\approx 1.2\) kcal/mol (Fig. 6.5).

There are three GH6 data points, two for cellooligosaccharide-bound complexes with an average \(E_{\text{Intra}}\) value of \(\approx 5.2\) kcal/mol (Fig. 6.5) and one for a thiocello-oligosaccharide-bound complex with a significantly different \(\phi\) value and an \(E_{\text{Intra}}\) value of \(\approx 1.8\) kcal/mol (Fig. 6.6). The last point in the \(\beta\)-thiocellobiose map is for a GH3-bound ligand with \(E_{\text{Intra}}\approx 2\) kcal/mol. Only one structure is available for GH7, and it binds a cello-oligosaccharide with \(E_{\text{Intra}}\approx 1.5\) kcal/mol.

GH12 and GH44 members bind cellooligosaccharides with \(\text{\(^1\)S}_3\)-distorted sugar rings to their active sites. GH9 members twist the sugar ring in the active site to a \(\text{\(^4\)E}\) conformation, which is very close to the \(\text{\(^1\)S}_3\) conformation. Therefore, we included the GH9 data point in the \(\beta-\)\(\text{\(^1\)S}_3\)-cellobiose energy contour map (Fig. 6.7). \(E_{\text{Intra}}\) for the GH9-bound cello-oligosaccharide substrate is \(\approx 1.5\) kcal/mol above the \(\phi/\psi\) minimum. For GH12, \(E_{\text{Intra}}\approx 1.9\) kcal/mol and for GH44, \(E_{\text{Intra}}\approx 2.1\) kcal/mol. There is only one available structure to plot on the \(\beta-\)\(\text{\(^2\)S}_0\)-cellobiose energy contour map (Fig. 6.8), and it is from GH8 with \(E_{\text{Intra}}\approx 2.5\) kcal/mol compared to the \(\text{\(^2\)S}_0\)-cellobiose minimum energy structure. In general, the \(E_{\text{Intra}}\) values of \(\text{\(^1\)S}_3\)-
or $^2S_O$-distorted structures are not significantly different from the $E_{\text{intra}}$ values of undistorted ($^1C_1$) structures bound to other GHs that break $\beta$-1,4 glycosidic bonds.

**Conclusions**

We have calculated hybrid QM::MM $\phi/\psi$ energy maps for $\alpha$-maltose, methyl $\alpha$-acarviosinide, and $\beta$-thiocellobiose, as well as for $\beta$-cellobiose in three different ring shapes, to measure the strain on different glycosidic bonds upon twisting. We have also plotted data obtained from protein–oligosaccharide complex crystal structures to estimate the protein-imposed strain on their substrates. In general, members of the same GH family twist the scissile bond similarly, while members of different families twist the bond to different regions on the energy contour maps. The GH-imposed strain ranges from 0.5 to 9.5 kcal/mol. GHs that break $\beta$-1,4 glycosidic bonds and have an anti-positioned proton donor distort the scissile glycosidic bond to higher $\phi$ values with respect to the minimum. GHs with syn proton donors twist the bond to lower $\phi$ values. The opposite is true in GHs that break $\alpha$-1,4 glycosidic bonds.

Distortion of the bond is mechanistically relevant, as it helps to orient one of the lone pairs in the glycosidic oxygen atom toward the proton donor, assisting proton transfer.

**Acknowledgments**

L.P. and P.J.R. thank the U.S. Department of Agriculture for its financial support through the Biotechnology Byproducts Consortium and through National Research Initiative Grant 2007-35504-18252 from the USDA Cooperative State Research, Education and Extension Service. G.P.J. and A.D.F. were supported by normal funds of the USDA Agricultural
Research Service. The authors thank Wim Nerinckx (University of Ghent) for his very helpful advice.

References


235–248.

Chapter 7: Molecular Mechanism of the Glycosylation Step Catalyzed by Golgi α-Mannosidase II

Luis Petersen, Albert Ardèvol, Carme Rovira, and Peter J. Reilly

Abstract

Golgi α-mannosidase II (GMII), a member of glycoside hydrolase family 38, cleaves two mannosyl residues from GlcNAcMan5GlcNAc2 as part of the N-linked glycosylation pathway. Inhibition of GMII has shown promise in developing anti-cancer therapies, but more selective inhibitors are required for clinical use. To gain insight into the molecular mechanism of the enzyme and the structural and electronic features of the reaction transition state (TS), we performed quantum mechanics/molecular mechanics Car–Parrinello metadynamics simulations of the glycosylation reaction catalyzed by GMII. The calculated energy barrier for mannosyl glycosylation (23 kcal/mol) agrees reasonably well with experimental estimates. Our simulations show that GMII features an oxacarbenium ion-like TS where the nucleophilic attack occurs before complete departure of the leaving group, consistent with a D_{N}A_{N} reaction mechanism. The conformation of the glycon mannosyl ring in the TS is close to a B_{2,5} boat, confirming experimental predictions. Evidence that the Zn ion has a catalytic role arises from the concerted variation of the Zn and anomeric charges along the reaction. We propose that the interaction between the Zn ion and the O2’ atom increases the acidity of the latter to facilitate the observed partial deprotonation of the O2’–H_{O2’} bond upon format-

1 Parc Cientific Barcelona, Barcelona, Spain
ion of the electron-deficient reaction oxacarbenium ion-like species, thereby providing a stabilizing effect. This information could be very valuable in designing TS mimics as powerful and selective inhibitors against GMII.

**Introduction**

N-linked glycosylation is a post-translational process that attaches a carbohydrate to a nascent protein and further modifies it.\(^1\) The process involves a series of well-defined carbohydrate trimming and addition steps performed by several glycoside hydrolases (GHs) and transferases, respectively as described chapter 2 (Figure 2.7). The carbohydrate trimming steps start in the endoplasmic reticulum, followed by additional trimming and modification in the Golgi apparatus, where high-mannosyl oligosaccharides are converted to more complex ones.

Breast, colon, and skin cancer cells have an unusual distribution of cell-surface complex oligosaccharides, with increased GlcNAc branching on the trimannosyl core\(^2\) correlated with disease progression.\(^3,4\) Inhibition of Golgi \(\alpha\)-mannosidase (GMII) with swainsonine blocks the abnormal formation of complex oligosaccharides, leading to reduced metastasis and tumor growth.\(^3,5\) However, swainsonine also inhibits the closely related lysosomal \(\alpha\)-mannosidase, limiting its clinical use.\(^3,6\)

GMII cleaves \(\alpha\)-1,3 and \(\alpha\)-1,6 glycosidic bonds between mannopyranosyl residues in GlcNAcMan\(_2\)GlcNAc\(_2\) (Figure 2.7).\(^7,8\) Its amino acid sequence places it in GH family 38.\(^9\) Several kinetic and crystallographic studies have clarified the general mechanism and mode of action of this enzyme.\(^8,10-14\) The enzyme operates by a double-displacement mechanism that leads to net retention of the anomeric configuration.\(^15\) In the rate-limiting glycosylation
step, a protonated carboxylate amino acid, Asp341 in *Drosophila melanogaster* GMII, transfers a proton to the oxygen atom of the scissile glycosidic bond. A second, negatively-charged carboxylate amino acid, Asp204, acts as the nucleophile, attacking the glycon C1’ atom to form a covalent intermediate with the protein. In the deglycosylation step, a water molecule replaces the aglycon molecule cleaved off during the glycosylation step. This water molecule, now coordinated with Asp341, acts as the nucleophile and attacks the C1’ atom, breaking the covalent bond to finish the reaction and regenerate the enzyme for the next catalytic round.

The available structural data on GMII originates from *D. melanogaster*, which has kinetic and inhibition properties similar to those of mammalian GMII. These two forms also have high sequence identity (41%), with their active-site amino acids being virtually identical. Crystal structures of the Michaelis complex and covalent intermediate of the mannosyl substrate and related structures bound in the protein active site are available. Nevertheless, structural information for the transition state (TS), of interest for the design of specific GMII inhibitors, is very difficult to obtain experimentally. Computational approaches, in particular first-principles simulations, can be very useful to obtain this information, as exemplified by recent studies for other GH enzymes. To the best of our knowledge, the catalytic mechanism of GMII has not previously been addressed by first-principles methods.

The active site of *D. melanogaster* GMII is shown in Figure 7.1. Asp341 and Asp204 are the catalytic residues. The active site also accommodates a catalytic zinc ion that coordinates with Asp204 and with two substrate glycon hydroxyl groups, C2’–OH and C3’–OH (Figure 7.1). The Zn ion also coordinates with His90, Asp92, and His471. The presence of a catalytic zinc ion is unprecedented in GHs, and its precise role during catalysis is unclear.
As observed in many other GHs, GMII binds the substrate in a distorted conformation. Specifically, the glycon mannosyl residue (M5) has $E_5^O\,H_5$ and $1\,S_5$ conformations rather than the undistorted $^4C_1$ form in the Michaelis complex and covalent intermediate, respectively. This has led to the suggestion that the glycosylation step catalyzed by GMII features a $B_{2,5}$-type TS.

![Figure 7.1. Active site of GMII complexed with GlcNAcMan$_3$GlcNAc. The pyranose rings are colored with respect to ring distortion, $^4C_1$ chairs are in green and the $6S_2/B_{2,5}$ shown in the active site is colored cyan.](image)

We report here a quantum mechanics/molecular mechanics (QM/MM) metadynamics simulation of the *D. melanogaster* GMII-catalyzed glycosylation reaction during hydrolysis of an α-1,6-mannosyl bond. By following the changes of the glycon mannosyl residue
throughout the reaction pathway, we obtain the conformational itinerary and describe the
main electronic changes during catalysis, in particular the conformation of the mannosyl
residue at the TS and the charge variation of the active-site zinc ion. This study unravels the
GMII molecular mechanism and is of interest in designing future potential GMII inhibitors as
anti-cancer agents.

Computational methods

Classical molecular dynamics simulations

The initial structure for the simulation was taken from a recently published
crystallographic study of *D. melanogaster* GMII (PDB 3CZN) complexed with its natural
substrate and with its catalytic nucleophile mutated (D204A). The mutation was
computationally reverted to Asp204, using other unmutated structures as a template.
Hydrogen atoms were added using the AmberTools package. The catalytic acid/base,
Asp341, was modeled in its protonated state, while Asp204 was modeled in its deprotonated,
charged state, consistent with the glycosylation step of the retaining mechanism. The
protonation states and hydrogen atom positions of all other ionizable amino acid residues
were selected based on their hydrogen bond network. Thirty-six histidine residues were
modeled in their neutral states, while the rest were modeled in their charged states. All the
water molecules were retained and extra water molecules were added to form a 10-Å water
box around the protein surface. Two chloride ions were also added to neutralize the enzyme
charge.

The system was then submitted to classical molecular dynamics (MD) using Amber9
software. The protein was modeled with the FF99SB force field. All carbohydrates,
including the substrate and GlcNAc glycosylation site at Asn194, were modeled with the GLYCAM06 force field. Finally, all water molecules were described with the TIP3P force field.

The MD simulation was carried out in several steps. First, the system was minimized, holding the protein and substrate fixed. Then, the entire system was allowed to relax. To gradually reach the desired temperature of 300 K in the MD simulation, weak spatial constraints were initially added to the protein and substrate, while the water molecules and chloride ions were allowed to move freely. The constraints were then removed and the MD was extended to 1 ns when the system had reached equilibrium. Because the force field was unable to retain some important active-site hydrogen bonds, two distance constraints were used during the classical MD simulation, one between the proton of the catalytic acid/base and the glycosidic oxygen atom (H_{Asp341}–O6), which is important for the glycosylation reaction, and a second between Asp92 and the C2’–OH hydroxyl group (H_{O2}–O_{Asp92}), a hydrogen bond experimentally observed in the Michaelis complex (PDB 3BUP). These constraints were removed for the subsequent QM/MM and metadynamics simulations, when these interactions were treated with Density Functional Theory (DFT). A snapshot of the MD-equilibrated system was taken for the subsequent QM/MM simulation.

**QM/MM molecular dynamics simulations**

The QM/MM method used here, developed by Laio et al., integrates Car–Parrinello MD, based on DFT, with force-field MD methodology. A detailed description of the QM/MM method can be found in chapter 3. Previous studies showed that this methodology provides an accurate description of energetic, dynamic, and structural features of biological
systems, including GHs.28

The QM region chosen in this work included the side chains of His90, Asp92, Asp204, Asp341, and His471, capped at their Cα atoms with a link-atom pseudopotential, and two active-site mannosyl residues, M3 and M5 (Figure 7.1), capped at the C1 of M4 and C6 of M1. The zinc ion was also included. The QM region (92 atoms) was enclosed in an isolated 19.5 x 15.7 x 19.7 Å supercell. Kohn–Sham orbitals were expanded in a plane-wave basis set with a kinetic energy cutoff of 80 Ry and Troullier–Martins ab initio pseudopotentials29 were used. The BLYP functional30 in the generalized gradient-corrected approximation of DFT was selected in view of its good performance in previous works of Zn-containing enzymes.31 A constant temperature of 300 K was reached by coupling the system to a Nosé–Hoover thermostat32 at 3500 cm⁻¹.

Structural optimizations were done using QM/MM MD with annealing of electronic velocities until the maximal component of the nuclear gradient was <10⁻⁴ au. The time step used for the annealing simulations was 0.072 fs and the fictitious electron mass was 1000 au. A time step of 0.12 fs and a fictitious electron mass of 700 au were used in the Car–Parrinello simulations.

QM/MM metadynamics simulations

The metadynamics technique33 was designed to overcome energy barriers and to reconstruct free energy landscapes of complex systems in MD simulations. The general details of this method are described Chapter 3.

The collective variables (CVs) used in this work were taken as a combination of coordination indices of the covalent bonds being formed/broken (Figure 7.2). Each CV is the differ-
ence of coordination number ($CN$)$^{34}$ of two bonds and is given by

$$CN_{ij} = \frac{1 - \left(\frac{d_{ij}}{d_0}\right)^p}{1 - \left(\frac{d_{ij}}{d_0}\right)^{p+q}}$$

where $d_{ij}$ is the internuclear distance of the atoms involved, $d_0$ is the threshold distance for bonding, and $p$ and $q$ are exponents that determine the steepness of $CN_{ij}$ decay with respect to $d_{ij}$. $CN$ values range from 0 (no bond) to 1 (a bond). There are two CVs in this system: CV1 is the difference in $CN$ between the scissile glycosidic bond and the bond between the nucleophile, Asp204, and the anomeric carbon atom C1’ ($CN_{C1'–O6} – CN_{C1'–OAsp204}$), while CV2 is the difference of $CN$ between the $H_{Asp341–OAsp341}$ and the O6–$H_{Asp341}$ bonds ($CN_{HAsp341–O6} – CN_{HAsp341–OAsp341}$). The latter models the proton transfer between the catalytic acid/base, Asp341, and the scissile glycosidic oxygen atom. The selected values for the $CN$ parameters were $d_0 = 4.72$ au, $p = 10$, and $q = 6$ for CV1, and $d_0 = 2.55$ au, $p = 14$, and $q = 6$ for CV2.

An extended Lagrangian version of the method is here used for a proper coupling with the QM/MM simulations.$^{27,34}$ The selected mass values of the fictitious particles were 15 and 5 amu for CV1 and CV2, respectively, while those of the force constant were 1.3 and 1.5 au, respectively, for CV1 and CV2. The height of the Gaussian terms was 1.25 kcal mol$^{-1}$, which ensures sufficient accuracy for reconstructing the free energy surface. The width of the Gaussian terms (0.05 Å) was selected from the oscillations of the CVs in a free Car–Parrinello QM/MM simulation. A new Gaussian-like potential was added every 300 MD steps. The simulation was performed in an AMD Opteron cluster and required about 144,000 CPU h.
Results and discussion

Metadynamics simulation of GMII

A QM/MM metadynamics simulation was used to model the GMII-catalyzed glycosylation step during hydrolysis of the α-1,6 glycosidic bond that links the M3 and M5 residues in GlcNAcMan$_5$GlcNAc$_2$ (Figure 7.1). A total of 40.6 ps of QM/MM MD was required to model the reaction and obtain the free energy surface (FES) (Figure 7.3), first 4.7 ps of free dynamics followed by 35.9 ps of metadynamics. The metadynamics simulation started with the substrate–GMII complex in the reactant state, which was explored for ~4.8 ps until the forward reaction occurred. The scissile glycosidic bond was cleaved and the
mannosyl–Asp204 covalent intermediate was formed. The system explored the product state (left side of the FES) for ~10.8 ps and then the middle region of the FES (corresponding to states with oxacarbenium ion (OCI)-like character) for a further 7.8 ps. Next, the system returned to the product state for another 8.4 ps. Finally, 4.1 ps later, the mannosyl glycosidic bond was reformed and the system reached the reactant state. The simulation was stopped at this point. Approximately 950 Gaussian hills were added during the metadynamics simulation.

The FES and the minimal energy reaction pathway

The reconstructed reaction (FES) is shown in Figure 7.3. Here, the relevant stationary points corresponding to reactant, TS (the maximal energy point along the reaction pathway), and product (the covalent intermediate) states can be identified. The calculated energy barrier of the reaction is ~23 kcal/mol, which reasonably agrees with the experimentally measured activation energy of 17 kcal/mol for GH38 \textit{Aspergillus fischeri} \(\alpha\)-mannosidase hydrolyzing the \(\alpha\)-Man-(1\(\rightarrow\)6)-Man bond.\footnote{The exothermicity of the reaction, computed as the free energy between the Michaelis complex and covalent intermediate, is ~22 kcal/mol.} A detailed description of the glycosylation reaction in GMII was obtained by following the minimal free energy pathway, as given by the intrinsic reaction coordinate.\footnote{Snapshots of average structures for the reactant, TS, and product states are shown in Figure 7.4. At the reactant state, the catalytic acid/base (Asp341) is in its protonated state, forming a hydrogen bond (1.74 Å) with the glycosidic oxygen atom (O6).}
When the reaction starts, the nucleophile (Asp204) is in its charged state and is significantly separated (3.25 Å) from the C1’ atom (Table 7.1). Two additional carboxylate groups form strong hydrogen bonds with the glycon. Asp92 interacts with the C2’–OH hydroxyl group and Asp442 forms a bidentate interaction with the C3’–OH and C4’–OH groups. The zinc ion is hexacoordinated by His90, Asp92, Asp204, His471, and two substrate oxygen atoms, O2’ and O3’, as previously observed in the crystal structures.\textsuperscript{13}

The reaction begins with the Asp341 O–H bond lengthening along the minimal free energy pathway from the reactant state (R) to R’ (Figure 7.3), indicating partial transfer of
the Asp341 proton. At R’, the Asp341 proton is 1.35 Å from the O6 glycosidic oxygen atom (Table 7.1). The system then reaches the TS when the scissile bond elongates to ~2 Å and the distance between the nucleophile, Asp204, and the C1’ shortens from 3.25 Å in the reactant state to 2.93 Å in the TS. The Asp341 proton transfers further to within 1.19 Å of O6. In addition, the intra-ring C1’–O5’ bond shrinks significantly from 1.41 Å in the reactant state to 1.32 Å in the TS, indicating the formation of a partial double bond and the increasing OCI character of the mannosyl substrate. The substrate is expected to reach its maximal OCI character at CV1 ~ 0, where both the C1’–OAsp204 and C1’–O6 bonds are broken (i.e. their coordination numbers vanish). This corresponds to P’ in the FES (Figure 7.3).

![Figure 7.4. Upper panel: Schematic of the GMII-catalyzed glycosylation reaction. Bottom: Snapshots along the reaction pathway. The Zn ligands (His90 and His471, Figure 2) have been omitted.](image-url)
Table 7.1. Lengths (Å) of Important Bonds along the Reaction Pathway

<table>
<thead>
<tr>
<th></th>
<th>Reactant (R)</th>
<th>R’</th>
<th>TS</th>
<th>P’</th>
<th>Product (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1'–O6</td>
<td>1.53 ± 0.05(^a)</td>
<td>1.56 ± 0.06</td>
<td>2.00 ± 0.23</td>
<td>2.57 ± 0.10</td>
<td>3.47 ± 0.11</td>
</tr>
<tr>
<td>C1'–OAsp204</td>
<td>3.25 ± 0.06</td>
<td>3.25 ± 0.12</td>
<td>2.93 ± 0.11</td>
<td>2.56 ± 0.14</td>
<td>1.50 ± 0.04</td>
</tr>
<tr>
<td>OAsp341–HAsp341</td>
<td>1.03 ± 0.03</td>
<td>1.12 ± 0.06</td>
<td>1.27 ± 0.02</td>
<td>1.73 ± 0.05</td>
<td>1.56 ± 0.03</td>
</tr>
<tr>
<td>HAsp341–O6</td>
<td>1.74 ± 0.05</td>
<td>1.35 ± 0.04</td>
<td>1.19 ± 0.03</td>
<td>1.02 ± 0.03</td>
<td>1.03 ± 0.02</td>
</tr>
<tr>
<td>C1’–O5’</td>
<td>1.41 ± 0.03</td>
<td>1.40 ± 0.03</td>
<td>1.32 ± 0.03</td>
<td>1.29 ± 0.02</td>
<td>1.41 ± 0.04</td>
</tr>
<tr>
<td>O2’–H2’</td>
<td>1.04 ± 0.02</td>
<td>1.04 ± 0.03</td>
<td>1.03 ± 0.04</td>
<td>1.06 ± 0.05</td>
<td>1.03 ± 0.03</td>
</tr>
<tr>
<td>H2’–OAsp92</td>
<td>1.55 ± 0.11</td>
<td>1.56 ± 0.10</td>
<td>1.56 ± 0.09</td>
<td>1.51 ± 0.13</td>
<td>1.60 ± 0.14</td>
</tr>
<tr>
<td>Zn–OAsp204</td>
<td>2.12 ± 0.06</td>
<td>2.10 ± 0.08</td>
<td>2.11 ± 0.09</td>
<td>2.19 ± 0.14</td>
<td>3.00 ± 0.08</td>
</tr>
<tr>
<td>Zn–OAsp92</td>
<td>2.37 ± 0.13</td>
<td>2.28 ± 0.12</td>
<td>2.23 ± 0.10</td>
<td>2.26 ± 0.12</td>
<td>2.08 ± 0.07</td>
</tr>
<tr>
<td>Zn–NHis90</td>
<td>2.17 ± 0.05</td>
<td>2.15 ± 0.07</td>
<td>2.18 ± 0.07</td>
<td>2.10 ± 0.07</td>
<td>2.12 ± 0.06</td>
</tr>
<tr>
<td>Zn–NHis471</td>
<td>2.13 ± 0.08</td>
<td>2.15 ± 0.07</td>
<td>2.18 ± 0.08</td>
<td>2.13 ± 0.07</td>
<td>2.11 ± 0.06</td>
</tr>
<tr>
<td>Zn–O2’</td>
<td>2.27 ± 0.06</td>
<td>2.35 ± 0.13</td>
<td>2.46 ± 0.16</td>
<td>2.29 ± 0.11</td>
<td>2.24 ± 0.10</td>
</tr>
<tr>
<td>Zn–O3’</td>
<td>2.32 ± 0.07</td>
<td>2.34 ± 0.16</td>
<td>2.37 ± 0.11</td>
<td>2.39 ± 0.11</td>
<td>2.26 ± 0.11</td>
</tr>
</tbody>
</table>

\(^a\) Standard deviation.

Table 7.2. Average Restrained Electrostatic Potential Charges\(^37\) of Relevant Atoms along the Reaction Pathway

<table>
<thead>
<tr>
<th></th>
<th>Reactant (R)</th>
<th>R’</th>
<th>TS</th>
<th>P’</th>
<th>Product (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1’</td>
<td>0.51 ± 0.12(^a)</td>
<td>0.51 ± 0.15</td>
<td>0.61 ± 0.03</td>
<td>0.51 ± 0.11</td>
<td>0.56 ± 0.07</td>
</tr>
<tr>
<td>O5’</td>
<td>−0.65 ± 0.07</td>
<td>−0.58 ± 0.10</td>
<td>−0.41 ± 0.05</td>
<td>−0.32 ± 0.08</td>
<td>−0.57 ± 0.10</td>
</tr>
<tr>
<td>O6</td>
<td>−0.41 ± 0.04</td>
<td>−0.48 ± 0.14</td>
<td>−0.66 ± 0.06</td>
<td>−0.74 ± 0.05</td>
<td>−0.81 ± 0.07</td>
</tr>
<tr>
<td>O2’</td>
<td>−0.55 ± 0.11</td>
<td>−0.59 ± 0.12</td>
<td>−0.52 ± 0.06</td>
<td>−0.48 ± 0.11</td>
<td>−0.52 ± 0.07</td>
</tr>
<tr>
<td>OAsp204</td>
<td>−0.83 ± 0.12</td>
<td>−0.83 ± 0.06</td>
<td>−0.74 ± 0.07</td>
<td>−0.71 ± 0.05</td>
<td>−0.54 ± 0.03</td>
</tr>
<tr>
<td>OAsp341</td>
<td>−0.68 ± 0.08</td>
<td>−0.76 ± 0.04</td>
<td>−0.88 ± 0.05</td>
<td>−0.92 ± 0.03</td>
<td>−0.90 ± 0.03</td>
</tr>
<tr>
<td>Zn</td>
<td>0.83 ± 0.22</td>
<td>0.82 ± 0.20</td>
<td>0.61 ± 0.14</td>
<td>0.51 ± 0.13</td>
<td>0.69 ± 0.15</td>
</tr>
<tr>
<td>δ(^an) b</td>
<td>0.23</td>
<td>0.25</td>
<td>0.37</td>
<td>0.40</td>
<td>0.09</td>
</tr>
</tbody>
</table>

\(^a\) Standard deviation. \(b\) Sum of charges of C1’, C2’, H1’, and O5’.

At this point, both the nucleophile and leaving group are ~2.5 Å from the C1’ atom. The C1’–O5’ distance reaches a minimum (1.29 Å), and the C1’ atom becomes sp2-hybridized.

Interestingly, the TS does not coincide with the point of maximal OCI character (P’) but is closer to the reactants (Figure 7.3). The glycosylation reaction in GMII is consistent with a
DNAN mechanism, where the nucleophilic attack occurs before the scissile glycosidic bond is fully broken.\textsuperscript{36}

To quantify the degree of OCI character in the mannosyl residue along the reaction, a partial charge analysis was performed (Table 7.2). Since the charge increase of the anomeric carbon atom along the reaction pathway delocalizes over its neighboring atoms, specially at O5', we summed up the charges of C1', O5', C2' and H1' (\(\delta^{an}\)) (Table 7.2). As expected from the shape of the FES, the maximal value of \(\delta^{an}\) occurs at P'. As the reaction progresses from R to the TS and further to P', \(\delta^{an}\) increases by 0.17 electrons, mainly due to the change at the ring oxygen atom (O5'). Interestingly, the Zn ion charge shows the opposite tendency (Table 7.2). Thus, the increase in anomeric charge is compensated by a decrease of positive charge of the Zn ion. This clearly indicates that the role of Zn is not just structural but is also catalytic.

Finally, in the product state, a covalent bond between Asp204 and the C1’ atom is formed (1.5 Å). At this point, the leaving group is completely dissociated (C1’…O6 = 3.47 Å), the anomeric carbon atom has recovered the sp3-hybridization, and the C1’–O5’ distance is as short as in the reactant state (1.41 Å). Following the reaction coordinate from reactant to product, the nucleophilic oxygen (O\textsubscript{Asp204}) loses electronic charge as the nucleophile forms a covalent bond with the C1’ atom (Table 7.2). In contrast, the leaving group oxygen atom (O6) becomes more negative along the reaction pathway. The formation of the covalent bond between the glycon and Asp204 lengthens the O\textsubscript{Asp204}…Zn distance from 2.12 Å to 3.00 Å (Table 7.1), which is compensated by a decrease of the distances between Zn and the rest of its coordinating atoms.
The role of zinc and the partial deprotonation of the O2’–H\textsubscript{O2'} bond

As mentioned in the Introduction, the presence of a catalytic Zn ion in the active site of GHs is unprecedented, and its precise role remains unclear. The Zn ion is coordinated with both O2' and O3' glycon atoms of the substrate (Figure 7.1). The hydrogen atom from the C2’–OH hydroxyl group, in turn, forms a hydrogen bond with Asp92. It is well known that interactions between the C2’–OH group and electron-donating groups stabilize the OCI-like TS more than other glycon hydroxyl groups do.\textsuperscript{38} Even in solution, electron-withdrawing groups at the C2’ atom cause larger decreases in reaction rate than similar substituents at other pyranosyl ring positions.\textsuperscript{39} This is because the C2’–OH group is closest to the anomeric center, which becomes electron-deficient at the TS. It was proposed\textsuperscript{20,40} that, for β-retaining GH enzymes where the carboxylate nucleophile is forming a hydrogen bond at the C2’–OH position, the stabilization occurs through a partial deprotonation of the O2’–H\textsubscript{O2'} bond. The rationale behind this proposal was that upon formation of the OCI-like TS, the acidity of the O2’ increased because of the lack of electrons in the pyranosyl ring, loosening the O2’–H\textsubscript{O2'} bond.

For GMII, we directly observed the partial deprotonation of the O2’–H\textsubscript{O2'} bond upon the formation of the mannosyl OCI species. Figure 7.5 shows the time evolution of the O2’–H\textsubscript{O2'} and C1’–O5’ distances throughout the metadynamics simulation. A decrease of the C1’–O5’ distance represents the formation of the mannosyl OCI-like species. The decrease of the C1’–O5’ distance from 15 to 25 ps is paralleled by an increase of the O2’–H\textsubscript{O2'} distance. At ~17 ps of the simulation, the O2’–H\textsubscript{O2'} bond breaks and the H\textsubscript{O2'} proton transfers to Asp92. The partial deprotonation of the O2’ atom provides electrons that can be delocalized to the anomeric center, producing a stabilizing effect.
Since the Zn ion is coordinated to the O2’ atom, it can stabilize the negative charge accumulated on the latter upon lengthening of the O2’–H bond and promote its partial deprotonation, thus relieving the electronic deficiency of the OCI-like species. As discussed in Section 3.2, the charge of the Zn ion varies in concert with the anomeric charge (Table 7.2), indicating that the positive charge of the Zn ion is tuned to provide sufficient stabilization of the O2’ atom, without having excessive positive charge to destabilize the OCI species, which is also positively charged. Therefore, the interaction of the Zn ion and O2' atom is crucial for catalysis in GMII.

**Conformational itinerary of the glycon mannosyl ring along the reaction path**

Crystal structures of GMII complexed with mannosyl substrates (or close analogs) in their covalent intermediate state (PDB 1QX1, 1QWU, and 1QWN) reveal distorted $^{1}S_{5}$
conformations of the mannosyl glycon residue (M5) (Figure 7.6). Based on these structural observations and by analogy with the conformational itinerary suggested for another mannoside-degrading enzyme (GH26 β-mannanase), it was proposed that the glycon residue bound to GMII follows the $^{0}S_{2} \rightarrow B_{2,5} \text{[TS]} \rightarrow ^{1}S_{5}$ itinerary for the glycosylation step.

The precise ring conformation of the Michaelis complex of GMII is not known. A structure of mutated GMII (D341N) complexed with mannose (PDB 3BUP) in its reactant state (Michaelis complex) displays an $E_{5}/O_{5}H_{5}$ conformation. Nevertheless, the authors warned that their assignment was not unambiguous and that different nearby puckering conformers could also be accommodated reasonably well. Interestingly, a mutated nucleophile (D204A) fails to distort the ring and a regular $^{4}C_{1}$ conformation is observed in its complexed structures.

During the metadynamics simulation presented here, we monitored the ring conformation through the glycosylation reaction. Figure 7.6 shows a two-dimensional projection of the Cremer–Pople sphere (also named the Stoddart diagram) viewed from the North Pole that shows many of the possible puckering conformations of a pyranosyl ring. The mannosyl ring conformations visited before the TS are shown in blue. They lie around the $B_{2,5}$ conformation, protruding into the $^{0}S_{2}$ region. The conformations visited after the TS are shown in red, populating the zone between the $B_{2,5}$ and $^{1}S_{5}$ conformations. Experimental points for the Michaelis complex and enzyme–mannosyl covalent intermediate are also plotted onto the Stoddart diagram for comparison. The average conformation found for the TS structures is close to a $B_{2,5}$ (the yellow star in Figure 7.6). Therefore, our simulations confirm that GMII follows a $^{0}S_{2}/B_{2,5} \rightarrow B_{2,5} \text{[TS]} \rightarrow ^{1}S_{5}$ itinerary for the glycosylation step.
Figure 7.6. Conformational itinerary of the glycon mannosyl ring along the metadynamics simulation, mapped onto a Cremer–Pople sphere. The conformations visited before the TS are shown in blue, while those visited after the TS are shown in red. The average TS conformation is shown with a yellow star. The experimentally observed conformation for the Michaelis complex in a mutated (D341N) GMII enzyme is shown with a purple dot (PDB 3BUP). Three conformations corresponding to enzyme–covalent intermediate structures are shown as brown (PDB 1QX1), green (PDB 1QWU), and light blue (PDB 1QWN) dots.

In summary, our simulation provides a detailed description of the glycosylation reaction step during GMII-catalyzed hydrolysis of a mannosyl $\alpha$-1,6 glycosidic bond. The reaction is most consistent with a $D_NA_N$ mechanism, and follows an itinerary similar to $^0S_2/B_{2,5} \rightarrow B_{2,5}$ [TS] $\rightarrow ^1S_5$. The reaction TS has OCI character and displays a conformation similar to a $B_{2,5}$, consistent with previous predictions based on structural observations. In addition, the unusual elongation of the O2'–H02 bond when the substrate acquires OCI character suggests a role
for the active-site Zn in facilitating this process to relieve the electron deficiency of the OCI-like species.

Acknowledgment

The authors thank the Iowa State University High Performance Computation System for making its facilities available to us. They also thank the National Research Initiative of the U.S. Department of Agriculture Cooperative State Research, Education and Extension Service (grant number 2007-35504-18252), the Generalitat de Catalunya (grant 2009SGR-1309), and the Ministerio de Ciencia e Innovación (MICINN) (grant FIS2008-03845) for its financial assistance. A. A. acknowledges an FPU research fellowship from MCINN.

References


Chapter 8: Conclusions and Future Directions

In this dissertation, we contribute to the general understanding of glycoside hydrolases (GHs) by presenting QM/MM simulations of GH-catalyzed hydrolyses including the enzyme and solvent. We present detailed mechanisms of two enzymes involved in plant cell wall deconstruction (i.e. GH8 endoglucanase and GH43 β-xylosidase) and one GH involved in the N-linked glycosylation pathway (i.e. GH38 Golgi α-mannosidase II). In addition, we provide a study of enzyme-imposed scissile glycosidic bond twisting in 14 different GH families to measure the energy of this distortion and analyze its mechanistic implications. Finally, the role of non-covalent interactions to lower the energy barrier was also analyzed.

The GHs studied in this work hydrolyze glycosidic bonds in different carbohydrates. GH8 endoglucanase and GH43 β-xylosidase break glycosidic bonds in cellulose and xylan chains, respectively, both using the inverting mechanism. GH38 Golgi α-mannosidase II breaks very specific mannosyl glycosidic linkages from a complex carbohydrate structure involved in the N-glycosylation pathway using the retaining mechanism and a catalytic zinc ion (very rare in GHs). Despite the difference in function and mechanism of these enzymes, the simulations reveal certain similarities that may apply to most GHs.

For instance, the free energy surfaces obtained for the three GHs have similar shapes (Figures 3.4, 5.1, 7.3). In all cases, protonation of the glycosidic bond is very advanced at the transition state (TS) (represented by the negative values of CV2 at the TS in their free energy surfaces). The reaction mechanism in the three enzymes studied is right in the borderline between concerted (i.e. both the nucleophile and leaving group participate significantly in the TS), and dissociative (i.e. the glycosidic bond breaks first, forming an unstable cation that is
immediately trapped by the nucleophile) mechanisms. In all three cases, the nucleophile participates less in the TS than does the leaving group. For the retaining enzyme studied (GH38), the nucleophile is more involved at the TS, probably because the carboxylate group is a better nucleophile than the water molecules used in the GH8 and GH43 inverting mechanisms.

The topic of ring distortion of substrates in GH active sites has been actively researched. Previous studies have found that many of the substrate distortions observed inside GHs are also conformations with minimal energies in solution. In addition, the majority of these ring distortions found in the Michaelis complexes of substrates bound by GHs are consistent with the anti-periplanar lone pair hypothesis (ALPH) (Chapter 2). An interesting question is whether it is necessary for a GH to be assisted by stereoelectronic effects (from ALPH) to be catalytically competent.

The three enzymes that we chose to study were proposed to follow nonoptimal ALPH conformational itineraries instead of a $^1S_3 \rightarrow ^4H_3 \rightarrow ^4C_1$ pathway for $\beta$-anomeric substrates and the reverse for $\alpha$-anomeric substrates, where a lone pair of the ring oxygen is fully anti-periplanar to the leaving group at the Michaelis complex). In all three enzymes studied, we confirmed itineraries that were not optimal to receive assistance from stereoelectronic effects. In the Michaelis complex of a substrate bound by a GH38 enzyme, which has an $^0S_2/B_{2,5}$ conformation, the ring oxygen lone pair is barely anti-periplanar to the leaving group. Although stereoelectronic effects may offer an energetic advantage, our simulations suggest that GHs can use other itineraries and still be catalytically competent. It is important to note also that evolution does not only select for fast enzymes, but also for other factors such as
selectivity. In enzymes, such as GMII where selectivity is crucial, reaction rates may be partially sacrificed for a more selective reaction.

As part of this work, we also measured the effect of non-covalent interactions to lower the energy barrier in GH8 endoglucanase (Chapter 4). In addition, the effect of strain imposed by GHs on the substrate’s glycosidic bond was measured for enzymes in 14 different families (Chapter 6). For GH8 endoglucanase, the energetic contribution towards lowering the energy barrier provided by non-covalent interactions was much larger than that of imposed strain. For instance, glycosidic bond strain in GH8 endoglucanase contributes only ~2.5 kcal/mol, while non-covalent interactions contribute at least 15 kcal/mol, especially by hydrogen bonds with negatively charged active-site amino acid residues.

The use of computational methods to study the enzymatic reactions in detail provided important insights, especially in research areas that are out of reach of experimental investigations. In GHs, computational methods are very useful in studying enzymes with non-standard characteristics in their active sites, such as a metal ion that plays a role in the reaction (as in GH38 GMII). In the future, these methods can be used to give insight into a variety of problems that are still unresolved, such as in identifying catalytic amino acid residues when the x-ray crystal structures are insufficient (e.g. GH48 cellulase with no apparent catalytic base), or important drug targets where knowledge of their mechanisms can help drug design (e.g. influenza GH34 neuraminidase).
Acknowledgements

I want to start by thanking my major professor, Dr. Peter J. Reilly for allowing me to follow my research interests, and his constant support and guidance throughout my PhD. I also want to thank my committee members, Dr. Mark S. Gordon, Dr. Brent H. Shanks, Dr. Monica H. Lamm, and Dr. Olga A. Zabotina for their helpful advice. I am grateful to my former and current group members, Dr. Chandrika Mulakala, Dr. Anthony Hill, Chris Warner, Dr. Blake Mertz and David Cantú for their valuable discussions and input on my work. I also want to thank my collaborators, Dr. Carme Rovira and Albert Ardevol, for sharing their knowledge and welcoming me to their research group. Finally, Dr. Shinya Fushinobu and Dr. Wim Nerinckx for discussions that were valuable for my graduate education.

I am extremely grateful to my family and friends from home, Rayo Morfin, Eduardo Rodriguez, Santiago Petersen, Luis Petersen and Fernando Carvajal for always being there for me. I also want to thank my friends in Ames, Ramón Morales, Larsinho Kulseng, Sikander Hakim, Carlos Ariza, Kerry Campbell, Nalee Xiong, Janine Cheng, Basak Cinlar, and Yin Yani for many, many good memories.