Memòria presentada per **Albert Ardèvol Grau**

llicenciat en química

per tal d'obtar al títol de Doctor per la Universitat Autònoma de Barcelona.

Programa de doctorat de *Bioquímica, Biologia molecular i Biomedicina*.

Aquesta tesi ha estat realitzada al Parc Científic de Barcelona,

sota la direcció de la **Dra. Carme Rovira Virgili**,

Professora d'Investigació ICREA,

i comptant amb la tutorial del **Dr. Xavier Parés Casasampera**,

Catedràtic d'Universitat

del Departament de Bioquímica i Biologia molecular

de la Universitat Autònoma de Barcelona.

Barcelona, Gener 2012

l'interessat.   vist-i-plau de la directora

Albert Ardèvol Grau   Carme Rovira Virgili

ii

Als meus pares,

Collaborations

The work in this thesis could not be done without the close collaboration with the research groups below. My most sincere gratitude to both of them:

- **Prof. Antoni Planas**
  Department de Bioenginyeria. IQS-Universitat Ramon Llull, Barcelona, Catalunya.

- **Prof. Gideon J. Davies**
  Department of chemistry. York Structural Biology Laboratory, The University of York, York, United Kingdom.

Additional works not presented in this Thesis, were done in collaboration with the following research groups, to whom I am very indebt.

- **Prof. Xavier Parés** and **Prof. Jaume Farrés**
  Department of Biochemistry and Molecular Biology. Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Catalunya.

- **Prof. Peter J. Reilly**
  Department of Chemical and Biological Engineering. Iowa State University, Ames, USA

- **Prof. Mauro Boero**
  IPCMS - Département de Chimie et des Matériaux Inorganiques (DCMI). Centre National de la Recherche Scientifique (CNRS), Strasbourg, France.

Fellowships

This thesis was done with the support of the following fellowships:

- Fellowship of the Spain government (FPU) to perform the PhD in the Barcelona Science Park under the supervision of Prof. Carme Rovira (2009-2011)

- Collaboration fellowship at the Barcelona Science Park (2006 – 2008).

Short stay in foreign research center:

- 02/2008 - 05/2008
  3 months training at the Center for Computational Sciences (Prof. Mauro Boero). University of Tsukuba (Japan).

Symbols and acronyms

Å	Angstrom
a.u.	atomic units
ALPH	antiperiplanar lone-pair hypothesis
AIMD	*ab initio* molecular dynamics
BOMD	Born-Oppenheimer molecular dynamics
CAZY	carbohydrate acting enzymes
CI	covalent intermediate
CP	Car-Parrinello method
CPMD	Car-Parrinello molecular dynamics
CPMD/MM	Car-Parrinello molecular dynamics / molecular mechanics
CV	collective variable
DFT	density functional theory
D-RESP	dynamically generated Restrained electrostatic potential derived charges
e/e$^-$	electron
E·S	enzyme-substrate complex
ESP	electrostatic potential derived charges
FES	free energy surface
fs	femtosecond
G	Gibbs free energy
GH	glycosyl hydrolase
GT	glycosyl transferase
H	Hartree
K	Kelvin
MC	Michaelis complex

MD	molecular dynamics
MM	molecular mechanics
MTD	metadynamics
MUF	4-methyl umbelliferone
nm	nanometer
ns	nanosecond
OtsA	trehalose 6-phosphate synthase
PBE	Perdew-Burke-Ernzerhof exchange-correlation functional
ps	picoseconds
QM	quantum mechanics
QM/MM	quantum mechanics / molecular mechanics
RESP	restrained electrostatic potential derived charges
RI	reaction intermediate
RMSd	root mean square deviation
Ry	Rydberg
TS	transition state
UDP	uridine diphosphate

Contents

Chapter IV - Influence of the enzyme-substrate interactions through the 2-OH on substrate distortion. **63**

Chapter V - Mechanism of reaction in 1,3-1,4-β-endoglucanase with the 2-deoxy-2-fluoro substrate derivative. **77**

Chapter VI - The conformational free energy landscape of β-D-mannopyranose and α-L-fucopyranose.. **91**

Outline

2

Outline

Carbohydrates had historically been associated to two biological functions: energy storage and structural support. However, in the last decades, new complex structures of oligosaccharides have been found to play vital roles in many biological processes, such as signal transduction, immune response, cell differentiation and cancer development, among others.

Advances in the functional understanding of carbohydrate-protein interactions represented a breakthrough in the field of glycobiology and glycochemistry, opening a new branch of potential therapeutic targets (carbohydrate acting enzymes), glycomimetic drugs and biomarkers. The bottleneck in the field of glycochemistry is the synthesis of complex saccharides; hence many efforts have been devoted to the development of novel enzymatic strategies for carbohydrate synthesis.

Glycoside transferases (GT) and glycoside hydrolases (GH) are the enzymes that catalyze the formation and the cleavage of the glycosidic linkage respectively. They are used in complex oligosaccharides synthesis, and recently they have been engineered to produce enzymes with particular substrate specificities or even activities (such as glycosynthases).

In spite of these advances, the understanding of the molecular mechanisms of enzymatic carbohydrate synthesis and degradation is far from complete. Structural studies have shown that the puckering of the sugar ring at the cleavage point must change during catalysis. Knowing the *conformational catalytic itinerary* has an impact in the design of GHs inhibitors. However, these itineraries are not known for all families of GHs. On the other hand, the saccharide puckering is not an issue in GTs, but the reaction mechanism is not known. In fact, the glycosidic bond formation in GTs remains one of the most intriguing and unanswered questions in the field of glycobiology.

The coming of age of powerful theoretical methods such as quantum mechanics / molecular mechanics (QM/MM) and *ab initio* molecular dynamics (AIMD) has enabled the elucidation of complex reactive processes in proteins and enzymes. In particular, the modeling of the Michaelis complex and the reaction mechanisms of GHs highlighted the interplay between electronic and structural (conformational) changes that preactivate the substrate for catalysis. Some of these changes can already be anticipated by analyzing the conformational energy landscape of the substrate.

Part of the research of this Thesis complements previous studies of our group by analyzing the factors that govern substrate distortion in GHs. In this respect, it extends the use of conformational free energy landscapes of simple sugars to predict the conformation of the substrate in Michalis complexes. Additionally, the molecular mechanism of one type of glycoside transferases is elucidated.

This Thesis is organized as follows:

Chapter I contains an introduction of the enzymes studied (GHs and GTs) and presents the main objectives of this work. The theoretical methods used are detailed in *Chapter II*.

Chapters III to *V* are focused on enzyme-substrate interactions affecting the conformation of the substrate in GHs. Concretely; in *Chapter III* we test how mutation of the acid/base catalytic residue, the use of a substrate-like thio-analogue inhibitor or fluorometric aglycons affects the distortion of the substrate. In *Chapter IV* we study the influence of the enzyme-substrate interactions through the 2-OH, in particular the effect of the commonly used 2-deoxy-2-fluoro substitution. The conformational itinerary of this inhibitor during catalysis is modeled in *Chapter V*.

In *Chapter VI*, the conformational flexibility of β-D-mannopyranose and α-L-fucopyranose molecules is investigated. The topologies of their corresponding conformational free energy landscapes are related with the observed crystallographic structures of β-mannosidases and α-fucosidases, and the predictive potential of such calculations is discussed.

Chapter VII focuses on trehalose 6-phosphate synthase (a family 20 retaining GT that belongs to fold type B). The mechanism of glycosidic bond formation in this enzyme is elucidated. Finally, in *Chapter VI*, the main conclusions of this work are summarized.

Chapter I - Introduction

Introduction

Glycoside hydrolases (GHs).

Overview

In Nature, many different carbohydrate monomers, oligomers and polymers play a key role as structural and energy reserve components. On the cell surface, carbohydrates (glycans) occur frequently as glycoconjugates, where they are covalently attached to proteins and lipids (aglycons). In the recent years, it has been shown that these glycoconjugates are involved in a vast array of biological processes, including (but not limited to) cell-cell interaction, cell adhesion, modulation of growth factor receptors immune defense, inflammation or viral and parasitic infections.[2, 3] As a consequence, glycobiology has gained significant attention in life science research and industry.[4]

O-glycoside hydrolases (GHs, EC 3.2.1.x) represent a ubiquitous group of enzymes responsible for the degradation or modification of polysaccharides and glycoconjugates by catalyzing the cleavage of the glycosidic bond.[5] GHs comprise a structurally diverse group, classified in 113 families according to their primary amino acid sequence identity in the Carbohydrate Acting enZymes database (www.cazy.org).[6, 7] Therefore, enzymes that belong to the same family, usually share a similar tertiary structure and mechanism. According to their mechanism of action, glycoside hydrolases can also be classified into retaining and inverting GHs (see the "GHs mechanism" section below).

GHs may also be usefully classified into *endo-* or *exo-*glycoside hydrolases. This classification is made on the basis of whether a GH cleaves an internal glycosidic linkage in a chain (*endo*) or at the end of a chain (*exo*; most commonly the non-reducing end). Whether an enzyme is *endo-* or *exo-* acting has no relationship to the mechanism utilized - thus cellulases, for example, can be *endo-* or *exo-* and retaining or inverting β-glucanases. However, it does have a relation with the topology of the substrate binding site. In this sense, exo-glycosidases usually have a *pocket* or *crater* active site topology (figure I-1 A) while endo-glycosidases are more likely to adopt a *cleft* or a *tunnel* 3D structure (figure I-1 B and C respectively)[5].
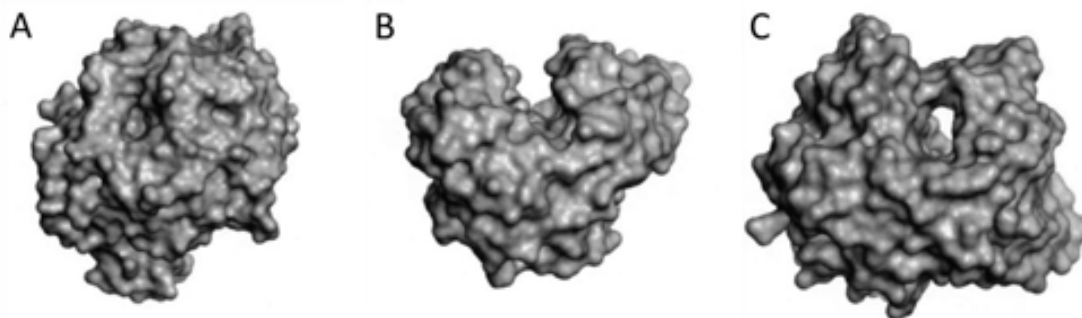

Figure I - 1: Topological structure of the active site of glycoside hydrolases. A) pocket or crater, B) cleft or groove and C) tunnel. Figure taken from reference 5.

Since GHs usually act on big polymeric saccharide chains (such as cellulose, amylose, pectin, …), they have big substrate binding site capable of accommodating more than one sugar units. The substrate binding site of GHs is therefore divided into different *subsites*, and each of them binds a different saccharide unit of the polysaccharide chain. Subsites are then labeled from *-n* to *+n* , with *-n* at the non-reducing end and *+n* the reducing end, and the hydrolysis occurs between the -1 and +1 subsites[8] (figure I-2). Therefore, the sugar unit of the polysaccharide that is occupying the binding subsite -1 is sometimes referred as the "sugar -1" or "subsite -1".
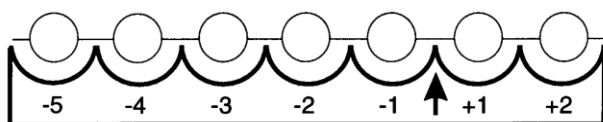

Figure I - 2: Schematic representation of the subsite nomenclature in GHs. The bond that is cleaved is marked with an arrow (between subsites -1 and +1) and each monosaccharide is represented with a circle. The non-reducing end of the polysaccharide chain is labeled with negative numbers and the reducing end with positive numbers. Figure taken from reference 8.

GHs mechanism

Polysaccharides are one of the most stable polymers in nature, with a room temperature uncatalyzed hydrolysis rate constant of $1.9 \cdot 10^{-15}$ s^{-1} (millions of years).[9] However, GHs are capable of accelerating the spontaneous hydrolysis reaction of sugars in water by up to 10^{17} fold,[9] becoming one of the most prominent class of enzymes in terms of catalytic enhancement.

Glycoside hydrolases catalyze the cleavage of the glycosidic bond with either inversion or retention of the configuration on the anomeric carbon, but in both cases it requires two essential catalytic residues: A general acid and a nucleophile (or a general base). These two residues are conserved in most families of GHs.[5, 10-13]

The hydrolysis in inverting GHs is performed in a single S_N2 (bimolecular nucleophilic substitution) displacement (figure I-3 A), in which the nucleophile is a water molecule. The general acid residue catalyzes the departure of the leaving group by transferring its proton to the glycosidic bond oxygen. The general base enhances the nucleophilicity of the attacking water by withdrawing one of its protons.

The reaction in retaining GHs[14] implies two S_N2 displacement reactions and the presence of a covalently linked glycosyl-enzyme intermediate (figure I-3 B). The first reaction (the *glycosylation* step) is very similar to the one of inverting GHs, but the nucleophile is not a water molecule but the Brønsted base enzyme residue (usually a deprotonated carboxylic acid aminoacid) and yield the covalent intermediate (CI) complex and the free positive subsites. The general acid residue acts then as a base to enhance the nucleophilicity of a water molecule that performs a second S_N2 reaction on the anomeric carbon (the *deglycosylation* step), yielding the product of hydrolysis (with net retention of the anomeric configuration) and recovering active form of the enzyme.

Interestingly, the distance between the general acid/base and the nucleophile residue is very conserved (around 5.5 Å and 10.5 Å in retaining and inverting GHs respectively).
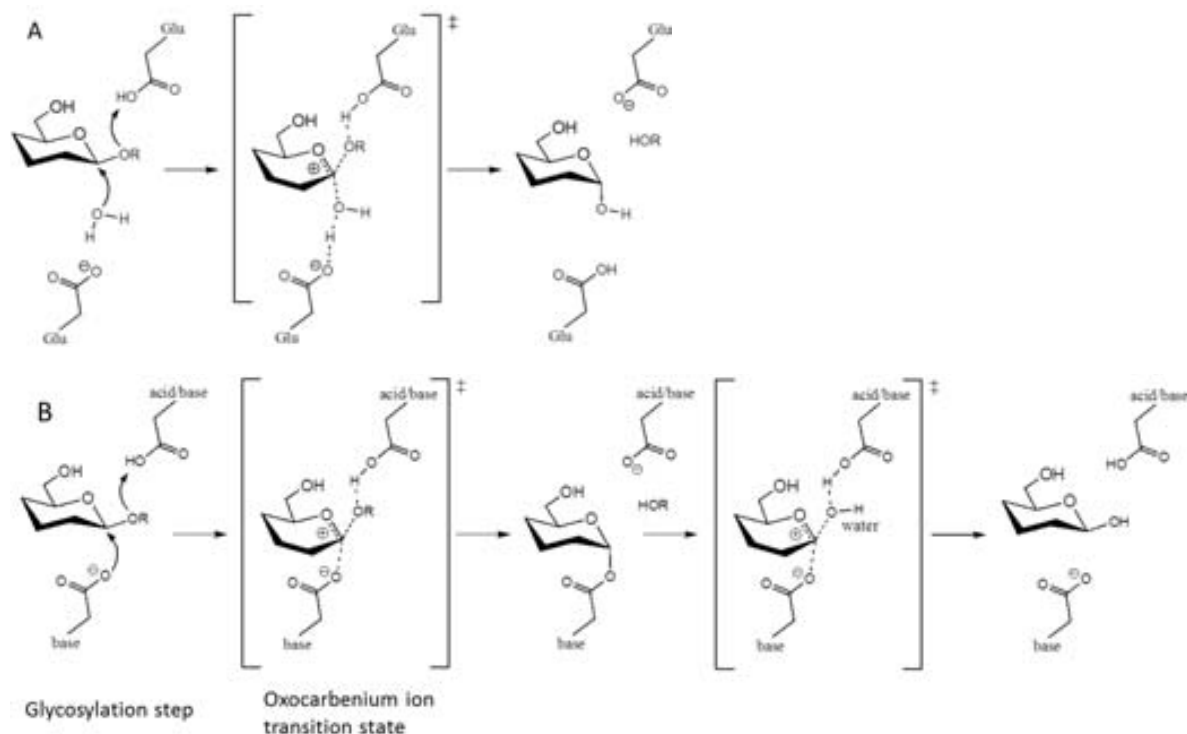
Figure I - 3: Reaction mechanism in A) inverting and B) retaining β-glycoside hydrolases.

The S_N2 mechanism in both retaining and inverting GHs involve a very similar oxocarbenium ion-like transition state (TS, see figure I-3 A and B). KIE experiments have revealed that the structure of the TS is highly dissociative, with low bond orders between the anomeric carbon (C1) and the glycosidic oxygen as well as the C1 and the incoming nucleophile.[12] Consequently, a great development of the positive charge on the C_{an}, which adopts a sp^2 type of hybridization is expected. This change in the stereoelectronics around the C1 has two consequences.

i.   In one hand, the anomeric carbon adopts a planar geometry, and the empty p orbital of C1 can therefore overlap with a lone pair from the pyranic oxygen (O5). This leads to a partial double bond formation between the two atoms (their bond distance shrinks) and a net charge transfer from O5 to C1 that stabilizes the transition state.

ii. On the other hand, the ring places the atoms C5-O5-C1-C2 on-the-plane. This limits the number of possible conformations that the subsite -1 pyranose ring can adopt in the transition state to six, namely: two half-chair $^4H_3$ and $^3H_4$, two envelopes ^4E and ^3E and two boat conformations $^{2,5}B$ and $B_{2,5}$ in the Stoddart[15] diagram (see the side panel *conformations of pyranose rings* and figure I-6). Except for the boats, these structures are not stable conformations, and hence the substrate must follow a certain *conformational itinerary* during catalysis.

Side panel: Conformations of pyranose rings

Pyranose (and six member rings in general) are very flexible molecules that can adopt several conformations. According to the IUPAC nomenclature,[16] there are 26 different conformations that can be classified into chair, boat, half or skew conformations (represented by the letters C, B, H and S respectively). These groups have four atoms on the same plane, and two atoms out-of-plane. Each canonical conformation can be described by the letter of its group classification and two indexes that indicate the atoms that are out-of-plane (subscript or superscript if they are down or up with respect to the plane).

i. The chair group exhibit the two out-of-plane atoms in opposite positions of the ring, one up and one down. There are two *chairs*: $^4C_1$ and $^1C_4$.

ii. The boat group displays the two out-of-plane atoms in opposite positions of the ring, both up or both down. There are six *boats*: $^{3,O}B$; $B_{1,4}$; $^{2,5}B$; $B_{3,O}$; $^{1,4}B$ and $B_{2,5}$.

iii. The half group displays four atoms on-the-plane that are consecutive. There are twelve *halfs*: $^OH_5$; $_OH^1$; $^2H_1$; $^2H_3$; $^4H_3$; $^4H_5$; $^3H_2$; ^1H_O; $^5H_4$; ^5H_O; $^3H_4$ and $^1H_2$.

iv. The skew group displays three atoms on-the-plane that are consecutive. There are six *skews*: $^3S_1$; $^5S_1$; ^2S_O; $^1S_3$; $^1S_5$ and $^OS_2$.

In 1971, J. F. Stoddart introduced a diagram[15] (figure I-4) that included all the possible conformations of a pyranose ring together with all the possible transformations between them. This diagram (namely, the Stoddart diagram) rapidly became very popular for representing the conformational itineraries of glycosides during catalysis.

It is noteworthy that the Stoddart diagram does not intend to give information about the relative energy of each conformation. *E.g.* in a cyclohexane ring, the halfs and the boats are not local minima, but the transition state between the chair and the skews and between skews conformations respectively.

Figure I - 4: Stoddart conformational diagram for a pyranose ring.

Indeed, high values of k_H/k_D from α and β-secondary ^2H kinetic isotope effects[17, 18] (KIE) for the glycosylation and the deglycosylation steps, suggested that the geometry on the anomeric carbon changed from tetrahedral (sp^3-hibridized) to trigonal (sp^2-hibridized) in both steps of the reaction. These results were further contrasted by linear free energy relationship experiments[19-22] (LFER), and protein structures of distorted Michaelis complex (first[23, 24]) and covalent intermediate (later, see *e.g.* reference [25]), gave a full picture of the conformational itinerary of subsite -1 during catalysis. More recently theoretical methods[1, 26-31] have given a minute description of the electronic structure of the reaction species.

X-ray resolved structures from the reaction intermediates of GHs are obtained by either co-crystallizing the wild type (wt) enzyme with an inhibitor, an enzyme mutant with the natural substrate or an enzyme mutant with an inhibitor. Usually, the mutated residues are either the catalytic nucleophile or the general acid/base catalyst, while the inhibitors may range from substrate-like or transition state-like inhibitors to suicide inhibitors such as the so-called Withers fluoro-inactivators (see chapter IV). As an example, the crystal structure of the Michaelis complex and the covalent intermediate of retaining *endo*-cellulase Cel5A[32] showed that the substrate adopted a $^1S_3$ and a $^4C_1$ respectively (see figure I-5). Given that the conformation between them ($^4H_3$) is a putative transition state, it is reasonable to conclude that the conformational itinerary of the substrate in this β-glucosidase is $^1S_3 \rightarrow {}^4H_3 \rightarrow {}^4C_1$.

The main conclusion from the aforementioned studies was that not only the ring at subsite -1 adopts a distorted conformation in the transition state, but it follows a complex conformational itinerary starting at the Michaelis complex and in some cases, even the covalent intermediate is not a $^4C_1$ conformation. The substrate distortion in the MC has been rationalized in terms of the induced fit theory.

Substrate distortion in the Michaelis complex.

GHs are indeed a paradigmatic class of enzyme concerning induce fit theory.[33] During the last decade there has been an increasing evidence that several GHs take advantage of the flexibility of the carbohydrates moieties by distorting the subsite -1 sugar ring upon binding, reaching a so-called pre-activated ES* complex (that constitutes the actual Michaelis complex). The energetic cost of the ring distortion is taken directly from the free energy of binding of the substrate, and it is invested to bring the substrate geometrically and electronically close to the transition state, or in other words, *en route* to the reaction pathway. This results in a preactivated MC with a lower reaction barrier.

Steric preactivation

An important aspect of the catalytic mechanism (S_N2 reaction) is that the nucleophile approaches the anomeric carbon in-line with leaving group departure. However, in β-glycosides with a $^4C_1$ conformation, the H1 atom adopts an axial orientation, directly pointing to the nucleophilic

11

residue, hence preventing the reaction (figure I-5 A). On the contrary, a distorted conformation such as the $^1S_3$ (figure I-5 B) has the H1 atom equatorial and the O1 atom axially oriented. Therefore, the distortion removes the steric hindrance of H1 and also places the glycosidic oxygen in a good orientation for catalysis.
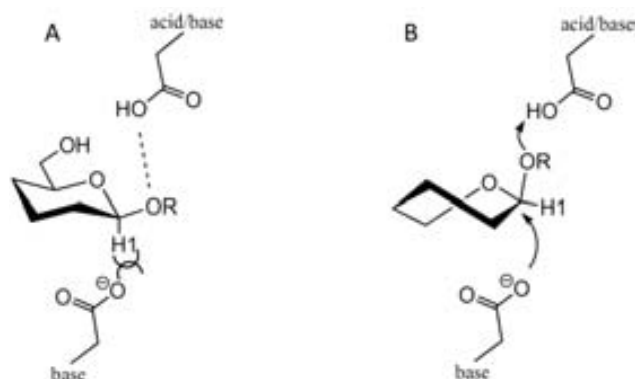

Figure I - 5: A) steric hindrance between H1 and the nucleophile prevents the hydrolysis reaction. B) The distorted conformation (^2S_O in the figure) place the leaving group with an axial orientation and H1 equatorial. This facilitates the in-line attack of the nucleophile and the proton transfer from the acid/base residue.

Structural preactivation

Since the reaction is very dissociative (the glycosidic bond is partially cleaved), the transition state exhibit a high oxocarbenium ion-like character (see figure I-3). The partial charge (δ^+) developed on the anomeric carbon is partially stabilized by a partial double bond formation with the pyranic oxygen. Therefore, a long C1-O1 distance, a short C1-O5 distance and a planar C5-O5-C1-C2 dihedral angle characterize the transition state structure.

Density Functional Theory-based studies[34, 35] have shown that the distorted conformations found in the Michaelis complexes of glucosidases display a longer C1-O1 bond distance and a shorter C1-O5 distance when compared to the $^4C_1$ conformation. Furthermore $^{2,5}B$, ^2S_O, $B_{3,O}$ and, to lesser extent $^1S_3$, $^{1,4}B$, $^1S_5$ and $B_{2,5}$ conformations, show a significant increase on the C1 charge with respect to $^4C_1$.

Stereoelectronic preactivation

A key factor for the reactivity of orthoesters, acetals and related compounds[36] (such as oligosaccharides) is the stereoelectronic effect, derived from the antiperiplanar lone-pair hypothesis[37, 38] (ALPH). The ALPH theory proposes that, highly dissociative S_N2 reactions are favored by those conformations in which the non-bonding lone pair electrons on α oxygen atom/s lay antiperiplanar to the bond with the leaving group (C1-O1). In this configuration, the O5 lone pair sp^3 orbital overlap with the antibonding σ* orbital of the C1-O1 bond is maximized. Consequently, the orbitals of the starting species (the MC) change hibridization into those of the transition state oxocarbenium ion with minimal structural organization. Therefore, in the

preactivated conformations, the C1-O1 bond is more labile and the charge transfer from O5 to C1 is more favorable.

The aglycone occupies an equatorial position in the ground state of a β-glycopyranoside, being antiperiplanar only to ring bonds. Therefore, C1-O1 cleavage is predicted to be stereoelectronically unfavorable. Instead, some distorted conformations (such as a $^1S_3$ or a ^2S_O) in which a lone pair on the ring oxygen is antiperiplanar to the C1-O1 bond, would be more favorable for hydrolysis.

Conformational itineraries

The substrate adopts a distorted conformation in the Michaelis complex to reach a preactivated state (closer to the transition state) that lowers the activation energy of the reaction. Structural information of the Michaelis complex can give insight of the geometry of the TS, and therefore provide valuable information in the understanding of the reaction mechanism and the rational design of TS-like GH inhibitors. In this area, many advances are expected to be achieved, since the best inhibitors found up to now have K_i in the nanomolar range, while giving the high efficiency of GHs in stabilizing the TS, it has been proposed that dissociation constants on the range of 10^{-22} M should be accomplished.[9] However, it has also been pointed out that when enzymatic effects other than TS stabilization play a role on the catalysis, the maximum inhibitor K_i cannot be directly inferred from the catalytic efficiency.[39]

To systematically classify the existing information about the conformational itineraries present in several GHs, Davies *et al.* used the Stoddart diagram to draw the complete transition pathways in a continuous representation[40] (see figure I-6). Interestingly, GHs that belong to the same family or that act on the same substrates, tend to use the same conformational itinerary. As previously said, only a reduced number of conformations are putative transition states for the hydrolysis reaction, which are those conformations that place the C5-O5-C1-C2 atoms on the same plane (namely $^4H_3/^3H_4$, $^{2,5}B/B_{2,5}$ and $^4E/E_3$). Necessarily, the Michaelis complex and the products (in inverting GHs) or the covalent intermediate (in retaining GHs) ought to adopt conformations around one of these "TS conformations". Therefore, the conformational itineraries can be classified according to the transition state conformation,[41] and usually the conformational itinerary implies a considerable movement of the anomeric carbon that is consistent with an electrophilic migration mechanism.

Glucosidases and xylanases:

X-ray structural studies have concluded that some retaining β-glycosidases (GH18 chitinases,[42] GH26 glucanases,[43] GH39 xylosidases[44] and GH84 GlcNAcases[45]) follow a longitudinal $^1S_3 \rightarrow {}^4H_3 \rightarrow {}^4C_1$ conformational itinerary. The $^1S_3$ conformation is preactivated in terms of structure and electronic arrangement, and places the C1-O1 bond in an axial orientation for β-glycosides.[35] However, in α-glycosides, the orientation of the C1-O1 bond is equatorial for $^1S_3$ and axial for $^4C_1$. Therefore, retaining α-glycosides follow a $^4C_1 \rightarrow {}^4H_3 \rightarrow {}^3S_1$ conformational itinerary.[46-48]

Nevertheless, there are some exceptions such as α-glucosidases, which may use a $B_{2,5}$ type of TS as evidenced by kinetic isotope effect experiments.[49] In this case, the substrate would follow a $^{1}S_{5} \rightarrow B_{2,5} \rightarrow {}^{O}S_{2}$ conformational itinerary, as most mannosidases do (see below). On the other hand, it has been proposed that some inverting β-glucosidases[26, 50-52] and retaining xylanases[53, 54] follow a $^{2}S_{O} \rightarrow {}^{2,5}B \rightarrow {}^{1}S_{5}$ conformational itinerary.

Mannosidases:

X-ray structures of the Michaelis complex and the covalent intermediate in retaining GH26[19, 55] and GH2[56] β-mannosidases were found to adopt a $^{1}S_{5}$ and $^{O}S_{2}$ conformations respectively, and transition state-like inhibitors adopted a $B_{2,5}$ conformation. Similarly, experimental[57, 58] and computational[28] evidences give support to a $^{O}S_{2} \rightarrow B_{2,5} \rightarrow {}^{1}S_{5}$ conformational itinerary for retaining α-mannosidases. The $B_{2,5}$ conformation in mannosidases has the C2 hydroxyl group in a pseudo-axial orientation, avoiding the *cis*-1,2 diaxial steric hindrance during the nucleophilic substitution.

The case of inverting α-mannosidases is more controversial, since a $B_{2,5}$ type of TS have been recently observed for GH97,[59] but a crystal structure of family GH47 inverting α-mannosidase was reported to adopt a flipped $^{1}C_{4}$ conformation,[60] indicating a Southern longitudinal itinerary around the $^{3}H_{4}$ TS conformation, as proposed for α-L-fucosidases and sialidases.[61, 62]
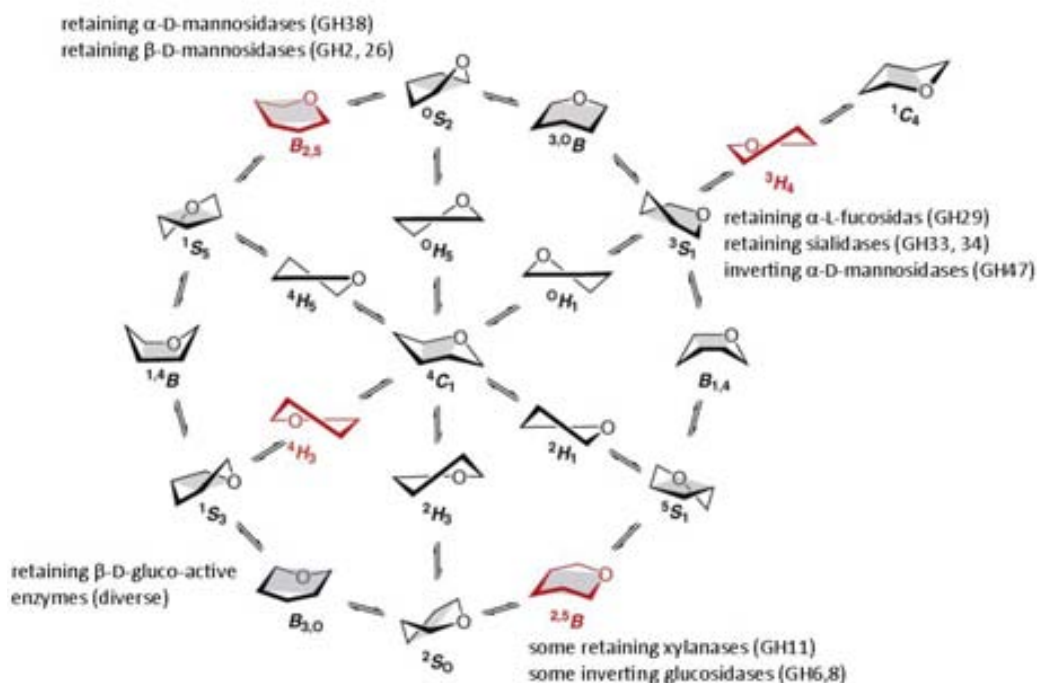

Figure I - 6: Stoddart diagram of pyranose ring conformation connectivity. In red are the conformations of the transition states of the enzymes specified besides (*e.g.* GH38 TS adopts a $B_{2,5}$ conformation). Figure taken from reference 40.

Understanding the catalytic mechanism and the substrate conformational itinerary in a particular GH can improve the rational design of specific and potent inhibitors for this enzyme. In this Thesis, we address the question of substrate flexibility from a theoretical perspective to decipher the reasons governing ring distortion. For this, we will investigate the particular case of three GHs and how the most commonly used experimental techniques affect ring distortion in the MC (chapters III and IV). Then we will study the particular case of two isolated sugar rings (mannopyranose and fucopyranose), and propose a general method to predict the conformational itinerary in GHs from QM calculations (chapter VI).

Glycoside transferases

Overview

Glycoside transferases (GTs) are the enzymes involved in the biosynthesis of oligosaccharides and glycoconjugates (EC. 2.4.x.y), and constitute an ubiquitous group of enzymes found in all living organisms. Concretely, GTs catalyze the formation of glycosidic linkages by the transfer of a saccharide, typically a monosaccharide from an activated donor substrate (usually a nucleotide sugar - *Leloir* GTs, see figure I-7 - or a sugar phosphate - *non-Leloir* GTs) to an acceptor substrate.[63-65] The possible acceptor molecules include other saccharides, lipids, proteins, nucleic acid and even unnatural products, thus GTs exhibit the greatest chemical diversity with respect to substrates and products of any enzyme class.
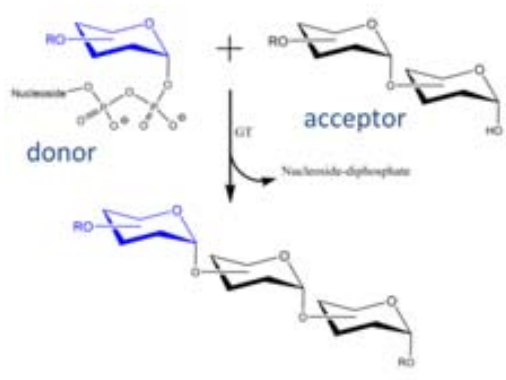

Figure I - 7: Saccharide transfer reaction catalyzed by Leloir GTs. The donor molecule (in blue) is an activated monosaccharide, the acceptor molecule (in black) may be a sugar (as represented), a lipid, a protein, or other biological molecules (see text).

Bacterial glycoside transferases are involved in the synthesis of bacterial cell wall and important saccharides present on their surface such as lipopolysaccharide (LPS) or capsular polysaccharide.[66] Numerous virulence factors of bacterial pathogens have been found to be covalently modified with carbohydrate residues, thereby identifying these factors as true glycoproteins. On the other hand, some bacterial GTs can be used in mass production of complex oligosaccharides of pharmaceutical interest.

In animals, GTs synthetize the sugars that constitute the cell-surface oligosaccharides (glycocalyx). These sugars act as specific binding sites with other cellular receptors, bacteria,[67] virus,[68] parasites[69] and toxins or hormones. GTs are also involved in important physiological functions such as fertilization,[70] cell growth and differentiation, cancer,[71] inflammation[72] processes and immune response.[73]

In plants, glycoside transferases participate in the biosynthesis of complex cell walls[74] as well as small oligosaccharides and glycoconjugates involved in important physiological functions such as hormones, flavonoids and growth factors.[75] As an example, trehalose-6-phosphate (synthetized by

trehalose-6-phosphate synthase) has been found in trace amounts on plants and there is evidence that it acts in the regulation of sugar metabolism[76] and stress protection metabolite.[77]

Glycosyl transferases have been shown to follow a sequential ordered bi-bi catalytic mechanism, whereby non-covalent binding of the donor substrate to the active site is followed by the binding of the acceptor substrate yielding an enzyme-substrates ternary complex.[78] After the chemical step has occurred, the glycosylated acceptor is released before the nucleoside diphosphate (NDP), (scheme I-1).

Scheme I - 1: Sequential ordered bi-bi mechanism

Classification

The cloning and sequencing of more than 500 genomes has now shown that glycoside transferases are a very prevalent enzyme type, representing 1 to 2 % of the genome. Just as for glycoside hydrolases, GTs are classified on the basis of their sequence similitude into more than 90 families in the carbohydrate acting enzyme database (www.cazy.org).[6, 7] Enzymes from the same family share the same folding, but enzymes with different donor/acceptor specificity are found in the same family.

Despite the large number of sequence families that have been defined, structural analysis has shown that GTs possess a limited number of fold types, namely fold type A[79-81] and fold type B[82-84] (figure I-8), although a recent X-ray structure of siayltransferase[85] (CstII) has been found to adopt a novel type of folding for GTs.

The catalytic domain of the GT-A fold enzymes consists in a single domain. The first approximately 120 amino acids show similarity to a *Rossmann* fold, which is found in proteins that bind nucleotides and are responsible for binding the nucleotide sugar donor substrate. The GT-A enzymes, commonly possess a metal binding motif (*DXD* motif) and are Mg^{2+}/Mn^{2+} dependent glycosyltransferases.

The GT-B-fold enzymes possess two distinct domains separated by a cleft that binds the acceptor. The carboxy-terminal domain is primarily responsible for binding the nucleotide sugar donor substrate, but both domains possess elements similar to those of the *Rossmann* fold. Unlike GT-A glycosyltransferases, the GT-B are metal-ion independent and do not possess a *DXD* motif.

The CstII type of folding is similar to GT-A, with a α/β domain similar to the *Rossmann* fold and a smaller α domain that forms a lid-like structure which folds over the active site. However, CstII type of folding does not contain the *DXD* motif.
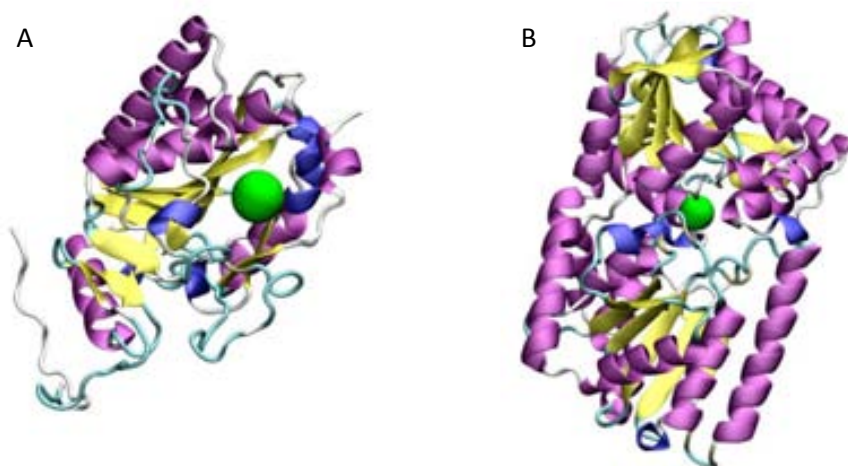

Figure I - 8: Cartoon representation of A) fold type A α-1,3-Galactosyl transferase (PDB code 2RJ7) and B) fold type B trehalose-6-phosphate synthase (PDB code 1GZ5). The active site of the enzymes is represented with a green sphere. Alpha helix are in purple, 3$_{10}$ helix in blue, beta sheets in yellow and loops in purple.

These tertiary structures are in agreement the sequential bi-bi mechanism of GTs. The active site represents a deep pocket, in which the nucleotide sugar substrate binds at the bottom and the acceptor substrate is stacked on top. If the acceptor substrate were to bind first, it would sterically hinder the binding of the donor substrate. Similarly, release of the glycosylated product must precede release of the nucleoside diphosphate.

The structure of GTs is not static but they show a great range of flexibility. In the case of fold type A GTs, X-ray crystal structures of the APO form[86] and the complex with the donor or acceptor substrates[87] shows a substrate-dependent ordering of a very flexible N-terminal loop. Typically, acceptor substrate binding orders the N-terminal loop that adopts a partial alpha-helix secondary structure. This loop folds over the acceptor substrate, and acts as a lid that "closes" the substrates binding site. In the absence of the substrate, the N-terminal loop is disordered and has not been solved by crystallographic studies. In figure I-9 A shows the loop flexibility obtained from a classical molecular dynamics simulation of the retaining GT-A α-(1,3)-galactosyltransferase. The case of fold type B GTs, the "open" to "closed" transition involve an approach of the two big Rossmann folds domains to create an "activated state".[88] In a recent structure of glycogen synthase solved by Ferrer *et al.* (PDB code 3FRO, not published), the "open" to "closed" transition involve an approach of more than 7 Å (figure I-9 B).
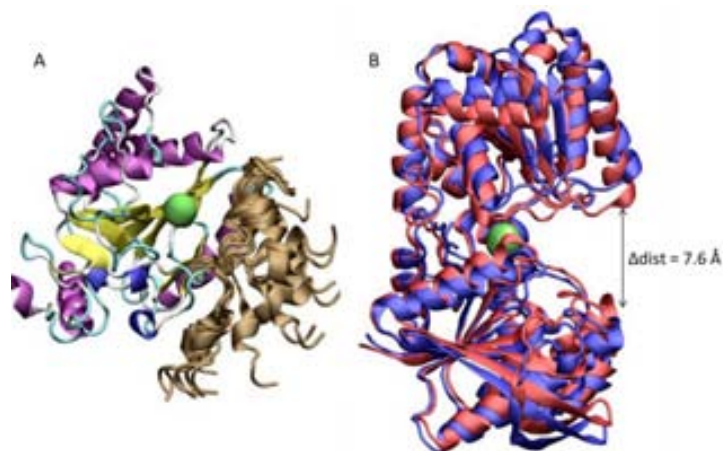
Figure I - 9: A) N-terminal loop flexibility in fold type A GTs. The flexible loop is shown in ochre in several snapshots of a classical MD of α3GalT. B) comparison of the "open" (in blue) and "closed" (in red) structures of glycogen synthase (fold type B GT). The structures correspond to two chains of the structure of the PDB code 3FRO. The active site of the enzymes are shown with green spheres.

Catalytic mechanism

In a similar way as glycoside hydrolases, glycoside transferases catalyze their reactions with either inversion or retention of stereochemistry at the anomeric carbon atom of the donor substrate.

By analogy with inverting GHs, GTs that act with inversion of stereochemistry follows an S_N2 (substitution nucleophilic bimolecular) reaction mechanism where, in contrast to GHs, the nucleophile is one of the acceptor molecule hydroxyl groups and the leaving group is the NDP (figure I-10).[89, 90] Typically, enzymes of this type possess an aspartic acid or glutamic general acid residue whose side chain serves to partially deprotonate the incoming acceptor hydroxyl group, rendering it a better nucleophile.

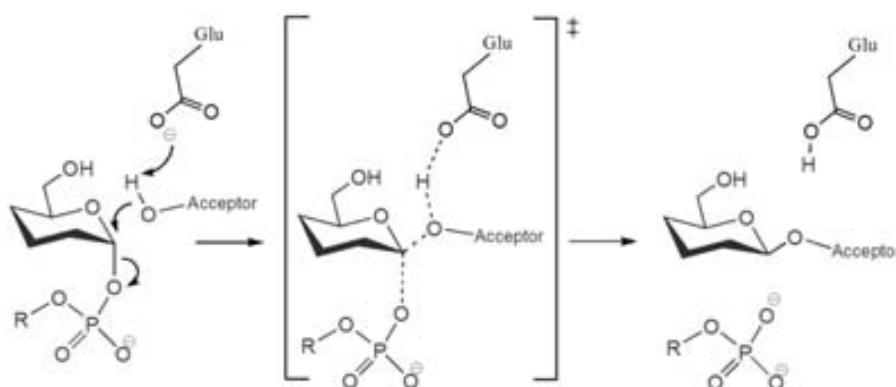**Figure I - 10: Reaction mechanism in inverting glycoside transferases. The reaction proceeds *via* a single S_N2 single displacement mechanism.**

In addition, these enzymes promote catalysis by helping leaving-group departure. In the GT-A enzymes, a metal ion, bound by the *DXD* motif, is typically positioned to interact with the diphosphate moiety. The positively charged metal ion serves to stabilize electrostatically the

additional negative charge that develops on the UDP leaving group during bond cleavage. In the one GT-A enzyme that is not metal ion dependent, positively charged side chains stabilize the leaving group, a strategy also used by some of the GT-B-fold enzymes.

Both kinetic isotope effects for β-1,4-GalTI[91] and α-1,3-FucTV[89], as well as inhibition[92] experiments are consistent with the presence of an oxocarbenium ion-like transition state, hence confirming a dissociative S_N2 or S_N1 type of reaction. Theoretical QM/MM studies on inverting α GT13[93] and β GT7[94] give support to a concerted S_N2 type of mechanism.

Nevertheless, the reaction mechanism of retaining GTs is not yet known. By analogy with retaining glycoside hydrolases, a double displacement mechanism involving a covalent glycosyl-enzyme intermediate was proposed. In the glycosylation step, an aspartic acid or glutamic acid sidechain in the active site makes the first attack and forms the glycosyl-enzyme intermediate. Following, the acceptor molecule performs a second nucleophilic attack on the anomeric carbon to give overall retention of stereochemistry (figure I-11).
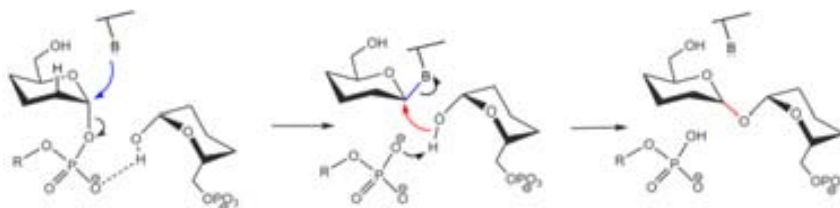

Figure I - 11: Double displacement mechanism proposed for retaining glycoside transferases.

Even though a glycosyl-enzyme intermediate has been trapped and analyzed by X-ray crystallography for a number of retaining GHs, similar attempts to trap the intermediate in retaining GTs have not been successful, with two noticeable exceptions:

i. Family GT8 lipopolysaccharyl-α-1,4-galactosyltransferase C (LgtC) is a retaining GT of fold type A. The only active site side chain suitably positioned to act as the nucleophile in a double displacement mechanism is that of Gln189.[80] Although not generally considered a good candidate as a nucleophile, an amide can indeed play this role, especially in intramolecular reactions such as in retaining hexosaminidases[95, 96] or chitinases[42]. In these cases, the reaction proceeds via an oxazolinium species intermediate. However, the Q189A variant of LgtC possesses relatively high residual activity, with k_{cat} value that is 3% that of the wild type enzyme.[80]

    To investigate the degree of nucleophilic character contributed by Gln-189 during catalysis, Lairson et al.[97] created a Q189E variant of LgtC, thereby replacing the side chain amide with a more nucleophilic carboxylate, and subjected it to kinetic, mass spectrometric (MS), and X-ray crystallographic analysis. Making the transferase more "glycosidase-like" might result in intermediate accumulation or even perhaps a more efficient enzyme. The results provide the first direct observation (by MS) of a glycosyl-enzyme intermediate covalently bound to the active site of a retaining

glycosyltransferase. However, the glycosyl-enzyme covalent bond was not with residue 189 but with aspartate 190, thereby implicating an alternative candidate catalytic nucleophile or a reorganization of the active site topology.

ii. Human ABO blood group GTs (GTA and GTB) are fold type A retaining glycoside transferase (α-1,3-GalNAcT and α-1,3-GalT respectively). Based on an analysis of X-ray crystal structures and the results of molecular modeling,[98] glutamate 303 was proposed to serve as the catalytic nucleophile in a double S_N2 displacement mechanism in human ABO blood group GTs. Furthermore, E303A mutation rendered a 30,000 fold decrease in the activity of GTB, hence stressing the critical role of this residue.[98]

In a recent study by Klassen *et al.*,[99] direct detection of a covalent glycosyl-enzyme intermediate by MS for E303C mutants of both GTA and GTB was reported. The site of glycosylation was found to be residue 303, and exposure of the glycosyl-enzyme intermediate to a disaccharide acceptor resulted in the formation of the corresponding trisaccharide.

Additional support for the double displacement mechanism was reported by Monegal *et al.*[100] in a chemical rescue study on Bovine α-1,3-Galactosyl transferase (α3GalT - a fold type A GT). As in human ABO blood group GTs, α3GalT has a glutamate sidechain (Glu317)[86] that can serve as a nucleophile in a double displacement mechanism. However, X-ray structure of the ternary complex (α3GalT + UDP + lactose)[87] suggested that it only plays a role in binding the acceptor molecule. Mutation to glutamine (E317Q) rendered a discrete 2400-fold reduction in k_{cat}.[101] In the mentioned study by Monegal *et al.*[100] Glu-317 residue was mutated to alanine, rendering an inactive enzyme with k_{cat} values of about 10^{-4} fold lower compared to the wild type enzyme. When sodium azide was added to the E317A mutant, the activity (k_{cat}) raised more than 10^2 fold compared to the residual activity of the mutant.

In conclusion, there are evidences that support a double displacement mechanism in those GTs for which there is a carboxylate that can act as a nucleophile, but further work is required to establish the mechanism[63, 90, 102] for those GTs where there is a glutamine (as in LgtC) or a backbone carbonyl group (as in trehalose-6-phosphate synthase or glycogen synthase).

In the last case (*i.e.* when there is not a putative nucleophile), an alternative mechanism, referred to as S_Ni (internal return nucleophilic substitution) has been proposed. The S_Ni reaction proceeds *via* a front-side single displacement forming an oxocarbenium ion-like transition state (figure I-12 top).[63, 80, 84, 102, 103] While such a mechanism can explain the retention of stereochemistry at the anomeric center without invoking a glycosyl-enzyme intermediate, the formation of this transition state would be sterically and entropically unfavorable.

The internal return mechanism was probed by Tvaroska *et al.*[104] for LgtC by means of *ab initio* calculations on a gas-phase model system of the enzyme. The model included 136 atoms whose positions were taken from the crystal structure of LgtC in complex with both the nucleotide-sugar donor and the oligosaccharide acceptor analogues. The potential energy profile obtained in the

calculation showed preference for a one step mechanism, with the nucleophilic attack of the acceptor oxygen onto the anomeric carbon and the proton transfer to a phosphate oxygen occurring simultaneously. The mechanism was regarded as very concerted ($A_N D_N A_H D_H$), with a unique transition state structure in which the attacking galactose group is more closely bound to the anomeric carbon than to the UDP leaving group and where the hydrogen bond between the nucleophile and the leaving group oxygen facilitates the attack of the acceptor O4' from the same side of the transferred galactose.

Recently, a new mechanism, referred to as stepwise $S_N i$-like or $S_N 1$-like (figure I-12, bottom), involving the formation of a short-lived ion pair intermediate has been suggested.[63, 105] The intermediate undergoes a small shift in its position in the enzyme-active site, allowing the acceptor to attack the donor on the same face that the nucleotide departs from.
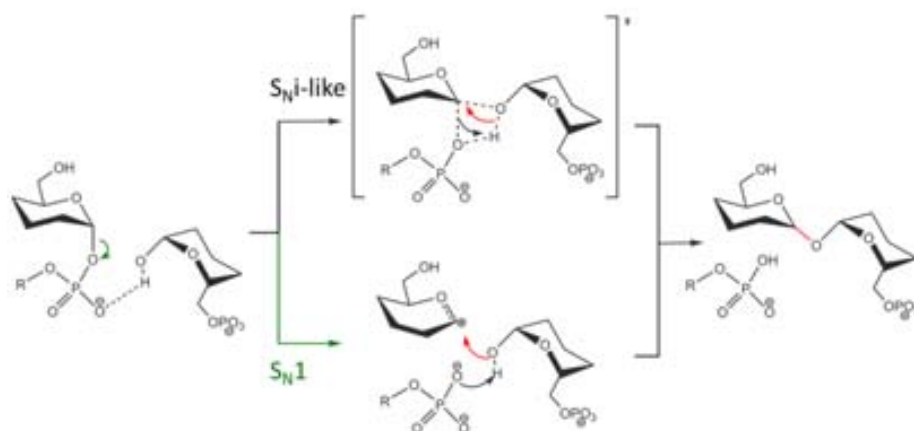

Figure I - 12: Front face reaction mechanism proposed for glycoside transferases.

The recent X-ray structure of trehalose-6-phosphate synthase (OtsA) in complex with a transition state-like inhibitor (validoxylamine-6-phosphate) was solved,[106] supporting the front face nucleophilic attack involving hydrogen bonding between leaving group and nucleophile. In addition to the structural, kinetic isotope effect experiments[107] indicated the presence of a remarkable dissociative transition state with a highly oxocarbenium ion-like character.

In conclusion, there are experimental evidences that support a front-face mechanism for those glycoside transferases in which there is no putative carboxylic acid residue that can serve as a nucleophile in a double displacement mechanism (as is the case of OtsA). However, the details of the mechanism, the relative stability of the transient oxocarbenium ion-like species and the structure of the transition state are not clear. This is of crucial importance in the rational design of potent and specific inhibitors for GTs, as well as for the tuning of the enzymes' activity. Therefore, we decided to address this question in the present Thesis. The results of this investigation are presented in Chapter VII.

Objectives

Objectives

The present Thesis is aimed at unraveling the molecular mechanisms of substrate recognition, preactivation and catalysis in glycoside hydrolases and transferases.

The following objectives have been pursued:

❖ Investigate the effect of substrate modifications and enzyme mutation on the Michaelis complex substrate conformation of glycoside hydrolases.

Previous studies on the group demonstrated that the subsite -1 sugar in 1,3-1,4-β-endoglucanase, adopts a $^1S_3$ conformation. It would be interesting to extend this investigation to enzymes for which there is an X-ray structure available, and study how the experimental strategies used to trap the Michaelis complex affect substrate distortion.

❖ Investigate the interactions between the exocyclic substituent at position 2 and the enzyme nucleophilic residue, and how these interactions affect substrate conformation.

Previous studies in the group suggested that substrate distortion is a global effect that depends on the shape of the binding pocket. One of the most important enzyme-substrate interactions (known to affect the reaction rate of the reaction) is the 2-OH ··· nucleophile hydrogen bond. It is interesting to analyze in detail how this interaction affects substrate distortion.

❖ Determine the conformational itinerary of a 2-deoxy-2-fluoro glucose derivative during the hydrolysis reaction in 1,3-1,4-β-endoglucanase.

The 2-deoxy-2-fluoro substitution prevents the 2-OH ··· nucleophile hydrogen bond, destabilizing the transition state of the reaction. Given the importance of the heterogroup at C2 in the reaction rate, such chemical modification might change the conformational itinerary of the substrate during the reaction mechanism as well.

❖ Obtain the conformational free energy landscape of a β-D-mannopyranose and a α-L-fucopyranose and compare them with the X-ray structures available for β-mannosidases and α-fucosidases.

The conformational free energy surface of β-D-glucopyranose associated to the Stoddart diagram showed a good correlation with the predicted conformational itinerary in β-glucosidases. It would be interesting to extend this analysis to other monosaccharides to investigate catalytic conformational itineraries in GHs.

❖ Unravel the reaction mechanism in retaining glycoside transferases.

Recent experimental evidences give support for a front side single displacement mechanism in a fold type B retaining glycoside transferase (trehalose 6-phosphate synthase). Nonetheless, these results are not conclusive regarding the energetics and the details of the mechanism.

Chapter II - Computational Methods

Computational Methods

Density Functional Theory (DFT)

DFT provides a framework to obtain the total energy of a polyatomic system given their atomic coordinates $\{R_N\}$. According to Hohenberg-Kohn theorems,[108] the ground state energy of a system of interacting electrons subject to an external potential *V(r)* is a unique functional of the electron density, *i.e.* it can be obtained by minimizing the energy functional with respect to the density:

$$E^{DFT} = \min_{\rho(r)}[\rho(r)] \qquad \text{Eq. II - 1}$$

The electronic density $\rho(r)$ can be expressed in terms of single-electron Kohn-Sham (KS) orbitals[109] $\Psi_i(r)$ defined as:

$$\Psi_i(r) = \sum_j c_j \varphi_{ij}(r) \qquad \text{Eq. II - 2}$$

And typically $\varphi_{ij}(r)$ is a Gaussian or a plane-wave function (equations II-3 and 4).

$$\varphi_{ij}(r) = \exp[-\alpha r^2] \qquad \text{Eq. II - 3}$$

$$\varphi_{ij}(r) = \exp[iGr] \qquad \text{Eq. II - 4}$$

In a Car-Parrinello MD, it is computationally convenient to use planewaves (PW) as the basis set in which the KS orbitals are expanded,[110, 111] hence the Kohn-Sham orbitals are expressed as:

$$\Psi_i(r) = \frac{1}{\sqrt{\Omega}} \sum_G^{G_{max}} c_{ij} \exp\left(iG_j(r)\right) \qquad \text{Eq. II - 5}$$

Where Ω is the volume of the cell, *G* is the planewave momentum, c_{ij} are the coefficients of the basis set expansion, and G_j are the reciprocal vectors (periodic boundary conditions are assumed). PW basis sets are denoted by an energy value E_{cut}, which is related to the maximum *G* value of the PW expansion (G_{max}). The number of plane waves N_{PW} can be approximated as:

$$N_{PW} \sim \frac{\Omega}{6\pi^2} E_{cut}^{3/2} \qquad \text{Eq. II - 6}$$

Compared to the Gaussian functions, PWs are not centered at the atoms but extend throughout all the space, and thus they do not suffer from basis set superposition error. However, PWs require the use of pseudopotentials to describe the effect of the core electrons (see the Pseudopotentials section below)

The electron density is calculated with the KS orbitals:

$$\rho(\boldsymbol{r}) = \sum_i^N |\Psi_i(\boldsymbol{r})|^2 \qquad \text{Eq. II - 7}$$

And the DFT energy:

$$E^{DFT} = \min_{\{\Psi_i\}} E^{KS}[\{\Psi_i(\boldsymbol{r})\}, \{\boldsymbol{R}_I\}] \qquad \text{Eq. II - 8}$$

Ab initio molecular dynamics (AIMD)

Assuming that the Born–Oppenheimer approximation holds, *i.e.* the electrons are moving in the field of fixed nuclei, *ab initio* molecular dynamics (AIMD)[110] can be viewed as a series of DFT calculations at different instants of time, each one for a different set of atomic positions $\{\boldsymbol{R}_I\}$. These atomic positions are related by the Newton's equations of motion:

$$M_I \ddot{\boldsymbol{R}}_I = -\frac{\partial E_{el}}{\partial \boldsymbol{R}_I} \qquad \text{Eq. II - 9}$$

Which can be derived from the Lagrangian:

$$\mathcal{L} = E_I^{kin} - E_{el} \qquad \text{Eq. II - 10}$$

Where the kinetic energy of the nuclei E_N^{kin} is:

$$E_I^{kin} = \sum_N \frac{1}{2} M_I \ddot{\boldsymbol{R}}_I \qquad \text{Eq. II - 11}$$

Where M_I and $\{\boldsymbol{R}_I\}$ are nuclear masses and positions respectively. The electronic energy E_{el} is the DFT energy E^{DFT} as given by equation II-8.

In this framework, the basic algorithm of AIMD consists in repeating two main steps (figure II-1):

i.   For a given set of atomic coordinates $\{\boldsymbol{R}_I\}$, the total energy E^{DFT} is calculated by minimizing the density functional.

ii. In the integration step, the Newton's equations of motion are solved.

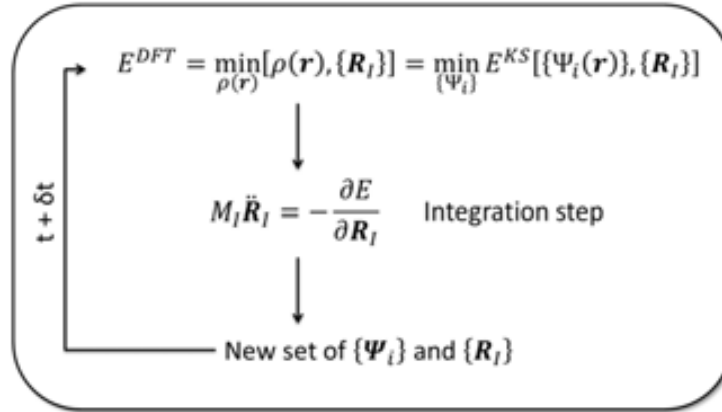iii.   The new ionic positions and velocities are passed to the next iteration step.

Figure II - 1: Schematic diagram of the Born-Oppenheimer molecular dynamics (BOMD) algorithm (adapted from reference [111]).

Car-Parrinello molecular dynamics (CPMD)

The basic difference between BOMD and the Car-Parrinello MD[112] lies in the integration step. In the CPMD, the electronic degrees of freedom are introduced in the Lagrangian as classical particles, and therefore are evolved simultaneously with the nuclei.

$$\mathcal{L} = E_N^{kin} + E_{el}^{kin} - E^{KS} + \sum_{ij} \Lambda_{ij} \left[\int d\mathbf{r} \Psi_j^*(\mathbf{r}) \Psi_j(\mathbf{r}) - \delta_{ij} \right] \qquad \text{Eq. II - 12}$$

Where $E_{el}^{kin} = \sum \mu \int d\mathbf{r} |\Psi_i(\mathbf{r})|^2$ is a fictitious classical kinetic energy term associated with the electronic subsystem $\Psi_i(\mathbf{r})$, μ is a parameter that controls the timescale of the electronic motion, E^{KS} is the electronic potential energy calculated with DFT and Λ_{ij} are Lagrangian multipliers that impose the orthonormality constraints between the orbitals. Therefore, the total energy of the Car-Parrinello Lagrangian is:

$$E_{tot}^{CP} = E_N^{kin} + E_{el}^{kin} + E^{KS} \qquad \text{Eq. II - 13}$$

And the equations of motions are:

$$\mu \ddot{\Psi}_i = \frac{\partial E^{KS}}{\partial \Psi_i^*} + \sum_j \Lambda_{ij} \Psi_j(\mathbf{r}) \qquad \text{Eq. II - 14}$$

$$M_N \ddot{\mathbf{R}}_N = \frac{\partial E^{KS}}{\partial \mathbf{R}_N} \qquad \text{Eq. II - 15}$$

31

The integrations of equations II-14 and 15 provides the time evolution of not only the atomic positions $\{R_I(t)\}$, but also the KS orbitals $\Psi_i(r,t)$. In practice, because the orbitals are expanded in a basis set, what is obtained from the integration is the value of the expansion coefficients at each time instant (the c_j parameters in equation II-2).

Therefore, the basic CPMD procedure (figure II-2) is the same as for AIMD (figure II-1), except that:

i. The electronic energy only needs to be calculated at the beginning of the simulation, because the electrons also evolve in time, following the nuclear motion.

ii. The equations of motion have to be integrated for both nuclei and electrons.

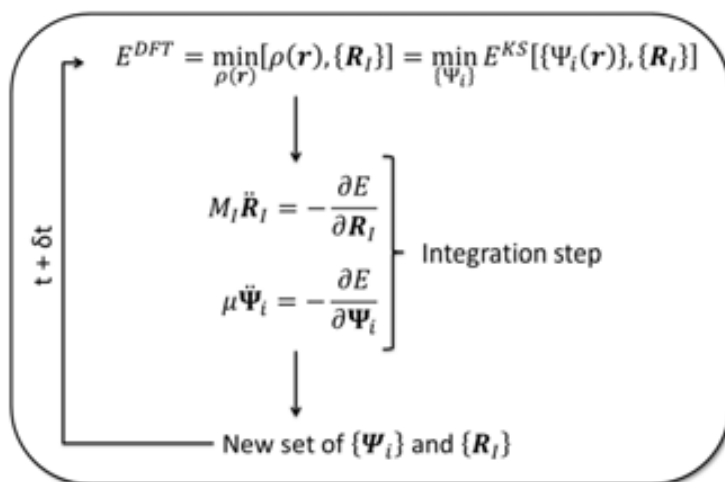iii. In each MD step we obtain a new set of orbital coefficients, besides the new nuclear positions.

Figure II - 2: Schematic diagram of the Car-Parrinello molecular dynamics algorithm (adapted from reference [111]).

The electronic energy obtained at a given instantaneous structure $\{R_I\}$ generally differs slightly from the exact DFT energy. However, if the energy exchange between the electronic and nuclear subsystems is small, the trajectory generated will be identical to the one obtained in a BOMD simulation.[110]

This decoupling of the two subsystems can be achieved by a suitable choice of the fictitious electronic mass μ, such that the frequencies of the electrons are well separated from the nuclear frequencies. In that case, the electronic subsystem remains on the Born-Oppenheimer surface, oscillating around the ground state, and the forces on the atoms are practically the same as those in a BOMD.[110] In particular, the maximum nuclear frequency (v_I^{max}) has to be well separated from the minimum electronic frequency (v_{el}^{min}). Since $v_{el}^{min} = \sqrt{E_{gap}/\mu}$ (where E_{gap} is the HOMO-LUMO separation), small values of μ give a better adiabaticity of the system.

 It can be demonstrated[110] that, provided that the electrons are initially in the ground state and μ is appropriately chosen, they will follow adiabatically the nuclear motion, remaining very close to

the instantaneous ground state. Therefore, in a CP simulation, the electronic energy only needs to be calculated at the beginning of the simulation, and not at each step as in BOMD. By contrast, the simulation time step is shorter (0.1-0.2 fs) compared to BOMD (≈1 fs), in order to describe properly the faster movement of the electrons.

Pseudopotentials

Since planewaves are not centered on nuclei, they are highly inefficient to describe the core electron wave functions. Indeed, they exhibit sharp spatial oscillations in closeness of the atomic nuclei, and therefore a very large set of plane waves is required for an accurate chemical description. On the other hand, the core levels are usually well separated in energy from the valence electrons, and they do not play a role in the chemical properties of the system.

Therefore, the core electron orbitals can be frozen in the KS equations and only the valence electrons are described explicitly. Then, the nucleus and the core electrons are approximated to an effective core potential (ECP or pseudopotential) that interacts with the valence electrons.[110] The pseudopotentials are constructed in such a way that the shape of the atomic valence orbitals outside a certain core radius is reproduced correctly with respect to the corresponding all-electron calculations.

The pseudopotentials used in this thesis are of the norm-conserving type[113] and have been generated using the recipe of Martins and Troullier.[114]

Hybrid methods

Whereas classical force fields allow to treat big systems, *ab initio* methods are needed to study electronic reorganizations, such those that occur in certain protein-ligand interactions and chemical reactions.

The hybrid methods combine the quantum mechanics and the molecular mechanics levels of theory. This enables first-principles calculations of the dynamic properties of the reactive center of a system in a classical force-field environment[115] (*e.g.* the active site of a protein).

In quantum mechanics / molecular mechanics (QM/MM) molecular dynamics, the overall system is divided in two regions: the QM region, (*i.e.* the active site) is treated with AIMD, whereas the rest of the system (*i.e.* the protein and solvent) is described by classical MD. The total energy of the hybrid system can be thus written as:

$$E = E_{QM} + E_{MM} + E_{QM-MM} \qquad \text{Eq. II - 16}$$

Where E_{QM} is the energy for the quantum subsystem, E_{MM} is the energy for the classical subsystem and E_{QM-MM} contains the interactions between the two regions.

In this thesis we have used the QM/MM methodology developed by Laio, Vandevondele and Röthlisberger,[116] which combines Car-Parrinello molecular dynamics with the classical molecular dynamics based on the GROMOS or AMBER force fields (hereafter CPMD/MM). E_{QM} is described by DFT and the potential created by the MM atoms, which polarizes the electronic density of the QM region, is included in the energy functional as an external potential. E_{MM} is computed using the AMBER parameters with the GROMOS force field equation. E_{QM-MM} is calculated as:

$$E_{QM-MM} = E_{QM-MM}^{bond} + E_{QM-MM}^{non-bond} \qquad \textbf{Eq. II - 17}$$

Non-bonding QM-MM interactions

The non-bonding interactions between the QM and the MM regions ($E_{QM-MM}^{non-bond}$) are divided into electrostatic and van der Waals interactions:

$$E_{QM-MM}^{non-bond} = E_{QM-MM}^{electrostatic} + E_{QM-MM}^{van\,der\,Waals} = \sum_{i\,\in\,MM} q_i \int \frac{\rho(r)}{|r-r_i|} dr + \sum_{\substack{i\,\in\,MM \\ j\,\in\,QM}} V_{vdW}(r_{ij}) \qquad \textbf{Eq. II - 18}$$

Where r_i is the position of the MM atom i, with charge q_i, ρ is the total charge of the quantum system, and $V_{vdW}(r_{ij})$ is the van der Waals interaction between atom i and atom j.

The steric non-bonded interactions due to the Pauli repulsion and the dispersion interactions are usually kept into account in a straightforward way by retaining the van der Waals interaction as described by the classical force field. The electrostatic interaction between the QM density and the point charges representing the charge distribution in the MM system often constitute the main environmental effect on the QM system and is technically more subtle.

The $E_{QM-MM}^{electrostatic}$ electrostatic interaction Hamiltonian described in equation II-18 poses serious theoretical and technical problems, related to both its short range and its long range behavior:

i. A first issue is related to the fact that positively charged classical atoms can act as traps for the electron if the basis set is flexible enough to allow for this. In fact, the Pauli repulsion from the electron cloud that would surround the classical atoms is absent, and therefore, the electron density is overpolarized, at short range, by an incorrect purely attractive potential, giving rise to the so-called *electron spill-out problem*. This effect is particularly pronounced in a plane-wave basis-set approach, in which the electrons are fully free to delocalize.

To solve this problem, in CPMD/MM, the point charges of the classical system are replaced by a charge distribution (v)[116] so that the $1/r$ behavior is maintained for large r, whereas for r values shorter than the covalent radius of atom i (r_{ci} in equation II-19) the Coulomb potential goes to a finite value.

$$E_{QM-MM}^{electronic} = \sum_{i \in MM} q_i \int \rho(r)v_i(|r - r_i|)dr \qquad \text{Eq. II - 19}$$

where: $$v_i(r) = \frac{r_{ci}^4 - r^4}{r_{ci}^5 - r^5} \qquad \text{Eq. II - 20}$$

ii.   A second problem is the computational cost. The number of operations that would be required for a direct evaluation of the electrostatic energy is of the order of $N_r\, N_{MM}$, where N_r is the number of real space grid points (of the order of 100^3) and N_{MM} is the number of classical atoms (usually of the order of 10^5 or more, in systems of biochemical relevance). In a real system, a straightforward computation of equation II-18 would therefore increase the computational cost by several orders of magnitude.

A consistent description of the electrostatic interaction between the QM subsystem and the nearest classical (MM) atoms is crucial in order to obtain an accurate prediction of the electronic and structural properties at the QM/MM interface. Instead, long range interactions can be computed as a classical Coulomb potential without introducing relevant errors. Therefore, the electrostatic interactions $E_{QM-MM}^{electrostatic}$ are treated using a multilayer approach (figure II-3).
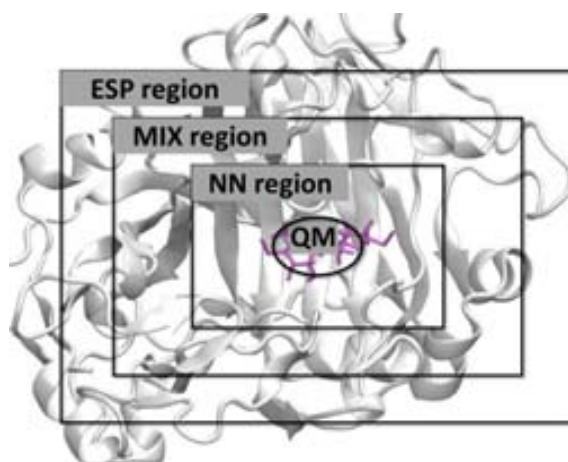

Figure II - 3: Schematic representation of the NN, MIX and ESP regions of electrostatic interaction used in the multilayer approach. NN region = equations II-19 and 20; ESP region = eq. II-21; MIX region = eqs. II-19 and 20 if $q_i \geq$ 0.1 e and eq. II-21 otherwise; outside the ESP region, the QM density is described as a quadrupole expansion.

In this scheme, the Coulombic electrostatic field is included exactly only for a set of MM atoms in the vicinity of the QM system using equations II-19 and 20. This set of atoms (NN region in figure II-3) is defined in such a way as to include all *non-neutral* atoms belonging to charge groups with at least one atom inside a shell of thickness R_{NN} around any quantum atom. Hence, care is taken that charge groups are included (or excluded) as a whole from the direct interaction with the quantum system.

The MM atoms within the ESP region, a classical Coulomb potential is used (equation II-21), using the dynamically generated RESP (D-RESP) charges[117] assigned to the QM atoms:

$$E_{QM-MM}^{electrostatic} \sim \sum_{\substack{i \in MM \\ j \in QM}} \frac{q_i Q_j^{RESP}}{|r_i - r_j|}$$ **Eq. II - 21**

The dynamical RESP charges (D-RESP) are RESP charges calculated "on-the-fly". To obtain the RESP charges, the electrostatic potential created by the electron density $\rho(r)$ is calculated at the position of the classical atoms inside the NN region (in contrast to the surface of a sphere as done in the standard ESP charges calculation). Then, the point charges that best reproduce this potential are calculated. This fitting is performed with a harmonic constraint such that the (RESP) charges do not deviate from a reference value, which is taken as the Hirshfeld charges obtained from the QM electronic density.

For the atoms that belong to the MIX region (figure II-3), equations II-19 and 20 are used if their charge is ≥ 0.1 e, or else the classical potential is used (equation II-21).

The electrostatic field on the MM atoms that do not belong to the ESP region is calculated by a multipolar expansion of the full interaction given in equation II-19.

Bonding QM-MM interactions

The partitioning between the QM and MM regions have to be chosen such that it is as far as possible from the region of interest, e.g. from the atoms involved in the chemical reaction to be studied. When the partitioning between the QM and the MM regions occurs at a covalent bond, the QM region needs to be saturated to have a good description of the electronic structure. In this thesis, we have used the link atom approximation,[118] a pseudopotential constructed with a reduced valence and empirically optimized to reproduce the structural and electronic description of the corresponding atom. In particular, we have used monovalent carbon pseudopotentials (C_{ps}) optimized to reproduce the CH_3-CH_3 bond length. Bonded, angular and dihedral terms involving

Figure II - 4: QM/MM partition through a covalent bond. Black circles are quantum atoms described by ordinary pseudopotentials. The gray circle is a boundary atom described by a monovalent pseudopotential to compensate the valence of the quantum atoms. Empty circles are classical atoms.

MM and QM atoms are treated as in the classical MM Hamiltonian. These terms permit to keep stable the geometry at the interface.

Metadynamics

Consider a system described by a set of coordinates *x* and a potential *V(x)* evolving under the action of a dynamics, whose equilibrium distribution is canonical at a temperature T (figure II-5). The probability to cross a barrier grows exponentially with the height of the barrier as

$\sim \exp[F^*/k_BT]$, hence escaping the minimum is a *rare event* that cannot be observed in standard molecular dynamics simulations. To study these metastable systems, a number of statistical mechanics methods such as adaptive force bias[119], umbrella sampling,[120] conformational flooding[121], blue moon ensemble[122, 123] or metadynamics[124] among others, have been developed.
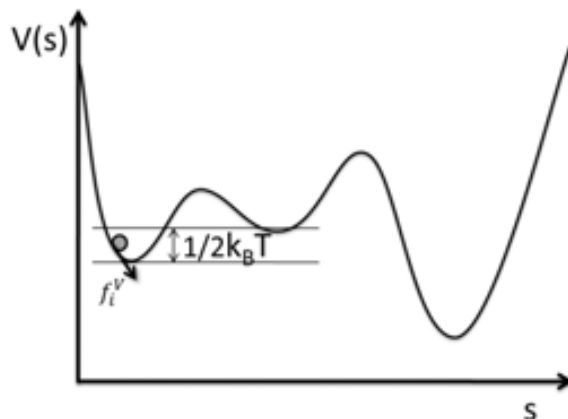

Figure II - 5: Standard MD in a metastable potential *V(s)*. The forces acting on the system (grey circle) are defined in equation II-24

In this thesis, we have used the Metadynamics technique. Metadynamics is a powerful algorithm that aims at enhancing the conformational sampling and at accelerating rare events in systems described by complex Hamiltonians, at the classical or at the quantum level.[125] The algorithm is based in the projection of the trajectories in a (reduced) set of well-chosen collective variables (CVs) which allow to define an effective free energy surface (FES).

Direct metadynamics

Consider that, for our previous system, we can define a set of collective variables *s* which is an explicit function of the coordinates *x* ($s = S(x)$) and that gives a good description of the transition that we want to study (*e.g.* a distance, a torsion, …). The equilibrium behavior of these variables is completely defined by the probability distribution:

$$P(s) = \frac{\exp\big(-(1/T)F(s)\big)}{\int ds \ \exp\big(-(1/T)F(s)\big)}$$

Eq. II - 22

Where *s* is a *d* dimensional vector (s_1, …, s_d) and the free energy $F(s)$ is given by:

$$F(s) = -T \ln\left[\int dx \exp\left(-\frac{1}{T}V(x)\right)\delta\big(s - S(x)\big)\right]$$

Eq. II - 23

In equation II-23 (and in the following) capital S is used for denoting the function of the coordinates S(x), while lower case s is used for denoting the value of the CVs.

The metadynamics simulation starts in one of the minima of the energy surface. At the beginning of the simulation, the force acting on the system is given by the gradient of the potential:

$$f_i^V = -\frac{\partial V}{\partial r_i}$$ **Eq. II - 24**

Every τ_G MD steps, a small Gaussian repulsive potential ($V_G(s)$) is added to $V(s)$. Therefore, the underlying potential is biased by a history dependent potential as $V_T = V + V_G$ where

$$V_G(S(x),t) = \sum_{t'=\tau_G,2\tau_G,\ldots} w \cdot \exp\left[-\frac{\left(S(x) - s(t')\right)^2}{2(\delta s)^2}\right]$$ **Eq. II - 25**

Where $s(t) = S\big(x(t)\big)$ is the value taken by the CV at time t, and the three parameters in V_G are:

i. The Gaussian height w.
ii. The Gaussian width δs.
iii. The frequency τ_G at which the Gaussians are added.

Then, the force acting on the system becomes the sum of two components, one coming from V and the other from V_G. As a consequence, the metadynamics trajectory is not an equilibrium trajectory.

$$f_i^{V_T} = f_i^V + f_i^{V_G} = -\frac{\partial V}{\partial r_i} - \frac{\partial V_G}{\partial r_i}$$ **Eq. II - 26**

In this approach therefore, the biasing potential is history dependent in a similar way to the adaptive force bias[119] or the Wang and Landau algorithm.[126] This manner of biasing evolution by discouraging the visited configurations was first introduced in the taboo search[127] and, in the context of molecular dynamics (MD) simulations, by the local elevation method[128]. In this way, the system is forced to explore conformations that were not previously visited, and thus allows it escaping the minima along low free energy paths and exploring other minima present in the free energy landscape (figure II-6). Obviously, the biasing potential performs a work on the system that needs to be dissipated by using a thermostat.
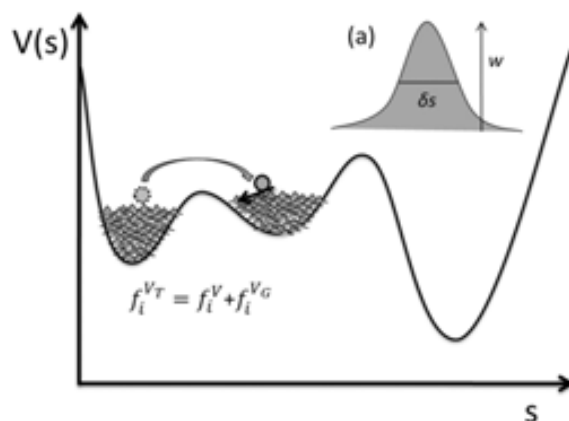
Figure II - 6: In the metadynamics simulation, small repulsive potentials are added on previously visited and pushes the system to escape local minima through the lowest free energy path. The forces acting on the system are defined in equation II-26. (a) repulsive Gaussian potential parameters δs and w.

It has been demonstrated that for long simulation times, the bias will converge to the negative of the free energy and then, since $V_G(s,t) = -\tilde{F}(s,t)$, it will oscillate around that profile (figure II-7 and equation II-27). At this point, the motion of the CVs becomes diffusive in the region of interest and the simulation should be stopped.

$$\lim_{t\to\infty} V_G(s,t) = \sum_{t'=\tau_G, 2\tau_G, \dots} w \cdot e^{-\frac{\left(s(t)-s(t')\right)^2}{2(\delta s)^2}} \approx F(s) \qquad \textbf{Eq. II - 27}$$

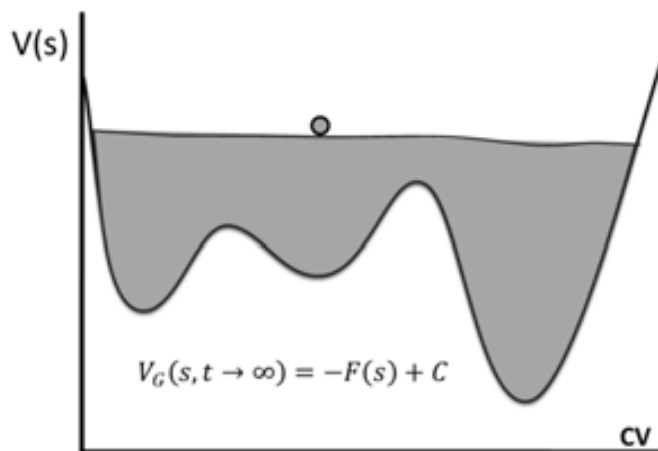**Figure II - 7: At the end of the metadynamics simulation the potential $V_T(s)$ is flat and the system behaves as in diffusive regime.**

However, $\tilde{F}(s,t)$ fluctuates around $F(s)$ with an amplitude that is dependent on both the diffusion coefficient in the CV space and on the repulsive Gaussian potential terms w and δs (the height and the width of the Gaussians respectively). Interestingly, the error in the FES

reconstruction does not depend on *F(s)* but has been proven, both empirically[129] and theoretically,[130] to be:

$$\varepsilon \propto \sqrt{\frac{w}{D\beta}} \qquad \text{Eq. II - 28}$$

where D is the intrinsic system diffusion coefficient in the CVs space. Clearly, a more accurate calculation can be obtained by decreasing the height of the Gaussians, but the convergence of the free energy landscape will then require a much longer simulation time.

Lagrangian metadynamics

If the metadynamics method is used for simulating chemical reactions by *ab initio* molecular dynamics (AIMD), the history-dependent potential has to force the system to cross barriers of several tenths of kcal·mol^{-1} in a short time, usually a few picoseconds. This implies that a lot of energy has to be injected in the degrees of freedom associated with the CVs. This might lead to a significant inhomogeneity in the temperature distribution of the system, and possibly to instabilities in the dynamics.

To address this problem, auxiliary variables $\tilde{s}$ were introduced.[131] These variables are coupled to the system by harmonic restraining potentials of the form $V_H = \frac{1}{2}k\big(\tilde{s} - S(x)\big)^2$. A fictitious kinetic energy of $\frac{1}{2}M\dot{\tilde{s}}^2$ is also assigned to the auxiliary variables. The dynamics of these extra degrees of freedom can be explicitly controlled by suitable thermostats and the trajectory of $\tilde{s}$ can be smoothened so as to control the stability of the algorithm.

The modified potential and forces for the system is:

$$\tilde{V}(x,\tilde{s}) = V(x) + \frac{1}{2}k[\tilde{s} - S(x)]^2 \qquad \text{Eq. II - 29 a}$$

$$f_i^{V_T} = f_i^V + f_i^{V_H} = -\frac{\partial V}{\partial r_i} - \frac{\partial V_H}{\partial r_i} \qquad \text{Eq. II - 29 b}$$

The value of the harmonic potential parameter *k* is assigned by performing an ordinary MD run on the extended system. *k* must be chosen in such a way that the typical value of the difference $\tilde{s} - S(x)$ is smaller than the length on which the free energy varies of approximately *T*.[125] This leads to the condition:

$$\langle\big(\tilde{s} - S(x)\big)^2\rangle \cong \frac{T}{k}(\langle\tilde{s}^2\rangle - \langle\tilde{s}\rangle^2) \qquad \text{Eq. II - 30}$$

where the averages are taken at a temperature T and in the absence of the metadynamics bias.

The value of the mass is a free parameter, that can be tuned in order to obtain a smooth evolution of the $\tilde{s}$. Within a Car–Parrinello scheme an important requirement is the adiabatic separation from the electronic degrees of freedom. Since the extra term in the Hamiltonian introduces frequencies of the order of $\sqrt{k/M}$, M should be relatively large, and the collective variable dynamics ideally should be adiabatically decoupled from the atomic motions.[132] In practice however, large values for M makes the exploration of the free energy surface very slow and computationally demanding in combination with CPMD.

This means that a balance between minimal energy transfer between the electronic, ionic, and collective variable dynamics subsystems on one hand, and a workable efficiency in free energy surface exploration on the other must be reached.

In this framework, the history-dependent potential V_G , instead of acting on the real system as in direct metadynamics, acts on the space of the auxiliary collective variable $\tilde{s}$ (figure II-8). Therefore, the dynamics in the space of the auxiliary CVs is driven by the forces coming from the harmonic potential V_H and the history-dependent potential V_G as:

$$f_{\tilde{s}}^{V_T} = f_{\tilde{s}}^{V_H} + f_{\tilde{s}}^{V_G} = -\frac{\partial V_H}{\partial \tilde{s}} - \frac{\partial V_G}{\partial \tilde{s}} \qquad \textbf{Eq. II - 31}$$
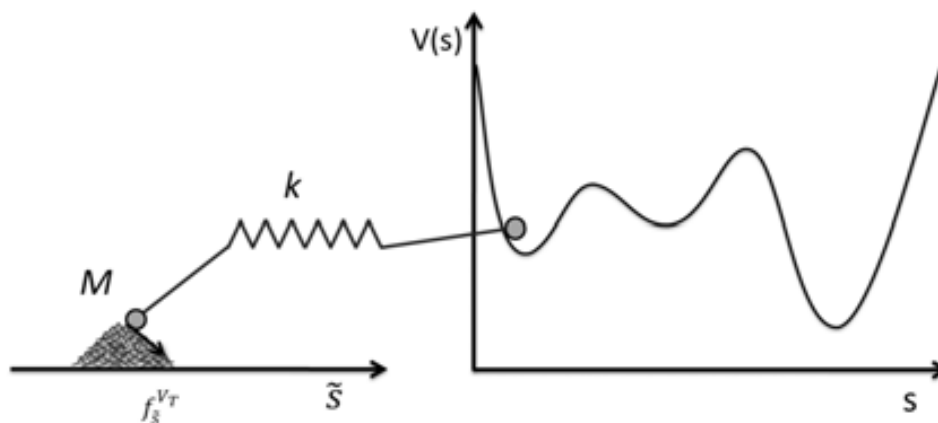
Figure II - 8: Schematic representation of the Lagrangian metadynamics coupled to Car-Parrinello MD. The dynamics of the real system (right side of the spring) in the *s* space is given by equation II-29b. The fictitious system (left side of the spring) is ruled by equation II-31.

Then the free energy as a function of the $\tilde{s}$ variables is given by:

$$G(s) = \lim_{t \to \infty} V_G(\tilde{s}, t) \cong \sum_{t'=\tau_G, 2\tau_G, \dots} w \cdot exp\left[-\frac{\left(\tilde{s}(t) - \tilde{s}(t')\right)^2}{2(\delta s)^2}\right] \qquad \textbf{Eq. II - 32}$$

The main advantages of metadynamics are:[133]

    i.   It accelerates the sampling of rare events by pushing the system away from local free energy minima.

    ii.   It allows exploring new reaction pathways as the system tends to escape the minima passing through the lowest free-energy saddle point.

    iii.   No *a priori* knowledge of the landscape is required. At variance with umbrella sampling, metadynamics inherently explores the low free-energy regions first.

On the other hand, it has two major drawbacks:

    i.   In a single run, V_G does not converge to a definite value of free energy, but oscillates around it, and therefore:

        a.   The bias potential overfills the underlying FES and pushes the system toward high energy regions of the CVs space.

        b.   It is not trivial to decide when to stop a simulation. However, in this respect, one can say, as a general rule, that, if metadynamics is used to find the closest saddle point, *it should be stopped as soon as the system exits from the minimum*. Otherwise, if one is interested in reconstructing an FES, *it should be stopped when the motion of the CVs becomes diffusive in the region of interest*.

    ii.   Identifying a set of CVs appropriate for describing complex processes is far from trivial.

Collective variables

A collective variable (CV) is a function of the microscopic coordinates of the system. To guarantee an effective application of metadynamics, the CVs must respect the following guidelines:[124, 125, 131, 133]

    i.   They must be a function of the microscopic coordinates of the system and the function must have a continuous derivative.

    ii.   They should distinguish between the initial and final state and describe all the relevant intermediates:
When we project the potential energy function on a FES, we operate a dimensional reduction. This transformation is not appropriate for studying a particular process if the CVs cannot discriminate between the configurations of the reactants, products, and relevant intermediates.

    iii.   They should include all the slow modes of the system:
We define as 'slow' those variables that cannot be satisfactorily sampled in the timescale

of the simulation. We also expect that the other 'fast' variables adjust rapidly to the evolution of the slow variables. If any of the latter is not added to the CVs list, the bias potential may not converge to the FES in a reasonable simulation time.

iv. They should be limited in number.

The collective variables used in this thesis (including distances, coordination number, puckering coordinates, …) are described in the chapter in which they are used.

Chapter III - The distortion on the Michaelis complex in β-glycoside hydrolases.

# The distortion on the Michaelis complex in β-glycoside hydrolases.

Introduction.

As mentioned in Chapter I, there are a number of experimental (structural) proofs of a distorted substrate in Michaelis complexes of GHs. These experiments are typically performed in inactive enzyme mutants; substrate-analog inhibitors; or unnatural leaving groups. In this Chapter we have selected one enzyme-substrate complex being representative of each strategy. We have investigated to which extent the changes on the enzyme/substrate might affect the distortion of the substrate.

The specific systems investigated are (1) an inactive mutant of family GH8 inverting endocellulase CelA, (2) a thio-saccharide substrate analog in GH7 retaining endoglucanase I (EGI), (3) a family GH16 retaining endoglucanase in complex with a substrate with a fluorophore aglycon (4-methyl umbellipherone).

Inverting family GH8 endoglucanase CelA from Clostridium thermocellum*.*

Endoglucanase CelA from *Clostridium thermocellum* (EC 3.2.1.4) is a family 8 cellulase that acts with inversion of the anomeric carbon configuration. This enzyme is part of cellulosome enzymatic complex and it is therefore an extracellular enzyme. From the structural point of view, the enzyme features a barrel $(\alpha/\alpha)_6$ type of folding, which is very common for inverting glycosidases such as GH9 endoglucanases, GH15 glucoamylases or GH48 cellobiohydrolases, but it has not been observed for retaining GHs. As explained in Chapter I, the reaction mechanism of inverting glycosidases consists in a single displacement nucleophilic substitution (S_N2), in which the nucleophile is a water molecule. The hydrolysis is assisted by two protein residues: A general acid catalyst that gives a proton to the glycosyl leaving group and a general base that withdraws a proton from the attacking water. By doing so, it enhances its nucleophilicity, as shown in scheme III-1.
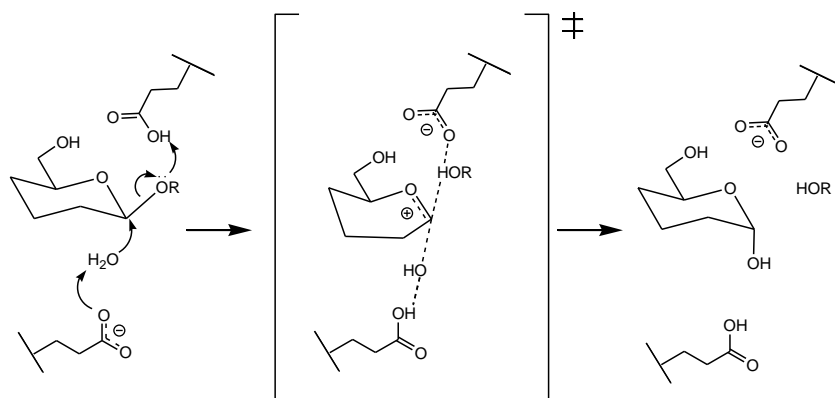
Scheme III - 1: General mechanism of inverting GHs.

In the case of the enzyme studied here, the general acid catalytic residue was identified as glutamate 95[134] and the general based residue as aspartate 278. They are located at 6.5 Å distance according to the 0.94 Å resolution X-ray structure (PDB code 1KWF) solved by **Guerin** and coworkers.[50] This structure was obtained by cocrystallizing the inactive E95Q mutant with cellopentaose as the substrate. It was found that the substrate binding site expands 6 subsites (from -3 to +3), which were identified from a bound cellopentaose molecule occupying subsites -3 to +2 and the hydrolysis product occupying subsites +1 to +3 (with similar occupancies).
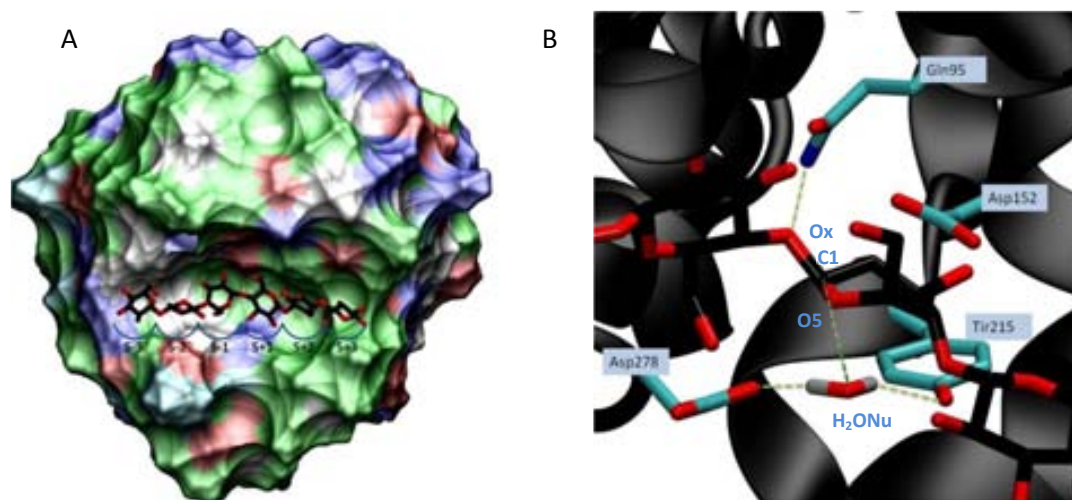

Figure III - 1: A) The substrate binding site in GH8 CelA is of "cleft" type, and can accommodate subsites -3 to +3. B) Active site of the E95Q mutant. The hydrogen bonds with the catalytic residues and the distance between the nucleophilic water and the anomeric carbon are marked in khaki green.

The structure evidences a distorted $^{2,5}B$ conformation of the -1 saccharide, while the rest of glucose rings exhibit the usual $^{4}C_1$ chair conformation. The enzyme-substrate complex is stabilized by hydrogen bonds between the hydroxyl exocyclic groups of the substrate and polar residues of the protein, as well as van der Waals interactions between the saccharide rings and enzyme's aromatic residues.

The distorted conformation in particular, is stabilized by two strong hydrogen bonds between the 2-OH and 3-OH groups and residue Asp152 (see figure III-2 B), a residue known to be essential for enzyme's activity.[135]

In this work, a cellohexaose substrate occupying all 6 subsites was modeled by combining the two saccharide molecules observed in the crystal structure (see figure III-2 A), and the influence of the E95Q on substrate distortion was investigated.

There are several features that make this system interesting for our purposes. In the first place, given the high resolution of the crystallographic structure (0.94 Å), it is a perfect system to test the reliability of the computational methodology used. In the second place, it is an example of a Michaelis complex with the natural substrate and only the enzyme (the general acid/base catalytic residue) has been modified (E95Q mutation). Whether the modification affects the distortion at subsite -1 can be tested by reversing the glutamine mutation back to glutamic acid in the

computations. The third reason for choosing this system is that, in contrast with the other two systems studied, it is an inverting glucosidase.

1,4-β-endoglucanase from Fusarium oxysporum*.*

The strategy of using a substrate thio-derivative has been widely used in the study of substrate distortions in MC complexes of GHs.[136, 137] The oxygen-to-sulfur modification lowers the proton affinity of the substrate, thereby slowing the hydrolysis reaction. Provided that a bad leaving group is used, the enzyme cannot hydrolyze the glycosidic bond with an efficient rate. However, the ability of thio-glycosides to mimic the glycoside-enzyme interactions has been called into question, and it has been claimed that the greater C-S bond length (compared to the C-O bond) might provoke differences in substrate binding and protein-substrate interactions. These structural considerations, together with the protonation state of the general acid/base catalyst and the alkyl sulfide pair (i.e. whether they exists as a glutamic acid and thio-glycoside neutral state or as a glutamate and sulfonium ion pair), have been addressed in a previous (QM) theoretical work by **Smith**[138] using gas-phase and protein active site models. It was concluded that the neutral species was more stable but that the protein environment (including the closest residues of the active site) has a substantial effect on the relative stability of the two species.

1,4-β-endoglucanase I from *Fusarium oxysporum* (EGI) belong to family 7 GH and catalyzes the hydrolysis reaction of 1,4-β linkages in cellulose with retention of the anomeric carbon configuration.[139] The reaction mechanism is the typical one for retaining GHs: a double S_N2 type displacement mechanism with residues glutamic acid 202 and glutamate 197 acting as general acid/base catalyst and nucleophile, respectively (see the introduction Chapter for a detailed explanation of the mechanism). Kinetic studies demonstrated that the EGI substrate binding site expands to four subsites (from -2 to +2). In the presence of short and soluble cellodextrin chains, EGI produces preferably cellobiose as leaving group, indicating the importance of subsites +1 and +2 for catalysis.[139] As observed for other GHs, at high concentration of polysaccharide, EGI has a remarkable transglicosidation activity that has been exploited for chemo-enzymatic synthesis of oligosaccharides.[140]

The tertiary structure of EGI is of jellyroll type, with six concave antiparallel β-sheet over six convex ones, stabilized by nine disulfide bonds. Residues asparagine 56 and 247 are glycosylated by *N*-acetylglucosamine, while the N-terminal glutamine is modified to pyroglutamic acid.

In 1997, **Sulzenbacher** and col. trapped the Michaelis complex of EGI by using a non- hydrolizable thio-derivative substrate: methyl-(*S*)-β-D-glucopyranosyl-(1→4)-(*S*)-4-*thio*-β-D-glucopyranosyl-(1→4)-(*S*)-4-*thio*-β-D-glucopyranosyl-(1→4)-(*S*)-4-*thio*-β-D-glucopyranosyl-(1→4)-4-*thio*-α-D-glucopyranose (or thio-DP5) published in the protein data bank with the code 1OVW.[141] Although the substrate used in the X-ray diffraction experiment was a pentasaccharide derivative, the electron density on subsites -3, -2 and +2 was not clear enough to be successfully solved in the diffraction pattern (the x-ray structure's resolution is 2.70 Å). Substrate atoms have a relatively high B factors (average of

42). As a result, only subsites -1 and +1 could be observed. Subsite -1 was assigned to adopt a $^{2,5}B$ type of conformation, positioning the glycosidic bond with subsite +1 in an axial orientation. The saccharide rings at subsites -3, -2 and +2 were manually added to model the MC of this enzyme.
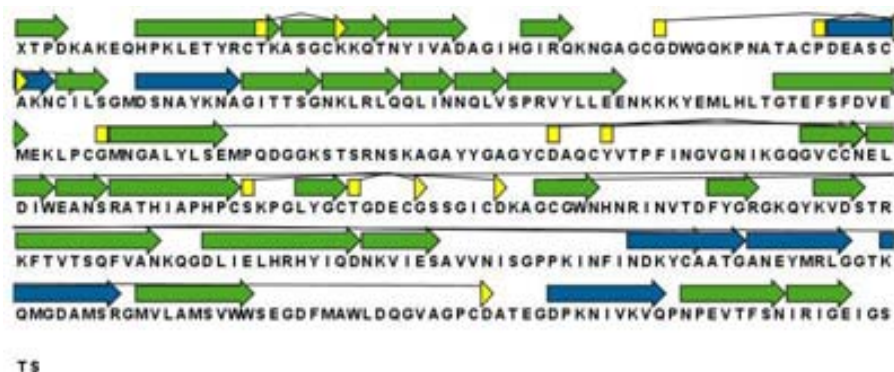

XTPDKAKEQHPKLETYRCTKASGCKKQTNYIVADAGIHGIRQKNGAGCGDWGQKPNATACPDEASC

AKNCILSGMDSNAYKNAGITTSGNKLRLQQLINNQLVSPRVYLLEENKKKYEMLHLTGTEFSFDVE

MEKLPCGMNGALYLSEMPQDGGKSTSRNSKAGAYYGAYCDAQCYVTPFINGVGNIKGQGVCCNEL

DIWEANSRATHIAPHPCSKPGLYGCTGDECGSSGICDKAGCGWNHNRINVTDFYGRGKQYKVDSTR

KFTVTSQFVANKQGDLIELHRHYIQDNKVIESAVVNISGPPKINFINDKYCAATGANEYMRLGGTK

QMGDAMSRGMVLAMSVWWSEGDFMAWLDQGVAGPCDATEGDPKNIVKVQPNPEVTFSNIRIGEIGS

TS

Figure III - 2: Sequence and secondary structure of EGI

Retaining 1,3-1,4-β-endoglucanase from Bacillus.

1,3-1,4-β-endoglucanase from *Bacillus* is a highly active retaining family GH16 endoglycosidase.[142] As other retaining GHs, the catalytic machinery is composed by two acidic residues, specifically glutamate 105 (the nucleophile) and glutamate 109 (the general acid/base residue). There is also an aspartic acid that helps in modulating their pKa[143], but it is not essential. The enzyme acts on linear β-glucans with an strict cleavage specificity for β-1,4 glycosidic bonds on 3-O-substituted glucosyl residues. Even if the Michaelis complex of 1,3-1,4-β-endoglucanase has not been yet characterized, previous studies in our group performed with the same hybrid QM/MM CPMD methodology used in the present thesis, concludd that the substrate sugar ring located at the -1 subsite adopts a distorted $^1S_3/^{1,4}B$ conformation.[34] Biarnés *et al.* found out that the main determinants of the substrate distortion have a structural origin. In particular, the enzyme's binding cavity shape forces the leaving group (subsite +1) to adopt an axial orientation that favores the -1 subsite ring distortion. The complex studied was the wild type enzyme with a 4-methylumbellyferyl (MUF) tetrasaccharide. *Id est*, the tetrasaccharide occupying the subsites -4 to -1 of the binding site and the MUF aglycon occupying subsites +1 and +2.

The MUF tetrasaccharide is a good substrate extensively used in enzyme kinetics for its solubility in water and the possibility of monitoring the 4-methylumbellyferone appearance by fluorometric assays. However, given the importance of the substrate interactions with the enzyme residues at subsites +1 and +2, we decided to model the complex of 1,3-1,4-β-endoglucanase with its natural substrate (i.e. a hexasaccharide molecule). Our aim was to check the possible implications of this modification (i.e. the change of the MUF aglycon by two glucose moieties) on the -1 subsite ring conformation. The initial structure of the Michaelis complex with the natural substrate was

obtained by manually removing the MUF aglycon from the previous structure and filling subsites +1 and +2 with two glucose molecules.

Computational details.

Classical mechanics calculations.

The initial structures for each of the systems were built from their corresponding X-ray structure. The protonation state for the acid and basic residues was chosen by analysis of their local environment, and a physiological pH (6.5) was considered in all cases. All simulations were performed in periodic boundary conditions (PBC) and therefore the net charge of the total system was neutralized with counterions (chloride or sodium) were necessary. All systems were enclosed in a 10 Å cubic cell of TIP3P water molecules. The Cornell *et al.* forcefield[144] was used for the protein and the solvent molecules while the Glycam04 forcefield[145-147] was used for the saccharide residues. All MM calculations were performed using the AMBER7.0 package.[148]

All systems were submitted to geometry optimization, using a combination of steepest descent and conjugate gradient algorithms, in two stages. Firstly, solvent molecules were optimized while keeping the protein and the substrate coordinates restrained. Then the complexes were further optimized without any restraint. The desired simulation temperature of 300 K was reached by coupling the system to a thermostatic bath in 300 ps of MD simulation at constant volume and 100 ps more at constant pressure, using a time step of 1fs.

The classical molecular dynamics simulations were extended up to ≥ 1.5 ns at constant pressure and temperature conditions. The root mean square deviation (RMSD) for the protein backbone atoms, the substrate and the subsite -1 glucopyranose ring puckering coordinates were monitored.

Figure III - 3: Atomic partial charges for the pyroglutamic acid residue.

In the particular case of 1,4-β-endoglucanase from *Fusarium oxysporum*, two non-standard residues had to be parameterized: an N-terminal pyroglutamic acid and a 4-thioglucose residue. Pyroglutamic acid was parameterized using the general amber force-field. Its RESP charges[149] were computed with the GAUSSIAN program[150] at the HF/6-311G** level of theory (see charges in figure I-3). Bonding and angle parameters, as well as atomic partial charges for the 4-thioglucose residue were obtained from a GAUSSIAN geometry optimization and frequencies calculation at the HF/6-31+G* basis set and scaled[151] to the Glycam04j force-field. Torsional parameters were adopted from the same force-field. Both parameters and the partial charges are shown in table III-1 and figure III-4.
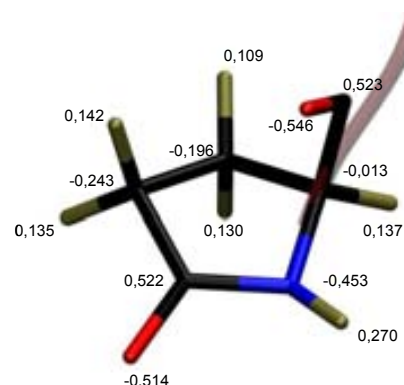
Table III - 1: bond and angle parameters for the thio-glycosidic linkage.

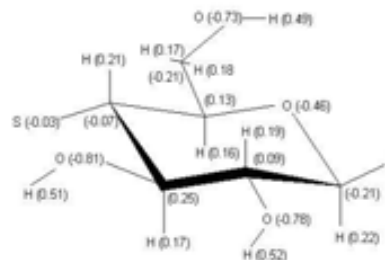| Parameter	equilibrium value	force constant
C-S bond	1,821Å	294,8 kcalmol^{-1}Å^{-1}
H-C-S angle	108,77°	40,5 kcalmol^{-1}rad^{-1}
C-S-C angle	100,94°	79,0 kcalmol^{-1}rad^{-1}
O-C-S angle	108,73°	37,5 kcalmol^{-1}rad^{-1}

Figure III - 4: atomic partial charges for the 4-thioglucosyl residue.

In the case of retaining 1,3-1,4-β-endoglucanase from *Bacillus*, the hexasaccharide (natural) substrate was treated using the Glycam2000[145] parameters, as used in the previous study by Biarnes *et al.*[34] for the complex with the MUF-tetrasaccharide substrate. As in the previous study, the distorted conformation on the -1 saccharide ($^1S_3$) was only stable when the partial charge on the anomeric carbon was increased from 0.2 to 0.4 (and neutralized by an equivalent decrease of the charge on the glycosidic oxygen). Otherwise the ring structure evolved to the most stable $^4C_1$ conformation. Noteworthy, the charge value of the anomeric carbon was increased from 0.2 to 0.384 in the actual generation of the Glycam parameters set (from Glycam04 forcefield to Glycam06).

Even though the $^4C_1$ conformation results in a long distance from the nucleophilic residue, the two conformations ($^1S_3$ and $^4C_1$) fit nicely inside the binding site. As in the previous work with the tetrasaccharide-MUF complex, we decided to perform two separate MD simulations. Both systems were submitted to geometry optimization and equilibration at 300 K using the protocol described before. The equilibrated structures were used as initial geometries for the subsequent CPMD/MM simulations.

CPMD/MM simulations.

CPMD/MM simulations were performed using the method developed by Laio *et al.*[116] which combines the Car-Parrinello molecular dynamics (CPMD)[112] based on the density functional theory (see the methods Chapter for a detailed description of the method). The QM/MM interface was modeled by a link-atom monovalent pseudopotential. The electrostatic interactions between the QM and MM regions were handled via a fully Hamiltonian coupling scheme[152] where the short-range electrostatic interactions between the QM and the MM regions are explicitly taken into account for all atoms. An appropriate 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 were treated via a multipole expansion. Bonded and van der Waals interactions between the QM and the MM regions were treated with the standard AMBER force field. Long range electrostatic interactions between MM atoms were described with the P3M implementation[153] using a 64 x 64 x 64 mesh.

In all the cases, the QM region was chosen as the substrate at subsites -1 and +1 (e.g. in figure III-5 is the GH8 system). The link atoms were located at the anomeric carbon of subsite -2 and the C4 atom of subsite +2. Such QM region was enclosed in an orthorhombic supercell that ensured that the minimum distance between an atom and its boundary is at least 7 a.u. (figure III-5). The rest of the substrate, the protein and the solvent atoms were treated with molecular mechanics. Kohn-Sham orbitals were expanded in a planewave basis set with a kinetic energy cutoff of 70 Ry. Norm-conserving Troullier-Martins *ab initio* pseudopotentials[114] were used for all elements. The calculations were performed using the Perdew-Burke-Ernzerhoff generalized gradient-corrected approximation (PBE).[154-156] This functional form has been proven to perform

Figure III - 5: Quantum atoms and supercell used in the simulation of the MC of the GH8 CelA.

well in the description of hydrogen bonds[157] and was already used in previous works on GHs in our group.[34] The time step (Δt) and the fictitious mass for the electronic degrees of freedom of the CP Lagrangian (μ) were chosen as described in table III-2 below.

Structure optimization was performed by means of annealing of the ionic velocities in a MD algorithm. Afterwards, the target temperature of the CPMD/MM simulations (300 K) was reached by applying a Nosé-Hoover thermostat[158] with a frequency of 3500 cm^{-1} on the nuclei. The CPMD/MM simulations were extended for at least 15.0 and the firsts ≈ 5 ps of simulation were taken as equilibration period. The CPMD/MM simulation parameters for each system are specified in table III-2.

Table III - 2: CPMD/MM simulations parameters

		wild type	E95Q mutant
inverting β-endoglucanase from *Clostridium Thermocellum*	QM Box	17.2 x 14.0 x 15.4 Å	17.2 x 14.2 x 15.6 Å
	NN/MIX/ESP regions	20/22/30 Å	10/12/14 Å
	μ	850 a.u.	850 a.u.
	time step	6 a.u.	6 a.u.
	total simulation time	20.4 ps	17.5 ps

		Natural pentasaccharide	thio-DP5 substrate
1,4-β-endoglucanase I	from *Fusarium oxysporum*		
		QM Box	10.2 x 9.6 x 6.0 Å	16.6 x 14.2 x 13.7 Å
		NN/MIX/ESP regions	20/22/30 Å	10/12/14 Å
		μ	850 a.u.	850 a.u.
		time step	6 a.u.	6 a.u.
		total simulation time	19.0 ps	17.4 ps
1,3-1,4-β-endoglucanase	from *Bacillus*	**chair**	**skew**
		QM Box	15.1 x 12.7 x 14.3 Å	11.7 x 13.3 x 16.4 Å
		NN/MIX/ESP regions	10/12/14 Å	10/12/14 Å
		μ	850 a.u.	850 a.u.
		time step	6 a.u.	6 a.u.
		total simulation time	15.0 ps	25.5 ps

Results and discussion

Inverting family GH8 endoglucanase CelA from Clostridium Thermocellum.

The conformation of subsite -1 glucose ring during the CPMD/MM simulation was analyzed by monitoring the **Cremer** and **Pople** puckering coordinates[159] during the simulation(see Chapter VI). The distortion of the substrate in the wild type enzyme is very similar to that of the E95Q mutant. It is very close to $^{2,5}B$ (figure III-6), in concordance with the crystallographic structure (of the E95Q mutant).
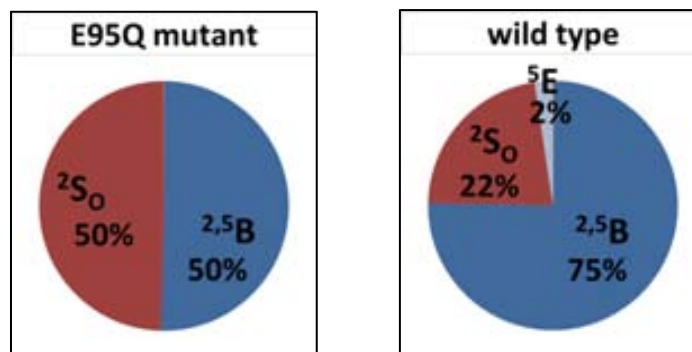

Figure III - 6: Histogram analysis of the subsite -1 ring puckering coordinates for the CPMD/MM simulations of the E95Q mutant and the wild type systems.

To check the reliability of the method used, structural parameters other than the substrate distortion were compared to experimental data. Table III-3 lists average structural values obtained from the MD simulations. The structural parameters considered include interactions among quantum atoms and between quantum and classical atoms. The results for the E95Q mutant differ, in average, less than 0.1 Å from the X-ray structure. Therefore, we can be confident when using this methodology in those situations in which there are no high resolution experimental structures available.

Table III - 3: Geometric values for the experimental and the average of E95Q and the wild type simulations. In parenthesis is the standard deviation. All distances are in Å. Hydrogen bond distances are calculated as the distance between heteroatoms. The planarity of C1 was calculated as the distance between the C1 atom and the C2-O5-H1 plane. Atoms and residues are labeled as in figure III-1.

	exp	E95Q mutant	wild type E.	$\Delta_{wt-E95Q}$
C1··Ox distance	1.40	1.48 (.004)	1.50 (.004)	+ 0,021
C1··O5 distance	1.40	1.43 (.008)	1.40_5 (.006)	- 0,011
C1··C2 distance	1.50	1.54 (.026)	1.54 (.003)	+ 0,001
Ox··Cx distance	1.43	1.44 (.030)	1.44_5 (.011)	+ 0,006
C1 planarity		0.42_5 (.030)	0.42 (.018)	- 0,009
C1··H_2ONu distance	3.32	3.40 (.327)	3.45 (.117)	+ 0,053
Ox··E/Q95 distance	2.75	2.95 (.050)	2.82 (.079)	- 0,132
O2··D152 distance	2.66	2.49 (.014)	2.53 (.078)	+ 0,048
O3··D152 distance	2.77	2.49_5 (.059)	2.50 (.064)	+ 0,002
H_2ONu··D278 distance	2.59	2.63 (.006)	2.70 (.142)	+ 0,069
H_2ONu··Y215 distance	2.97	2.99 (.181)	3.00 (.027)	- 0,009
-1 subsite conformation	$^{2,5}B$	$^{2,5}B/^2S_O$	$^{2,5}B$	

The effect of the mutation on the enzyme-substrate interactions was analyzed by comparing the results obtained for both systems. Most of the structural parameters are very similar. There are, however, some interesting differences that deserve consideration.

The hydrogen bond between the glycosidic oxygen (Ox) and residue 95 is stronger (shorter by 0.13 Å) for glutamic acid (wt) than for glutamine (mutant). This is an expected result, given the negative charge of the acid. A stronger H-bond might favor a slightly higher electron density on the glycosidic oxygen and thus a greater polarization of the glycosidic bond, reflected by a longer bond distance (by .021 Å). As a consequence, the charge transfer from the pyranic oxygen to the anomeric carbon increases, which shrinks the C1-O5 distance (by .011 Å) and enhances the planarity around the anomeric center (by .009 Å). These are, certainly, very subtle changes but chemically consistent with the E95Q mutation electrostatic effects. The overall structure of the Michaelis complex, and in particular the distortion at the subsite -1,is not affected by the mutation.

1,4-β-endoglucanase from Fusarium oxysporum

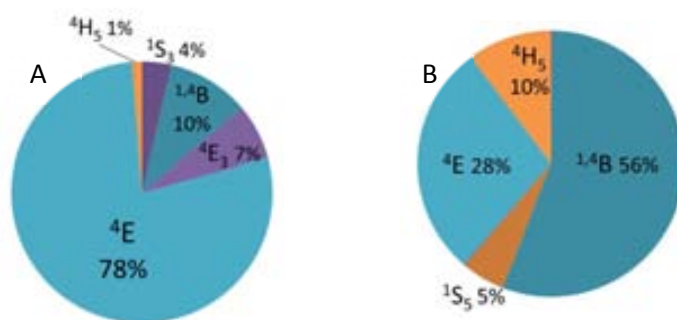The analysis of the -1 subsite ring conformation is shown in figure III-7.

Figure III - 7: natural substrate (A) and thio-DP5 (B) inhibitor subsite -1 ring distortion distribution analysis.

The saccharide ring at subsite -1 adopts a similar conformation in the two simulated systems (the natural substrate and the thio-analogue), since ^{4}E and the $^{1,4}B$ conformations only differ in the value of the θ angle, which is $\pi/4$ rad in the former and $\pi/2$ in the latter. Therefore, both systems can be considered "equivalent" in terms of substrate distortion.

It is noteworthy that our results are in good agreement with the distortion observed in the crystallographic structure ($^{1,4}B$). Given the excellent agreement we had previously found for the atomic-resolution GH8 structure we conclude that the small differences between the x-ray experimental data and the CPMD/MM calculation are due to the limited resolution of the crystal structure (2.70 Å).
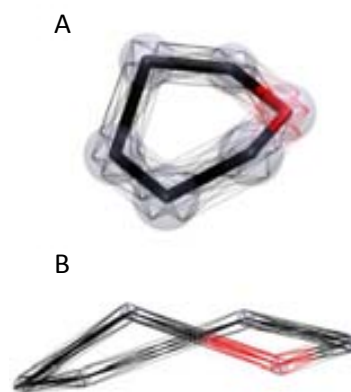
Figure III - 8: Atomic anisotropy and possible ring conformations for a B-factor of 42 (A) and a B-factor of 6.5 (B).

It is worth analyzing in further detail the consequences of the low resolution in the conformation of the -1 subsite ring. The most disordered region (i.e. the residues that have the highest B factor) is the binding site of the substrate and especially, the substrate itself. Actually, only two of the five rings of the thio-saccharide inhibitor were ordered enough to be solved and their B-factors are relatively high (42 Å^2 in average). Considering that $B = 8\pi^2 \langle u^2 \rangle$ where $u^2 = \frac{1}{N} \sum (x_i - \bar{x})^2$ then it means that the "real" atom position for the substrate atoms can be within 0.5 Å from the "assigned" position (figure III-8 A).

In comparison, the previous case (the GH8) in which the resolution was of 0.94 Å and the average B-factors of the substrate atoms was of 6.5 Å^2, the expected atom deviance was of 0.08 Å, significantly much more precise (figure III-8 B). Of course, other considerations such as bond length, angles restraints or energy minimization are taken into account when modeling the

substrate from the electron density observed in the X-ray experiment. Nevertheless, the reliability of the assigned ring conformation is highly dependent on the resolution of the crystal and can be ambiguous if the electron density is not clear. This becomes evident in figure III-9, where high and low resolution structures (GH8 and GH7 respectively) are compared.
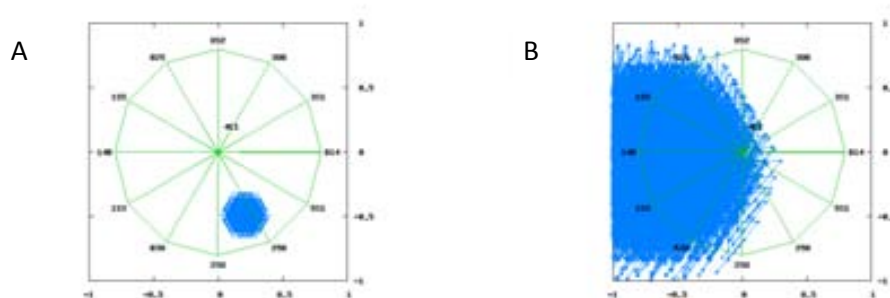

Figure III - 9: Possible ring distortion for a high resolution structure (A) and a low resolution structure (B) taking into account the B-factors and distance and angle restraints only.

In spite of these likely sources of error, it is remarkable that the crystal and the computed structures converge to a very similar conformation (close to $^{1,4}B$) for the -1 subsite glucose ring.

Structural parameters describing enzyme-substrate interactions and substrate geometry are shown in table III-4.

The main differences between the MC with the natural substrate and with the thio-derivative are obviously found on those distances involving the glycosidic oxygen/sulfur atom (i.e. the C1··Ox and the Ox··Cx distances are ≈ 0.4 Å shorter than the C1··Sx and the Sx··Cx distances). The glycosidic angle is smaller for the inhibitor complex compared with the natural substrate. Instead, the glycosidic torsional angles (φ_H and Ψ_H) are very similar in both systems, probably because they are dictated by the saccharides at subsites -1 and +1 relative positions, which are tightly anchored to the binding cavity.

The C1-O5 distance is slightly longer in the thio-DP5 ligand compared with the natural substrate. This indicates that the degree of double bond character is mainly influenced by the ring distortion, whereas the structure and chemistry of the glycosidic bond has little effect. The distances with the catalytic residues also show small variations between the substrates (except for the Ox/Sx ··· E202 distance). Electronically, sulfur is less electronegative than oxygen, resulting in a lower affinity for the acidic proton (the distance with the carboxylic acid group is 0.35 Å longer). The higher charge development on the anomeric carbon for the complex with the natural substrate is reflected by a shorter distance (by 0.21 Å) with the catalytic nucleophile.

In view of these results, it can be concluded that the thio-saccharide inhibitor forms a similar pattern of enzyme-substrate interactions and adopts the same ring distortion as the natural substrate.

Table III - 4: Main structural parameters measured from the CPMD/MM simulations of the natural substrate and the thio-derivative inhibitor respectively. In parenthesis is the standard deviation. All distances are in Å. Hydrogen bond distances are measured as the distance between heteroatoms. φ_H and Ψ_H are a widely used variation of the φ and Ψ torsion angles recommended by the IUPAC-IUB to indicate the relative orientation of two adjacent pyranose rings. Concretely φ_H is the H1-C1-Ox-Cx torsion and Ψ_H is the torsion between the C1-Ox-Cx-Hx atoms.

Parameter	Natural substrate	Thio-derivative
C1··Ox distance	1.52 (.013)	1.93 (.101)
C1··O5 distance	1.39 (.042)	1.41 (.041)
C1··C2 distance	1.56 (.022)	1.55 (.000)
Ox··Cx distance	1.45 (.029)	1.86 (.005)
Ox··E202 distance	2.76 (.094)	3.10 (.017)
Cx··E197 distance	3.34 (.080)	3.56 (.057)
O2··E197 distance	2.49 (.081)	2.57 (.123)
C1-Ox-Cx angle	115.6 (0.2)	102.4 (3.5)
φ_H	-47.9 (3.5)	-32.0 (12.1)
Ψ_H	-21.3 (9.3)	-20.1 (1.4)
-1 subsite conformation	^4E	$^{1,4}B$

Retaining 1,3-1,4-β-endoglucanase from Bacillus.

Analysis of the relative weighs of all conformations (figure III-10) shows that the "chair" and the "skew" conformers are both stable. This was also observed in the case of the substrate with the MUF aglycon.[34] Both conformations are possible, but a transition between them does not occur.

A   B
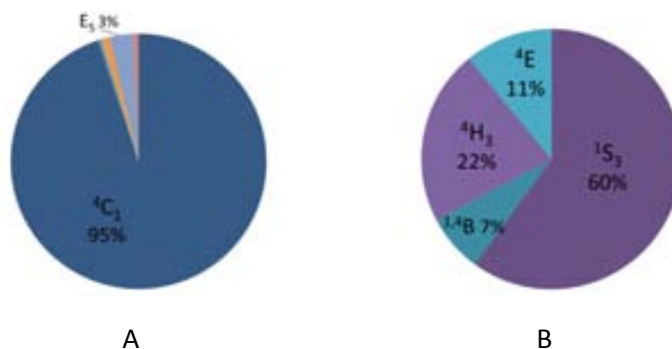
Figure III - 10: Subsite -1 ring distortion in 1,3-1,4-β-endoglucanase when the substrate is the natural hexasaccharide. A) Simulation starting from the chair conformation. B) Simulation starting from the distorted $^1S_3$ conformation.

Figure III-11 compares the relative weight of the distorted conformations of the natural hexasaccharide substrate (B in figure III-10) with the case of the substrate with the MUF aglycon. It can be noted that there are subtle changes in the relative weight of the conformations: It is $^1S_3/^{1,4}B$ for the MUF substrate but $^1S_3$ for the natural substrate.
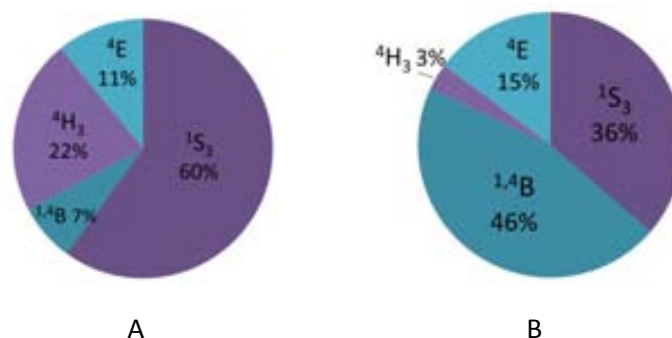
Figure III - 11: Subsite -1 ring distortion in the Michaelis complex with the hexasaccharide substrate (A) and the tetrasaccharide-MUF moiety (B).

Substrate-enzyme interactions from subsites -1 to +2 are analyzed in table III-5. For an easier understanding, enzyme-substrate interactions found in both systems are shown in blue, while interactions only found in one of the systems are shown in red. *E.g.* the +1 subsite in the complex with the natural substrate (G4G4G4G3-G4G) shows a hydrogen bond O3⋯Tyr123. This interaction is marked in blue because it is also found with the G4G4G4G3-MUF substrate: +1 subsite O1⋯Tyr123.

Table III - 5: Enzyme-substrate interactions for subsites from -1 to +2 observed in the natural substrate and the MUF-tetrasaccharide substrate. Equivalent interactions are in blue while exclusive interactions are in red.

	subsite	type of Hbond	substrate atom					
G4G4G3G-G4G			o2	o3	o4	o5	o6	vdw
	s -1	donor	Glu105 (1.5)		s -2 (1.8)		Glu109 (1.8)	
		acceptor			Asn182 (2.2)		Trp184 (2.2)	
			o2	o3	o4	o5	o6	vdw
	s +1	donor	Asn129 (1.8)	Glu109 (2.2)				Trp192 (3.7)
		acceptor	Gln119 (2.2)	Tyr123 (2.1)	Glu109 (2.1)	s +2 (2.1)		
			o2	o3	o4	o5	o6	vdw
	s +2	donor		s +1 (2.1)			Glu131 (1.7)	Trp192 (3.7)
		acceptor					Asn129 (2.2)	
G4G4G3G-MUF			o2	o3	o4	o5	o6	vdw
	s -1	donor	Glu105 (1.7)				Glu109 (1.9)	
		acceptor			Asn182 (2.2)		Trp184 (2.9)	
			o1	o2	o8	vdw		
	s +1+2	donor				Trp192 (3.7)		
		acceptor	Tyr123 (2.1)	Asn129 (2.5)	Glu109 (3.5)			

Both systems have a glucose moiety in subsite -1, and its interactions with the enzyme are also the same (table III-5). It does not make any stacking interaction with the enzyme, but it makes strong

hydrogen bond interactions with the catalytic residues Glu105 and Glu109, as well as with Asn182 and Trp184.

Both leaving groups are able of making strong stacking interactions with Trp192, but the disaccharide is much more functionalized than the MUF aglycon and therefore can make more polar and hydrogen bonds interactions with the protein (in particular with residues Glu131 and Gln119) as well as intramolecular hydrogen bonds.

Overall, table III-5 shows that, in spite of the changes in the leaving group (MUF *vs* G4G), the modified substrate forms similar interactions with the enzyme compared to the natural substrate. In particular, the important stacking interaction with Trp192 is maintained. As a result, both substrates adopt a similar distortion ($^1S_3$/$^{1,4}B$ for G4G4G4G3-MUF and $^1S_3$ for G4G4G4G3-G4G) and, therefore, it is expected that they will follow the same catalytic itinerary.

Conclusions

1. CPMD/MM simulations minutely reproduce the high resolution structure of endoglucanase CelA from *Clostridium Thermocellum* in complex with a cellohexaose substrate.

2. The mutation of the general acid/base catalytic residue (glutamic acid) to glutamine in GH8 inverting endoglucanase does not have a significant effect on substrate ring distortion.

3. The oxygen to sulfur substitution on the glycosidic bond in retaining GH7 1,4-β-endoglucanase from *Fusarium oxysporum* perturbs the MC structure but has a little effect on the -1 substrate ring distortion (from ^4E to $^{1,4}B$).

4. Both $^1S_3$ and the $^4C_1$ conformations are stable in the active site of 1,3-1,4-β-endoglucanase from *Bacillus* in complex with an hexasaccharide substrate. By analogy with previous calculations performed with a G4G4G4G3-MUF substrate,[34] we can infer that the distorted $^1S_3$ conformation is the most stable and that both conformers cannot interconvert.

5. The change of a disaccharide by a MUF aglycon does not have any significant effect on the -1 substrate ring distortion (from $^1S_3/^{1,4}B$ to $^1S_3$).

Chapter IV - Influence of the enzyme-substrate interactions through the 2-OH on substrate distortion.

Influence of the enzyme-substrate interactions through the 2-OH on substrate distortion.

Introduction.

In the previous Chapter, we have shown that CPMD/MM simulations reproduce the distortion of the -1 saccharide ring on GHs. By investigating specific experimental complexes, we also showed that three of the most used modifications to study substrate catalytic itineraries, namely mutation of the general acid/base residue, the thio-glycoside chemical modification and the leaving group substitution (by a MUF aglycon); lead to just small changes on the conformation of the substrate in the Michaelis complex (MC).

Another common strategy to investigate catalytic itineraries consists on crystalizing the wild type enzyme with a so-called Withers fluoro-sugars inactivators.[40] Because of the importance of these inhibitors in mechanistic studies of GHs, we decided to analyze them in a separate chapter.

<div style="border:1px solid">

Side panel: Withers' inactivators

Fluoro-sugar inactivators

Withers inhibitors consist in the natural glucoside substrate, with one of the exocyclic hydroxyl groups substituted by a fluorine atom. The influence of the chemical nature of each exocyclic group is well known and reflected by the relative rate of hydrolysis of 2-deoxy > 4-deoxy > 3-deoxy ≈ 6-deoxy > parent > 6-deoxy-6-fluoro > 3-deoxy-3-fluoro > 4-deoxy-4-fluoro > 2-deoxy-2-fluoro glycosides.[160] Since the transition state is known to have a high oxocarbenium ion-like character (see the Introduction Chapter), this effect has been rationalized on the basis of the relative stability of the TS due to the different σ-acceptor effect of the ring substituents[161, 162] (see figure IV-2).

2-deoxy-2-fluoro gluco-derivatives (figure IV-1) are relatively good substrate-like inhibitors for retaining glucoside hydrolases. On top of that, Withers and coworkers,[163] showed that by introducing a good leaving aglycon group (*e.g. o-p*-dinitrophenyl) on the fluoro-sugar inactivator, it is possible to accelerate the first step of the reaction again (the glycosylation step) without affecting the deglycosylation step. As a result, the glycosylation step can proceed but deglycosylation is inhibited. Therefore, the covalent intermediate of the overall hydrolysis reaction is isolated.

Figure IV - 1: 2-deoxy-2-fluoro glucose

However, if the same experiment is performed at low pH conditions, the catalytic nucleophile becomes protonated, being unable to attack the anomeric carbon. In these conditions, even if a good leaving group is provided, the glycosylation reaction cannot proceed and only Michaelis complex is isolated.
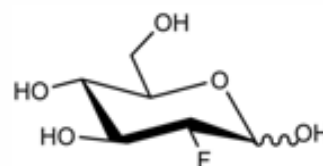
</div>

In summary, by synthetizing a single Withers fluoro-sugar inactivator with a good leaving group, it is possible not only to trap the MC (by crystalizing the binary complex at low pH conditions), but also the covalent intermediate (by crystalizing it at neutral pH). By identifying the subsite -1 ring distortion at both the MC and the covalent intermediate, the distortion at the transition state could be inferred. This powerful and versatile strategy has been used,[13] for instance, in the design of potent TS-like inhibitors, which are more efficient than substrate-like inhibitors.
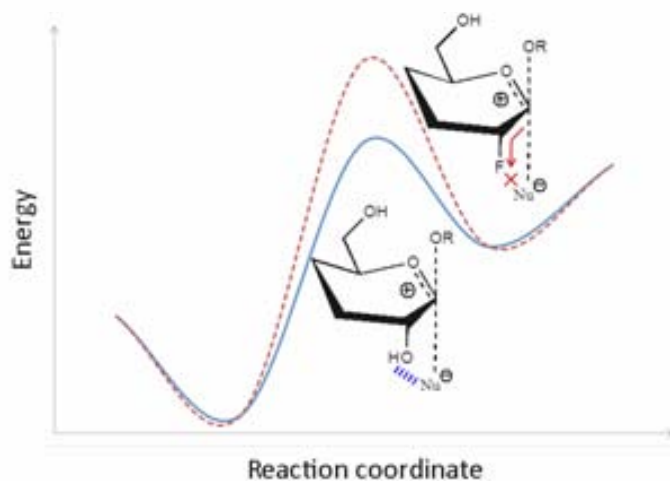

Figure IV - 2: TS destabilization by the 2-deoxy-2-fluoro substrate modification (in red) and C2(OH) hydrogen bond stabilization (in blue).

Nevertheless, it has made evident that for many enzymes, the hydrogen bond interaction of the C(2)OH substituent of subsite -1 overpowers the inductive effect of this group.[160] Therefore, by substituting the hydroxyl group by a fluorine atom, this hydrogen bond is also removed and this effect is added to the fluorine inductive effect in increasing the TS energy (figure IV-2).

The stabilizing hydrogen bond with the nucleophile residue has an important role in catalysis. It has been pointed out that the sole contribution of this hydrogen bond lowers the reaction energy barrier energy by up to 12 kcal·mol^{-1}.[160, 164, 165]

In this Chapter we studied the importance of the interaction between the C2 exocyclic group and the nucleophile residue in the MC substrate distortion of 1,3-1,4-β-endoglucanase. We will do this by analyzing how electrostatic; hydrogen bonding and steric hindrance affect this interaction. At the same time, we will assess the reliability of the 2-deoxy-2-fluoro inhibitors in mimicking the structure of the natural substrate.

Model systems

A total of seven models were used to study the interactions of the nucleophile residue glutamate 105 with the C2 exocyclic group of subsite -1 sugar moiety and its implications for the ring

distortion in the MC of 1,3-1,4-β-endoglucanase (summarized in table IV-1 below). The Michaelis complex of 1,3-1,4-β-endoglucanase with the MUF-tetrasaccharide substrate was used as a template to build the starting structures for the subsequent simulations.

This complex was taken as the reference system (model 1 in table IV-1). Two models were used to study the effect of the 2-deoxy-2-fluoro glucose substitution on substrate distortion at low and neutral pH (models 2 and 3). The only difference between them is the protonation state of the nucleophilic residue Glu105 (protonated at low pH but deprotonated at neutral pH). Concerning enzyme modifications, four mutants E105Q, E105D, E105A and E105I were considered (models 4 to 7 respectively). All the mutant simulations and simulations 1 and 3 were performed at neutral pH (when residue 105 was an acidic aminoacid, its conjugate base form was used).

Table IV - 1: Scheme of the seven simulations performed.

Model	Enzyme	Substrate[a]
1	Wild type	Glucose
2	Wild type, low pH	2-deoxy-2-fluoroglucose
3	Wild type, neutral pH	2-deoxy-2-fluoroglucose
4	E105Q mutant	Glucose
5	E105D mutant	Glucose
6	E105A mutant	Glucose
7	E105I mutant	Glucose

a) substrate refers only to the pyranose ring occupying subsite -1

The choice of the enzyme mutants was made on the basis of the kind of interaction with the glucopyranose moiety located at subsite -1 could they affect. We differentiated three possible interactions:

<u>Electrostatic</u>: Negatively charged residues can stabilize de positive charge developed on the anomeric carbon upon substrate distortion.[34] In this sense, mutations at residue 105 can have a formal -1 charge (WT and E105D) or be neutral (E105Q, E105A and E105I). Positively charged aminoacids have not been considered.
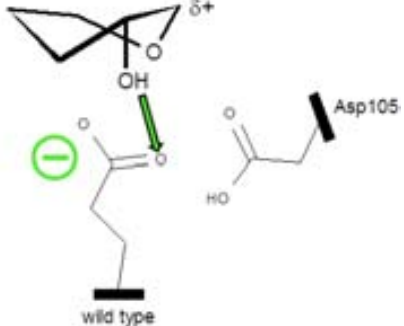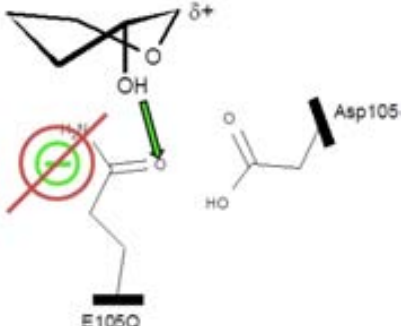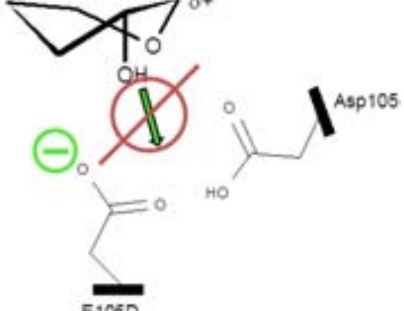
<u>Hydrogen bond</u>: Polar residues at position 105 can accept a hydrogen bond from the hydroxyl at position C2 (e.g. WT and E105Q) resulting in a stabilizing interaction or not (e.g. E105D, E105A and E105I).

<u>Steric hindrance</u>: Big hydrophobic residues establish repulsive steric interactions with the subsite -1 glucose ring, while small residues render a bigger binding cavity. Mutations were made into big (WT, E105Q and E105I) or small residues (E105D and E105A). Aromatic residues were not considered because for structural reasons, residue 105 cannot establish stacking interactions with the substrate.

The mutants chosen (together with the WT model) would allow us to decipher which interactions are important, and to which extent, for the distortion of the -1 subsite glucopyranose. Table IV-2 gives a description of how each of the mutants is expected to affect the interactions above.

Table IV - 2: Description of the mutants chosen and the effect of each mutation on the -1 subsite ring distortion.

Model 1	**Glutamate (wt)** H-bond acceptor; negatively charged. The wild type enzyme is the *reference experiment*. The expected MC distortion is $^1S_3$ (by analogy with the previous calculations[34]), with a strong hydrogen bond with the C2 hydroxyl group. The negative charge of the glutamate should have an effect on the stabilization of the partial δ^+ charge development on the anomeric carbon upon ring distortion.	

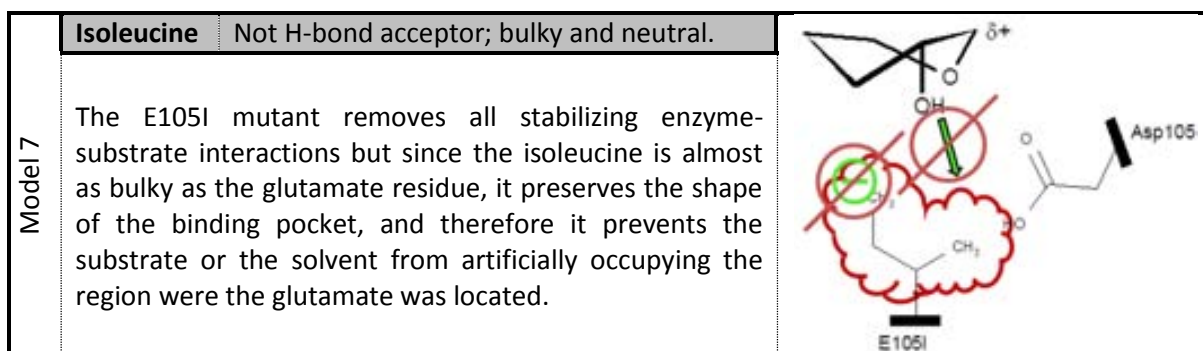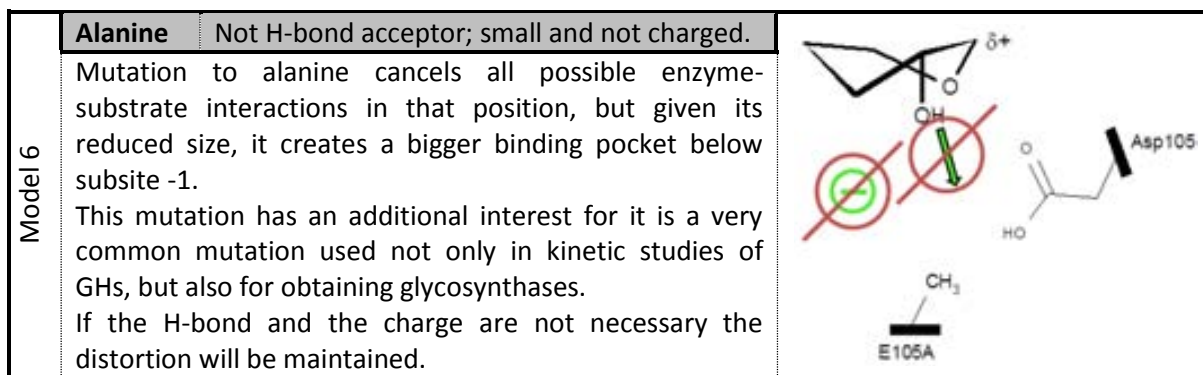|		
Model 4	**Glutamine** H-bond acceptor; neutral. Glutamine has the same size and it is capable of making the same type of hydrogen bond as the glutamate residue, albeit somewhat weaker. However, since it is not negatively charged, it does not have any stabilizing effect on the positive partial charge developed on the anomeric carbon. If the charge is necessary, the distortion will be lost. If the H-bond is enough, the distortion will be maintained.	

Model 5	**Aspartate** Not H-bond acceptor; negatively charged. Aspartate is one carbon shorter than the glutamate so it cannot make the hydrogen bond with the C2 hydroxyl substituent. Therefore, it keeps the effect of the negative charge but retrieving the H-bond. If the H-bond is necessary, the distortion will be lost. If the charge is enough, the distortion will be maintained.	

Model 6	**Alanine** Not H-bond acceptor; small and not charged.	
	Mutation to alanine cancels all possible enzyme-substrate interactions in that position, but given its reduced size, it creates a bigger binding pocket below subsite -1. This mutation has an additional interest for it is a very common mutation used not only in kinetic studies of GHs, but also for obtaining glycosynthases. If the H-bond and the charge are not necessary the distortion will be maintained.	

Model 7	**Isoleucine** Not H-bond acceptor; bulky and neutral.	
	The E105I mutant removes all stabilizing enzyme-substrate interactions but since the isoleucine is almost as bulky as the glutamate residue, it preserves the shape of the binding pocket, and therefore it prevents the substrate or the solvent from artificially occupying the region were the glutamate was located.	

Each of these systems was submitted to CPMD/MM simulation and the -1 subsite ring distortion evolution was analyzed.

Computational details

The Michaelis complex of 1,3-1,4-β-endoglucanase with a 4-methyl umbelliferyl tetrasaccharide, previously studied in the group, was taken for model 1 in table IV-2 and as a template for building the mutants E105Q, E105D, E105A and E105I (models 4 to 7) as well as the 2-fluoro inhibitor complexes (models 1 and 2). The low pH conditions were simulated by manually protonating the nucleophile glutamate 105 to its acid form, while in the simulations at physiologic pH, the glutamate form was used.

All the simulations were performed using the Car-Parrinello Molecular Dynamics (CPMD) method in a hybrid QM/MM framework, as in Chapter III. The part of the enzyme included in the QM region consisted in residues 105, 107 and 109 truncated on the alpha carbon (residues 105 and 107) and on the beta carbon (residue 109) by a monovalent pseudopotential.[166] The rest of the QM atoms were the part of the substrate occupying subsites -1 to +2 (a glucopyranose moiety and the MUF aglycon) truncated on the anomeric carbon of subsite -2 using also a monovalent pseudopotential. Therefore, the size of the QM system ranged from 63 to 72 atoms, depending on the system studied. The MM part of the system consisted in the rest of the substrate and protein, and the TIP3P water solvent and chloride counterions. The NN/MIX/ESP regions cutoff for the

electrostatic interactions between the QM and the MM parts of the system were set to 10/12/14 a.u. respectively. In all cases, the QM region was enclosed in a supercell big enough to ensure a minimum distance of 7 a.u. from any atom to its closest boundary. Troullier-Martins *ab initio* pseudopotentials[114] and the Perdew-Burke-Ernzerhoff generalized gradient-corrected approximation (PBE)[154-156] were used. The Cornell *et al.* forcefield[144] was used to describe the MM part of protein and the solvent molecules while the Glycam04 forcefield[145-147] was used for the saccharide residues. The time-step chosen for all the simulations was of 5 a.u. and the electron mass μ used was of 850 a.u. to ensure the adiabaticity of the system. The simulations were performed at constant temperature by coupling the system to a Noseé-Hoover thermostat[158] at 300 K.

All the systems were initialized using the following protocol: First, an optimization of the starting structure was performed by means of molecular dynamics with annealing of the ionic velocities until the highest nuclear gradient was of $5 \cdot 10^{-4}$ H/a.u. Afterwards, the system was heated up to 300 K and the MD was extended until a stable structure was reached. The total simulation time ranged from 6.70 ps to 20.68 ps depending on the system studied. The first two picoseconds of each simulation were discarded as equilibration time.

Results

For all systems, the protein backbone atoms remained very close to their starting position and very few side-chain protein residue rearrangements were observed. Most of the side-chain movements affected residues 105 and 107, which are the mutated residue and the aspartic acid next to it, respectively. The general acid/base catalyst residue Glu109 forms a stable hydrogen bond with the glycosidic oxygen in all the simulations, independently of the subsite -1 sugar ring conformation.

Model 1: Reference system

The subsite -1 glucose ring conformation was tightly conserved on a $^{1,4}B/^{1}S_3$ distortion, as expected by comparison with X-ray structures of other endoglucanases (see the Introduction Chapter). This distortion was stabilized with a strong hydrogen bond (1.87 Å in average) between the catalytic nucleophile Glu105 and the sugar hydroxyl group. Aspartate 107 kept a low barrier (1.57 Å) H-bond with the catalytic nucleophile. In the figure IV-4 a snapshot of the MD with the relevant H-bond distances is shown. Embedded in the figure is the representation of the puckering coordinates calculated for the subsite -1 sugar ring along the whole trajectory and projected
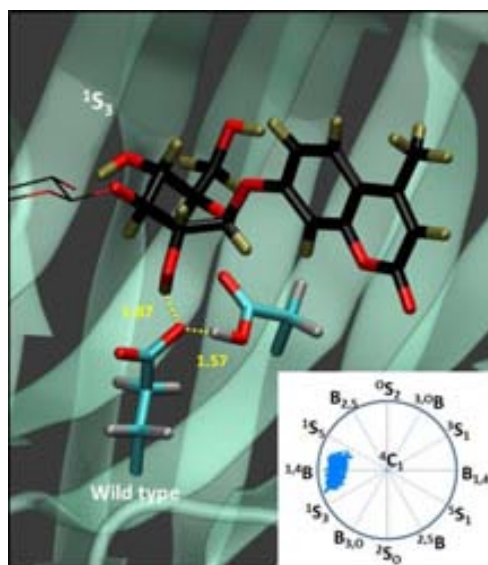

Figure IV - 3: Snapshot of the wild type simulation and the projection of the computed puckering coordinates.

onto the Stoddart diagram.[15]

Model 2: WT enzyme (low pH) in complex with 2-deoxy-2-fluoro glucose.

In models 2 and 3 (see table IV-1), the enzymes were not mutated and the substrate used was the 2-deoxy-2-fluoro derivative. As explained in the *Introduction*, this modification combined with a good leaving group, provides crystallographers with a suicide inhibitor which can overcome the first nucleophilic displacement if the pH media is physiologic (i.e. the nucleophile Glu105 is unprotonated) but it is not reactive at all if the buffer pH is low (and Glu105 is in its acid form). Both conditions were studied in models 3 and 2 respectively.

Even though there is no crystal structure of 1-3,1-4-β-endoglucanase with a 2-deoxy-2-fluoro substrate, by analogy with other X-ray experimental structures[13] and previous calculations with the natural substrate, we could expect a $^1S_3$ distortion for the 2-fluoro derivative at low pH. Indeed, the simulation of model 2 revealed a very stable $^1S_3$ distorted conformation for subsite -1 (Figure IV-3 A). The glutamic acid (residue 105) is H-bonded with the fluorine atom at position 2 at an average distance of 2.92 Å (distance measured between heteroatoms).
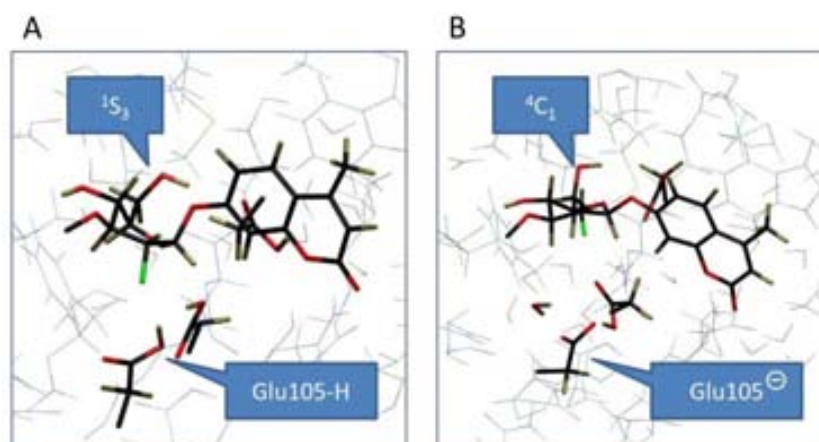
Figure IV - 4: Representative snapshots of simulations 1 (A) and 2 (B). In licorice is represented the quantum atoms while in lines are other MM of the system.

Model 3: WT enzyme (neutral pH) in complex with 2-deoxy-2-fluoro glucose.

The -1 subsite 2-fluoro-glucose moiety rapidly evolved towards a $^4C_1$ undistorted conformation during the simulation at a neutral pH (simulation 2). It fluctuated around this conformation during most of the simulation (20.68 ps) as shown in figure IV-3 B above. The reason behind this result is that the interaction between residue 105 and the fluorine atom at position 2 became no longer stabilizing but repulsive. This drove the sugar upwards on the binding site, with the concomitant loss of the distorted conformation. The average distance between the fluorine atom and the carboxylic acid was of 4.36 Å (1.44 Å longer than in simulation 1). Due to the new accommodation of the substrate inside the binding pocket, the nucleophilic residue is not capable of attacking the

anomeric carbon, for the distance between them is of 4.30 Å. As a consequence, the substrate becomes no longer preactivated for catalysis.

The results for model 3 show that the 2-deoxy-2-fluoro inhibitor does not reproduce the structure of the natural substrate (at neutral pH). This is something that had not been observed experimentally because all available structures of E·S complexes with a 2-deoxy-2-fluoro inhibitor are at low pH (i.e. with the nucleophilic residue protonated, Glu105 in our case). Our calculations suggest that, if the same experiment was done at neutral pH, the sugar ring will not be distorted. Of course, this experiment needs to be done with a bad leaving group (e.g. *p*-bromo-phenol) to avoid the formation of the covalent intermediate.

In summary, the results for model 2 showed that the **Withers** 2-Fluoro inactivator crystalized at low pH are capable of mimicking the same enzyme-substrate interactions and reproduces very accurately the same subsite -1 ring distortion as adopted by the natural substrate at catalytically competent conditions (as shown by the results of the reference model 1 above). Therefore, it can be concluded that the subsite -1 distortion in the MC can be inferred from this structure. However, this might not be the same if the inhibitor was crystallized at physiological pH (see results of model 3). This emphasizes the significance of the hydrogen bond between the C2 substituent and the nucleophile residue: its presence preserves the $^1S_3$ conformation, but its absence cancels it out.

Model 4: E105Q mutant in complex with the natural substrate.
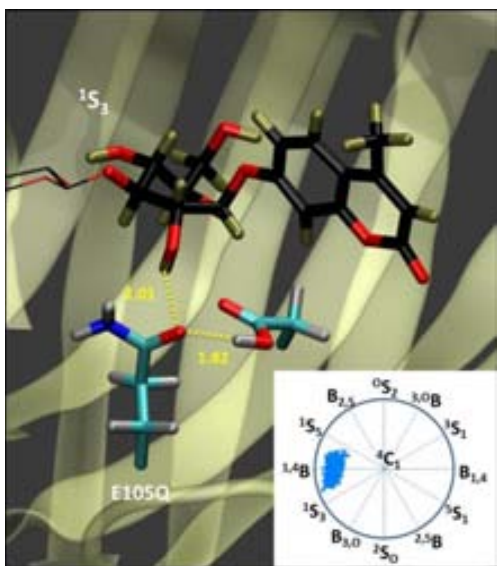
Figure IV - 5: Snapshot of the E105Q mutant (simulation 4) and the computed puckering coordinates.

The results of the MD simulation of model 4 (with the glutamate 105 residue mutated to glutamine) are shown in figure IV-5. The hydrogen bond between the sidechain amide oxygen and the sugar alcohol is maintained, yet it is longer (2.01 Å) than in the wild type case, as expected by the neutral character of the glutamine residue (compared to a glutamate). Nonetheless, the hydrogen bond interaction is enough to preserve the $^1S_3$ distortion of the subsite -1 substrate pyranose ring. Aspartate 107 is also hydrogen bonded with Gln105, although again, at a longer distance (1.82 Å) compared to the wild type enzyme. This indicates that the stabilization of the δ^+ developed on the anomeric carbon by the negative charge on the aminoacid at position 105 is not necessary to maintain the distortion.

Model 5: E105D mutant in complex with the natural substrate.

Glutamate to aspartate mutation is usually a very conservative one, but since aspartate is shorter, the H-bond between residue 105 and the substrate is broken. In this situation, the sugar manages itself to bend over by changing its distortion from $^1S_3$ to $^1S_5/^{1,4}B$ (all of them are very closed on the **Stoddart's** diagram). This distortion is stabilized by the formation of a short hydrogen bond between the C2 hydroxyl substituent and the sidechain of Asp107 (1.52 Å). As a result, Asp107 loses its proton in favor of Asp105 (although both residues still form a short barrier H-bond of 1.54 Å distance), which further strengthens the enzyme-substrate interaction (see figure IV-6). The MUF aglycon is in an axial orientation, as for the $^1S_3$ or the $^{1,4}B$ conformation.

The E105D mutation maintains the substrate distortion. However, because of the formation of a new enzyme-substrate H-bond interaction, we cannot conclude whether the negative charge of the residue 105 is enough to keep the substrate distortion.
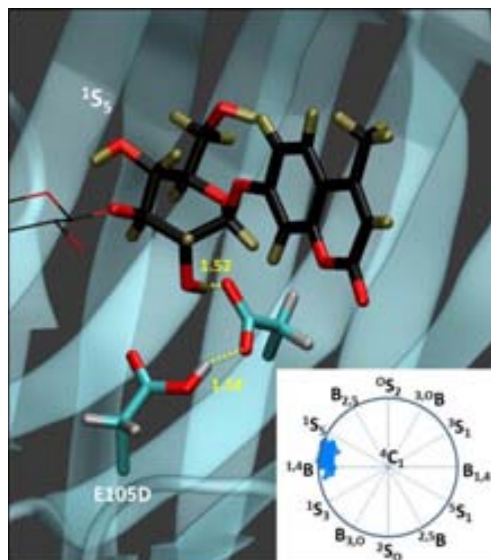
Figure IV - 6: Snapshot of the E105D mutant (simulation 5). The ring distortion is $^1S_5/^{1,4}B$.

Model 6: E105A mutant in complex with the natural substrate.

Unexpectedly, even if the alanine sidechain is chemically very different from the aspartic acid, E105A led to similar results as E105D. The subsite -1 ring adopts a conformation close to a boat $^{1,4}B$ by establishing a new interaction with Asp107 (figure IV-7). However, in this case, Asp107 cannot transfer the proton to residue 105, hence its formal charge remains zero and the interaction with the substrate is weaker than in the previous case (E105D model). This can be inferred from the increase of the H-bond distance, which is 1.88 Å compared to 1.54 Å in the alanine mutant. The only thing that the alanine and the aspartate mutation have in common is that the reduced size of their sidechains disables them from forming a H-bond with the substrate, creating a bigger cavity in the subsite -1 binding site. The substrate uses this cavity to bend over and adopts a distorted conformation by forming an H-bond with the residue Asp107. In the next section, it will be

Figure IV - 7: Snapshot of the E105A MD (simulation 6).

shown that this does not occur when residue 105 is an isoleucine.

Model 7: E105I mutant in complex with the natural substrate.

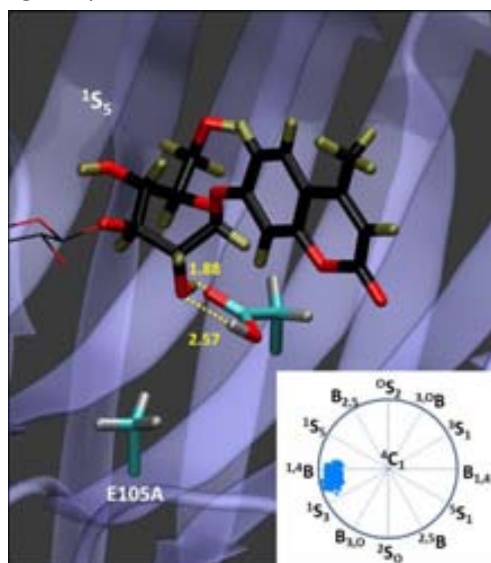Given the hydrophobic character of the isoleucine sidechain, it cannot make a hydrogen bond with the substrate and its neutral charge cannot stabilize the charge on the anomeric carbon. Moreover, given that its size is very similar to the size of the glutamic acid, the substrate could not bend over and interact with Asp107 as in models 4 and 5 (E105D and E105A mutants). As a result of this lack of stabilizing interactions, the subsite -1 saccharide ring eventually (after ≈ 1.5 ps) loses its $^1S_3$ distortion in favor of the lowest energy $^4C_1$ chair conformation (figure IV-8). In the chair conformation, the aglycon has an equatorial orientation that is incompatible with the hydrolysis reaction.

Figure IV - 8: Snapsoht of the E105I MD (simulation 7). The evolution of the subsite -1 pyranose ring changing from $^{1,4}B$ to $^4C_1$ is shown on the Stoddart diagram.

Discussion

Our results demonstrate that the ring conformation at subsite -1 is very sensitive to a perturbation of the C2 substituent with the enzyme. In other words, the interaction through the C2 exocyclic group is crucial in stabilizing the ring distortion. The most important interaction is the hydrogen bond with residue 105, while the charge of this residue is not necessary to maintain the distortion. In the lack of this hydrogen bond, the distortion is lost, unless the ring can bend over, adopting a distortion close to $^1S_5$, and establish a new hydrogen bond interaction with aspartate 107.

The E105I mutant (model 7) resembles the case of then complex between the 2-fluoro inhibitor and the wild type enzyme (model 3), in which the exocyclic group at ring carbon 2 cannot interact with any protein residue. In model 3, this was achieved by modifying the substrate while in model 7 the enzyme was modified. However, the effect of perturbing this crucial interaction is similar in both systems: the -1 subsite pyranose moves upwards and loses its distortion (figure IV-9). It could be argued that the change in subsite -1 ring conformation in model 3 is partially due to either the inductive acceptor effect of the fluorine atom or the repulsive fluorine-carboxylic acid repulsion. Nonetheless, this is not the case in model 7 where only the protein has been modified. Therefore, it can be concluded that this hydrogen bond is absolutely necessary to maintain the distortion.

The results obtained for model 2 show that the 2-fluoro substrate derivative at low pH is a good mimic of the "natural" Michaelis complex, but this is not the case at physiological pH. Our calculations predict that, different substrate conformations would be observed if crystallizing a 2-deoxy-2-fluoro substrate derivative at neutral pH. This experiment should be performed using a bad leaving group to avoid the cleavage of the glycosidic bond. Still, our prediction awaits experimental verification.

It is usually assumed that the non-reactivity of the 2-fluoro inhibitors is only due to the destabilizing inductive effect of the fluorine atom and its impossibility of forming a hydrogen bond with the nucleophile in the transition state. However, the fact that the substrate is not distorted in the Michaelis complex, as our simulations show, introduces an additional factor.

Figure IV - 9: Alignment of snapshots from simulation 2 and 7 (carbon atoms in cian and black respectively). The 4C1 conformation is correlated with an upwards reallocation of the substrate.

Part of the enzyme-substrate binding energy of GHs is invested in the -1 subsite ring distortion. Upon distortion, the system acquires a high energy preactivated state that intrinsically lowers the reaction barrier. However, if the substrate is not distorted, this contribution to the induced fit effect is lost and hence the reaction could be inhibited. To study these effects, we considered interesting to analyze the mechanism of inhibition of 2-deoxy-2-fluoro inhibitors. The results are presented in the next Chapter.
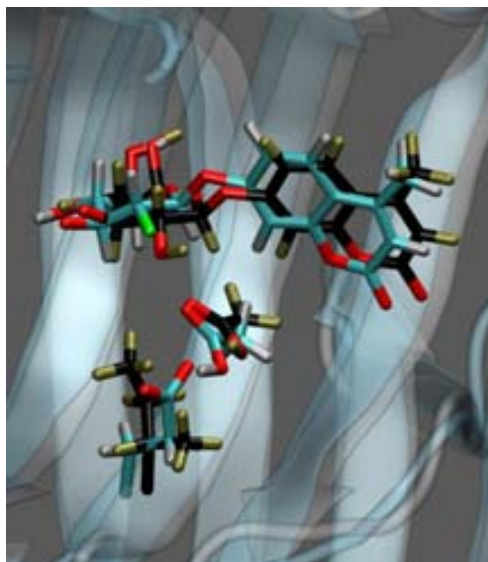
Conclusions

1. The substrate distortion in the E105Q mutant of 1,3-1,4-β-endoglucanase with a 4-methyl umbelliferyl tetrasaccharide substrate is the same as in the wild type enzyme ($^1S_3$). Therefore, the hydrogen bond between C2(OH) and residue 105 is sufficient for maintaining the substrate distortion.

2. The substrate in the E105D and the E105A mutants lack the H-bond interaction with residue 105. However, the larger binding cavity causes the substrate to change conformation from $^1S_3$ to $^1S_5$, which is stabilized *via* a hydrogen bond between C2(OH) and aspartate 107.

3. The E105I mutant also lacks the H-bond interaction with residue 105, but the binding cavity does not change. In this case, the substrate loses its distorted conformation and adopts an undistorted $^4C_1$. The hydrogen bond between C2(OH) and residue 105 is thus necessary to maintain the $^1S_3$ distortion.

4. The 2-deoxy-2-fluoro substrate derivative adopts a $^1S_3$ distorted conformation at low pH conditions. The hydrogen bond between the C2(F) and residue 105 (protonated carboxylic acid) mimics the hydrogen bond between the C2(OH) and residue 105 (carboxylate). Therefore, the 2-deoxy-2-fluoro inhibitor reproduces the distortion of the natural substrate.

5. The 2-deoxy-2-fluoro substrate derivative adopts a not distorted conformation ($^4C_1$) at neutral pH. Therefore, it does not mimic the Michaelis complex. Specific experiments are proposed to verify this prediction.

Chapter V - Mechanism of reaction in 1,3-1,4-β-endoglucanase with the 2-deoxy-2-fluoro substrate derivative.

Mechanism of reaction in 1,3-1,4-β-endoglucanase with the 2-deoxy-2-fluoro substrate derivative.

Introduction.

The mechanism of hydrolysis of glycosides has been historically investigated from experimental[167, 168] and theoretical[169-175] perspectives. This interest is not surprising, since polysaccharide structures play a role in several biological systems and processes, from energy storage polymers to posttranslational modification of proteins to name just a few. However, glycoside chemistry is extremely difficult due to the great stability of the glycosidic bond and the structural complexity of carbohydrates (*e.g.* a hexasaccharide has 10^{12} possible isomers).

In solution, the rate-limiting step of the acid-catalyzed hydrolysis of the glycosidic bond is the cleavage of the C-O bond, and proceeds via a transition state with a high oxocarbenium ion-like character in which the anomeric carbon adopts an sp^2 hybridization and in which there is little participation of the attacking nucleophile. The activation barrier has been estimated to be of around 30 kcal·mol^{-1} by means of Arrhenius plot analysis.[9, 176]

Experimental[167, 169, 170] and theoretical[171-174, 177] studies revealed that the standard acid-catalyzed solvolysis reaction proceeds via a dissociative transition state of a D_N*A_N reaction mechanism (see the side panel below) with a very short lived[178] ($\approx$ 1-3 ps) oxocarbenium ion-like transition state.

However, it is known that the type of mechanism depends on the nature of the sugar and the aglycon:[162, 179-182] Experiments performed with substrates with bad leaving groups[183] or with good nucleophiles,[179] gave a higher k_{13_C}/k_{14_C} ratio (see the *nucleophilic substitution mechanisms* side panel) and were interpreted as the reaction having more S_N2 character. Instead, other substrates with very good leaving groups[181] lead to reactions with a stronger S_N1 character. On the same line, stabilization of the charge developed on the anomeric carbon by using a 2-deoxy-glycoside[162, 182] or a methyl 5-thioxylopyranoside[183] as a substrate favored a more $D_N + A_N$ like dissociative mechanism.

Similar parameters are used to describe the enzyme-catalyzed reaction by glycoside hydrolases (GHs). Experimental KIE studies[170] on almond β-glucosidase revealed that the reaction proceeds via a dissociative D_N*A_N mechanism (k_{13_C}/k_{14_C} ratio of $\approx$ 1.03) with a $^4H_3/E_3$ distortion of the -1 subsite saccharide in the transition state. However, in yeast α-glucosidase the mechanism was more concerted (A_ND_N) with a k_{13_C}/k_{14_C} ratio of $\approx$ 1.01. Similarly as observed in solution, the same enzyme yeast α-glucosidase fed with a glucosyl pyridinium salt (with a better leaving group) resulted in a more dissociative D_N*A_N mechanism.

Side panel: Nucleophilic substitution mechanisms

According to the IUPAC nomenclature[184, 185] (Figure V-1), the transition state of the nucleophilic substitution can go from a fully "solvent loose ion pair" species named as $D_N + A_N$ (or unimolecular nucleophilic substitution S_N1) to a totally concerted mechanism named as A_ND_N (known as bimolecular nucleophilic substitution S_N2). In between there is a continuous degree of charge separated species such as the "tight ion pair" transition state involved in the D_N*A_N mechanism.

$$A_ND_N \qquad D_N*A_N \qquad D_N + A_N$$

Less dissociative   More dissociative

Figure V - 1: Nucleophilic substitution mechanisms according to the IUPAC nomenclature.

One of the most used techniques to unambiguously study the nature of the transition state in this reaction was the ^2H and ^{13}C kinetic isotope effects on the α hydrogen and on anomeric carbon respectively. A high k_H/k_D ratio is indicative of a planar geometry around the anomeric carbon that accounts for its sp^2 hybridization and hence for a high oxocarbenium ion-like character of the transition state. On the other hand, a low k_{13_C}/k_{14_C} ratio on the anomeric carbon (1.00 to 1.01) is indicative of a dissociative mechanism, (close to a $D_N + A_N$) while a k_{13_C}/k_{14_C} ratio of 1.03 to 1.06 is indicative of a more concerted transition state (A_ND_N mechanism).[168]

Previous studies of our group on the hydrolysis reaction in a family GH8 retaining endocellulase[26] and a family GH38 retaining α-mannosidase[28] concluded that hydrolysis occurred after protonation of the glycosidic bond by the catalytic acid residue and that the mechanism was also dissociative (D_N*A_N) with a boat distortion on the transition state ($^{2,5}B$ and $B_{2,5}$ respectively).

However, the mechanism of reaction of 1,3-1,4-β-endoglucanase[1] takes place via a dissociative type of S_N2 reaction (D_N*A_N following the recommended IUPAC nomenclature[184, 185]) with little assistance from the general acid catalyst in the TS. The conformational itinerary of subsite -1 sugar ring was analyzed during the reaction in terms of its puckering coordinates and a $^1S_3/^{1,4}B$ (Reactants) - $^4H_3/^4H_5$ (TS) - $^4C_1$ (Products) pathway was found (figure V-2).
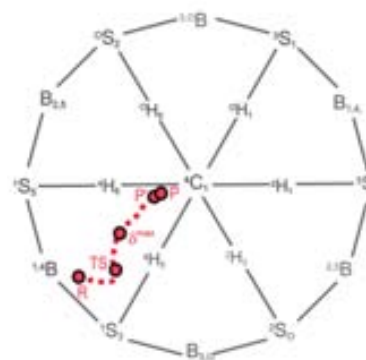

Figure V - 2: Conformational itinerary in the case of a substrate with an unmodified glucose ring at -1 subsite. Figure taken from reference [1].

In the previous Chapter we have shown that enzyme-ligand interactions govern conformation of the substrate in the Michaelis complex of 1,3-1,4-β-endoglucanase from *Bacillus*. In particular, the hydrogen bond between the exocyclic group at

C2 and the nucleophilic residue is necessary to maintain the $^1S_3$ distortion that preactivates the substrate for catalysis. Therefore, the Michaelis complex with a 2-deoxy-2-fluoro substrate showed a non-distorted $^4C_1$ conformation.

In view of the importance of this interaction in the distortion of the substrate, it is likely that the 2-deoxy-2-fluoro modification changes the conformational itinerary of the substrate during catalysis. In this Chapter, we extend the previous studies of the properties of the mechanism of β-glucanases by analyzing the reaction with the 2-deoxy-2-fluoro glucose inhibitor.

Computational details.

CPMD/MM simulations

The system was simulated at constant pressure and temperature conditions (0.1 MPa and 300 K) using a Nosé-Hoover chain thermostats[158] as in the previous calculations. The QM system was enclosed on an orthorhombic supercell of dimensions 34.0 x 28.0 x 31.0 a.u. and consisted of 67 atoms that correspond to the substrate subsites -1 and +1 (truncated at the anomeric carbon of subsite -2 by a monovalent carbon-like pseudopotential), the catalytic residues Glu105, Asp107 and Glu109 (glutamates truncated at the beta carbon and the aspartate truncated at the alpha carbon by a monovalent pseudopotential), and a close water molecule; as shown by the atoms represented in licorice in the figure V-3.
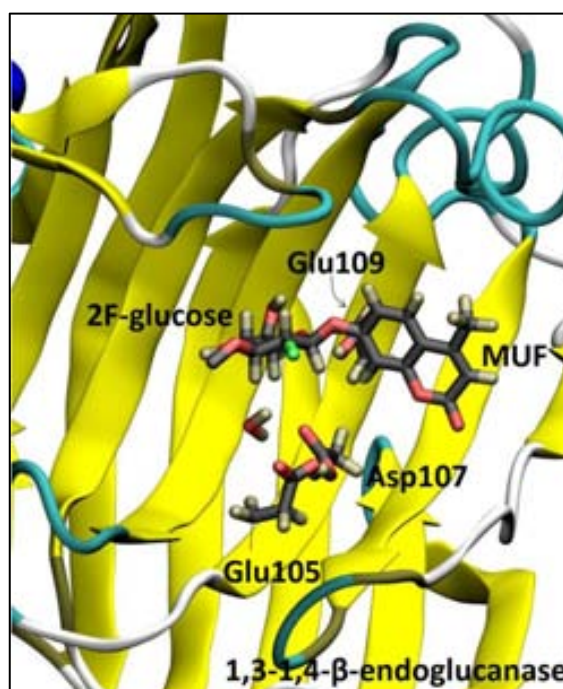

Figure V - 3: Active site of 1,3-1,4-β-endoglucanase in complex with 2-deoxy-2-fluoro substrate derivative. The QM atoms are represented in licorice.

The MM part of the system consisted in the rest of the substrate and the protein, and the solvent molecules and counterions. The classical box size was of 50.4 x 61.0 x 48.8 Å. The starting structure for the metadynamics simulation was obtained from the previous simulated Michaelis complex of this system (simulation 2 in Chapter IV). The first 0.5 ps of the metadynamics simulation were considered as equilibration time and the height of the repulsive Gaussian potentials was set to zero.

All calculations were performed within spin-restrained Kohn-Sham density functional theory in its plane wave formulation[110], with a kinetic energy cutoff of 70 Ry. The PBE exchange-correlation functional was chosen and the core electrons were taken into account using Martin-Troullier pseudopotentials.[114] The timestep for the MD was set to 5 a.u. (≈0.12 fs) and the fictitious mass
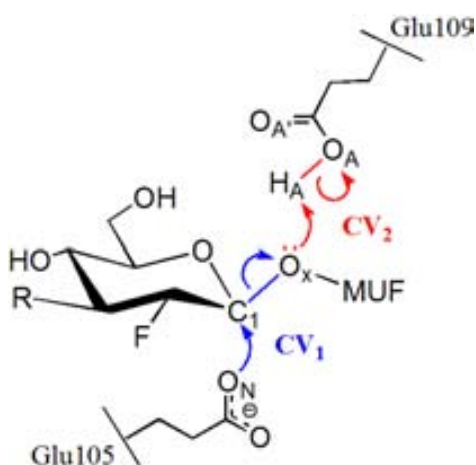
for the electronic degrees of freedom was set to 850 a.u. The quantum hydrogen masses were substituted by deuterium masses to enhance the adiabaticity of the system.

Metadynamics

The metadynamics approach as developed by Laio and Parrinello,[125] in its extended Lagrangian version[131] was used to accelerate the reaction of the glycosylation step of the enzymatic hydrolysis reaction.

The collective variables (CVs) used were taken as a linear combination of the coordination numbers between the atoms involved in the covalent bonds that are formed and cleaved during the hydrolysis reaction. Since there are up to four bonds being formed/cleaved, a bidimensional (two CVs) metadynamics was called for. CV_1 accounts for the nucleophilic attack by glutamate 105 on the anomeric carbon and the departure of the MUF aglycon (see figure V-4). CV_2 activates the hydrogen transfer from the general acid/base residue glutamate 109 to the glycosidic oxygen.

Each CV is the difference in coordination numbers (CN_i) defined as:[131]

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

Where d_{ij} is distance between the nuclei i and j, d_0 is the threshold bonding distance, and p and q are exponents that determine the steepness of CN_{ij} decay with respect to d_{ij}. Therefore, CN_{ij} values range from 0 (not bonded) to 1 (bonded).

Figure V - 4: Collective variables used during the metadynamics simulations. CV1 in blue, CV2 in red.

The selected p, q and d_0 parameters were chosen to effectively separate all the important species along the reaction path, and are specified in table V-1:

In this way, both CVs take negative values in the Michaelis complex configuration (reactants), and positive values in the covalent intermediate state (products). Transient configurations halfway between the bonds formation and cleavage will be reflected with CV values around 0.

Table V - 1: Definition and parameters of the CVs used in the metadynamics simulations. Atoms are labeled as in figure V-4.

CV	Definition	Parameters	
1	$CV_1 = CN_{C1-ON} - CN_{C1-Ox}$	d_0	3.03 a.u.
		p	12
		q	2
2	$CV_1 = CN_{HA-Ox} - CN_{HA-OA}$	d_0	2.08 a.u.
		p	12
		q	2

The selected parameters for the mass of the fictitious particles (MCV) and for the fictitious spring constant (KCV) needed for the Lagrangian formulation of the metadynamics were of 10 and 1.3 respectively for CV_1 and to 5 and 1.4 for CV_2 (in atomic units). The height of the Gaussian used was of 0.003 a.u. ($\approx$ 1.88 kcal·mol^{-1}) in the beginning of the simulation (the firsts 14.3 ps) and was then reduced to 0.0015 ($\approx$ 0.94 kcal·mol^{-1}) for a fine determination of the transition state energy. Similarly, the width of the repulsive Gaussian potentials was set to 0.15 and 0.075 respectively (in units of coordination number). The Gaussian deposition time τ was set to 200 MD steps (24 fs). The CVs phase space boundaries were delineated by repulsive potentials of 1.5 Hartrees located at -0.89 and +0.89 for CV_1 and at -0.95 and +0.9 for CV_2.

In figure V-5 the evolution of the collective variables along the metadynamics simulation is shown. In the reactants state both CVs' value are of -1, while in the products state they are close to 1. The simulation was stopped before the products well was completely filled (the total simulation time was of 23.3 ps and 969 Gaussian hills had been added). As a result, we cannot quantify the energy difference between reactants and products. However, since the metadynamics simulation naturally follows the lowest free energy path, the mechanistic features of the FES are a good approximation to the reaction pathway. In the products state, the proton H_A is bonded to the MUF aglycon (O_G). Going from the products to the reactants, it is transferred to glutamate 109. However, chemically, both oxygen atoms from the carboxylate (O_A and $O_{A'}$ in figure V-4) are equivalent, but CV_2 discriminate between them. In consequence, a new minimum located at $CV_1 \approx$ -1 and $CV_2 \approx$ 0 was filled.
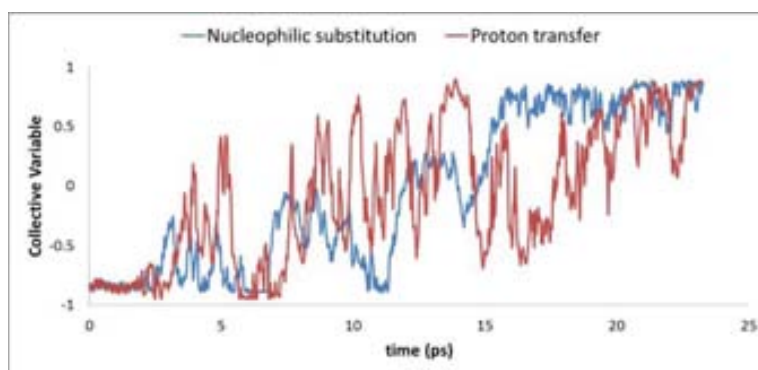

Figure V - 5: Evolution of the CVs along the metadynamics simulation time. CV1 is in blue and CV2 in red.

Results and discussion

The free energy surface (FES) of the hydrolysis reaction, and is shown in figure V-6. The most stable minimum in the FES corresponds to the Michaelis complex well (R), located in the lower-left corner. The products well (P) is in the upper right corner, and corresponds to the covalent glycosyl-enzyme intermediate. Below the products minimum, there are two more minima with the same value of CV_1 but with different values of CV_2 (P' and P''). In all these structures, the covalent intermediate is formed, but the H_A proton jumps from the leaving group to the glutamate residue and *vice versa*. This is indicative that the pKa of the MUF aglycon and the glutamate residue in the protein environment is similar and was also observed in the simulation of the hydrolysis reaction with a standard 2-hydroxy-glucose at the -1 subsite.[1]

Following the lowest free energy path that connects these two minima (the dotted line in figure V-6) there are three different species that correspond to two saddle points and a reaction intermediate: TS1, TS2 and IR respectively (displayed in figure V-7), with a free energy barrier of ≈ 118 kcal·mol[-1]. The activation barrier of the reaction is much higher compared to the results obtained for the natural substrate[1] on the reaction with the standard glucose moiety at subsite -1 (32 kcal·mol[-1]), consistent with the fact that the 2-deoxy-2-fluoro substrate derivative is an inhibitor.
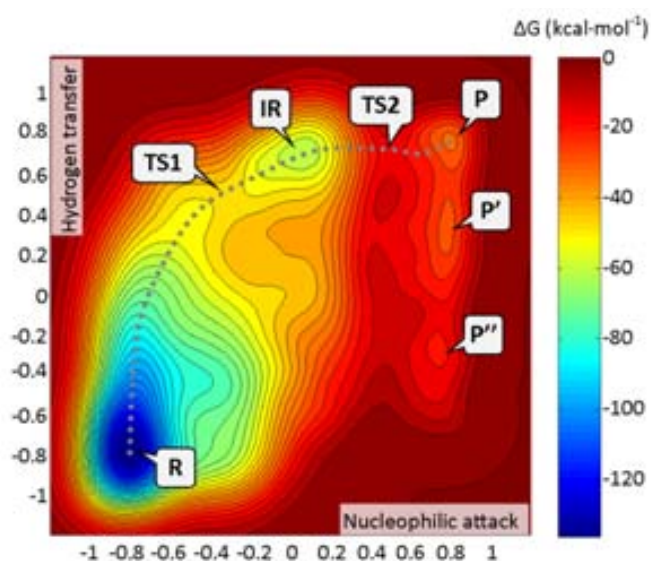

Figure V - 6: Free energy surface for the hydrolysis reaction of the 2-deoxy-2-fluoro substrate derivative. Contour lines represents 4 kcal·mol[-1]. The dotted line represents the minimum energy path.

Snapshots of the most relevant structures along the minimum free energy reaction path are shown in figure V-7. The reaction starts with the proton transfer from glutamate 109 to the glycosidic bond (see the red and the orange lines in figure V-8), together with a significant elongation of the C_1-O_X distance.

Following the minimum free energy reaction path, the first saddle point (TS1) corresponds to the cleavage of the glycosidic bond (the C_1-O_X distance is of 2.27 Å). At this point the acid proton has completely been transferred to the glycosidic bond (the H_A-O_X distance is of 0.97 Å) and hence the structure corresponds to an oxonium ion and the leaving group is an alcohol.

In the TS1 structure, glutamate 105 is still far from the anomeric carbon (the C_1-O_N distance is still of 3.98 Å) and the nucleophilic attack is not possible. Therefore, following the glycosidic bond cleavage is the formation of an oxocarbenium ion species. In this reaction intermediate, the anomeric carbon adopts a sp[2] hibridization and it forms a partial double bond with the pyranic oxygen as shown by a shrink in the distance between them (1.42 Å in the reactants and 1.25 Å in

the intermediate). This intermediate has a certain lifetime to allow it to migrate from the leaving group to approach the glutamate 105 (the position of the nucleophilic residue does not change). The next step in the reaction path is TS2, which corresponds to the formation of the glycosyl-enzyme covalent bond (the C_1-O_N distance is of 1.93 Å). At this point the anomeric carbon changes again its orbital hybridization to the usual tetrahedral sp^3 stereoelectronic structure. The last snapshot is the products state.
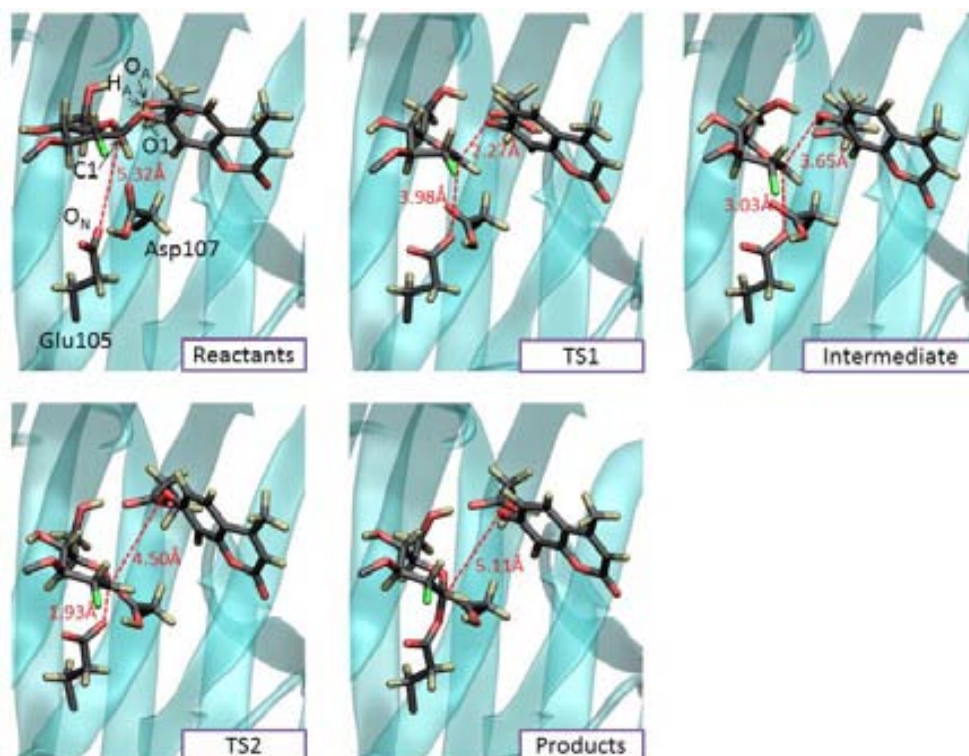
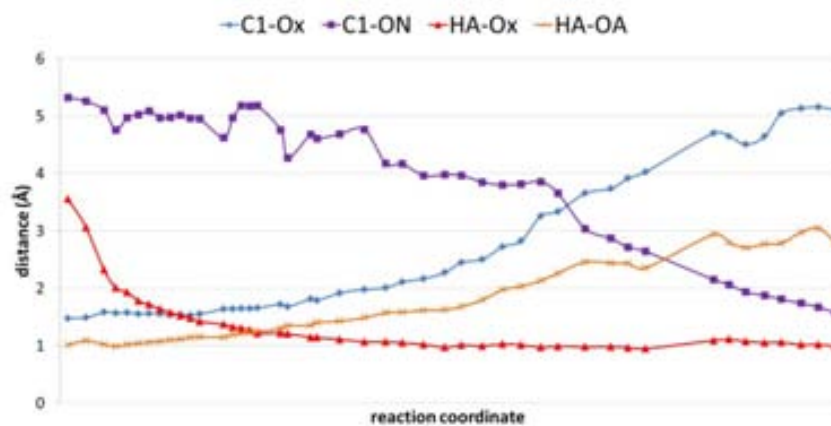**Figure V - 7: Snapshots along the minimum free energy reaction path.**

Figure V - 8: Main distances measured along the minimum free energy path. The abscissa axis (the reaction coordinate) is minus the distance to the products state in units of coordinate number.

Analysis of the subsite -1 ring distortion was performed along the reaction path. Surprisingly, the conformational itinerary found was to be $^4C_1$ (Reactants) - $^4H_3/E_3$ (TS and Reaction Intermediate) - $^4C_1$ (Products) as shown in figures V-9 and 10. This circular conformational itinerary is significantly different from the linear one found by Biarnés *et al.* with the usual glucose at subsite -1 (see figure V-2).

It is interesting that the reaction itinerary starts from the non-distorted $^4C_1$ conformation (the same structure found in the modeling of the Michaelis complex chapter IV) and it goes to a $^4H_3$ and ^4E distorted conformation in the transition state without passing through an equatorial conformation such as $^1S_3$ or $^{1,4}B$. This implies that the glycosidic bond is cleft starting from a $^4C_1$ chair and not a $^1S_3$ conformation, and consequently, the pre-activation effects of the induced-fit distortion of the substrate are not applicable on the reaction with the inhibitor.
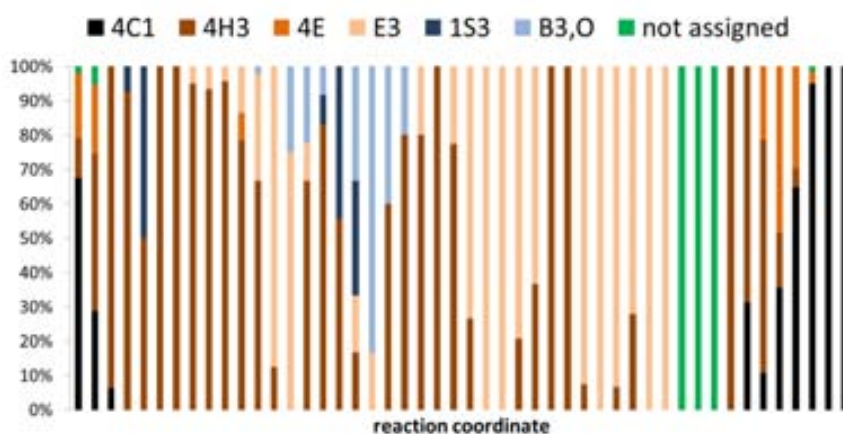

Figure V - 9: Histogram analysis of the puckering coordinates. In black is the $^4C_1$ conformation. Those conformations that have the C5-O5-C1-C2 dihedral planar ($^4H_3$, ^4E and E_3) are in brownish color. The conformations on the equatorial ring $^1S_3$ and $B_{3,O}$ (those found in the Michaelis complex with the natural substrate are in dark and light blue respectively. In green are represented those points for which the sampling was not enough to assign a conformation.

As a consequence of the $^4C_1$ conformation, the glycosidic bond does not adopt an axial conformation and the hydrogen H1 is not in a semi-equatorial conformation. Instead, H1 is pointing right down to the nucleophilic residue and there is a high steric hindrance for the nucleophilic attack. This explains that the distance between C_1 and O_{Nuc} atoms in the MC structure is 5.32 Å.

Moreover, the ALPH theory (see the introduction Chapter) does not apply in the $^4C_1$ conformation. That is, there cannot be a charge transfer between the free electron pair of the pyranic oxygen with the antibonding σ* orbital of the glycosidic bond because the orbital overlap is structurally not possible. As a result, the glycosidic bond is not polarized and there is not a partial double bond formation between C_1 and O5 atoms.

Noteworthy, the collective variables chosen do not impose any conformational itinerary. Hence, the system could have undergone any other energetically favorable mechanism. For instance, the subsite -1 could approach the nucleophilic residue without breaking the glycosidic bond, changing

its conformation from $^4C_1$ to $^1S_3$ and starting the chemical step from this preactivated state, following a longitudinal $^1S_3 \rightarrow {}^4H_3 \rightarrow {}^4C_1$ conformational itinerary. Instead, the circular itinerary has the advantage that the oxocarbenium ion ($^4H_3$) displays the fluorine atom in an equatorial orientation (in contrast to the axial orientation displayed by the $^4C_1$ conformation). Therefore, the steric hindrance between C2(F) and residue 105 vanishes and the nucleophilic attack is easier.
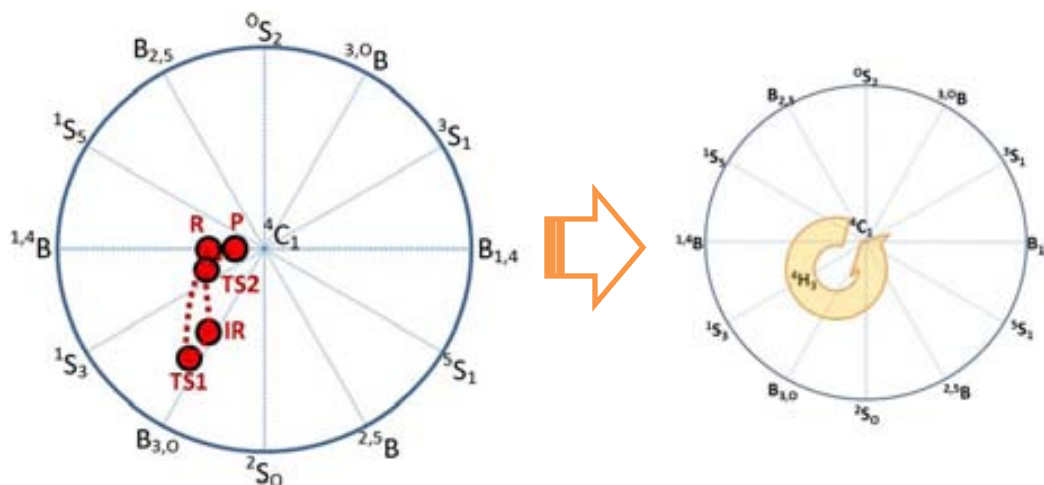

Figure V - 10: Puckering coordinates of the main reaction species projected on the Stoddart's diagram: R → TS1 → IR → TS2 → P. The conformational itinerary is circular.

All these factors place the Michaelis complex much further (geometrically and energetically) from the transition state than compared to the MC with the natural substrate. As a consequence, the activation energy increases.

This has to be added to the fact that i) the hydrogen bond between the hydroxyl substituent at C2 in the natural substrate with the nucleophilic residue is not present in the 2-deoxy-2-fluoro inhibitor and ii) the greater inductive effect of the fluoride substituent with respect to the hydroxyl group. The sum of all these terms greatly increases the activation barrier.

Nevertheless, it is noteworthy that these effects not only modify the activation energy but also the reaction mechanism itself. In the study by Biarnés *et al.* on the glycosylation step with a glucose moiety in subsite -1,[1] we found that the reaction is a very dissociative S_N2 type of nucleophilic substitution with very little protonation of the leaving group (a MUF aglycon). Although the present simulation is performed with the same leaving group (MUF), we observed that the glycosidic bond needs protonation prior to its cleavage. We also observed the presence of a short lived oxocarbenium intermediate. This intermediate is needed due to the large distance between the anomeric carbon and the nucleophilic residue when the bond glycosidic is cleaved. Consequently, the mechanism resembles more a unimolecular nucleophilic substitution (S_N1).

Of course, if the reaction follows a S_N1 mechanism, the slow step is the cleavage of the glycosidic bond to yield the ion pair intermediate. Therefore, the reaction rate will depend mostly on the

nucleofugicity of the aglycon (compared to the S_N2 mechanism in which the nucleophilicity of the attacking group is the most important). This is in agreement with the fact that the reaction rate can be accelerated by using a much better leaving group such as 2,4-dinitrophenol, to reach the covalent glycosyl-enzyme intermediate (even with a fluoro-inhibitor).

The mechanism found is in agreement with the results obtained in the study of the Michaelis complex (chapter IV). The subsite -1 pyranose ring of the 2-deoxy-2-fluoro substrate derivative adopts a non-distorted $^4C_1$ chair conformation. This places the anomeric carbon at a large distance from the nucleophilic residue and precludes a S_N2 type of reaction. In this sense, the reaction can only be described by a S_N1 mechanism, and only when the glycosidic bond is labile, that is, when a good leaving group is present.

Conclusions

1. The reaction mechanism of 1,3-1,4-β-endoglucanase with a 2-deoxy-2-fluoro glycoside follows a very dissociative $D_N + A_N$ type of nucleophilic substitution. Most of the activation energy is invested to cleave the covalent glycosidic bond and the formation of a short lived tight ion-pair intermediate.

2. Protonation of the glycosidic bond occurs necessarily prior to the cleavage of the glycosidic bond, in contrast to what was observed with a standard glucose in subsite -1. However, in the products state, the acidic proton can be found indistinctly in the MUF aglycon or the glutamate 109 residue.

3. The conformational itinerary of the subsite -1 saccharide ring is circular. It starts and ends in a non-distorted $^4C_1$ chair conformation, while the transition state is a $^4H_3/E_3$ with a planar geometry on the anomeric carbon.

4. The fact that the Michaelis complex adopts a non-distorted $^4C_1$ conformation, places the reactants very far from the nucleophilic residue and the substrate is not preactivated. Therefore, the reaction is inhibited.

Chapter VI - The conformational free energy landscape of β-D-mannopyranose and α-L-fucopyranose

Ardèvol, A.; Biarnés, X.; Planas, A.; Rovira, C.

van Bueren, A.L.; Ardèvol, A.; Fayes-Kerr, J.; Luo, B.; Zhang, Y.; Sollogoub, M.; Blériot, Y.; Rovira, C.; Davies, G.J.

The conformational free energy landscape of β-ᴅ-mannopyranose and α-ʟ-fucopyranose.

Introduction

In the previous chapters, we have studied the substrate distortion in the Michaelis complex and during the catalytic itinerary in glycoside hydrolases. We focused on the identification of the main E·S interactions that govern ring distortion and how perturbation of these interactions may affect the substrate conformation.

Previous theoretical studies have demonstrated that the distorted conformation is both structurally (elongation of the leaving-group distance) and electronically (charge increase at the anomeric center) similar to the oxocarbenium ion-like TS of the GH hydrolysis reaction[34, 186] (scheme VI-1). Therefore, the distorted conformation of the substrate in the Michaelis complex (MC) is "on the path" towards the TS and gives information about the catalytic itinerary. This is of relevant importance when designing TS-analogue specific inhibitors for GHs, which are known to be far more potent than substrate-analogue inhibitors.[187]
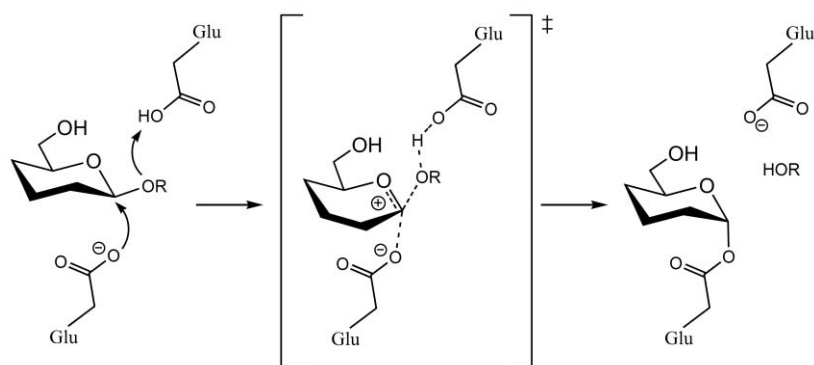
Scheme VI - 1: Glycosylation reaction mechanism in retaining GHs.

Understanding of *how* the distortion is acquired and *what* are the effects of this distortion on the properties of the substrate is interesting for design of novel GHs inhibitors, protein engineering, etc. In this chapter, the question is moved to the understanding of *why* the enzyme *decides to* stabilize one particular ring conformation in front of another.

Given the large flexibility of pyranosides, it might be expected that of different GHs would favor a wide variety of different sugar conformations. Instead, all MC are encountered in a restricted number of conformations. In addition, enzymes acting on the same substrate (*e.g.* β-mannose) tend to stabilize the same ring distortion and to use the same conformational itinerary (see the chapter I).

To try to explain this, we previously[35] applied CPMD metadynamics to obtain the free-energy surface (FES) of β-D-glucopyranose with respect to all of the conformations in the Stoddart diagram (scheme VI-2). It was shown that the FES of the monosaccharide correlate well with the conformations observed in ligand-enzyme complexes (*i.e.* the experimental X-ray structures of MC analogues are located in the most stable distorted minima found on the calculated FES).

Our study suggests that the glucosidases has evolved to induce the substrate to adopt a distortion that is energetically most accessible. As a corollary, the FES has a predictive value to address the substrate distortion on the Michaelis complex, and hence the complete conformational itinerary during catalysis.

In this Chapter, we have applied the same methodology to obtain the FES of β-D-mannopyranose and α-L-fucopyranose saccharides. We have analyzed the structural and electronic rearrangements with respect to ring conformation. In addition, we have introduced a simple index that measures the suitability of a specific conformation to be the Michaelis complex and investigate its relation to the conformations observed in the X-ray structures of E·S complexes of β-D-mannosidases and α-L-fucosidases.
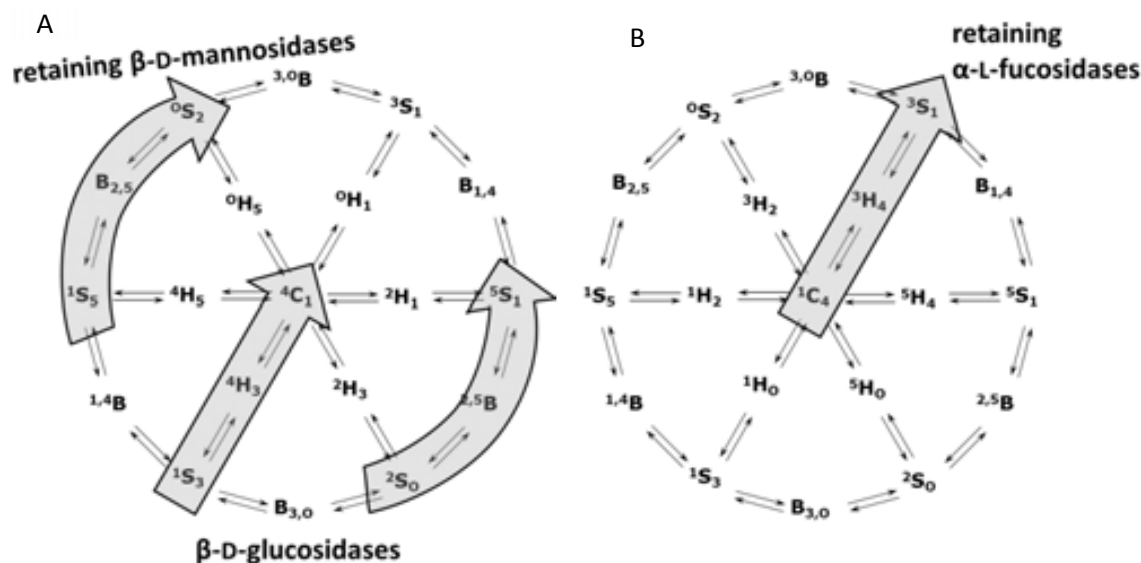
β-D-mannose

Enzymatic mannoside hydrolysis is of paramount importance for its role in the regulation of the high-mannose and complex type glycan post-translational modifications involved in cell biology. Its malfunction can lead to a variety of diseases, including cancer.[188] This has stimulated the field of mannoside synthesis, which faces considerable challenges, mainly in the case of the β-anomers. Potential steric clashes, unsympathetic anomeric effects, and participation of the C2 neighboring group make β-anomers kinetically and thermodynamically unfavorable.[189] Therefore, parallel to smart synthetic strategies that have been applied,[190, 191] mannosidase and mannosyltransferase biochemistry is blooming.[56, 192]

In 2002, it was first proposed that enzymatic mannoside hydrolysis uses an unusual substrate conformational pathway.[19] For retaining β-mannosidases, structural studies of families GH2 and GH26 have revealed that the β-mannosyl substrate adopts a $^{1}S_5$ conformation in the Michaelis complex and an $^{O}S_2$ conformation in the covalent intermediate of the hydrolysis reaction.[40, 193] In addition, structural studies of GH5 inhibitors have suggested a $B^{2,5}$ type distortion at the TS.[194] This is further supported by a recent study by Tailford *et al.*[56] reporting linear free-energy relationships (LFERs) and crystal structures of TS analogues. Therefore, the proposed conformational itinerary of β-D-mannosidases is $^{1}S_5 \rightarrow B_{2,5} \rightarrow {}^{O}S_2$ (as shown using the Stoddart's diagram[15] in scheme VI-2).

Detailed information concerning the energetic/structural/electronic relations of mannose is necessary for the understanding of mannosidase mechanisms and to rationalize the unfavorable effects involved in β-mannoside organic synthesis. Several theoretical studies have aimed at quantifying the relative energies of different mannopyranose ring conformations using force-field based approaches.[195-197] It is worth mentioning the study of Dowd, French, and Reilly,[195] who

computed potential energy surfaces of 16 different aldopyranose rings, including α- and β-D-mannopyranose; their work has served as a reference for subsequent studies.[35, 198]
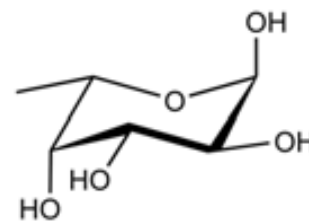

Scheme VI - 2: Stoddart diagram as the projection of the Cremer and Pople sphere (A = northern hemisphere, B = southern hemisphere) onto the equatorial plane. The proposed reaction conformational itineraries for retaining β-D-mannosidases, β-D-glucosidases and α-L-fucosidases are highlighted.

Electronic rearrangements upon ring distortion, however, can be analyzed only by going beyond the force-field approximation. In recent years, a number of studies of sugar ring conformations using density functional theory (DFT) have appeared (*e.g.* see refs [199-206]), including evaluations of different density functionals[206] and basis sets.[205] However, because of the complexity of the conformational phase space, these studies have typically been limited to a set of conformations that are stable upon geometry optimization (*e.g.* boats, skews, and chairs), and energy/structure/atomic charges have not been investigated in relation to experimental data for GHs.

α-L-fucose

Human glycans including N-linked glycans, ABO blood group and Lewis antigens, are decorated with terminal α-L-fucose residues (scheme VI-3). Moreover, fucosylated glycoconjugates can be encountered in a variety of tissues, including liver, brain and spleen, among others. In man, the dynamic process of α-L-fucoside removal is catalyzed by two α-L-fucosidases (FucA1 and FucA2) which liberate α-L-fucose as a product.[207, 208] These enzymes are important in regulating bacterial

Scheme VI - 3: α-L-fucose in the $^1C_4$ conformation.

infections,[207] and a deficiency in these enzymes leads to fucosidosis, a debilitating neurodegenerative lysosomal storage disorder.[209, 210] Some cancers are identified by increased fucosylation levels on their surface, which promotes growth and metastasis of cancerous cells.[209, 210] Increased levels of α-L-fucosidase expression have been important in cancer prognosis[211, 212]

and are a potential target for cancer therapy.[213] α-L-fucosidases are found exclusively in family GH29 according to the CAZy database (www.cazy.org),[7] and they follow the usual double displacement mechanism of retaining glycoside hydrolases.[214-218] Understanding the details of the catalytic mechanism of α-L-fucosidases is central to an understanding of their function.

In 2004, the crystal structure of a covalent intermediate of α-L-fucosidase from *T. maritima* was solved with a 2-deoxy-2-fluoro fucose inhibitor, which was found to adopt a $^3S_1$ conformation (PDB code 1HL9).[61] More recently, a series of snapshots of the reaction coordinate of fucosidase Bt2970 from *B. thetaiotaomicron* (that shows high homology to human enzymes) were described.[62] Concretely:

a) The Michaelis complex was obtained by crystallizing a *p*-nitrophenyl α-L-fucoside substrate in an inactive form of the enzyme (the nucleophilic residue Asp299 was mutated to asparagine). In the MC, the fucoside moiety adopted a $^1C_4$ conformation (PDB code 2WVU).

b) The structure of the covalent intermediate (CI) was facilitated through the use of the 2-deoxy-2-fluoro fucosyl fluoride in complex with an acid/base mutant (E288Q) of the α-L-fucosidase. In the CI, the conformation of the fucose ring was $^3S_1$ (PDB code 2WVS)

c) The wild type enzyme in complex with an α-L-fucose like iminosugar that mimics the E·S complex showed a $^1C_4$ pyranose ring conformation (PDB code 2WVT).

Together, these structures provided structural evidence for a "latitudinal" Southern hemisphere $^1C_4 \rightarrow {}^3H_4 \rightarrow {}^3S_1$ pathway for the terminal α-L-fucoside hydrolysis by this family of enzymes (glycosidase pathways have recently been reviewed[25, 40, 41]).

Computational details

Metadynamics simulations

To enhance the conformational phase space exploration of the simulated system (*i.e.*, the β-D-mannopyranose and the α-L-fucopyranose monosaccharides) we used the metadynamics technique[130, 219] (see the introduction Chapter for a detailed description of the method).

The collective variables used to obtain the conformational FES of the sugar molecules were defined as a variation of the Cremer and Pople puckering coordinates[159] in their polar formulation (Q, θ, and φ, see figure VI-1 and the appendix). Concretely, two Cartesian coordinates (q_x and q_y in figure VI-2) that correspond to the projection of the Q, θ, and φ onto the equatorial plane as:

$$\begin{cases} q_x = Q \sin\theta \sin\phi \\ q_y = Q \sin\theta \cos\phi \end{cases} \qquad \text{Eq. VI - 1}$$

These collective variables are explicit function of the atomic positions, and they fulfill the requirements for suitable CVs in the metadynamics procedure, as described by Laio and Gervasio:[125]

    i.   They clearly distinguish the different states of the system (*i.e.*, all of the conformers in the Stoddart diagram).

   ii.   They do describe the main slow modes that are relevant to the process of interest.

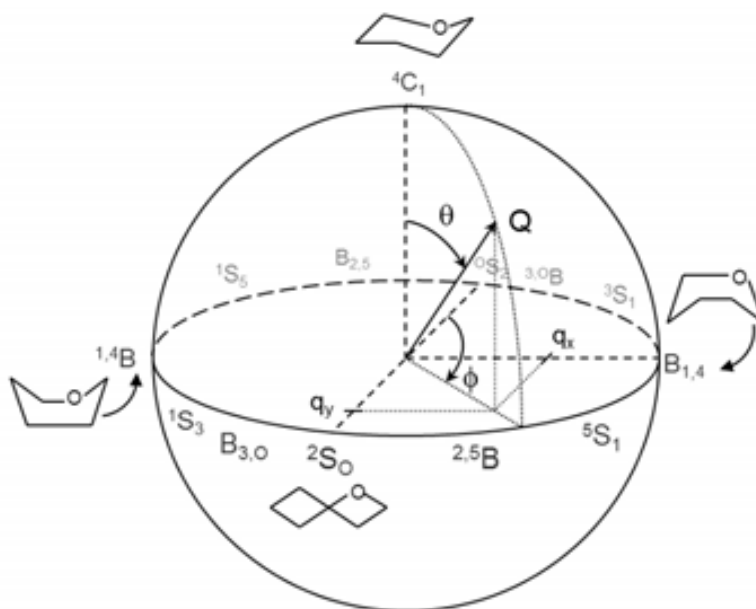 iii.   They result in a bidimensional FES that is affordable with reasonable computational resources.

Figure VI - 1: Cremer and Pople puckering coordinates Q, θ, and φ for a six membered ring and the collective variables q_x and q_y used in the metadynamics run.

The fast variation of Q ensures that the metadynamics simulation explores all possible Q values for each conformer in the FES (see figure VI-2) and that their contributions to the distortion entropy are taken into account.

The two-dimensional projection (q_x, q_y) of the spherical representation of the pyranose ring conformational space (Q, θ, φ) facilitates the computation and visualization of the computed FES in a continuous space. In this way, itineraries among any three consecutive boat and skew-boat conformations can easily be drawn, as is done in structural and mechanistic studies of GHs.[40, 56]
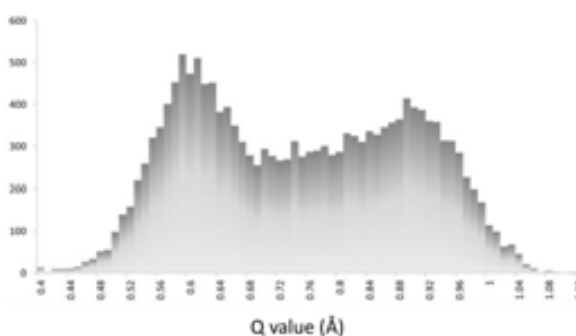

Figure VI - 2: Distribution of Q values sampled during the metadynamics simulation. The peak at Q ≈ 0.6 corresponds to the $^4C_1$ conformation, whereas the other conformers show a higher Q value.

In the study of the conformational FES of β-D-mannopyranose, we focused on the Northern hemisphere of the Cremer and Pople sphere (figure VI-2) because the substrate distortions present in all of the currently available structures of MC of β-mannosidases are found in this region.[40] In addition, the conformational itineraries of the mannosyl moiety during the hydrolysis reaction in either retaining or inverting β-mannosidases are expected to be limited to this region of the conformational space. Therefore, the simulation was started from the most stable $^4C_1$ conformation.

Instead, in the case of the α-L-fucopyranose we focused on the Southern hemisphere, because the most stable conformation is the $^1C_4$ and the structural studies on the MC of α-fucosidases show that the substrate sugar ring adopts this chair conformation. Therefore, the metadynamics simulation was started from the $^1C_4$ conformation.

In both cases, analysis of the projection of the obtained trajectory on the puckering sphere-like volume (figure VI-2) confirmed that only one hemisphere (the Northern for the β-mannose and the Southern for the α-fucose) was sampled.

It should be noted that the use of the coordinates q_x and q_y makes it very unlikely that the system could sample more than one hemisphere, as transitions from north to south (*e.g.* $^4C_1$ to $^1C_4$) and *vice versa* are not activated, and it is expected that there is a sizable barrier for this transition.

The *ab initio* simulations were done within the Car-Parrinello extended Lagrangian formalism.[131] A fictitious electron mass of 850 au and a time step of 0.12 fs were used. The system temperature was set to 300 K by coupling it to a thermostat using the Nosé algorithm.[220] The Kohn-Sham orbitals were expanded in a plane-wave (PW) basis set with a kinetic energy cutoff of 70 Ry. *Ab initio* pseudopotentials generated within the Troullier-Martins scheme were employed.[114] The calculations were performed using the Perdew-Burke-Ernzerhoff (PBE) generalized gradient-corrected approximation,[221] as previously used in Car-Parrinello simulations of isolated carbohydrates and GHs.[34, 35]

The β-D-mannopyranose system consists in 24 atoms enclosed in an orthorhombic box with dimensions 13.0 Å × 14.0 Å × 12.0 Å. The α-L-fucopyranose system lacks the hydroxyl group at C6 and therefore it consists in 23 atoms in a 13.0 Å × 12.5 Å × 12.5 Å isolated box. For both systems, about 6 ps of free dynamics at 300 K were performed prior to the metadynamics simulations.

Values of 10.0 a.u. for the mass of the fictitious particles and 0.2 a.u. for the force constant were used in the metadynamics simulations.

For the β-mannose system, Gaussian like functions with a width (*δs*) of 0.15 a.u. and a height (*w*) of 0.31 kcal·mol^{-1} were used with a deposition time of 12 fs during the firsts 96 ps of simulation. After that, the height (*w*) and the Gaussian deposition time (*τ*) were set to 0.06 kcal·mol^{-1} and 18 fs respectively, for a better convergence of the FES. A total of 13,369 Gaussians were added, which corresponds to ≈ 190 ps.

The parameters for the α-fucose system were set to: δs = 0.1 a.u., w = 0.19 kcal·mol^{-1} and τ = 36 fs (300 MD steps). The simulation was stopped after 9,613 Gaussians had been added ($\approx$ 347 ps of simulation time) and a satisfactory converged FES was reconstructed.

Results and discussion

β-D-mannose

Conformational free energy surface

The computed FES of β-D-mannopyranose as a function of the coordinates q_x and q_y is shown in figure VI-3. It contains seven local minima, which are labeled as **M1** to **M7** according to their relative stability (table IV-2). As expected, the most stable conformer is the undistorted $^4C_1$ chair conformation (**M1**). There are four minima (**M2** to **M5**) within 5 kcal·mol^{-1} of $^4C_1$, and the remaining ones (**M6** and **M7**) are 6 and 7 kcal·mol^{-1} higher than $^4C_1$, respectively. It is also apparent from figure VI-3 that among the distorted conformations (**M2** to **M7**), the most stable ones fall on the left-hand side of the diagram, whereas the right-hand side contains the highest energy minima. Interconversion from **M1** ($^4C_1$) to any of the other local minima involves energetic barriers of $\geq$ 6 kcal·mol^{-1}, whereas some transitions around the equatorial belt might encounter barriers as low as 1 kcal·mol^{-1}.
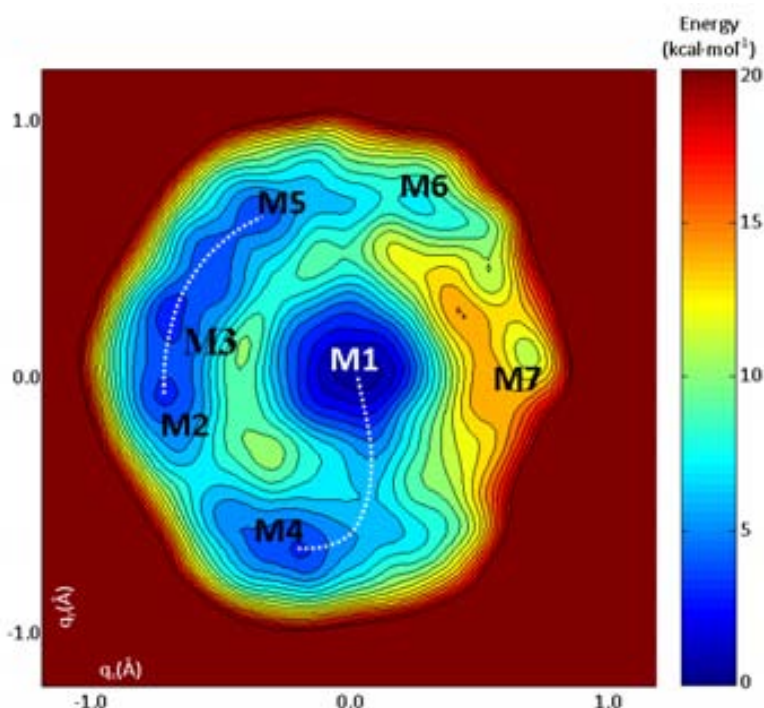

Figure VI - 3: Computed free energy surface of β-D-mannopyranose with respect to ring distortion. Energy values are in kcal·mol^{-1} and each contour line correspond to 0.5 kcal·mol^{-1}. Possible reaction itineraries are marked with white dotted lines.

It is of interest to superimpose the computed FES upon the ideal conformation map given by the Stoddart diagram (scheme VI-2). To this aim, we divided the computed FES into 12 radial regions according to the φ puckering coordinate (Figure VI-4 A). Each division line corresponds to one of the 12 different canonical conformations around the circumference of the Stoddart diagram. We assumed that a local minimum corresponds to a canonical conformation when its φ value is within ± π/24 of a division line. Intermediate conformations were assigned to minima that did not fulfill this criterion.

Table VI - 1: Relative free energies (ΔG_{rel}) among the main stationary points of the FES.

minimum	conformation	ΔG_{rel} (kcal·mol^{-1})
M1	$^4C_1$	0.0
M2	$^1S_5$ / $^{1,4}B$	3.8
M3	$^1S_5$ / $B_{2,5}$	3.8
M4	$^1S_3$ / $B_{3,0}$	4.2
M5	$^OS_2$	4.8
M6	$^3S_1$	6.0
M7	$^5S_1$	6.8

As previously observed for β-D-glucopyranose,[35] not all stationary points of the mannose FES have a direct correspondence to the ideal conformations represented in the Stoddart diagram. For instance, **M3** corresponds to a conformation between $^1S_5$ and $B_{2,5}$, while **M4** lies between $^1S_3$ and $B_{3,0}$. Moreover, there are several canonical conformers (*e.g.*, ^2S_O, $^{2,5}B$, and $B_{1,4}$) with no direct correspondence to a computed local minimum. As a consequence, there are fewer local minima (seven) than conformations in the diagram periphery (twelve).

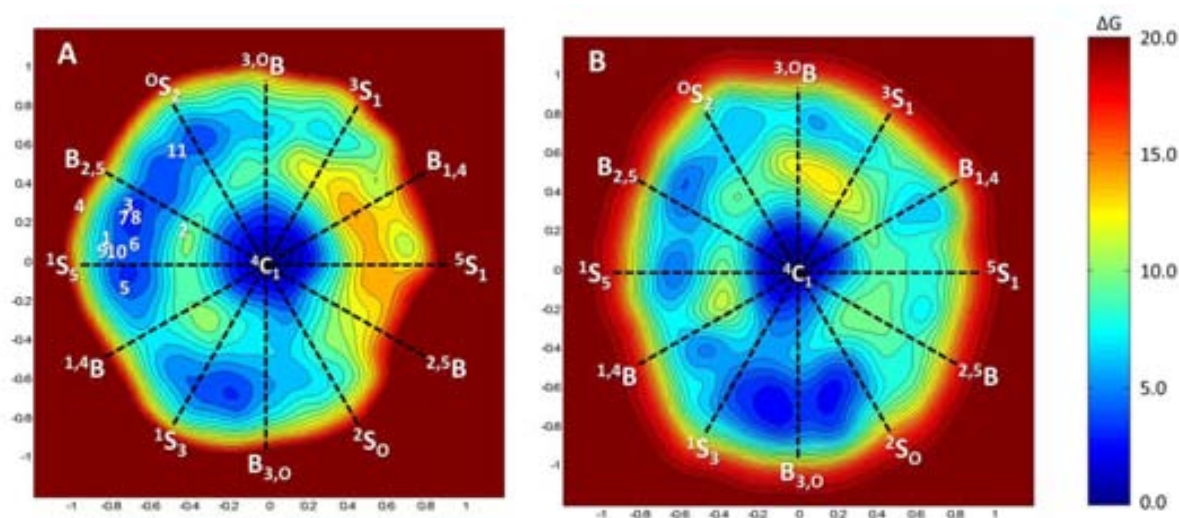**Figure VI - 4: Distribution of the canonical conformations around the Stoddart diagram on the computed free energy surface of β-D-mannopyranose (A, this work) and β-D-glucopyranose (B, results from reference 35). The conformations found in experimental structures of β-D-mannosidases are represented by numbers that correspond to the structures listed in table IV-2.**

Comparison of the β-mannose FES with the one previously computed for β-glucose (figure VI-4 B) reveals qualitative differences among the energy maps. In both cases, the most stable minima (apart from the $^4C_1$ conformation) lie on the left side of the diagram, namely from the $^OS_2$ to the ^2S_O conformation, with a small gap at $^{1,4}B$ conformation that splits this "low energy region" in two: the upper ($^OS_2$ to $^1S_5$) and the lower sides ($^1S_3$ to ^2S_O). Interestingly, in the FES of β-mannose, the upper side is more stable than the lower side, while for β-glucose is the opposite.

The differences in these FES might seem surprising a priori, as the two rings differ only in the orientation of one hydroxyl group, 2-OH, which is equatorial in $^4C_1$-glucose but axial in $^4C_1$-mannose.

Scheme VI - 4: ^2S_O conformation of β-D-mannose (A) and β-D-glucose (B).

The particular 2-OH orientation of β-D-mannose is most likely to provoke unfavorable interactions between the C2 exocyclic group and the hydroxymethyl (CH2OH) substituent at position 5 (as highlighted in scheme VI-4) when the ring adopts conformations near ^2S_O. As a result, the ^2S_O to $^{1,4}B$ region loses stabilization in the FES of β-mannose when compared to the β-glucose, for which the 2-OH group is equatorially oriented.

Scheme VI - 5: $^OS_2$ conformation of β-D-mannose (A) and β-D-glucose (B).

Similarly, the higher energy of $^OS_2$-glucose relative to $^OS_2$-mannose is due to the steric interaction between the 2-OH group and the aliphatic hydrogen at position 5 from the α-side of the sugar ring, an interaction that does not occur in mannose (scheme VI-5). In contrast, the $^1S_3$ distortion places the CH2OH substituent in a quasi-equatorial orientation in both saccharides, hence this conformation is equally stable for both FES.

In order to check whether there is a relation between the positions of the local minima and the occurrence of distorted conformations in enzyme-bound mannosides, we analyzed the substrate ring distortions of GH Michaelis complexes from all available β-mannosidase X-ray structures in which the sugar ring located at subsite -1 is either a β-D-mannose derivative or a TS-analogue inhibitor. The q_x and q_y coordinates of the saccharide at subsite -1 were computed as described in the methods section, and the corresponding value was located on the FES (see figure VI-4 A).

Table VI-2 lists the available experimental structures of β-mannoside complexes together with the conformations adopted by the substrate. Structures 1, 9, and 10 are true Michaelis complexes, whereas structures 2 to 8 correspond to complexes with TS analogues. Structure 11 (approximately $^OS_2$) corresponds to the covalent intermediate of the hydrolysis reaction.

Table VI - 2: Available complexes of β-mannosidases and the conformations adopted by the saccharide at subsite -1. (see also figure VI-4 A).

structure	PDB code	enzyme/family	resolution (Å)	substrate conformation	type of structure
1	2WBK[a]	β-Man2A/GH2	2.10	$^1S_5$	Michaelis complex
2	2VJX[b]	β-Man2A/GH2	1.85	E_5	TS analogue
3	2VL4[b]	β-Man2A/GH2	1.90	$^1S_5$ / $B_{2,5}$	TS analogue
4	2VO5[b]	β-Man2A/GH2	2.30	$^1S_5$ / $B_{2,5}$	TS analogue
5	2VR4[b]	β-Man2A/GH2	1.80	$^1S_5$	TS analogue
6	2VQT[b]	β-Man2A/GH2	2.10	$^1S_5$	TS analogue
7	2VOT[b]	β-Man2A/GH2	1.95	$^1S_5$ / $B_{2,5}$	TS analogue
8	2VMF[b]	β-Man2A/GH2	2.10	$^1S_5$ / $B_{2,5}$	TS analogue
9	1GVY[b]	β-Man26A/GH26	1.70	$^1S_5$	Michaelis complex
10	1GW1[c]	β-Man26A/GH26	1.65	$^1S_5$	Michaelis complex
11	1GW1[c]	β-Man26A/GH26	1.65	$^OS_2$	Covalent Interm.

[a]Data from ref [193]. [b]Data from ref [56]. [c]Data from ref [19].

Interestingly, all of the experimental structures fall around one of the two most stable regions of the FES ($^1S_5$ in figure VI-4 A or **M2** in figure VI-3). This suggests, as previously found for β-D-glucopyranose, that β-mannosidases have evolved to preferentially select those conformations of the mannosyl substrate that require less energy to be distorted.

As previously mentioned, the TS of the reaction catalyzed by GHs is known to be oxocarbenium ion-like.[11, 12] Therefore, optimum catalysis relies on the condition that the C5, O5, C1, and C2 atoms are almost coplanar at the TS, favoring a sp^2-like hybridization of the anomeric carbon (C1) and partial double-bond formation between C1 and O5. Among all conformations of a pyranose ring, only four of them fulfill this requirement: $^4H_3/^4E$, $^3H_4/^3E$, $^{2,5}B$, and $B_{2,5}$.

The two most stable distorted conformations of β-D-mannopyranose ($^1S_5$ and $^1S_3$ or minima **M2**/**M3** and **M4** in figure VI-3 and table VI-1) are next to putative TS conformers in the Stoddart diagram ($B_{2,5}$ and $^4H_3$, respectively). Interestingly, the $^OS_2$ and $^4C_1$ conformers (neighbors of $B_{2,5}$ and $^4H_3$, respectively) also correspond to local minima on the FES.Therefore, two possible reaction itineraries on the computed FES of β-mannose can be drawn: $^1S_5 \rightarrow B_{2,5} \rightarrow {}^OS_2$ and $^1S_3 \rightarrow {}^4H_3 \rightarrow {}^4C_1$ (marked as dotted white lines in figure VI-3).

Experimentally, however, only structures of Michaelis complexes around $^1S_5$ have been observed (figure VI-4 A), whereas no experimental structure around $^1S_3$ has been found. Most importantly, the experimentally observed conformation of the covalent intermediate (structure 11 in table VI-

2) corresponds to $^OS_2$, which determines the itinerary adopted by retaining β-mannosidases to be $^1S_5 \rightarrow B_{2,5} \rightarrow {}^OS_2$. This suggests that factors other than the relative free energy are needed in order to understand the preferred catalytic itinerary of β-mannosidases. This will be discussed below, when an index integrating several molecular properties is defined.

Structural and electronic changes

Previous studies of our group have found that the main changes in the substrate, as it evolves from the Michaelis complex to the reaction transition state, during the glycosylation step in retaining GHs (scheme VI-1) are: the partial charges of C1, O1, and O5; the C1-O1 and C1-O5 bond distances; and the orientation of the C1-O1 bond (axial or equatorial).[1, 186] Since substrate distortion in the MC causes the substrate to resemble the TS,[34, 40, 193] these electronic and structural parameters are a good measure of substrate preactivation.

Therefore, we focused on these parameters to analyze structural and electronic changes upon ring distortion.

Table VI - 3: Main structural and electronic parameters that change from the Michaelis complex to the transition state. The nature of the change and the reason are specified (see scheme VI-1).

	parameter	change from the MC to TS	reason
1	C1-O1 bond orientation	Increase	Orientation of the leaving group and the nucleophile for the S_N2 reaction.
2	C1-O1 bond distance	Increase	Bond cleavage
3	C1-O5 bond distance	Decrease	Partial double bond formation
4	C1 charge (q_{C1})	Increase	Formation of an oxocarbenium TS
5	O1 charge (q_{O1})	Decrease	Leaving group departure
6	O5 charge (q_{O5})	Increase	Charge transfer to C1

We selected a number of structures (240) from the metadynamics simulation that lay on the equatorial belt (our criterion was that their θ coordinate be within π/8 of the equatorial value π/2) and submitted them to geometry optimization. The resulting structures were grouped around the canonical distortions according to their final φ values. Each of the six parameters was computed for each structure within the group, and the average values were assigned to the corresponding canonical conformation.

The corresponding values for the undistorted $^4C_1$ structure were obtained from 20 optimized structures taken from a separate equilibrium simulation starting from this conformation. The changes observed in each of the six parameters are analyzed below:

1. C1-O1 bond orientation: The orientation of the C1-O1 bond was analyzed by measuring the angle Ω between the C1-O1 bond and the average plane of the ring. As shown in figure VI-5, $^1S_3$, $^{1,4}B$, $^1S_5$, and $B_{2,5}$ display the highest Ω values (~80°, *i.e.*, an axial orientation of the C1-

O1 glycosidic bond), whereas $^3S_1$ and $B_{1,4}$ with an equatorial C1-O1 bond have the lowest Ω (~12°). Interestingly, the most stable conformations of the FES (**M2** to **M4**) feature an axial C1-O1 bond.
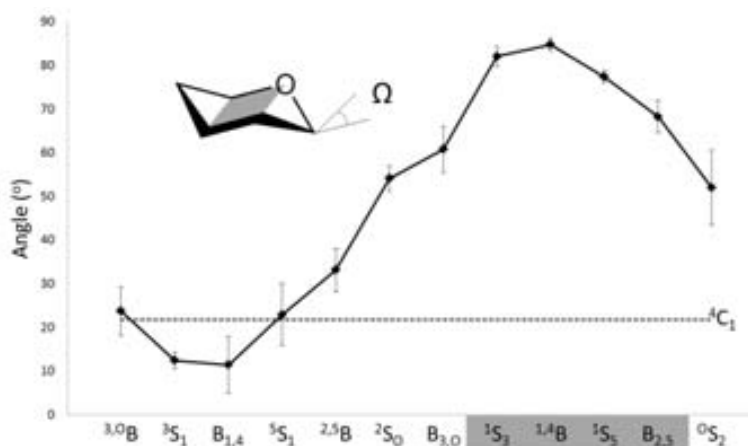

Figure VI - 5: Orientation of the bond between the anomeric carbon (C1) and its exocyclic oxygen (O1), as measured by the angle Ω between the C1-O1 bond and the average plane of the ring, plotted as a function of the ring conformation. The dotted line represents the corresponding value obtained for the $^4C_1$ conformation. The standard deviation of each point is showed as error bars.

2,3. <u>C1-O1 and C1-O5 bond distances:</u> The intra-ring distances show significant variations (0.06 Å) with ring conformation (figure VI-6 A and B). Distortions from $^{1,4}B$ to $B_{2,5}$ show a major increase in the C1-O1 distance and a major decrease in the C1-O5 distance. Therefore, $^{1,4}B$, $^1S_5$, and $B_{2,5}$ are the conformations that most resemble the TS of the glycosylation reaction in terms of the C1-O1 and C1-O5 distances.
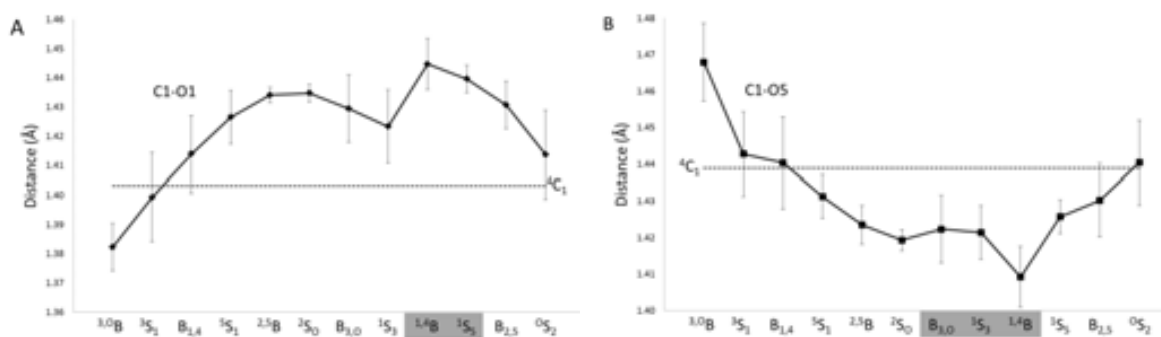

Figure VI - 6: C1-O1 (A) and C1-O5 (B) distances as a function of ring conformation (ϕ). The structures highlighted in grey are the most preactivated for catalysis.

4-6. C1, O1 and O5 charges: The conformations having a larger charge on the anomeric carbon (q_{C1}) are clearly $^3S_1$ and $B_{2,5}$ (figure VI-7), but the conformation has a minor effect on the charge of atoms O1 and O5 (q_{O1} and q_{O5}).
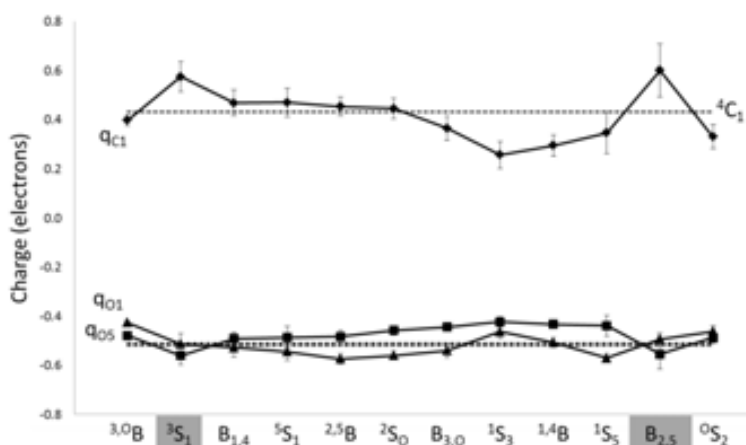

Figure VI - 7: atomic charges on the anomeric carbon (q_{C1}), the hydroxyl oxygen of the anomeric carbon (q_{O1}) and the ring oxygen (q_{O5}) as functions of the ring conformation. The dotted line represents the average charges on the three atoms obtained for the $^4C_1$ conformation (the O1 and the O5 atoms have -0.51 and -0.52 e respectively).

Therefore, the structural and electronic analysis indicates that conformations around $^1S_5$ are preactivated for catalysis to a greater extent than those around $^1S_3$.

To complement this analysis, we combined all of the parameters analyzed above (q_{C1}, q_{O5}, q_{O1}, d_{C1-O1}, d_{C1-O5}, and Ω) along with the relative free energy (ΔG_{rel}) into a unique index that could reflect the likelihood that a given conformation would be adopted in the Michaelis complex of β-mannosidases. This was done by assigning for each conformation j a score for each parameter x_i using equations VI-2 and 3:

$$score(x_{i,j}) = \frac{x_{i,j} - x_{i,j}^{min}}{x_{i,j}^{max} - x_{i,j}^{min}} \cdot 100 \quad \text{for } x_i = (d_{C1-O1}, q_{C1}, q_{O5}, \Omega) \qquad \textbf{Eq. VI - 2}$$

$$score(x_{i,j}) = \frac{x_{i,j}^{max} - x_{i,j}}{x_{i,j}^{max} - x_{i,j}^{min}} \cdot 100 \quad \text{for } x_i = (d_{C1-O5}, q_{O1}, \Delta G) \qquad \textbf{Eq. VI - 3}$$

The first formula is used for those parameters for which the score increases with increasing parameter value, whereas the second is used when the score increases with decreasing parameter value (see table VI-3). For instance, the conformation with the maximum C1-O1 distance ($^{1,4}B$, for which d_{C1-O1} = 1.45 Å) has the maximum d_{C1-O1} score (100), whereas the minimum d_{C1-O1} score (0) is assigned to $^{3,0}B$, which displays the minimum C1-O1 distance (1.33 Å). The values of the parameters and the corresponding scores are given in table VI-4. Since the score for each parameter is normalized, the scores can be directly compared, which was our main objective.

We then defined an index ξ_j as the average of the scores for the n parameters (7 in our case) for a given conformation j:

$$\xi_j = \frac{1}{n} \sum_{i=1}^{n} score(x_{i,j}) \qquad \text{Eq. VI - 4}$$

In this formulation, the conformations displaying the highest values of ξ are the most likely candidates to be the MC.
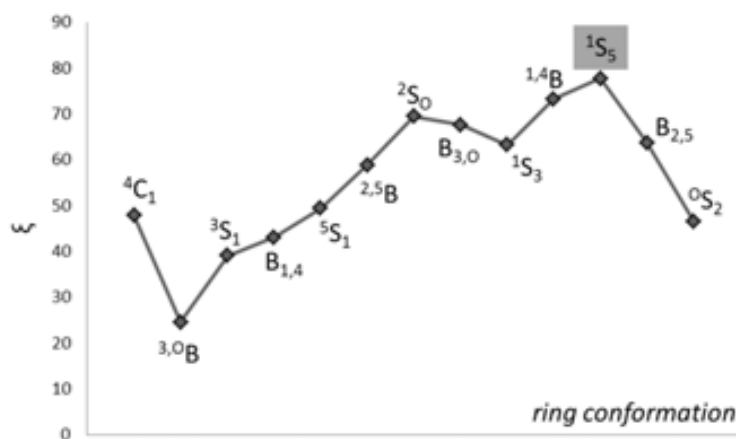

Figure VI - 8: Variation of the preactivation index ξ as a function of the ring conformation obtained for β-D-mannopyranose.

Figure VI-8 (and table VI-4) shows the variation of the ξ index with ring distortion. There is no single conformation with the optimum values of every parameter ($\xi_j = 100$). The highest values of ξ occur for conformations on the left-hand side of the Stoddart diagram (^2S_O to $B_{2,5}$), similar to the low-energy region of the FES (figures VI-2 and 3 A). The conformation with the maximum index is $^1S_5$ ($\xi = 77.8$), whereas the value for its MC competing low-energy minimum ($^1S_3$) is significantly smaller ($\xi = 63.2$).

Therefore, we conclude that $^1S_5$ is the conformation that is best prepared for catalysis in terms of energy, structure (elongation/shortening of the C1-O1/C1-O5 bonds and axiality of the C1-O1 bond), and charge development at the anomeric center. This could explain why β-mannosidases preferentially recognize the $^1S_5$ conformation and follow a $^1S_5 \rightarrow B_{2,5} \rightarrow {}^OS_2$ itinerary (figure VI-3), in contrast to the $^1S_3 \rightarrow {}^4H_3 \rightarrow {}^4C_1$ itinerary.

Similar analyses for β-D-glucopyranose (figure VI-9) showed that the conformations with maximum ξ values (^2S_O and $^1S_3$) are those observed in the experimental structures of Michaelis complexes of β-glucosidases. Therefore, the itineraries of β-glucosidases shown in scheme VI-2 ($^1S_3 \rightarrow {}^4H_3 \rightarrow {}^4C_1$ and $^2S_O \rightarrow B_{2,5} \rightarrow {}^5S_1$) are explained by the degree of preactivation in the MC.

Table VI - 4: Electronic and structural parameters defining the degree of preactivation of a given β-D-mannopyranose conformation: RESP atomic charges around the anomeric carbon (q_{CA}, q_{O5} and q_{O1}), C1-O1 and C1-O5 distances, orientation of the C1-O1 bond (Ω angle) and relative free energy (ΔG_{rel}). For each conformation, the rate for a given parameter (equations VI-2 and 3) appears in parenthesis. The average preactivation index (ξ) as computed with equation VI-4 is given in the last column. Distances are given in angstroms, charges in electrons angles in degrees and energy in kcal·mol^{-1}.

| conformer	q_{C1}	q_{O5}	q_{O1}	d_{C1-O1}	d_{C1-O5}	Ω_{C1-O1}	ΔG_{rel}	ξ
$^{3,O}B$	0,397 (41)	-0,479 (60)	-0,426 (0)	1,382 (0)	1,468 (0)	23,6 (17)	6,4 (54)	**24.5**
$^{3}S_{1}$	0.575 (93)	-0.561 (0)	-0.514 (59)	1.399 (27)	1.443 (43)	12.3 (1)	7.0 (50)	**39.0**
$B_{1,4}$	0.468 (61)	-0.491 (51)	-0.528 (69)	1.414 (51)	1.440 (47)	11.4 (0)	10.8 (23)	**43.1**
$^{5}S_{1}$	0.471 (62)	-0.487 (54)	-0.545 (80)	1.427 (71)	1.431 (63)	22.8 (16)	14.0 (0)	**49.4**
$^{2,5}B$	0.454 (57)	-0.484 (56)	-0.575 (100)	1.434 (83)	1.424 (76)	33.1 (30)	12.6 (10)	**58.8**
$^{2}S_{O}$	0.446 (55)	-0.458 (75)	-0.560 (91)	1.435 (84)	1.419 (83)	54.1 (58)	8.4 (40)	**69.4**
$B_{3,O}$	0.366 (32)	-0.444 (85)	-0.541 (77)	1.429 (75)	1.422 (78)	60.7 (67)	5.8 (59)	**67.6**
$^{1}S_{3}$	0.257 (0)	-0.424 (100)	-0.464 (25)	1.423 (66)	1.421 (79)	82.0 (96)	3.4 (76)	**63.2**
$^{1,4}B$	0.295 (11)	-0.433 (93)	-0.507 (54)	1.445 (100)	1.409 (100)	84.7 (100)	6.4 (54)	**73.2**
$^{1}S_{5}$	0.347 (26)	-0.439 (89)	-0.570 (97)	1.440 (92)	1.426 (72)	77.3 (90)	3.0 (79)	**77.8**
$B_{2,5}$	0.601 (100)	-0.555 (4)	-0.493 (45)	1.431 (78)	1.430 (64)	68.2 (77)	3.2 (77)	**63.7**
$^{O}S_{2}$	0.331 (22)	-0.488 (53)	-0.463 (25)	1.414 (50)	1.440 (47)	52.0 (55)	3.8 (73)	**46.4**
$^{4}C_{1}$	0.432 (51)	-0.520 (30)	-0.510 (57)	1.403 (33)	1.439 (50)	21.7 (14)	0.0 (100)	**47.8**
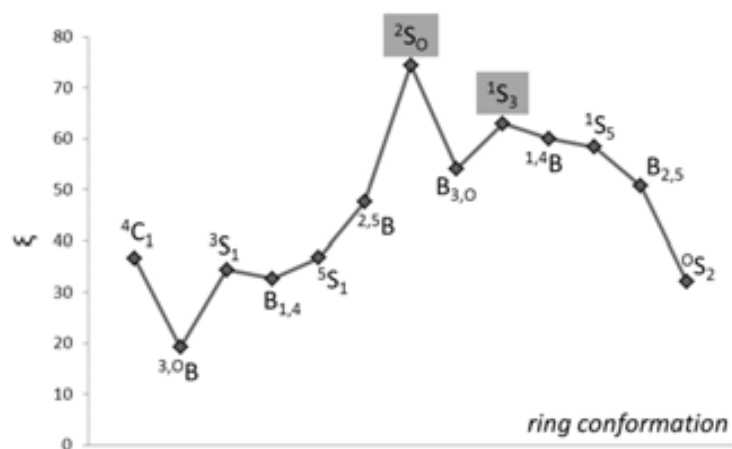
Figure VI - 9: Variation of the preactivation index ξ as a function of the ring conformation for β-D-glucopyranose.

α-L-fucose

Conformational free energy surface

Figure VI-10 shows the conformational free energy landscape of α-L-fucopyranose as a function of q_x and q_y (defined in equation VI-1). Notice that, in contrast to β-D-mannopyranose, only the southern hemisphere of the Cremer and Pople puckering sphere was sampled, and hence the chair conformation is $^1C_4$ instead of $^4C_1$.

The FES shows nine minima, labeled according to their stability from **M1** to **M9** (figure VI-10). The most stable conformer is the $^1C_4$ chair (**M1**) whereas the remaining local minima (**M2** to **M9**) are located on the periphery of Stoddart diagram (*i.e.* the equatorial plane of the Cremer-Pople sphere). There are four local minima (**M2** to **M5**) at less than 5 kcal·mol^{-1} above it, and six other local minima (**M6** to **M9**) which are ≥ 6 kcal·mol^{-1} higher in free energy. Interconversion from **M1** to any of the other local minima involves energetic barriers ≥ 8 kcal·mol^{-1}, whereas the barriers associated with the transition between peripheral conformations can be as low as 0.5 kcal·mol^{-1}.

It is also apparent that the most stable distorted conformations (**M2** to **M5**) fall on the right half of the diagram, whereas the left half contains the shallowest minima. Therefore, the "low-energy" region in α-L-fucose is the opposite of β-D-mannose and β-D-glucose. The reason for this difference is the opposite orientation of the bulky exocyclic group at position C5 (hydroxymethyl in glucose and mannose and methyl in fucose) for D and L sugars. In all cases, the orientation of this substituent in the "low-energy region" is equatorial or pseudoequatorial.
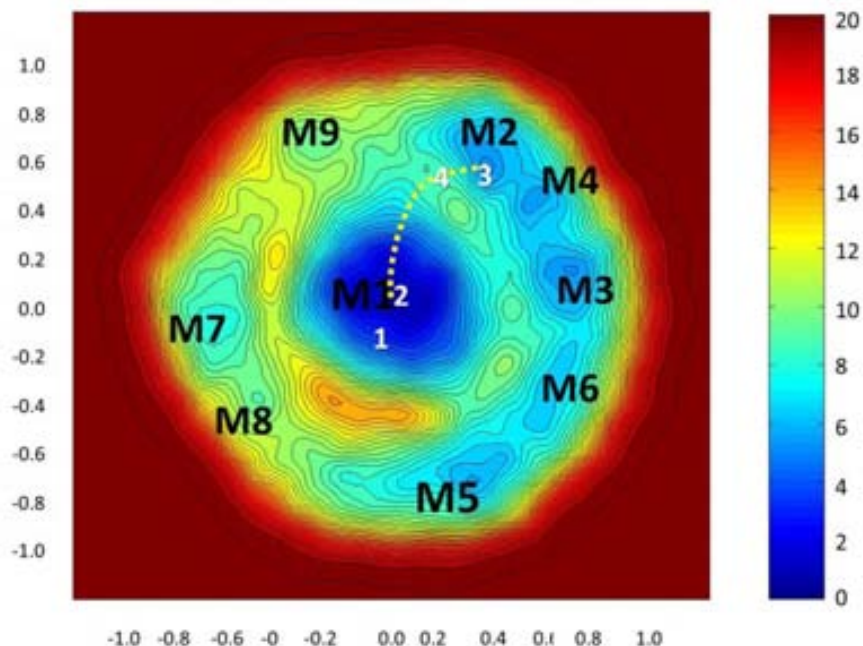

Figure VI - 10: Computed free energy landscape of α-L-fucose with respect to ring distortion (southern hemisphere of the Cremer and Pople sphere). Each contour line is 0.5 kcal·mol^{-1}. The proposed conformation itinerary of α-fucosidases is shown in dotted yellow line.

Table VI - 5: Relatives free energies (ΔG_{rel}) among the main stationary points of the FES.

minimum	conformation	ΔG_{rel} (kcal·mol^{-1})
M1	$^1C_4$	0.0
M2	$^3S_1$	4.5
M3	$B_{1,4}$ / $^5S_1$	4.5
M4	$^5S_1$ / $B_{1,4}$	5.0
M5	^2S_O	5.0
M6	$^{2,5}B$	6.0
M7	$^1S_5$	7.5
M8	$^{1,4}B$	8.5
M9	$^OS_2$	9.0

Interestingly, the FES of α-L-fucose is also consistent with the conformations of enzyme-ligand complexes observed by X-ray crystallography (figure VI-11), and with the mechanism proposed for α-fucosidases[61, 62] ($^1C_4 \rightarrow {}^OH_1 \rightarrow {}^3S_1$). Concretely, the Michaelis complex and the covalent intermediate structures exhibit the $^1C_4$ and the $^3S_1$ conformation, which correspond to the lowest minima (**M1** and **M2**) of the FES.
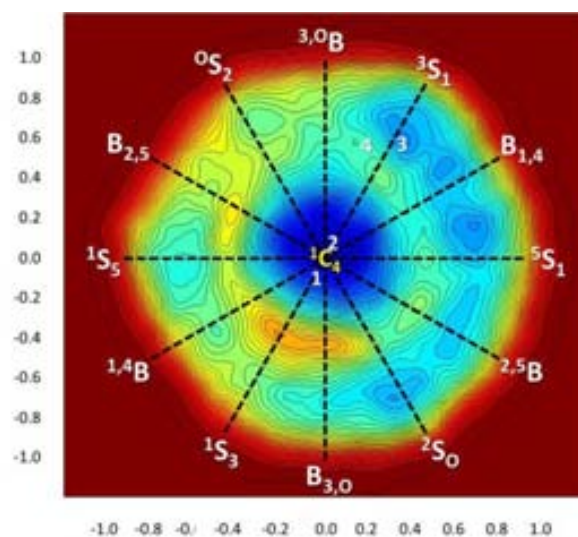

Figure VI - 11: Stoddart diagram superimposed to the FES calculated for α-L-fucose. The white numbers correspond to the X-ray structures of table VI-6

Table VI - 6: Available complexes of α-fucosidases and the conformations adopted by the saccharide at subsite -1. (see also figure VI-11).

structure	PDB code	enzyme/family	resolution (Å)	substrate conformation	type of structure
1	2WVU[a]	Bt 2970/GH29	1.95	$^1C_4$	Michaelis complex
2	2WVS[a]	Bt 2970/GH29	2.19	$^1C_4$	Michaelis complex
3	2WVT[a]	Bt 2970/GH29	1.80	$^3S_1$	Covalent intermediate
4	1HL9[b]	Tm 0306/GH29	2.25	$^3S_1$	Covalent intermediate

[a]Data from ref.[62] [b]Data from ref.[61]

Structural and electronic changes

As for β-D-mannopyranose, the relative free energy of the isolated sugar can be complemented with other structural and electronic parameters to score a "preactivation index" for each conformation. These parameters (see table VI-3) are: the partial charges of C1, O1, and O5; the C1-O1 and C1-O5 bond distances; and the orientation of the C1-O1 bond (axial or equatorial).

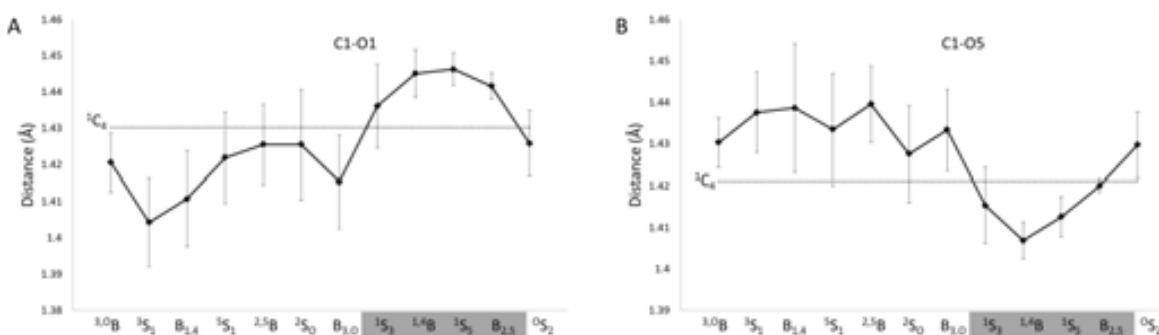To track these changes, the same procedure as in β-D-mannopyranose was performed.

Figure VI - 12: C1-O1 (A) and C1-O5 (B) distances as a function of ring conformation (φ). The structures highlighted in grey are the most preactivated for catalysis. The values for the $^1C_4$ conformation are in dotted lines.

Figures VI-12 A and B show that the C1-O1 and C1-O5 bond distances display significant variations (up to 0.04 Å) with ring conformation. Distortions in the range from $^1S_3$ to $B_{2,5}$ show the major increase in C1-O1 distance and decrease in C1-O5 distance, with respect to an undistorted $^1C_4$ conformation (broken lines).

The variations in the atomic (RESP) charges[149] for the atoms C1, O1 and O5 (figure VI-13 A) clearly show that the conformations having a larger q_{C1} are $^3S_1$, $^{1,4}B$ and $^5S_1$, while $^1S_3$ to $B_{2,5}$ are the ones with fewest charge development on the anomeric center. Only slight changes are observed on the values of q_{O1} and q_{O5}. The C1-O1 bond Ω angle (figure VI-13 B) shows a higher value (more preactivated) for conformations in the region of $B_{3,O}$, $^1S_3$ and $^{1,4}B$ as well as for the $^1C_4$ conformation.
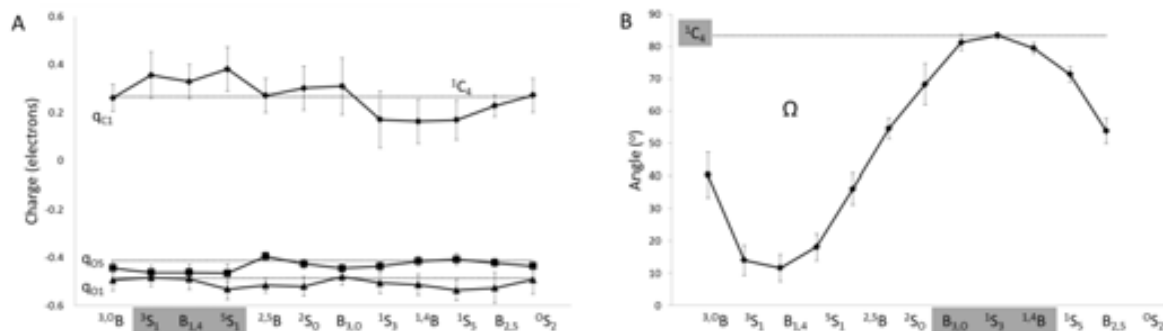
Figure VI - 13: A) Atomic charges on the anomeric carbon (q_{C1}), the O1 (q_{O1}) and the pyranic oxygen (q_{O5}) as a functions of the ring conformation. B) Orientation of the bond between the anomeric carbon (C1) and its exocyclic oxygen (O1), as measured by the angle Ω. The dotted line represents the average values for the $^1C_4$ conformation. The standard deviation is represented with error bars.

Again, there are no conformers with optimum values for all parameters. For instance, the region from $^1S_3$ to $^1S_5$ region of the diagram is more preactivated in terms of $d_{C1\text{-}O1}$, $d_{C1\text{-}O5}$ and $\Omega_{C1\text{-}O1}$, but it is less favorable in terms of charge distribution and free energy. To consider all the parameters together, the preactivation index (ξ) was calculated for the α-L-fucopyranose using equations VI-2 to 4, as previously done for β-D-mannose and β-D-glucose.
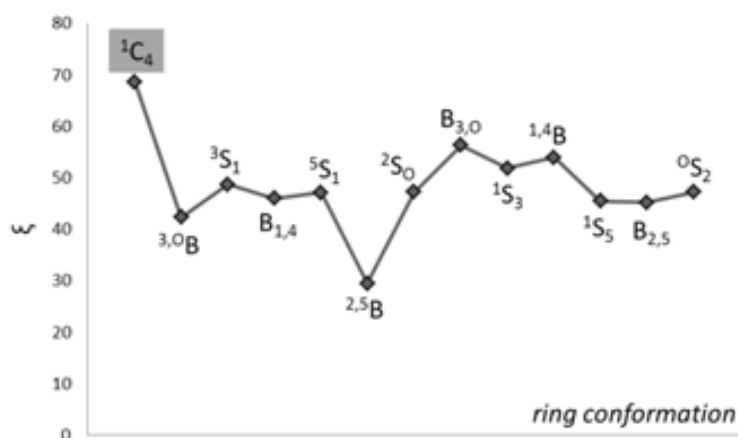

Figure VI - 14: Variation of the preactivation index ξ for each canonical conformation for α-L-fucopyranose. The most preactivated conformation is the $^1C_4$ (highlighted in grey).

The ξ score obtained for the canonical conformations in the Stoddart diagram is shown in figure VI-14 and table VI-7. The $^1C_4$ conformation is clearly the one with a higher ξ value (69), and it is therefore, the most preactivated for catalysis. The other conformations have ξ values significantly lower (from 30 in $^{2,5}B$ to 56 in $B_{3,O}$).

The results suggest that the best Michaelis complex for α-fucosidases is the $^1C_4$ conformation, and therefore give support to the $^1C_4 \rightarrow {}^3H_4 \rightarrow {}^3S_1$ conformational itinerary during catalysis. This is in good agreement with the structural experimental data available[61, 62] (see figure VI-11).

Therefore, we propose that the preactivation index ξ is an efficient score of the overall structural, electronic and energetic factors involved in the preactivation of the substrate. As such, it can be used to have an idea of the conformation of the substrate in the Michaelis complex of GHs, as well as the conformational itinerary of the substrate during catalysis.

Nevertheless, it is presumable that the exact formulation of the preactivation index might not be optimal for all sugars, and it is likely to be improved by further studies. For instance, the number of parameters included or their weight in equation VI-4 are, in a sense, arbitrary and could be easily modified in future studies.

Table VI - 7: Values of the different preactivation parameters, along with its rate as calculated in equations VI-2 and 3 (in parenthesis). The resulting preactivation index ξ associated to each canonical conformation is calculated with equation VI-4. Distances are in angstroms, RESP charges in electrons, ΩC1-O1 angle in degrees and free energy (ΔG$_{rel}$) in kcal·mol^{-1}.

conformer	q_{C1}	q_{O5}	q_{O1}	d_{C1-O1}	d_{C1-O5}	Ω_{C1-O1}	ΔG_{rel}	ξ
$^{3,O}B$	0,26 (45)	-0,49 (75)	-0,44 (70)	1,42 (39)	1,43 (28)	40,3 (40)	9.5 (0)	42
$^{3}S_{1}$	0.36 (88)	-0.48 (94)	-0.46 (97)	1.40 (0)	1.44 (6)	14.0 (3)	4.5 (53)	49
$B_{1,4}$	0.33 (76)	-0.49 (83)	-0.46 (97)	1.41 (16)	1.44 (3)	11.6 (0)	5.0 (4)	46
$^{5}S_{1}$	0.38 (100)	-0.53 (7)	-0.46 (100)	1.42 (42)	1.43 (19)	18.1 (9)	4.5 (53)	47
$^{2,5}B$	0.27 (49)	-0.52 (36)	-0.39 (0)	1.43 (51)	1.44 (0)	36.0 (34)	6.0 (37)	30
$^{2}S_{O}$	0.30 (64)	-0.52 (28)	-0.42 (44)	1.43 (51)	1.43 (37)	54.6 (60)	5.0 (47)	47
$B_{3,O}$	0.31 (67)	-0.48 (100)	-0.44 (71)	1.42 (26)	1.43 (19)	68.3 (79)	6.5 (31)	56
$^{1}S_{3}$	0.17 (3)	-0.51 (54)	-0.44 (59)	1.44 (76)	1.41 (74)	81.2 (97)	9.5 (0)	52
$^{1,4}B$	0.16 (0)	-0.51 (41)	-0.41 (29)	1.44 (97)	1.41 (100)	83.4 (100)	8.5 (11)	54
$^{1}S_{5}$	0.17 (3)	-0.53 (0)	-0.41 (18)	1.45 (100)	1.41 (82)	79.5 (94)	7.5 (21)	45
$B_{2,5}$	0.23 (30)	-0.53 (16)	-0.42 (38)	1.44 (89)	1.42 (60)	71.3 (83)	9.5 (0)	45
$^{O}S_{2}$	0.27 (50)	-0.49 (78)	-0.43 (58)	1.43 (51)	1.43 (30)	53.8 (59)	9.0 (5)	47
$^{4}C_{1}$	0.27 (47)	-0.48 (92)	-0.41 (23)	1.43 (62)	1.42 (56)	83.3 (100)	0.0 (100)	69

Conclusions

1. By means of *ab initio* metadynamics simulations, we have demonstrated that the low-energy minima of the conformational free energy surface (FES) of a free β-D-mannopyranose and a α-L-fucopyranose correlate well with the observed structures of enzyme-ligand complexes.

2. The FES of β-D-mannopyranose and α-L-fucopyranose gives support to the conformational itineraries proposed for the glycosylation reaction in retaining β-mannosidases and α-fucosidases ($^{1}S_5 \rightarrow B_{2,5} \rightarrow {}^{O}S_2$ and $^{1}C_4 \rightarrow {}^{3}H_4 \rightarrow {}^{3}S_1$ respectively).

3. The substrate distortion in the Michaelis complex of glycoside hydrolases is not arbitrary, but the protein environment induces it to adopt the distortion that is best preactivated for catalysis.

4. We have developed a scoring function to calculate the degree of preactivation of a given conformation. The preactivation index ξ includes six electronic and structural parameters that measure the molecule approach to the hydrolysis transition state (namely: q_{C1}, q_{O1}, q_{O5}, d_{C1-O1}, d_{C1-O5} and Ω_{C1-O1}), and the relative free energy.

5. The enzymatic selection of a concrete pathway presumably reflects evolutionary strain on the enzyme. It can be concluded that the intrinsic properties of a small single sugar unit play a role in this process.

Chapter VII - The molecular mechanism of enzymatic glycosyl transfer with retention of configuration.

Ardèvol, A. Rovira, C.

The molecular mechanism of enzymatic glycosyl transfer with retention of configuration.

Introduction.

The catalytic mechanism of nucleotide-sugar dependent glycosyltransferases (GTs), especially those that act with retention of anomeric configuration, remains one of the most intriguing unanswered questions in the field of glycobiology. In contrast to the well-characterized mechanistic strategies used by glycoside hydrolases (GHs) to catalyse the cleavage of glycosidic bonds,[40] the mechanisms of retaining GTs remain unclear. GTs catalyse glycosyl group transfer with either inversion or retention of the anomeric stereochemistry with respect to the donor sugar[63] (scheme VII-1).
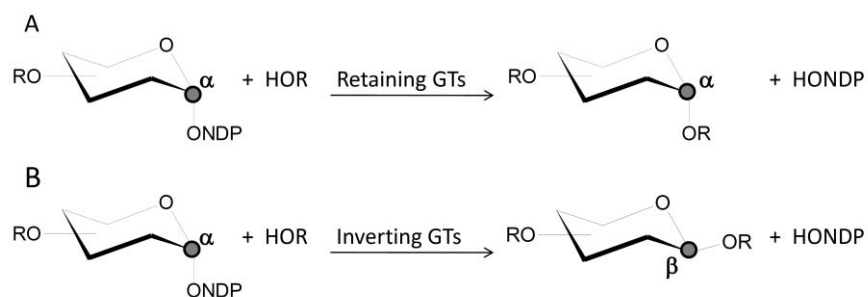

Scheme VII - 1: Classification of GTs according to their mechanism. A) In retaining GTs the stereochemistry on the anomeric carbon (marked as a dot) is maintained, while B) in inverting GTs it is not. HOR is the acceptor molecule and NDP is a nucleoside diphosphate.

In the case of inverting GTs, the mechanism is clearly established: a S_N2 reaction in a single displacement step with a general base catalyst that increases the nucleophilicity of the attacking group,[63] analogous to that of inverting GHs. In contrast, the mechanism of retaining GTs remains controversial.

A double displacement mechanism (two consecutive S_N2 reactions) was early proposed by analogy to retaining GHs.[14, 86] This mechanism involves the presence of an enzymatic residue that acts as a nucleophile and attacks the anomeric carbon, resulting of the formation of a glycosyl-enzyme covalent intermediate (glycosylation step). Then, the acceptor molecule attacks the anomeric carbon of the donor sugar *via* a second S_N2 reaction (deglycosylation step) to yield the products with overall retention of the anomeric carbon configuration.

Indeed, evidence of the formation of covalent glycosyl-enzyme adducts (scheme VII-1, top) has been reported for some fold type A GTs (family 6 mammalian α-3-galactosyltransferase, α3GalT and blood group GTs) by means of chemical rescue,[100] mass spectrometry[99] and theoretical calculations[104] (see the Introduction chapter).
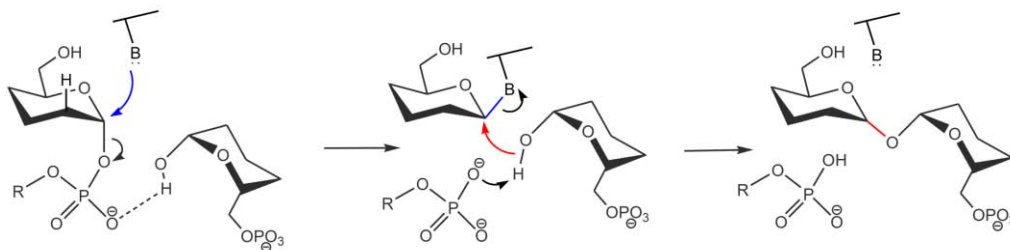
Scheme VII - 2: Double displacement mechanism proposed for retaining GTs. The reaction proceeds *via* two S_N2 nucleophilic substitution and hence a glycosyl-enzyme covalent intermediate is formed.

Another scenario is that of those GTs for which a putative nucleophile protein residue has not been proposed, such as glycogen phosphorylase (a retaining GT that does not use nucleotide-sugars[222]), lipopolysaccharyl α-galactosyltransferase (LgtC) and trehalose-6-phosphate synthase (OtsA).

This prompted some authors to suggest an unusual mechanism (scheme VII-3, top),[80] in which the reaction proceeds via a front side single displacement, similar to the solvolysis reaction of glycosyl fluoride[223] that was early regarded as a variation of the internal return nucleophilic substitution (S_Ni-like) mechanism. In this mechanism, the nucleophilic hydroxyl group of the acceptor attacks the anomeric carbon from the same side from which the leaving group departs, explaining the retention of stereochemistry. This new mechanism was proposed to avoid the presence of an stable oxocarbenium ion intermediate species that would be present in a standard S_N1 mechanism (scheme VII-3, bottom), which had been controversial for some time in retaining GHs,[224] being finally discarded.[225]
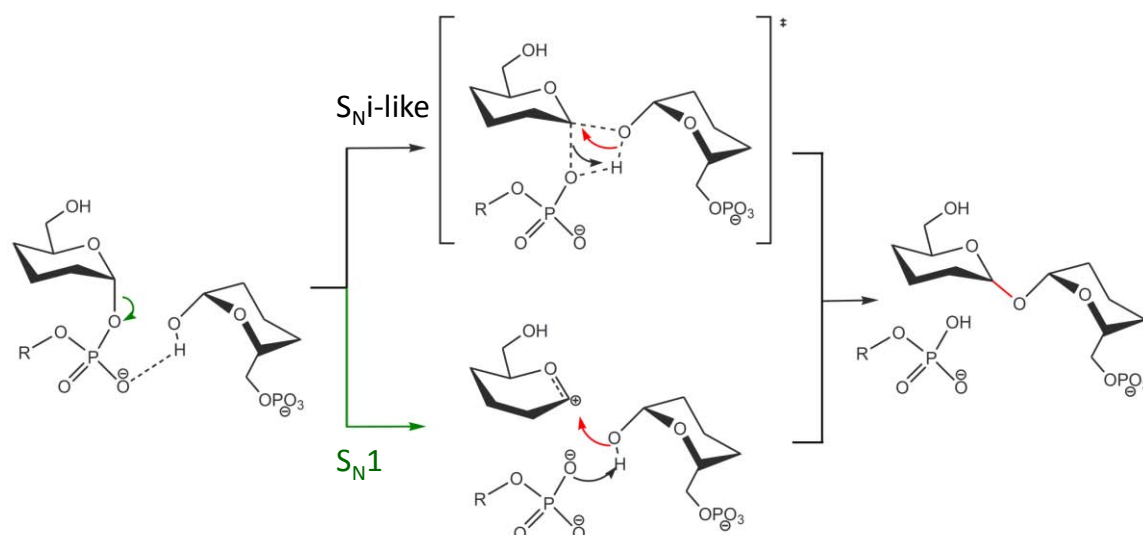

Scheme VII - 3: Front face reaction mechanism proposed for retaining GTs. The details of the mechanism are not known: the reaction can proceed *via* a S_Ni type of transition state (top) or an oxocarbenium intermediate (bottom).

Early theoretical analyses on simplified gas-phase models of the active site of LgtC (a type A fold GT) claimed that the front face mechanism and that it takes place in one step[226, 227] (scheme VII-3, top). Other authors have further elaborated on this idea[63, 105] alluding to a mechanism in which a short-lived oxocarbenium ion-like species is formed, thus avoiding the high steric strain expected for a $S_N i$ mechanism. In this respect, the reaction could also be regarded as a type of $S_N 1$ mechanism in which the outgoing leaving group enhances the basicity of the acceptor nucleophile (scheme VII-3, bottom).

Trehalose-6-phosphate synthase

E.C 2.4.1.15 trehalose-6-phosphate synthase (OtsA, figure VII-1) is a classical retaining glycosyltransferase of the GT-B fold that catalyses the formation of the α,α-1,1 linkage between uridyldiphosphate-glucose (UDP-Glc) and glucose-6-phosphate (Glc-6P), leading to trehalose-6-phosphate[228] (scheme VII-4).
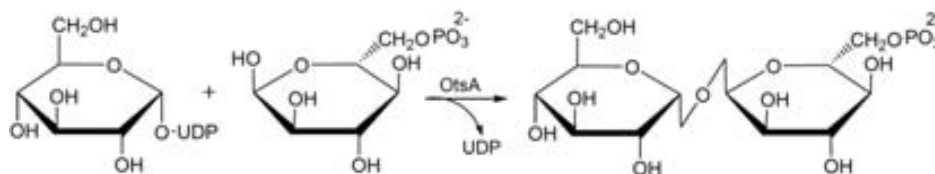
Scheme VII - 4: Reaction catalyzed by trehalose-6-phosphate synthase reaction.

A recent structural study on OtsA inhibitor complexes[106] shows that there is a hydrogen bond between the amine of the inhibitor (to be replaced by the O1' for the natural substrate, atom labelling as in figure VII-6) and the UDP, suggesting that the phosphate group acts as a Brønsted base in the glycosyl transfer reaction. This gives support to a $S_N i$-type of mechanism, without excluding the presence of a metastable (*i.e.* short-lived) oxocarbenium ion-like intermediate (see scheme VII-3). Using the IUPAC recommended nomenclature the latter is defined as D_N*A_{Nss} mechanism.[63]

A recent kinetic isotope-labelling and linear free energy relationship investigation[107] gives further support for this type of mechanism. However, a current lack of mechanistic insight does not allow conclusions regarding the molecular details of the mechanism.

Here we use Car-Parrinello quantum mechanics / molecular mechanics (CPMD/MM)[116] metadynamics simulations to decipher the reaction mechanism of
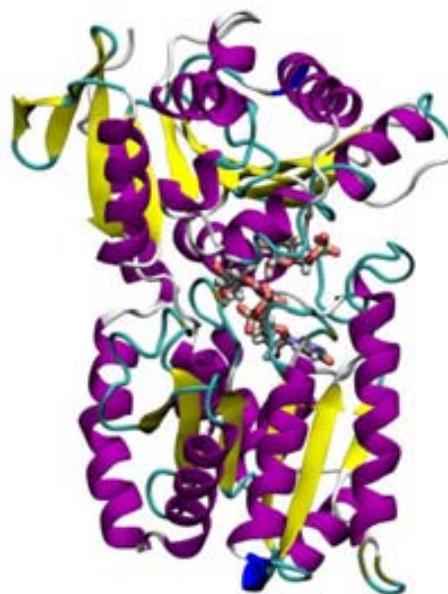

Figure VII - 1: Model of the MC of OtsA.

OtsA. The DFT description of the active site allows for an electronic characterization of the system. Together with the use of the metadynamics technique, we can depict the reaction process in an affordable computer time.

Computational details.

Initial structure

A major barrier to the modelling study of OtsA (and GTs in general) is the limited structural data for the ternary complex between the enzyme, the donor and the acceptor residues (UDP-Glc and Glc-6P in OtsA). Among the complexes that were recently reported by **Errey** *et al.*,[106] there is one (OtsA complex with UDP and validoxylamine-6-phosphate, VA6P) that structurally resembles one of the reaction products (trehalose-6-phosphate). Therefore, we took this ternary OtsA complex with VA6P and UDP (PDB 2WTX) and we manually converted into the natural products of the reaction (UDP and trehalose-6-phosphate). This structure was then used to generate the Michaelis complex (MC) (*i.e.* the OtsA·UDP-Glc·Glc-6P ternary complex) for which there is no crystal structure available as explained below. The building of the Michaelis complex using this procedure will be detailed in a forthcoming section named "building the Michaelis complex".

The protonation states and hydrogen atom positions of all other ionisable aminoacid residues were selected based on their hydrogen bond network. Thirteen histidine residues were modelled in their neutral states. All the crystallographic water molecules were retained and extra solvent was added to form a 10 Å water box around the protein surface. Nine sodium ions were also added to neutralize the enzyme charge.

Classical MD

A classical molecular dynamics using NAMD[229] software was initially performed. The protein was modelled with the FF99SB force-field.[144] All carbohydrates, including the substrate and GlcNAc glycosylatioin site at Asn194, were modelled with the GLYCAM06 force-field[230] and water molecules were described with the TIP3P force-field.[153] 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 sodium ions were allowed to move freely. The constraints were then removed and the MD was extended to 5 ns when the system had reached equilibrium. The overall structure did not change significantly from the X-ray structure (the RMSD of the backbone atoms is within 1 Å). Analysis of the trajectories was carried out using standard tools of VMD.[231] A snapshot of the MD-equilibrated system was taken for the subsequent CPMD/MM simulations.

Quantum mechanics / molecular mechanics MD

CPMD/MM calculations were performed using the method developed by **Laio** *et al.,*[116] which combines Car–Parrinello MD,[112] based on DFT, with force-field MD. The CPMD/MM interface was modeled by the use of a link-atom pseudopotential that saturates the QM region.[166] The electrostatic interactions between the QM and MM regions were handled via a fully Hamiltonian coupling scheme, where the short-range electrostatic interactions between the QM and the MM regions are explicitly taken into account for all atoms.[116] 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 were treated *via* a multipole expansion. Bonded and van der Waals interactions between the QM and the MM regions were treated with the standard AMBER force-field.[230] Long-range electrostatic interactions between MM atoms were described with the P3M implementation[232] using a 64 x 64 x 64 mesh.

The QM region chosen in this work included the Glc-UDP donor and the Glc-6-P acceptor. The sidechain of the nearby Arg262 was also included, capped at its C_α atom with a link-atom pseudopotential. The QM region (72 atoms) was enclosed in an isolated 18.5 x 16.4 x 18.0 Å supercell. Kohn-Sham orbitals were expanded in a planewave basis set with a kinetic energy cutoff of 70 Ry. Norm-conserving Troullier-Martins *ab initio* pseudopotentials[114] were used for all elements. The calculations were performed using the Perdew, Burke and Ernzerhoff generalized gradient-corrected approximation (PBE).[221] The fictitious mass for the electronic degrees of freedom of the CP Lagrangian was set at 500 a.u., and the simulation time-step at 5 a.u. (0.12 fs). The simulation of the OtsA complex with UDP and trehalose-6-phosphate was equilibrated at 300 K for 6.4 ps, and the resulting structure (figure VII-2) is in good agreement with the OtsA·VA6P binary complex of **Errey** *et al.*[106]
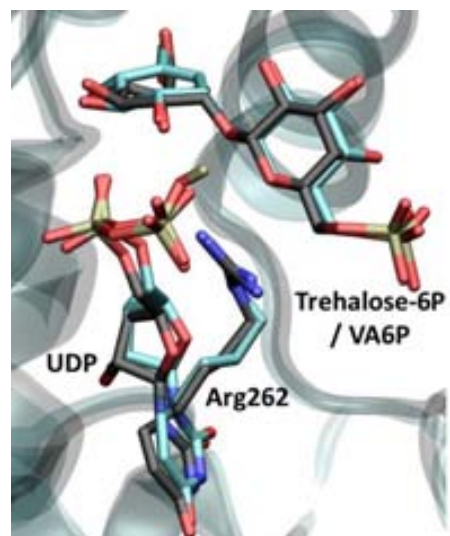
Figure VII - 2: Superimposition of the computed structure of the products complex with the crystallographic structure of OtsA with the VA6P inhibitor (PDB code 2WTX). C atoms of the X-ray and the computed structure are in cyan and black respectively.

Building the Michaelis complex

Starting from the OtsA complex with UDP and trehalose-6P, exploratory metadynamics simulations were performed to drive the reaction in the reverse direction thus reaching the reactants state (the Michaelis complex). The reaction was activated in two steps (scheme VII-5).
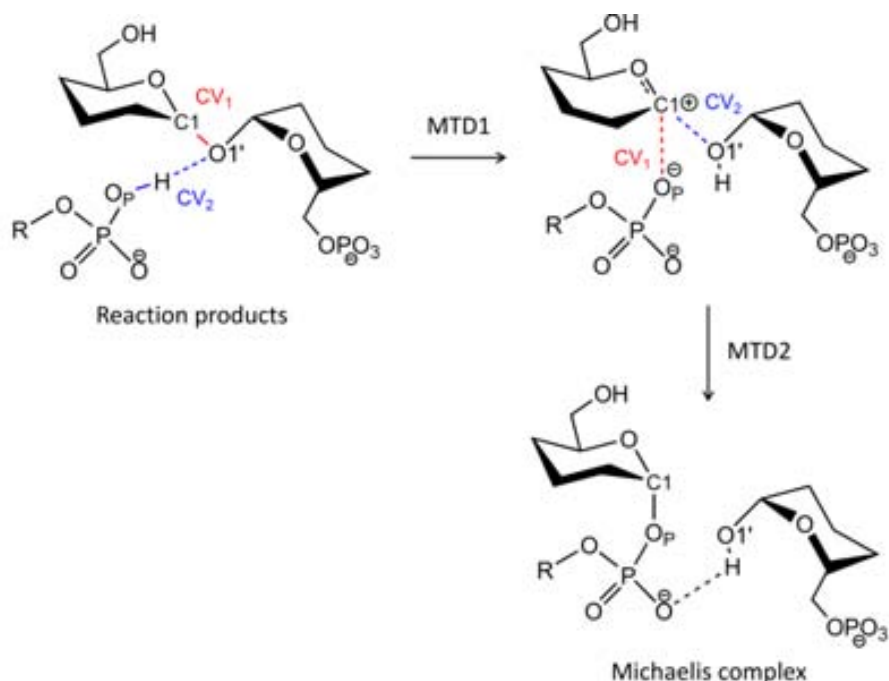
Scheme VII - 5: Design of the two exploratory metadynamics simulations used to build the Michaelis complex. The starting configuration are the reaction products. The collective variables described in tables VII-1 and VII-2 are highlighted in red and blue.

The first metadynamics simulation (MTD1) was aimed at breaking the glycosidic bond of the acceptor molecule was performed. Two collective variables were used. The first one (CV_1 in table VII-1) was taken as the coordination number (CN) between the two atoms forming the glycosidic bond. The second one (CV_2 in table VII-1) activates the transfer of the proton from O1' to O_P that is necessary to stabilize a putative ion-pair species upon glycosidic bond cleavage. The CV was defined as in Chapter V:

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

The parameters used for each CV in MTD1 are as follows.

Table VII - 1: Definition and parameters of the CVs used in the metadynamics simulation MTD1. Atom labeling as in scheme VII-5.

MTD1	Definition	Parameters	
CV_1	$CV_1 = CN_{C1,O1'}$	d_0	3.65 a.u.
		p	14
		q	14
CV_2	$CV_2 = CN_{O1',H} - CN_{O_P,H}$	d_0	2.50 a.u.
		p	10
		q	16

The selected mass values of the fictitious particles (μ) were 8 and 4 a.u. for CV_1 and CV_2 respectively, while those of the force constant (k) were 2.0 and 1.8 a.u. respectively. The height of the Gaussian used was of 1.255 kcal·mol^{-1}. The width of the Gaussian terms (0.15 in coordination number units) and the deposition time of the Gaussian-like potentials hills (τ = 200 MD steps or 24 fs) were selected from the oscillations of the CVs in a free Car-Parrinello CPMD/MM simulation. These values are similar to the ones used in other CPMD/MM metadynamics simulations of chemical reactions performed in our group and other research groups.[1, 26-28, 233, 234] The simulation was stopped after 19.51 ps, once the free energy surface (FES) of the reaction simulated was sampled. A total of 807 Gaussian hills had been deposited.

The FES reconstructed from the MTD1 metadynamics simulation (figure VII-3 A) revealed a minimum structure corresponding to an oxocarbenium ion. This intermediate is probably very short-lived (**I** in figure VII-4 A and B) in view of the FES. The glycosidic bond is already cleaved in **I** (C1···O1' distance of 2.6 Å), the proton is on the glycosidic bond (O$_P$···H distance of 3.8 Å), and the anomeric carbon adopts a sp^2 hybridization.

Another metastable structure (**P'** in figure VII-4 A) corresponding to an oxonium ion (figure VII-3), in which the proton has already been transferred to the glycosidic bond has not yet been cleaved, was found. Interestingly, the chemical species (**I**, **P** and **P'**), as well as the free energy profile are in good agreement (qualitatively and quantitatively) with a similar study by Liang *et al.*[175] on the cellobiose glycosidic bond cleavage.
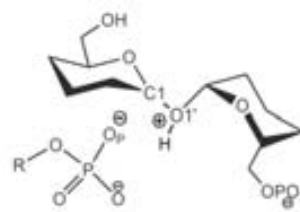

Figure VII - 3: Oxonium ion (P' in figure VII-4 A) found in MTD1

Starting from the structure of the oxocarbenium ion (**I** in figure VII-4 A), a second metadynamics (MTD2 in scheme VII-5) was performed using the distance between C1 and the O$_P$ phosphate atom as a collective variable (CV$_2$, *"phosphate bond formation/cleavage"* in figure VII-4 B). However, to avoid that the system goes back to the products (*i.e.* forming again the C1-O1' glycosidic bond), an addition collective variable (CV$_1$, *"glucose bond formation/cleavage"*) was considered as the coordination number between the C1 and O1' atoms ($CV_1 = CN_{C1,O_P}$).

By introducing a repulsive wall at CV_1 = 0.09 in the first part of the simulation, glycosidic bond formation was avoided. The parameters used for CV_1 were: p = 14, q = 14 and d_0 = 3.65 a.u. The values of height (w) and width (δs) for the repulsive Gaussian potentials, the μ and k values for the fictitious particles and the Gaussian deposition time τ, were the same as in MTD1. The simulation was stopped after 12.50 ps and a total of 515 Gaussian hills had been deposited.

Table VII - 2: Definition and parameters of the CVs used in the metadynamics simulation MTD2. Atom labeling as in scheme VII-5.

MTD2	Definition	Parameters	
CV$_1$	$CV_1 = Distance_{C1,O_P}$	---	
CV$_2$	$CV_2 = CN_{C1,O1'}$	d$_0$	2.50 a.u.
		p	10
		q	16
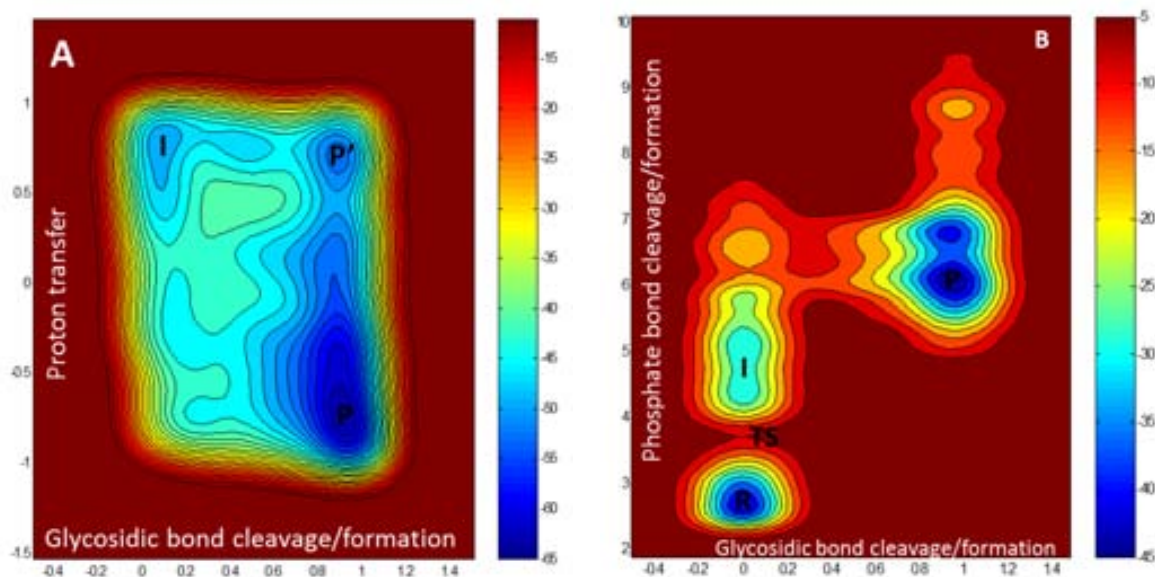
Figure VII - 4: FES reconstructed from MTD1 (A) and MTD2 (B). Each iso-energy line represents 2 and 4 kcal·mol^{-1} respectively. P = reaction products (OtsA + UDP + trehalose-6-phosphate ternary complex). R = Reactants (OtsA + Glc-UDP + Glc-6P, *i.e.* the Michaelis complex). P' = metastable oxonium ion (figure VII-3). I = oxocarbenium-ion species. TS = transition state.

The computed structure for the Michaelis complex of the reaction (the **R** state) superimposes very well with the crystallographic structure of the binary complex of OtsA with a 2-fluoro-Glc-UDP inhibitor[235] (PDB code 1UQT) shown in figure VII-5. The slightly different position of the phosphate group is probably due to the lack of hydrogen bond between the phosphate and the acceptor molecule that is not present on the X-ray structure. This agreement gives confidence that the model of the MC is a reliable structure for the subsequent simulations.

Figure VII - 5: Superimposition of the average structure of the modeled MC (carbon atoms in black) with the binary complex X-ray structure 1UQT (carbons in cian).

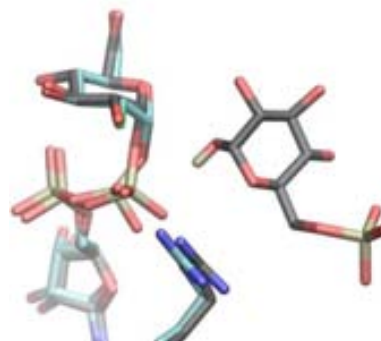Going from the reactants to the products, the rate limiting step is the cleavage of the glucose-phosphate bond (figure VII-4 B) and has a free energy barrier of 30.7 kcal·mol^{-1}. It should be noted though, that because of our particular choice of the collective variables, the reaction was forced to take place in two steps; following a typical S_N1 type of reaction in which the Glc-UDP bond breaking takes place without assistance from the incoming acceptor (H transfer is not activated in the rate limiting step), as is shown in scheme VII-3 bottom.

Metadynamics simulation of the glycosyl transfer reaction

The metadynamics simulation of the glycosyl transfer reaction was initiated from the modelled Michaelis complex (the **R** state in figure VII-4 A). Two collective variables, defined as functions of the distances involved in the bonds that are being formed or cleaved, during the reaction were considered as collective variables (figure VII-5).
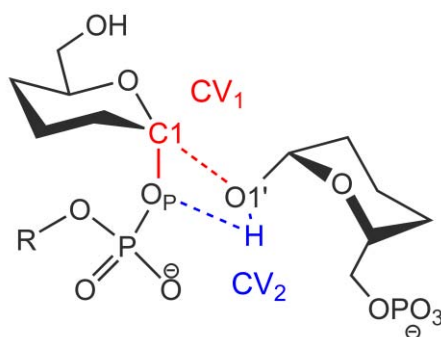
Figure VII - 6: Collective variables used in the CPMD/MM metadynamics simulation of the trehalose-6-phosphate synthesis.

The first collective variable (CV$_1$) is the difference in coordination number (CN) between the anomeric carbon of the donor with the phosphate oxygen (O$_P$) minus the O1' ($CV_1 = CN_{C1,O_P} - CN_{C1,O1'}$). The second collective variable (CV$_2$) involves the O1', H and OP atoms in a similar way ($CV_2 = CN_{O1',H} - CN_{OP,H}$). Therefore, CV$_1$ describes the cleavage of the glucose-UDP covalent bond and the formation of the glycosidic bond, while CV$_2$ describes the proton transfer from the acceptor molecule to the phosphate group.

It is important to note that by using the above defined collective variables, both the fully concerted reaction (scheme VII-3, top) or a step-wise S_N1 reaction (with a stable intermediate) could be obtained. The simulation does not impose any of these mechanisms, but can differentiate between them.

The p, q and d$_0$ parameters used for each CV are defined in table VII-2. The mass (μ) of the fictitious particles were set to 8 and 10 a.u. for CV$_1$ and CV$_2$ respectively and the spring force constants (k) were set to 1.9 and 1.4 a.u.

Table VII - 3: Definition and parameters of the CVs used in the metadynamics simulation of the trehalose-6-phosphate synthesis. Atom labeling as in figure VII-6

CV	Definition	Parameters	
1	$CV_1 = CN_{C1,O_P} - CN_{C1,O1'}$	d_0	3.53 a.u.
		p	16
		q	16
2	$CV_2 = CN_{O1',H} - CN_{O_P,H}$	d_0	2.25 a.u.
		p	16
		q	16

The Gaussian height used was of 1.255 kcal·mol^{-1} during the firsts 3.7ps and was reduced to 0.628 kcal·mol^{-1} for the rest of the metadynamics simulation. The width of the Gaussians was set to 0.10. A new Gaussian like potential was added every τ = 200 MD steps (24 fs) and the simulation was stopped after 6.04 ps of simulation once the products well was reached. A total of 250 Gaussian hills potentials had been deposited.

The metadynamics simulation was further extended up to 24.20 ps (998 metasteps) but the system did not come back to the reactants state (instead, a very wide and deep products well appeared). This is indicative that there are subtle structural changes (possibly, a slight displacement of the UDP group away from the trehalose-6P in the products) that preclude the products from going back to the reactants within the subspace of collective variables used. As a consequence, the relative free energy between reactants and products cannot be deduced from the metadynamics simulation.

Results

The free energy surface reconstructed from the metadynamics simulation is shown in figure VII-7. The two most stable minima are the reactants (in the upper left corner) and the products (in the lower right corner). The reactants well is 23 kcal·mol^{-1} below the maximum point along the minimum energy reaction pathway, in reasonable agreement with the values obtained from experimental measurements of reaction rate constants for retaining GTs (17-20 kcal mol^{-1}).[99, 100, 107, 236] The energy barrier is also similar to the ones reported for glycosyl hydrolases investigated with a similar computational methodology.[1, 26-28]

Two deep minima were found in the FES, corresponding to the Michaelis complex (**R**) and the products (**P**). The shallow minimum (**P'**) corresponds to an oxonium ion (the glycosidic bond is formed but the proton is at atom O1' instead of the phosphate). A detailed description of the reaction can be obtained by following the minimum free energy pathway[237] from the reactants to the products, represented as a dashed line in figure VII-7. Four structurally different species (**1** to **4** in figure VII-7) that were very close in energy were identified along the reaction path.
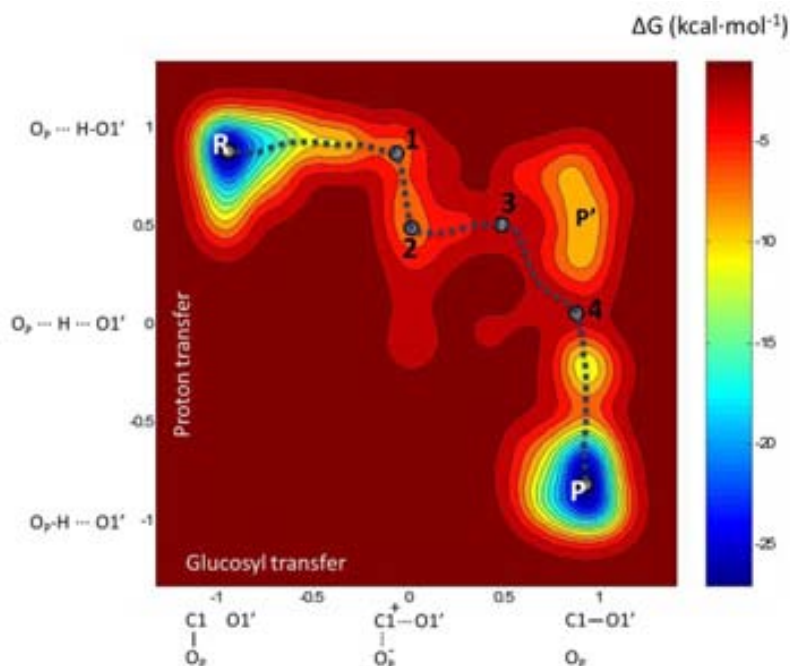
Figure VII - 7: Computed free energy landscape with respect to the two collective variables described in table VII-2. Each contour line corresponds to 2 kcal·mol^{-1}.

Figure VII-9 shows representative snap-shots of relevant configurations along the minimum energy pathway. At the reactants state (**R**), there is a hydrogen bond between the O1' acceptor atom and one of the non-bridging UDP oxygen atom. Consequently, the distance between the H atom and the leaving phosphate oxygen (OP) is rather long (OP···H > 2.5 Å, Figure VII-8 and table VII-3).
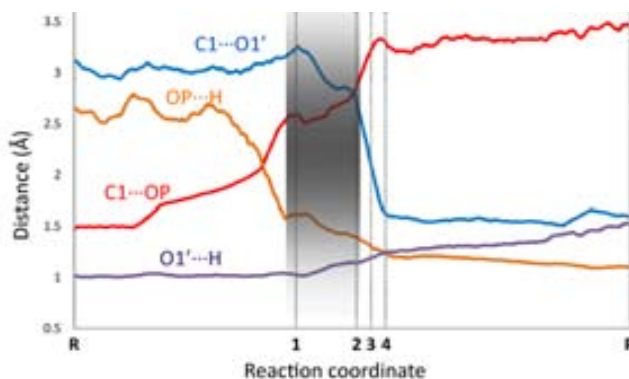
Figure VII - 8: Evolution of the most relevant distances involving the donor and acceptor along the reactioin path. (Atom labeling as in figure VII-6).

The reaction starts with the elongation of the C-O bond between the UDP and the glucose molecule of the donor (*e.g.* the C1···O$_P$ distance increases more than 1 Å from **R** to **1**). Simultaneously, the O1'-H bond changes hydrogen bond partner to O$_P$ atom (see *e.g.* the dramatic decrease in the O$_P$···H distance from 2.5 Å to 1.6 Å, figure VII-8 and table VII-3), placing the hydrogen atom at the proper position to assist the departure of the UDP leaving group. The O1'-H···O$_P$ hydrogen bond not only stabilizes the negative charge being developed at the phosphate

127

group, it also places the acceptor molecule in a position that favors the forthcoming front side nucleophilic attack. Therefore, the first part of the reaction can be described as the cleavage of the $C1-O_P$ bond together with the formation of a hydrogen bond between the acceptor and the phosphate oxygen of the bond being broken.
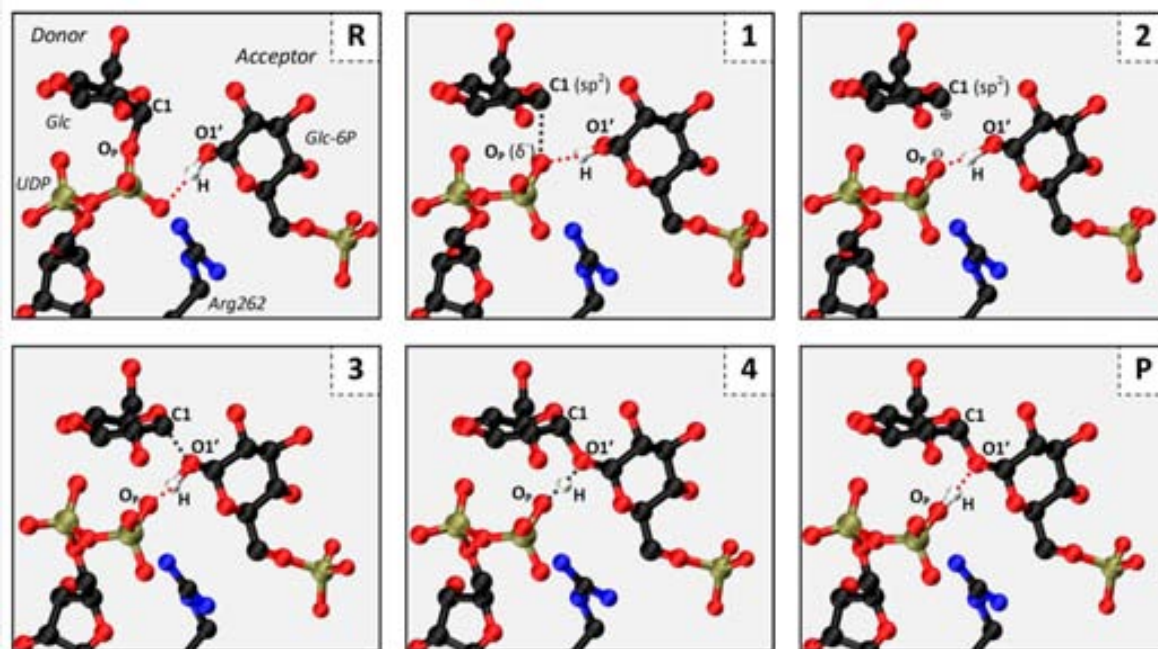
Figure VII - 9: Atomic rearrangement along the minimum free energy reaction path. Hydrogen atoms have been omitted for clarity, except the one being transferred. Bonds being broken or formed are represented by a dotted black line whereas highlighted hydrogen bonds are represented by a dotted red line.

The $C1-O_P$ bond becomes completely broken in **2** ($C1-O_P$ = 2.84 Å, table VII-3). Remarkably, the distance between the donor and the acceptor ($C1 \cdots O1'$) is still rather long (2.79 Å), indicating the presence of an oxocarbenium-phosphate ion pair. In fact, there is a portion of the reaction pathway (*i.e.* the shadowed region in figure VII-8) in which both O_P and O1' atoms are quite separated from the anomeric carbon (> 2.5 Å). All complexes in this region, which expands around **1** and **2**, approximately, can be taken as representatives of oxocarbenium ion-like species.

Further evidence for the sp^3 to sp^2 change in the hybridization of the anomeric carbon is the conformational change of the glucose ring from $^4C_1$ chair in the reactants to a $^4H_3$ half-chair conformation (the C2, C1, O5 and C5 atoms are almost coplanar) at **1**, **2** and **3**. This is accompanied by a decrease of the intra-ring C1-O5 distance (from 1.39 to 1.27 Å, table VII-3) and an increase of the ESP charge[149] of the anomeric center (by 0.45 e from **R** to **2**, Figure VII-10). These structural changes are in remarkable agreement with recent kinetic isotope-labeling experiments, which indicate that a species with substantial oxocarbenium ion character, with the pyranose ring flattened through C5-O5-C1-C2, forms.[107]

Table VII - 4: Relevant distances along the minimum free energy reaction path. Distances are in Å and energy in kcal·mol^{-1}. Standard deviations are given in parenthesis.

	C1 - OP	C1 - O1'	C1 - O5	H - O1'	H - OP	ΔG
R	1.51 (0.04)	3.04 (0.11)	1.39 (0.03)	1.03 (0.03)	1.87 (0.27)	0
1	2.56 (0.13)	3.20 (0.12)	1.28 (0.01)	1.03 (0.03)	1.60 (0.05)	19.24
2	2.84 (0.18)	2.79 (0.09)	1.27 (0.01)	1.15 (0.02)	1.38 (0.06)	18.36
3	3.40 (0.00)	1.89 (0.00)	1.28 (0.01)	1.16 (0.00)	1.35 (0.00)	23.17
4	3.34 (0.16)	1.63 (0.02)	1.35 (0.02)	1.24 (0.01)	1.27 (0.02)	23.00
P	3.48 (0.15)	1.52 (0.07)	1.39 (0.03)	1.56 (0.13)	1.09 (0.03)	1.14

The decrease of CV$_2$ from **1** to **2** reflects the increasing contribution of the O1'-H···O$_P$ hydrogen bond to the phosphate-oxocarbenium ion-pair stabilization. In fact, the H···O$_P$ distance becomes quite short at **2** (1.38 Å). Afterward, a slight displacement of the glucose donor and acceptor molecules takes place, facilitating the interaction between the C1 and O1' atoms at **3**. Concomitant with the formation of the C1-O1' bond, the proton transfers to the phosphate group (at **4**, the proton is shared between O1' and O$_P$). The process **3 → 4 → P** is thus the formation of the donor-acceptor glycosidic bond in concert with the protonation of the UDP phosphate group.

As a further characterization of the oxocarbenium ion-like species, we extracted several snapshots of the metadynamics simulation that correspond to structures around **1** and **2** and performed geometry optimizations followed by room temperature *ab initio* molecular dynamics simulations. Although the ion-pair species turned out to be stable under optimization, in all cases the structure evolved either towards reactants or towards products within 1 picosecond of molecular dynamics at room temperature.
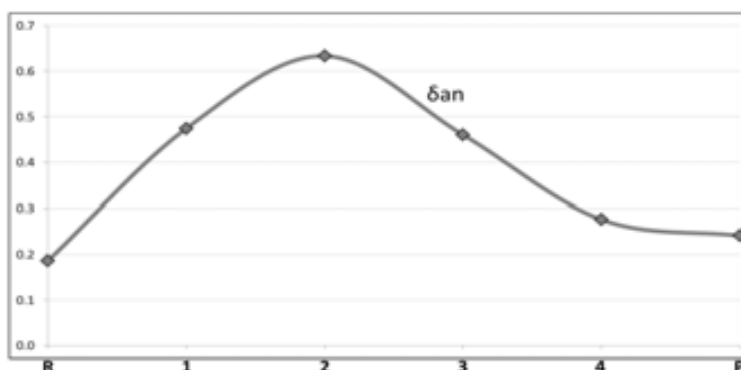
Figure VII - 10: Evolution of the ESP charges at the anomeric center (δan). Because the charge increase on the anomeric carbon atom partially delocalizes over its neighbor atoms, especially at O5, δan is computed by summing up the charges of C1, O5, H1 and C2 atoms. The same procedure was used in our previous work on glycoside hydrolases.[31b]

Discussion

Several conclusions can be drawn from the above mechanistic analysis. Firstly, the cleavage of the glucose-UDP bond (C1-O$_P$) and the formation of the donor-acceptor bond (C1-O1') are highly

asynchronous. The cleavage of the UDP-Glc bond precedes the formation of the Glc-Glc bond and thus an intimate ion-pair intermediate is formed. Secondly, there is a crucial participation of the acceptor hydrogen during the whole process. At the first stage of the reaction, the acceptor hydrogen interacts with the O_P atom, assisting the departure of the UDP leaving group. In the second stage, it transfers to the phosphate group, enhancing the nucleophilicity of the acceptor sugar O1' and thus facilitating the formation of the donor-acceptor glycosidic bond.

Interestingly, when the H···O_P interaction is not included as part of the collective variables (see the two exploratory metadynamics simulations in the Methods section), the **TS** of the minimum free energy pathway is 30.7 kcal·mol^{-1} above the reactants (**R** in figure VII-4 B), and corresponds to the non-catalyzed cleavage of the glucose-phosphate bond. In contrast, when the H···O_P interaction is included, although the rate limiting step is the same, the $\Delta G^{\ddagger}$ is 23.3 kcal·mol^{-1}. Therefore, we conclude that even though the mechanism in which the acceptor group does not participate in the rate limiting step of the reaction is possible, it is less favored (by 7.4 kcal·mol^{-1}), and that the critical stabilizing interaction of the acceptor hydrogen with the UDP leaving group acts as a driving force for the reaction.

The FES of figure VII-7 could be interpreted as a two steps mechanism (similar to that of a S_N1 reaction, but with participation of the acceptor molecule) in which an oxocarbenium ion-pair intermediate forms. However, the energy difference among configurations **1** to **4** is very small (± 2 kcal·mol^{-1}). The oxocarbenium ion (**1** to **3**) corresponds to just a shallow minimum in the free energy profile (see figure VII-11) and, therefore, it represents an extremely short-lived species. Notably, such flat reaction free energy profile was already hypothesized by Sinnott and Jencks for the solvolysis of D-glucosyl fluoride by mixtures of ethanol and trifluorethanol.[223]
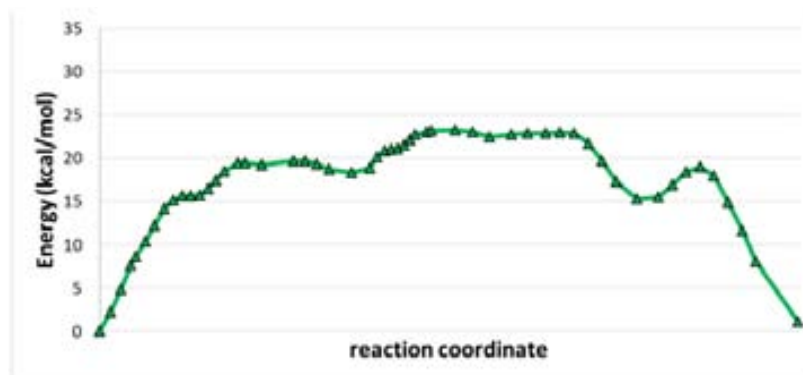
Figure VII - 11: Free energy profile of the metadynamics of the overall reaction (using the CVs specified in table VII-2). The activation energy is 23.3 kcal·mol^{-1}.

As a consequence, the lifetime of the intermediate is short (in the range of 1 ps according to free CPMD/MM simulations starting at several points in the region from **1** to **3**) and it should be very difficult to observe experimentally. However, it is long enough to allow a reorganization of the active site: Whereas the C1···O_P distance increases, the H-bond that forms between the acceptor and the phosphate (H···O_P) drives the acceptor towards the oxocarbenium ion with whom it collapses.

It is important to point out that the identification of a unique transition state structure for the enzymatic reaction is, in this case, not trivial. The highest point along the free energy pathway corresponds to the formation of the Glc-Glc bond (**3**), but most of the activation energy is invested in breaking the UDP-Glc bond (**R → 1**). As mentioned above, there is a wide window along the reaction pathway (from **1** to **4**, approximately) in which the free energy does not change significantly. Any structure in this region can be taken as representative of the transition state and therefore an experimental measurement aimed at characterizing the transition state would probably reflect a mixture of all configurations. The TS-like inhibitor of **Errey** *et al.*, for instance, resembles configuration **4** from its donor-acceptor covalent bond distance. Our calculations suggest that molecules with a longer donor-acceptor bond, *i.e.* moving towards configuration **3** or **2** (C1-O1' ≥ 2 Å) in the FES, such as those with a thio-glycosidic bond, might also be good inhibitors.

The electronic structure of the ion-pair species can be analyzed from its localized Wannier orbitals,[238] shown in figure VII-12. The partial double bond formation between the pyranic oxygen (O5) and the anomeric carbon (C1) in the oxocarbenium ion can be appreciated. The C1-O_P bond is completely cleaved and the C1-O1' bond is not formed, as can be seen from the sp^3 lone pairs on the O_P and O1' atoms. The acceptor hydrogen (H) has not been transferred to the phosphate, but there is a clear charge transfer from O_P to H that stabilizes the leaving group.
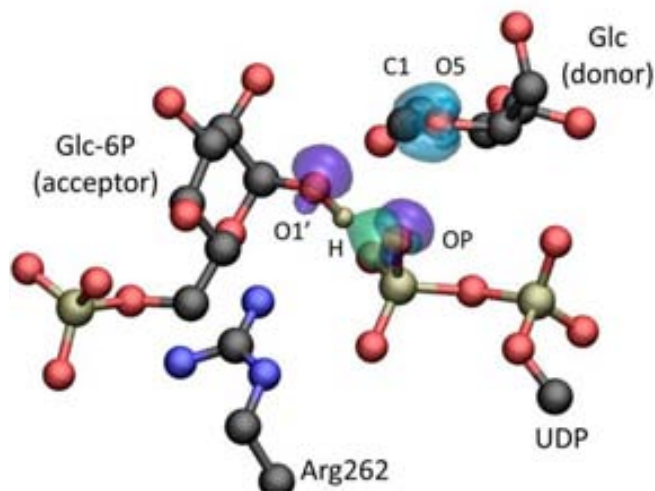
Figure VII - 12: Analysis of the electronic structure of the ion-pair species as localized Wannier orbitals. In blue is the double bond formed between the p orbitals of C1 and O5 atoms (on the upper right side of the figure). In green is the sp^3 orbital of O_P to which the hydrogen is transferred. In purple are the sp3 orbitals that leave/attack the anomeric carbon (C1).

In summary, our QM/MM metadynamics simulations support a single displacement but two step type of mechanism with a very short lived ion-pair intermediate for the formation of trehalose-6-phosphate from UDP-Glc and Glc-6P catalyzed by OtsA.

This process, which can be regarded as a type of S_N1 with a) an extremely short-lived intermediate b) acid catalysis by the acceptor molecule and c) base catalysis by the leaving group, is in

remarkable agreement with the proposals of recent kinetic and structural studies on OtsA.[106, 107] Altogether, our results show that glycosyl transfer with retention of the anomeric configuration can take place in the absence of a protein nucleophile, provided that a hydrogen atom of the acceptor is properly oriented to assist the cleavage of the Glc-UDP bond. It would be extremely interesting to probe whether this mechanism is also possible for glycosyltransferases that, unlike OtsA, have a putative nucleophile residue near the anomeric carbon, such as α3GalT[100] and blood group transferases,[99] *i.e.* whether the mechanism highlighted in this study can compete with the proposed S_N2 reaction for these structurally different GTs. These studies are underway in our laboratory.

Conclusions

1. The computed Michaelis complex of trehalose-6-phosphate synthase obtained by CPMD/MM metadynamics starting from the X-ray structure of a products-like complex is in good agreement with the available experimental data.[235]

2. The glucosyl transfer reaction from Glc-UDP to Glc-6P in OtsA follows a very dissociative mechanism with an extremely short-lived ion-pair intermediate. The cleavage of the Glc-UDP bond is assisted by the acceptor molecule and the phosphate group acts as a general base after the formation of the glycosidic bond.

3. The oxocarbenium ion-like species is characterized by a large development of charge on the sp^2 hybridized anomeric carbon. This charge development is stabilized by the pyranic oxygen.

4. The glucosyl transfer reaction in OtsA can take place *via* a front-face mechanism in the absence of an enzymatic nucleophile. However, we cannot conclude on the mechanism for glycoside transferases in which a competent nucleophilic residue is present.

Chapter VIII - Summary and conclusions

Summary and conclusions

Ab initio and hybrid quantum mechanics / molecular mechanics simulations have been performed to study substrate binding and reactivity in carbohydrate active enzymes (glycosyl hydrolases and transferases) as well as the conformational flexibility in isolated monosaccharides. *In silico* modeling of such systems gives an atomistic and electronic description of the process "as-it-is". In one hand, the simulations are not limited by the requirement to use non-active mutants or inhibitors. On the other hand, they give an atomic-electronic description of the processes studied.

The reliability of the methodology used was tested against high resolution X-ray structural data of the Michaelis complex in GH8 endocellulase (*chapter I*). The computed structure was in excellent agreement with the experimental structures.

Subsequently, the effect of enzyme mutations and substrate modifications was studied. Only the hydrogen bond between the exocyclic group at position C2 and the nucleophilic residue was found to affect the subsite -1 ring conformation. When this hydrogen bond is prevented (either by using a Withers fluoro inhibitor or by glutamate to isoleucine mutation), the substrate distortion is completely lost in favor of the $^4C_1$ conformation (see *chapter II*). Experimental X-ray structures of the Michaelis complex of a glucosidase with a 2-deoxy-2-fluoro inhibitor with a bad leaving group aglycon have been proposed to verify our prediction.

Substrate distortion in glycoside hydrolases is known to be very important in terms of substrate preactivation and reaction rate enhancement. Therefore, the reaction mechanism and the conformational itinerary of a non-distorted substrate might be significantly different from a preactivated one. To this end, the glycosylation step of the hydrolysis reaction of the non-distorted 2-deoxy-2-fluoro inhibitor was simulated by means of CPMD/MM metadynamics simulations (see *chapter III*). Indeed, the conformational itinerary of the substrate during the reaction showed to be circular ($^4C_1 \rightarrow {}^4H_3 \rightarrow {}^4C_1$) instead of the usual longitudinal itinerary ($^1S_3 \rightarrow {}^4H_3 \rightarrow {}^4C_1$) predicted for the natural substrate.

Because the substrate structure in the Michaelis complex, which dictates the conformational itinerary of the substrate during the reaction, can be predicted from the analysis of the conformational flexibility of isolated sugars, in *chapter IV* we studied the conformational free energy surface of β-D-manopyranose and a α-L-fucopyranose. We introduce a simple preactivation index integrating several structural, electronic, and energetic properties that can be used to predict the conformation of the substrate in the Michaelis complex of any GH.

The glycosidic bond formation in retaining glycoside transferases was simulated in *chapter VII*. The results are in agreement with a front-face single displacement mechanism, solving historical controversies on the feasibility of such mechanism.

The main conclusions reached in the present Thesis are the following:

❖ Three of the common strategies used to trap the Michaelis complex of GHs [namely: i) mutation of the acid/base catalytic residue; ii) use of a thio-glycoside inhibitor; or iii) use of a Withers fluoro inactivator at low pH conditions], do not affect the distortion of the substrate significantly.

❖ The subsite -1 glucose ring structure of a retaining β-glucosidase in complex with a 2-deoxy-2-fluoro inhibitor at neutral pH conditions adopts a non-distorted $^4C_1$ conformation.

❖ The subsite -1 glucose ring structure of a retaining 1,3-1,4-β-endoglucosidase in which the nucleophilic residue has been mutated to isoleucine adopts a non-distorted $^4C_1$ conformation. Glutamine, aspartate, and alanine mutants of the same residue result in a $^1S_3$, $^1S_5$ and $^1S_5$ distortions respectively.

❖ The 2-deoxy-2-fluoro glucose derivative follows a cyclic conformational itinerary ($^4C_1 \rightarrow {}^4H_3 \rightarrow {}^4C_1$) during the glycosylation step of the enzymatic hydrolysis. This is in contrast to the longitudinal ($^1S_3 \rightarrow {}^4H_3 \rightarrow {}^4C_1$) itinerary that exhibits the natural substrate.

❖ The conformational FES of β-D-mannopyranose and α-L-fucopyranose gives support to the proposed conformational itineraries proposed for the glycosylation reaction in retaining β-mannosidases and α-fucosidases ($^1S_5 \rightarrow B_{2,5} \rightarrow {}^OS_2$ and $^1C_4 \rightarrow {}^3H_4 \rightarrow {}^3S_1$ respectively).

❖ Ghs bind substrates in specific conformations that preactivate them for catalysis (structurally and electronically). The degree of preactivation can be estimated by *ab initio* calculations of the isolated sugar ring, thereby predicting the distortion in the Michaelis complex. The intrinsic properties of the isolated sugar have exerted an important evolutionary pressure in glycoside hydrolases to induce a particular distortion to the substrate.

❖ The α,α-(1→1) linkage formation in trehalose 6-phosphate synthase follows a very dissociative mechanism with a very short lived oxocarbenium ion intermediate.

❖ The cleavage of the Glc-UDP bond is assisted by the acceptor molecule and the phosphate group acts as a general base after the formation of the glycosidic bond.

❖ The driving force for the reaction is the cleavage of the glycoside-phosphate bond with the formation of a key hydrogen bond between the acceptor and the leaving phosphate.

❖ The mechanism can be regarded as a type of S_N1 reaction, with the presence of a very short lived oxocarbenium ion intermediate. Nevertheless, these results are not generalizable to all glycoside transferases, and other enzymes that show a putative nucleophilic residue might undergo a different mechanism.

Publications and presentations in congresses

Publications

The present thesis has given rise to the following publications:

- van Bueren, A.L.; <u>Ardèvol, A.</u>; Fayes-Kerr, J.; Luo, B.; Zhang, Y.; Sollogoub, M.; Blériot, Y.; Rovira, C.; Davies, G.J. "Analysis of the Reaction Coordinate of α-L-Fucosidases: A Combined Structural and Quantum Mechanical Approach." *J. Am. Chem. Soc.* **2010** 132(6), 1804-1806.

- <u>Ardèvol, A.</u>; Biarnés, X.; Planas, A.; Rovira, C. "The Conformational Free-Energy Landscape of β-D-Mannopyranose: Evidence for a $^1S_5 \rightarrow B_{2,5} \rightarrow {}^OS_2$ Catalytic Itinerary in β-Mannosidases." *J. Am. Chem. Soc.* **2010** 132(45), 16058-16065

- <u>Ardèvol, A.</u>; Rovira, C. "The molecular mechanism of enzymatic glycosyl transfer with retention of configuration: evidence for a short-lived oxocarbenium ion-like species." *Angew. Chem. Int. Ed.* **2011**, 51(46), 10897-901.

- Ardèvol, A.; Rovira, C. "Mechanism of reaction in 1,3-1,4-β-endoglucanase from Bacillus with the 2-deoxy-2-fluoro substrate derivative." In preparation.

- Ardèvol, A.; Planas, A.; Rovira, C. "Influence of the enzyme-substrate interactions on substrate distortion". In preparation.

Besides, the work during the PhD training has given rise to other publications not included in this thesis:

- Biarnés, X.; <u>Ardèvol, A.</u>; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. "The Conformational Free Energy Landscape of β-d-Glucopyranose. Implications for Substrate Preactivation in β-Glucoside Hydrolases" *J. Am. Chem. Soc.* **2007** 129(35), 10686–10693

- Gallego, O.; Ruiz, F. X.; <u>Ardèvol, A.</u>; Domínguez, M.; Alvarez, R.; de Lera, A.R.; Rovira, C.; Farrés, J.; Fita, I.; Parés, X. "Structural basis for the high all-trans-retinaldehyde reductase activity of the tumor marker AKR1B10" *Proc. Natl. Acad. Sci. USA* **2007** 104(52), 20764-20769

- Ruiz, F. X.; Gallego, O.; <u>Ardevol, A.</u>; Moro, A.; Domínguez, M.; Alvarez, S.; Alvarez, R.; de Lera, A.R.; Rovira, C.; Fita, I.; Parés, X.; Farrés, J. "Aldo-keto reductases from the AKR1B subfamily: Retinoid specificity and control of cellular retinoic acid levels" *Chem. Biol. Inter.* **2009** 178(1-3), 171-177

- Petersen, L.; <u>Ardèvol, A.</u>; Rovira, C.; Reilly, P.J. "Mechanism of Cellulose Hydrolysis by Inverting GH8 Endoglucanases: A QM/MM Metadynamics Study" *J. Phys. Chem. B* **2009** 113(20), 7331-7339

- Biarnés, X.; <u>Ardèvol, A.</u>; Planas, A.; Rovira. C. "Substrate conformational changes in glycoside hydrolase catalysis. A first-principles molecular dynamics study." *Biocatal. Biotransform.* **2010** 28(1), 33-40.

- Petersen, P.; <u>Ardèvol, A.</u>; Rovira, C.; Reilly, P.J. "Molecular Mechanism of the Glycosylation Step Catalyzed by Golgi α-Mannosidase II: A QM/MM Metadynamics Investigation" *J. Am. Chem. Soc.* **2010** 132(24), 8291-8300.

- Ruiz, F. X.; Gallego, O.; Moro, A.; Gallego, O.; <u>Ardèvol, A.</u>; Rovira, C.; Petrash, J.M.; Parés, X.; Farrés, J. "Human and rodent aldo-keto reductases from the AKR1B subfamily and their specificity with retinaldehyde" **2011** *Chem. Biol. Interact.* 191(1-3), 199-205

- Ruiz, F.X.; Porté, S.; Gallego, O.; Moro, A.; <u>Ardevol, A.</u>; Del Rio, A.; Rovira, C.; Farrés, J.; Parés, X. "Retinaldehyde is a substrate for human aldo-keto reductases of the 1C subfamily" *Biochem. J.* **2011** 440(3), 335-44.

- Biarnés, X.; <u>Ardèvol, A.</u>; Planas, A.; Rovira, C. "Catalytic itinerary in 1,3-1,4-β-glucanase unravelled by QM/MM metadynamics. Charge is not yet fully developed at the oxocarbenium ion-like transition state." *J. Am. Chem. Soc.* **2011** Accepted (DOI: 10.1021/ja207113e)

Presentations in congresses

The work of this thesis has been presented in the following congresses:

- CPMD 2011. "How does nature make glycosidic bonds. A metadynamics investigation". **Invited talk.** Barcelona Science Park (Catalonia), September 2011

- 9th Carbohydrate Bioengineering Meeting (CBM9). **Poster presentation.** Technical University of Lisbon, Lisbon (Portugal) – May 2011

- XXVI meeting of the XRQTC. **Poster presentation.** Reference network on theoretical and computational chemistry, Barcelona (Spain) – July 2010

- Expanding the frontiers of molecular dynamics simulations in biology. **Poster presentation.** Barcelona Supercomputing Center - CNS, Barcelona (Spain) – November 2009

- Theoretical chemistry: Modeling reactivity from gas phase to bibomolecules and solids. **Poster presentation.** Reference Network on Theoretical and Computational Chemistry, Barcelona (Spain) – july 2009

- 8th Carbohydrate Bioengineering Meeting (CBM8). **Poster presentation.** Institute of Protein Biochemistry - CNR, Ischia (Italy) – May 2009

- CPMD 2008. **Poster presentation.** International Center for Theoretical Physics (ICTP), Trieste (Italy) – June 2008

- "Reaction mechanism in glycoside transferases & heme binding in heme oxigenase". **Invited seminar.** Institute of Multidisciplinary Research for Advanced Materials, Tohoku University (Japan), April 2008

- Understanding molecular dynamics tutorial. **Workshop & Poster presentation.** Centre Européen de Calcul Atomique et Moléculaire (CECAM), Amsterdam (Netherlands) – January 2008

- 7th Carbohydrate Bioengineering Meeting (CBM7). **Poster presentation.** Technical University of Braunschweig, Braunschweig (Germany) – April 2007

Appendix

Appendix

Cremer and Pople puckering coordinates

In 1975 Cremer and Pople introduced a puckering coordinates system to unambiguously assign the conformation of an N-atoms ring from the Cartesian coordinates of these atoms.[159] It is given a set of coordinates $\{R_j\}$ that satisfy equation A-1 (i.e. the origin is the geometrical center of the ring).

$$\sum_j^N R_j = 0 \qquad \text{Eq. A - 1}$$

To set up a system of puckering coordinates, it is desirable to specify the displacement of each nucleus from a suitably defined mean plane. This plane is chosen to pass through the geometrical center. The z axis is taken as orthogonal to the mean plane and the y axis as the one passing through the projection of nuclear position 1 onto this plane. Thus, the new atomic positions (x_j, y_j, z_j) are a simple linear transformation of the initial coordinates (X_j, Y_j, Z_j).

The orientation of the mean plane ($z = 0$) can be specified as follows:

The puckering with respect to the plane $z = 0$ can be partly described by the N coordinates z_j. Given eq. A-1 and the requirement that the new origin is at the geometrical center, then:

$$\sum_j^N z_j = 0 \qquad \text{Eq. A - 2}$$

We now impose two additional conditions

$$\sum_j^N z_j \cos\left[2\pi(j-1)/N\right] = 0 \qquad \text{Eq. A - 3}$$

$$\sum_j^N z_j \sin\left[2\pi(j-1)/N\right] = 0 \qquad \text{Eq. A - 4}$$

Equations A-3 and 4 are sufficient to fix the mean plane uniquely. In the special case of small puckering displacement of a regular planar polygon, they correspond to the condition that the displacements z_j are such as to lead to no overall angular momentum. However, the same conditions may be used more generally for finite displacements, non-equivalent atoms, and any

bond lengths and angles. Futher, it is easily confirmed that these conditions are the same whichever atom in the ring is number one.

The orientation of the mean plane can now be determined for the position vectors R_j in the following manner. Define new vectors

$$R' = \sum_{j}^{N} R_j \sin\ [2\pi(j-1)/N] \qquad \text{Eq. A - 5}$$

$$R'' = \sum_{j}^{N} R_j \cos\ [2\pi(j-1)/N] \qquad \text{Eq. A - 6}$$

Then the unit vector

$$n = R' \times R''/|R' \times R''| \qquad \text{Eq. A - 7}$$

will be perpendicular to R' and R''. This will be chosen as the molecular z axis. Since the components of R' and R'' along n are zero, it follows that eq A-3 and 4 are satisfied. The positive direction of n defines a "topside" of the ring (above the face with clockwise numbering).

The components of the unit vector n with respect to the space-fixed axes may be obtained directly from the components (X_j, Y_j, Z_j) of R_j using eq. A-5, 6 and 7. The full set of displacements from the mean plane are then given by the scalar products

$$z_j = R_j \cdot n \qquad \text{Eq. A - 8}$$

These will automatically satisfy equations A-2, 3 and 4.

We may now define generalized ring-puckering coordinates in the following manner. Considering that the number of atoms of the ring is $N = 6$ (as it is in a pyranose), then we define two amplitudes q_m and $q_{N/2}$, and a phase φ_m ($m = 2$) by:

$$q_m \cos \phi_m = \sqrt{2/N} \sum_j^N z_j \cos [2\pi m(j-1)/N] \qquad \text{Eq. A - 9}$$

$$q_m \sin \phi_m = -\sqrt{2/N} \sum_j^N z_j \sin [2\pi m(j-1)/N] \qquad \text{Eq. A - 10}$$

$$q_{N/2} = \frac{1}{\sqrt{N}} \sum_j^N z_j \cos [(j-1)\pi] = \frac{1}{\sqrt{N}} \sum_j^N (-1)^{j-1} z_j \qquad \text{Eq. A - 11}$$

Therefore, the number of puckering coordinates to describe the conformation of a six membered ring is 3 (N - 3). Equations A-2 to 4 and A-9 to 11 constitute a set of N linear equations for the N displacements z_j. They may be solved to give expressions for z_j in terms of the puckering coordinates q_2, φ_2 and q_3. The result is that:

$$z_j = \sqrt{2/N} \sum_{m=2}^{N-1} q_m \cos[\phi_m + 2\pi m(j-1)/N] + \qquad \text{Eq. A - 12}$$

The normalization factors in the definitions in eq. A-9 to 11 are such that

$$\sum_j^N z_j = q_2{}^2 + q_3{}^2 = Q^2 \qquad \text{Eq. A - 13}$$

And the quantity $Q \geq 0$ may be termed a *total puckering amplitude*.

Finally, it is useful to replace[239, 240] the q_2, φ_2 and q_3 coordinates by a "spherical polar set" (Q, θ, and ɸ) such as:

$$\begin{cases} q_2 = Q \sin \theta \\ q_3 = Q \cos \theta \end{cases} \qquad \text{Eq. A - 14}$$

Then these polar coordinates can be expressed as a function of the atomic displacements from the mean plane z_j as:

$$Q \sin \theta \cos \phi = +\sqrt{1/3} \sum_{j}^{6} z_j \cos\left[\frac{2\pi}{6} 2(j-1)\right] \qquad \text{Eq. A - 15}$$

$$Q \sin \theta \sin \phi = -\sqrt{1/3} \sum_{j}^{6} z_j \sin\left[\frac{2\pi}{6} 2(j-1)\right] \qquad \text{Eq. A - 16}$$

$$Q \cos \theta = \sqrt{1/6} \sum_{j}^{6} (-1)^{j-1} z_j \qquad \text{Eq. A - 17}$$

This coordinate system permits the mapping of all types of puckering on the surface of the "Cremer and Pople puckering sphere" (figure A-1). In this framework, the polar positions ($\theta = 0$, π) correspond to a chair conformation ($^1C_4$ and $^4C_1$ respectively - in this work). Similarly, the positions on the equator ($\theta = \pi/2$) contains six boat and skew conformations separated at constant intervals ($\pi/6$) of ϕ. The conversion from the chair to the skew/boat conformations involves a half-chair/envelope that are located at a value of $\theta = \pi/4, 3\pi/4$.
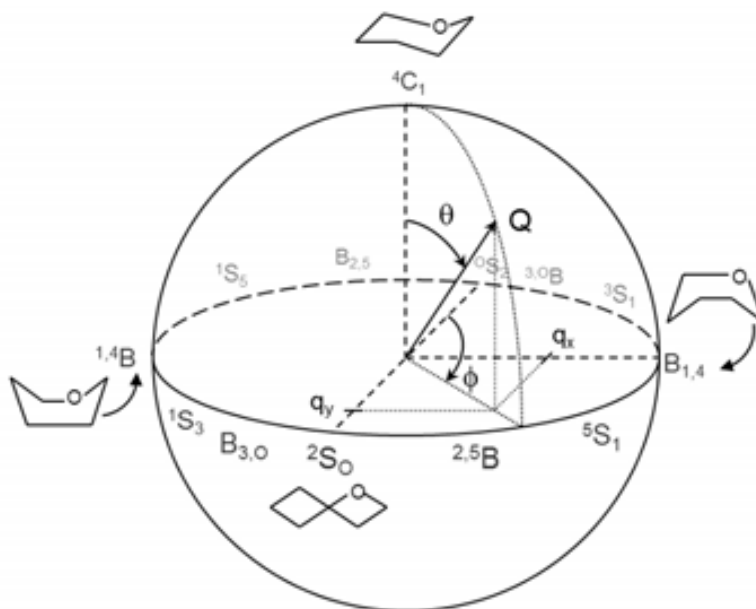

Figure A - 1: Cremer and Pople puckering coordinates Q, θ, and φ for a six membered ring and the collective variables q_x and q_y used in the metadynamics run.

Bibliography

Bibliography

Bibliography

1.   X. Biarnes, A. Ardevol, J. Iglesias-Fernandez, A. Planas and C. Rovira, "Catalytic Itinerary in 1,3-1,4-beta-Glucanase Unraveled by QM/MM Metadynamics. Charge Is Not Yet Fully Developed at the Oxocarbenium Ion-like Transition State." *J Am Chem Soc* **2011**.

2.   S. Hurtley, R. Service and P. Szuromi, "Cinderella's coach is ready." *Science* **2001,** *291*, 2337-2337.

3.   N. E. Zachara and G. W. Hart, "The emerging significance of O-GlcNAc in cellular regulation." *Chemical reviews* **2002,** *102*, 431-438.

4.   H. Li and M. d'Anjou, "Pharmacological significance of glycosylation in therapeutic proteins." *Current opinion in biotechnology* **2009,** *20*, 678-84.

5.   G. Davies and B. Henrissat, "Structures and mechanisms of glycosyl hydrolases." *Structure* **1995,** *3*, 853-859.

6.   B. Henrissat and A. Bairoch, "Updating the sequence-based classification of glycosyl hydrolases." *The Biochemical journal* **1996,** *316 (Pt 2)*, 695-6.

7.   B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat, "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics." *Nucleic acids research* **2009,** *37*, D233-8.

8.   G. J. Davies, K. S. Wilson and B. Henrissat, "Nomenclature for sugar-binding subsites in glycosyl hydrolases." *The Biochemical journal* **1997,** *321 (Pt 2)*, 557-9.

9.   R. Wolfenden, X. D. Lu and G. Young, "Spontaneous hydrolysis of glycosides." *J. Am. Chem. Soc.* **1998,** *120*, 6814-6815.

10.   J. D. McCarter and S. G. Withers, "Mechanisms of enzymatic glycoside hydrolysis." *Current opinion in structural biology* **1994,** *4*, 885-92.

11.   C. S. Rye and S. G. Withers, "Glycosidase mechanisms." *Curr Opin Chem Biol* **2000,** *4*, 573-80.

12.   D. L. Zechel and S. G. Withers, "Glycosidase mechanisms: anatomy of a finely tuned catalyst." *Accounts of chemical research* **2000,** *33*, 11-8.

13.   A. Vasella, G. J. Davies and M. Bohm, "Glycosidase mechanisms." *Curr Opin Chem Biol* **2002,** *6*, 619-29.

14.   D. E. Koshland, "Stereochemistry and the mechanism of enzymatic reactions." *Biol. Rev. Cambridge Philosophic. Soc.* **1953,** *28*, 416-436.

15.   J. F. Stoddart, *Stereochemistry of carbohydrates*. Wiley-Interscience: New York, 1971; p xi, 249 p.

16.   "IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Conformational nomenclature for five and six-membered ring forms of monosaccharides and their derivatives: recommendations 1980." *European journal of biochemistry / FEBS* **1980,** *111*, 295-8.

17.   M. L. Sinnott and I. J. Souchard, "The mechanism of action of beta-galactosidase. Effect of aglycone nature and -deuterium substitution on the hydrolysis of aryl galactosides." *The Biochemical journal* **1973,** *133*, 89-98.

18.	T. Selwood and M. L. Sinnott, "A solvent-isotope-effect study of proton transfer during catalysis by Escherichia coli (lacZ) beta-galactosidase." *The Biochemical journal* **1990,** *268*, 317-23.

19.	V. M. A. Ducros, D. L. Zechel, G. N. Murshudov, H. J. Gilbert, L. Szabo, D. Stoll, S. G. Withers and G. J. Davies, "Substrate distortion by a beta-mannanase: Snapshots of the Michaelis and covalent-intermediate complexes suggest a B2,5 conformation for the transition state." *Angewandte Chemie-International Edition* **2002,** *41*, 2824-+.

20.	R. L. Nath and H. N. Rydon, "The influence of structure on the hydrolysis of substituted phenyl beta-D-glucosides by emulsin." *Biochem. J.* **1954,** *57*, 1-10.

21.	C. S. Jones and D. J. Kosman, "Purification, properties, kinetics, and mechanism of beta-N-acetylglucosamidase from Aspergillus niger." *The Journal of biological chemistry* **1980,** *255*, 11861-9.

22.	M. M. Mader and P. A. Bartlett, "Binding Energy and Catalysis: The Implications for Transition-State Analogs and Catalytic Antibodies." *Chemical reviews* **1997,** *97*, 1281-1302.

23.	G. Sulzenbacher, H. Driguez, B. Henrissat, M. Schulein and G. J. Davies, "Structure of the Fusarium oxysporum endoglucanase I with a nonhydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group." *Biochemistry* **1996,** *35*, 15280-7.

24.	I. Tews, A. Perrakis, A. Oppenheim, Z. Dauter, K. S. Wilson and C. E. Vorgias, "Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease." *Nat Struct Biol* **1996,** *3*, 638-48.

25.	G. J. Davies, V. M. Ducros, A. Varrot and D. L. Zechel, "Mapping the conformational itinerary of beta-glycosidases by X-ray crystallography." *Biochemical Society transactions* **2003,** *31*, 523-7.

26.	L. Petersen, A. Ardevol, C. Rovira and P. J. Reilly, "Mechanism of cellulose hydrolysis by inverting GH8 endoglucanases: a QM/MM metadynamics study." *The journal of physical chemistry. B* **2009,** *113*, 7331-9.

27.	I. J. Barker, L. Petersen and P. J. Reilly, "Mechanism of xylobiose hydrolysis by GH43 beta-xylosidase." *The journal of physical chemistry. B* **2010,** *114*, 15389-93.

28.	L. Petersen, A. Ardevol, C. Rovira and P. J. Reilly, "Molecular mechanism of the glycosylation step catalyzed by Golgi alpha-mannosidase II: a QM/MM metadynamics investigation." *J Am Chem Soc* **2010,** *132*, 8291-300.

29.	A. L. Bowman, I. M. Grant and A. J. Mulholland, "QM/MM simulations predict a covalent intermediate in the hen egg white lysozyme reaction with its natural substrate." *Chem Commun (Camb)* **2008**, 4425-7.

30.	N. F. Bras, S. A. Moura-Tamames, P. A. Fernandes and M. J. Ramos, "Mechanistic studies on the formation of glycosidase-substrate and glycosidase-inhibitor covalent intermediates." *Journal of computational chemistry* **2008,** *29*, 2565-74.

31.	J. Jitonnom, V. S. Lee, P. Nimmanpipug, H. A. Rowlands and A. J. Mulholland, "Quantum mechanics/molecular mechanics modeling of substrate-assisted catalysis in family 18 chitinases: conformational changes and the role of Asp142 in catalysis in ChiB." *Biochemistry* **2011,** *50*, 4697-711.

32.　　G. J. Davies, L. Mackenzie, A. Varrot, M. Dauter, A. M. Brzozowski, M. Schulein and S. G. Withers, "Snapshots along an enzymatic reaction coordinate: analysis of a retaining beta-glycoside hydrolase." *Biochemistry* **1998,** *37*, 11707-13.

33.　　D. E. Koshland, "The key-lock theory and the induced-fit theory." *Angew. Chem.-Int. Edit. Engl.* **1995,** *33*, 2375-2378.

34.　　X. Biarnes, J. Nieto, A. Planas and C. Rovira, "Substrate distortion in the Michaelis complex of Bacillus 1,3-1,4-beta-glucanase. Insight from first principles molecular dynamics simulations." *The Journal of biological chemistry* **2006,** *281*, 1432-41.

35.　　X. Biarnes, A. Ardevol, A. Planas, C. Rovira, A. Laio and M. Parrinello, "The conformational free energy landscape of beta-D-glucopyranose. implications for substrate preactivation in beta-glucoside hydrolases." *J. Am. Chem. Soc.* **2007,** *129*, 10686-10693.

36.　　A. J. Kirby, "Stereoelectronic effects on acetal hydrolysis." *Accounts of chemical research* **1984,** *17*, 305-311.

37.　　P. Deslongchamps, "Stereoelectronic control in cleavage of tetrahedral intermediates in hydrolysis of esters and amides." *Tetrahedron* **1975,** *31*, 2463-2490.

38.　　P. Deslongchamps, *Stereoelectronic effects in organic chemistry*. Pergamon: Oxford, 1983; p xi, 375 p.

39.　　V. L. Schramm, "Enzymatic transition states: thermodynamics, dynamics and analogue design." *Archives of biochemistry and biophysics* **2005,** *433*, 13-26.

40.　　D. J. Vocadlo and G. J. Davies, "Mechanistic insights into glycosidase chemistry." *Curr Opin Chem Biol* **2008,** *12*, 539-55.

41.　　G. J. Davies, A. Planas and C. Rovira, "Conformational Analyses of the Reaction Coordinate of Glycosidases." *Accounts of chemical research* **2011**.

42.　　D. M. van Aalten, D. Komander, B. Synstad, S. Gaseidnes, M. G. Peter and V. G. Eijsink, "Structural insights into the catalytic mechanism of a family 18 exo-chitinase." *Proceedings of the National Academy of Sciences of the United States of America* **2001,** *98*, 8979-84.

43.　　V. A. Money, N. L. Smith, A. Scaffidi, R. V. Stick, H. J. Gilbert and G. J. Davies, "Substrate distortion by a lichenase highlights the different conformational itineraries harnessed by related glycoside hydrolases." *Angew Chem Int Ed Engl* **2006,** *45*, 5136-40.

44.　　M. Czjzek, A. Ben David, T. Bravman, G. Shoham, B. Henrissat and Y. Shoham, "Enzyme-substrate complex structures of a GH39 beta-xylosidase from Geobacillus stearothermophilus." *Journal of molecular biology* **2005,** *353*, 838-46.

45.　　Y. He, M. S. Macauley, K. A. Stubbs, D. J. Vocadlo and G. J. Davies, "Visualizing the reaction coordinate of an O-GlcNAc hydrolase." *J Am Chem Soc* **2010,** *132*, 1807-9.

46.　　A. L. Lovering, S. S. Lee, Y. W. Kim, S. G. Withers and N. C. Strynadka, "Mechanistic and structural analysis of a family 31 alpha-glycosidase and its glycosyl-enzyme intermediate." *The Journal of biological chemistry* **2005,** *280*, 2105-15.

47.　　Y. W. Kim, A. L. Lovering, H. Chen, T. Kantner, L. P. McIntosh, N. C. Strynadka and S. G. Withers, "Expanding the thioglycoligase strategy to the synthesis of alpha-linked thioglycosides allows structural investigation of the parent enzyme/substrate complex." *J Am Chem Soc* **2006,** *128*, 2202-3.

48.   A. I. Guce, N. E. Clark, E. N. Salgado, D. R. Ivanen, A. A. Kulminskaya, H. Brumer, 3rd and S. C. Garman, "Catalytic mechanism of human alpha-galactosidase." *The Journal of biological chemistry* **2010,** *285*, 3625-32.

49.   L. Hosie and M. L. Sinnott, "Effects of deuterium substitution alpha and beta to the reaction centre, 18O substitution in the leaving group, and aglycone acidity on hydrolyses of aryl glucosides and glucosyl pyridinium ions by yeast alpha-glucosidase. A probable failure of the antiperiplanar-lone-pair hypothesis in glycosidase catalysis." *The Biochemical journal* **1985,** *226*, 437-46.

50.   D. M. A. Guerin, M. B. Lascombe, M. Costabel, H. Souchon, V. Lamzin, P. Beguin and P. M. Alzari, "Atomic (0.94 angstrom) resolution structure of an inverting glycosidase in complex with substrate." *Journal of molecular biology* **2002,** *316*, 1061-1069.

51.   J. Zou, G. J. Kleywegt, J. Stahlberg, H. Driguez, W. Nerinckx, M. Claeyssens, A. Koivula, T. T. Teeri and T. A. Jones, "Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Ce16A from trichoderma reesei." *Structure* **1999,** *7*, 1035-45.

52.   A. Varrot, S. Leydier, G. Pell, J. M. Macdonald, R. V. Stick, B. Henrissat, H. J. Gilbert and G. J. Davies, "Mycobacterium tuberculosis strains possess functional cellulases." *The Journal of biological chemistry* **2005,** *280*, 20181-4.

53.   G. Sidhu, S. G. Withers, N. T. Nguyen, L. P. McIntosh, L. Ziser and G. D. Brayer, "Sugar ring distortion in the glycosyl-enzyme intermediate of a family G/11 xylanase." *Biochemistry* **1999,** *38*, 5346-54.

54.   E. Sabini, G. Sulzenbacher, M. Dauter, Z. Dauter, P. L. Jorgensen, M. Schulein, C. Dupont, G. J. Davies and K. S. Wilson, "Catalysis and specificity in enzymatic glycoside hydrolysis: a 2,5B conformation for the glycosyl-enzyme intermediate revealed by the structure of the Bacillus agaradhaerens family 11 xylanase." *Chemistry & biology* **1999,** *6*, 483-92.

55.   A. Cartmell, E. Topakas, V. M. Ducros, M. D. Suits, G. J. Davies and H. J. Gilbert, "The Cellvibrio japonicus mannanase CjMan26C displays a unique exo-mode of action that is conferred by subtle changes to the distal region of the active site." *The Journal of biological chemistry* **2008,** *283*, 34403-13.

56.   L. E. Tailford, W. A. Offen, N. L. Smith, C. Dumon, C. Morland, J. Gratien, M. P. Heck, R. V. Stick, Y. Bleriot, A. Vasella, H. J. Gilbert and G. J. Davies, "Structural and biochemical evidence for a boat-like transition state in beta-mannosidases." *Nature chemical biology* **2008,** *4*, 306-12.

57.   S. Numao, D. A. Kuntz, S. G. Withers and D. R. Rose, "Insights into the mechanism of Drosophila melanogaster Golgi alpha-mannosidase II through the structural analysis of covalent reaction intermediates." *J. Biol. Chem.* **2003,** *278*, 48074-48083.

58.   W. Zhong, D. A. Kuntz, B. Ember, H. Singh, K. W. Moremen, D. R. Rose and G. J. Boons, "Probing the substrate specificity of Golgi alpha-mannosidase II by use of synthetic oligosaccharides and a catalytic nucleophile mutant." *J Am Chem Soc* **2008,** *130*, 8975-83.

59.   Y. Zhu, M. D. Suits, A. J. Thompson, S. Chavan, Z. Dinev, C. Dumon, N. Smith, K. W. Moremen, Y. Xiang, A. Siriwardena, S. J. Williams, H. J. Gilbert and G. J. Davies, "Mechanistic insights into a Ca2+-dependent family of alpha-mannosidases in a human gut symbiont." *Nature chemical biology* **2010,** *6*, 125-32.

60.   F. Vallee, K. Karaveg, A. Herscovics, K. W. Moremen and P. L. Howell, "Structural basis for catalysis and inhibition of N-glycan processing class I alpha 1,2-mannosidases." *The Journal of biological chemistry* **2000,** *275*, 41287-98.

61.   G. Sulzenbacher, C. Bignon, T. Nishimura, C. A. Tarling, S. G. Withers, B. Henrissat and Y. Bourne, "Crystal structure of Thermotoga maritima alpha-L-fucosidase. Insights into the catalytic mechanism and the molecular basis for fucosidosis." *The Journal of biological chemistry* **2004,** *279*, 13119-28.

62.   A. Lammerts van Bueren, A. Ardevol, J. Fayers-Kerr, B. Luo, Y. Zhang, M. Sollogoub, Y. Bleriot, C. Rovira and G. J. Davies, "Analysis of the reaction coordinate of alpha-L-fucosidases: a combined structural and quantum mechanical approach." *J Am Chem Soc* **2010,** *132*, 1804-6.

63.   L. L. Lairson, B. Henrissat, G. J. Davies and S. G. Withers, "Glycosyltransferases: structures, functions, and mechanisms." *Annual review of biochemistry* **2008,** *77*, 521-55.

64.   M. M. Palcic, Glycosyltransferases in glycobiology. In *Methods in Enzymology; Guide to techniques in glycobiology*, W. J. Lennarz and G. W. Hart, Eds. Academic Press, Inc., 1250 Sixth Ave., San Diego, California 92101, USA 14 Belgrave Square, 24-28 Oval Road, London NW1 70X, England, UK: 1994; Vol. 230, pp 300-316.

65.   X. Qian, K. Sujino, M. M. Palcic and R. M. Ratcliffe, *Glycosyltransferases in oligosaccharide synthesis*. Marcel Dekker AG, Hutgasse 4, CH-4001, Postfach 812, Basel, Switzerland 270 Madison Avenue, New York, NY, 10016, USA: 2001; p 535-565.

66.   I. Benz and M. A. Schmidt, "Never say never again: protein glycosylation in pathogenic bacteria." *Molecular microbiology* **2002,** *45*, 267-76.

67.   P. L. DeAngelis, "Microbial glycosaminoglycan glycosyltransferases." *Glycobiology* **2002,** *12*, 9R-16R.

68.   N. Markine-Goriaynoff, L. Gillet, J. L. Van Etten, H. Korres, N. Verma and A. Vanderplasschen, "Glycosyltransferases encoded by viruses." *J. Gen. Virol.* **2004,** *85*, 2741-2754.

69.   B. P. Dias, T. Ueda-Nakamura, C. H. Lopez, L. T. Tsuneto, B. A. Abreu and C. V. Nakamura, "Cell surface glycoproteins in Crithidia deanei: Influence of the endosymbiont." *Acta Protozool.* **2005,** *44*, 13-18.

70.   P. Talbot, B. D. Shur and D. G. Myles, "Cell adhesion and fertilization: steps in oocyte transport, sperm-zona pellucida interactions, and sperm-egg fusion." *Biology of reproduction* **2003,** *68*, 1-9.

71.   S. Hakomori, "Glycosylation defining cancer malignancy: new wine in an old bottle." *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99*, 10231-3.

72.   D. J. Becker and J. B. Lowe, "Fucose: biosynthesis and biological function in mammals." *Glycobiology* **2003,** *13*, 41R-53R.

73.   P. M. Rudd, T. Elliott, P. Cresswell, I. A. Wilson and R. A. Dwek, "Glycosylation and the immune system." *Science* **2001,** *291*, 2370-6.

74.   W. R. Scheible and M. Pauly, "Glycosyltransferases and cell wall biosynthesis: novel players and insights." *Current opinion in plant biology* **2004,** *7*, 285-95.

75.   K. Keegstra and N. Raikhel, "Plant glycosyltransferases." *Current opinion in plant biology* **2001,** *4*, 219-24.

76.   P. J. Eastmond, A. J. van Dijken, M. Spielman, A. Kerr, A. F. Tissier, H. G. Dickinson, J. D. Jones, S. C. Smeekens and I. A. Graham, "Trehalose-6-phosphate synthase 1, which catalyses the first step in trehalose synthesis, is essential for Arabidopsis embryo maturation." *The Plant journal : for cell and molecular biology* **2002,** *29*, 225-35.

77.   O. J. Goddijn and K. van Dun, "Trehalose metabolism in plants." *Trends in plant science* **1999,** *4*, 315-319.

78.   S. Zuccotti, D. Zanardi, C. Rosano, L. Sturla, M. Tonetti and M. Bolognesi, "Kinetic and crystallographic analyses support a sequential-ordered bi bi catalytic mechanism for Escherichia coli glucose-1-phosphate thymidylyltransferase." *Journal of molecular biology* **2001,** *313*, 831-43.

79.   S. J. Charnock and G. J. Davies, "Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms." *Biochemistry* **1999,** *38*, 6380-5.

80.   K. Persson, H. D. Ly, M. Dieckelmann, W. W. Wakarchuk, S. G. Withers and N. C. Strynadka, "Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs." *Nat Struct Biol* **2001,** *8*, 166-75.

81.   U. M. Unligil, S. Zhou, S. Yuwaraj, M. Sarkar, H. Schachter and J. M. Rini, "X-ray crystal structure of rabbit N-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily." *The EMBO journal* **2000,** *19*, 5269-80.

82.   A. Vrielink, W. Ruger, H. P. Driessen and P. S. Freemont, "Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose." *The EMBO journal* **1994,** *13*, 3413-22.

83.   S. Ha, D. Walker, Y. Shi and S. Walker, "The 1.9 A crystal structure of Escherichia coli MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis." *Protein science : a publication of the Protein Society* **2000,** *9*, 1045-52.

84.   R. P. Gibson, J. P. Turkenburg, S. J. Charnock, R. Lloyd and G. J. Davies, "Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA." *Chemistry & biology* **2002,** *9*, 1337-46.

85.   C. P. Chiu, A. G. Watts, L. L. Lairson, M. Gilbert, D. Lim, W. W. Wakarchuk, S. G. Withers and N. C. Strynadka, "Structural analysis of the sialyltransferase CstII from Campylobacter jejuni in complex with a substrate analog." *Nature structural & molecular biology* **2004,** *11*, 163-70.

86.   L. N. Gastinel, C. Bignon, A. K. Misra, O. Hindsgaul, J. H. Shaper and D. H. Joziasse, "Bovine alpha1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltransferases." *The EMBO journal* **2001,** *20*, 638-49.

87.   E. Boix, Y. Zhang, G. J. Swaminathan, K. Brew and K. R. Acharya, "Structural basis of ordered binding of donor and acceptor substrates to the retaining glycosyltransferase, alpha-1,3-galactosyltransferase." *The Journal of biological chemistry* **2002,** *277*, 28310-8.

88.   S. Baskaran, P. J. Roach, A. A. DePaoli-Roach and T. D. Hurley, "Structural basis for glucose-6-phosphate activation of glycogen synthase." *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107*, 17563-8.

89.   B. W. Murray, V. Wittmann, M. D. Burkart, S. C. Hung and C. H. Wong, "Mechanism of human alpha-1,3-fucosyltransferase V: glycosidic cleavage occurs prior to nucleophilic attack." *Biochemistry* **1997,** *36*, 823-31.

90.   L. L. Lairson and S. G. Withers, "Mechanistic analogies amongst carbohydrate modifying enzymes." *Chem Commun (Camb)* **2004**, 2243-8.

91.   S. C. Kim, A. N. Singh and F. M. Raushel, "Analysis of the galactosyltransferase reaction by positional isotope exchange and secondary deuterium isotope effects." *Archives of biochemistry and biophysics* **1988,** *267*, 54-59.

92.   M. D. Burkart, S. P. Vincent, A. Duffels, B. W. Murray, S. V. Ley and C. H. Wong, "Chemo-enzymatic synthesis of fluorinated sugar nucleotide: useful mechanistic probes for glycosyltransferases." *Bioorganic & medicinal chemistry* **2000,** *8*, 1937-46.

93.   M. Krupicka and I. Tvaroska, "Hybrid Quantum Mechanical/Molecular Mechanical Investigation of the beta-1,4-Galactosyltransferase-I Mechanism." *Journal of Physical Chemistry B* **2009,** *113*, 11314-11319.

94.   S. Kozmon and I. Tvaroska, "Catalytic mechanism of glycosyltransferases: Hybrid quantum mechanical/molecular mechanical study of the inverting N-acetylglucosaminyltransferase I." *J. Am. Chem. Soc.* **2006,** *128*, 16921-16927.

95.   B. L. Mark, D. J. Vocadlo, S. Knapp, B. L. Triggs-Raine, S. G. Withers and M. N. James, "Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase." *The Journal of biological chemistry* **2001,** *276*, 10330-7.

96.   B. L. Mark and M. N. G. James, "Anchimeric assistance in hexosaminidases." *Can. J. Chem.-Rev. Can. Chim.* **2002,** *80*, 1064-1074.

97.   L. L. Lairson, C. P. Chiu, H. D. Ly, S. He, W. W. Wakarchuk, N. C. Strynadka and S. G. Withers, "Intermediate trapping on a mutant retaining alpha-galactosyltransferase identifies an unexpected aspartate residue." *The Journal of biological chemistry* **2004,** *279*, 28339-44.

98.   S. I. Patenaude, N. O. Seto, S. N. Borisova, A. Szpacenko, S. L. Marcus, M. M. Palcic and S. V. Evans, "The structural basis for specificity in human ABO(H) blood group biosynthesis." *Nat Struct Biol* **2002,** *9*, 685-90.

99.   N. Soya, Y. Fang, M. M. Palcic and J. S. Klassen, "Trapping and characterization of covalent intermediates of mutant retaining glycosyltransferases." *Glycobiology* **2011,** *21*, 547-52.

100.   A. Monegal and A. Planas, "Chemical rescue of alpha3-galactosyltransferase. Implications in the mechanism of retaining glycosyltransferases." *J Am Chem Soc* **2006,** *128*, 16030-1.

101.   Y. Zhang, G. J. Swaminathan, A. Deshpande, E. Boix, R. Natesh, Z. Xie, K. R. Acharya and K. Brew, "Roles of individual enzyme-substrate interactions by alpha-1,3-galactosyltransferase in catalysis and specificity." *Biochemistry* **2003,** *42*, 13512-21.

102.   C. Breton, L. Snajdrova, C. Jeanneau, J. Koca and A. Imberty, "Structures and mechanisms of glycosyltransferases." *Glycobiology* **2006,** *16*, 29R-37R.

103.   G. J. Davies, "Sweet secrets of synthesis." *Nature Structural Biology* **2001,** *8,* 98-100.

104.   I. Andre, I. Tvaroska and J. P. Carver, "On the reaction pathways and determination of transition-state structures for retaining alpha-galactosyltransferases." *Carbohydr Res* **2003,** *338*, 865-77.

105.   C. Goedl and B. Nidetzky, "Sucrose phosphorylase harbouring a redesigned, glycosyltransferase-like active site exhibits retaining glucosyl transfer in the absence of a covalent intermediate." *Chembiochem : a European journal of chemical biology* **2009,** *10*, 2333-7.

106.   J. C. Errey, S. S. Lee, R. P. Gibson, C. M. Fleites, C. S. Barry, P. M. J. Jung, A. C. O'Sullivan, B. G. Davis and G. J. Davies, "Mechanistic Insight into Enzymatic Glycosyl Transfer with Retention of Configuration through Analysis of Glycomimetic Inhibitors." *Angewandte Chemie-International Edition* **2010,** *49*, 1234-1237.

107.   S. S. Lee, S. Y. Hong, J. C. Errey, A. Izumi, G. J. Davies and B. G. Davis, "Mechanistic evidence for a front-side, SNi-type reaction in a retaining glycosyltransferase." *Nature chemical biology* **2011,** *7*, 631-8.

108.   P. Hohenberg and W. Kohn, "Inhomogeneous electron gas." *Physical Review B* **1964,** *136*, B864-&.

109.   W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects." *Physical Review* **1965,** *140*, 1133-&.

110.   D. H. Marx, J, "Modern methods and algorithms of quantum chemistry." *NIC, FZ Jülich* **2000**.

111.   G. U. Nienhaus, *Protein-ligand interactions : methods and applications*. Humana Press: Totowa, N.J., 2005; p xi, 568 p.

112.   R. Car and M. Parrinello, "Unified approach for molecular dynamics and density-functional theory." *Phys Rev Lett* **1985,** *55*, 2471-2474.

113.   D. R. Hamann, "Generalized norm-conserving pseudopotentials." *Physical review. B, Condensed matter* **1989,** *40*, 2980-2987.

114.   N. Troullier and J. L. Martins, "Efficient pseudopotentials for plane-wave calculations." *Physical review. B, Condensed matter* **1991,** *43*, 1993-2006.

115.   A. Warshel and M. Levitt, "Theoretical studies of enzymatic reactioins - Dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme." *Journal of molecular biology* **1976,** *103*, 227-249.

116.   A. Laio, J. VandeVondele and U. Rothlisberger, "A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations." *Journal of Chemical Physics* **2002,** *116*, 6941-6947.

117.   A. Laio, J. VandeVondele and U. Rothlisberger, "D-RESP: Dynamically generated electrostatic potential derived charges from quantum mechanics/molecular mechanics simulations." *Journal of Physical Chemistry B* **2002,** *106*, 7300-7307.

118.   O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger and D. Sebastiani, "Variational optimization of effective atom centered potentials for molecular properties." *The Journal of chemical physics* **2005,** *122*, 14113.

119.   E. Darve, D. Rodriguez-Gomez and A. Pohorille, "Adaptive biasing force method for scalar and vector free energy calculations." *Journal of Chemical Physics* **2008,** *128*.

120.   G. N. Patey and J. P. Valleau, "Monte-Carlo method for obtaining interionic potential of mean force in ionic solution." *Journal of Chemical Physics* **1975,** *63*, 2334-2339.

121.   H. Grubmuller, "Predicting slow structural transitions in macromolecular systems: Conformational flooding." *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics* **1995,** *52*, 2893-2906.

122.   M. Sprik and G. Ciccotti, "Free energy from constrained molecular dynamics." *Journal of Chemical Physics* **1998,** *109*, 7737-7744.

123. E. A. Carter, G. Ciccotti, J. T. Hynes and R. Kapral, "Constrained reaction coordinate dynamics for the simulation of rare events." *Chemical Physics Letters* **1989,** *156*, 472-477.

124. B. Ensing, M. De Vivo, Z. Liu, P. Moore and M. L. Klein, "Metadynamics as a tool for exploring free energy landscapes of chemical reactions." *Accounts of chemical research* **2006,** *39*, 73-81.

125. A. Laio and F. L. Gervasio, "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science." *Reports on Progress in Physics* **2008,** *71*, 126601.

126. F. Wang and D. P. Landau, "Efficient, multiple-range random walk algorithm to calculate the density of states." *Phys Rev Lett* **2001,** *86*, 2050-3.

127. D. Cvijovicacute and J. Klinowski, "Taboo search: an approach to the multiple minima problem." *Science* **1995,** *267*, 664-6.

128. T. Huber, A. E. Torda and W. F. van Gunsteren, "Local elevation: a method for improving the searching properties of molecular dynamics simulation." *Journal of computer-aided molecular design* **1994,** *8*, 695-708.

129. A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli and M. Parrinello, "Assessing the accuracy of metadynamics." *Journal of Physical Chemistry B* **2005,** *109*, 6714-6721.

130. G. Bussi, A. Laio and M. Parrinello, "Equilibrium free energies from nonequilibrium metadynamics." *Phys Rev Lett* **2006,** *96*, 090601.

131. M. Iannuzzi, A. Laio and M. Parrinello, "Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics." *Phys Rev Lett* **2003,** *90*, 238302.

132. B. Ensing, A. Laio, M. Parrinello and M. L. Klein, "A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions." *The journal of physical chemistry. B* **2005,** *109*, 6676-87.

133. A. Barducci, M. Bonomi and M. Parrinello, "Metadynamics." *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2011,** *1*, 826-843.

134. P. M. Alzari, H. Souchon and R. Dominguez, "The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from Clostridium thermocellum." *Structure* **1996,** *4*, 265-275.

135. Q. Yao, T. Sun, G. Chen and W. Liu, "Heterologous expression and site-directed mutagenesis of endoglucanase CelA from Clostridium thermocellum." *Biotechnology letters* **2007,** *29*, 1243-7.

136. P. Isorna, J. Polaina, L. Latorre-Garcia, F. J. Canada, B. Gonzalez and J. Sanz-Aparicio, "Crystal structures of Paenibacillus polymyxa beta-glucosidase B complexes reveal the molecular basis of substrate specificity and give new insights into the catalytic machinery of family I glycosidases." *Journal of molecular biology* **2007,** *371*, 1204-18.

137. M. Hrmova, J. N. Varghese, R. De Gori, B. J. Smith, H. Driguez and G. B. Fincher, "Catalytic mechanisms and reaction intermediates along the hydrolytic pathway of a plant beta-D-glucan glucohydrolase." *Structure* **2001,** *9*, 1005-16.

138. B. Smith, "Can thioglycosides imitate the oxonium intermediate in glycosyl hydrolases?" *Journal of Molecular Graphics and Modelling* **2003,** *22*, 151-159.

139.   C. Schou, G. Rasmussen, M. Schulein, B. Henrissat and H. Driguez, "4-thiocello-oligosaccharides. Their synthesis and use as inhibitors of cellulases." *Journal of Carbohydrate Chemistry* **1993,** *12*, 743-752.

140.   S. Armand, S. Drouillard, M. Schulein, B. Henrissat and H. Driguez, "A bifunctionalized fluorogenic tetrasaccharide as a substrate to study cellulases." *The Journal of biological chemistry* **1997,** *272*, 2709-13.

141.   G. Sulzenbacher, M. Schulein and G. J. Davies, "Structure of the endoglucanase I from Fusarium oxysporum: native, cellobiose, and 3,4-epoxybutyl beta-D-cellobioside-inhibited forms, at 2.3 A resolution." *Biochemistry* **1997,** *36*, 5902-11.

142.   A. Planas, "Bacterial 1,3-1,4beta-glucanases: Structure, function and protein engineering." *Biochimica et Biophysica Acta* **2000,** *1543*, 361-382.

143.   J. L. Viladot, E. de Ramon, O. Durany and A. Planas, "Probing the mechanism of Bacillus 1,3-1,4-beta-D-glucan 4-glucanohydrolases by chemical rescue of inactive mutants at catalytically essential residues." *Biochemistry* **1998,** *37*, 11332-11342.

144.   W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, "A 2nd generation force-fied for the simulation of proteins, nucleic acids and organic molecules." *J. Am. Chem. Soc.* **1995,** *117*, 5179-5197.

145.   K. N. Kirschner and R. J. Woods, "Solvent interactions determine carbohydrate conformation." *Proceedings of the National Academy of Sciences of the United States of America* **2001,** *98*, 10541-10545.

146.   K. N. Kirschner and R. J. Woods, "Quantum mechanical study of the nonbonded forces in water-methanol complexes." *Journal of Physical Chemistry A* **2001,** *105*, 4150-4155.

147.   M. Basma, S. Sundara, D. Calgan, T. Vernali and R. J. Woods, "Solvated ensemble averaging in the calculation of partial atomic charges." *Journal of computational chemistry* **2001,** *22*, 1125-1137.

148.   D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel and P. Kollman, "Amber, a package of computer programs for applying molecular mechanics, normal-mode analysis, molecular dynamics and free-energy calculations to simulate the structural and energetic properties of molecules." *Computer Physics Communications* **1995,** *91*, 1-41.

149.   C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atom charges - the RESP model " *Journal of Physical Chemistry* **1993,** *97*, 10269-10280.

150.   G. W. T. M. J. Frisch, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q.

Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, , "Gaussian 03, Revision C.02." **2004**.

151.   K. Gaedt and H.-D. Höltje, "Consistent valence force-field parameterization of bond lengths and angles with quantum chemical ab initio methods applied to some heterocyclic dopamine D3-receptor agonists." *Journal of computational chemistry* **1998,** *19*, 935-946.

152.   V. Hornak and C. Simmerling, "Generation of accurate protein loop conformations through low-barrier molecular dynamics." *Proteins-Structure Function and Genetics* **2003,** *51*, 577-590.

153.   W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, "Comparison of simple potential functions for simulating liquid water." *Journal of Chemical Physics* **1983,** *79*, 926-935.

154.   C. T. Lee, W. T. Yang and R. G. Parr, "Development of the colle-salvetti correlation energy formula into a functional of the electron density." *Physical Review B* **1988,** *37*, 785-789.

155.   A. D. Becke, "Density functional thermochemistry 3. The role of exact exchange." *Journal of Chemical Physics* **1993,** *98*, 5648-5652.

156.   A. D. Becke, "Density-functional thermochemistry .4. A new dynamical correlation functional and implications for exact-exchange mixing." *Journal of Chemical Physics* **1996,** *104*, 1040-1046.

157.   J. Ireta, J. Neugebauer and M. Scheffler, "On the accuracy of DFT for describing hydrogen bonds: Dependence on the bond directionality." *Journal of Physical Chemistry A* **2004,** *108*, 5692-5698.

158.   G. J. Martyna, M. L. Klein and M. Tuckerman, "Nosé-Hoover chains - the canonical ensamble via continous dynamics." *Journal of Chemical Physics* **1992,** *97*, 2635-2643.

159.   D. Cremer and J. A. Pople, "A general definition of ring puckering coordinates." *J. Am. Chem. Soc.* **1975,** *97*, 1354-1358.

160.   M. N. Namchuk and S. G. Withers, "Mechanism of Agrobacterium beta-glucosidase: Kinetic analysis of the role of noncovalent enzyme/substrate interactions." *Biochemistry* **1995,** *34*, 16194-16202.

161.   M. N. Namchuk, J. D. McCarter, A. Becalski, T. Andrews and S. G. Withers, "The role of sugar substituents in glycoside hydrolysis." *J. Am. Chem. Soc.* **2000,** *122*, 1270-1277.

162.   J. Zhu and A. J. Bennet, "Solvolyses of 2-deoxy-alpha- and beta-D-glucopyranosyl 4 '-bromoisoquinolinium tetrafluoroborates." *Journal of Organic Chemistry* **2000,** *65*, 4423-4430.

163.   S. G. Withers, I. P. Street, P. Bird and D. H. Dolphin, "2-deoxy-2-fluoroglucosides. A novel class of mechanism-based glucosidase inhibitors." *J. Am. Chem. Soc.* **1987,** *109*, 7530-7531.

164.   A. White, D. Tull, K. Johns, S. G. Withers and D. R. Rose, "Crystallographic observation of a covalent catalytic intermediate in a beta-glycosidase." *Nature Structural Biology* **1996,** *3*, 149-154.

165.   V. Notenboom, C. Birsan, M. Nitz, D. R. Rose, R. A. J. Warren and S. G. Withers, "Insights into transition state stabilization of the beta-1,4-glycosidase Cex by covalent intermediate accumulation in active site mutants." *Nature Structural Biology* **1998,** *5*, 812-818.

166.   Y. K. Zhang, T. S. Lee and W. T. Yang, "A pseudobond approach to combining quantum mechanical and molecular mechanical methods." *Journal of Chemical Physics* **1999,** *110*, 46-54.

167. Y. Zhang, J. Bommuswamy and M. L. Sinnott, "Kinetic Isotope Effect Study of Transition States for the Hydrolyses of .alpha.- and .beta.-Glucopyranosyl Fluorides." *J. Am. Chem. Soc.* **1994,** *116*, 7557-7563.

168. L. C. S. Melander and W. H. Saunders, *Reaction rates of isotopic molecules*. 2nd ed.; New York &c.,, 1980.

169. A. J. Bennet and M. L. Sinnott, "Complete kinetic isotope effect description of transition states for acid-catalyzed hydrolyses of methyl .alpha.- and .beta.-glucopyranosides." *J. Am. Chem. Soc.* **1986,** *108*, 7287-7294.

170. J. K. Lee, A. D. Bain and P. J. Berti, "Probing the transition states of four glucoside hydrolyses with C-13 kinetic isotope effects measured at natural abundance by NMR spectroscopy." *J. Am. Chem. Soc.* **2004,** *126*, 3769-3776.

171. J. M. Stubbs and D. Marx, "Glycosidic bond formation in aqueous solution: On the oxocarbenium intermediate." *J. Am. Chem. Soc.* **2003,** *125*, 10960-10962.

172. N. Buckley and N. J. Oppenheimer, "Reactions of Charged Substrates. 7. The Methoxymethyl Carbenium Ion Problem. 2. A Semiempirical Study of the Kinetic and Thermodynamic Stabilities of Linear and Cyclic Oxo- and Thiocarbenium Ions Generated from Aldehyde Hydrates, Hemiacetals, Acetals, and Methyl Ribosides and Glucosides." *The Journal of organic chemistry* **1996,** *61*, 8048-8062.

173. N. Buckley and N. J. Oppenheimer, "Reactions of Charged Substrates. 6. The Methoxymethyl Carbenium Ion Problem. 1. A Semiempirical Study of the Kinetic and Thermodynamic Stabilities of Linear and Cyclic Oxo- and Thiocarbenium Ions Generated from Pyridinium and Dimethylanilinium Ions." *The Journal of organic chemistry* **1996,** *61*, 8039-8047.

174. B. J. Smith, "A conformational study of 2-oxanol: Insight into the role of ring distortion on enzyme-catalyzed glycosidic bond cleavage." *J. Am. Chem. Soc.* **1997,** *119*, 2699-2706.

175. X. Liang, A. Montoya and B. S. Haynes, "Local site selectivity and conformational structures in the glycosidic bond scission of cellobiose." *The journal of physical chemistry. B* **2011,** *115*, 10682-91.

176. B. Capon, "Mechanism in carbohydrate chemistry." *Chemical reviews* **1969,** *69*, 407-&.

177. A. Berces, G. Enright, T. Nukada and D. M. Whitfield, "The conformational origin of the barrier to the formation of neighboring group assistance in glycosylation reactions: a dynamical density functional theory study." *J Am Chem Soc* **2001,** *123*, 5460-4.

178. T. L. Amyes and W. P. Jencks, "Lifetimes of oxocarbenium ions in aqueous solution from common ion inhibition of the solvolysis of alpha-azido ethers by added azide ion." *J. Am. Chem. Soc.* **1989,** *111*, 7888-7900.

179. N. S. Banait and W. P. Jencks, "Reactions of anionic nucleophiles with .alpha.-D-glucopyranosyl fluoride in aqueous solution through a concerted, ANDN (SN2) mechanism." *J. Am. Chem. Soc.* **1991,** *113*, 7951-7958.

180. X. Huang, C. Surry, T. Hiebert and A. J. Bennet, "Hydrolysis of (2-Deoxy-.beta.-D-glucopyranosyl)pyridinium Salts." *J. Am. Chem. Soc.* **1995,** *117*, 10614-10621.

181. X. Huang, K. S. E. Tanaka and A. J. Bennet, "Glucosidase-Catalyzed Hydrolysis of α-d-Glucopyranosyl Pyridinium Salts: Kinetic Evidence for Nucleophilic Involvement at the Glucosidation Transition State." *J. Am. Chem. Soc.* **1997,** *119*, 11147-11154.

182.   J. Zhu and A. J. Bennet, "Hydrolysis of (2-Deoxy-α-d-Glucopyranosyl)pyridinium Salts: The 2-Deoxyglucosyl Oxocarbenium Is Not Solvent-Equilibrated in Water." *J. Am. Chem. Soc.* **1998,** *120*, 3887-3893.

183.   D. Indurugalla and A. J. Bennet, "A kinetic isotope effect study on the hydrolysis reactions of methyl xylopyranosides and methyl 5-thioxylopyranosides: oxygen versus sulfur stabilization of carbenium ions." *J Am Chem Soc* **2001,** *123*, 10889-98.

184.   R. D. Guthrie, "System for symbolic representation of reaction-mechanisms." *Pure and Applied Chemistry* **1989,** *61*, 23-56.

185.   R. D. Guthrie and W. P. Jencks, "IUPAC recommendations for hte representation of reaction mechanisms." *Accounts of chemical research* **1989,** *22*, 343-349.

186.   M. E. Soliman, G. D. Ruggiero, J. J. Pernia, I. R. Greig and I. H. Williams, "Computational mutagenesis reveals the role of active-site tyrosine in stabilising a boat conformation for the substrate: QM/MM molecular dynamics studies of wild-type and mutant xylanases." *Organic & biomolecular chemistry* **2009,** *7*, 460-8.

187.   M. E. Caines, S. M. Hancock, C. A. Tarling, T. M. Wrodnigg, R. V. Stick, A. E. Stutz, A. Vasella, S. G. Withers and N. C. Strynadka, "The structural basis of glycosidase inhibition by five-membered iminocyclitols: the clan a glycoside hydrolase endoglycoceramidase as a model system." *Angew Chem Int Ed Engl* **2007,** *46*, 4474-6.

188.   M. Granovsky, J. Fata, J. Pawling, W. J. Muller, R. Khokha and J. W. Dennis, "Suppression of tumor growth and metastasis in Mgat5-deficient mice." *Nature medicine* **2000,** *6*, 306-12.

189.   J. J. Gridley and H. M. I. Osborn, "Recent advances in the construction of beta-D-mannose and beta-D-mannosamine linkages." *J. Chem. Soc.-Perkin Trans. 1* **2000**, 1471-1491.

190.   D. Crich and L. Li, "4,6-O-benzylidene-directed beta-mannopyranosylation and alpha-glucopyranosylation: the 2-deoxy-2-fluoro and 3-deoxy-3-fluoro series of donors and the importance of the O2-C2-C3-O3 interaction." *The Journal of organic chemistry* **2007,** *72*, 1681-90.

191.   D. Crich and N. S. Chandrasekera, "Mechanism of 4,6-O-benzylidene-directed beta-mannosylation as determined by alpha-deuterium kinetic isotope effects." *Angew Chem Int Ed Engl* **2004,** *43*, 5386-9.

192.   L. E. Tailford, V. A. Money, N. L. Smith, C. Dumon, G. J. Davies and H. J. Gilbert, "Mannose foraging by Bacteroides thetaiotaomicron: structure and specificity of the beta-mannosidase, BtMan2A." *The Journal of biological chemistry* **2007,** *282*, 11291-9.

193.   W. A. Offen, D. L. Zechel, S. G. Withers, H. J. Gilbert and G. J. Davies, "Structure of the Michaelis complex of beta-mannosidase, Man2A, provides insight into the conformational itinerary of mannoside hydrolysis." *Chem. Commun.* **2009**, 2484-2486.

194.   F. Vincent, T. M. Gloster, J. Macdonald, C. Morland, R. V. Stick, F. M. Dias, J. A. Prates, C. M. Fontes, H. J. Gilbert and G. J. Davies, "Common inhibition of both beta-glucosidases and beta-mannosidases by isofagomine lactam reflects different conformational itineraries for pyranoside hydrolysis." *Chembiochem : a European journal of chemical biology* **2004,** *5*, 1596-9.

195.   M. K. Dowd, P. J. Reilly and A. D. French, "Relaxed-residue conformational mapping of the three linkage bonds of isomaltose and gentiobiose with MM3 (92)." *Biopolymers* **1994,** *34*, 625-38.

196.   V. Krautler, M. Muller and P. H. Hunenberger, "Conformation, dynamics, solvation and relative stabilities of selected beta-hexopyranoses in water: a molecular dynamics study with the GROMOS 45A4 force field." *Carbohydr Res* **2007,** *342*, 2097-124.

197.   C. B. Barnett and K. J. Naidoo, "Free Energies from Adaptive Reaction Coordinate Forces (FEARCF): an application to ring puckering." *Molecular Physics* **2009,** *107*, 1243-1250.

198.   V. Spiwok, B. Kralova and I. Tvaroska, "Modelling of beta-D-glucopyranose ring distortion in different force fields: a metadynamics study." *Carbohydr Res* **2010,** *345*, 530-7.

199.   S. E. Barrows, F. J. Dulles, C. J. Cramer, A. D. French and D. G. Truhlar, "Relative stability of alternative chair forms and hydroxymethyl conformations of beta-D-glucopyranose." *Carbohydrate Research* **1995,** *276*, 219-251.

200.   M. Appell, G. Strati, J. L. Willett and F. A. Momany, "B3LYP/6-311++G** study of alpha- and beta-D-glucopyranose and 1,5-anhydro-D-glucitol: 4C1 and 1C4 chairs, (3,O)B and B(3,O) boats, and skew-boat conformations." *Carbohydr Res* **2004,** *339*, 537-51.

201.   U. Schnupf, J. L. Willett, W. B. Bosma and F. A. Momany, "DFT study of alpha- and beta-D-allopyranose at the B3LYP/6-311++G ** level of theory." *Carbohydr Res* **2007,** *342*, 196-216.

202.   A. R. Ionescu, A. Berces, M. Z. Zgierski, D. M. Whitfield and T. Nukada, "Conformational pathways of saturated six-membered rings. A static and dynamical density functional study." *The journal of physical chemistry. A* **2005,** *109*, 8096-105.

203.   F. A. Momany, M. Appell, J. L. Willett, U. Schnupf and W. B. Bosma, "DFT study of alpha- and beta-D-galactopyranose at the B3LYP/6-311++G** level of theory." *Carbohydr Res* **2006,** *341*, 525-37.

204.   M. Appell, J. L. Willett and F. A. Momany, "DFT study of alpha- and beta-D-mannopyranose at the B3LYP/6-311++G** level." *Carbohydr Res* **2005,** *340*, 459-68.

205.   G. I. Csonka, A. D. French, G. P. Johnson and C. A. Stortz, "Evaluation of Density Functionals and Basis Sets for Carbohydrates." *J. Chem. Theory Comput.* **2009,** *5*, 679-692.

206.   S. Grimme, H. Kruse, L. Goerigk and G. Erker, "The mechanism of dihydrogen activation by frustrated Lewis pairs revisited." *Angew Chem Int Ed Engl* **2010,** *49*, 1402-5.

207.   T. W. Liu, C. W. Ho, H. H. Huang, S. M. Chang, S. D. Popat, Y. T. Wang, M. S. Wu, Y. J. Chen and C. H. Lin, "Role for alpha-L-fucosidase in the control of Helicobacter pylori-infected gastric cancer cells." *Proceedings of the National Academy of Sciences of the United States of America* **2009,** *106*, 14581-14586.

208.   S. W. Liu, C. S. Chen, S. S. Chang, K. K. T. Mong, C. H. Lin, C. W. Chang, C. Y. Tang and Y. K. Li, "Identification of Essential Residues of Human alpha-L-Fucosidase and Tests of Its Mechanism." *Biochemistry* **2009,** *48*, 110-120.

209.   P. J. Willems, H. C. Seo, P. Coucke, R. Tonlorenzi and J. S. O'Brien, "Spectrum of mutations in fucosidosis." *Eur. J. Hum. Genet.* **1999,** *7*, 60-67.

210.   K. Yuan, D. Kucik, R. K. Singh, C. M. Listinsky, J. J. Listinsky and G. P. Siegal, "Alterations in human breast cancer adhesion-motility in response to changes in cell surface glycoproteins displaying alpha-L-fucose moieties." *International journal of oncology* **2008,** *32*, 797-807.

211.   N. R. De Martinez and N. Castro, "Serum alpha-fucosidase and hexosaminidase as biological markers in human cancer." *Medicina* **1982,** *42*, 36-42.

212.   J. Fernandez-Rodriguez, D. Ayude, M. P. de la Cadena, V. S. Martinez-Zorzano, A. de Carlos, A. Caride-Castro, G. de Castro and F. J. Rodriguez-Berrocal, "Alpha-L-fucosidase enzyme in the prediction of colorectal cancer patients at high risk of tumor recurrence." *Cancer detection and prevention* **2000,** *24*, 143-9.

213.   M. Shah, S. Telang, G. Raval, P. Shah and P. S. Patel, "Serum fucosylation changes in oral cancer and oral precancerous conditions: alpha-L-fucosidase as a marker." *Cancer* **2008,** *113*, 336-46.

214.   W. J. White, Jr., K. J. Schray, G. Legler and J. A. Alhadeff, "Further studies on the catalytic mechanism of human liver alpha-L-fucosidase." *Biochim Biophys Acta* **1987,** *912*, 132-8.

215.   E. V. Eneyskaya, A. A. Kulminskaya, N. Kalkkinen, N. E. Nifantiev, N. P. Arbatskii, A. I. Saenko, O. V. Chepurnaya, A. V. Arutyunyan, K. A. Shabalin and K. N. Neustroev, "An alpha-L-fucosidase from Thermus sp. with unusually broad specificity." *Glycoconjugate journal* **2001,** *18*, 827-34.

216.   O. Berteau, I. McCort, N. Goasdoue, B. Tissot and R. Daniel, "Characterization of a new alpha-L-fucosidase isolated from the marine mollusk Pecten maximus that catalyzes the hydrolysis of alpha-L-fucose from algal fucoidan (Ascophyllum nodosum)." *Glycobiology* **2002,** *12*, 273-82.

217.   B. Cobucci-Ponzano, A. Trincone, A. Giordano, M. Rossi and M. Moracci, "Identification of the catalytic nucleophile of the family 29 alpha-L-fucosidase from Sulfolobus solfataricus via chemical rescue of an inactive mutant." *Biochemistry* **2003,** *42*, 9525-31.

218.   C. A. Tarling, S. He, G. Sulzenbacher, C. Bignon, Y. Bourne, B. Henrissat and S. G. Withers, "Identification of the catalytic nucleophile of the family 29 alpha-L-fucosidase from Thermotoga maritima through trapping of a covalent glycosyl-enzyme intermediate and mutagenesis." *The Journal of biological chemistry* **2003,** *278*, 47394-9.

219.   A. Laio and M. Parrinello, "Escaping free-energy minima." *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99*, 12562-6.

220.   S. Nose, "A molecular dynamics method for simulations in the canonical ensemble." *Molecular Physics* **1984,** *52*, 255-268.

221.   J. P. Perdew, K. Burke and M. Ernzerhof, "Generalized Gradient Approximation Made Simple." *Phys Rev Lett* **1996,** *77*, 3865-3868.

222.   M. L. Sinnott, "Catalytic mechanisms of enzymatic glycosyl transfer." *Chemical reviews* **1990,** *90*, 1171-1202.

223.   M. L. Sinnott and W. P. Jencks, "Solvolysis of D-glucopyranosyl derivatives in mixtures of ethanol and 2,2,2-trifluoroethanol." *J. Am. Chem. Soc.* **1980,** *102*, 2026-2032.

224.   D. C. Phillips, "Henn egg white lysozyme molecule." *Proceedings of the National Academy of Sciences of the United States of America* **1967,** *57*, 484-&.

225.   D. J. Vocadlo, G. J. Davies, R. Laine and S. G. Withers, "Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate." *Nature* **2001,** *412*, 835-838.

226.   I. Tvaroska, "Molecular modeling insights into the catalytic mechanism of the retaining galactosyltransferase LgtC." *Carbohydrate Research* **2004,** *339*, 1007-1014.

227.   I. Tvaroska, Molecular modeling of retaining glycosyltransferases. In *Nmr Spectroscopy and Computer Modeling of Carbohydrates: Recent Advances*, J. F. G. Vliegenthart and R. J. Woods, Eds. Amer Chemical Soc: Washington, 2006; Vol. 930, pp 285-301.

228.   E. Cabib and L. F. Leloir, "Biosynthesis of trehalose phosphate." *J. Biol. Chem.* **1958,** *231*, 259-275.

229.   J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, "Scalable molecular dynamics with NAMD." *Journal of computational chemistry* **2005,** *26*, 1781-1802.

230.   K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. Gonzalez-Outeirino, C. R. Daniels, B. L. Foley and R. J. Woods, "GLYCAM06: a generalizable biomolecular force field. Carbohydrates." *Journal of computational chemistry* **2008,** *29*, 622-55.

231.   W. Humphrey, A. Dalke and K. Schulten, "VMD: Visual molecular dynamics." *Journal of molecular graphics* **1996,** *14*, 33-&.

232.   P. H. Hunenberger, "Optimal charge-shaping functions for the particle-particle-particle-mesh (P3M) method for computing electrostatic interactions in molecular simulations." *Journal of Chemical Physics* **2000,** *113*, 10464-10476.

233.   M. Boero, T. Ikeda, E. Ito and K. Terakura, "Hsc70 ATPase: an insight into water dissociation and joint catalytic role of K+ and Mg2+ metal cations in the hydrolysis reaction." *J Am Chem Soc* **2006,** *128*, 16798-807.

234.   M. Boero, "LeuRS Synthetase: A First-Principles Investigation of the Water-Mediated Editing Reaction." *The journal of physical chemistry. B* **2011,** *115*, 12276-86.

235.   R. P. Gibson, C. A. Tarling, S. Roberts, S. G. Withers and G. J. Davies, "The donor subsite of trehalose-6-phosphate synthase - Binary complexes with UDP-glucose and UDP-2-deoxy-2-fluoro-glucose at 2 angstrom resolution." *J. Biol. Chem.* **2004,** *279*, 1950-1955.

236.   H. S. Seo, Y. J. Koo, J. Y. Lim, J. T. Song, C. H. Kim, J. K. Kim, J. S. Lee and Y. D. Choi, "Characterization of a bifunctional enzyme fusion of trehalose-6-phosphate synthetase and trehalose-6-phosphate phosphatase of Escherichia coli." *Applied and environmental microbiology* **2000,** *66*, 2484-2490.

237.   K. Fukui, "The path of chemical reactions. The IRC approach." *Accounts of chemical research* **1981,** *14*, 363-368.

238.   N. Marzari and D. Vanderbilt, "Maximally localized generalized Wannier functions for composite energy bands." *Physical Review B* **1997,** *56*, 12847-12865.

239.   H. M. Pickett and H. L. Strauss, "Conformational structure, energy and inversion rates of cyclohexane and some related oxanes." *J. Am. Chem. Soc.* **1970,** *92*, 7281-&.

240.   H. M. Pickett and H. L. Strauss, "Symmetry and conformation of cycloalkanes." *Journal of Chemical Physics* **1971,** *55*, 324-&.