Elucidating catalytic mechanisms of glycoside hydrolases and transferases by means of *ab initio* molecular dynamics simulations

Javier Iglesias Fernández

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA

PROGRAMA DE DOCTORAT EN BIOTECNOLOGIA

TÍTOL DE LA TESI

Elucidating catalytic mechanisms of glycoside hydrolases and transferases by means of *ab initio* molecular dynamics simulations.

Memòria presentada per **Javier Iglesias Fernández** per optar al títol de doctor per la Universitat de Barcelona

Nom, cognoms i signatura del Director

Carme Rovira Virgili

Nom, cognoms i signatura del doctorand/a

Javier Iglesias Fernández

Nom, cognoms i signatura del tutor/a

Josefa Badía Palacín

Acknowledgements

¡Venga va! ¡Ya han pasado cuatro años! ¡Voy a cumplir 30! ¡Estoy escribiendo los agradecimientos del doctorado! Ufff…que estrés. Esto no puede estar pasando…es hora de ver como he llegado a esta situación, depurar responsabilidades. Pensándolo bien, me doy cuenta de que…

los principales responsables de que todo esto esté sucediendo son *mis padres*. Creo que sin todo el apoyo que me han dado (en esta parte también incluyo las broncas por no estudiar) esto no habría pasado. No hay que olvidarse de la familia, que siempre ha confiado en mí.

la *Carme* también tiene bastante parte de culpa. Si no me hubiese dado una oportunidad única para hacer un doctorado, consejo y confianza durante seis años no estaría en esta situación.

Albert no tendría que haber tenido tanta paciencia conmigo cuando parametrizaba mi primer sistema. ¡Maldito 2F-glucosa ahí empezó todo!.

si no hubiese entrado a formar parte del *Trio Calavera* aún seguiría en primero de carrera y no hubiese descubierto la química computacional unos años más tarde. Vaya tela Victoria, Maestre y Lorena.

todos los *compañeros de carrera y máster* son responsables de esta situación y, por lo tanto, culpables. Especialmente Rubén por todas las cervezas y Chicagos que hemos compartido.

sin todos los momentos compartidos con *los Chunguitos* (fiestas, risas, viajes, casas rurales, barbacoas, carnavales, etc.) el resultado de este viaje habría sido muy diferente.

todos los compañeros *de "laboratorio" (Víctor, Lluís, Santi, Alba, Marc W., Oriol, Javi, Mertxe, Fermín, Marc M., Pietro*, *Xevi) y personas que he conocido durante la tesis* tienen su parte de culpa. Estáis todos fichados.

la *Cris*, por su paciencia y por estar presente durante tantos años, se merece un capítulo especial en esta sección. Eres una de las mayores responsables de todo esto.

solo hay una persona capaz de aguantar mis ataques de cosquillas, momentos de estrés, embolados, etc. sin salir corriendo. Eres increíble *Moflins*.

Muchas gracias a todos.

Collaborations.

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

- **Prof. Gideon J. Davies**.
  Department of Chemistry. University of York (United Kingdom).

- **Prof. Spencer Williams**.
  School of Chemistry. University of Melbourne (Australia).

- **Prof. Ramón Hurtado**.
  Institute of Biocomputation and Physics of Complex Systems (BIFI). University of Zaragoza (Spain).

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

- **Prof. Antoni Planas**.
  Department of Bioengineering. IQS-Universitat Ramon Llull (Spain).

- **Prof. Giovanni Bussi**.
  Molecular and Statistical and Biophysics group. Scuola Internazionale Superiore di Studi Avanzati (Italy).

- **Dr. Alba Díaz Rodríguez**.
  Department of Organic and Inorganic Chemistry. University of Oviedo (Spain).

- **Prof. James Ketudat**.
  School of Biochemistry. University of Technology (Thailand).

Short stay in foreign research center:

- 12/2013 – 02/2014.

  3 months training at the Molecular and Statistical and Biophysics group (Prof. Giovanni Bussi). Scuola Internazionale Superiore di Studi Avanzati (SISSA), Italy.

Symbols and acronyms.

Å	Angstrom
a. u.	atomic units
a. m. u.	atomic mass units
ALPH	Antiperiplanar Lone-Pair Hypothesis
CAZy	Carbohydrate-active enzymes
CP	Car-Parrinello method
CPMD	Car-Parrinello molecular dynamics
CPMD/MM	Car-Parrinello molecular dynamics / molecular mechanics
CV	collective variable
DCACP	dispersion-corrected atom-centered potential
DFT	density functional theory
e/e^-	electron
ESP	electrostatic potential derived charges
FEL	free energy landscape
fs	femtosecond
G	Gibbs free energy
GalNAc-T	N-acetylgalactosaminyl transferase
GH	glycosyl hydrolase
GT	glycosyl transferase
K	Kelvin
KIE	Kinetic Isotope Effects
MC	Michaelis complex
MD	molecular dynamics

MM	molecular mechanics
MTD	metadynamics
nm	nanometer
ns	nanosecond
OtsA	trehalose 6-phosphate synthase
PBE	Perdew-Burke-Ernzerhof exchange-correlation functional
pNPβGal	*p*-nitrophenil β-D-galactopyranoside
ps	picosecond
QM	quantum mechanics
QM/MM	quantum mechanics / molecular mechanics
rmsd	root mean square deviation
RESP	restrained electrostatic potential derived charges
Ry	Rydberg
SsβG	*Sufolobus solfataricus* β-glucosidase
TS	transition state
UDP	uridine diphosphate
α3GalT	alpha-3-galactosyltransferase
δt	time step
δs	width of the Gaussian terms
μ	fictitious electronic mass
w	height of the Gaussian terms

Contents.

Outline.

Outline

Carbohydrates play a central role in transport and storage of energy and as molecular building blocks. Glycoconjugates, specifically glycoproteins and glycolipids, are important components of cell surfaces and the extracellular environment that mediate cellular and molecular interactions. Defects in glycosylation are associated with human diseases and the ability of glycans to modulate immune response confers them a critical role in susceptibility and resistance to pathogens.

The large amount of glycan structures requires the existence of a diverse group of degrading and remodeling enzymes, such as glycoside hydrolases (GHs) and glycoside transferases (GTs). These are highly specific enzymes responsible of the hydrolysis (GHs) and formation (GTs) of glycosidic bonds in carbohydrates. Knowledge of their enzymatic mechanism at a molecular level is crucial to understand how carbohydrates are assembled/degraded in organisms, as well as to develop new drugs. The detailed characterization of the transition state (TS) of the chemical reaction that they catalyze, for instance, is crucial for the development of TS-analog inhibitors, which are known to be very efficient.

In recent years, our group has investigated the implications of substrate sugar conformations in catalysis for several GHs. By adapting sugar puckering coordinates as collective variables in *ab initio* metadynamics simulations, it was found that the conformational changes *on-enzyme* can be traced back to the conformational properties of isolated saccharides (e. g. β-D-glucose and β-D-mannose). This provides a basis for predicting substrate catalytic itineraries for GHs. In this Thesis, we extend these analyses to α-D-mannose to corroborate the catalytic itineraries that have been proposed for α-mannosidases and investigate the dependence of the free energy landscape with respect to the choice of collective variables. To gain insights into TS mimicry, we also explore whether the analyses previously developed for GH substrates can be applied to GH inhibitors and sugar cations (oxocarbenium ions).

Unlike retaining GHs, known to operate by means of a double displacement mechanism, the reaction mechanism of retaining GTs is still controversial. Both a double displacement mechanism (by analogy to retaining GHs) and a front-face mechanism have been proposed in previous experimental and computational studies. In this Thesis, we apply quantum mechanics/molecular mechanics (QM/MM) metadynamics to elucidate the catalytic mechanism of glycosyl transfer in two different enzymes: an engineered retaining β-glucosidase and a N-acetylgalactosaminyl transferase.

The Thesis is organized as follows:

After a brief introduction to carbohydrate-active enzymes and the methods employed for their theoretical study (Chapters I and II), we test the reliability of the collective variables employed for simulations of ring puckering by calculating the free energy landscape of all possible conformations of cyclohexane using two different set of collective variables (Chapter III).

In Chapter IV, the effect of substrate distortion in GHs (specifically, in α-mannosidases) is studied by analyzing the conformational flexibility of an isolated α-D-mannose molecule. We use a simple preactivation index integrating several structural, electronic, and energetic properties to predict the conformation of the substrate in the Michaelis complex and the conformational itineraries of α-mannosidases. The same methodology is later applied to GH inhibitors (same Chapter IV) and sugar cations (Chapter V) to find out why some inhibitors are able to mimic the TS conformation (good TS-mimics) better than others.

Finally, we investigate the molecular mechanism of glycosidic bond formation in two cases: an engineered GH (chapter VI) and a GT (chapter VII). The results for both enzymes are consistent with a front-face single displacement mechanism, increasing the number of evidences for this mechanism as an important strategy for sugar bond formation in nature.

Chapter I – Introduction.

Introduction.

1. Carbohydrates. Classification and structure.

Carbohydrates, as the most abundant biomolecules on Earth, have a huge diversity of roles in nature. They serve as structural and protective elements in the cell walls of bacteria and plants and in the connective tissues of animals, participate in cell-cell recognition and adhesion processes and are the main energy-yielding molecules of nonphotosynthetic cells.

There are three major classes of carbohydrates in terms of size: monosaccharides, oligosaccharides, and polysaccharides. Monosaccharides, or simple sugars, consist of a single polyhydroxy aldehyde or ketone unit, being D-glucose the most abundant monosaccharide in nature. Oligosaccharides consist of short chains of monosaccharide units joined by characteristic linkages called glycosidic bonds. Finally, polysaccharides are sugar polymers containing more than 20 monosaccharide units. As an example, cellulose is a polysaccharide formed by hundreds or thousands of glucose units.

In aqueous solution, monosaccharides with five or more carbon atoms tend to form cyclic structures, predominantly (Figure I-1). This process is achieved thanks to the reaction between the aldehyde or ketone unit of the linear monosaccharide with an alcohol of the same molecule to form a covalent bond, a hemiacetal or hemiketal, respectively. The chemical reaction leads to the formation of an additional asymmetric center, called anomeric carbon, that can exist in two stereoisomeric forms or anomers (α or β) interchangeable in aqueous solution (mutarotation process; Figure I-1). It is noteworthy that the equilibrium mixture generally do not contain equal amounts of both anomers.

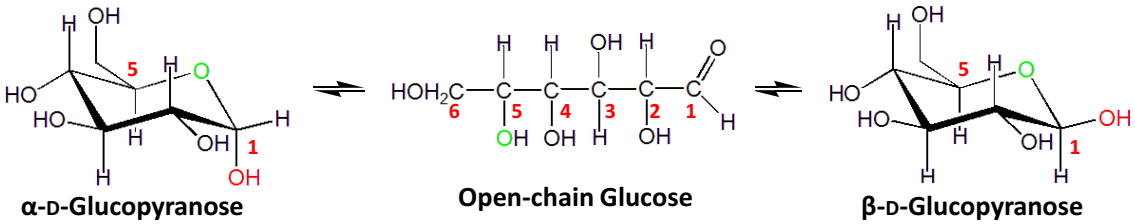

Figure I - 1. Mutarotation process: Formation of the two cyclic forms of D-Glucose.

2. Conformations of pyranose rings.

Pyranose rings (six-membered carbohydrate molecules consisting of five carbon atoms and one oxygen atom; Figure I-2) are very flexible molecules that can adopt a wide range of conformations.

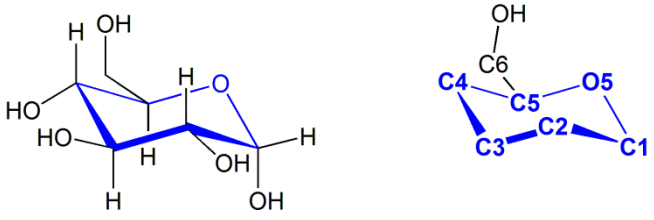

Figure I - 2. α-D-glucopyranose representation and atom numbering.

The nomenclature to define the different conformations of six-membered rings is regulated by the IUPAC (IUPAC, 1980). There are 38 possible conformations that can be grouped into five different types: chair, boat, skew-boat, half-chair and envelope (represented by the letters C, B, S, H and E, respectively)(Figure I-3). To distinguish between each of the five possible types of conformations is necessary to define a reference plane formed by four atoms. The numerals associated with each conformation (i. e. $^{4}C_{1}$) are the ring-atoms that lie outside of the reference plane (subscript or superscript if they are down or up with respect to the plane).

- Chair: The reference plane is defined by two parallel ring sides, with the lowest-numbered carbon atom located exoplanar. Possible conformations are $^{4}C_{1}$ and $^{1}C_{4}$.
- Boat: The reference plane is also defined by two parallel ring sides, but the two exoplanar atoms are on the same side of the reference plane. Possible conformations are: $^{1,4}B$, $B_{1,4}$, $^{2,5}B$, $B_{2,5}$, $^{O,3}B$ and $B_{O,3}$.
- Skew-Boat: The reference plane is formed by three adjacent atoms and one remaining non-adjacent atom. Possible conformations are: $^{1}S_{3}$, $^{3}S_{1}$, $^{1}S_{5}$, $^{5}S_{1}$, $^{O}S_{2}$ and $^{2}S_{O}$.
- Half: The reference plane is defined by four adjacent coplanar atoms. The two remaining atoms are located above and below the plane, respectively. Possible conformations are: $^{O}H_{1}$, $^{1}H_{O}$, $^{1}H_{2}$, $^{2}H_{1}$, $^{2}H_{3}$, $^{3}H_{2}$, $^{3}H_{4}$, $^{4}H_{3}$, $^{4}H_{5}$, $^{5}H_{4}$, $^{5}H_{O}$, and $^{O}H_{5}$.
- Envelope: The reference plane is defined by five atoms and the remaining one lie outside. Possible conformations are: ^{3}E, E_{4}, ^{5}E, E_{O}, ^{1}E, E_{2}, ^{O}E, E_{1}, ^{2}E, E_{3}, ^{4}E, E_{5}.

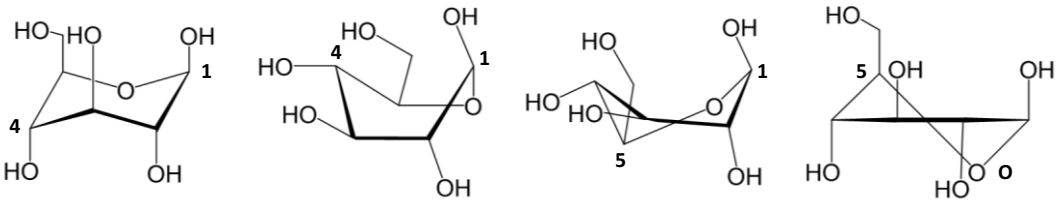

Figure I - 3. Nomenclature for a few different conformations of a pyranose ring.

2.1 Stoddart and Mercator diagrams.

Following Cremer and Pople spherical coordinates of ring puckering (Cremer & Pople, 1975; discussed in detail in Chapter III in the context of the conformations of cyclohexane), all possible conformations of six-membered rings can be represented on the surface of a sphere, with both chair conformations located on the poles (Figure I-4, center) and the remaining conformations at different latitudes. Several 2D plots can be derived from this spherical representation. As an example, the Stoddart or Cartesian diagrams (Figure I-4, left) are often used as an "itinerary map", since they easily inform of the possible conformational routes that might be followed by a pyranoside ring as its moves from one conformation to another. The Mercator projection (Figure I-4, right), analogous to the global cartography, may also be used. None of the diagrams gives any information relative to the stability of the different ring

conformations, but it can be determined using suitable theoretical approaches, as exemplified by the results presented in Chapters III-IV of this Thesis.

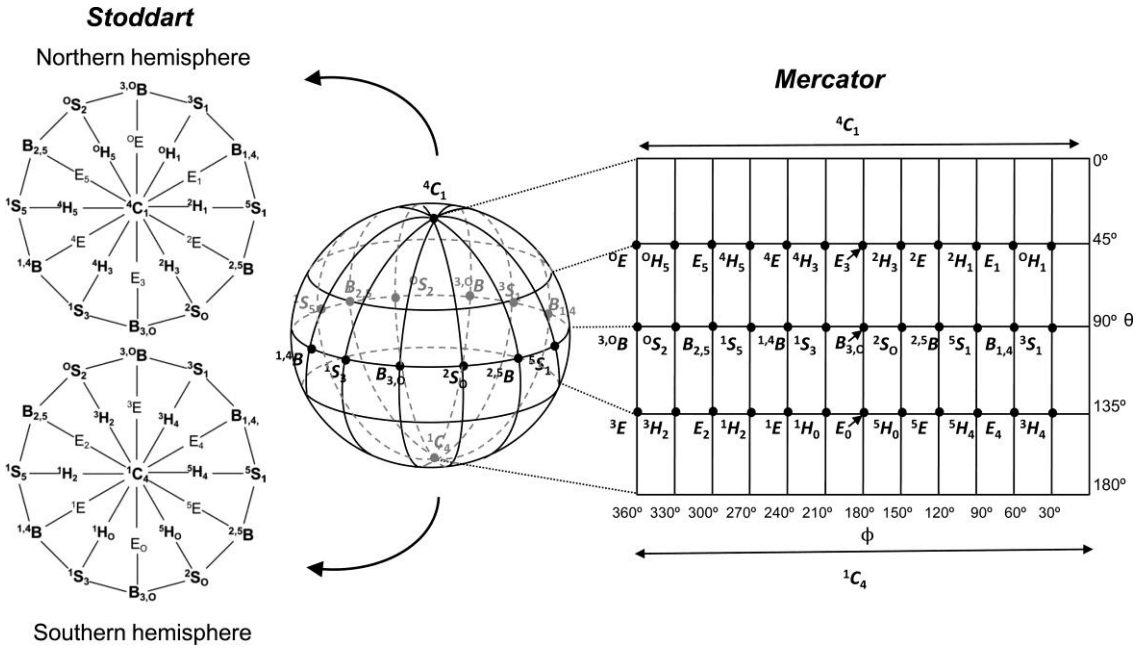

Figure I - 4. Spherical representation of all possible conformations of a pyranose ring along with two of their most used 2D representations: Stoddart or Cartesian (Northern and Southern projections) and Mercator diagrams. Figure taken from reference (Davies, 2012).

3. Glycoside hydrolases or glycosidases.

The vast array of carbohydrate structures present in nature necessitates the existence of a diverse group of degrading and remodeling enzymes, such as glycoside hydrolases (GHs), lyases or glycoside transferases (GTs). Typically, these enzymes make up approximately 1-2% of the genome of any organism (Davies, 2005).

GHs (EC 3.2.1.x) represent a ubiquitous group of enzymes responsible for the cleavage of the glycosidic bond between a carbohydrate subunit and another sugar molecule or aglycon leaving group. This family of enzymes plays essential roles in a wide range of biological processes such as control of protein folding and viral infections (Hurtley, 2001). Additionally, GHs have attracted much interest because of the huge occurrence of genetically inherited disorders associated to these enzymes and the potential for GHs inhibitors to act as new therapeutic agents for viral infections, such as HIV or influenza (Gloster, 2007).

To date, GHs have been systematically classified on the basis of amino acid sequence similarity into 133 families. This classification system is known as the CAZy (Carbohydrate-Active Enzymes, www.cazy.org) database (Lombard, 2014). The idea underlying this classification is that enzymes belonging to the same family share a similar 3D structure and mechanism of action. According to their mechanism, glycoside hydrolases can also be classified into retaining and inverting (see further details in the section 3.1).

GHs can also be classified on the basis of whether they cleave an internal glycosidic linkage (*endo*-GH) or a terminal one (*exo*-GH), most commonly at the non-reducing end. Whether and enzyme is *endo* or *exo* has no relation with the mechanism of action, but with the topology of the active site. Although there is a huge amount of GHs families and hence protein folds, the overall topologies of the active sites of all these enzymes fall into only three general classes (Davies & Henrissat, 1995). The *pocket or crater* topology (Figure I-5a) is optimal for the recognition of a saccharide non-reducing end (*exo*-GHs), whereas the *cleft or groove* topology (Figure I-5b) is commonly found in *endo*-GHs and consist of a more open structure, with a substrate binding cleft that allows random binding of several sugar units and make internal cuts within the polysaccharide chain. Finally, the *tunnel* topology (Figure I-5c) is found only in processive enzymes, in which the resulting tunnel enables consecutive hydrolysis reactions without releasing the polysaccharide substrate.

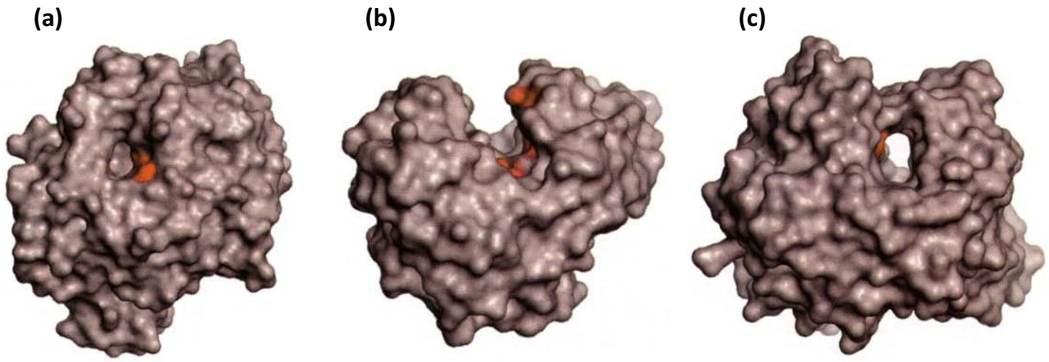

Figure I - 5. Topological structure of the active site of glycoside hydrolases: a) pocket or crater, b) cleft or groove and c) tunnel. Figure taken from reference (Davies & Henrissat, 1995).

Since GHs are enzymes able to act on big polymeric substrates (e. g. cellulose), their active center is able to accommodate several monosaccharide residues. Therefore, the active site of GHs is divided into different *subsites*, each one accommodating one monosaccharide residue. Subsites are labeled from *–n* to *+n* (where *n* is an integer)(Figure I-6). *–n* represents the non-reducing end and *+n* the reducing end, with the cleavage taking place between -I and +I subsites (Davies, 1997). The sugar saccharide located at the -I subsite will be hereafter named as "-I sugar molecule".

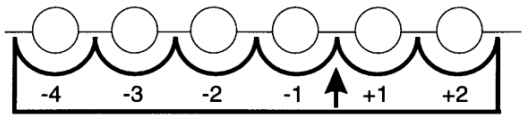

Figure I - 6. Schematic representation of the sugar-binding subsites in GHs. Each monosaccharide is represented with a circle and the bond that is cleaved is located between subsites –I and +I (indicated with an arrow). Figure taken from reference (Davies, 1997).

3.1 GHs reaction mechanism.

Despite the large number of GHs known, most of them share a similar acid/base catalytic mechanism that involves two essential residues: a proton donor (general acid) and a nucleophile (general base). Additionally, the hydrolysis of the glycosidic bond can occur with either retention or inversion of the anomeric configuration (Figure I-7), as outlined by Koshland (Koshland, 1953).

Hydrolysis with inversion of stereochemistry is performed in a single S_N2 (bimolecular nucleophilic substitution) displacement (Figure I-7a), in which the nucleophile is a water molecule. The general acid residue facilitates the departure of the +I sugar or aglycon unit (-OR), by transferring a proton, whereas the general base activates the water molecule to attack at the anomeric center.

The retaining mechanism follows a double S_N2 displacement (Figure I-7b) with the formation of a covalent glycosyl-enzyme intermediate. First, the nucleophile (general base) residue directly attacks the anomeric carbon and the leaving group departs thanks to the assistance of the acid residue. The resulting covalent glycosyl-enzyme intermediate is hydrolyzed by an activated water molecule generating a product with the same stereochemistry as the substrate.

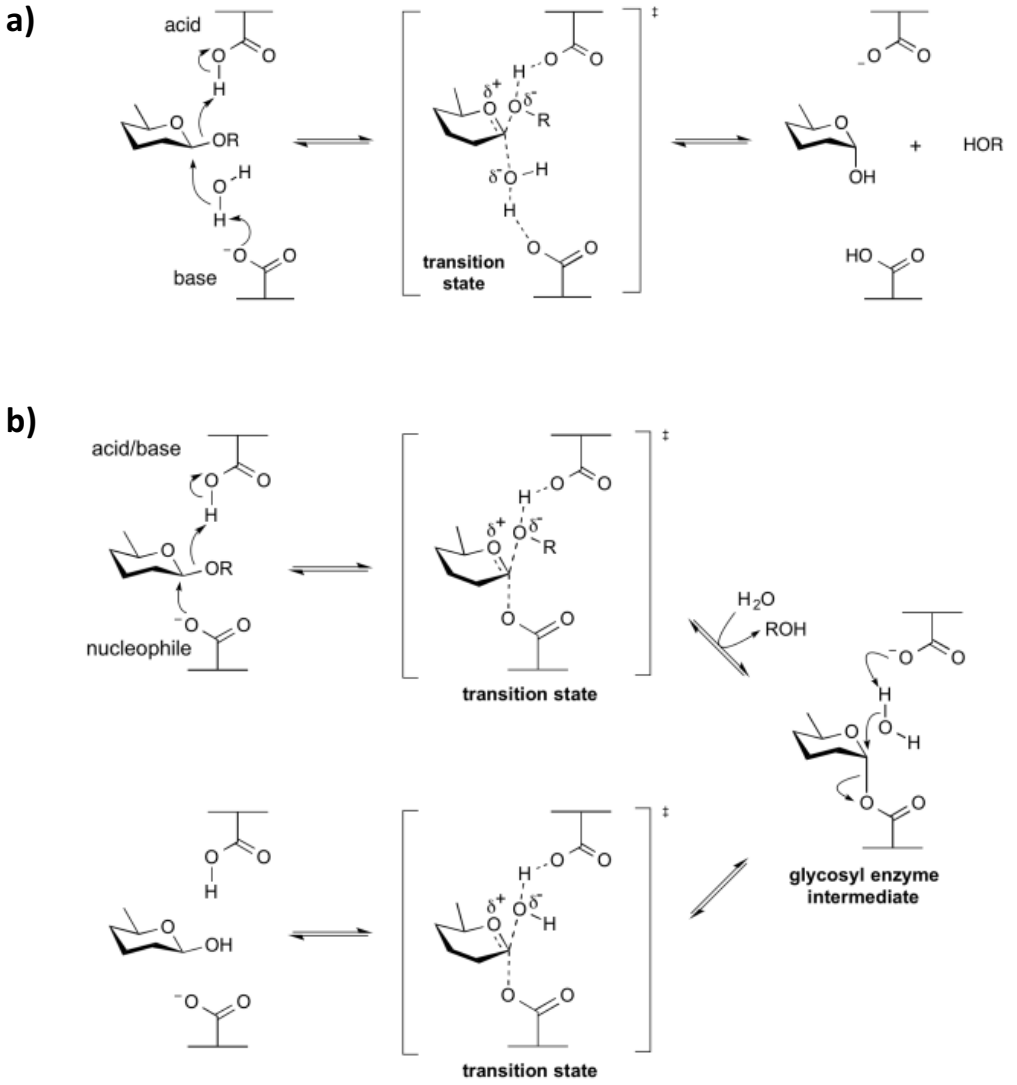

Figure I - 7. Reaction mechanism of an inverting (a) and a retaining (b) β-glycoside hydrolase. Figures taken from the CAZy web page (www.cazy.org).

It has been found that the distance between both catalytic residues (general acid and general base) is very well conserved, being 5.5 Å and 10.5 Å in retaining and inverting GHs, respectively. The higher distance observed for inverting enzymes accounts for the extra space

needed for the accommodation of a water molecule at the active center (Figure I-7a) (McCarter & Withers, 1994).

Finally, Kinetic Isotope Effects (KIE) studies reveal that both reaction mechanism of GHs (with retention or inversion of configuration) involve a very similar sugar cation or oxocarbenium ion transition state (TS)(Zechel & Withers, 2000)(Figure I-7). Such a TS presents a highly dissociative character, with low bond orders between the anomeric carbon and the glycosidic oxygen as well as the nucleophile. As a consequence, significant positive charge is developed at the anomeric carbon, which is partially compensated by an electron donation from the endocyclic O5 oxygen atom. This electron donation gives rise to a C1-O5 partial double bond character (sp^2 geometry), forcing the -I sugar molecule to adopt a planar geometry (C5, O5, C1 and C2 atoms arrange in a coplanar geometry). The requirement of a planar geometry limits the number of possible conformations that a sugar ring can adopt in the TS, namely: two half-chair $^4H_3$ and $^3H_4$, and two boat $^{2,5}B$ and $B_{2,5}$ (see "Conformations of pyranose rings" above).

3.2 Substrate distortion in the Michaelis complex.

The first enzyme whose crystal structure was solved by X-ray techniques was lysozyme, providing an excellent model for studying substrate-enzyme interactions (Blake, 1965). These results proved that the lock-and-key model, postulated in 1894 by Emil Fischer (Fischer, 1894), was too rigid to describe the dynamic and flexible structure of this kind of proteins. In contrast, the "induced fit theory" proposed by Daniel Koshland (Koshland, 1958) provided a most satisfactory explanation for the experimental results. This theory states that the enzyme active site changes shape to bind the substrate into the proper alignment, approaching the substrate to the TS of the reaction.

The first evidence of substrate distortion in GHs was proposed by David Philips during his studies on hen-egg white lysozyme (Philips, 1967), proving that sugar conformations play a role in GH catalysis. Additionally, during the last decades several studies have shown that sugar molecules bind to GHs in a distorted conformation, such as boat ($^{1,4}B$), skew-boat ($^1S_3$ or $^1S_5$) or other conformations instead of adopting the relaxed chair conformation (Davies, 2012). It is nowadays established that substrate distortion has significant benefits on catalysis (Vocadlo & Davies, 2008), as discussed in the next sections.

3.2.1 Steric hindrance.

The nucleophilic attack at the anomeric carbon and the concomitant leaving group departure (see section 3.1) require an *in-line* approach of the general base protein residue. For β-glucopyranose molecules in the chair conformation, such attack is unfavorable due to the steric hindrance by the hydrogen atom bound to C1 (Figure I-8a). Therefore, *prior* to the nucleophilic substitution, the enzyme must distort the -I sugar molecule away from its ground state $^4C_1$ conformation toward a boat or skew-boat distorted conformation. In this conformation, the glycosidic and the C1-H1 bonds become axial and equatorial, respectively (Davies, 1998)(Figure I-8b).

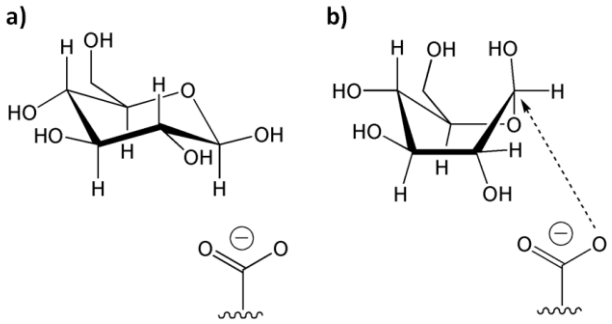

Figure I - 8. Minimization of steric hindrance due to substrate distortion.

3.2.2 Structural and electronic "approach" to the transition state.

As mentioned above, active sites of enzymes are complementary to the TS of the enzymatic reaction. Therefore, GHs bind substrates in a distorted conformation that is closer to the TS of the enzymatic process (sugar cation or oxocarbenium ion-like). There are four distorted structures that place the C2-C1-O5-C5 atoms in a coplanar geometry ($^{2,5}B$, $B_{2,5}$, $^{4}H_{3}$ and $^{3}H_{4}$; see section 2) allowing orbital overlap and electron delocalization between O5 and C1 atoms. Calculations, based on *ab initio* molecular dynamics simulations, reveal that such distorted conformations are characterized by an increase of the charge at the anomeric carbon, an increase of the C1-O1 distance and, finally, a decrease of the C1-O5 distance (partial double bond character). Therefore, suggesting that substrate distortion resembles structurally and electronically the oxocarbenium TS of the reaction (Biarnés, 2006)(Biarnés, 2007).

3.2.3 Stereoelectronic theory.

As mentioned before, the TS of the reaction is characterized by a planar configuration of the -I sugar molecule that increases electron delocalization between O5 and C1. This configuration displaces most of the positive charge that would otherwise develop at the anomeric carbon due to leaving group departure. Therefore, the oxocarbenium ion-like TS is dependent on the stabilization capabilities of the O5 oxygen atom. This ability of oxygen to donate its nonbonding electrons is also the key factor for the reactivity of orthoesters, acetals and other related compounds (Kirby, 1984). The stereoelectronic theory, derived from the APLH (Antiperiplanar Lone-Pair Hypothesis)(Deslongchamps, 1975), proposes that such hydrolysis reactions are favored by conformations in which the nonbonding (lone-pair) electrons of the intra-ring oxygen atom (O5 oxygen atom for sugar molecules) lie antiperiplanar to the leaving group bond. In this configuration, the overlap between the O5 lone pair orbital and the antibonding σ* orbital of the leaving group bond is maximized. Consequently, the orbitals of the saccharide ring at reactants state (i. e. Michaelis complex, MC) change hybridization into those of the oxocarbenium ion-like TS with minimal structural reorganization.

The above considerations have important implications for the mechanism of action of GHs and other related enzymes. As an example, in the ground state of a β-glucopyranose molecule ($^{4}C_{1}$ conformation), C-OR cleavage is expected to be stereoelectronically unfavorable: due to the equatorial orientation of the aglycone leaving group, none of the O5 lone pairs are antiperiplanar with it. In order to stabilize the charge development at the anomeric carbon,

the reaction must occur via an alternative non-ground state conformation displaying an axial aglycon leaving group and an antiperiplanar orientation with the O5 lone pairs (Figure I-9).

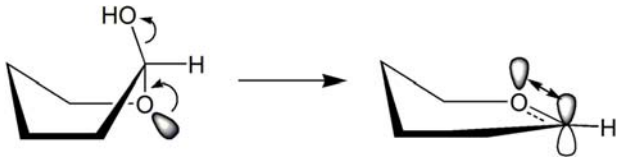

Figure I - 9. Schematic representation of a sugar hydrolysis according to the stereoelectronic theory.

3.2.4 Conformational itineraries.

As mentioned above, the substrate adopts a distorted conformation in the Michaelis complex (MC) of most GHs that approaches it to the TS of the reaction. Therefore, the MC and the products of the reaction (inverting GHs) or the covalent intermediate (retaining GHs) should adopt a conformation close to one of the four possible TS conformations ($^{4}H_{3}$, $^{3}H_{4}$, $^{2,5}B$ and $B_{2,5}$). Therefore, the conformational itineraries followed by the substrate of GHs can be classified according to the TS conformation adopted by the -I sugar molecule (Davies, 2012).

The Cartesian or Stoddart diagram (see section 2.1) has been used to systematically classify the existing information about the conformational itineraries of different GHs (Vocadlo & Davies, 2008)(Figure I-10). Interestingly, GHs that belong to the same family or act on the same type of substrate tend to use the same conformational itinerary (i. e. same TS conformation). For example, retaining β-D-mannosidase enzymes use a $B_{2,5}$ transition state conformation to perform the hydrolysis reaction (Figure I-9), whereas retaining β-D-glucosidases use a $^{4}H_{3}$ TS.

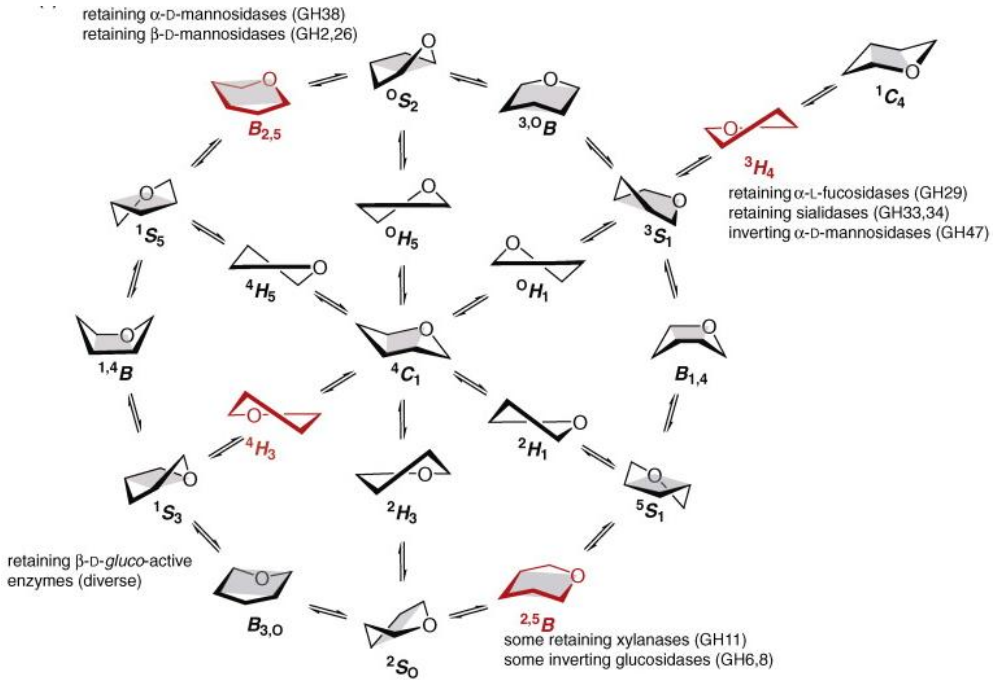

Figure I - 10. Cartesian or Stoddart diagram (Northern projection, including one itinerary that extends toward the South Pole, see Figure I – 4) of all possible pyranose ring conformations and their interconversions. In red are the TS conformations employed by different GH types and families. Figure taken from reference (Vocadlo & Davies, 2008).

4. Glycoside Transferases.

Glycoside transferases (GTs; E.C. 2.4.x.y) constitute a large family of enzymes that are involved in the biosynthesis of oligosaccharides and glycoconjugates (Taniguchi, 2002). In particular, they catalyze the formation of glycosidic linkages by the transfer of a saccharide, typically a monosaccharide from an activated donor substrate, to an acceptor substrate (Lairson, 2008)(Figure I-11). The most common sugar donors used by GTs are nucleotide sugars, often named as "Leloir" donors, and lipid or phospho-sugar donors ("non-Leloir" donors). Possible acceptor molecules include other saccharides, lipids, proteins (N- and O-glycosylation), nucleic acids, and numerous small molecules such as antibiotics, flavonols, and so on (Breton, 2012).

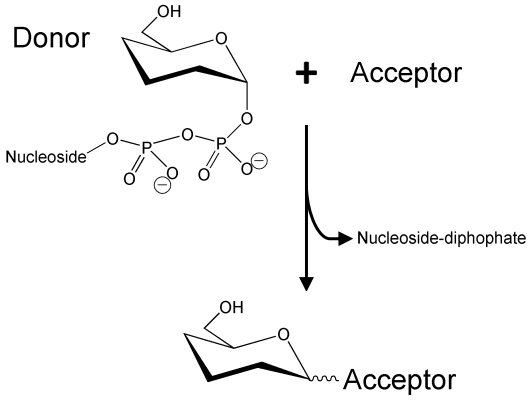

Figure I - 11. Glycosyl transfer reaction catalyzed by Leloir GTs. The donor molecule is an activated monosaccharide and the acceptor molecule can be a sugar, a lipid, a protein, or other small molecules.

4.1 Structure of GTs.

Similarly to GHs, GTs have been classified into families based on amino acid sequence similarities. More than 90 families are currently included in the CAZy database (Lombard, 2014). Despite the large number of enzyme families, GTs exhibit a limited number of fold types, namely GT-A (Charnock, 1999)(Persson, 2001), GT-B (Vrielink, 1994)(Gibson, 2002) and GT-C folds.

The GT-A fold consists of an α/β/α sandwich that resembles a Rossmann fold (Figure I-12a). The central β-sheet is flanked by a smaller one and the association of both creates the active site. The N-terminal region, which recognizes the sugar-nucleotide donor, comprises several β-strands flanked by α-helices, whereas the C-terminal, which contains the acceptor binding site, consists largely of mixed β-sheets. A general feature of all enzymes of the GT-A family is the presence of a common DXD motif and the requirement of a divalent cation (Mg^{2+} or Mn^{2+}) for activity.

The GT-B fold consists of two separate Rossmann fold domains with a connecting linker region and a catalytic site located between the two domains (Figure I-12b). There is an excellent high structural conservation between protein members of the GT-B family, particularly in the C-terminal domain, which corresponds to the nucleotide-binding domain. Variations are more pronounced in the N-terminal domains, in particular in the loops and

helices that point towards the active site, which have evolved to accommodate very different acceptors. Unlike GT-A GTs, GT-B GTs are metal-ion independent and do not possess any DXD motif.

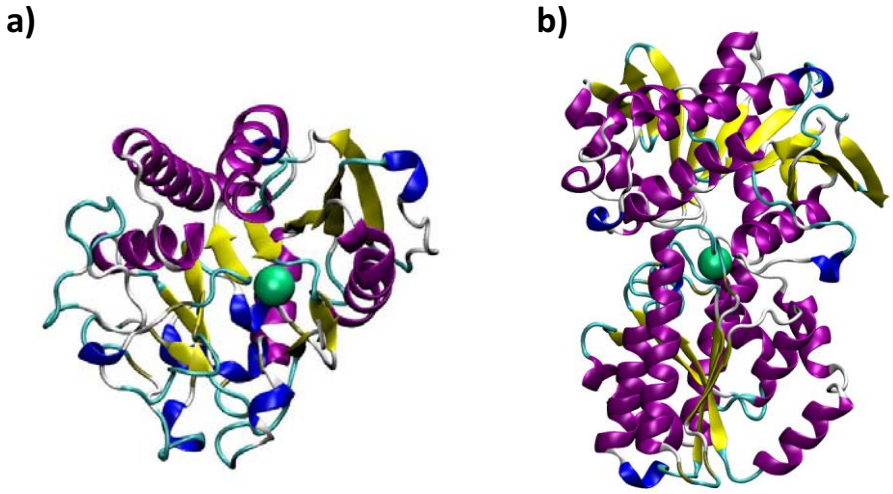

Figure I - 12. Cartoon representation of a) fold type A α-1,3-Galactosyl transferase (PDB code 2VFZ) and b) fold type B trehalose-6-phosphate synthase (PDB code 1GZ5). The active site of the enzymes is represented with a green sphere. Alpha helices are in purple, 3_{10} helices in blue, beta sheets in yellow and loops in cyan.

Both GT topologies (GT-A and GT-B) include enzymes that catalyze glycosyl transfer with inversion and retention of the anomeric configuration and, therefore, there is no correlation between the possible folds presented by GTs and the reaction stereochemistry.

A third fold type has also emerged, which comprises a bacterial sialyl transferase (CstII)(Chiu, 2004). This protein is composed of a single Rossmann domain and, therefore, is similar to a GT-A fold but presents some structural differences in the connectivity of the secondary structural elements, being considered as a new fold (GT-C).

4.2 Catalytic mechanism of GTs.

In contrast to the well characterized mechanistic strategies used by GHs to catalyze glycosidic bond hydrolysis (see section 3.1), the mechanisms of GTs remain less clear. Like GHs, GTs catalyze glycosyl transfer with either inversion or retention of the donor anomeric stereochemistry (Lairson, 2008). The mechanism of inverting GTs seems clearly established: an S_N2 reaction in a single displacement step involving a general base catalyst that increases the nucleophilicity of the attacking group, a mechanism that is analogous to the one of inverting GHs (Murray, 1997)(Lairson, 2008)(Figure I-12a). However, the mechanism of retaining GTs is still very controversial. A double displacement mechanism was early proposed by analogy to retaining GHs (Gastinel, 2001)(Figure I-12b). However, the observation that most GTs do not have a protein residue properly located to act as a nucleophile in a double displacement reaction leads to the suggestion of an alternative mechanism (Persson, 2001), in which the reaction proceeds via a front-face single displacement (Figure I-12c). In this mechanism, the nucleophilic hydroxyl group of the acceptor attacks the anomeric carbon at the same side from which the leaving group departs.

The front-face (or front-side) type of mechanism has little chemical precedence. It was first invoked to explain the retention of configuration in the decomposition of alkyl chlorosulfites (Lewis & Boozer, 1952) and the reaction was termed S_Ni (internal nucleophilic substitution). The front-face reaction has also been used for glycosyl transfer. Sinnott and Jencks (Sinnott & Jencks, 1980) proposed this reaction (but did not term it S_Ni) for the solvolysis of glycosyl fluorides in a mixture of ethanol and chloroethanol. The front-face mechanism in this system was later confirmed by measurements of the kinetic isotope effect (Chan, 2012). In the context of retaining GTs, several studies have proposed the front-face mechanism on the basis of structures of pseudo-ternary complexes that show a proper orientation of the donor and the acceptor (Persson, 2001)(Lairson, 2008).

The front-face reaction in GTs has been sometimes indistinctively referred as S_Ni, S_Ni-like or S_Ni-type (Persson, 2001)(Monegal & Planas, 2006)(Errey, 2010). In our opinion, this is a confusing terminology as the reaction is not "internal" (see discussion in Chapter VII section 3) as in the alkyl chlorosulfites reaction. Some authors use the term S_Ni to indicate a fully concerted reaction (i. e. with only one step) but S_Ni-like/type to refer to a step-wise reaction (Gómez, 2012). Nevertheless, Lairson, Henrissat, Davies and Withers only consider the term S_Ni-like for the front-face reaction in retaining GTs in the most comprehensive GT review to date, stating that: *"an enzyme-catalyzed S_Ni-like mechanism would in all likelihood proceed with the required formation of a short-lived ion pair intermediate with a lifetime longer than a bond vibration"* (Lairson, 2008).

Characterization of inhibitors of trehalose-6-phosphate synthase (OtsA), a classical glycosyl transferase of the GT-B fold, point towards a front-face mechanism (Errey, 2010). Furthermore, a KIE study on OtsA indicate a highly dissociative oxocarbenium ion TS or an oxocarbenium ion-like intermediate (both mechanisms cannot be excluded from an experimental point of view) with partial protonation of the leaving group by the incoming nucleophile (Lee, 2011). Independently, a computational study performed in our group confirms the experimental observations and show that a short-lived oxocarbenium ion-like TS is formed (Ardèvol, 2011), as predicted by Lairson *et al*. (Lairson, 2008).

A different scenario is provided by GTs that exhibit a possible nucleophile residue in the β face of the donor sugar. Recent experimental (Patenaude, 2002)(Monegal & Planas, 2006) (Soya, 2011) and computational (Gómez, 2013)(Rojas-Cervellera, 2013) investigations in family 6 GTs (mammalian alpha-3-galactosyl transferase (α3GalT) and blood group GTs) support the two-step displacement mechanism for glycosyl transfer in this GT family, with one study supporting also a front-face reaction (Gómez, 2013).

Concerning the overall catalytic process (binding of substrates and reaction), the reaction catalyzed by both, retaining an inverting GT enzymes, follows a sequential bi-bi catalytic mechanism (Lira-Navarrete, 2014) in which the sugar nucleotide binds to the enzyme, followed by the acceptor to yield the reactive ternary complex. After the transfer of the glycosyl moiety to the acceptor the product is released, followed by the nucleoside diphosphate and returning the enzyme to its original state.

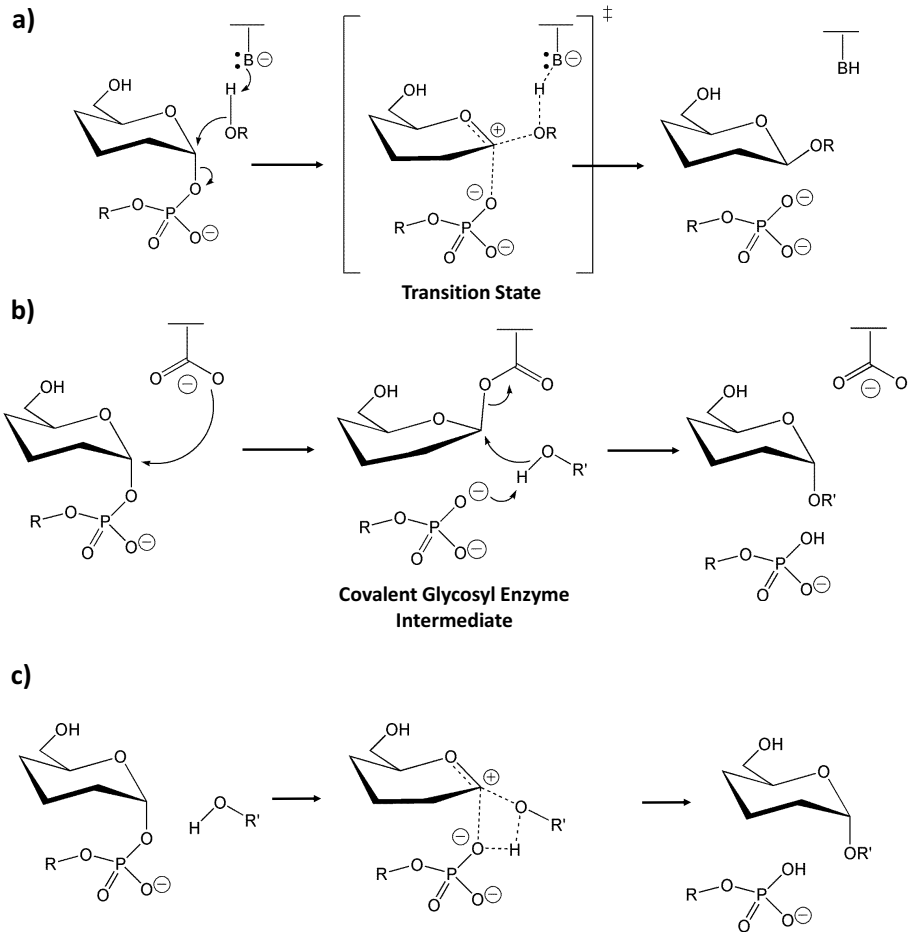

Figure I - 13. Reaction mechanisms proposed for inverting and retaining GTs. Inverting GTs operate via a single displacement with the formation of an oxocarbenium ion-like transition state (a). For retaining enzymes both double displacement (b) and front-face (c) mechanisms have been proposed.

5. Transition state mimic inhibitors.

Small molecule inhibitors are not only invaluable tools for elucidating enzyme function, but also, in a limited number of cases, may have applications as therapeutic compounds. Malfunction of carbohydrate processing enzymes have implications in a variety of diseases, such as cancer (Lau, 2008), diabetes (Bischoff, 1995) or viral infections (Gamblin, 2010). Therefore, both GHs and GTs are crucial targets for the design of compounds that modulate the activity of these enzymes.

Considering that enzymes catalyze reactions by a very tight binding and stabilization of the TS, inhibitors able to mimic features of this intermediate species should exhibit high affinities (Pauling, 1946)(Pauling, 1948). As a consequence, huge efforts have been made to develop inhibitors which incorporate features of the particular TS of each kind of enzyme. As mentioned before, for the particular case of the hydrolysis of glycosides, the TS formed during the reaction possess an oxocarbenium ion character, with substantial charge development at the anomeric carbon and partial C1-O5 double bond formation (see section 3.1). Therefore, TS mimicry inhibitors are being designed to mimic the positive charge of the TS and/or the double bond character of one intraring bond (i. e. enforcing ring planarity).

Applications of the above “recipe” lead to the synthesis of several compounds with a huge range of inhibitor potencies. This diversity of results made some scientist to question whether all these inhibitors present “true” TS mimicry capabilities, i. e., whether they mimic the TS or instead they mimic the ground-state substrate or form fortuitous interactions with the enzyme (Gloster, 2010). Different experimental approaches (linear free energy relationships (LFERs)(Ermert, 1993), thermodynamic measures of the inhibitor binding (Gloster, 2007), slow onset inhibition experiments (Schramm, 2005) and measurements of the inhibition pH dependence (Leger, 1999)) have been proposed to test TS mimicry. Of these, LFERs seems the most appropriate technique to test the likelihood of an inhibitor to mimicking the TS, although still gives a qualitative answer and different conclusions were drawn from the same ‘classes’ of inhibitors (see discussion in Gloster, 2012b).

Based structural analyses, it is known that isofagomine type inhibitors (Figure I-13a), which protonate inside the enzymatic catalytic cavity, resemble a glycosyl cation (i. e. charge is localized at C1) but are poor mimics of the TS conformation. Most of the X-ray structures show isofagomine bound in a $^{4}C_{1}$ conformation, which is none of the four possible TS conformations ($^{3}H_{4}$, $^{4}H_{3}$, $B_{2,5}$, $^{2,5}B$) discussed above for GHs (section 3.2). In contrast, mannoimidazole inhibitors (Figure I-13b) normally adopt one of these conformations, they display an sp^{2} hybridization and the imidazole functionality can be protonated (Heightman, 1999). Kinetic and thermodynamic analyses suggest that this compound is a true TS mimic (Tailford, 2008)(Thomson, 2012). Additionally, complexes with mannoimidazole can also report on both catalytic itineraries follow by α-mannosidases by mimicking the conformation of the TS ($B_{2,5}$ or $^{3}H_{4}$)(Tailford, 2008)(Thomson, 2012). However, all these considerations do not allow one to conclude which features of the inhibitor imbue TS mimicry capabilities. This aspect will be investigated in Chapter V of this Thesis.

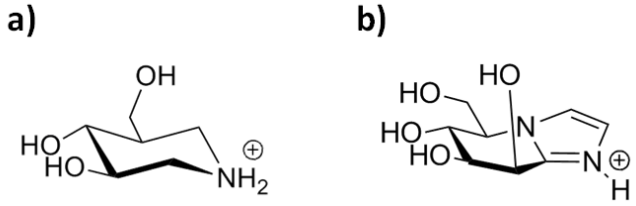

Figure I - 14. Isofagomine (a) and mannoimidazole (b) TS mimic inhibitors.

Objectives.

Objectives.

The present Thesis is aimed at unraveling the molecular basis for substrate recognition, preactivation and catalysis in several glycoside hydrolases (GHs) and transferases (GTs).

The following objectives have been pursued:

- Quantify the dependence of the free energy landscape of six-membered rings with respect to the choice of collective variables (Cartesian or Mercator coordinates) .

- Obtain the conformational free energy landscape of an isolated α-D-mannopyranose to understand the occurrence of mannoside conformations in X-ray structures of α-mannosidases. Determine the effect of the protein environment on the accessible conformations of the mannose molecule.

- Investigate the transition state mimicry properties of two common GH inhibitors (mannoimidazole and isofagomine) by computing its conformational free energy landscape.

- Find out whether the specific transition state conformations used by glucosidases, mannosidases, xylosidases and fucosidases, have any relation with the conformational preferences of the glucosyl, mannosyl, xylosyl and fucosyl cations, respectively.

- Unravel the reaction mechanism of an engineered glycoside hydrolase (*Sufolobus solfataricus* β-glucosidase E387Y) that is able to synthesize glycosidic bonds from activated sugar molecules with retention of the anomeric configuration.

- Find out the mechanism by which the enzyme N-acetylgalactosaminyl transferase GalNAc-T2 is able to form the glycosidic linkage between an activated sugar donor (UDP-GalNAc) and a peptide acceptor molecule with retention of the anomeric configuration.

Chapter II – Methods.

Methods.

1. Introduction to molecular dynamics.

In molecular dynamics (MD) simulations, successive configurations for a given system are generated by integrating Newton's laws of motion. The result is a trajectory that specifies how the positions and velocities of the particles vary with time for a given system. As an example, we can follow protein dynamics in space and time and obtain a valuable amount of information concerning structural and dynamic properties, such as molecular geometries and energies, local fluctuations of structure, enzyme substrate binding processes, etc.

MD techniques can be classified based on the type of equations used to calculate the energy of our system. In *classical* or molecular mechanics MD simulations (MM MD) the electrons are not considered explicitly and the system under study is represented as a collection of masses centered at the nuclei of the atoms. MM MD simulations allow to study the dynamics of large systems during long simulation times but they are unable to describe electronic processes such as chemical reactions or charge reorganizations. On the other hand, *ab initio* or quantum mechanical MD simulations (QM MD) are based on solving the Schrödinger equation and, therefore, they take into account the electronic structure of our system. These simulations are computationally expensive and, therefore, can only be applied to relatively small systems and short simulation times.

1.1 Setting up and running a Molecular Dynamics Simulation.

Once the initial configuration of the system is selected, which may be obtained from experimental data, from a theoretical model or a combination of both, the next step consists in the determination of the initial velocities for all the atoms of our system. This can be done by selecting random velocities from a Maxwell-Boltzmann distribution at the temperature of interest. The Maxwell-Boltzmann equation provides the probability that an atom i of mass m_i, has a velocity v_{ix} in the *x* direction at a temperature T (Eq. II-1).

$$p(\boldsymbol{v}_{ix}) = \left(\frac{m_i}{2\pi kT}\right)^{1/2} e^{\left(-\frac{m_i \boldsymbol{v}_{ix}^2}{2kT}\right)} \quad (\mathbf{Eq.\,II-1})$$

Once the initial velocities were assigned, we can compute the force acting on each atom using a potential function that is different for each type of MD simulation. As an example, classical methods relay on force field equations where the energy depends only on the atomic coordinates (Eq. II-2).

$$\boldsymbol{E} = \boldsymbol{f}(\boldsymbol{r_i}) \quad (\mathbf{Eq.\,II-2})$$

Considering Newton's laws of motion, the acceleration that one particle experiences at each step is determined by the force acting upon it (Eq. II-3).

$$-\frac{\delta E}{\delta r_i} = \boldsymbol{F_i} = \boldsymbol{m_i} \cdot \boldsymbol{a_i} \quad (\mathbf{Eq.\,II-3})$$

For complex systems, the integration of the equations of motion cannot be solved analytically and, therefore, numerical techniques must be used. The integration algorithm can be derived from a Taylor series expansion of the atom coordinates at time *t* (Eq. II-4 and 5):

$$\boldsymbol{r}_i(t+\delta t) = \boldsymbol{r}_i(t) + \boldsymbol{v}_i\delta t + \frac{1}{2}\boldsymbol{a}_i\delta t^2 + \frac{1}{6}\boldsymbol{a}_i'\delta t^3 + \cdots \quad (\mathbf{Eq.\,II-4})$$

$$\boldsymbol{v}_i(t+\delta t) = \boldsymbol{v}_i(t) + \boldsymbol{a}_i\delta t + \frac{1}{2}\boldsymbol{a}_i'\delta t^2 + \frac{1}{6}\boldsymbol{a}_i''\delta t^3 + \cdots \quad (\mathbf{Eq.\,II-5})$$

where r_i, v_i, and a_i are the position, velocity and acceleration vectors, respectively, and δt is the time step. The length of this time interval (δt) is chosen so to ensure that the variation in the potential energy function of the atoms within each time step is small. When simulating flexible molecules, a useful guide is that the time step should be approximately one-tenth of the shortest period of motion. In organic molecules, such as enzymes, the highest frequency vibrations are due to bond stretches (compression/extension of bonds), especially those involving hydrogen atoms. A C—H bond vibrates with a period of approximately 10 fs, so a time step of 1 fs should be a reasonable value (Schlick, 2000).

In MD applications, the most widely used method for integrating the equations of motion is, probably, the *Verlet* algorithm (Verlet, 1967). This method requires the positions and accelerations at time *t*, and the positions from the previous step, $\boldsymbol{r}(t-\delta t)$, to calculate the new positions at time, $\boldsymbol{r}(t+\delta t)$. From Eq. II-4 we can also write the following relationship (Eq. II- 6).

$$\boldsymbol{r}_i(t-\delta t) = \boldsymbol{r}_i(t) - \boldsymbol{v}_i\delta t + \frac{1}{2}\boldsymbol{a}_i\delta t^2 - \frac{1}{6}\boldsymbol{a}_i'\delta t^3 + \cdots \quad (\mathbf{Eq.\,II-6})$$

Adding Eq. II-4 and 6 results in Eq. II-7:

$$\boldsymbol{r}_i(t+\delta t) = 2\boldsymbol{r}_i(t) - \boldsymbol{r}_i(t-\delta t) + \boldsymbol{a}_i\delta t^2 + \cdots \quad (\mathbf{Eq.\,II-7})$$

Velocities do not appear explicitly in the *Verlet* algorithm, but they can be calculated in a variety of ways. A simple approach is dividing the difference of the expansions $r_i(t+\delta t)$ and $r_i(t-\delta t)$ by 2δt (Eq. II-8):

$$\boldsymbol{v}_i(t) = \frac{[\boldsymbol{r}_i(t+\delta t) - \boldsymbol{r}_i(t-\delta t)]}{2\delta t} \quad (\mathbf{Eq.\,II-8})$$

Several variations on the Verlet algorithm (Verlet, 1967) have been developed. The simplest among these is the so-called *leap-frog* algorithm (Hockney, 1970), which is currently implemented in the AMBER software. In this algorithm, first the velocities are calculated at half integer time steps t+1/2δt, then these are used to calculate the position at time t+δt (Eq. II-9 and 10).

$$\boldsymbol{v}_i\left(t+\frac{1}{2}\delta t\right)=\boldsymbol{v}_i\left(t-\frac{1}{2}\delta t\right)+\boldsymbol{a}_i(t)\delta t^2 \quad (\mathbf{Eq.\,II-9})$$

$$\boldsymbol{r}_i(t+\delta t)=\boldsymbol{r}_i(t)+\boldsymbol{v}_i(t+\frac{1}{2}\delta t)\delta t \quad (\mathbf{Eq.\,II-10})$$

2. Classical or Molecular Mechanics Molecular Dynamics (MM MD).

2.1 Introduction to MM MD methods.

MM MD methods ignore the electronic motions and calculate the energy of our system based on simple functions of the atomic positions $E=f(r_i)$. These methods are able to simulate large systems during long periods of time thanks to the validity of several assumptions, which enable to reduce computational costs (Leach, 2001). The first one is the Born-Oppenheimer approximation, which allows computing the energy as a function of just the nuclear coordinates. Electrons are not considered, but point charges, localized on each nuclear center, are used to reproduce the electrostatic properties of each molecule. Additionally, simple functions (e. g. Hooke's law) are used to describe processes such as the stretching of bonds, the opening and closing of angles and the rotations about single bonds. These functions performed quite acceptably and are easy to compute. Transferability is also a key attribute of a force field and refers to the utilization of parameters developed from data on small molecules (e. g. single amino acid) into much larger molecules such as polymers or proteins. In other words, the effect of the molecular environment on these parameters is neglected (i. e. assumed to be not relevant), allowing for a simple calculation.

In MM MD methods, the energy of a molecule in a particular configuration is calculated from the ideal energy of the entities that constitute the system. The energy is relative to the coordinates of a reference state and is calculated as the sum of different terms to indicate the penalty for departure from ideality of bond distances, angles, torsions, etc.

2.2 Force fields.

A force field is defined as the set of equations employed to calculate the different contributions to the energy, along with the values of structural and other adjustable parameters. These parameters are obtained by fitting a potential function with experimental data and high level quantum chemical calculations performed on small molecules. In this Thesis, we have used the AMBER force field (Cornell, 1995), where the additive potential energy function has the following form (Eq. II-11):

$$E = E_{bonds} + E_{angles} + E_{dihedrals} + E_{VdW} + E_{electrostatic} =$$

$$= \sum_{bonds} k_b \cdot (r - r_{eq})^2 + \sum_{angles} k_\theta \cdot (\theta - \theta_{eq})^2 + \sum_{dihed.} \frac{V_n}{2} \cdot [1 + \cos(n\varphi - \alpha)]$$
$$+ \sum_{i<j} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \sum_{i<j} \frac{q_i \cdot q_j}{\varepsilon \cdot R_{ij}} \quad (\mathbf{Eq.\,II-11})$$

where E_{bonds} and E_{angles} represent the energy of the summation over all bonded atoms and angles between three bonded atoms, respectively, modeled by and harmonic potential (r_{eq} and θ_{eq} refer to the equilibrium bond lengths and angles, k_b and k_θ are the vibrational constants). The third term is a torsional potential that describes how the energy changes as the bond rotates (V_n is the torsional barrier corresponding to *n* multiplicity and α phase).

Interatomic interactions between atoms not directly linked are also considered, which have an important contribution to the total energy of the system. These are "through space" interactions and are usually modeled as a function of some inverse power of the distance. The *van der Waals* interactions (E_{VdW}) are described by a Lennard-Jones potential (Eq. II-11), containing an attractive term due to dispersion ($1/R^6$) and a repulsive term due to exchange repulsion ($1/R^{12}$). The electrostatic interaction energy ($E_{electrostatic}$), between pairs of atoms, is then calculated as a sum of the interactions for pairs of atomic point charges using Coulomb's law (Eq. II-11).

3. Quantum mechanics (QM).

3.1 The Schrödinger equation.

MM MD methods do not allow for studying chemical reactions, since electrons are not taken explicitly into account. Inclusion of electrons requires however a quantum mechanical description. The starting point for any discussion of quantum mechanics is, of course, the Schrödinger equation (Leach, 2001). The full, time-dependent form of this equation is:

$$\left\{-\frac{\hbar^2}{2m}\left(\frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2} + \frac{\delta^2}{\delta z^2}\right) + v\right\}\psi(\boldsymbol{r}, t) = i\hbar\frac{\delta\psi(\boldsymbol{r}, t)}{\delta t} \quad (\mathbf{Eq.\,II-12})$$

The equation above refers to a single particle (e. g. an electron) of mass *m*, which is moving through the space at time *t* under the influence of an external field *v* (which might be the electrostatic potential due to the nuclei of a molecule). ħ is the Planck's constant divided by 2π and ψ is the wavefunction, which characterizes the particle's motion. When the external potential is constant and, therefore, independent of time then the wavefunction can be written as the product of a spatial part and a time part ($\psi(r,t) = \psi(r)\cdot T(t)$). This characteristic enables to write the Schrödinger equation in a time-independent form (Eq. II-13). Here *E* refers to the energy of the particle.

$$\left\{-\frac{\hbar^2}{2m}\nabla^2 + v\right\}\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}) \quad (\mathbf{Eq.\,II-13})$$

It is usual to abbreviate the left-hand side of the Schrödinger equation as $H\psi$, where H is the Hamiltonian operator. As can be deduced from the previous equation, the Hamiltonian operator is composed of two terms that reflect the contributions of the kinetic and potential energies to the total energy of the system (Eq. II-14 and 15). Here, Z refers to the nuclear charge and r is the distance of the electron from the nucleus.

$$H = -\frac{\hbar^2}{2m}\nabla^2 - \frac{Ze^2}{4\pi\varepsilon_0 r} \quad (\mathbf{Eq.\,II-14}) \quad \text{and in atomic units} \quad H = -\frac{1}{2}\nabla^2 - \frac{Z}{r} \quad (\mathbf{Eq.\,II-15})$$

This reduces the Schrödinger equation to $H\psi = E\psi$. To solve the Schrödinger equation it is necessary to find values of E and functions ψ such that, when the wavefunction is operated upon by the Hamiltonian, it returns the wavefunction multiplied by the energy.

In quantum chemistry, for each particle described by a wavefunction, the product $\psi^*\cdot\psi$ (where ψ^* is the complex conjugate of the wavefunction) represents the probability of finding the particle at a position r. More generally, for each classical observable, such as position or momentum, there is a corresponding quantum mechanical operator, which yields the exact value for this operator when acting on the wavefunction. The most important operator, the Hamiltonian, enables to calculate the energy of a given system. For a molecule composed of several nuclei and electrons, the Hamiltonian has an additional term, the electron-electron potential (v_{ee}), due to the electron's repulsion (Eq. II-16).

$$H = T + v_{ext} + v_{ee} \quad (\mathbf{Eq.\,II-16})$$

The wavefunction depends on both the electronic and nuclear degrees of freedom, but the electrons are by orders of magnitude lighter than the nuclei and move much faster. Within the Born-Oppenheimer approximation, nuclear and electronic degrees of freedom could be separated and the problem is reduced to the calculation of the energy of the system for a given nuclear configuration.

3.2 Density Functional Theory.

As a result of the Born-Oppenheimer approximation, the Coulomb potential arising from the nuclei is treated as a static external potential v_{ext}. The remainder of the electronic Hamiltonian includes the kinetic part and the electron-electron interaction (Eq. II-17).

$$H_e = -\frac{1}{2}\sum_i \nabla_i^2 + \frac{1}{2}\sum_{i<j}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \quad (\mathbf{Eq.\,II-17})$$

Density functional theory (DFT) provides a framework to obtain the total energy of a polyatomic system given their atomic coordinates. In the mid of 1960s, Hohenberg and Kohn demonstrated that the ground state energy of a system of interacting electrons subject to an external potential, v_{ext}, is a unique functional of the electron density (Hohenberg & Kohn, 1964). In DFT, the energy functional is written as a sum of two terms (Eq. II-18), where the first term refers to the interaction of the electron density and the external potential, v_{ext}, and the latter is the electronic Hamiltonian.

$$E[\rho(\boldsymbol{r})] = \int v_{ext}(\boldsymbol{r})\rho(\boldsymbol{r})\delta(\boldsymbol{r}) + H_e[\rho(\boldsymbol{r})] \quad (\mathbf{Eq.\,II-18})$$

In order to obtain the exact value of the energy, one can apply a variational principle approach. It states as follows: *"The expectation value of a Hamiltonian, the energy E_i, calculated with a trial wavefunction, is never lower in value than the true ground state energy, E_0, calculated with the true ground state wavefunction" (Griffiths, 1995)*. This principle is also valid for DFT and, therefore, by minimizing E_i with respect to the electron density the energy could be calculated (Eq. II- 19).

$$E^{DFT} = \min_{\rho(\boldsymbol{r})} \left(E[\rho(\boldsymbol{r})]\right) \quad (\mathbf{Eq.\,II-19})$$

The difficulty within the previous description is that we do not know what the function $H_e[\rho(\boldsymbol{r})]$ is. Later, Kohn and Sham demonstrated that there is an equivalence between the electronic density of our system (real system) and that of a model system of non-interacting electrons which are subjected to an effective potential, v_{eff} (Kohn & Sham, 1965). This provided a way to solve the problem of finding the density of the many-electron system, via obtaining the electron density of the non-interacting system. The electronic density $\rho(\boldsymbol{r})$ can be expressed in terms of single-electron orbitals $\psi_i(\boldsymbol{r})$ (Eq. II-20), known as Kohn-Sham (KS) orbitals,

$$\rho(\boldsymbol{r}) = 2\sum_{i}^{occ.} |\psi_i(\boldsymbol{r})|^2 \quad (\mathbf{Eq.\,II-20})$$

where the sum extends over the occupied single-particle orbitals.

In conclusion, the energy functional can be expressed in terms of the density of the single-electron orbitals (ψ_i) for a given nuclear coordinates ($\boldsymbol{R_N}$) (Eq. II-21).

$$E^{DFT} = \min_{\{\psi_i\}} \left(E^{KS}[\{\psi_i(\boldsymbol{r})\}, \{\boldsymbol{R_N}\}]\right) \quad (\mathbf{Eq.\,II-21})$$

3.3 Kohn-Sham energy functional.

The energy functional, E^{KS}, as proposed by Kohn and Sham (Kohn & Sham, 1965), can be written as:

$$E^{KS} = 2\sum_{i}^{occ} \int \psi_i^*(\boldsymbol{r}) \left(-\frac{\nabla^2}{2}\right) \psi_i(\boldsymbol{r})\delta(\boldsymbol{r}) + \int V(\boldsymbol{r})\rho(\boldsymbol{r})\delta(\boldsymbol{r}) + \frac{1}{2}\int \frac{\rho(\boldsymbol{r})\rho(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\delta\boldsymbol{r}\delta\boldsymbol{r}' + E_{xc}[\rho(\boldsymbol{r})] \quad (\mathbf{Eq.\,II-22})$$

The first term on the left-hand side of this expression is the kinetic energy of the non-interacting electrons. The second term corresponds to the interaction of the electrons with the nuclear charges ($V(\boldsymbol{r})$ is the potential as a result of the nuclei). In case only valence electrons are explicitly considered in the calculation, $V(\boldsymbol{r})$ would be a pseudopotential (see more details in section 3.5). The third term corresponds to the classical Coulomb interaction of a density distribution ρ. Finally, the fourth term, $E_{xc}[\rho(\boldsymbol{r})]$, is a functional of the density that accounts for the remaining contributions to the electron-electron interaction.

All terms in the expression of the energy are exact and calculable according to the density functional theory except the last term, $E_{xc}[\rho(r)]$. In this case, DFT do not provide an explicit form for its calculation and it has to be approximated. The most common approximations used are:

- *Local density approximation (LDA).*

Usually, the $E_{xc}[\rho(r)]$ is approximated by the exchange correlation energy of a uniform electron gas with density $\rho(r)$, which is precisely known. In this approximation, it is assumed that the exchange and correlation energy of an electron at a point depends on the density at that point instead of the density at all points in the space (Koch & Holthausen, 2000).

One of the main drawbacks of LDA is that the *van der Waals* interactions, which originated from motions of the electrons clouds caused by the interactions between atoms, cannot be properly described. Therefore, special care should be taken when addressing problems in which *van der Waals* interactions might play a relevant role, such as stacking interactions between π-systems and the diffusion of ligands in purely hydrophobic cavities (Kriegl, 2003).

- *Generalized Gradient Approximation (GGA).*

The accuracy provided by LDA is not enough for most applications in chemistry and biology. One of its main drawbacks is that the bond distances and binding energies can have large errors that appear in a nonsystematic way. This represents a serious problem for the study of ligand-protein interactions.

A step forward with respect to LDA is the generalized gradient approximation methodology (GGA). This approach is based on using not only the density at a point, but also the gradient of the density, $\nabla\rho(\boldsymbol{r})$, in the functional expression in order to account for the nonhomogeneity of the true electron density. In practice, E_{xc}^{GGA}, is usually split into two terms corresponding to its exchange and correlation contributions and separate forms for each term are provided. Among the most popular GGA exchange and correlation functionals used in biological applications are the ones denoted as BP86 [exchange part by Becke (Becke, 1984) and correlation by Perdew (Perdew, 1986)] and BLYP [combination of Becke exchange and correlation developed by Li, Yang and Parr (Lee, 1988)]. Hybrid functionals, which include to some extent the exact exchange energy in the functional expression, are widely used. One of the most popular used to be the B3LYP [exact exchange developed by Becke (Becke, 1993) combined with the LYP correlation functional], but other hybrid functionals are nowadays also widely used (e. g. hybrid meta-GGA functionals (Zhao & Truhlar, 2004)).

The use of the GGA improves considerably the description of the bonding interactions with respect to pure LDA with a very low additional computational cost. The description of weak *van der Waals* interactions, however, remains problematic. In this work, we have used the generalized gradient correction of Perdew, Burke and Ernzerhoff (PBE functional)(Perdew, 1996), which is suitable for biological studies due to its good description of the hydrogen bond interactions (Ireta, 2004) and has been previously used with success for the calculation of mechanisms in carbohydrate-active enzymes (Biarnés, 2006)(Biarnés, 2011)(Ardèvol, 2010). For sugar conformations, the PBE functional gives similar results as hybrid functionals (Biarnés, 2007).

3.4 Kohn-Sham Equations.

Kohn and Sham (KS) proposed a coupled set of differential equations (Eq. II-23 and 24), known as Kohn-Sham equations (Kohn & Sham, 1965), which permit the calculation of the single electron orbitals, ψ_i, (Eq. II-24)

$$\left(-\frac{\nabla^2}{2} + V(\boldsymbol{r}) + \underbrace{\int \delta \boldsymbol{r} \frac{\rho(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{r}'|} + V_{xc}(\boldsymbol{r})}_{v_{eff}} \right) \psi_i(\boldsymbol{r}) = \varepsilon_i \psi_i(\boldsymbol{r}) \quad (\mathbf{Eq.\,23})$$

where ε_i are the eigenvalues of the matrix of Lagrange multipliers and are called the Kohn-Sham eigenvalues or Kohn-Sham orbital energies. The exchange correlation potential $V_{xc}(\boldsymbol{r})$ follows from equation:

$$V_{xc}(\boldsymbol{r}) = \frac{\delta E_{xc}[\rho(\boldsymbol{r})]}{\delta \rho(\boldsymbol{r})} \quad (\mathbf{Eq.\,II-24})$$

The KS equations can be solved iteratively given an initial guess for the set of single electron orbitals [$\psi_i(\boldsymbol{r})$] to obtain the energy of the system. Alternatively, the total energy can be minimized with respect to the $V_{xc}(\boldsymbol{r})$ using gradient search techniques.

In summary, DFT provides a framework to find the total energy of a many-electron interacting system by means of solving one-electron equations of a model non-interacting system that shares the same density.

3.5 Basis set.

To solve numerically the KS equations (Eq. II-23 and 24), the KS single electron orbitals (Eq. II-20) are expanded in a basis set (Eq. II-25).

$$\psi_i(\boldsymbol{r}) = \sum_j c_j \cdot \emptyset_{ij}(\boldsymbol{r}) \quad (\mathbf{Eq.\,II-25})$$

This expansion should in principle extend to infinity, but it is generally truncated that only a limited set of basis functions is used. In the chemistry community, Gaussian functions (Eq. II-26) are very popular:

$$\psi_i(\boldsymbol{r}) = \sum_j c_j \cdot e^{-\alpha_j r^2} \quad (\mathbf{Eq.\,II-26})$$

Several notations are used to specify a particular set of atomic Gaussian functions, such as Pople's split valence basis set [e. g., 3-21G, 6-31G* or 6-311++G(2d,2p)]. These notations usually depend on the number of functions representing every atomic angular momentum and the spread of the Gaussian function, which is given by the exponent value α (large/small values result in compact/diffuse functions).

In the physics community, plane waves (PW) (Eq. II-27) are commonly used to expand the KS orbitals,

$$\psi_i(\boldsymbol{r}) = \frac{1}{\Omega^{1/2}} \sum_{G}^{G_{max}} C_G \cdot e^{i\boldsymbol{G}\boldsymbol{r}} \quad (\mathbf{Eq.\,II-27})$$

where Ω is the volume of the cell and G is the plane wave momentum. PW basis set is 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 can be approximated as

$$N_{PW} \approx \frac{\Omega}{6\pi^2} E_{cut}^{3/2} \quad (\mathbf{Eq.\,II-28})$$

PWs are not centered at the atoms but extend throughout all the space and usually a large number of basis functions is necessary. Also, PWs are highly inefficient to describe the core electron wavefunctions and addition of a very large extra set of functions is necessary to overcome this limitation. On the other hand, the core levels are well separated in energy from the valence electrons and they do not play a role in the chemical bonding properties of the system. Thus, in order to reduce computational costs, the core electron orbitals are frozen into the KS equations and only the valence electrons are described explicitly. Their interactions with the nuclei and valence electrons are described by using effective potentials, the pseudopotentials.

The pseudopotentials are usually derived from all electron atomic calculations and several types have been proposed to date. Here, we used norm-conserving pseudopotentials derived with the Martins-Troullier method (Troullier & Martins, 1991).

4. Quantum Mechanical MD simulations (QM MD).

Assuming the validity of the Born-Oppenheimer approximation (i. e. the electrons are moving in the field of a fix nuclei), QM MD simulations can be view as a series of DFT calculations at different instants of time, each one for a different set of atomic positions. At the beginning of the simulation, the Kohn-Sham equations have to be solved in an iterative process, from an initial set of one-electron orbitals $\{\psi_i(\boldsymbol{r})\}$, and the total energy of the system is calculated to subsequently evaluate the gradients on the nuclear coordinates. At this point, the coordinates of the atoms can be propagated through the equations of motion. From the new positions of the nuclei, it is necessary to repeat again the iterative process to obtain the total energy of the system (Figure II-2a). In a QM methodology the energy depends on the atomic positions and electron density (Eq. II-29).

$$\mathcal{L} = E_N^{kin} - E_{el} \quad (\mathbf{Eq.\,II-29})$$

The process of minimizing the functional, and get the new one-electron orbitals, is the bottleneck step of this kind of MD simulations due to its high computational cost, limiting its applicability.

5. Car-Parrinello MD Simulations (CPMD).

In the 80's Car and Parrinello (CP)(Car & Parrinello, 1985) developed a method that combines MD and DFT, in which the electronic and nuclear degrees of freedom evolve simultaneously according to a modified set of classical equations of motion (a fictitious electron mass (μ) is assigned to the electronic Kohn-Sham orbitals $\{\psi_i(\boldsymbol{r})\}$). Forces are evaluated on both the nuclei and the electrons applying the Kohn-Sham equations (Figure II-2b). The CP method is based on the description of N interacting atoms via the following Lagrangian:

$$\mathcal{L} = E_N^{kin} + E_{el}^{kin} - E^{KS} + \sum_{ij} \Lambda_{ij} \left(\int \delta \boldsymbol{r} \cdot \psi_i^*(\boldsymbol{r}) \cdot \psi_j^*(\boldsymbol{r}) - \delta_{ij} \right) \quad (\boldsymbol{Eq.II-30})$$

where $E_{el}^{kin} = \sum \mu \int \delta \boldsymbol{r} \left| \dot{\psi}_i(\boldsymbol{r}) \right|^2$ term is the "fictitious" kinetic energy associated with the electronic subsystem $\{\psi_i(r)\}$ of electronic mass μ, E_N^{kin} is the classical kinetic energy of the nuclei and Λ_{ij} are the Lagrangian multipliers that impose the orthonormality constraints between orbitals along the simulation. In this case, the Lagrangian is expressed according to the DFT formalism and, therefore, the term E^{KS} corresponds to the potential energy (Kohn-Sham energy). The total energy in the Car-Parrinello method is a constant of motion and is given by the following equation (Eq. II-31).

$$E_{tot}^{CP} = E_{el}^{kin} + E_N^{kin} + E^{KS} \quad (\mathbf{Eq.II-31})$$

The basic procedures for a QM and CPMD methodologies are illustrated in Figure II-2.

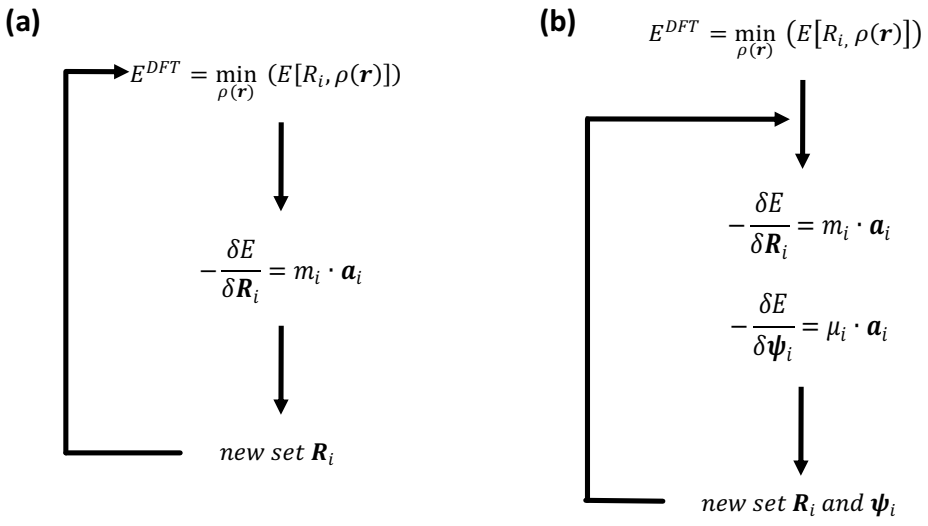

Figure II - 1. Schematic diagrams of a QM DFT (a) and CP (b) molecular dynamics. Adapted from Rovira, 2005.

In a CPMD simulation, the integration of the equations of motion provides the time evolution of not only the atomic positions (R_i) but also of the KS orbitals (ψ_i) and, therefore, both electrons and nuclei evolve simultaneously. This allows for the calculation of the subsequent steps of the simulation without additional minimization of the density functional and thus reducing the computational cost of the simulation.

The stability of a CPMD simulation has a strong dependence on the value of the electronic mass (μ) and the time step used along the simulation. As mentioned before, the time step is related to the shortest period of motion, which corresponds to the electronic degrees of freedom for CPMD simulations. To ensure that the electronic energy differs slightly from the exact DFT energy, electronic and nuclear frequencies must be separated to ensure an adiabatic behavior between them. 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. If the energy exchange between the electronic and nuclear subsystems is small (adiabatic behavior), the trajectory generated will be identical to the one obtained in a QM MD simulation (Marx, 2000).

6. Hybrid QM/MM methods.

Whereas MM methods allow to treat large biological systems, through the use of parameterized force fields, QM methods enable calculations on small systems taking into account the electronic properties, without prior parameterization. Therefore, hybrid techniques that use the advantages of both methods are very useful to model reactive processes in proteins.

The interest in multiscale approaches, which are able to take into account different levels of accuracy in the description of large systems, is continuously growing and has been honored by last year's Nobel Prize in Chemistry that was jointly awarded to M. Karplus, M. Levitt and A. Warshel. In fact, the combination of quantum mechanics and molecular mechanics (QM/MM), introduced by the seminal work of Warshel and Levitt (Warshel & Levitt, 1976), together with the increasing computational power of modern parallel, vector-parallel and hybrid CPU-GPU platforms, has lead to a real breakthrough in the simulations of biomolecular reactions (for a review, see for instance reference (Senn & Thiel, 2009)).

Hybrid quantum mechanics/molecular mechanics (QM/MM) techniques enable to treat the region of biological interest, normally the active site of the protein, at the QM level (QM part) while the remaining part of the protein, assumed to be chemically inert, as well as solvent and counter ions are treated with MM methods (MM part).

The Hamiltonian for a hybrid system can be written as:

$$H = H_{QM} + H_{MM} + H_{QM/MM} \quad (\mathbf{Eq.\,II-32})$$

where H_{QM} is the Hamiltonian for the quantum subsystem, H_{MM} is the Hamiltonian for the classical subsystem and $H_{QM/MM}$ contains the interactions between the two regions. The crucial issue regarding hybrid methods is the definition of the $H_{QM/MM}$ and, more precisely, the treatment of the interactions involving both the classical and the quantum parts of the system.

In this Thesis we have used the QM/MM methodology developed by Laio, Vandevondele and Röthlisberger (Laio, 2002a), which combines Car-Parrinello and classical (GROMOS or AMBER force fields) molecular dynamics (hereafter CPMD/MM). In this implementation, H_{QM} is described at the DFT level and also includes the polaritzation effects of the MM atoms on the

electronic density. H_{MM} calculation is based on the AMBER parameters with the GROMOS force field equation. The interactions coupling QM and MM degrees of freedom ($H_{QM/MM}$) can be grouped in bonded and nonbonded (Laio, 2002a).

6.1 Bonded interactions.

Bonded interactions between the QM and MM regions, which include bonds, angles and dihedrals, are taken into account by the classical force field. Classical parameters provide stable geometries at the interfaces but they have limited accuracy when the boundary is distorted upon a chemical reaction. As a consequence, the QM/MM boundary must be place far from the active site of the enzyme. Additionally, when the QM/MM boundary cuts through a covalent bond, unsaturated valence orbitals are present, which strongly perturb the electronic structure. Several approaches have been developed to address this problem, such as the link-atom approach or the hydrogen capping methodology.

In this work, we have employed the link-atom approximation (von Lilienfeld, 2005) based on monovalent pseudopotentials that saturate the QM region. This approach is conceptually simple and does not introduce unphysical atoms in the system but, on the other hand, increases the computational cost in a PW scheme due to the QM treatment of the link-atom.

6.2 Non-bonded interactions.

The non-bonding interactions between the QM and the MM regions are divided into electrostatic and van der Waals interactions (Eq. II-33). The steric non-bonded interactions due to Pauli repulsion and the dispersion interactions are usually kept into account in a straightforward way by retaining the van der Waals interactions as described by the classical force field (see section 2.1). The electrostatic interactions between the QM electronic density and the MM point charges are more subtle to compute.

$$H_{QM/MM}^{non-bonded} = \sum_{i\,\epsilon\,MM} q_i \int \frac{\rho(\boldsymbol{r})}{|\boldsymbol{r}-\boldsymbol{r}_i|}\delta(\boldsymbol{r}) + \sum_{\substack{i\,\epsilon\,MM \\ j\,\epsilon\,QM}} V_{vdW}(\boldsymbol{r}_{ij}) \quad (\mathbf{Eq.\,II-33})$$

Calculation of the electrostatic interactions in a QM/MM scheme has two main problems: the *electron spill-out phenomenon* and the very high computational cost. The *spill-out* consists on the anomalous rearrangement of the QM electron density that tends to localize near positive classical point charges of the MM region. This effect is particularly pronounced in a PW basis set approach, in which the electrons are fully free to delocalize. To solve this problem, in the CPMD/MM implementation the Coulomb potential in the core region is modified. Whereas the 1/r behavior is maintained for large r, for values of r shorter than the covalent radius of atoms the Coulomb potential goes to a finite value. The electrostatic term becomes:

$$H_{QM/MM}^{electrostatic} = \sum_{i\,\epsilon\,MM} q_i \int \rho(\boldsymbol{r}) v_i(|\boldsymbol{r}-\boldsymbol{r}_i|)\delta(\boldsymbol{r}) \quad \text{where } v_i(r) = \frac{r_{ci}^4 - r^4}{r_{ci}^5 - r^5} \quad (\mathbf{Eq.\,II-34})$$

The problem of the high computational cost is related to the huge number of operations that would be required for the direct evaluation of the electrostatic energy in a PW scheme (of the order of $N_r \cdot N_{MM}$, where N_r is the number of real space grid points, ~100^3, and N_{MM} is the number of classical atoms, 10^5 or more in systems with biochemical relevance). In the CPMD/MM approach used here, the electrostatic interactions are taken into account within a multilayer approach, in which the MM region is treated in different degrees of simplification (Laio, 2002a) (Figure II-2).

For the classical atoms present in the first subregion (NN region), the electrostatic interactions between each classical point charge and the electron density of the QM system are explicitly calculated. The electrostatic interaction is also computed explicitly for atoms in the second subregion (MIX region) with a point charge higher than 0.1 e^- (absolute value).

Atoms located in the third subregion (ESP region), together with those in the MIX region with a point charge ≤ 0.1 e^-, interact with the electron density through classical coulomb interactions. These interactions are calculated between the MM partial charges, obtained from the force field, and the dynamically generated RESP charges (D-RESP)(Laio, 2002b) of the QM system. Outside the ESP region, interactions with the QM electronic density are simply described through multipole expansions in order to reduce the computational cost. The radii of each subregion needs to be tested for each system at the beginning of the project.

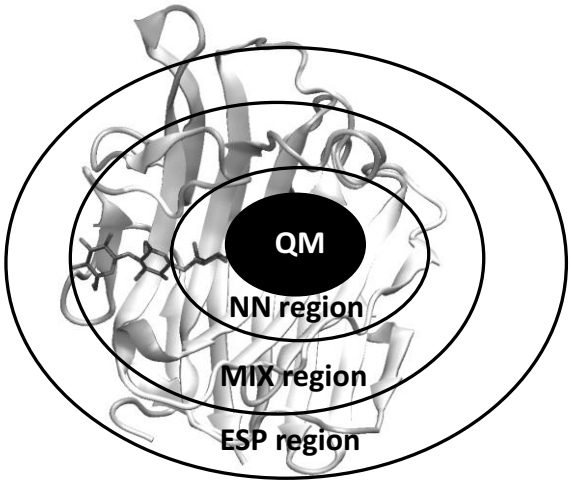

Figure II - 2. Schematic representation of the NN, MIX and ESP regions of electrostatic interactions used in the multilayer approach.

7. Metadynamics

Metadynamics is a novel MD based technique aimed at enhancing the conformational sampling of the phase space and the estimation of the free energy landscape (Laio & Gervasio, 2008). The algorithm is based on a dimensional reduction, defining a set of well-chosen collective variables (CVs) that allow to describe the essential modes associated with the transition of interest.

7.1 Direct metadynamics.

Let us consider (Laio & Gervasio, 2008) a system described by a set of coordinates *x* which is moving under the influence of an interatomic potential *V(x)*. If the energy barrier is higher than the thermal fluctuations (1/2·k_b·T), the system will remain oscillating around the energy minima and the crossing event will be unlikely to occur (*rare event*) within the typical MD simulation times (Figure II-3a). In order to accelerate the exploration of the conformational space we can use the metadynamics approach. For this particular system, we can define a CV *s,* which is an explicit function of a subset of coordinates *x'* [*s* = f(*x'*)], e. g. a distance or a coordination number.

The MD metadynamics simulation starts in one of the minima of the energy surface (Figure II-3a). At the beginning of the simulation, the force acting on the system is given by the gradient of the potential *V*:

$$f_i^V = -\frac{\delta V}{\delta r_i} \qquad (\mathbf{Eq.\,II-37})$$

After a time interval *t'*, a small Gaussian-like repulsive potential term (*V'*) is added, centered at the value of *s* at a time *t'* (Figure II-3b). This deposition process is repeated every time interval *t'* along the MD simulation. Therefore, at a time *t* the total potential is the sum of the interatomic potential *V* and the sum of the *V'* terms added up to that time, i. e. V_G:

$$V_T = V + V_G \quad \text{where} \quad V_G = \sum_{t\prime} w \cdot \exp\left(-\frac{[s(t)-s(t\prime)]^2}{2(\delta s)^2}\right) \qquad (\mathbf{Eq.\,II-38})$$

where *w* is the height of the Gaussian potential term, *δs* is its width and V_G is the history-dependent potential added (i. e. the number of Gaussian terms added during the simulation). As a result, the force acting on the system under study is now the sum of two components, one corresponding to the interatomic potential, *V*, and the other from the Gaussian potential, V_G:

$$f_i^{V_T} = f_i^V + f_i^{V_G} = -\frac{\delta V}{\delta r_i} - \frac{\delta V_G}{\delta r_i} \qquad (\mathbf{Eq.\,II-39})$$

The added potential V_G progressively counterbalance the underlying free energy, penalizing those regions of the phase space already visited (Figure II-3c). As a result, the system tends to escape from already visited regions to another energy minima, through the nearest and most accessible transition state. Obviously, the biasing potential performs a work on the system that needs to be dissipated by using a thermostat.

It has been demonstrated that for long simulation times, the bias potential will converge to the negative of the free energy (Figure II-3c) and then, since $V_G(s,t) = -F(s,t)$, it will oscillate around that profile (Eq. II-40). At this point, the motion of the CVs becomes diffusive in the region of interest and the simulation should be stopped.

$$\lim_{t\to\infty} V_G(s,t) \approx \sum_{t\prime} w \cdot \exp\left[-\frac{[\mathrm{s(t)} - \mathrm{s(t')}]^2}{2(\delta \mathrm{s})^2}\right] \qquad (\mathbf{Eq.\,II-40})$$

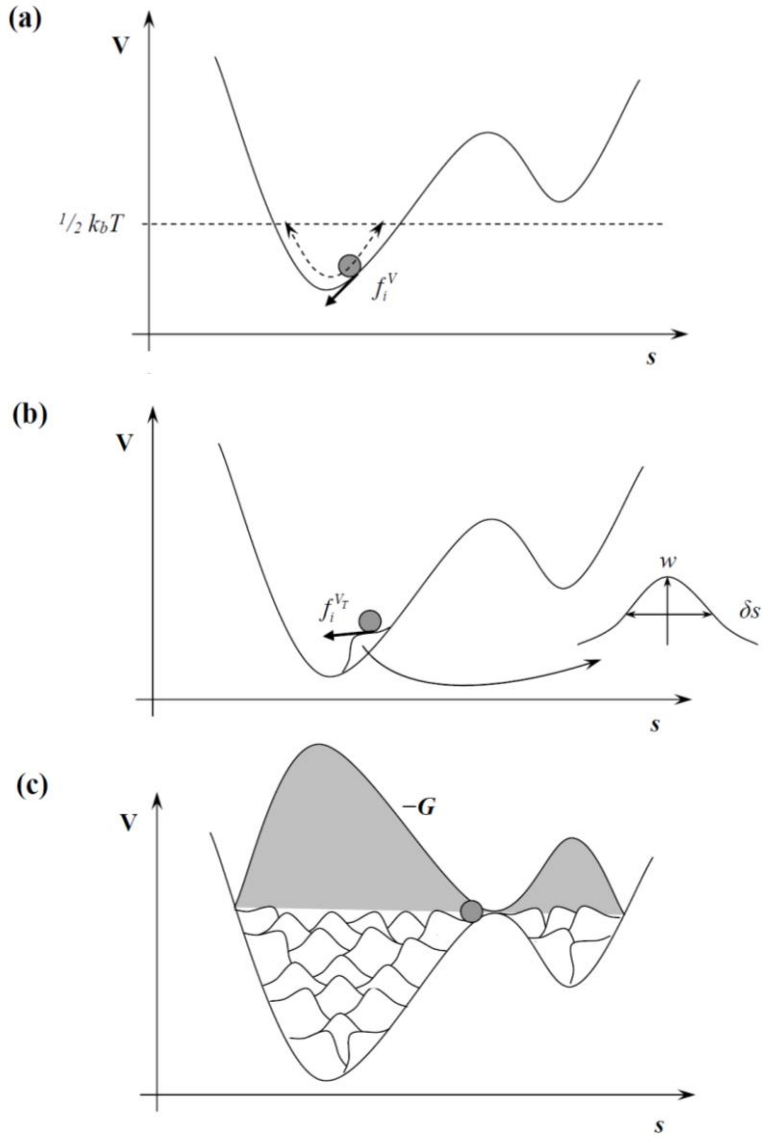

Figure II - 3. a) Standard molecular dynamics simulation. The system evolves following the interatomic potential V, and thus the force acting on the system is f_i^V. (b) Metadynamics simulation (i. e. a Gaussian potential term with height *w* and with *ds* is added every certain *t'* time centered at the value of s at that moment). c) End of the metadynamics simulation and reconstruction of the free energy landscape (Eq. II-40). Taken from reference (Biarnés, 2007).

7.2 Lagrangian metadynamics.

If the metadynamics approach is used in combination with QM/MM techniques (e. g. for simulating chemical reactions), the CVs are enforced to cross high barriers (~tenths of kcal·mol^{-1}) in short times (a few ps) by adding a history-dependent potential V_G, which injects a lot of energy into 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, the extended lagrangian metadynamics method has been introduced (Iannuzzi, 2003). This methodology introduce auxiliary variables *s'* that are coupled to the system by an harmonic potential of the form $V_H = ½k(s'-s(x))^2$ (where *k* is the spring

constant of the harmonic potential). These auxiliary CVs are treated as additional degrees of freedom with a fictitious mass *M* and an associated fictitious kinetic energy $K_{s'}$ given by:

$$K_{s'} = \frac{1}{2} M v_{s'}^2 \quad (\mathbf{Eq.\,II-41})$$

In this way, the dynamics of these extra degrees of freedom can be explicitly controlled using thermostats and the trajectory of *s'* will be sufficiently soft to maintain the stability of the dynamics (Ensing, 2005).

Because we are introducing an extra potential V_H, the total potential acting on the real system is modified:

$$V_T = V + V_H = V + \frac{1}{2} k[s' - s(x)]^2 \quad (\mathbf{Eq.\,II-42})$$

$$f_i^{V_T} = f_i^V + f_i^{V_H} = -\frac{\delta V}{\delta r_i} - \frac{\delta V_H}{\delta r_i} \quad (\mathbf{Eq.\,II-43})$$

Therefore, the dynamics of the real system in the real space (*s*) is driven by the sum of the interatomic potential *V* and the harmonic potential V_H. The mass for this fictitious particle and the force constant of the coupling potential need to be tested to ensure that the coupled particle follows naturally the value of the associated collective varialbe in the real system. The force constant has to be choosen large enough to keep *s'* close to the actual coordinates of the system *s(x)*. However, a large value of *k* requires a small time step for the integration of the equations of motion, which make the simulation more computationally demanding (Ensing, 2005). Within a Car-Parrinello scheme, another important requirement is the adiabatic separation between the auxiliary variables s' and the electronic degrees of freedom. Since the extra term in the Hamiltonian introduces frequencies of the order $\sqrt{k/M}$, *M* should be relative large, and the CV dynamics should be adiabatically decoupled from the atomic motions (Ensing, 2005). In practice however, large values of *M* makes the exploration of the free energy landscape (FEL) 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 the FEL exploration on the other must be reached.

In this framework, the hystory-dependent potential V_G, instead of acting on the real system as in direct metadynamics, acts on the space of the auxiliary collective variables *s'*. 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:

$$f_{s'}^{V_T} = f_{s'}^{V_H} + f_{s'}^{V_G} = -\frac{\delta V_H}{\delta s'_i} - \frac{\delta V_G}{\delta s'_i} \quad (\mathbf{Eq.\,II-44})$$

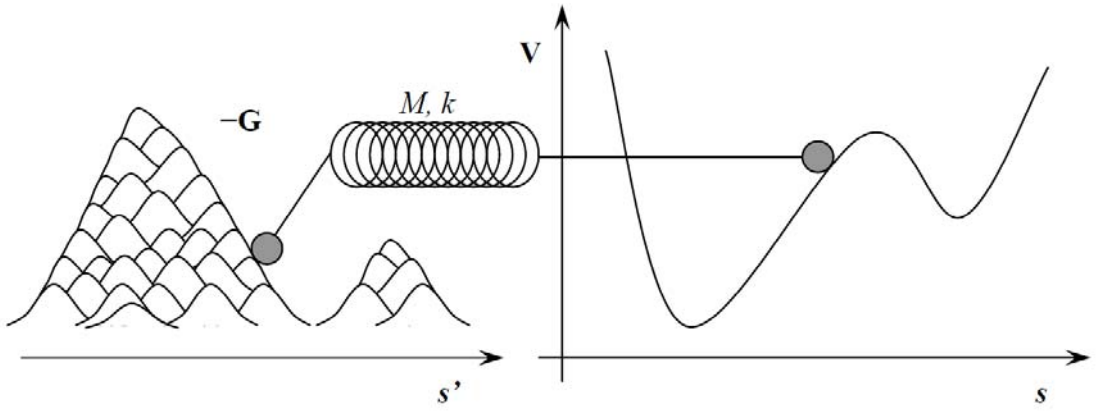

Figure II - 4. Completed simulation and reconstruction of the free energy landscape (FEL) for extended-Lagrangian metadynamics. Taken from reference (Biarnés, 2007).

Then the free energy as a function of the *s'* variables is given by:

$$G(s) = \lim_{t\to\infty} V_G(s',t) \approx \sum_{t'} w \cdot \exp\left[-\frac{[\mathrm{s}'(\mathrm{t}) - \mathrm{s}'(\mathrm{t}')]^2}{2(\delta \mathrm{s})^2}\right] \quad (\mathbf{Eq.\,II-45})$$

7.3 Well-tempered metadynamics.

In spite of its success, the metadynamics algorithm presents two major drawbacks. First of all, it is not trivial to decide when to stop a metadynamics run. In fact, in a single run V_G does not converge to a definite value of the free energy but oscillates around the correct result. This leads to an average error that is proportional to the square root of the deposition rate *t'* (Laio, 2005). Decreasing this rate (i. e. Gaussian functions are deposited with less frequency) leads to an increase of the time required to fill the FEL. Secondly, long metadynamics runs carry the risk that the system is irreversibly pushed to high energy regions of the CVs space, with no physical relevance.

Well-tempered metadynamics (Barducci, 2008) provides a framework to compute free energies that are converged to the exact result in the long time limit and control the regions of the FEL that are meaningful to explore. To achieve these purposes, the height of the Gaussian terms added during the simulation is modified according to the relationship:

$$w = \omega e^{-[V(s,t)/\Delta T]} t' \quad (\mathbf{Eq.\,II-46})$$

where ω is the initial height with units of energy per unit time, *V(s,t)* is the estimate of the free energy at the current CV position and time, and ΔT is a tunable temperature-like parameter that controls how quickly w is reduced as the wells are filled. In practice, ΔT limits the exploration of the FEL to an energy range of the order of $T+\Delta T$. In this scheme, the bias potential V_G does not fully compensate for the underlying free energy landscape; rather, we have:

$$G(s) + V_G(s,t) = G(s) - \frac{\Delta T}{T+\Delta T} G(s) = \frac{T}{T+\Delta T} G(s) \quad (\mathbf{Eq.\,II-47})$$

As a resume, well-tempered metadynamics overcomes the convergence problems of standard metadynamics approaches and allow the computational efforts to be focused on the physical relevant regions of the FEL.

7.4 Collective variables.

The CVs are functions of the microscopic coordinates of the system. To guarantee an effective application of the biasing potential, the CVs must fulfill the following requirements (Ensing, 2005)(Iannuzzi, 2003)(Laio & Gervasio, 2008):

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 states and describe all the relevant intermediates of the process under study.
iii. They should include all the relevant slow modes of the system, or at least the most relevant components of the (a priori unknown) reaction coordinate.
iv. They should be limited in number.

The CVs used in this Thesis, which include distances, coordination numbers, puckering coordinates, etc., are described in the respective chapters.

Chapter III – The conformational free energy landscape of cyclohexane.

Iglesias-Fernández, J.; Ardevol, A.; Raich, L.; Rovira, C.

The conformational free energy landscape of cyclohexane and β-D-xylopyranose using both Cartesian and Mercator representations.

Submitted.

The conformational free energy landscape of cyclohexane.

1. Introduction

Cyclic and heterocyclic compounds play fundamental roles in many biological processes. Carbohydrates, an example of highly flexible molecules built with different heterocyclic constituents, have a huge diversity of roles ranging from energy storage and transport, to cell recognition and signaling processes. The reaction mechanism for the synthesis or hydrolysis of the glycosidic bond, between two carbohydrate monomers, involves important changes in the ring conformation (Vocadlo, 2008). As observed in both NMR and X-Ray analyses (Espinosa, 1998)(Garcia-Herrero, 2002), the sugar ring located at the –I enzyme subsite distorts away from its $^4C_1$ conformation in solution towards a higher-energy one (e. g., boat or skew-boat) upon binding to GHs. Therefore, knowledge of ring conformations along the enzymatic reaction, the catalytic itinerary, is a key factor for the understanding of how GHs or GTs function (Davies, 2012).

A proper description of the conformations of cyclic compounds is a necessary prerequisite for the study of carbohydrate molecules and, in turn, their catalytic itineraries. In this chapter, we characterize the conformational energy landscape (FEL) of the simplest six-membered ring, cyclohexane, for which experimental information of relative energies among conformations is available. The calculations will be performed by means of *ab initio* metadynamics, using different sets of collective variables (CVs). One of the aims of these calculations is to demonstrate the validity of the methodology employed and the invariance of the relative free energies obtained with respect with the CVs used.

1.1 Cyclohexane.

The various conformations of cyclohexane have been the object of intensive study since the suggestion of the existence of different conformations (chair and skew-boat) made by Sachse in the 1890s (Sachse, 1890). Direct information on the skew-boat conformation was first obtained by Squiallacote *et al*., (Squiallacote, 1975) who obtained the matrix-isolated infrared spectrum relative to this conformation by condensing hot cyclohexane vapor. They trapped the skew-boat structure in an argon matrix at 20K and by slowly warming the matrix, they determined a value of $\Delta G^{\ddagger}$ for the isomerization to a chair conformation ($\Delta G^{\ddagger}$ = 5.27 ± 0.05 kcal·mol^{-1}). Additionally, a value of $\Delta G^{\ddagger}$=10.3 kcal·mol^{-1} was found for the inverse conversion, from chair to skew-boat conformations, by proton magnetic resonance (Jense, 1960)(Figure III-1b).

A substantial amount of experimental and computational efforts have been invested in the conformational analysis of cyclohexane, to establish the fundamental features of its energy landscape. Cyclohexane can exist in a variety of conformational states that have been depicted in Figure III-1a. The molecule presents two global energy minima, which correspond to both chair structures ($^4C_1$ and $^1C_4$, which are equivalent) and another local minimum, higher in energy, which is a skew-boat structure. As mention in the introduction (Chapter I, section 2), for a six-membered sugar ring there are six possible skew-boat conformations ($^3S_1$, $^5S_1$, ^2S_O, $^1S_3$, $^1S_5$, $^OS_2$), but due to the high symmetry of the cyclohexane molecule all of them are

indistinguishable. The transition state between the chair and the skew-boat corresponds to a half-chair conformation (Hendrickson, 1961). It has also been established that the boat structures are not stable conformers but they act as transition states between two skew-boat structures (Johnson, 1961)(Margrave, 1963). Skew-boat conformations lie about 6 kcal·mol^{-1} above the chair (Figure III- 1b)(Allinger, 1959)(Hendrickson, 1961)(Allinger, 1960)(Dixon, 1990) whereas half-chair and envelope conformations are 10.8 and 11.2 kcal·mol^{-1} higher in energy, respectively (Anet, 1966)(Hendrickson, 1967). The boat conformations have been reported to be 1.4 kcal·mol^{-1} higher in energy than the skew-boat structures (Kakhiani, 2009)(Ionescu, 2005).

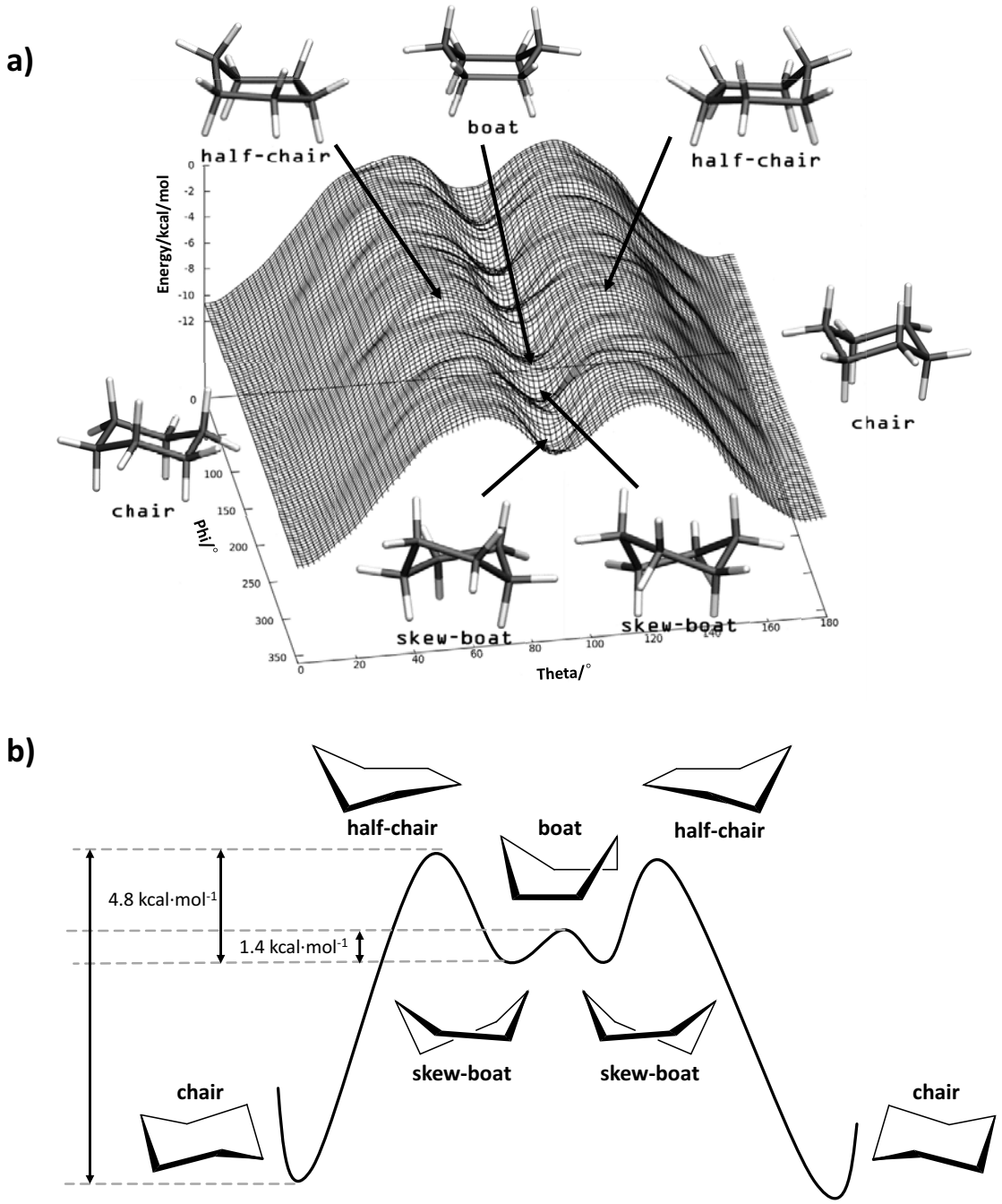

Figure III - 1. (a) Schematic representations of the energy surface of cyclohexane with the representative conformations (data obtained in this chapter). (b) Experimental energy profile (see text).

1.2 Cartesian and Mercator representations of the puckering sphere.

Cremer and Pople introduced a set of puckering coordinates to unequivocally describe the conformation of a ring of any size (Cremer & Pople, 1975). The conformations of a six-membered ring (such as cyclohexane or β-D-glucopyranose) are defined by three coordinates: a radius Q and two phase angles ϕ and θ (Figure III-2). The *Q* coordinate is the sum of the perpendicular distance of each ring atom (j) to the ring average plane ($Q = \sum_j^6 z_j$). The ϕ and θ coordinates are obtained by solving the following system of equations.

$$\begin{cases} Q\,\sin\theta\,\cos\phi = \sqrt{\frac{1}{3}}\sum_{j=1}^{6} z_j\,\cos\left[\frac{2\pi}{6}2\,(j-1)\right] \\ Q\,\sin\theta\,\sin\phi = \sqrt{\frac{1}{3}}\sum_{j=1}^{6} z_j\,\sin\left[\frac{2\pi}{6}2\,(j-1)\right] \\ Q\,\cos\theta = \sqrt{\frac{1}{6}}\sum_{j=1}^{6}(-1)^{j-1} z_j \end{cases} \qquad (\text{Eq. III} - 1 \text{ to } 3)$$

Since these are polar coordinates, any ring conformation would fall within the puckering sphere-like volume (Figure III-2). On the poles (θ = 0 or π) are located the two chair conformers ($^4C_1$ and $^1C_4$, respectively); on the equatorial region (θ = π/2) the 6 boat and 6 skew structures are sequentially placed in steps of θ = π/6.

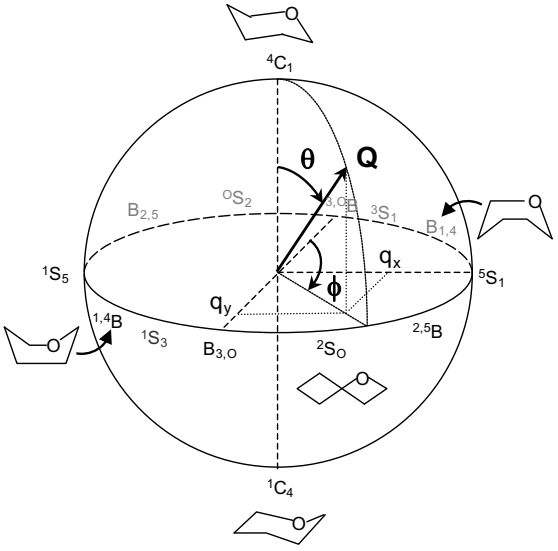

Figure III - 2. Cremer-Pople puckering coordinates of a six-membered ring (Q, θ, and ϕ) and their projection in the x,y plane (Cartesian coordinates: q_x and q_y).

Two representations have been historically used to map the Cremer and Pople 3D plot into a simpler two dimensional plot (Figure III-3). The Stoddart's or Cartesian diagram (Figure III-3a), corresponds to the projection of the polar coordinates onto the equatorial plane (q_x = Q·sin(θ)·sin(ϕ) and q_y = *Q*.sin(θ)·cos(ϕ); Figure III-2)(see Introduction; section 3.2.4). The projection can be taken either from the Northern or Southern hemispheres. Alternatively, the so-called plate carrée or Mercator representation is an equidistant cylindrical projection that results in a rectangular map with respect to θ and ϕ (Dowd, 1994).

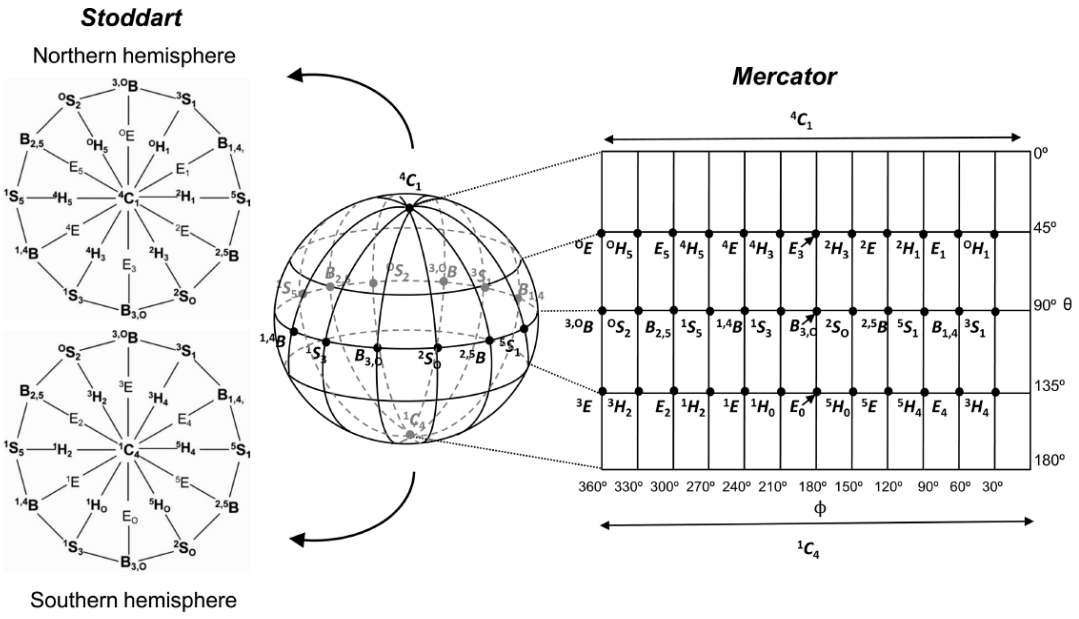

Figure III - 3. Main conformations on the Cremer-Pople sphere along with two of their most used 2D representations: Cartesian diagram (Northern and Southern projections are shown) and Mercator projection. Taken from reference (Davies, 2012).

Each type of representation has its pros and cons. Normally, catalytic GH itineraries are easy to interpret when they are drawn on a Cartesian representation, as there are no discontinuities between conformations. Moreover, since GHs normally follow straight itineraries either on the Northern or Southern hemispheres, but not on both (i. e. they do not follow transitions from $^4C_1$ to $^1C_4$ or vice versa), only one diagram (i. e. hemisphere) is sufficient to sample the experimentally relevant conformations, which further simplifies the problem. On another hand, the Mercator representation represents all conformations on a single diagram, thus mapping the complete Cremer-Pople sphere. This can be most convenient in a more general context, if one is interested in sampling all conformations of the sugar ring.

From a computational point of view, both representations (Cartesian and Mercator) can be implemented as CVs for free energy calculations (see Metadynamics section; Methods) but, do they give the same results? Sega *et al*. (Sega, 2009) reported, from a classical metadynamics simulation, that the Mercator representation is the only proper set of CVs to be used. They argued that the Cartesian representation had unphysical values of the radius Q for conformations close to the equatorial line and reported strong biases for the energy barriers between Northern and Southern hemispheres (errors of the order of 3 kcal·mol^{-1} were reported for energy barriers). In this chapter we demonstrate that, contrary to the conclusion of the above mentioned study, the conformational FEL obtained by *ab initio* metadynamics is invariant with respect to the type of CVs used in the calculations.

Additionally, Sega *et al*. (Sega, 2009) reported a unimodal behavior of the Q parameter for all conformations along the metadynamics simulation, with a narrow peak around 0.5 Å. This unimodal distribution contrasts with our previous results (Biarnés, 2007)(Ardèvol, 2010), which show a bimodal distribution with peaks around 0.54 Å and 0.72 Å for the chairs and skew-boat conformations, respectively. Here, we demonstrate that cyclohexane features a bimodal

distribution for the radius Q, as found in our calculations of the puckering conformations previously performed in our group (Ardèvol, 2010).

2. Objectives.

The main objectives of the present chapter are:

1) Determine how the FEL of a six membered-ring varies with respect to the chosen CVs. To this aim, we computed the conformational FEL of an isolated (gas-phase) cyclohexane with different sets of CVs (Cartesian and Mercator).
2) Analyze the variations in the amplitude (Q) values for different cyclohexane conformations and determine the reasons for a bimodal distribution of this parameter.
3) Determine which are the main consequences of choosing Cartesian or Mercator CVs for the determination of the FEL of a six-membered ring.

3. Computational details.

Gas-phase *ab initio* metadynamics simulations of cyclohexane, at room temperature, were performed within the Car-Parrinello approach (Car & Parrinello, 1985), as implemented in the CPMD 3.15.1 program (CPMD, 1990). The system analyzed consists of a single cyclohexane unit enclosed in an orthorhombic box of size 11 Å x 11 Å x 11.5 Å. The electronic structure was computed within the DFT, using the PBE generalized gradient-corrected approximation (PBE)(Perdew, 1996). This functional was found to give reliable results in previous Car-Parrinello simulations of isolated carbohydrates and carbohydrate-active enzymes (Biarnés, 2006)(Biarnés, 2007)(Ardèvol, 2010). In particular, the error on relative energies due to the density functional was found to be ± 0.6 kcal·mol^{-1} for β-D-glucopyranose (Biarnés, 2007). Kohn-Sham orbitals were expanded in a PW basis set with a kinetic energy cutoff of 70 Ry. *Ab initio* pseudopotentials were employed, generated within the Troullier-Martins scheme (Troullier, 1991). The fictitious mass for the electronic degrees of freedom was set to 1000 a. u. together with a time step of 0.12 fs.

The representation of the FEL with respect to the Mercator diagram ("Mercator FEL") was obtained using directly the ϕ and θ polar puckering coordinates as CVs in the metadynamics simulation (Direct metadynamics; see Chapter II, section 7.1). On another hand, the "Cartesian FEL" was obtained from the Cartesian coordinates q_x and q_y (Figure III-2) by extending the Car-Parrinello Lagrangian with extra terms describing the fictitious dynamics of the CVs (Lagrangian metadynamics; see Chapter II, section 7.2)(Iannuzzi, 2003). Both types of CVs were used for the simulation of all possible conformations of cyclohexane.

The height of the Gaussian terms (w) was set to 0.13 kcal·mol^{-1}, which ensures sufficient accuracy for the reconstruction of the FEL. The width (δs) was set to 0.1 (Å or rad, depending on the type of CV), according to the oscillations of the selected CVs observed in the free dynamics. The extended Car-Parrinello Lagrangian (Iannuzzi, 2003) was used to describe the dynamics of the CVs of the simulations with Cartesian coordinates, as done in our previous works (Biarnés, 2007)(Ardèvol, 2010). The mass of the fictitious particle and the force constant of the coupling potential were tested to ensure that the coupled particle follows the value of

the associated CV in the real system. Values of 5.0 a. m. u for the mass of the fictitious particle and 0.5 a. u. for the force constant were found to fulfill these conditions.

A new Gaussian potential was added every 200-400 MD steps at the first stage of the simulation (Cartesian and Mercator FEL, respectively), increasing it up to 1000 MD steps to ensure a proper convergence of the simulation. At the beginning of the simulation different deposition times were employed because for the direct metadynamics approach larger equilibration times, which allow the system to reach the configuration with the maximum probability, are needed. This is due to the absence of a fictitious particle coupled to the Mercator CVs. A total number of 3500/9000 Gaussian functions were added to completely explore the free energy landscape of cyclohexane for the Mercator and Cartesian FELs, respectively. Convergence of the metadynamics simulations was further assessed by checking the invariance of the FEL with the progression of the simulation.

4. Results and Discussion.

4.1 The conformational free energy landscape of cyclohexane.

The conformational free energy landscape of cyclohexane, reconstructed from the metadynamics simulation using θ, ϕ as CVs (Mercator FEL), is shown in Figure III- 4a. The surface is quite symmetric, with the most stable conformers being $^4C_1$ and $^1C_4$. The skew-boat conformers correspond to local minima, 6.5 kcal·mol^{-1} higher in energy than the chair conformers, whereas boat conformations lay 1 kcal·mol^{-1} higher in energy and correspond to transition states between skew-boats. Interconversion from a chair to any skew-boat crosses a half-chair conformer, with an energy barrier of 10.5 kcal·mol^{-1}. Half-chair and envelope conformations are energetically equivalent. As a summary of the results obtained for the Mercator FEL: $\Delta G^{\ddagger}_{chair \rightarrow skewboat}$ = 10.5 kcal·mol^{-1}; $\Delta G_{skewboat-chair}$ = 6.5 kcal·mol^{-1} and $\Delta G^{\ddagger}_{skewboat \rightarrow boat}$ = 1 kcal·mol^{-1}.

The next step is the calculation of the FEL using the Cartesian coordinates (q_x and q_y) as CVs (Figure III-4b). Remarkably, the FEL exhibits the same features resulting in identical energy differences as the calculation using Mercator coordinates (Figure III-3)($\Delta G^{\ddagger}_{chair \rightarrow skewboat}$ = 10.5 kcal·mol^{-1}; $\Delta G_{skewboat-chair}$ = 6.5 kcal·mol^{-1} and $\Delta G^{\ddagger}_{skewboat \rightarrow boat}$ = 1 kcal·mol^{-1}). Therefore, and contrary to Sega *et al*. (Sega, 2009), we conclude that metadynamics simulations performed with both CVs give identical results.

The relative energy differences among the various conformers, obtained in both metadynamics simulations, are in good agreement with previous experimental and theoretical estimates ($\Delta G^{\ddagger}_{chair \rightarrow skewboat}$ = 10.4-10.8 kcal·mol^{-1}; $\Delta G_{skewboat-chair}$ = 4.7-6.2 kcal·mol^{-1}). Therefore, we can conclude that both sets of CVs report identical free energies for the various conformations of cyclohexane, which are in excellent agreement with the available literature on the field.

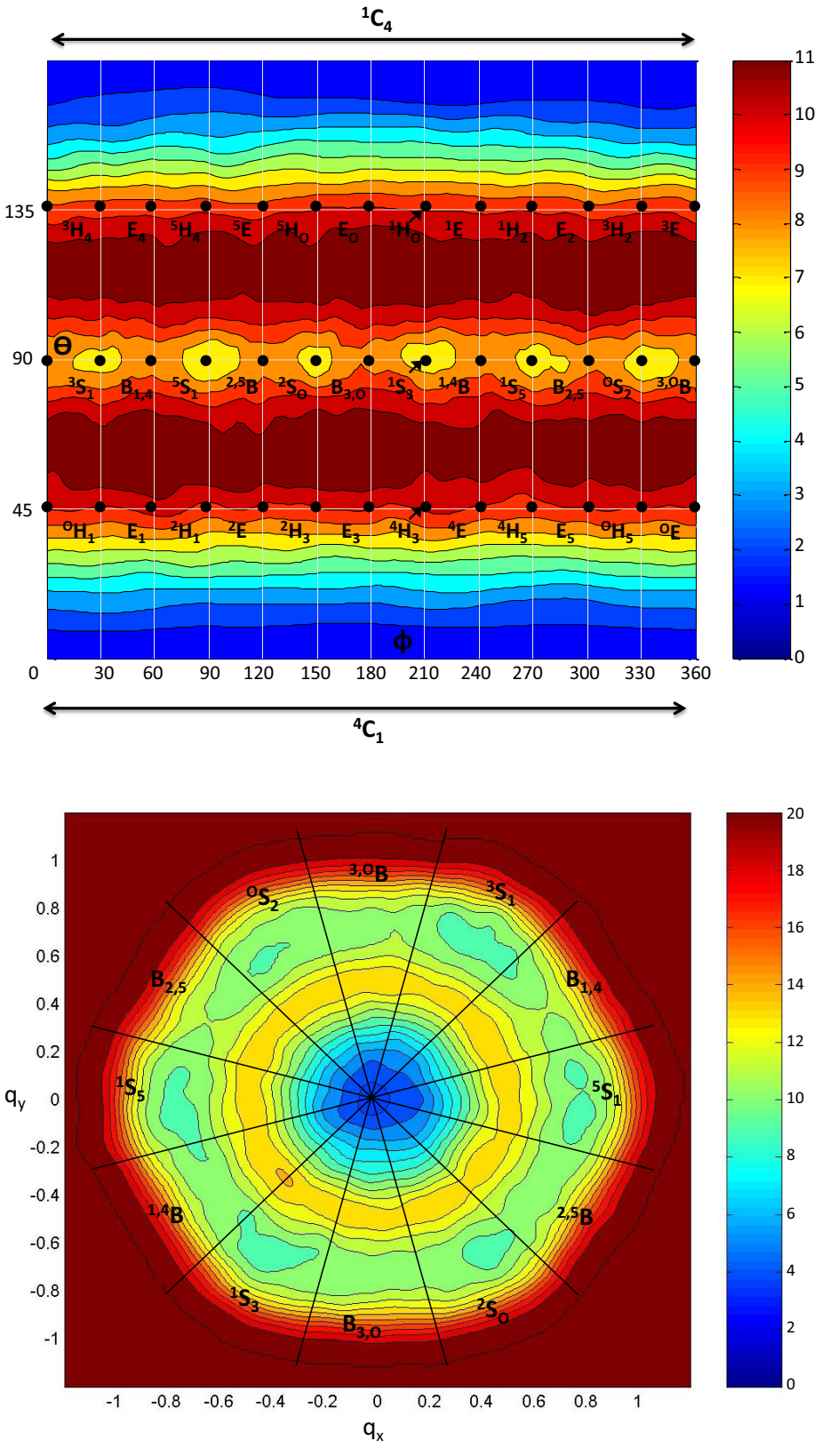

Figure III - 4. a) Computed free energy landscape of cyclohexane resulting from *ab initio* metadynamics using θ, ϕ as collective variables. b) Computed free energy landscape of cyclohexane resulting from *ab initio* metadynamics using q_x/q_y as collective variables. The continuous lines, separated by 30° in ϕ, indicate regions corresponding to different canonical conformations. Energy values are given in kcal·mol^{-1} and each contour line of the diagram corresponds to 1 kcal·mol^{-1}.

4.2 Analysis of the amplitude (Q) values for different cyclohexane conformations.

It is interesting to analyze not only the variation of the θ and ϕ coordinates, which determine ring conformation, but also the puckering amplitude *Q* (Figure II-2), which gives an idea of the degree of real "puckering" of the ring. Analysis of the Q values, obtained from structures along the metadynamics simulation, evidences that the distribution of Q values, *P(Q)*, features a bimodal behavior (Figure III-5a). The probability distribution shows two peaks around 0.54 Å and 0.72 Å which correspond to the chairs and boat/skew-boat conformations, respectively. Therefore, the puckering volume is not spherical but ellipsoidal, with the radii Q increasing slowly from the poles to the equator (Figure III-5b). These results are in contradiction with the ones obtained in previous metadynamics simulations using force-fields (Sega, 2009), which show a unique narrow peak around 0.5 Å for the probability distribution. This value is similar to the one that we obtained for chair conformations (0.54 Å) but very different to the value for the boat/skew-boat conformations (0.72 Å). As it will be shown in next paragraphs, forcing a value of Q ~ 0.5 Å for all conformers would lead to anomalous values of the angles between carbon atoms of cyclohexane (C-C-C angles), creating substantial ring strain on the boat and skew-boat conformations.

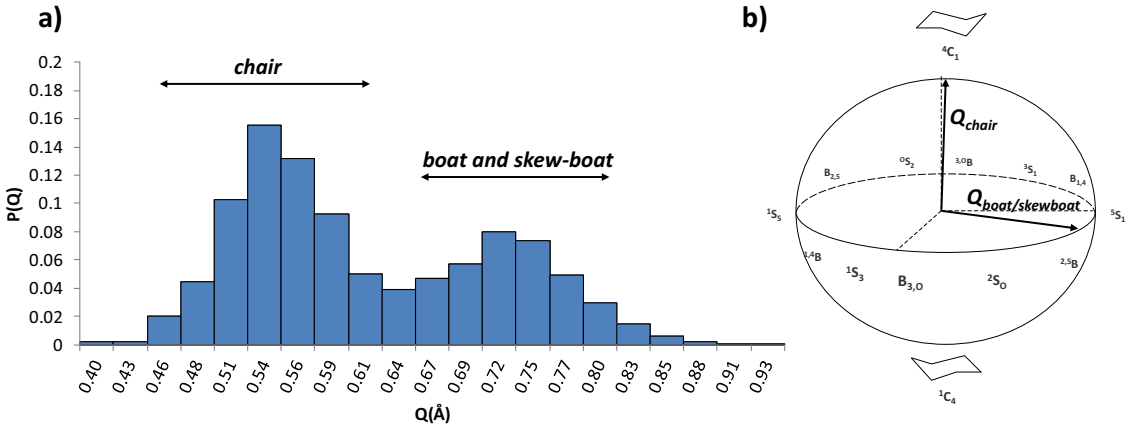

Figure III - 5. a) Probability distribution for the radial coordinate Q in cyclohexane. b) Ellipsoidal representation of the CV space.

The mathematical definition of the radius Q corresponds to the sum of the absolute perpendicular distance of each ring atom to the mean plane of the ring ($Q = \sum_j^6 z_j$). Therefore, conformations having small values of Q adopt more planar geometries (i. e. the distances between atoms and the mean plane are very short; Figure III-6b) whereas high Q values correspond to more "puckered" or compressed structures. Figure III-6a shows an example of how a given conformation of cyclohexane changes as a function of Q. For both conformations (chair and boat) low values of Q correspond to planar-like structures whereas higher values correspond to more compressed or "*puckered*" conformations.

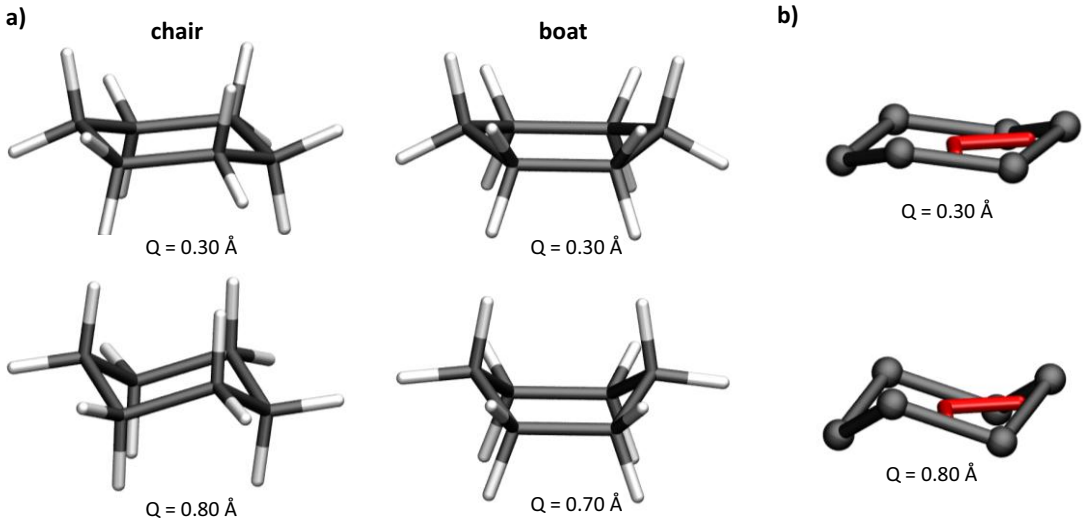

Figure III - 6. a) Chair and boat geometries at different values of the parameter Q. b) Chair conformations at different values of the parameter Q. In red it is shown the mean plane of the ring.

An easy way to understand differences in Q values, for a given conformation (e. g. chair, boat or skew-boat), is by looking at the internal bond angles of the ring. For a propane molecule the C-C-C bond angle is 111.7°. Therefore, it seems reasonable that values near 111.7° are also the most stable for cyclohexane and other rings with sp^3 carbon atoms. Table III-1 shows the values of the internal C-C-C angle for 3 conformations of cyclohexane at different Q values.

	Conformation		
Q/Å	Chair	Boat	Skew
0.3	116.55	118.45	118.46
0.4	114.72	117.32	117.33
0.5	112.74	115.94	115.96
0.6	110.66	114.37	114.37
0.7	108.51	112.69	112.63
0.8	106.32	----	110.78

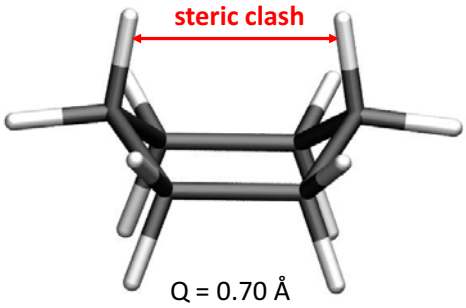

Table III - 1 and Figure III - 7. The table shows the values of the internal C-C-C bond angle for different conformations of cyclohexane at values of Q ranging from 0.3 to 0.8. Each value of the bond angle corresponds to the average of the six C-C-C angles present in cyclohexane. Angles are in degrees. Figure III-7 illustrates the close contact, between hydrogen atoms, which causes boat structures with Q values higher than 0.7 Å to be unstable.

Table III-1 shows that, for the chair conformation, the optimum value of the C-C-C angle (111.7°) occurs when Q is in the range of 0.5-0.6 Å. This value of Q corresponds exactly with the maximum of the probability observed in Figure III-5a. In contrast, for the skew-boat conformation the optimum C-C-C angle corresponds to a Q parameter between 0.7 and 0.8 Å, which also agrees with the results obtained in Figure III-5a. The same behavior is observed for the boat conformation, although in this case it was not possible to obtain a conformation with C-C-C angle lower than 112.69° due to steric clashes among ring hydrogen atoms (Figure III-7). This fact agrees with the observation (experimental and theoretical) of the boat conformations being transition states between skew-boat conformations.

Sega *et al*. (Sega, 2009) reported a distribution of Q values with a unique peak for glucuronic acid (Figure III-8). According to our calculations, this would result in anomalous values of the C-C-C angles creating substantial ring strain on the boat and skew-boat conformations.

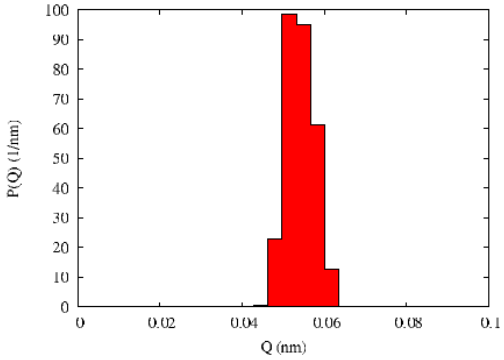

Figure III - 8. Probability density for the radial coordinate Q. Taken from reference (Sega, 2009).

4.3 Extent of Q values explored in metadynamics simulations using Cartesian or Mercator CVs.

The range of sampled conformations at *Q* values near the equatorial line is different in calculations using Cartesian (q_x/q_y) CVs with respect to those using Mercator (ϕ, θ) CVs. The reason for this relies on how forces are applied during the metadynamics calculation. For Cartesian calculations, forces are applied on the plane of the CVs (Figure III-9a) whereas for Mercator CVs forces are applied directly on the sphere (Figure III-9b). Therefore, when the bias potential accumulated during a Cartesian metadynamics simulation is too high, the system can explore conformations with Q values larger than those sampled in calculations with Mercator CVs (i .e. conformations outside the sphere, with higher values of Q)(Figure III-9a).

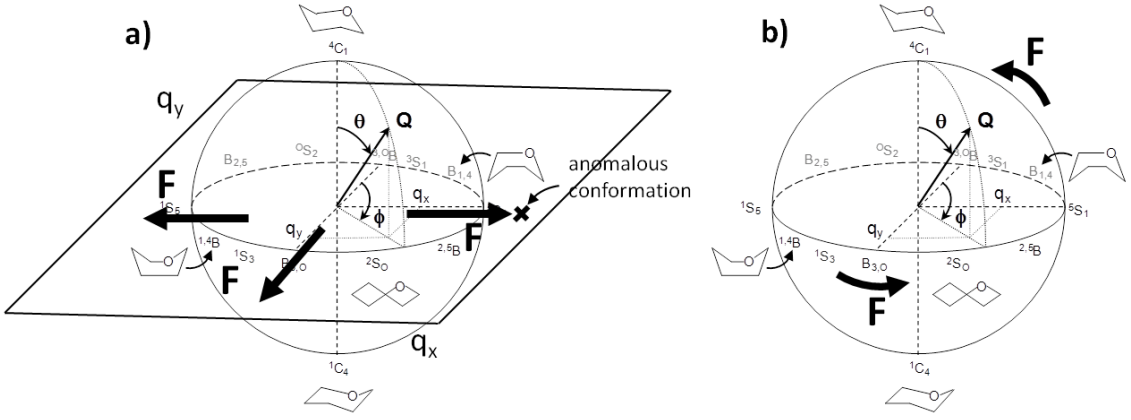

Figure III - 9. Direction of the forces applied during a metadynamics simulation: a) Cartesian CVs b) Mercator CVs.

Sega *et al*. (Sega, 2009) reported strong biases during classical metadynamics simulations using Cartesian CVs. This conclusion was based on the results obtained by plotting the puckering amplitude *Q* vs. q_r, where $q_r = \sqrt{q_x^2 + q_y^2}$ is the projected puckered vector (Figure III-10a). Because this plot shows a linear correlation for q_r values larger than 0.05 nm, the authors hypothesized that the Q values are driven by the metadynamics forces and explore unphysical conformations. Contrary to these results, both metadynamics simulations performed in this work present a dual behavior for the correlation between q_r and Q (Figure III-10b and c). Therefore, it is not true that this linear correlation is introduced by the Cartesian CVs and due to the exploration of unphysical values of Q. These results are in agreement with the ones presented in Figure III-5, in which the puckering amplitude presents a bimodal distribution.

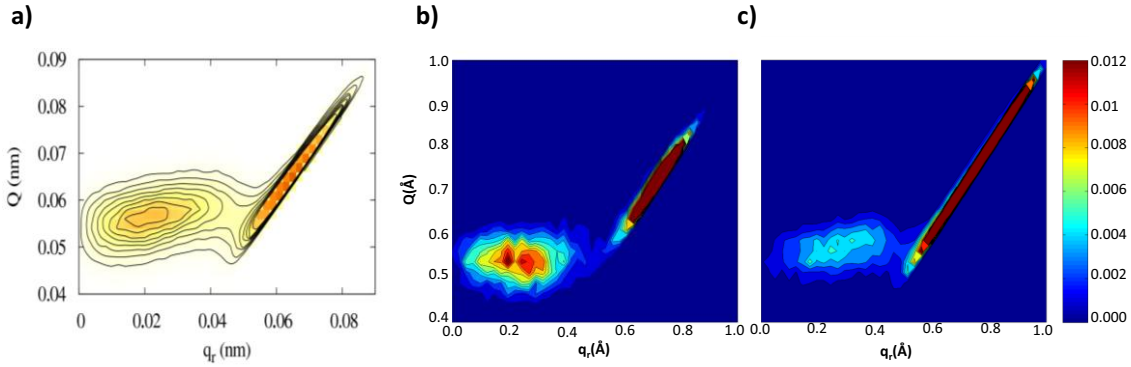

Figure III - 10. Distribution *P(q_r,Q)* of the projection $q_r = \sqrt{q_x^2 + q_y^2} = Q \cdot sen(\theta)$ vs the total puckering amplitude Q for different metadynamics simulations: a) Classical metadynamics simulation of Glucoronic acid, taken from reference (Sega, 2009); b) Mercator FEL of cyclohexane, this work; c) Cartesian FEL of cyclohexane, this work.

Analyzing Figure III-10 in more detail, we can observe that the Cartesian FEL presents a range of Q values for distorted conformations (Q = 0.5-1.0 Å; Figure III-10c) higher than the Mercator FEL (Q = 0.6-0.9 Å; Figure III-10b). This fact seems to agree with the argument, discussed at the beginning of this section, that the Cartesian CVs affect the range of Q values due to how forces are applied during the simulation. In order to check whether there is a relation between the range of Q values explored and the length of the simulation, we plotted the Cartesian *P(q_r,Q)* distribution at different simulation times (Figure III-11). This plot also contains the energy difference among all distorted conformations and both chair conformations. Therefore, it is indicative of the degree of convergence of the simulation.

The analysis of Figure III-11 allows us to conclude that the extent of Q values explored in a metadynamics with Cartesian CVs depends on the length of the simulation. We can clearly see that at the beginning of the simulation the range of Q values is shorter (Q = 0.5-0.9 Å at 250 ps) and slowly increases as the simulation proceeds (Q = 0.5-0.95 Å at 400 ps and Q = 0.5-1.0 Å at 825 ps). Therefore, it is important to consider when to stop the simulation to avoid introducing artifacts. For cyclohexane, the simulation is converged at 400 ps and, therefore, it should be stopped at this point to avoid an unbalance sampling of Q values in the puckering sphere.

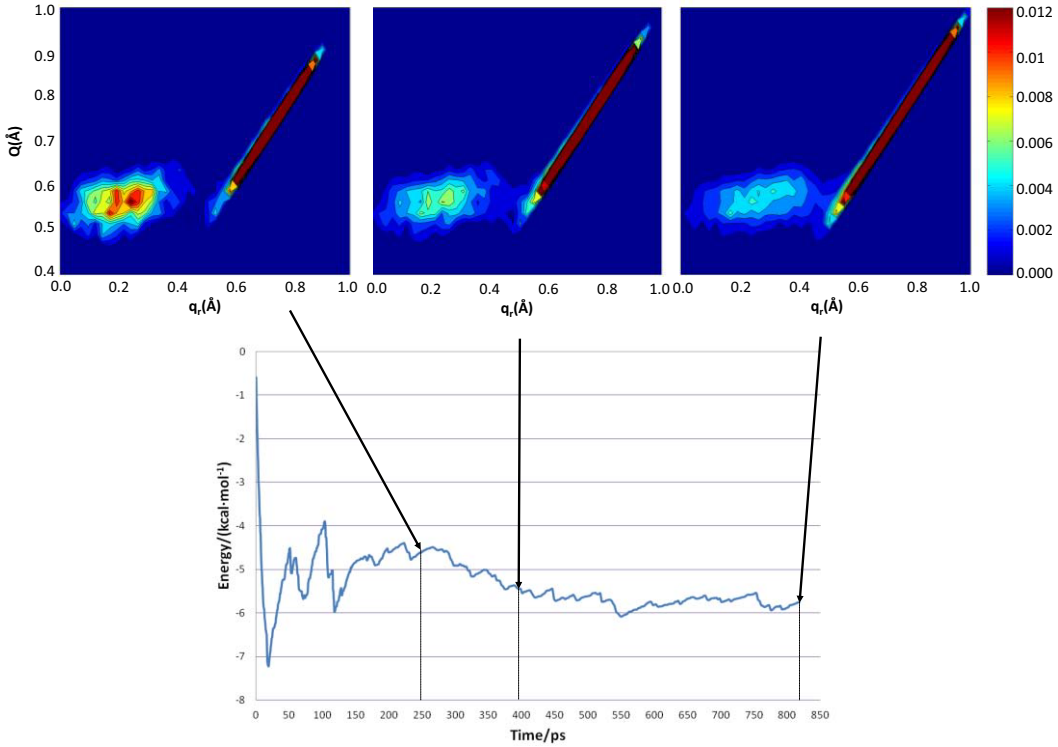

Figure III - 11. ***Distribution P(q$_r$,Q)*** **of the projection** $q_r = \sqrt{q_x^2 + q_y^2} = Q \cdot sen(\Theta)$ **vs the total puckering amplitude Q at different simulation times of the Cartesian FEL of cyclohexane.**

This is of special relevance for classical simulations, where time scales of hundreds of nanoseconds are easily accessible and, therefore, overconvergence of the simulation is often attained. As it was mentioned in Chapter II, long metadynamics runs carry the risk that the system is irreversibly pushed to high energy regions of the CVs space, with no physical relevance. For *ab initio* simulations, the time scales will rarely exceed one ns and, therefore, these effects will be very small. It is also important to remark that, even though higher values of Q were sampled in the calculations using Cartesian CVs, compared to Mercator CVs, the relative energies among conformations are invariant.

5. Conclusions.

The main conclusions of the present chapter are the following:

1. Both sets of CVs (Cartesian and Mercator) show identical results for the free energy calculation of all possible conformations of cyclohexane.

2. Cyclohexane shows a bimodal distribution of the amplitude (Q) due to the need to optimize the internal C-C-C angles. Chair conformations acquire an optimal value of the C-C-C angle (~111.7°) for Q values in the range of 0.5-0.6 Å, whereas skew-boat/boat conformations achieve the optimum angle for Q=0.7-0.8 Å.

3. Cartesian CVs introduce an unbalance in the range of Q values explored, during the metadynamics simulations, due to the procedure to apply forces. This effect, which does not affect energy differences, is dependent on the simulation length and, therefore, one must stop the simulation at the proper time. This is of special relevance for classical simulations where long simulations times are accessible with low computational costs.

Chapter IV – The conformational free energy landscape of α-D-mannopyranose and mannosidase inhibitors.

Thompson, A. J.; Dabin, J.; Iglesias-Fernández, J.; Ardevol, A.; Dinev, Z.; Williams, S. J.; Bande, O.; Siriwardena, A.; Moreland, C.; Hu, T. C.; Smith, D. K.; Gilbert, H. J.; Rovira, C.; Davies, G. J.

The Reaction Coordinate of a Bacterial GH47 α-Mannosidase: A Combined Quantum Mechanical and Structural Approach.

Angew. Chem. Int. Ed. **2012**, 51: 10997-11001.

Williams[*], R. J.; Iglesias-Fernández[*], J.; Stepper, J.; Jackson, A.; Thompson, A. J.; Lowe, E. C.; White, J. M.; Gilbert, H. J.; Rovira, C.; Davies, G. J.; Williams, S. J.

Combined Inhibitor Free-Energy Landscape and Structural Analysis Reports on the Mannosidase Conformational Coordinate.

Angew. Chem. Int. Ed. **2014**, 53: 1087-1091.

[*] **Equal contribution**

The conformational free energy landscape of α-D-mannopyranose and mannosidase inhibitors.

1. Introduction.

Previous theoretical studies in the group have demonstrated that substrate distortion approaches the substrate both structurally (elongation of the bond with the leaving group and shortening of the C1-O5 bond) and electronically (increase of the charge at the anomeric position) to the transition state (TS) of the hydrolysis reaction (Biarnés, 2006)(Soliman, 2009)(Biarnés, 2011). Therefore, distortion to form the Michaelis complex (MC) places the substrate "on the path" towards the TS of the reaction (Figure IV-1) and allow us to extract information about the catalytic itinerary followed by the enzyme. This is of paramount importance because knowing the exact conformation of the TS of the reaction allow us to rationalize the rate enhancement achieved by and enzyme and to develop TS inhibitors which have enormous potential as drug candidates (Caines, 2007)(Gloster, 2012).

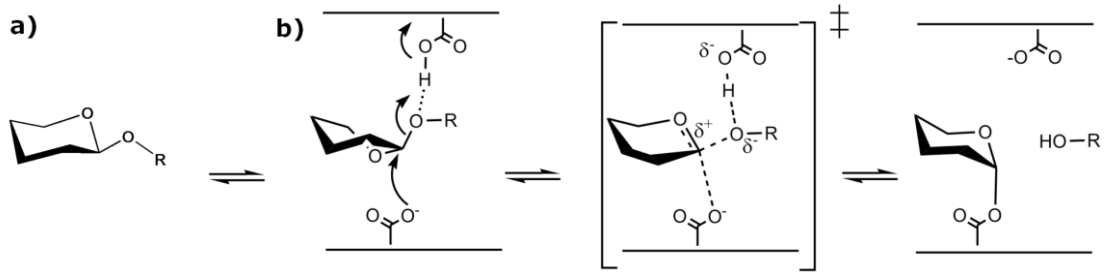

Figure IV - 1. a) Free glycoside in a chair conformation. b) Enzyme-Substrate complex (MC) with a distorted substrate conformation. First step (glycosylation) of the reaction mechanism catalyzed by retaining 1,3-1,4-β-glucanase.

How can these itineraries be predicted for a given substrate (e. g. glucose- or mannose-based polysaccharides)? Previous computational studies, performed in our group, for the isolated sugars β-D-glucopyranose (Biarnés, 2007), β-D-mannopyranose (Ardèvol, 2010), and α-L-fucopyranose (Lammerts van Bueren, 2010) have revealed that only a small number of possible conformations are available, as reflected in lower energy values for these conformations in the free energy landscape (FEL). Additionally, it was shown that the FEL of isolated monosaccharides correlates well with the conformations observed in ligand-enzyme complexes being, therefore, a useful tool for the prediction of the MC of the reaction.

The fact that the FEL of a given sugar "*in vacuo*" informs on the behavior of this sugar inside the enzyme gave us another idea: Can the energy landscape of a GH inhibitor gave us information about the most stable conformations inside the enzyme? If yes, can we extract information about transition state mimicry (i. e. why some inhibitors resemble the TS conformation more than others)?

In this Chapter, we apply the methodology that was previously developed in the group (Biarnés, 2007)(Ardèvol, 2010)(Lammerts van Bueren, 2010), to obtain the FEL of α-D-mannopyranose. Together with the analysis of electronic and structural changes upon ring

distortion, we determine the most suitable conformations for the MC of α-mannosidase enzymes. We also compute the FEL of α-D-mannopyranose bound to a specific enzyme (GH47 α-1,2-mannosidase) to determine the effect of the protein environment on substrate flexibility. Finally, we compute the FEL for two mannosidase inhibitors to report features about inhibition and its TS mimicry capabilities.

This work was done in collaboration with the experimental groups of Prof. Gideon Davies from the University of York and Spencer J. Williams from the University of Melbourne. They have synthesized and characterized several inhibitors for α- and β-mannosidases which provide essential information about the structure and catalytic itineraries of these enzymes (Thompson, 2012)(Williams, 2014).

1.1 α-Mannosidases.

Mannosidases are GHs that catalyze the cleavage of the glycosidic bond present in mannose-containing glycoconjugates and polysaccharides. The reaction coordinates followed by mannosidases are of fundamental mechanistic interest and are increasingly relevant to develop new chemical synthetic strategies of mannosides (Crich, 2010). Furthermore, mannosidase catalysis is crucial in several biochemical events, in both the healthy cell and in the context of disease. For example, α-mannosidases are important in N-glycan biosynthesis and protein quality control and their inhibition may allow intervention in diseases related with the degradation of folding-defective glycoproteins (Alzheimer's, Parkinson's or Creutzfeldt–Jakob diseases) (Lederkremer, 2009)(Dobson, 2004).

Biochemical studies for different sequence-base mannosidases have highlighted that a variety of conformational itineraries are employed by these enzymes for the cleavage of glycosidic bonds (Davies, 2012)(Figure IV-2). A $^{1}S_{5}\leftrightarrow[B_{2,5}]^{\ddagger}\leftrightarrow{}^{O}S_{2}$ conformational itinerary[1] have been identified for mannosidases of families GH2 (Tailford, 2008), 38 (Numao, 2003), and 92 (Zhu, 2010) and a $^{3}S_{1}\rightarrow[^{3}H_{4}]^{\ddagger}\rightarrow{}^{1}C_{4}$ itinerary for the α-mannosidases of family GH47 (Thompson, 2012).

One particularly important group of family GH47 enzymes are the so-called mannosidase I Golgi and endoplasmic reticulum (ER) α-mannosidases (Cantarel, 2009). These enzymes are involved in the biosynthetic remodeling of mannose containing N-glycans and the degradation of misfolded proteins as part of the protein folding quality control apparatus (Molinari, 2007) (Moremen, 2006). The first insight of an $(\alpha/\alpha)_7$ fold and the presence of an essential Ca^{2+} ion, for GH47 enzymes, was provided thanks to the determination of the structure of the yeast *Saccharomyces cerevisiae* class I α-1,2-mannosidase (Vallée, 2000). Subsequently, a ring-flipped $^{1}C_{4}$ conformation was observed for the inhibitors deoxymannojirimycin and kifunensine bound at the active center of the human class I enzyme. Additionally, the MC was observed in a $^{3}S_{1}$ conformation thanks to an S-linked disaccharide substrate mimic (Karaveg, 2005). All together, these data support a $^{3}H_{4}$ or close conformation as the TS for GH47 enzymes (Figure IV-2).

[1] The double arrow indicates that the conformational itinerary is bidirectional.

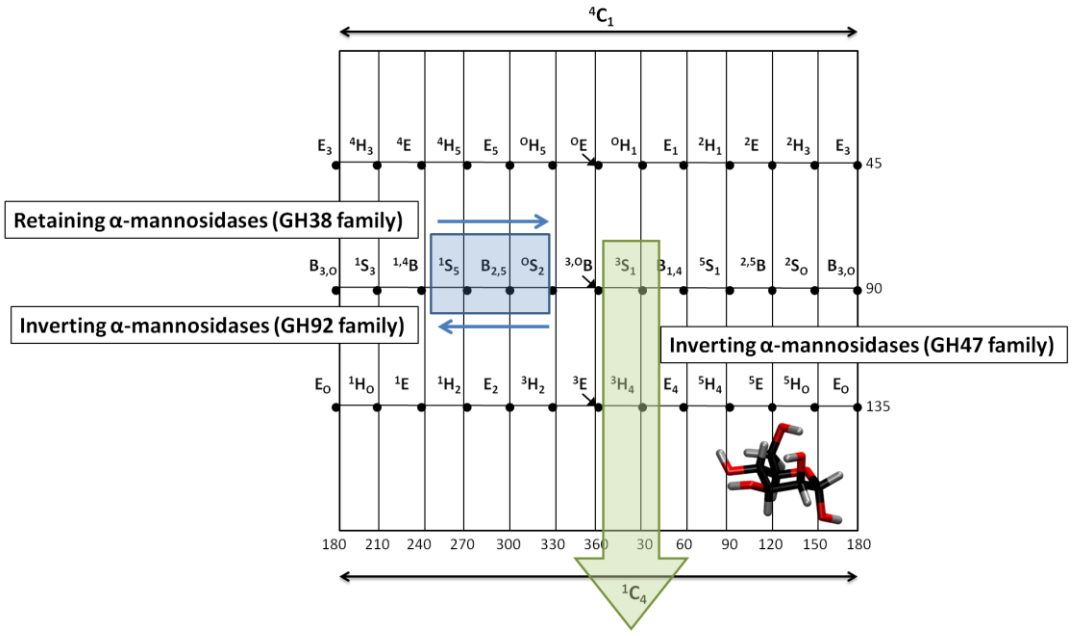

Figure IV - 2. Proposed reaction conformational itineraries for retaining and inverting α-mannosidases.

1.2 Mannosidase inhibitor complexes.

GH inhibitors are not only invaluable as therapeutic compounds (e. g. Relenza and Tamiflu) but also as tools for elucidating catalytic itineraries. NMR or X-ray characterization of an inhibitor bound to an enzyme (inhibitor-enzyme complex) in a particular conformation, such as $^{3}S_{1}$ or $^{1}C_{4}$, provides information about the catalytic itinerary followed by the enzyme.

The reaction coordinate of α-mannosidases could be studied through complexes 1-7 (Table 1; Figure IV-3) which define the different conformational reaction coordinates of these enzymes (Figure IV-2). For the GH47 family, the non-hydrolyzable disaccharide (1) binds to the -I subsite in a distorted $^{3,O}B/^{3}S_{1}$ conformation, mimicking the MC of the reaction (Thompson, 2012). Available complexes with mannoimidazole (2) and noeuromycin (3) define the remaining part of the catalytic cycle (Thompson, 2012). Mannoimidazole (2) is a putative TS mimic and bind to the enzyme in a $^{3}E/^{3}H_{4}$ conformation, which provides evidence for the TS conformation of the reaction. In solution, Noeuromycin (3) exists in α- and β-configured hemiaminal forms (Liu, 2001). For the inverting enzyme α-1,2-mannosidase CkGH47, noeuromycin binds as the less favored D-manno-configured hemiacetal form in a ring-flipped $^{1}C_{4}$ conformation, which mimics the products of the reaction. All these complexes define a clear $^{3,O}B/^{3}S_{1} \rightarrow [^{3}E/^{3}H_{4}]^{\ddagger} \rightarrow {}^{1}C_{4}$ itinerary.

Kifunensine (5)(Shah, 2003) and noeuromycin (7)(Kuntz, 2006) were used as inhibitors for the retaining GH38 α-mannosidase II. Both compounds exhibit a distorted conformation around $^{1,4}B/^{1}S_{5}$ for the MC/TS of the reaction. Additionally, the presence of a $^{1}S_{5}$ conformation for the covalent intermediate, obtained with a 5-fluoro derivative (6)(Numao, 2003), defines a $^{1}S_{5} \rightarrow [B_{2,5}]^{\ddagger} \rightarrow {}^{O}S_{2}$ catalytic itinerary for this family of enzymes. This itinerary has also been assessed by QM/MM methods (Petersen, 2010). For the inverting GH92 enzyme Bt3990, mannoimidazole (4) was used to assess the TS of the reaction (Zhu, 2010). The observation of a $^{1}S_{5}/B_{2,5}$ conformation defines a catalytic itinerary around $^{O}S_{2} \rightarrow [B_{2,5}]^{\ddagger} \rightarrow {}^{1}S_{5}$ pathway, which corresponds to the reverse pathway observed for the retaining GH38.

Struct.	PDB code(Reference)	Enzyme/family	Resol. (Å)	Subst. conform.	Type struct.
1	4AYP (Thompson, 2012)	α-1,2-Mann/GH47	0.85	$^{3,O}B/^{3}S_1$	MC
2	4AYQ (Thompson, 2012)	α-1,2-Mann/GH47	1.10	$^{3}E/^{3}H_4$	TS analogue
3	4AYR (Thompson, 2012)	α-1,2-Mann/GH47	1.10	$^{1}C_4$	Product comp.
4	2WZS (Zhu, 2010)	α-1,2-Mann/GH92	2.25	$^{1}S_5/B_{2,5}$	TS analogue
5	1PS3 (Shah, 2003)	α-Mann/GH38	1.80	$^{1,4}B$	TS analogue
6	1QWN (Numao, 2003)	α-Mann/GH38	1.20	$^{1}S_5$	Covalent Int.
7	2ALW (Kuntz, 2006)	α-Mann/GH38	1.86	$^{1,4}B/^{1}S_5$	TS analogue

Table IV - 1. Available inhibitor-enzyme complexes of retaining and inverting α-mannosidases and the conformations adopted by the saccharide at the –I subsite.

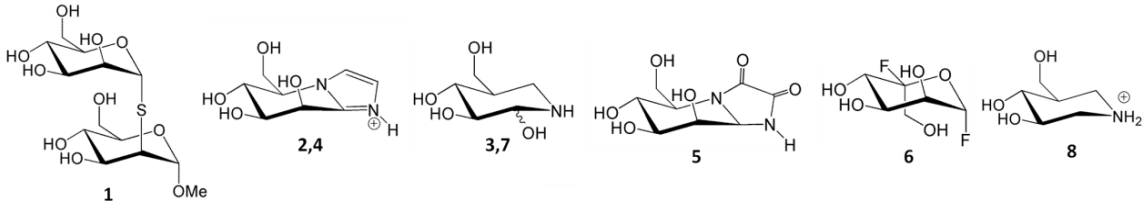

Figure IV - 3. 1) S-linked disaccharide (methyl-2-S-(alpha-D-mannopyranosyl)-2-thio-alpha-D-mannopyranoside). 2,4) Mannoimidazole. 3,7) Noeuromycin. 5) Kifunensine. 6) 5-fluoro-β-L-gulosyl fluoride (5FGulF).

In this chapter we studied the FEL of two α-mannosidase inhibitors, isofagomine and mannoimidazole (see Introduction, section 5). Based on structural analyses, it is known that isofagomine type inhibitors (Figure IV-3(8)), upon protonation inside the enzymatic catalytic cavity, resemble the glycosyl cation with charge localized at C1, but are poor mimics of the TS conformation. Isofagomine binds in a $^{4}C_1$ conformation in most X-Ray structures of GH complexes and, therefore, do not report information about the TS of the reaction. On another hand, mannoimidazole inhibitors (Figure IV-3(2-4)) are qualitatively good models of a mannopyranosyl oxocarbenium ion-like TS, with sp^2 hybridization and the potential for anti protonation (Heightman, 1999) of the imidazole functionality. Kinetic and thermodynamic analyses suggest that this compound is a true TS mimic (Tailford, 2008)(Thomson, 2012). Therefore, complexes with mannoimidazole are expected to report on the two catalytic itineraries followed by α-mannosidases (passing through either $B_{2,5}$ or $^{3}H_4$)(Tailford, 2008)(Thomson, 2012). However, these studies do not report which features of the inhibitor imbue TS mimicry capabilities.

2. Objectives.

The main objectives of the present chapter are:

1) Compute the conformational FEL of an isolated (gas-phase) α-D-mannose to gain information about the conformational itineraries followed by α-mannosidases.
2) Determine the effect of the protein environment on the accessible conformations of a sugar molecule by computing the FEL for an α-D-mannose inside the catalytic cavity of the GH47 α-1,2-mannosidase.
3) Analyze the FELs of two α- and β-mannosidase inhibitors (mannoimidazole and isofagomine) to get insight into their TS mimicking properties.

3. Computational details.

3.1 Gas-phase metadynamics simulations of α-D-mannose and mannosidase inhibitors.

Ab initio simulations were done within the Car-Parrinello formalism (Car & Parrinello, 1985). A fictitious mass of 850 a. u. 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 (Nosé, 1984). The Kohn-Sham orbitals were expanded in a PW basis set with a kinetic energy cutoff of 70 Ry. *Ab initio* pseudopotentials generated within the Troullier-Martins scheme were employed (Troullier, 1991). The calculations were performed using the PBE generalized gradient-corrected approximation (Perdew, 1996), as previously used in Car-Parrinello simulations of isolated carbohydrates (Ardèvol, 2010)(Biarnés, 2007) and GHs (Biarnés, 2006)(Biarnés, 2011).

The α-D-mannopyranose consists in 24 atoms enclosed in an orthorhombic box with dimensions 12.0 Å × 12.3 Å × 13.2 Å. The inhibitors isofagomine (22 atoms) and mannoimidazole (27 atoms) were enclosed in boxes of dimensions 13.0 Å × 15.0 Å × 13.0 Å and 14.0 Å × 14.0 Å × 13.0 Å , respectively. For all systems, 10 ps of free MD at 300 K were performed prior to the metadynamics simulations.

For all gas-phase metadynamics simulations, ϕ and θ puckering coordinates were used as CVs (Mercator FEL). Gaussian-like functions with width (δs) of 0.10 rad (selected according to the oscillations of the collective variables in a free dynamics) and a height (*w*) of 0.18 kcal·mol^{-1} were used with a deposition time of 48 fs. Once the full conformational space was explored, *w* and the deposition time were set to 0.13 kcal·mol^{-1} and 120 fs, respectively, for a better convergence of the FEL. A total of 6980, 5050 and 2500 Gaussians were added for α-mannose, isofagomine and mannoimidazole, which corresponds to ≈ 600 ps, 300 ps and 180 ps of total simulation time, respectively.

3.2 Force field based MD simulations for the α-D-mannose Enzyme-Substrate Complex.

A classical MD simulation was performed to equilibrate the structure of the MC with the natural substrate (α-1,2-mannobiose). The initial structure was taken from the crystallographically-determined enzyme-substrate complex (MC; pdb code: 4AYP). This structure was selected as starting point because it contains a S-linked substrate mimic, which is easily replaced by an oxygen atom *in silico*, and its high resolution (1.43 Å). The protonation states and hydrogen atom positions of all histidine residues were selected based on their hydrogen bond network. All Asp and Glu residues were taken as deprotonated (i. e. negative charge) except Glu121 (the general acid residue). Three sodium ions were added to achieve neutrality of the protein structure. All the crystallographic water molecules were retained and extra water molecules were added to form a 10 Å water box around the protein surface. The system was initially subjected to classical MD using Amber10 software (Pearlman, 1995). The protein was modeled with the FF99SB force field (Hornak, 2003) and the GLYCAM06 parameter set (Kirschner, 2008) was used for the carbohydrate molecules. Finally, all water molecules were described with the TIP3P force field (Jorgensen, 1983).

The system was equilibrated in several steps. First, all water molecules were relaxed with a gradient minimizer, holding the protein and substrate fixed. Next, the whole system was allowed to relax. To gradually reach the desired temperature of 300 K in the MD simulation, spatial constraints were initially added to the protein and substrate, while water molecules and sodium ions were allow to move freely. The constraints were then removed and the whole system was allowed to reach the desired temperature. The simulation was continued for 50 ps at constant pressure, allowing the cell volume to evolve. The MD simulation was extended to 2 ns until the system had reached equilibrium. Because the force field was unable to retain the substrate distortion, present in the initial X-ray structure, small restraints were used for the atoms of the –I sugar ring during the classical simulation.

A separate simulation was performed considering Asp249 (Figure IV-10) as the protonated acid residue (Cantu, 2008). However, the structure obtained after the equilibration phase did not maintained important active-site interactions (i. e. calcium mannose interactions) and therefore this scenario was not considered further. Instead, a perfect match was obtained with a protonated Glu121 and a unprotonated Asp249 (the r.m.s.d of the backbone atoms was ≈ 0.95 Å). One snap-shot from the end of the simulation was taken as starting structure for the subsequent enzyme-substrate CPMD/MM simulations.

3.3 Metadynamics simulations for the α-D-mannose Enzyme-Substrate Complex.

A metadynamics simulation was performed to explore the conformational FEL of α-D-mannose in an enzyme-substrate complex (MC). Initially, the *w* term was set at 0.38 kcal·mol^{-1} and, once the FEL was fully explored, it was reduced to 0.19 kcal·mol^{-1} to ensure sufficient accuracy for the reconstruction of the FEL. The δs term was set to 0.15 rad. At the beginning of the metadynamics simulation, a new Gaussian-like potential was added every 24 fs, which was increased up to 48 fs towards the end of the simulation. The metadynamics calculation was stopped after having added 140 Gaussians, which in terms of simulation time this corresponds to ≈ 4 ps. To ensure that the enzyme restricts the FEL to only two main substrate conformations, a long simulation was performed (keeping the initial parameters) for 16 ps (660 Gaussians added). The results further confirmed that there are only two minima in the FEL.

The CPMD/MM calculations of the MC used the following partition: the QM region was composed of an α-1,2-mannobiose and the Ca^{2+} ion coordinated to it (46 atoms in total). The rest of the protein and the solvent were included in the MM region and were treated with the AMBER force filed. All QM atoms were enclosed in a isolated orthorhombic box of size 13.0 Å x 14.0 Å x 18.0 Å. Due to the presence of the Ca^{2+} ion the Kohn-Sham orbitals were expanded to an energy cutoff of 80 Ry. *Ab initio* pseudopotentials generated within the Troullier-Martins scheme (Troullier, 1991) were employed. The calculations were performed using the PBE generalized gradient-corrected approximation (Perdew, 1996). The fictitious mass for the electronic degrees of freedom of the Car-Parrinello Lagrangian was set to 650 a. u. and the simulation time step to 0.12 fs. The system was equilibrated at 300 K for ≈ 10 ps before starting the metadynamics run.

4. Results and discussion

4.1 α-D-mannopyranose free energy landscape.

The computed FEL of α-D-mannopyranose as a function of the CVs θ and ϕ is shown in Figure IV-4. It contains six minima, which are labelled as **M1** to **M6** according to their relative stabilities (Figure IV-4b). As expected, the most stable conformer is the undistorted $^4C_1$ chair conformation (**M1**). The remaining local minima correspond to the $^1C_4$ inverted chair (**M2**) and distorted conformations located on the central part of the Mercator diagram (Θ=90°)(**M3-M6**). Local minima (**M2 –M6**) are within 5 to 6 kcal·mol^{-1} higher in energy than the most stable conformation $^4C_1$. Interconversion from **M1** to any of the other local minima involves energetic barriers ≈ 7 kcal·mol^{-1}, whereas the barriers associated with the transition between distorted conformations can be as low as 1-3 kcal·mol^{-1}.

As previously observed for different isolated sugars (Biarnés, 2007)(Ardèvol, 2010), not all the stationary points of the mannose FEL have a direct correspondence to the ideal conformations represented in the Mercator diagram. For example, **M3** corresponds to a conformation in between $B_{1,4}$ and $^5S_1$, while **M4** includes both $B_{2,5}$ and $^OS_2$ conformations. Moreover, there are several canonical conformers (e. g., $^{2,5}B$, ^2S_O, $B_{3,O}$) with no direct correspondence to a computed local minima. As a consequence, there are fewer local minima (six) than conformations in the diagram.

To check whether there is a relation between the shape of the computed FEL and the reaction pathway followed by α-mannosidases, we analyzed representative complexes from available X-ray structures along the reaction pathway (either an α-D-mannose derivative, TS-analogue inhibitors, covalent intermediate or product complexes; Table IV-1; Figure IV-3). The θ and ϕ values for the saccharide at -I subsite were computed and the corresponding values were located on the FEL (Figure VI-5).

Interestingly, all the experimental X-ray structures fall around energetically accessible regions of the FEL. Additionally, both catalytic itineraries of retaining and inverting α-mannosidases pass through allow regions of the FEL in the process of going from reactants to products (Figure IV-5). This suggest, as previously found for β-D-glucopyranose (Biarnés, 2007), β-mannopyranose (Ardèvol, 2010) and α-L-fucopyranose (Lammerts van Bueren, 2010) that α-mannosidases have evolved to preferentially select those conformations of the mannosyl substrate that require less energy to be distorted for catalysis to occur.

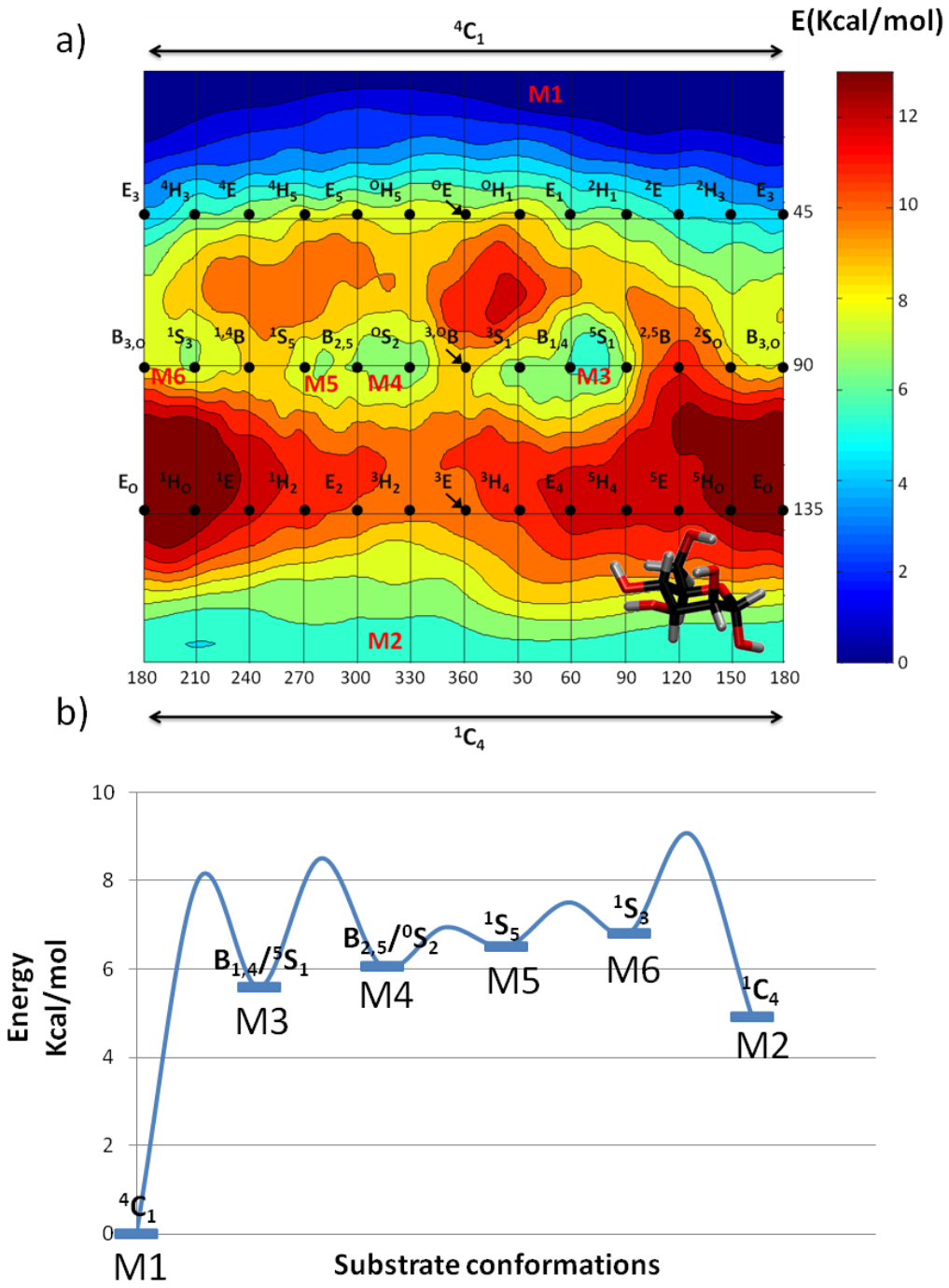

Figure IV - 4. a) Conformational FEL of an isolated α-D-mannose. Energy values are in kcal·mol^{-1} and each contour line corresponds to 1 kcal·mol^{-1}. b) Free energy of the different minima relative to the $^4C_1$ conformation.

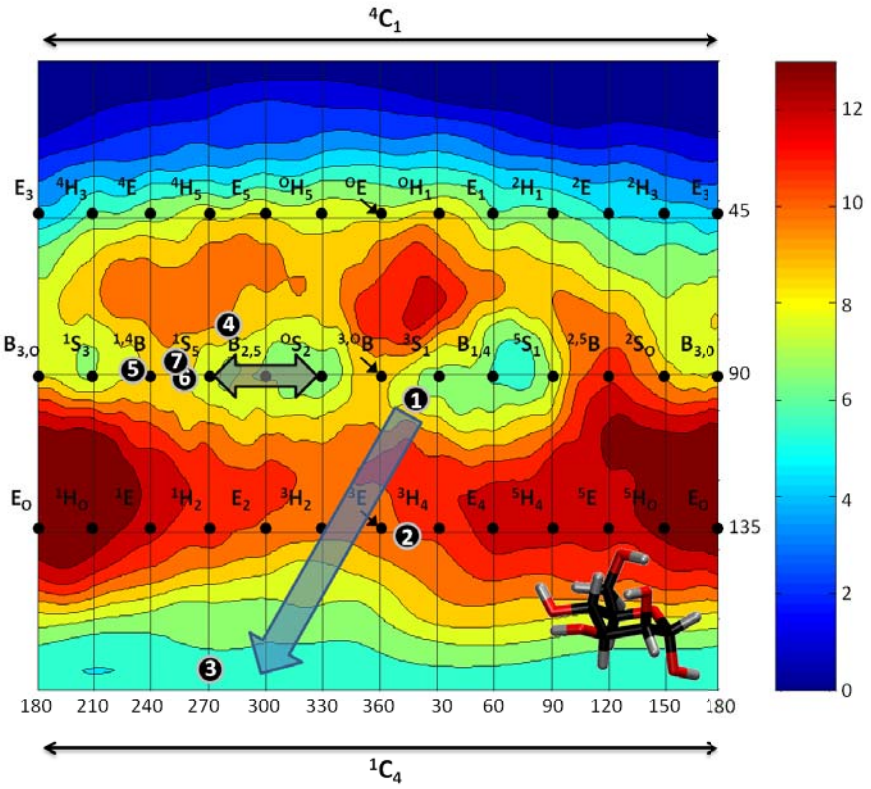

Figure IV - 5. Conformational FEL of α-D-mannopyranose with respect to ring distortion. Energy values are in kcal·mol^{-1} and each contour line corresponds to 1 kcal·mol^{-1}. The two possible reaction itineraries are marked with arrows (Grey for both catalytic itineraries $^{1}S_{5}\leftrightarrow[B_{2,5}]^{\ddagger}\leftrightarrow{}^{O}S_{2}$ and blue for $^{3,O}B/{}^{3}S_{1}\rightarrow[{}^{3}E/{}^{3}H_{4}]^{\ddagger}\rightarrow{}^{1}C_{4}$).

4.2 Degree of preactivation of α-D-mannopyranose conformations.

It is interesting to note that there are other allow free energy regions of the α-mannopyranose FEL (Figure IV-5) with no correlation with the catalytic itineraries known ($^{4}C_{1}$; $^{3}S_{1}$-$^{5}S_{1}$). Additionally, previous studies of our group (Ardèvol, 2010) have shown that in addition to the relative energy of each conformation, other relevant properties (structural and electronic) should be taken into account in order to know which conformations are the best preactivated for catalysis (best suited to be present in the MC of the reaction).

From a structural point of view, the two most relevant parameters are the C1-O5 and C1-O1 distances because these are the ones that change during catalysis (the C1-O5 acquires partial double bond character and the C1-O1 distance lengthens in the oxocarbenium ion-like TS (Whitfield, 2007)). The orientation of the C1-O1 bond (axial for an oxocarbenium ion) is another structural factor to be taken into account, as well as the cis-1,2 diaxial steric hindrance (axial orientations of the O2 hydroxyl group are disfavored)(Davies, 2012). From an electronic point of view, the most relevant parameters are the relative charge at the anomeric carbon (q_{C1}), ring oxygen (q_{O5}) and glycosidic oxygen (q_{O1}).

To track how these parameters change within each conformation, a set of approximately 500 structures was selected from the MTD simulation and submitted to geometry optimization. The optimized structures were clustered to the canonical distortions according to their θ and ϕ final values and their distances and charges were analyzed, as well as the orientation of C1-O1 and C2-O2 (both orientations were measured as the angle between the

bond of interest and the average ring plane, $\Omega_{C1\text{-}O1}$ and $\Omega_{C2\text{-}O2}$). Each of the seven parameters was computed for each structure within the group, and the average values were assigned to the corresponding canonical conformation.

Figure IV-6 shows 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 $^{3,O}B$ to $^{2,5}B$ show the major increase in C1-O1 distance and decrease in C1-O5 distance, with respect to an undistorted $^{4}C_1$ conformation (broken lines). Therefore, these conformations are the ones that most resemble the TS of the hydrolysis reaction in terms of bond distances.

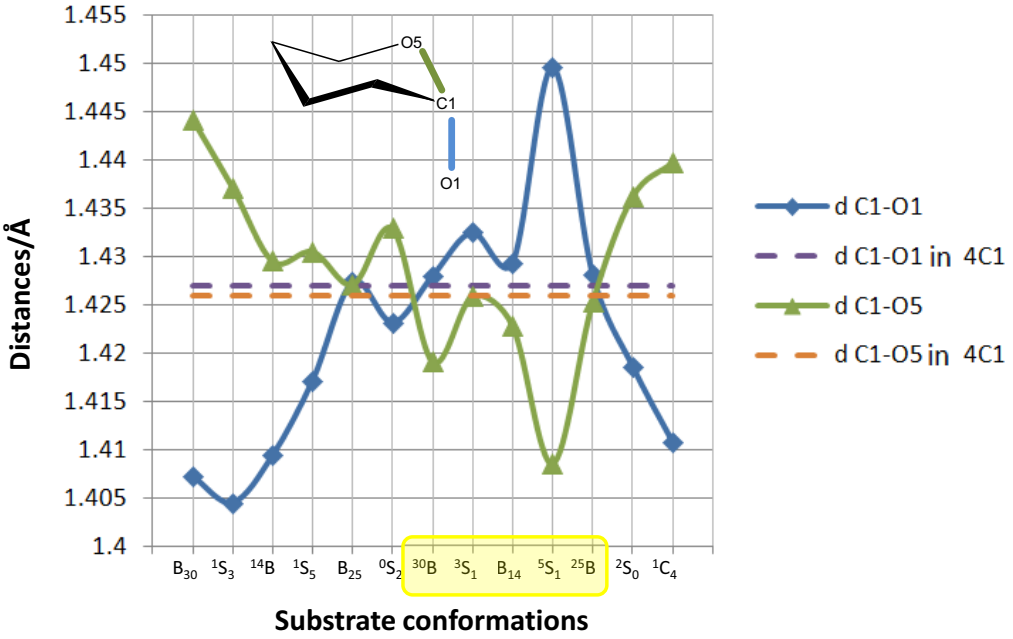

Figure IV - 6. Distances as a function of ring conformation. The structures highlighted in yellow are the most preactivated for catalysis.

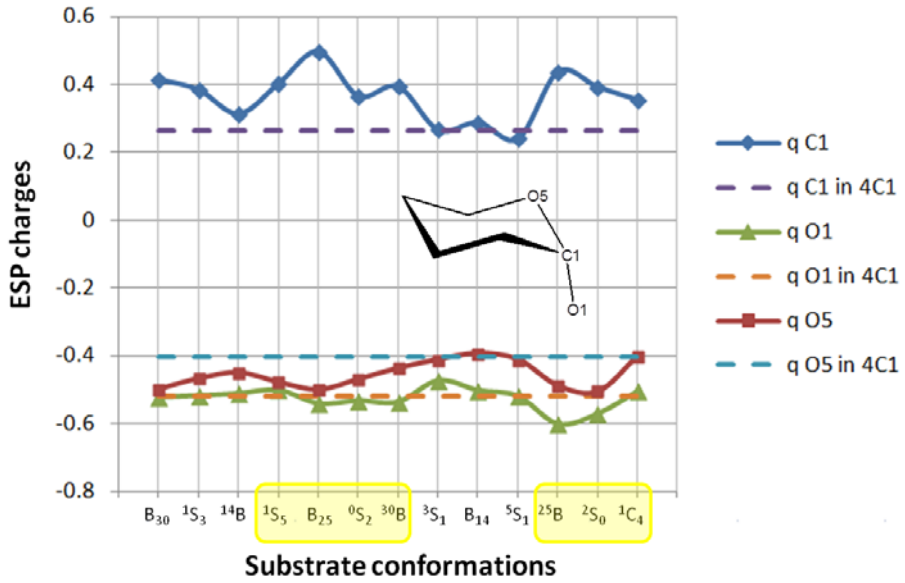

Figure IV - 7. Atomic charges on the anomeric carbon (q_{C1}), the hydroxyl oxygen of the anomeric carbon (q_{O1}) and the ring oxygen (q_{O5}) as a function of ring conformation. The structures highlighted in yellow are the most preactivated for catalysis.

The variations in the atomic (ESP) charges for the atoms C1, O1 and O5 (Figure IV-7) show that the conformations having a larger q_{C1} are the ones that fall near $B_{2,5}$ and $^{2,5}B$. Only slightly changes are observed on the values of q_{O1} and q_{O5}.

The orientation of the C1-O1 and C2-O2 bonds was analyzed by measuring the angle Ω between these bonds and the average plane of the ring (Figure IV-8a). For $\mathbf{\Omega}_{C1\text{-}O1}$, which is negative due to the α-sugar configuration, more negative values indicate a high tendency to an axial configuration, which favors catalysis. In contrast, for $\mathbf{\Omega}_{C2\text{-}O2}$ (β-configuration) more positive values indicate an axial configuration, which inhibits catalysis due to the steric clash with the nucleophile of the reaction. As shown in Figure IV-8b, conformations around $^{3,0}B$ show the best compromise between both parameters.

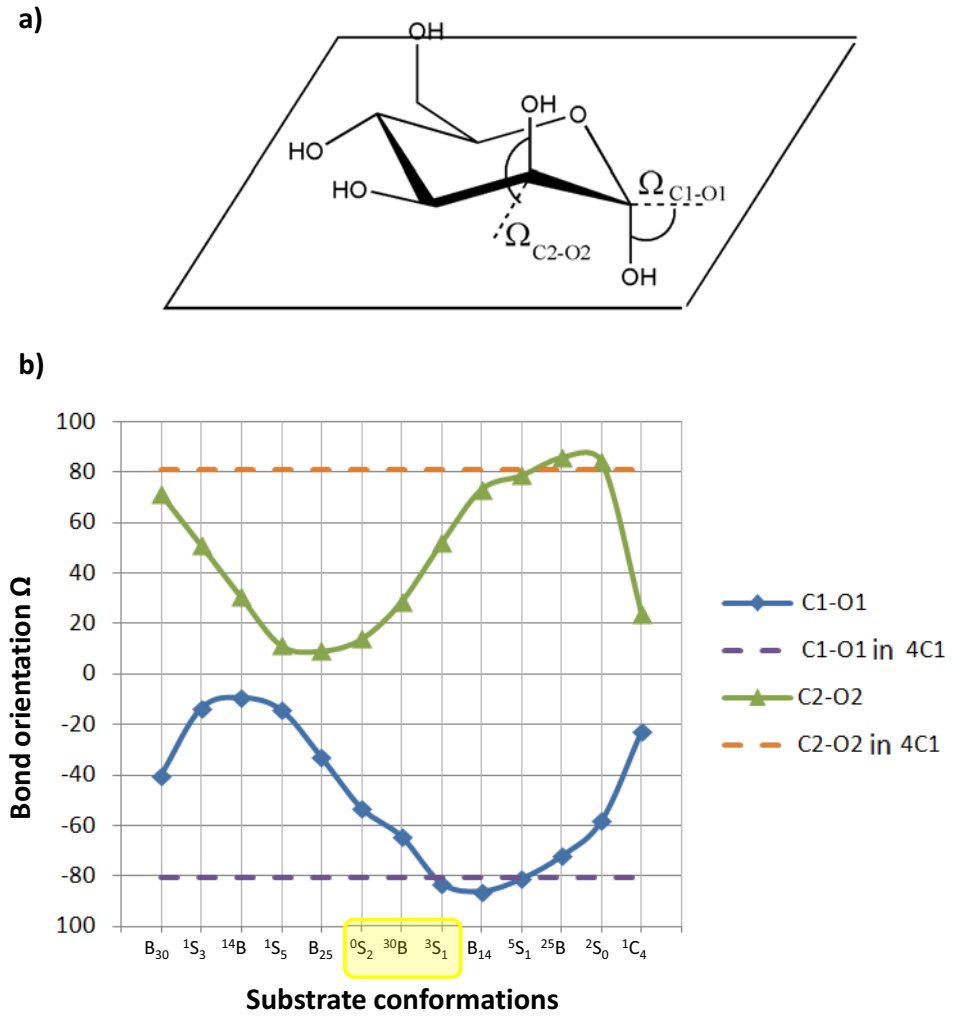

Figure IV - 8. a) Schematic representation of the $\Omega_{C1\text{-}O1}$ and $\Omega_{C2\text{-}O2}$ bond orientation. b) Orientation of the C1-O1, C2-O2 bonds with respect to the mean plane of the ring for each the ring conformation.

Unlike what was observed for β-D-glucopyranose (Biarnés, 2007), there are no conformers with optimum values of all parameters determining the degree of preactivation. Therefore, conclusions cannot be based on one parameter separately but all of them need to be considered together. For this reason, we combined the seven parameters discussed above ($q_{C1,}$ q_{O5}, q_{O1}, $d_{C1\text{-}O1}$, $d_{C1\text{-}O5}$, $\Phi_{C1\text{-}O1}$ and $\Phi_{C2\text{-}O2}$) along with the relative free energy (ΔG_{rel}) into a unique index that would reflect the likelihood that a given conformation would be adopted in the MC of an α-mannosidase. This was done by assigning for each conformation j a score for each parameter x_i using the following formulas:

$$score(x_{i,j}) = \frac{x_{i,j} - x_{i,j}^{min}}{x_{i,j}^{max} - x_{i,j}^{min}} \times 100 \quad for\ x_i = d_{C1-O1}, q_{C1}, q_{O1,} |\phi_{C1-O1}| \quad (\mathbf{Eq.\,IV-1})$$

$$score(x_{i,j}) = \frac{x_{i,j}^{max} - x_{i,j}}{x_{i,j}^{max} - x_{i,j}^{min}} \times 100 \quad for\ x_i = d_{C1-O5}, q_{O5,} \Delta G_{rel}, |\phi_{C2-O2}| \quad (\mathbf{Eq.\,IV-2})$$

The values of the parameters and the corresponding scores are given in Table IV-2. Since the score for each parameter is normalized, they can be directly compared. We then defined an index ξ_j as the average of the scores for the n parameters (n = 8 in our case) for a given conformation j:

$$\xi_j = \sum_j \frac{score(x_{i,j})}{n} \qquad (\mathbf{Eq.\,IV-3})$$

Figure IV-9a shows the variation of ξ with ring distortion. There is no single conformation with the optimum values for every parameter (ξ = 100). For α-D-mannopyranose conformations $B_{2,5}$, $^{O}S_2$, $^{3,O}B$, $^{3}S_1$, $B_{1,4}$, $^{5}S_1$ and $^{2,5}B$ (dotted square; Figure IV-9a,b), as well as $^{4}C_1$, are among the most stable, in terms of free energy (Figure IV-4a). Moreover, these conformations were found to be preactivated for catalysis in terms of energy, charge, and structural parameters. Interestingly, the expected MC conformations of α-mannosidases ($^{O}S_2$ for retaining enzymes and $^{3}S_1$-$^{1}S_5$/$B_{2,5}$ for the inverting ones) are among this reduced set of conformations and thus are preactivated for catalysis.

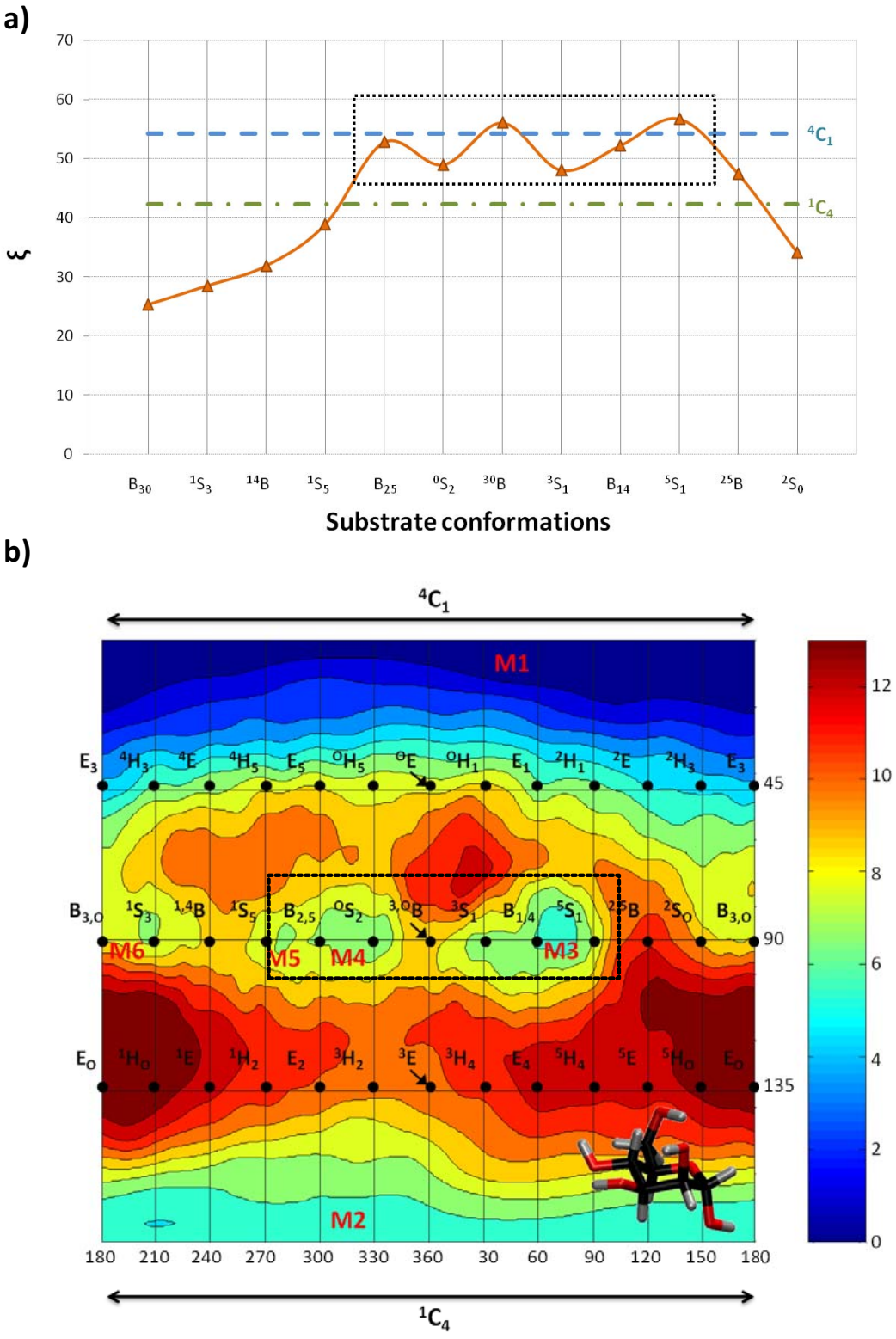

Figure IV - 9. a) Variation of the preactivation index ξ as a function of the ring conformation obtained for α-D-mannopyranose. b) Conformational FEL of an isolated α-D-mannopyranose, with preactivated conformations enclosed by a dotted line. Energy values are in kcal·mol^{-1} and each contour line corresponds to 1 kcal·mol^{-1}.

Table IV - 2. Values of the different structural, electronic and energy parameters, along with its score (in grey) and the resulting preactivation index (ξ), associated to each canonical conformation (Figure IV-9a). Distances are given in angstroms, ESP charges in electrons, Φ_{C1-O1} and Φ_{C2-O2} in degrees and free energy (ΔG_{rel}) in kcal mol^{-1}.

conf.	q (C1)	q (O5)	q (O1)	dist. C1-O1	dist. C1-O5	ϕ_{C1-O1}	ϕ_{C2-O2}	ΔG_{rel}	ξ
$^{4}C_{1}$	0.26	-0.40	-0.52	1.43	1.43	-80.67	81.17	0.0	54
	9	90	37	48	51	93	6	100	
$^{1}C_{4}$	0.36	-0.40	-0.50	1.41	1.44	-22.57	23.88	4.9	42
	45	94	24	14	13	17	81	51	
$B_{3,O}$	0.42	-0.50	-0.52	1.41	1.44	-40.17	71.27	7.6	25
	68	6	39	6	0	40	19	24	
$^{1}S_{3}$	0.39	-0.47	-0.52	1.40	1.44	-13.62	51.06	6.8	28
	56	34	35	0	20	5	45	32	
$^{1,4}B$	0.31	-0.45	-0.51	1.41	1.43	-9.43	30.52	7.8	32
	29	49	30	11	41	0	72	22	
$^{1}S_{5}$	0.40	-0.48	-0.50	1.42	1.43	-14.33	11.41	6.9	39
	63	24	22	28	38	6	97	31	
$B_{2,5}$	0.50	-0.50	-0.54	1.43	1.43	-33.05	9.24	6.5	53
	100	5	53	51	48	31	100	35	
$^{O}S_{2}$	0.37	-0.47	-0.53	1.42	1.43	-53.20	13.94	6.0	49
	49	32	46	41	32	57	94	40	
$^{3,O}B$	0.40	-0.44	-0.54	1.43	1.42	-64.46	28.68	7.9	56
	60	61	49	52	70	72	75	22	
$^{3}S_{1}$	0.27	-0.41	-0.47	1.43	1.43	-83.02	51.93	6.4	48
	10	84	0	62	51	96	44	37	
$B_{1,4}$	0.29	-0.39	-0.50	1.43	1.42	-86.20	73.17	5.6	52
	19	100	23	55	60	100	16	45	
$^{5}S_{1}$	0.24	-0.41	-0.52	1.45	1.41	-81.19	78.67	7.0	57
	0	84	36	100	100	93	9	31	
$^{2,5}B$	0.44	-0.49	-0.60	1.43	1.43	-72.18	85.72	10.1	47
	77	15	100	53	53	82	0	0	
$^{2}S_{O}$	0.39	-0.50	-0.57	1.42	1.44	-58.16	84.04	8.3	34
	59	0	76	31	22	63	2	18	

4.3 α-D-mannopyranose Enzyme-Substrate free energy landscape.

To investigate the effect of the protein environment on the conformations of α-D-mannopyranose, we computed the FEL of the substrate on a mannosidase-substrate complex, using ϕ and θ puckering coordinates as CVs. We selected the enzyme GH47 α-1,2-mannosidase, whose structure in complex with an S-linked substrate mimic is available.

The strategy of using an S-linked substrate mimic with the wild-type enzyme has been widely used in the conformational studies of Michaelis complexes of GHs (Isorna, 2007)(Hrmova, 2001)(Zhong, 2008). Replacement of the glycosidic oxygen by a sulfur atom lowers the proton affinity of the substrate and, therefore, inhibits the hydrolysis reaction, allowing the crystallographic characterization of the complex. However, the ability of thio-glycosides to mimic the substrate-enzyme interactions has been called into question due to the greater C-S bond length (compared to the C-O bond). In this section we first address the question of whether S-linked substrate inhibitors mimic the interactions of the natural substrate (O-linked).

Figure IV-10(left) shows the CPMD/MM optimized structure of the MC with the natural substrate. The structure displays the reactive conformation expected for an inverting GH, as well as a water molecule ideally oriented for an *in-line* nucleophilic attack on the anomeric carbon. An alignment of the initial X-ray structure with the thio-derivative is shown in Figure IV-10, right. Both structures are in excellent agreement and, therefore, the utilization of S-substrate inhibitors seems a valid approach to gain insights into the MC structure in this case.

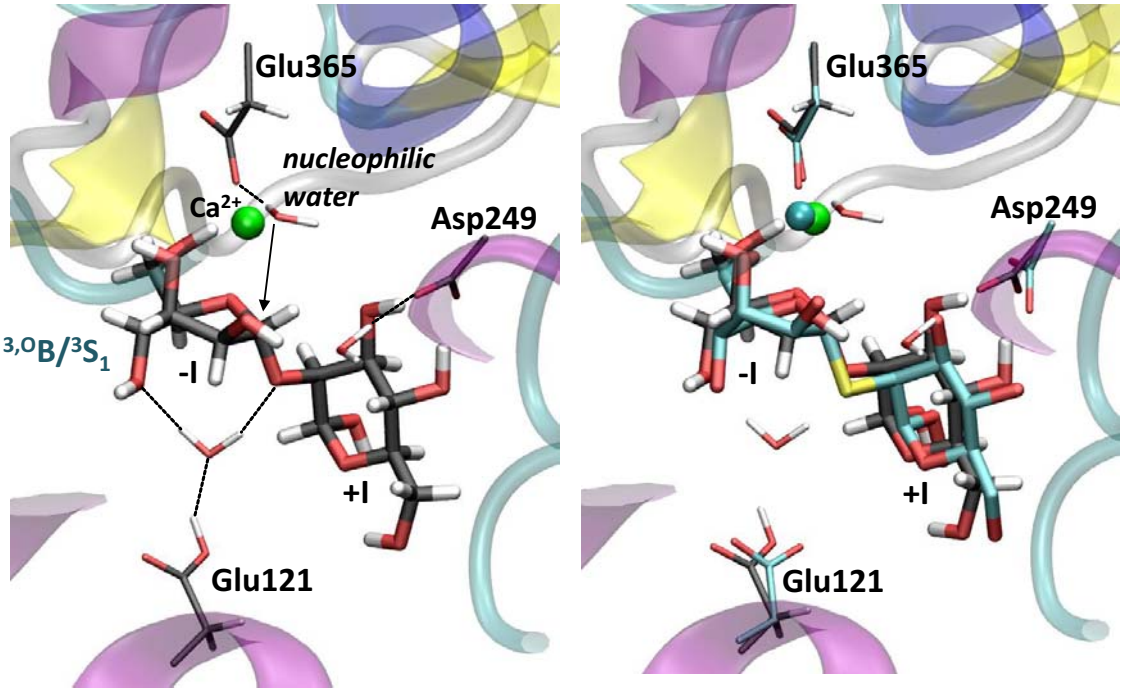

Figure IV - 10. Left) Active site structure of *Caulobacter* α-1,2-mannosidase in complex with α-1,2-mannobiose obtained from CPMD/MM simulations. Right) Structural alignment of the optimized CPMD/MM structure with the initial X-ray structure (thio-derivative).

Figure IV-11 shows the metadynamics simulation of all possible conformations of the -I subsite α-mannose residue, inside protein's cavity. It is clear that the enzyme acts to fundamentally restrict the accessible conformational landscape of the molecule. Of particular note, the undistorted $^{4}C_{1}$ conformer is no longer an energy minimum. On enzyme, the $^{3,O}B/^{3}S_{1}$ conformations are the only stable distorted conformation, defining a clear $^{3,O}B/^{3}S_{1} \rightarrow [^{3}E/^{3}H_{4}]^{\ddagger} \rightarrow {}^{1}C_{4}$ conformational pathway for the reaction coordinate. This conformational pathway is in excellent agreement with the experimentally determined X-ray structures of the enzyme GH47 α-1,2-mannosidase with a series of inhibitors (labels 1-3 Figure IV-11).

It is interesting to note that the FEL presents two energy minima (corresponding to $^{3,O}B/^{3}S_{1}$ and $^{1}C_{4}$ conformations) which are almost of the same stability, although the former is approximately 0.5 kcal·mol^{-1} more stable and wider. This indicates a conformational equilibrium between these two structures in the MC; however, nucleophilic substitution can only occur through $^{3,O}B/^{3}S_{1}$ conformations, as only in these conformations is the C1-H bond in a pseudo-equatorial orientation.

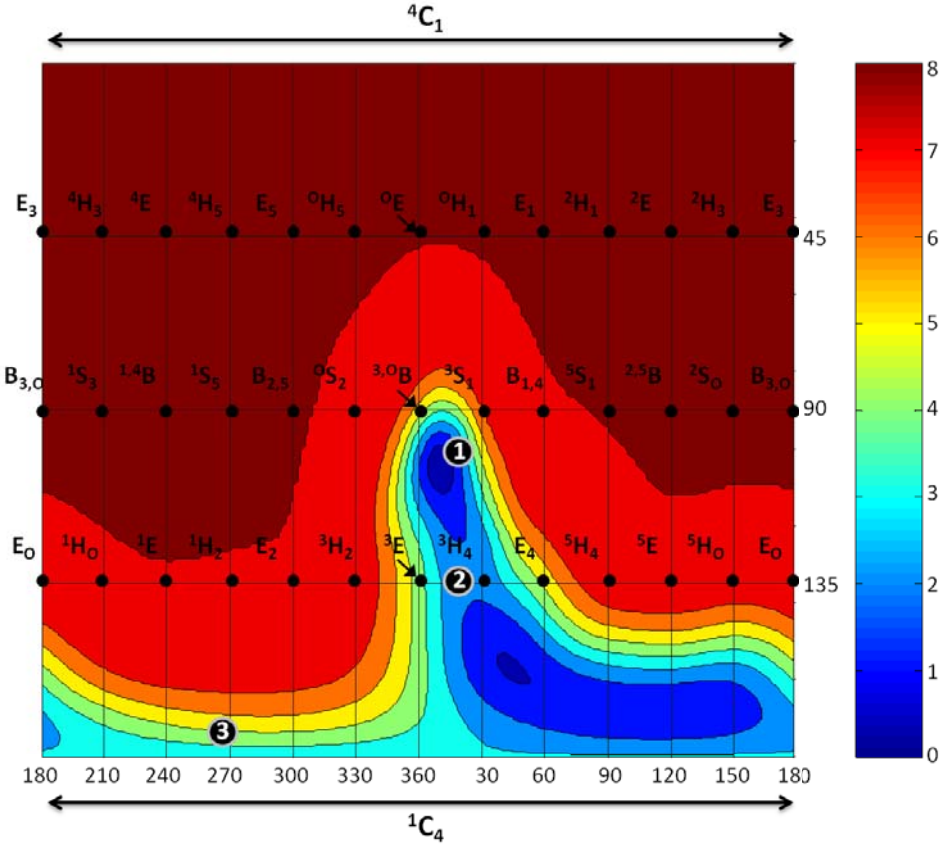

Figure IV - 11. Conformational FEL of the α-mannosyl residue at the –I enzyme subsite of Caulobacter α-1,2-mannosidase, contoured at 1 kcal·mol^{-1}. The labels represent the observed conformations with an S-linked disaccharide (1; pdb code: 4AYO), mannoimidazole (2; pdb code: 4AYQ) and noeuromycin (3; pdb code: 4AYR).

In summary, the CPMD/MM calculation of the enzyme GH47 α-1,2-mannosidase in complex with its natural substrate shows that the enzyme confines the mannose substrate to a reduced set of conformations, which completely agrees with the catalytic itinerary followed by this particular GH enzyme.

4.4 The conformational free energy landscape of mannosidase inhibitors.

In this section, we analyze the conformational properties of isofagomine and mannoimidazole inhibitors. As discussed in section 1.2 of this chapter, mannoimidazole inhibitors are good models of the mannopyranosyl oxocarbenium ion-like TS (true TS mimic), reporting on both the transition state conformation and the conformational itinerary (Thompson, 2012). On the contrary, isofagomine inhibitors resemble the glycosyl cation with charge localized at C1, but are poor mimics of the TS conformation.

To find out the intrinsic conformational properties of isofagomine and mannoimidazole, we computed their conformational free energy landscapes. The FEL of isofagomine (Figure 12, a) reveals that the $^{4}C_{1}$ conformation is the preferred one, with the mechanistically relevant distorted conformations (the TS for mannosidases are $^{4}H_{3}$ and $B_{2,5}$; see Figure IV-2) located at least 8 kcal·mol^{-1} higher in energy. This data suggest that isofagomine is a poor TS mimic for mannosidases because of the high energy barrier for the interconversion to other conformations other than the $^{4}C_{1}$. In fact, isofagomine-type inhibitors exhibit this undistorted conformation in all X-Ray structures of mannosidase complexes to date (1-3 in Figure IV-12a) .

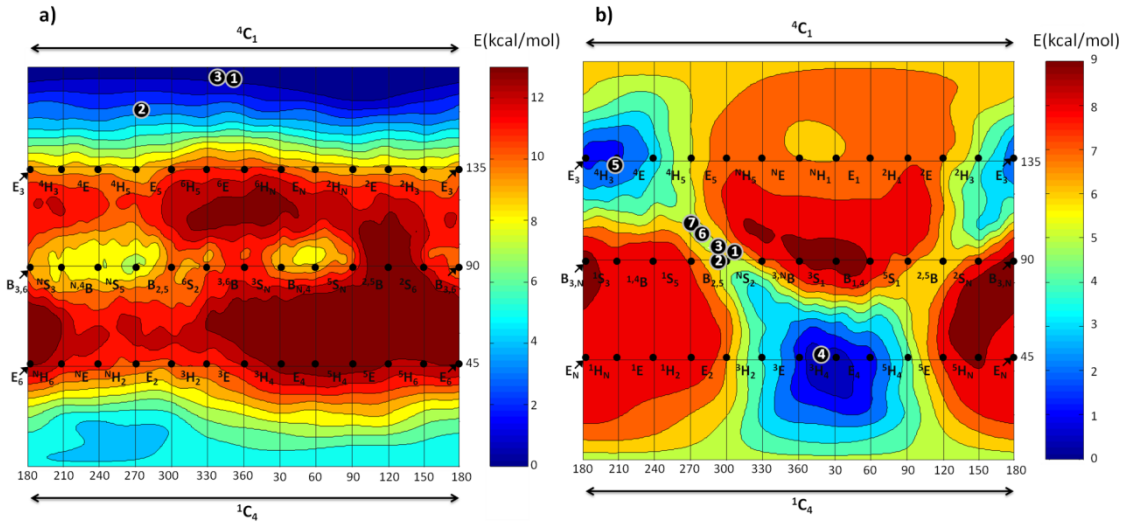

Figure IV - 12. Conformational free energy landscapes of isofagomine (a) and mannoimidazole (b), contoured at 1 kcal·mol^{-1}. FELs have been annotated with the conformations of isofagomine-type (a) and mannoimidazole-type (b) inhibitors that have been observed on-enzyme. a) 1: isofagomine bound to GH26 *Cj*Man26C (pdb code: 4CD4); isofagomine bound to GH113*Aa*ManA (pdb code: 4CD6); 3: α-Glc-1,3-isofagomine bound to *Bx*GH99 (PDB code: 4AD2). b) 1: mannoimidazole bound to GH26 *Cj*Man26C (pdb code: 4CD5); 2: mannoimidazole bound to GH113*Aa*ManA (pdb code: 4CD7))(Williams, 2014); 3: mannoimidazole bound to GH2 *Bt*Man2A (PDB code: 2VMF)(Tailford, 2008); 4: mannoimidazole bound to GH47 *Ck*Man47 (PDB code 4AYP)(Thompson, 2012b); 5: mannoimidazole bound to GH38 *Dm*GManII (PDB code 3D4Y) in half-chair conformer (Kunts, 2008); 6: mannoimidazole bound to GH38 *Dm*GManII (PDB code 3D4Y) in boat conformer; 7: mannoimidazole bound to GH92 *Bt*Man3990 (PDB code 2WZS)(Zhu, 2010).

In striking contrast, the FEL of mannoimidazole (Figure IV-12b) is consistent with a good TS shape mimicry, with all the mechanistically relevant conformations for GHs (half-chair: $^{4}H_{3}$ and $^{3}H_{4}$; envelope: ^{3}E, E_{3}, ^{4}E and E_{4}; and boat: $^{2,5}B$ and $B_{2,5}$) energetically accessible. A global minimum is found near the $^{4}H_{3}$ conformation with a second local minimum, near the $^{3}H_{4}$, approximately 1 kcal·mol^{-1} higher in energy. Both half-chair conformations of mannoimidazole

are observed in inhibitor-protein complexes (labels 4,5; Figure IV-12b). The other conformation of mannoimidazole, observed on-enzyme, the $B_{2,5}$, is near a saddle point between these two minima and was 5 kcal·mol^{-1} higher in energy than the $^{3}H_{4}$ (labels 1-3 and 6-7; Figure IV-12b).

Labels 1-7 in Figure IV-12b, show the conformations of mannoimidazole-type inhibitors bound to mannosidases of diverse families, which report on the TS of the hydrolysis reaction. This highlights that mannosidases are able to modulate the inhibitor conformation to adapt it to its TS. The observation of a $B_{2,5}$-mannoimidazole is of special significance, with the enzyme inducing the inhibitor to adopt a conformation which is not an energy minima but a saddle point between them.

In conclusion, the FEL for mannoimidazole reveals that all the conformations relevant for catalysis are energetically accessible and, therefore, it is easy for the enzyme to distort the compound to the TS of the reaction. In contrast, the strong bias of the isofagomine inhibitor towards a $^{4}C_{1}$ conformation hinders the distortion towards the TS of the hydrolysis reaction. This explains why experimentally determined mannoimidazole-enzyme complexes are found in a variety of conformations, which agree with the TS of the reaction, whereas isofagomine complexes are found in a relaxed chair conformation.

5. Conclusions.

The main conclusions reached in this chapter are the following:

1. By means of *ab initio* metadynamics simulations, we have demonstrated that the low-energy minima of the conformational FEL of an isolated α-mannose correlate with the observed structures of enzyme-ligand complexes.

2. The FEL for α-mannose gives support for the conformational itineraries proposed experimentally for α-mannosidases ($^{1}S_{5} \leftrightarrow [B_{2,5}]^{\ddagger} \leftrightarrow {}^{O}S_{2}$ conformational itinerary for GH38 (Numao, 2003) and GH92 (Zhu, 2010) and $^{3}S_{1} \rightarrow [^{3}H_{4}]^{\ddagger} \rightarrow {}^{1}C_{4}$ for GH47 (Thompson, 2012)).

3. α-mannosidases act on reshaping the conformational energy landscape of the α-mannoside substrates, restricting the accessible conformations to the ones useful for catalysis.

4. One important factor that dictates TS mimicry properties of α-mannosidase inhibitors are the intrinsic conformational properties of the inhibitor molecules.

Chapter V – The conformational free energy landscape of oxocarbenium ions.

The conformational free energy landscape of oxocarbenium ions.

1. Introduction.

In the previous chapter we have studied two examples of mannosidase inhibitors (isofagomine and mannoimidazole) to determine which features imbue transition state (TS) mimicry. We concluded that mannoimidazole is a "true" TS mimic because all possible conformations that conform the stereoelectronic requirements of an oxocarbenium ion ($^{4}H_{3}$, $^{3}H_{4}$, $^{2,5}B$, and $B_{2,5}$, hereafter referred as TS^{GH})(Figure V-1) are energetically accessible whereas isofagomine is strongly biased towards a $^{4}C_{1}$ conformation and, therefore, it does not mimic the TS.

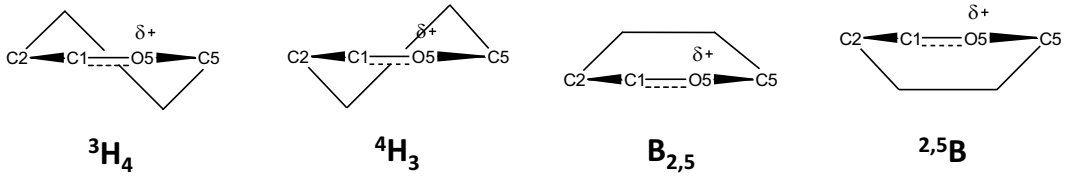

Figure V - 1. Candidate conformations for the TS of the reactions catalyzed by GHs (TS^{GH}). These conformations present the C2, C1, O5 and C5 atoms in a coplanar arrangement.

A question now arises: Which are the conformational preferences of a pure oxocarbenium ion for a given sugar (e. g. glucosyl or mannosyl cation)? Previous computational studies on this topic are not conclusive. Two stable conformations ($^{4}H_{3}$ and $^{3}H_{4}$) for cyclohexene, which interconvert via boat conformations ($^{2,5}B$ and $B_{2,5}$), have been reported (Allinger, 1972) with a force field method. For the particular case of the mannosyl cation, *ab initio* calculations showed different relative stabilities between the $^{3}H_{4}$ and $^{4}H_{3}$ conformations depending on the reference (Nukada, 2005)(Amat, 2000), whereas experiments suggest that the $^{3}H_{4}$ conformation it the favored one (Davis, 1998).

Bowen and coworkers (Woods, 1992) reported, by molecular mechanics calculations, that electrostatic interactions and steric effects exert an important effect on the conformational preferences of six-membered oxocarbenium ions. In particular, they showed that the orientation of hydroxyl groups at positions C3 and C4 play important roles in the conformational stability of cyclic oxocarbenium ions, being the axial orientations the most stable ones (axial configurations were predicted to be 2.3-5.0 kcal·mol^{-1} more stable than the equatorial ones). This was attributed to stabilizing interactions between the partial negative charge of the hydroxyl oxygen atom and the carbocationic carbon, which are positioned close together in the axial conformation. Additionally, it was found that an axial orientation of C4-OH results in a higher stabilization effect compared to C3-OH, whereas C2-OH prefers a pseudo-equatorial orientation.

Here, we apply *ab initio* metadynamics, following the procedure described in the previous chapters, to obtain the FEL of four different sugar cations (D-glucose, D-mannose, xylose and L-fucose oxocarbenium ions). Together with the analysis of electronic and structural changes upon ring distortion, we determine the most preactivated conformations, for each sugar oxocarbenium, and compare them with the expected TS conformations deduced from

experimental data. Finally, we compare these results with the ones obtained for mannoimidazole in the previous chapter to gain further insights about TS mimicry.

1.1 Experimental data on oxocarbenium ions conformations.

Nowadays, it is known that the four TS^{GH} conformations ($^4H_3$, $^3H_4$, $^{2,5}B$, $B_{2,5}$) are used by GHs to perform the hydrolysis reaction. The evidence comes from analysis of X-ray structures with modified substrates (i. e. fluorinated, or non-hydrolyzable substrates), TS mimic inhibitors or enzyme active mutants, which are able to trap relevant species along the GH reaction mechanism (e. g. Michaelis complex, TS or covalent intermediate). Localization of these species on the Stoddart or Mercator diagrams allow to identify the catalytic itinerary and the TS of a given GH.

1.1.1 D-Glucose ($^4H_3$ and $^{2,5}B$ conformations).

For retaining β-glucosidases, a $^4H_3$ TS conformation was initially deduced from experiments on two cellulose complexes: endo-glucanase I (EGI; GH7)(Sulzenbacher, 1996) in complex with a tiooligosaccharide ($^1S_3$ conformation) and Cel5A GH5 (Davies, 1998) in complex with a fluoroglycoside ($^1S_3$ conformation for the Michaelis complex (MC) and $^4C_1$ for the covalent intermediate). Observation of these conformations ($^1S_3$ and $^4C_1$), which flank the transition state of the reaction, support the hypothesis of a $^4H_3$ conformation as TS. Other subsequent studies also gave support to this itinerary (Money, 2006)(Sandgren, 2004). Based on a similar structural evidence (Lovering, 2005), it was determined that retaining α-glycosidases (glucosidases, xylosidases, etc.) simply follow the reverse pathway $^4C_1 \rightarrow [^4H_3]^{\ddagger} \rightarrow {}^1S_3$ (Figure V-2).

Additionally, support for a $^{2,5}B$ TS for inverting β-glucosidases comes from the distortion of the MC towards a $^2S_O/^{2,5}B$ conformation observed in GH6 inverting β-glucosidases (Zou, 1999)(Guerin, 2002), subsequently confirmed by QM/MM metadynamics simulations (Petersen, 2009)(Figure V-2).

1.1.2 D-Mannose ($^3H_4$ and $B_{2,5}$ conformations).

As mentioned in Chapter IV, two different catalytic itineraries are experimentally supported for α-mannosidases (Figure V-2). A $^1S_5 \leftrightarrow [B_{2,5}]^{\ddagger} \leftrightarrow {}^OS_2$ conformational itinerary[3] has been identified for inverting and retaining α-mannosidases of families 2 (Tailford, 2008), 38 (Numao, 2003), and 92 (Zhu, 2010) and a $^3S_1 \rightarrow [^3H_4]^{\ddagger} \rightarrow {}^1C_4$ itinerary for the inverting α-mannosidases of family 47 (Thompson, 2012).

Identification of a $^1S_5$ conformation for the Michaelis complex of a GH26 retaining β-mannosidase and $^OS_2$ for the covalent intermediate provided the first insight into a new conformational itinerary (Ducros, 2002) passing through a $B_{2,5}$ TS. Additionally, the observation of a large panel of TS mimic inhibitors (demonstrated by linear free energy relationships) close to a $B_{2,5}$ conformation for a GH2 retaining β-mannosidase (Tailford, 2008), provides support for the $B_{2,5}$ conformation as TS.

[3] The double arrow indicates that the reaction mechanism is bidirectional: It goes from left to right for retaining α-mannosidases and in the opposite direction for inverting α-mannosidases.

1.1.3 *Xylose ($^{4}H_3$ and $^{2,5}B$ conformations).*

Two main itineraries have been predicted for retaining β-xylanases (Figure V-2): one that involves a $^{2,5}B$ TS ($^{2}S_0 \rightarrow [^{2,5}B]^{\ddagger} \rightarrow {}^{5}S_1$) and another one that passes through a $^{4}H_3$ ($^{1}S_3 \rightarrow [^{4}H_3]^{\ddagger} \rightarrow {}^{4}C_1$). Support for the first itinerary comes from reports on covalently bound substrates of family 11 GHs in which the xylose saccharide located at the -I subsite (2-deoxy-2-fluoroxylopyranosyl substrate) adopts a $^{2,5}B$ conformation (Sabini, 1999), as well as experiments on synthetic xyloside analogues locked on a $^{2,5}B$ conformation (Amorim, 2011). Support from the second itinerary ($^{1}S_3 \rightarrow [^{4}H_3]^{\ddagger} \rightarrow {}^{4}C_1$) is based on the observation of covalently bound substrates (2-deoxy-2-fluoroxylopyranosyl substrate) exhibiting a $^{4}C_1$ conformation (Noteboom, 1998) and the observation of a $^{1}S_3$ conformation for the MC of a GH10 xylanase (Suzuki, 2009). As observed for the glucosidases, retaining α-xylosidases follow the reverse pathway $^{4}C_1 \rightarrow [^{4}H_3]^{\ddagger} \rightarrow {}^{1}S_3$ (Lovering, 2005) (Figure V-2).

1.1.4 *L-Fucose ($^{3}H_4$ conformation).*

A series of X-ray structures along the reaction coordinate of a GH29 α-L-fucosidase, unraveled a MC in a $^{1}C_4$ conformation and a covalent glycosyl-enzyme intermediate in a $^{3}S_1$ conformation (Lammerts van Bueren, 2010). This data, together with *ab initio* metadynamics simulations on isolated α-L-fucose strongly support a $^{1}C_4 \rightarrow [^{3}H_4]^{\ddagger} \rightarrow {}^{3}S_1$ conformational itinerary for the glycosylation step of the reaction mechanism (Figure V-2).

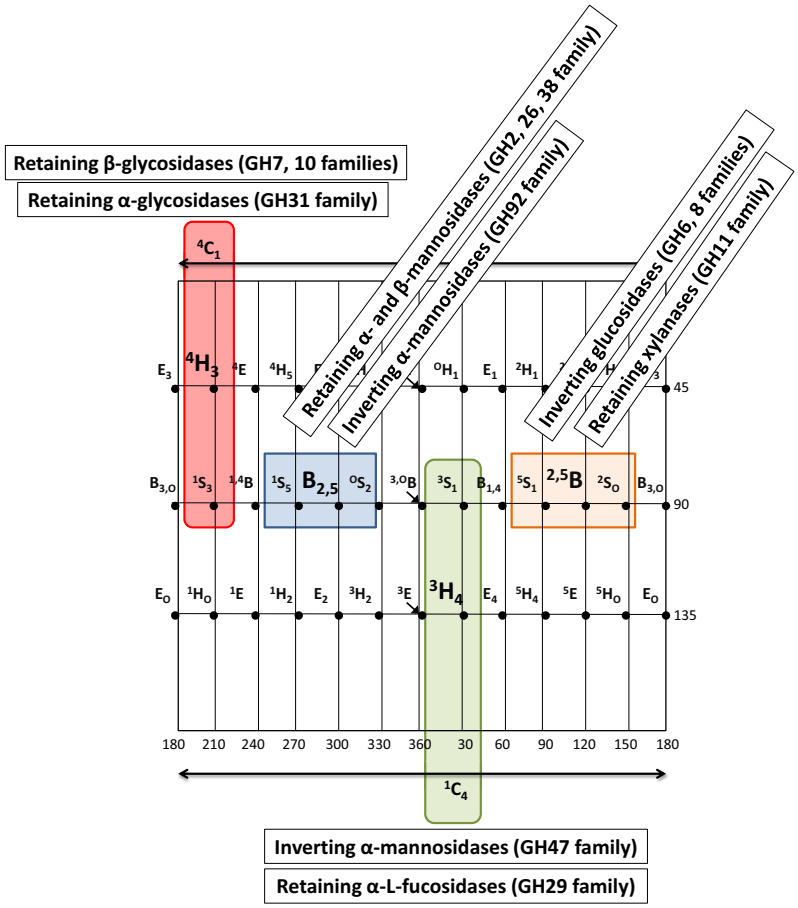

Figure V - 2. Mercator diagram showing the conformational itineraries followed by various GHs. Predicted transition states are highlighted in bold.

2. Objectives

The main objectives of this chapter are the following:

1) Compute the conformational free energy landscape (FEL) of four isolated (gas-phase) sugar oxocarbenium ions to quantify the relative energies among conformations.
2) For a given sugar oxocarbenium ion (e. g. mannosyl cation), determine which are the best conformations to act as TS and compare the results with expected conformations derived from experimental information.
3) Compare the results obtained for a pure mannose oxocarbenium ion (i. e. mannosyl cation) with the ones obtained for mannoimidazole to get further insights into the TS mimic properties of the latter.

3. Computational details.

3.1 Metadynamics simulations.

Ab initio simulations were done within the Car-Parrinello formalism (Car & Parrinello, 1985) using the version 3.15.1 of the CPMD package (CPMD, 1990). A fictitious mass of 500 a. u. and a time step of 0.12 fs were used for all systems studied. The system temperature was set to 300 K by coupling it to a thermostat using the Nosé algorithm (Nosé, 1984). The Kohn-Sham orbitals were expanded in a PW basis set with a kinetic energy cutoff of 70 Ry. *Ab initio* pseudopotentials generated within the Troullier-Martins scheme were employed (Troullier, 1991). The calculations were performed using the PBE generalized gradient-corrected approximation (Perdew, 1996), as in the previous chapters.

The glucosyl, mannosyl, xylosyl and fucosyl cations consist in 22, 22, 18 and 21 atoms, enclosed in orthorhombic boxes with dimensions 16.0 Å × 15.0 Å × 15.0 Å, 15.0 Å × 15.0 Å × 16.0 Å, 15.0 Å × 14.0 Å × 15.0 Å and 14.0 Å × 16.0 Å × 15.0 Å , respectively. All molecules were simulated with a positive charge of +1 e^-. For all systems, 10 ps of free dynamics at 300 K were performed prior to the metadynamics simulations. Glucosyl and mannosyl cations required the addition of a restraint in the dihedral angle C4-C5-C6-O6 to avoid the interaction between the O6 oxygen of the hydroxymethyl group and the positive charge of the C1 carbon atom.

The metadynamics simulations were performed using the version 3.15.1 of the CPMD package (CPMD, 1990) coupled to Plumed2 (Tribello, 2014), previously modified to allow the usage of the Mercator CVs. For all simulations, Gaussian-like functions with width (δs) of 0.10 rad, an initial height (w) of 0.20 kcal·mol^{-1} and a temperature window of ΔT = 2100 K for the well-tempered algorithm were used together with a deposition time of 60 fs. A total of 19716, 12700, 14381, and 14620 Gaussians were added for glucosyl, mannosyl, xylosyl and fucosyl cations, which correspond to ≈ 1200 ps, 760 ps, 860 ps and 880 ps of total simulation time, respectively.

4. Results and discussion

4.1 D-glucosyl cation FEL.

The FEL for the D-glucosyl cation (i. e. D-glucose oxocarbenium ion) (Figure V-3a, b) reveals that the four possible TS^{GH} conformations are energetically accessible being the $^{3}H_{4}$ the most stable one. The $^{4}H_{3}$ conformation is also an energy minima, approximately 1.1 kcal·mol^{-1} higher in energy, whereas the $^{2,5}B$ and $B_{2,5}$ conformations are intermediate structures between both energy minima and are 2.9 and 4.0 kcal·mol^{-1} above the $^{3}H_{4}$, respectively. Contrary to what we expected, none of the two TS conformations for glucose hydrolysis deduced experimentally ($^{4}H_{3}$, $^{2,5}B$) is the most stable conformation of the FEL ($^{3}H_{4}$).

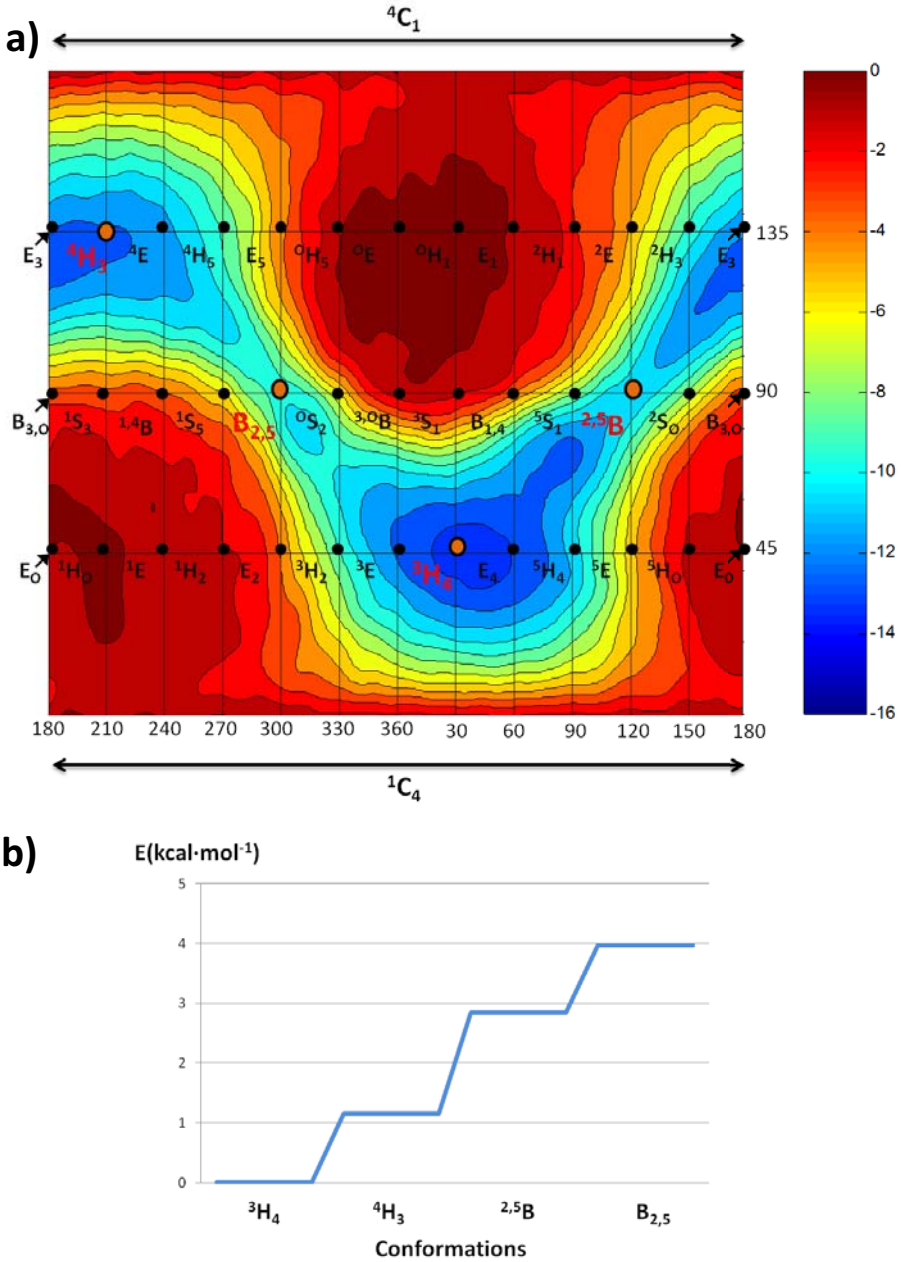

Figure V - 3. a) Conformational free energy landscape of an isolated D-glucosyl cation. Energy values are in kcal·mol^{-1} and each contour line corresponds to 1 kcal·mol^{-1}. b) Free energy of the four D-glucosyl cation TS^{GH} conformations relative to $^{3}H_{4}$.

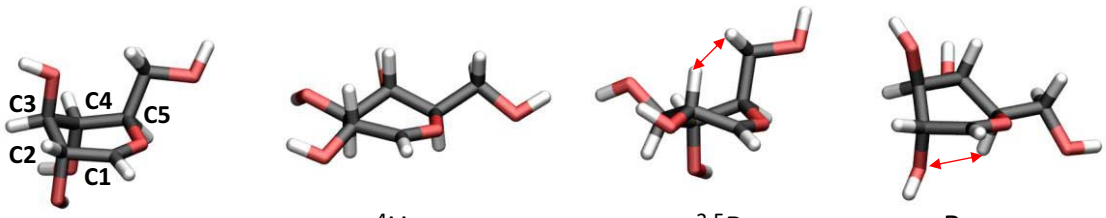

Figure V - 4. Four TS^{GH} conformations for a D-glucosyl cation. Red arrows indicate close contacts between atoms.

The $^{3}H_4$ conformation is the most stable because, as mentioned in the introduction, it has both C3 and C4 hydroxyl groups in an axial orientation, which stabilize the positive charge developed at C1 (the distance between the C1 carbon atom and the hydroxyl groups at positions 3 and 4, is shorter for an axial bond orientation than it is for an equatorial bond) (Figure V-4). On the contrary, the hydroxyl group at C2 exhibits the less preferred axial configuration. There is an additional stabilization effect coming from the intramolecular hydrogen bond interaction between C2 and C4 hydroxyl groups. The $^{2,5}B$ and $B_{2,5}$ conformations also display an axial orientation for some hydroxyl groups, but both structures have hydrogen-hydrogen or hydrogen-hydroxyl group steric clashes that result in a destabilization effect (Figure V-4). Finally, the $^{4}H_3$ conformation, the second most stable geometry, lacks the steric clashes observed in both boat conformations but has the C3 and C4 hydroxyl groups in the less stable equatorial orientation. A "clockwise" hydrogen bond pattern between C2-C3-C4 hydroxyl groups additionally stabilizes this conformation.

4.2 D-mannosyl cation FEL.

The FEL for the D-mannosyl cation (Figure V-5a, b) is qualitatively similar to the one of the D-glucosyl cation, although the relative energies between representative TS conformations are different. In this case, there is one main energy minima, centered on the $^{3}E/^{3}H_4$ conformation, which also covers the $B_{2,5}$ conformation, being 1.6 kcal·mol^{-1} higher in energy. The $^{4}H_3$ conformer corresponds to a small minima 3.5 kcal·mol^{-1} above in energy with respect to the global minimum ($^{3}E/^{3}H_4$ conformation) and the $^{2,5}B$ is an intermediate conformation between half-chair conformations ($^{3}H_4$ and $^{4}H_3$), the highest in energy of the four possible TS's (7.6 kcal·mol^{-1}). Interestingly, the two TS conformations used by mannosidases ($^{3}H_4$ and $B_{2,5}$) are the ones with the lowest energies on the FEL (Figure V-5b).

As observed for the D-glucosyl cation, the $^{3}H_4$ conformation is the most stable one due to the axial orientation of both C3 and C4 hydroxyl groups, which stabilize the positive charge developed at C1, and the presence of intramolecular hydrogen bonds (Figure V-6). Contrary to the D-glucose ion, the $B_{2,5}$ D-mannosyl cation is the second most stable conformation, thanks to the equatorial orientation of the C2 hydroxyl group that prevents the steric clashes observed for the $B_{2,5}$ D-glucosyl cation (Figure V-4) and allows formation of an intramolecular hydrogen bond between the C2 and C3 hydroxyl groups. The $^{4}H_3$ conformation is higher in energy than the corresponding D-glucose conformation due to the absence of the hydrogen bond pattern observed previously (Figures V-4 and 6). Finally, the $^{2,5}B$ conformation is the less stable due to the steric clashes between the C2 hydroxyl group and the C5 hydrogen atom (Figure V-6).

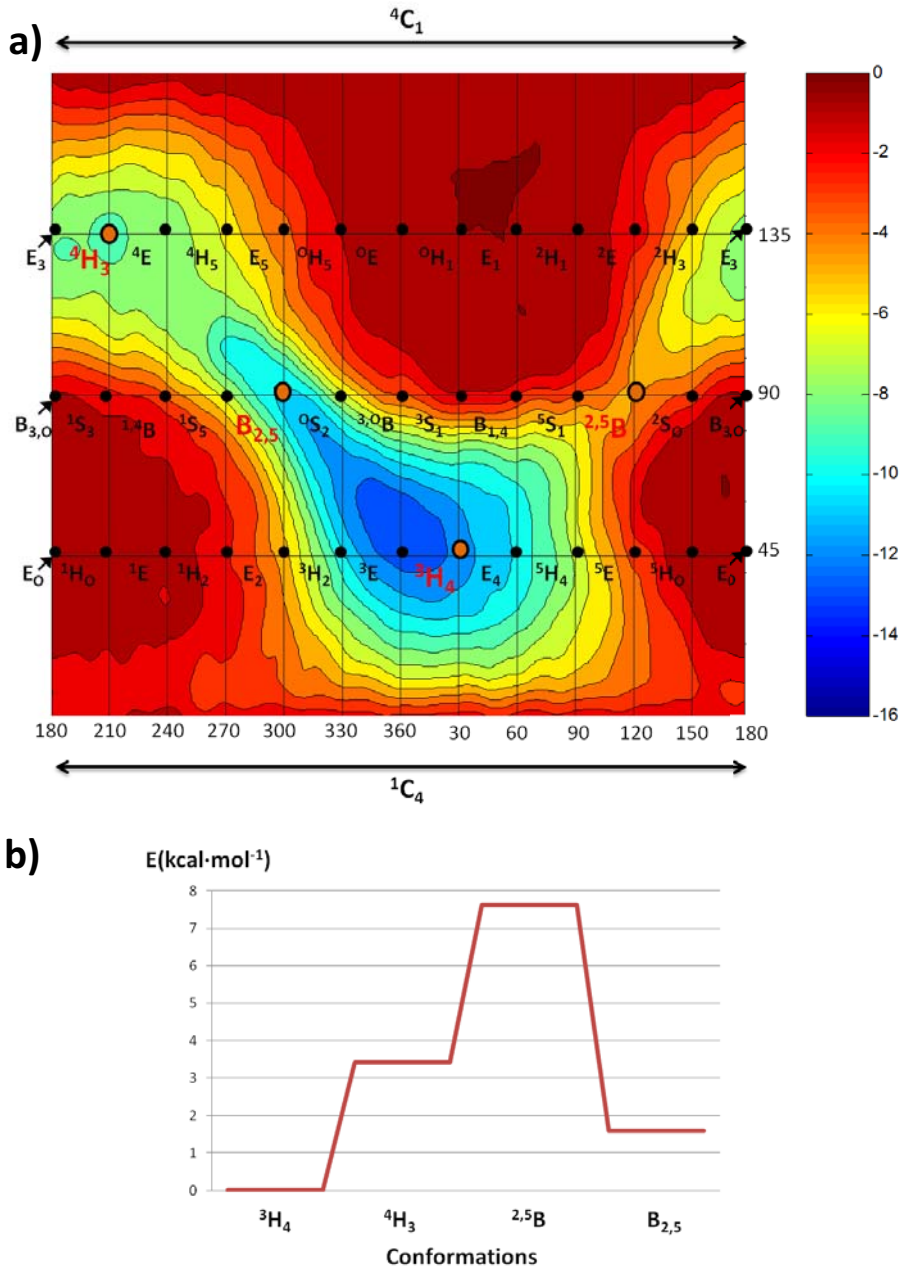

Figure V - 5. a) Conformational free energy landscape of an isolated D-mannosyl cation. Energy values are in kcal·mol^{-1} and each contour line corresponds to 1 kcal·mol^{-1}. b) Free energy of the four TS^{GH} conformations of D-mannosyl cation ion relative to $^3H_4$.

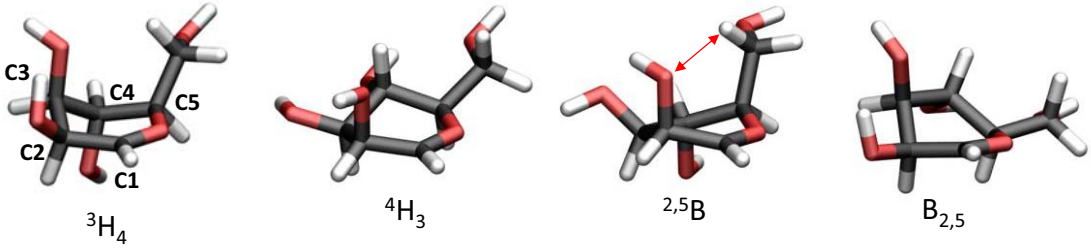

Figure V - 6. Four TS^{GH} conformations for a D-mannosyl cation. Red arrows indicate close contacts between atoms.

4.3 Xylosyl cation FEL.

The FEL for the xylosyl cation also shows the four relevant TS^{GH} conformations ($^{4}H_3$, $^{3}H_4$, $^{2,5}B$, $B_{2,5}$) in low energy regions (Figure V-7a). As observed previously, the $^{3}H_4$ is the global energy minima covering also other TS^{GH} conformations (the $^{2,5}B$ in this case; 1.4 kcal·mol^{-1} higher in energy)(Figure V-7b). The $^{4}H_3$ and $B_{2,5}$ conformations are 2.9 and 6.1 kcal·mol^{-1} above in energy with respect to the global minima. As observed for the D-glucosyl cation, but not for the D-mannose ion, the most stable conformation of the FEL ($^{3}H_4$) does not correspond to the TS for the hydrolysis reaction predicted experimentally ($^{4}H_3$ and $^{2,5}B$).

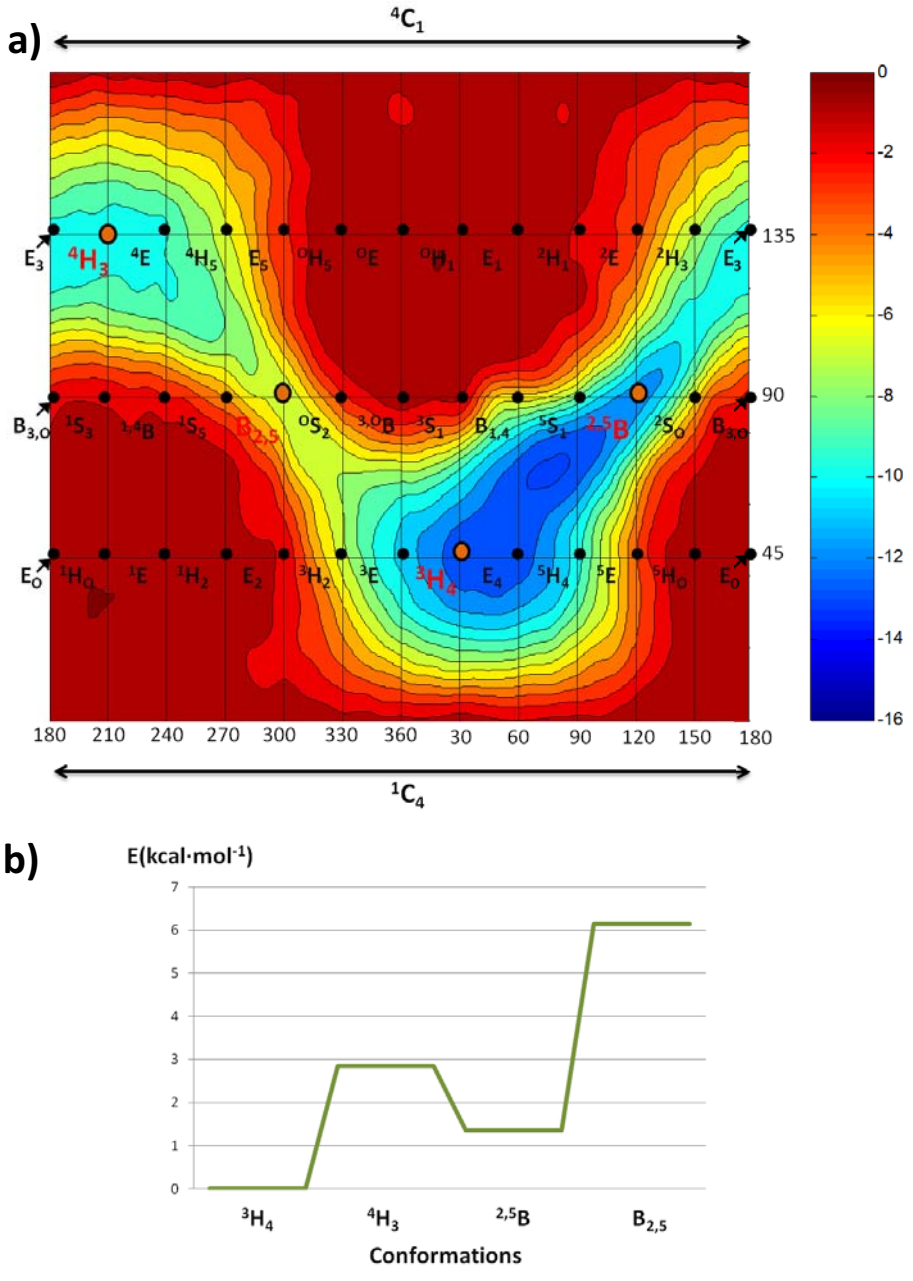

Figure V - 7. a) Conformational free energy landscape of an isolated xylosyl cation. Energy values are in kcal·mol^{-1} and each contour line corresponds to 1 kcal·mol^{-1}. b) Free energy of the four TS^{GH} conformations of xylosyl cation relative to $^{3}H_4$.

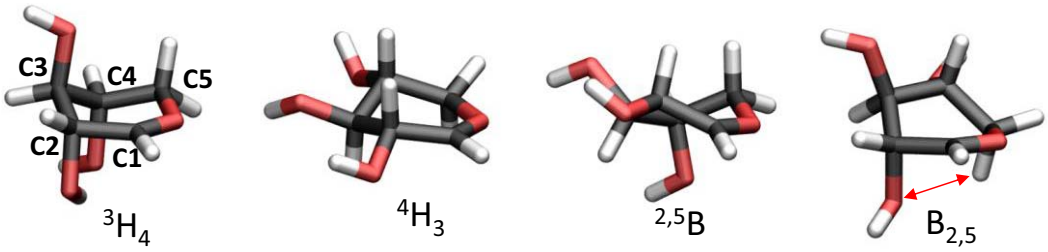

Figure V - 8. Four TS^{GH} conformations for a D-xylosyl cation. Red arrows indicate close contacts between atoms.

The xylose molecule is identical to glucose but lacks the hydroxymethyl group attached to the C5 carbon atom (Figures V-4 and V-8). Relative energies, among the four conformers are also similar to the glucosyl cation ($^{3}H_4$ is the global minimum)(Figures V-3b), with the exception of $^{2,5}B$ being particularly stable for the xylosyl cation (Figure V-7b). The absence of the hydroxymethyl group at C5 avoids the steric clashes with the C2 hydrogen atom observed for the D-glucosyl cation (Figures V-4 and V-8), which explains the higher stability of the latter conformation. As observed for other sugar ions, the $^{4}H_3$ conformation lacks an axial orientation for the hydroxyl groups, but is able to form intramolecular interactions that stabilize this conformation. The $B_{2,5}$ conformation shows the highest relative energy (Figure V-7) due to steric clashes between the C2-OH and the hydrogen atom at C5 carbon atom. This destabilizing effect was also reported for the glucosyl cation (Figure V-4).

4.4 L-Fucose FEL.

Fucose is the only sugar with an L-configuration studied in this chapter. As shown in Figure V-9a, the shape of the FEL is qualitatively similar to the previous sugars, but the relative energies show a different behavior (Figure V-9b). The fact that a single conformational itinerary, passing through $^{3}H_4$, has been predicted experimentally is in perfect agreement with the FEL (this conformation is the main energy minima and all other conformations are at least 6 kcal·mol^{-1} higher in energy). Therefore, the simulations predict that a catalytic itinerary involving a $^{3}H_4$ conformation is the only possible itinerary.

The characteristic energy profile observed for the L-fucosyl cation can be attributed to the orientation of the methyl group attached to the C5 carbon atom, which is different than the orientation observed for other sugars. This methyl group destabilizes both $^{4}H_3$ and $B_{2,5}$ conformations (Figure V-10) because of its close proximity to the other hydrogen atoms or hydroxyl groups. The $^{2,5}B$ conformation is also higher in energy with respect to $^{3}H_4$, due to a steric clash among internal atoms of the ring (Figure V-10).

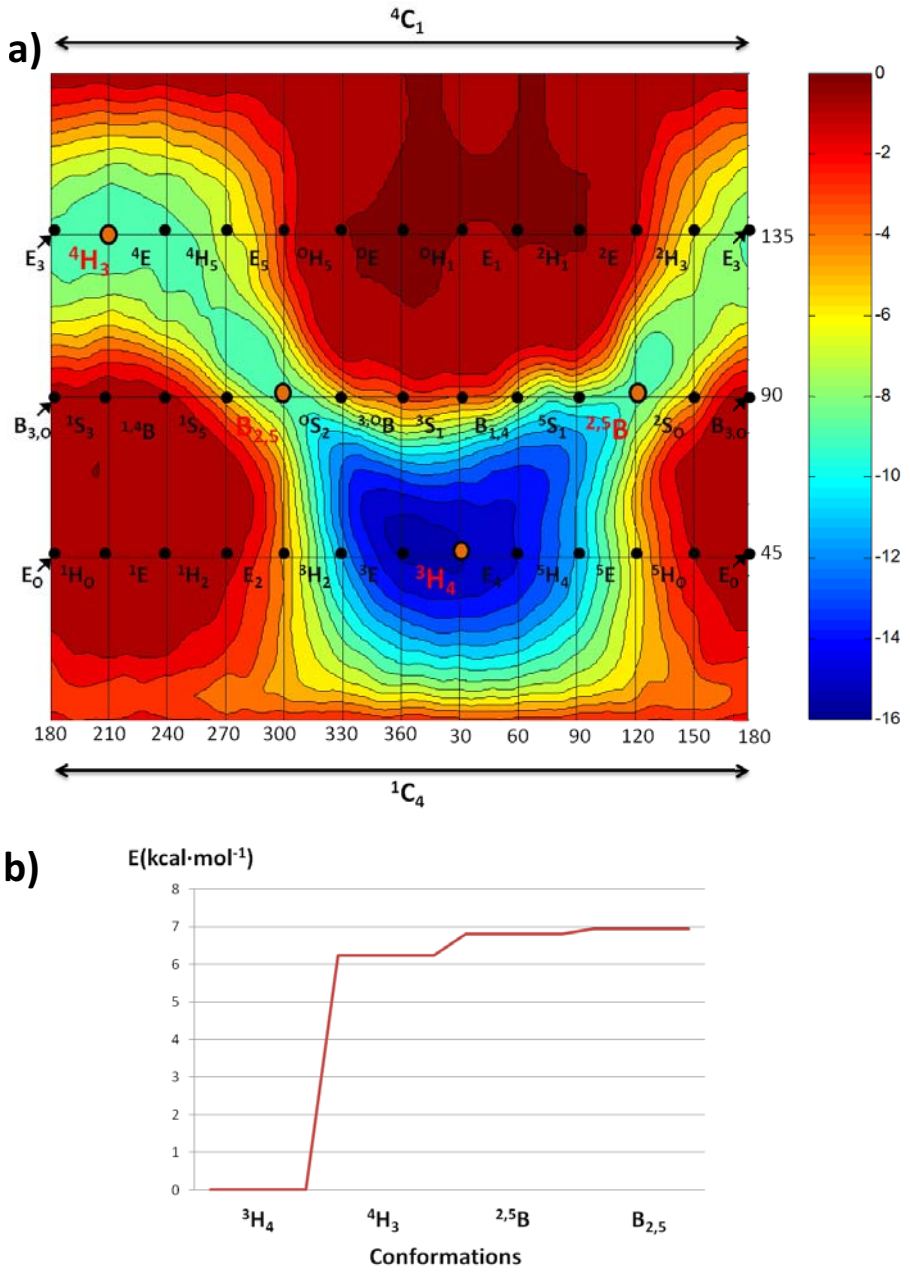

Figure V - 9. a) Conformational free energy landscape of an isolated L-fucosyl cation. Energy values are in kcal·mol^{-1} and each contour line corresponds to 1 kcal·mol^{-1}. b) Free energy of the four TS^{GH} conformations of L-fucosyl cation relative to $^3H_4$.

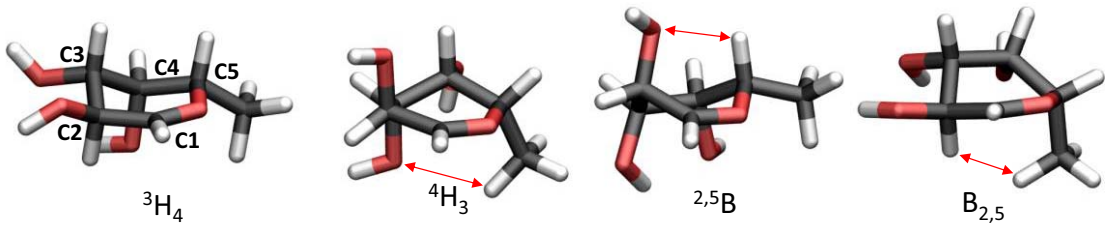

Figure V - 10. Four TS^{GH} conformations for L-fucosyl cation. Red arrows indicate close contacts between atoms.

As a summary of this section, all sugar cations studied present a qualitatively similar conformational FEL, but the relative energies among TS^{GH} conformations differ. The $^{3}H_{4}$ conformation is always the most stable, due to the axial orientation of its C3 and C4 hydroxyl groups, whereas the relative energies of the other conformations depend on the particularities of each sugar. Close contacts between different carbon substituents and hydrogen bond patters also play an important role on the relative energies. It is interesting to mention that the most stable conformation of the FEL does not always agree with the TS predicted experimentally (see D-glucosyl and D-xylosyl cations). Therefore, other factors, in addition to the free energy, must play a role in selecting one TS or another for the hydrolysis reaction in GH enzymes.

4.5 Degree of preactivation of sugar TS conformations.

As demonstrated for α-D-mannose (Chapter IV) and β-D-mannose (Ardèvol, 2010), in addition to the relative energy of each conformation, other relevant properties (structural and electronic) should be taken into account in order to know which conformations are the best pre-activated for catalysis (in the case of the simulations of neutral isolated sugars, the conformation best suited to be present in the MC of the reaction). In this section, we apply the same ideas to elucidate which are the best conformations to act as a TS for each one of the sugar cations studied.

For a given TS^{GH} conformation, the two most relevant structural parameters are the C1-O5 distance (the C1-O5 acquires partial double bond character in the oxocarbenium ion-like TS (Whitfield, 2007)) and the C2-C1-O5-C5 dihedral angle (transition states have a planar geometry in which C5, O5, C1 and C2 atoms adopt a coplanar arrangement (Vocadlo & Davies, 2008)) because these are the ones that change during catalysis. The orientation of the hydroxyl group attached to the C2 carbon atom is also of especial relevance because placing O2 pseudo-equatorially would alleviate cis-1,2 diaxial repulsions at the TS (Taylor, 2005)(Davies, 2012). Additionally, from an electronic point of view, the most relevant parameters are the charge at the anomeric carbon (q_{C1}) and ring oxygen (q_{O5}).

To track how these parameters change within each conformation, a set of approximately 100 structures was selected for each sugar cation and submitted to geometry optimization. The optimized structures were clustered to one of the TS^{GH} conformations ($^{4}H_{3}$, $^{3}H_{4}$, $^{2,5}B$, and $B_{2,5}$) according to their θ and ϕ final values. Distances, angles, dihedral angles and charges were analyzed for all the conformations grouped in each TS^{GH} conformer and the average values were assigned to the corresponding canonical conformation.

4.5.1 Changes on the C1-O5 distance.

Figure V-11 shows the C1-O5 bond distance with respect to ring distortion. This parameter display small variations (~0.01Å) within each one of the sugars studied. As expected, glucosyl and xylosyl cations show an identical behavior, being the experimentally predicted $^{4}H_{3}$ conformation the one with the lower values of the distance. Surprisingly, mannosyl and fucosyl cations also show a similar trend for this parameter, with no correlation with the experimentally observed TS ($^{3}H_{4}$ for L-fucosidases and α-D-mannosidases and $B_{2,5}$ for α-D-mannosidases).

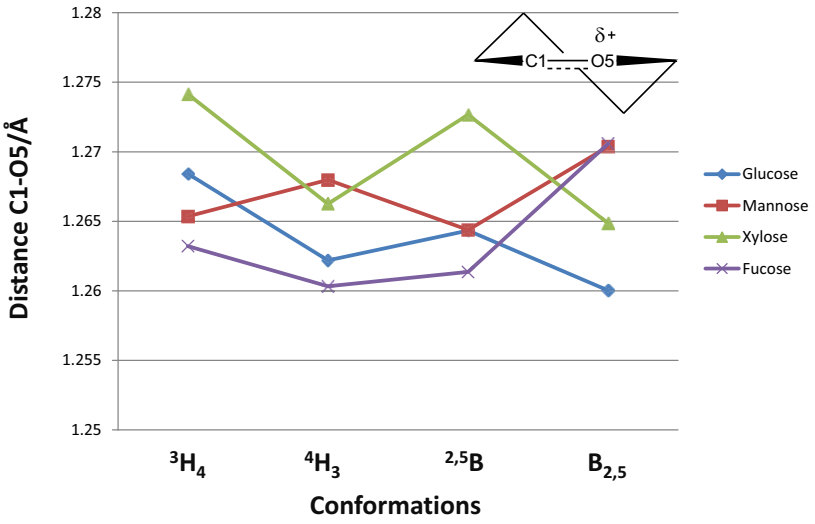

Figure V - 11. C1-O5 distance as a function of ring conformation.

4.5.2 Changes on the C2-C1-O5-C5 dihedral angle.

As observed in Figure V-12a, the C2-C1-O5-C5 dihedral angle ($\phi_{C2\text{-}C1\text{-}O5\text{-}C5}$) deviates from the optimum value of 0° for a perfect, fully planar, oxocarbenium ion. This is a consequence of the different sugar substituents and their intramolecular interactions. As an example, the $^{3}H_4$ D-glucosyl cation conformation (Figure V-12b) has an average value of the dihedral angle of approximately 9° due to the internal hydrogen bond between C2 and C4 hydroxyl groups, which lowers the position of the C2 carbon atom and distorts the dihedral angle from its 0° reference value.

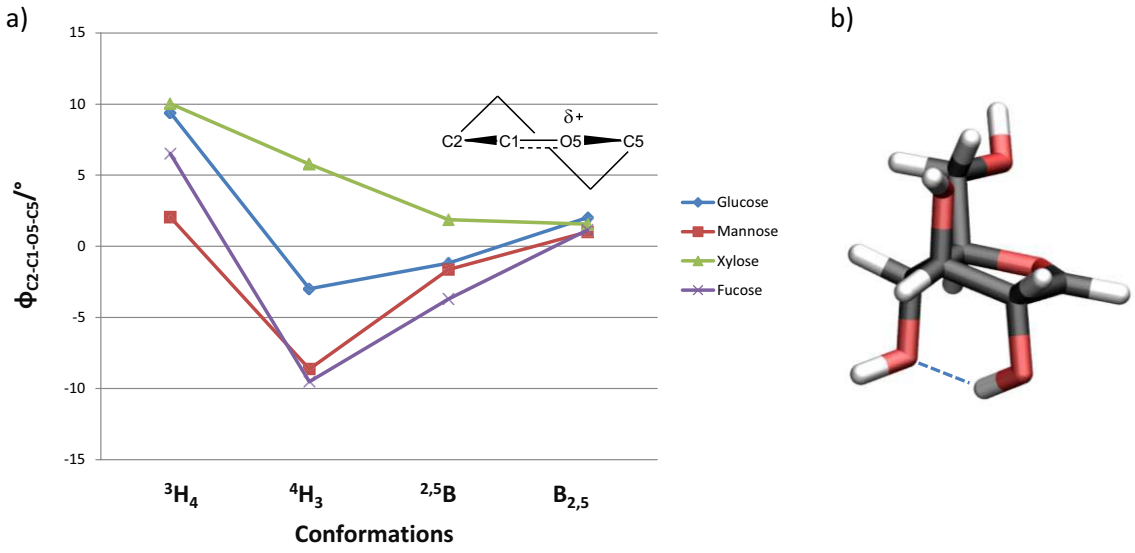

Figure V - 12. C2-C1-O5-C5 dihedral angle ($\phi_{C2\text{-}C1\text{-}O5\text{-}C5}$) value as a function of TS conformation.

4.5.3 Change on the atomic charges.

Positive charge development at the anomeric carbon (C1 atom) is important for the hydrolysis reaction because it favors the nucleophilic substitution at this position (see General Introduction). Variations in the atomic (ESP) charges for the C1 atom (Figure V-13) show a correlation with the TS conformations used by GHs acting on mannose, glucose, xylose and fucose-derived sugars. As shown in Figure V-13, conformations having a larger q_{C1} value are $^{4}H_{3}$ for D-glucose ($^{4}H_{3}$ and $^{2,5}B$ predicted experimentally), $^{3}H_{4}$ and $B_{2,5}$ for D-mannose ($^{3}H_{4}$ and $B_{2,5}$ predicted experimentally), $^{4}H_{3}$ and $^{2,5}B$ for D-xylose ($^{4}H_{3}$ and $^{2,5}B$ found experimentally), and $^{3}H_{4}$ and $B_{2,5}$ for L-fucose ($^{3}H_{4}$ predicted experimentally). Therefore, this data suggest that GHs distort sugars at TS conformations with a high C1 charge development.

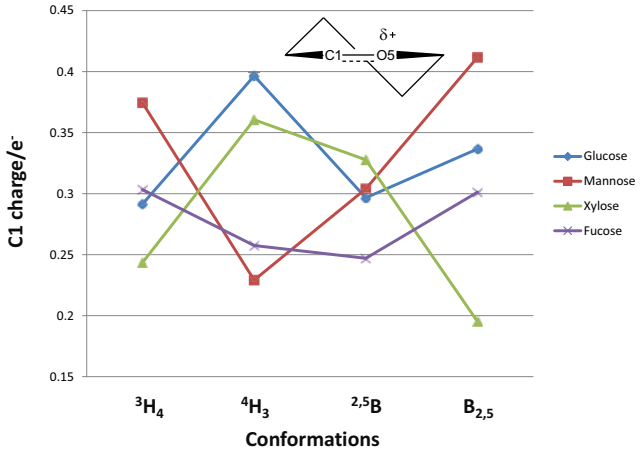

Figure V - 13. Atomic charges on the anomeric carbon (q_{C1}) as a function of the TS conformation.

Concerning the O5 ring oxygen, slight changes are observed on the charge values (q_{O5}). They seem to follow the changes observed at the anomeric carbon position (q_{C1}): higher charge at C1 corresponds to a lower charge at O5 (i. e. more negative value). This indicates a worse electron donation and, therefore, less double bond character. However, comparison of the double bond character with the variations in the C1-O5 distance (Figure V-11) does not show any special relation.

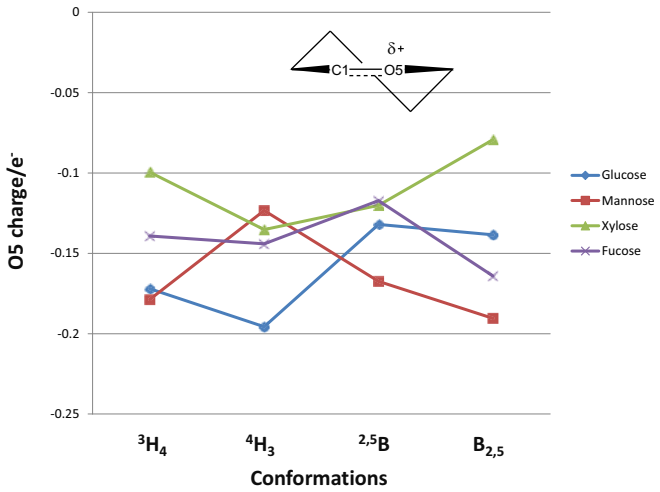

Figure V - 14. Atomic charges on the ring oxygen (q_{O5}) as a function of ring conformation.

4.5.4 Changes on the C2-O2 bond orientation.

The orientation of the C2-O2 bond was analyzed by measuring the angle Ω between this bond and the average plane of the ring (Figure V-15a). Values of the bond orientation near 0° indicate an equatorial orientation, with no cis-1,2 diaxial repulsions, whereas values near 100° or -100° indicate axial configurations, which are less favored for catalysis. As shown in Figure V-15b, for sugars with a β configuration of the C2-O2 bond (D-mannosyl and L-fucosyl cations) the conformations $^{3}H_{4}$ and $B_{2,5}$ present an equatorial orientation, which are the ones experimentally predicted as TS conformations for the hydrolysis of L-fucosidases and D-mannosidases. For sugars with an α configuration of the bond (D-glucosyl and xylosyl cations), the conformations with an equatorial configuration of the bond ($^{4}H_{3}$ and $^{2,5}B$) also agree with the ones predicted as a TS for the hydrolysis reaction. Therefore, an equatorial configuration of the C2-O2 bond seems to be an important factor determining the TS of the reaction.

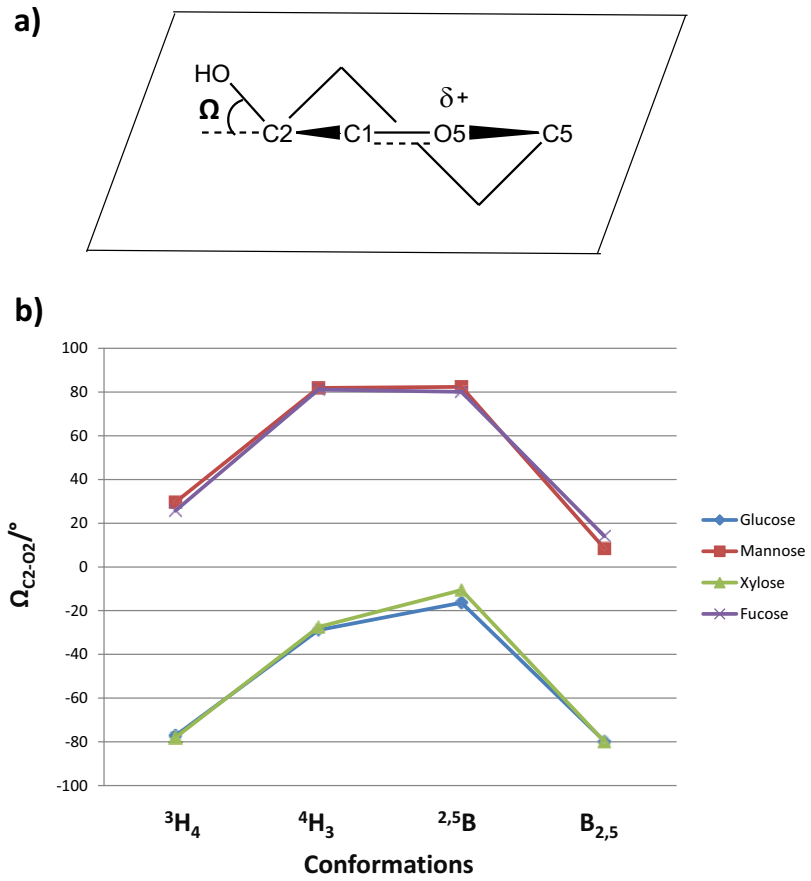

Figure V - 15. Orientation of the bond between the C2 carbon atom and its exocyclic oxygen (O2), as measured by the angle Ω between the C2-O2 bond and the average plane of the ring, plotted as a function of the TS conformation.

4.5.5 Calculation of the preactivation index.

Following the same procedure applied for α-mannose (Chapter IV), we combined all parameters (q_{C1}, q_{O5}, $d_{C1\text{-}O5}$, $\Omega_{C2\text{-}O2}$ and $\Phi_{C2\text{-}C1\text{-}O5\text{-}C5}$) along with the relative free energy (ΔG_{rel}) into a unique index that would reflect the likelihood that a given conformation would be adopted as the TS of the hydrolysis reaction for a given sugar. This was done by assigning for each conformation *j* a score for each parameter x_i using the following formulas:

$$\text{score}\left(x_{i,j}\right) = \frac{x_{i,j} - x_{i,j}^{\min}}{x_{i,j}^{\max} - x_{i,j}^{\min}} \times 100 \quad \text{for } x_i = q_{C1} \qquad \textbf{Eq. V - 1}$$

$$\text{score}\left(x_{i,j}\right) = \frac{x_{i,j}^{\max} - x_{i,j}}{x_{i,j}^{\max} - x_{i,j}^{\min}} \times 100 \quad \text{for } x_i = d_{C1\text{-}O5}, q_{O5}, \Delta G_{rel}, \left|\Phi_{C2-C1-O5-C5}\right|, \left|\Omega_{C2-O2}\right| \qquad \textbf{Eq. V - 2}$$

The values of the parameters and the corresponding scores are given in Table V-1. Since the score for each parameter is normalized, the scores can be directly compared. We then defined an index ξ_j as the average of the scores for the n parameters (n = 6 in our case) for a given conformation j:

$$\xi_j = \sum_i \text{score}\left(x_{i,j}\right) / n \qquad \textbf{Eq. V - 3}$$

Figure V-16 shows the variation of ξ for the four sugar cations studied. There is no single conformation with the optimum values for every parameter (ξ = 100) but, in perfect agreement with experimental data, the geometries with higher ξ values correspond with the TS conformations used by GHs. For D-glucosyl and xylosyl cations, conformations $^4H_3$ and $^{2,5}B$ (Figure V-16) show the higher ξ values. This agrees with the experimental data available (two main itineraries have been predicted for D-glucosidases and xylanases: one that involves a $^4H_3$ TS conformation and another one that passes through a $^{2,5}B$ conformation). For D-mannose and L-fucose, the $^3H_4$ conformation is the one with the higher ξ value in agreement with the TS used by D-mannosidases and L-fucosidases. For these two sugar ions, the $B_{2,5}$ conformation exhibits the second highest ξ value, which agrees with the other catalytic itinerary predicted for D-mannosidases. For L-fucosidases the $B_{2,5}$ conformation was not predicted as TS for the hydrolysis reaction and, as a consequence, its ξ value is lower than the one observed for D-mannosidases.

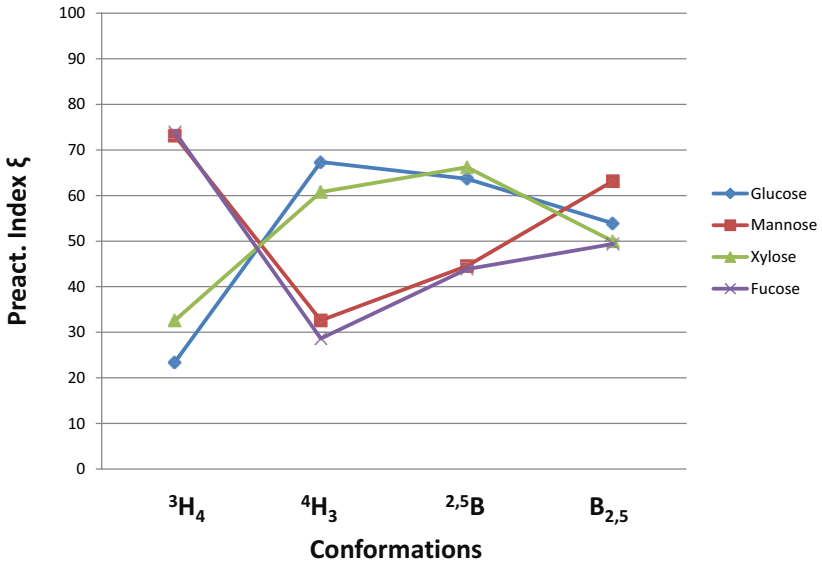

Figure V - 16. Variation of the preactivation index ξ as a function of the TS ring conformation for D-glucosyl, D-mannosyl, xylosyl and L-fucosyl cations.

Table V - 1. Electronic, structural and energetic parameters defining the degree of preactivation of each TS^{GH} conformation for the D-glucosyl, D-mannosyl, xylosyl and L-fucosyl cations: q_{C1} and q_{O5} ESP charges, C1-O5 distance, orientation of the C2-O2 bond (Ω angle), planarity of the ring ($\phi_{C2\text{-}C1\text{-}O5\text{-}C5}$) and relative free energy (ΔG_{rel}). For each conformation, the rate for a given parameter (equations V-1 and V-2) appears below the value of the corresponding parameter in *italics*. The average preactivation index (ξ), computed from equation V-3, is given in the last column. Distances are given in Angstroms, charges in electrons, angles in degrees and energies in kcal·mol^{-1}.

	conformer	q_{C1}	q_{O5}	$d_{C1\text{-}O5}$	$\Omega_{C2\text{-}O2}$	$\phi_{C2\text{-}C1\text{-}O5\text{-}C5}$	ΔGrel	ξ
D-glucosyl cation	$^{3}H_{4}$	0.292 *0*	-0.172 *37*	1.268 *0*	-77.150 *4*	9.406 *0*	0.0 *100*	**23.4**
	$^{4}H_{3}$	0.396 *100*	-0.196 *0*	1.262 *74*	-28.886 *81*	-2.993 *78*	1.2 *71*	**67.3**
	$^{2,5}B$	0.297 *5*	-0.132 *100*	1.264 *49*	-16.354 *100*	-1.183 *100*	2.8 *28*	**63.6**
	$B_{2,5}$	0.337 *43*	-0.138 *91*	1.260 *100*	-79.656 *0*	2.018 *90*	4.0 *0*	**53.9**
D-mannosyl cation	$^{3}H_{4}$	0.374 *80*	-0.179 *17*	1.265 *84*	29.587 *71*	2.043 *86*	0.0 *100*	**73.1**
	$^{4}H_{3}$	0.229 *0*	-0.123 *100*	1.268 *40*	81.872 *1*	-8.627 *0*	3.4 *55*	**32.6**
	$^{2,5}B$	0.304 41	-0.167 34	1.264 100	82.303 0	-1.631 92	7.6 *0*	**44.5**
	$B_{2,5}$	0.412 *100*	-0.190 *0*	1.270 *0*	8.527 *100*	1.010 *100*	1.6 *79*	**63.2**
xylosyl cation	$^{3}H_{4}$	0.244 *29*	-0.100 *64*	1.274 *0*	-77.987 *3*	10.033 *0*	0.0 *100*	**32.6**
	$^{4}H_{3}$	0.361 *100*	-0.135 *0*	1.266 *85*	-27.480 *76*	5.772 *50*	2.9 *54*	**60.7**
	$^{2,5}B$	0.328 *80*	-0.120 *27*	1.273 *16*	-10.587 *100*	1.870 *96*	1.4 *78*	**66.2**
	$B_{2,5}$	0.195 *0*	-0.079 *100*	1.265 *100*	-79.886 *0*	1.554 *100*	6.2 *0*	**50.0**
L-fucosyl cation	$^{3}H_{4}$	0.303 100	-0.139 53	1.263 72	25.721 83	6.513 36	0.0 *100*	**73.9**
	$^{4}H_{3}$	0.258 19	-0.144 43	1.260 100	81.129 0	-9.519 0	6.2 *10*	**28.6**
	$^{2,5}B$	0.247 0	-0.117 100	1.261 90	80.005 2	-3.709 69	6.8 *2*	**43.8**
	$B_{2,5}$	0.301 96	-0.164 0	1.271 0	14.100 100	1.153 100	7.0 *0*	**49.4**

In summary, our results suggest that GH enzymes select those transition states and, therefore, catalytic itineraries that are most suitable for catalysis in terms of energies and structural and electronic parameters. The question regarding why enzymes acting on one type of substrate (e. g. α-D-mannosides) perform catalysis using two different TS (or catalytic itineraries) needs further investigation, but probably reflects different evolutionary restraints on the different types of GH enzymes.

4.6 Transition state mimic inhibitors.

In the previous chapter we have shown that mannoimidazole compounds are good TS mimic inhibitors because all relevant TS^{GH} conformations are energetically accessible. Therefore, the enzyme is able to modulate the conformation of this inhibitor and to adapt it to its TS. On the contrary, the conformational FEL of isofagomine is strongly biased towards a $^4C_1$ conformation and, as a consequence, it is more difficult for the enzyme to modify it.

On the context of this chapter, it is interesting to compare the FELs of mannoimidazole and D-mannosyl cation to shed more light on the reasons for its TS mimicry. Although this is beyond the scope of this Thesis, design of novel molecules with energy landscapes close to the one of a pure transition state could be an interesting approach to improve inhibitor's potency.

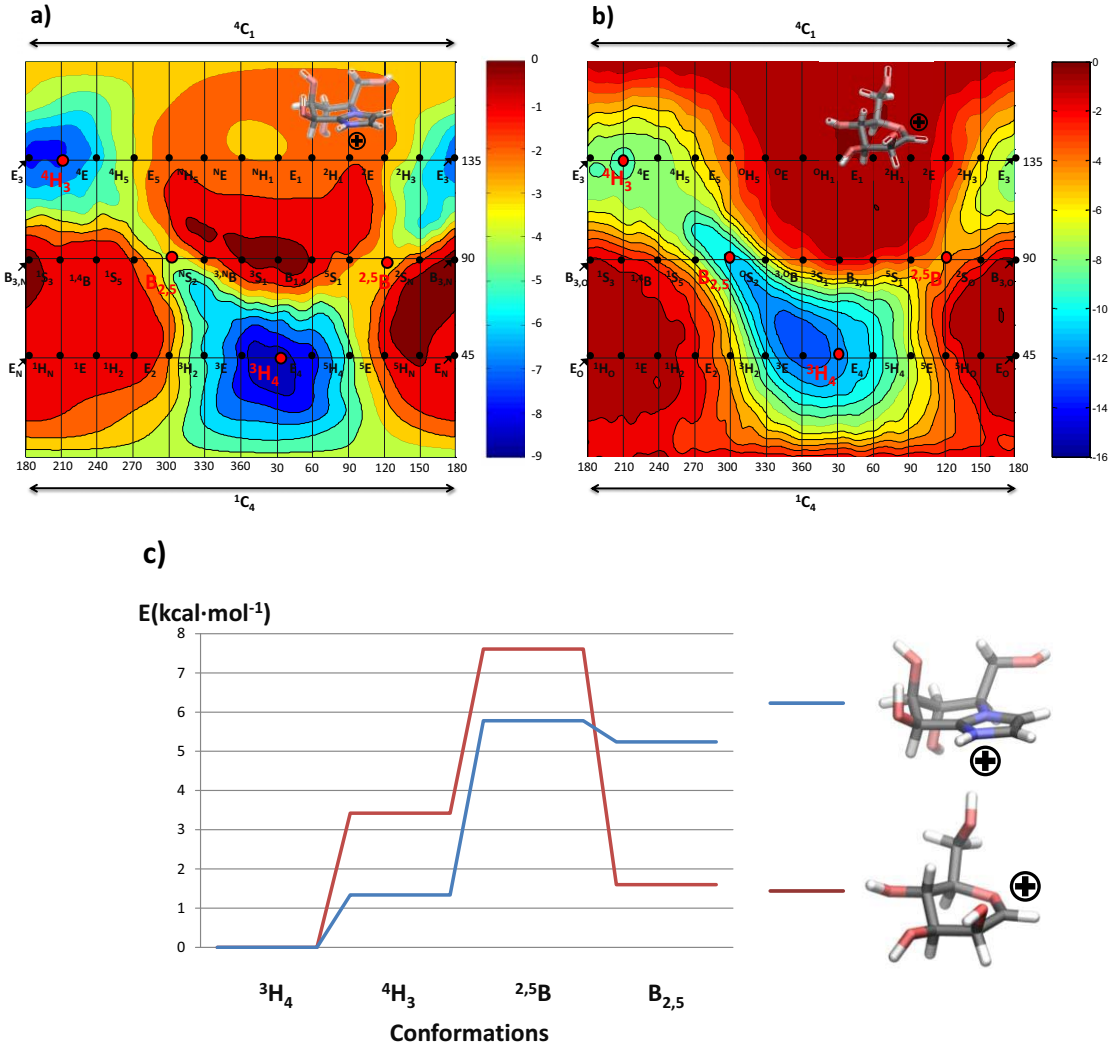

Figure V - 17. Distribution of conformations on the computed free energy surface of mannoimidazole (a, Chapter IV) and D-mannosyl cation (b). c) Free energy of the different TS^{GH} conformations for mannoimidazole (blue line) and D-mannosyl cation (red line) relative to the $^3H_4$ conformation.

Figure V-17a shows the FEL of mannoimidazole presented in the previous chapter and the corresponding energy landscape for the D-mannosyl cation (Figure V-17b), discussed above. It is clear that the two landscapes are very similar, with identical number of accessible conformations. Although the $^3H_4$ conformation is the global energy minima for both molecules, relative energies between conformations (Figure V-17c) differ. For the mannoimidazole

inhibitor, both $^{4}H_{3}$ and $^{2,5}B$ conformations are almost 2 kcal·mol^{-1} more stable than the same conformations in the mannosyl cation, indicating a higher degree of interconversion of both half conformations. The reason for this does not lie on the properties of these two conformations, but in the fact that the $^{3}H_{4}$ conformation (reference value for the energies) is less stable for mannoimidazole than for the D-mannosyl cation. Mannoimidazole has a positive charge delocalized over the imidazole functionality, which is far from the axial hydroxyl groups that have an stabilization effect (Figure V-18). Therefore, as the positive charge is less stabilized by the axially oriented hydroxyl groups in $^{3}H_{4}$, the conformation becomes more unstable compared to the D-mannose case. An identical explanation applies to the higher free energy of the $B_{2,5}$ conformation.

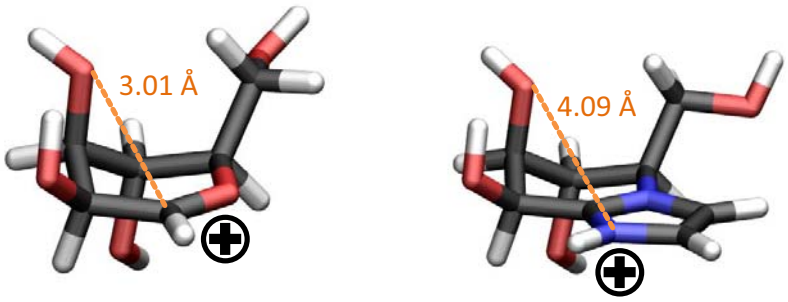

Figure V - 18. $^{3}H_{4}$ structures for D-mannose oxocarbenium ion and mannoimidazole molecules. Dotted lines indicate the distance between the O3 hydroxyl group and the atom which has the positive charge. For mannoimidazole, positive charge is delocalized over the imidazole ring, therefore, assignation of an individual atom must be taken with caution.

Although the capabilities of mannoimidazole inhibitors to inform on TS conformations have been demonstrated, some inconsistent results have also been found (Williams, 2014). The $^{4}H_{3}$ conformation reported for the complex of mannoimidazole and *Drosophila melanogaster* Golgi GH38 α-mannosidase II (Kuntz, 2008) is inconsistent with a $^{1}S_{5}$ glycosyl-enzyme intermediate (Numao, 2003) and with the TS of the reaction computed by CPMD/MM calculations (Petersen, 2010). The fact that $^{3}H_{4}$ and $^{4}H_{3}$ conformations are almost isoenergetic for mannoimidazole could be the reason for this anomalous result. Therefore, mannosidase inhibitors with a high energy difference between these two conformations are expected to be better TS mimics than inhibitors in which both $^{3}H_{4}$ and $^{4}H_{3}$ have a similar energy. First, the equilibrium will be more biased towards a $^{3}H_{4}$ conformation and, as a consequence, the enzyme will invest less energy to modify alternative conformations. Secondly, the higher stability of the $^{3}H_{4}$ conformation will avoid trapping the inhibitor in anomalous conformations, such as the one found for *Drosophila melanogaster* Golgi GH38 α-mannosidase II.

5. Conclusions.

The main conclusions of the present chapter are:

1. By means of *ab initio* metadynamics simulations, we have demonstrated that the conformational FEL of four isolated sugar oxocarbenium ions (D-glucosyl, D-mannosyl, xylosyl and L-fucosyl cations) are qualitatively similar, although relative energies between TS^{GH} conformations differ.

2. The $^{3}H_{4}$ conformation is the most stable conformation for the four sugar oxocarbeniums studied, due to the axial orientation of the C3 and C4 hydroxyl groups. For D-glucose and xylose, this is in contradiction with the experimentally predicted $^{4}H_{3}$ and $^{2,5}B$ transition states, whereas the results for D-mannose and L-fucose agree with the TS conformations predicted experimentally for D-mannosidases and L-fucosidases.

3. The preactivation index ξ, which includes electronic, structural, and energetic parameters, allows to predict the most suitable TS conformations for the hydrolysis reaction catalyzed by GHs.

4. GHs transition state mimic inhibitors have a conformational energy landscape that resembles the one of a pure sugar cation. This explains the ability of these molecules to adapt to the TS of the hydrolysis reaction.

Chapter VI – Observation of a S_Ni-Like mechanism in an engineered retaining glycoside hydrolase.

Observation of a S_Ni-Like mechanism in an engineered retaining glycoside hydrolase.

1. Introduction.

β-Hexosaminidases catalyze the hydrolysis of N-acetyl-β-hexosaminides (N-acetylgalactosamine, GalNAc and N-acetylglucosamine, GlcNAc) from the non-reducing end of glycoconjugates and oligo or polysaccharides. For retaining enzymes following a double displacement mechanism, the sugar in the -I subsite deforms into a distorted conformation, allowing the 2-acetamido oxygen to attack the anomeric C1 carbon and releasing the +I subsite sugar molecule (substrate asisted catalysis; Figure VI-1)(Vocadlo, 2008)(He, 2009).

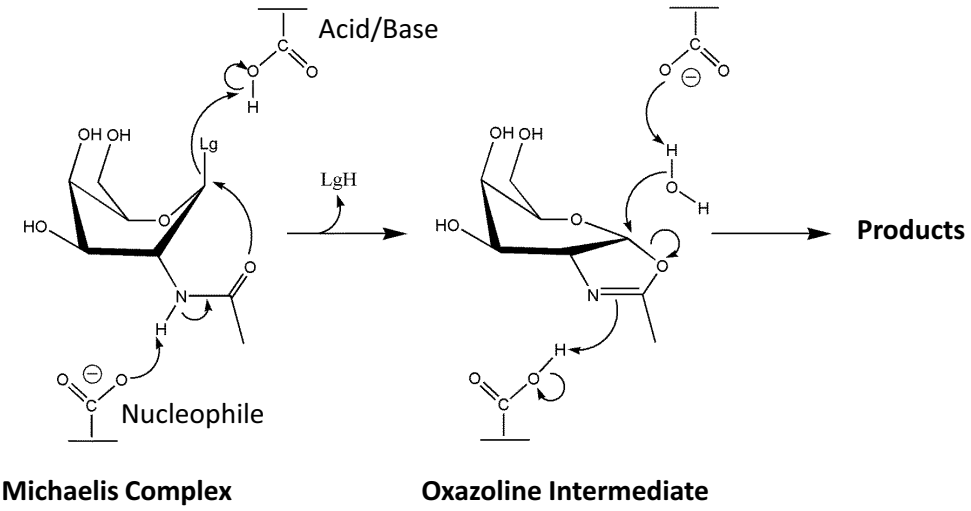

Figure VI - 1. The catalytic mechanism of retaining β-hexosaminidases proceeds via a two-step reaction mechanism through a distorted Michaelis complex and an oxazoline intermediate. In the first step of the reaction, the general acid residue aids to the departure of the leaving group while a general base (nucleophile) activates the 2-acetamido group for nucleophilic attack. In the second step, the general acid (now acting as a base) increases the nucleophile character of a water molecule, pointed to attack the anomeric center, while the other catalytic residue aids opening of the oxazoline ring to form the products with retained stereochemistry at the anomeric carbon.

The *Sulfolobus solfataricus* β-glycosidase (SsβG) is a family 1 GH that operates via the typical double displacement mechanism (see General Introduction; Section 3.1), forming a glycosyl-enzyme covalent intermediate. The enzyme has a broad specificity, being able to hydrolyze both gluco and galacto substrates with a $^{1}S_{3} \rightarrow [^{4}H_{3}]^{\ddagger} \rightarrow {}^{4}C_{1}$ catalytic itinerary for the glycosilation step (Gloster, 2004).

In an attempt to create a hexosaminidase type of an enzyme utilizing a substrate-assited catalysis from a classical glycoside hydrolase, the experimental group of Prof. Ben Davis from the University of Oxford mutated the catalytic nucleophile (E387) of SsβG to a tyrosine residue, in the hope that it may stabilise the formation of an oxazolinium ion intermediate. However, they observed that this mutant SsβG E387Y did not show hexosaminidase activity even with an activated substrate, such as an oxazoline. However and very interestingly, the E387Y mutant gave an enzyme with highly specific transglycosylation activity.

Transglycosylation reactions conducted with *p*-nitrophenyl β-D-galactopyranoside (pNPβGal) as the donor (Figure VI-3b) and a variety of monosaccharides as acceptors mostly resulted in the donor reacting with itself to give 1,3 and 1,6-linked products (Figure VI-2a, b). Additionally, lower yields of disaccharide were isolated with phenyl β-D-glucopyranoside acceptor molecules (Figure VI-2c, d). Therefore, this mutant enzyme was stereospecific; interestingly only creating β-glycosidic linkages from activated β-donors such as *p*-nitrophenyl glycosides, and it could not hydrolyze the glycosidic linkages in the products due to the absence of the nucleophile.

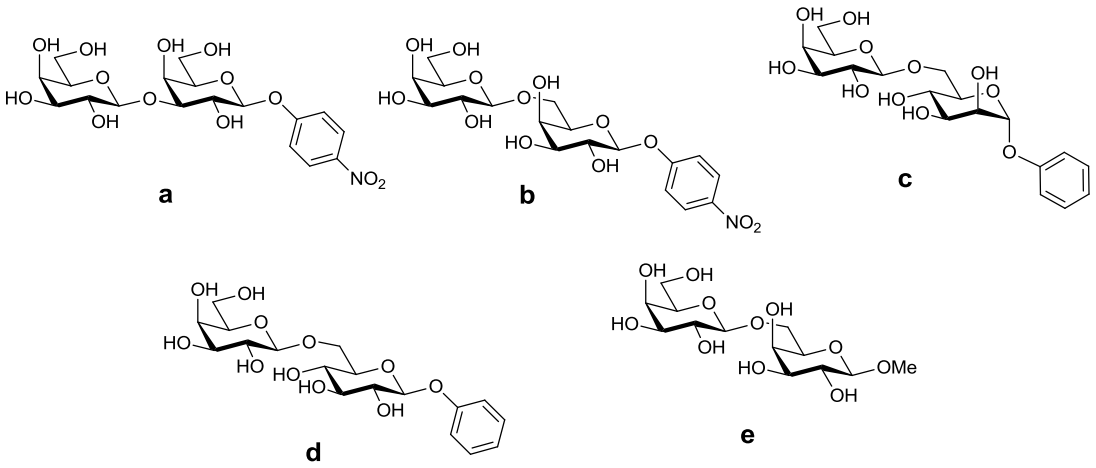

Figure VI - 2. Products of the reaction of *p*-nitrophenyl β-D-galactopyranoside with various acceptors catalysed by E387Y SsβG. a) *p*-nitrophenyl 3-*O*-(β-D-galactopyranosyl)-β-D-galactopyranoside b) *p*-nitrophenyl 6-*O*-(β-D-galactopyranosyl)-β-D-galactopyranoside c) phenyl 6-*O*-(β-D-galactopyranosyl)-α-D-mannopyranoside d) phenyl 6-*O*-(β-D-galactopyranosyl)-β-D-glucopyranoside e) methyl 6-*O*-(β-D-galactopyranosyl)-β-D-galactopyranoside. Results obtained in the group of Prof. Ben Davis (University of Oxford).

Thorough investigations including kinetic, biochemical, mutagenic and structural studies (data not published) led the experimentalist to hypothesize that this mutant enzyme utilizes a front-face nucleophilic substitution, often named as S_Ni or S_Ni-like in the literature, to perform the transglycosylation reaction. This type of mechanism was first proposed for the solvolysis of D-glucopyranosyl derivatives (Sinnott, 1980) and it was recently demonstranted, by kinetic isotope effects (Lee, 2011) and a CPMD/MM metadynamics study (Ardèvol, 2011), that it is operative for retaining glycosyl transferases (Figure VI-3a).

Crucial features of this mechanism revealed in these previous studies are as follows (Figure VI-3a): Firstly, the reaction proceeds through a highly dissociative transition state (TS), with a considerable oxocarbenium ion character, or there is a short-lived oxocarbenium ion intermediate (Ardèvol, 2011). This feature is important since it can provide sufficient space to accommodate the nucleophile and the leaving group on the same face. Secondly, the leaving group acts as a general base, deprotonating the incoming nucleophile when it is a hydroxyl group. For retaining glycosyl transferases, the nucleoside diphosphate leaving group (i. e. uridine diphosphate) acts as a general base (Figure VI-3a). For the SsβG E387Y enzyme it is supposed that the donor leaving group (*p*-nitrophenil) is the general base (Figure VI-3b). Thus, a hydrogen bonding at the TS between the leaving group oxygen and the incoming hydroxyl group should be present, which also plays the role of guiding the nucleophile to the same face as the leaving group (Errey, 2010)(Lee, 2011).

In this chapter, we perform an *ab initio* CPMD/MM (Laio, 2002a) metadynamics study of the reaction mechanism of SsβG E387Y.

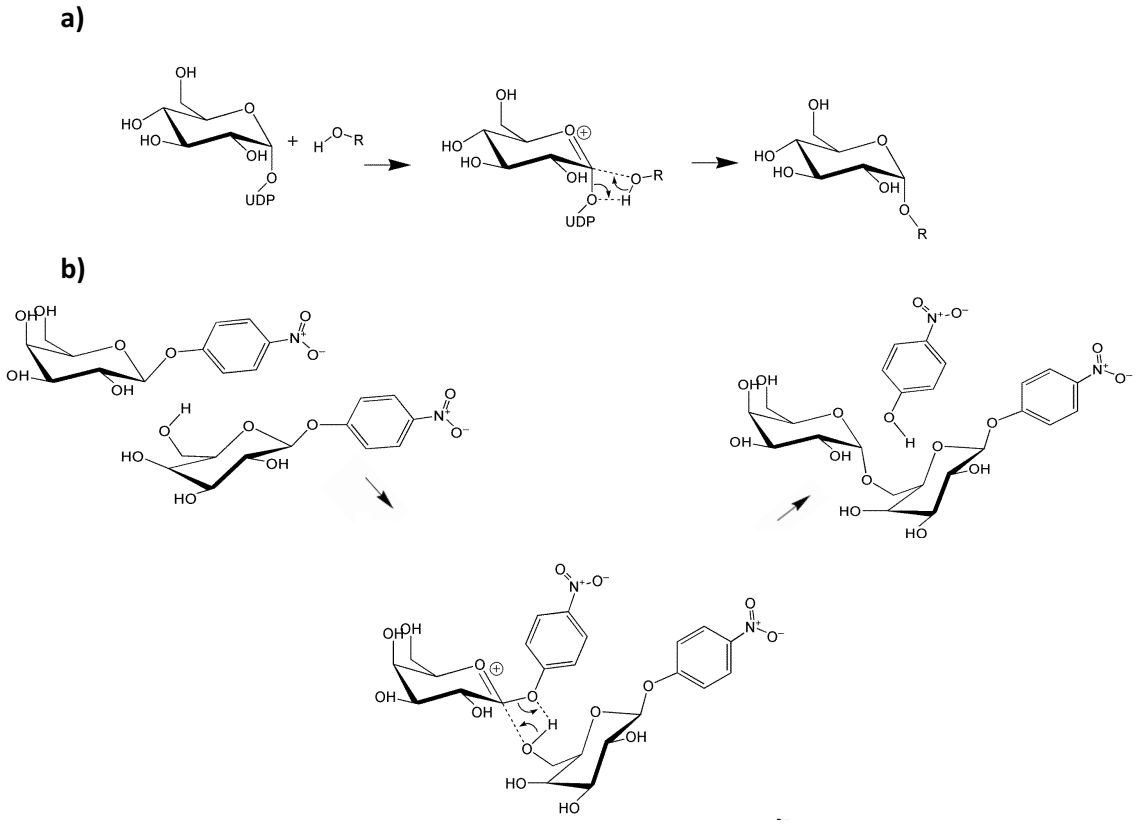

Figure VI - 3. a) Front-face mechanism demonstrated for retaining GTs. b) Proposed front-face mechanism for the SsβG E387Y enzyme. The middle species can be either a transition state or a very short-lived oxocarbenium ion intermediate.

2. Computational details.

2.1 Initial structure.

The initial protein structure for the simulations was taken from the X-ray crystal structure of the E387Y SsβG mutant (data not published). Two pNPβGal molecules were placed manually near the catalytic groove of the enzyme and flexible docking of the ligands was performed with a combination of classical MD and metadynamics (Laio & Parrinello, 2002)(see next sections). Protonation states and hydrogen atom positions of all ionizable amino acids residues were selected base on their hydrogen bond environment. Eleven histidine residues were modeled in their neutral states, while two were modeled in their protonated states. All the crystallographic water molecules were retained and extra water molecules were added to form a 10 Å water box around the protein surface. Three sodium ions were also added to neutralize the enzyme charge.

2.2 Classical molecular dynamics simulations.

MD simulations of the enzyme (E387Y SsβG mutant) in complex with two molecules of pNPβGal were performed with the Amber10 software package (Pearlman, 1995). The protein was modeled with the FF99SB force field (Cornell, 1995), whereas all carbohydrate molecules were modeled with the GLYCAM06 force field (Kirschner, 2008). The partial charges for the aglycon leaving group were obtained using a RESP fit to a HF/6-31G* calculation, which was performed with the Gaussinan03 software package (Frisch, 2004). The MD simulations were carried out in several steps. First, the system was minimized, maintaining the protein and both substrate molecules fixed. In a second step, the entire system was allowed to relax. Weak spatial constraints were initially added to the protein and substrates to gradually reach the desired temperature of 300 K, while the rest of the system was allowed to move freely. The constraints were subsequently removed and the system was subjected to 100 ps of constant pressure MD simulation to adjust the density of the water environment. Afterwards, 200 ns of constant volume MD simulation were performed. During this time, one of the two pNPβGal molecules partially entered the catalytic groove, whereas the other one remained at its entrance. In particular, one pNPβGal molecule accommodated near the -I subsite (at ~8 Å from the catalytic residues) and the second one remained far from it (more than 20 Å). As complete entrance of the two molecules in the active site did not occur during the 200 ns time-scale window, it was subsequently activated using metadynamics (Laio & Parrinello, 2002)(Barducci, 2011).

2.3 Classical metadynamics simulations of substrate binding.

A snapshot of the equilibrium MD simulation was taken for the metadynamics simulations, which were performed with NAMD2.9 software (Phillips, 2005). Two collective variables (CVs) were chosen (Figure VI-4). The first one (CV_1) was taken as the distance between the center of mass of the galactose ring of the inner pNPβGal molecule and the center of mass of a few residues defining the -I enzyme subsite (Tyr387, Trp425 and His150, as identified in structural studies of WT SsβG in complex with inhibitors). Thus, this CV measures the degree of penetration of the first pNPβGal molecule into the active site. This pNPβGal molecule will act as the *donor* of the transglycosylation reaction. The second collective variable (CV_2) was taken as the distance between the glycosidic oxygen (O1) of the donor and the terminal hydrogen atom (H') of the hydroxymethyl substituent of the outer pNPβGal molecule (the *acceptor* of the transglycosylation reaction) (Figure VI-4). Therefore, CV_2 accounts for the formation of an O1···H' interaction or, in other words, it measures the distance between the *donor* and *acceptor* molecules.

The values of the height and width of the Gaussian-like biasing potential were selected as 1.0 kcal·mol^{-1} and 0.15 Å, respectively. A new Gaussian-like potential was added every 2 ps. Reflective walls were placed at proper distances (15 Å in CV_1 and 25 Å in CV_2) to avoid excessive sampling of configurations with the ligands out of the enzyme, in a fully solvated environment, with loss of the substrate-protein interactions. The simulation was continued until the system completely explored the free energy landscape (FEL) which, in terms of the simulation time, corresponds to approximately 100 ns.

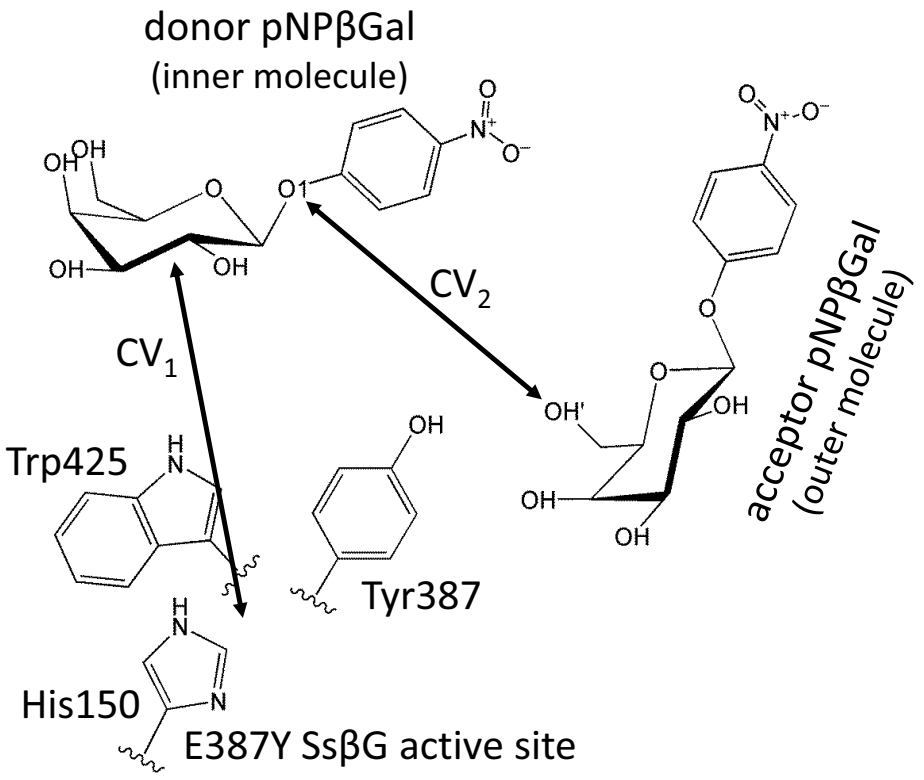

Figure VI - 4. Collective variables used during the metadynamics simulation. CV_1 was taken as the distance between the center of mass of the Gal ring of the donor pNPβGal molecule and the center of mass of a few residues defining the –I enzyme subsite (Tyr387, Trp425 and His150). CV_2 was taken as the distance between the glycosidic oxygen (O1) of the donor and the terminal hydrogen atom (H') of the hydroxymethyl substituent of the acceptor pNPβGal molecule.

2.4 CPMD/MM metadynamics simulations.

To further understand the transglycosylation process, *ab initio* methods were used to model the front-face transglycosylation reaction. CPMD/MM calculations were performed using the method developed by Laio *et al*., (Laio, 2002a) which combines Car−Parrinello molecular dynamics (Car & Parrinello, 1985) based on DFT, with force field MD methodology. This methodology has been used with satisfactory results in the description of energetic, dynamic, and structural features of biological systems, including GHs (Carloni, 2002)(Rovira, 2013).

The QM region was chosen to include both *p*-nitrophenil glycosides (72 atoms) and was enclosed in an isolated 18 × 18.5 × 18.5 $Å^3$ supercell. Kohn−Sham orbitals were expanded in a PW basis set with a kinetic energy cutoff of 70 Ry. Troullier-Martins *ab-initio* pseudopotentials with dispersion-corrected atom-centered potentials (DCACPs)(Lin, 2007) were used for all elements. The PBE functional in the generalized gradient-corrected approximation of DFT was used. As mentined in other chapters of this Thesis, this fuctional had been already used in previous works on glicosyl hydrolases (Biarnes, 2011) and transferases (Ardèvol, 2011)(Rojas-Cervellera, 2013) performed in the group. Structural optimizations were done by MD with annealing of the ionic velocities until the maximal component of the nuclear gradient was < $5 \cdot 10^{-4}$ a. u. A constant temperature of 300 K was reached by coupling the system to a Nosé–

Hoover thermostat (Nosé, 1984). A time step of 0.12 fs and a fictitious electron mass of 650 a.u. were used in the Car–Parrinello simulations.

The metadynamics algorithm (Laio & Parrinello, 2002) was used to overcome energy barriers and to reconstruct the FEL associated with the transglycosylation reaction. Although the most intuitive way to model an enzymatic reaction is by starting from the Michaelis complex (MC)(obtained previously from a classical metadynamics simulation; section 2.3) and going to the products of the reaction, all our attempts (using different sets of CVs, metadynamics parameters, etc.) following this procedure were unsuccessful. Therefore, as an alternative approach, we decided to model the enzymatic reaction starting from the products of the reaction. The structure was obtained by steered molecular dynamics (Weinan, 2002) coupled to the CPMD/MM algorithm (pulling four CVs: distances C1-OpNP, C1-O6', OpNP-H and O6'-H) and equilibrated with 10 ps of unbiased simulation. Starting from this products structure, three CVs, corresponding with the main bonds undergoing breaking or formation, were used (See Figure VI-5). The first collective variable (CV_1) was taken as the distance between the donor anomeric carbon (C1) and the O6 atom. Therefore, this CV_1 measures the degree of formation of the glycosidic bond. The second collective variable (CV_2), taken as the distance between C1 and the oxygen atom of the leaving group (O_{PNP}), measures the degree leaving group departure. Finally, the third collective variable (CV_3) was taken as the difference of coordination number (CN) between the bonds involved in the transfer of the hydrogen atom ($CV_3 = CN_{OPNP\text{-}H} - CN_{O6'\text{-}H}$). Chosen in this way, the CV_3 take negative values when the hydrogen atom is bound to the O6' of the acceptor molecule and positive values when it is bound to the oxygen atom of the leaving group (Figure VI-5). The CN is given by the following formula:

$$\boldsymbol{CN_{ij}} = \frac{\mathbf{1}-\left(\frac{\boldsymbol{d_{ij}}}{\boldsymbol{d_0}}\right)^{\boldsymbol{p}}}{\mathbf{1}-\left(\frac{\boldsymbol{d_{ij}}}{\boldsymbol{d_0}}\right)^{\boldsymbol{p+q}}} \qquad \textbf{Eq. VI - 1}$$

where d_{ij} is the internuclear distance of the atoms involved, d^0 is the threshold distance for bonding, and p and q are exponents that determine the steepness of CN_{ij} decay with respect to d_{ij}. CN values range from 0 (not bonded) to 1 (bonded). The p, q and d^0 values were chosen, by plotting with different sets of values, to effectively separate all the important species along the process under study. The selected values for the CN parameters were d^0 = 2.50 a. u., p = 12, and q = 4.

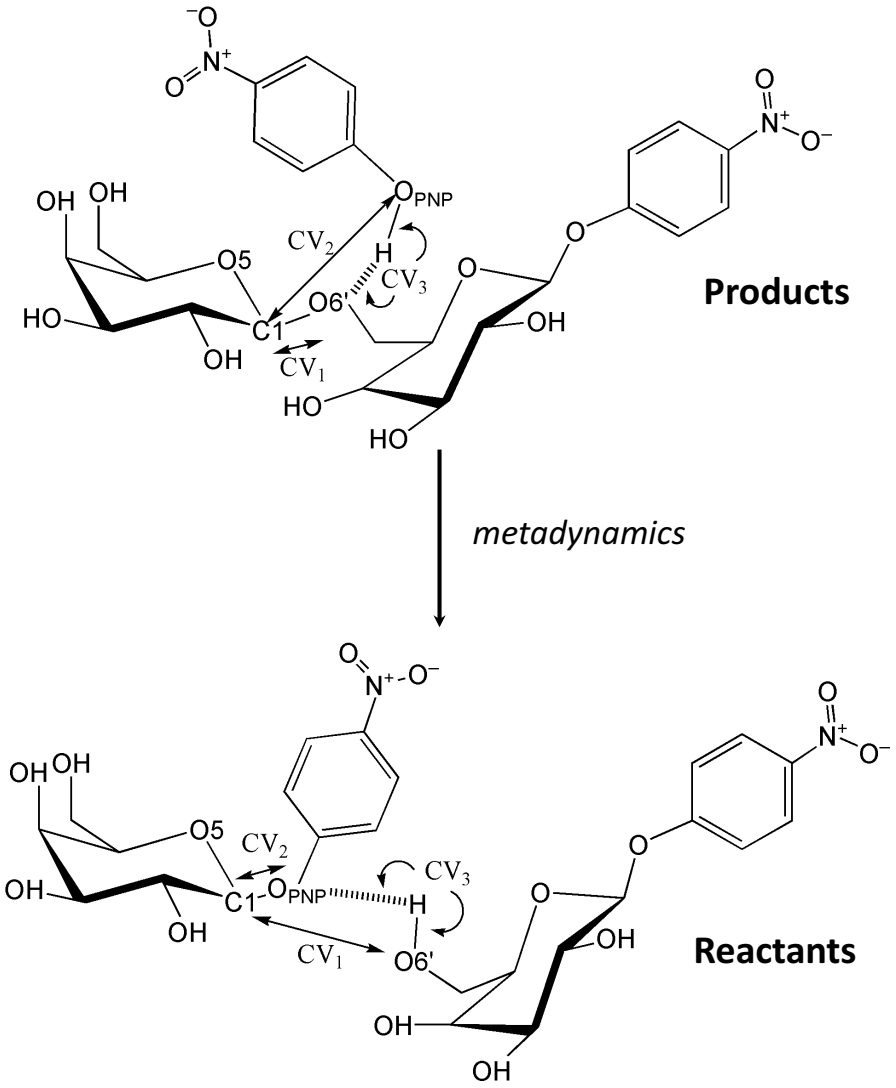

Figure VI - 5. Collective variables used during the CPMD/MM metadynamics simulation.

The extended Lagrangian version of the metadynamics method was used for a proper coupling with the CPMD/MM simulations (Iannuzzi, 2003). The selected mass values of the fictitious particles were 2 a. m. u. (CV_1 and CV_2) and 1 a. m. u. for the difference in coordination numbers (CV_3), while those of the force constant were 20 a. u. for all CVs. The height of the Gaussian terms was 2.0 kcal·mol^{-1} and a width of 0.3 a. u, 0.4 a. u. and 0.1 arbitrary units for CV_1, CV_2 and CV_3 were used, respectively. A new Gaussian-like potential was added every 200 MD steps and the simulation was stopped after one recrossing of the first TS. Approximately 940 Gaussian hills were added during the metadynamics simulation which, in terms of simulation time, corresponds to 23 ps.

3. Results and discussion.

3.1 Metadynamics simulations of substrate binding.

The FEL for substrate binding obtained from the classical metadynamics simulation (Figure VI-6) show two main energy minima (stable configurations). The first one (1 in Figure VI-6; CV_1 ≈ 8 Å and CV_2 ≈ 25 Å) corresponds to the initial structure obtained from the initial equilibrium MD simulation (Figure VI-7A), in which the acceptor molecule partially enters the active site whereas the donor remains at its entrance. The second energy minimum (2 in Figure VI-6; CV_1 ≈ 4 Å and CV_2 ≈ 2.5 Å) corresponds to the ternary complex, in which the two pNPβGal molecules are inside the enzyme active site (Figure VI-7A). Close look at the orientation of the two molecules in the active site (Figure VI-7B) shows that the hydroxymethyl group of the acceptor molecule is in an optimum configuration for a front-face mechanism, which could ultimately lead to a transglycosylation product with retention of configuration. The terminal hydrogen atom of the hydroxymethyl group points towards the glycosidic oxygen of the donor molecule, favoring the frontal-face transglycosylation reaction with the formation of a 1,6-glycosidic linkage. This interaction is probably important to guide the nucleophile to the same face as the leaving group, as observed in a retaining glycosyl transferases (Ardèvol, 2011). Furthermore, this implies that the leaving group glycosidic oxygen may act as a general base that deprotonates the incoming 6-hydroxyl, consistent with the non-detrimental effect of the mutation of the general base (nucleophile) residue in SsβG E387Y.

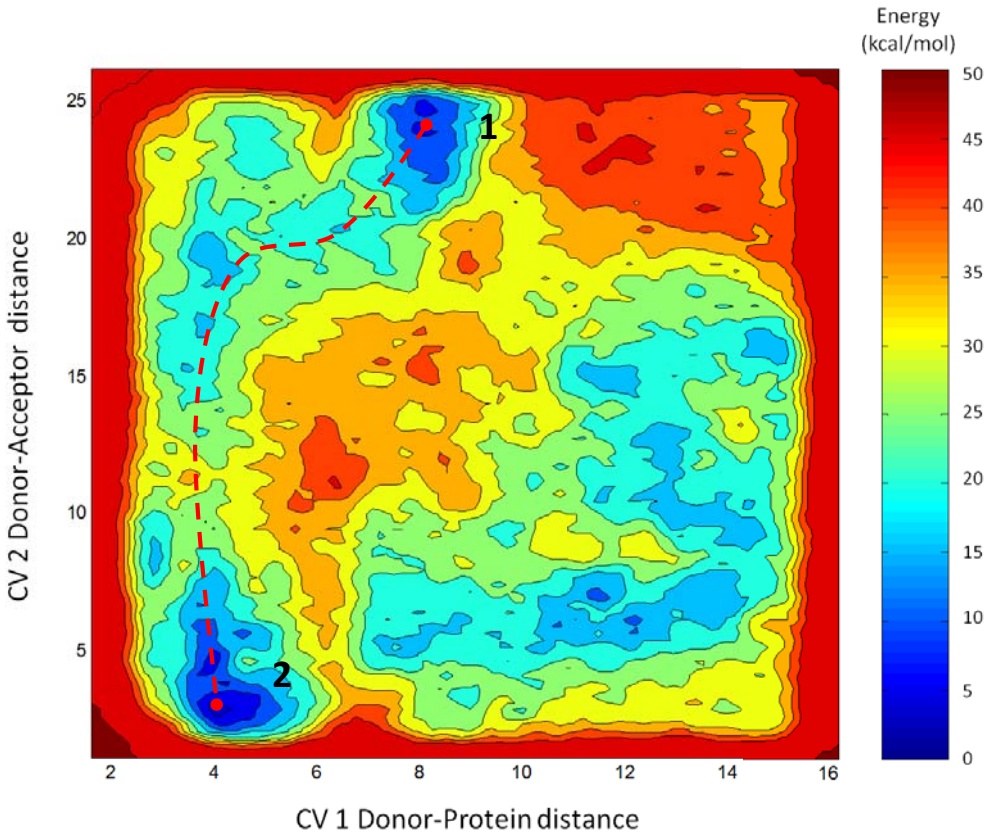

Figure VI - 6. Free energy landscape for the binding of two pNPβGal molecules into the active site of the SsβG E387Y enzyme mutant.

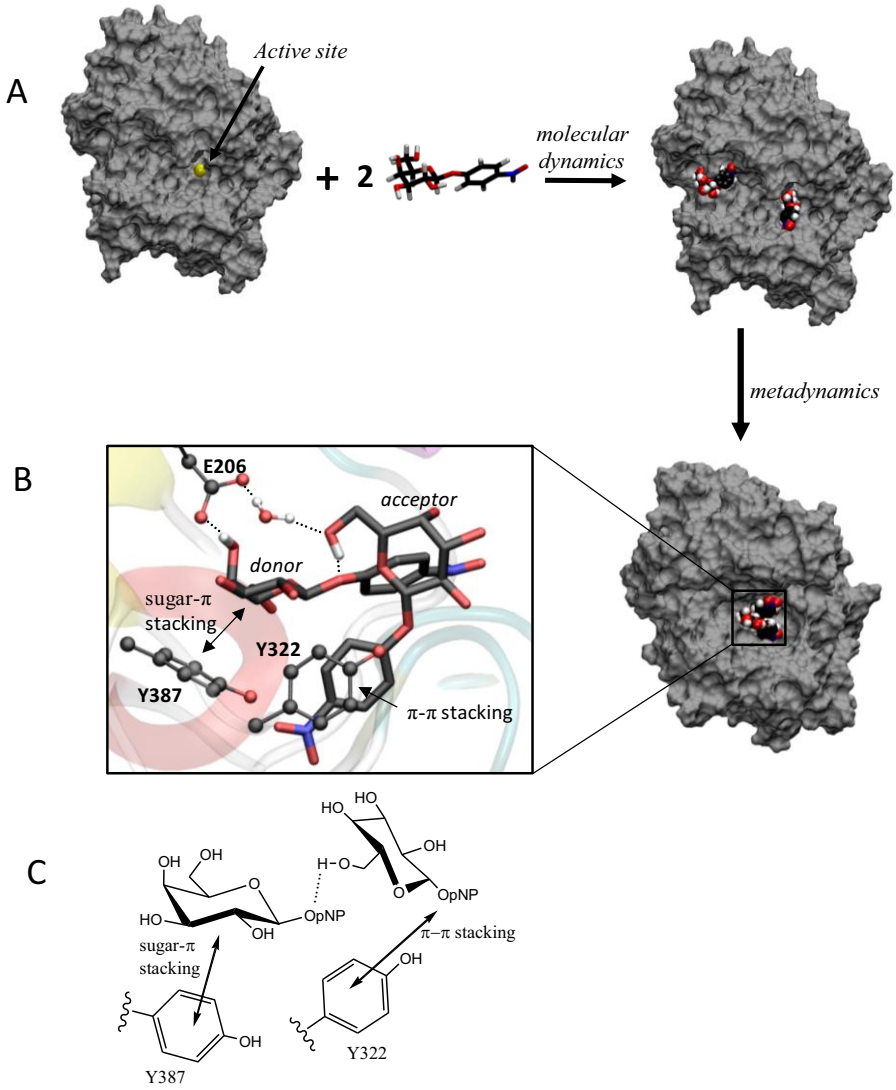

Figure VI - 7. A) Procedure used to model the complex of the E387Y mutant enzyme with two pNPβGal molecules. B,C) Detailed structure of the active site ternary complex.

There are crucial substrate-protein interactions that contribute to the stability of the above "front-face configuration" (Figure VI-7B,C). First of all, the mutated amino acid Y387 forms stabilizing sugar···π interactions with the donor molecule (sugar hydrogen atoms point towards the center of the Y387 phenol ring, with distances < 3 Å). Second, Y322 forms π···π stacking interactions with the acceptor pNP ring (the distance between carbon atoms of both six-membered rings amounts to ~3.5 Å). Most likely this places the 6-hydroxyl group in an optimum orientation to attack the anomeric carbon of the sugar donor. These π···π stacking interactions explain why activated sugars, such as pNPβGal, are good substrates for the SsβG E387Y enzyme and why non-aromatic acceptor substrates are not processed to any extent (see the Introduction of this Chapter).

In summary, classical simulations show that the relative orientation of the donor and acceptor sugar rings, in the complex of SsβG with two molecules of pNPβGal, is such that the donor anomeric carbon is spatially accesible to the acceptor O6 hydroxyl group. This suggest a direct nucleophillic subtitution on the same face of the leaving group, i. e. a frontal-face mechanism for the transglycosylation reaction of E387Y SsβG.

3.2 CPMD/MM metadynamics simulations.

The FEL for the transglycosylation reaction, obtained by the CPMD/MM metadynamics simulations, is shown in Figure VI-8. The most stable minimum of the FEL corresponds to the products well (**P**), located on the upper-left corner. The reactants well (**R**) is in the lower-right corner and agrees well with the ternary complex obtained by classical metadynamics (Figure VI-7B). This well is flanked by another small well (**R'**) that displays a larger C1-O6' distance. There is another energy minimum that corresponds to the intermediate of the reaction (**2** in Figure VI-8) and is separated from reactants and products by two saddle points (**1** and **3** in Figure VI-8), which are ~25 kcal·mol^{-1} above the reactants well. This value is similar to the ones reported for glycosyl transferases investigated previously in the group with a similar methodology (Ardèvol, 2011)(Rojas-Cervellera, 2013).

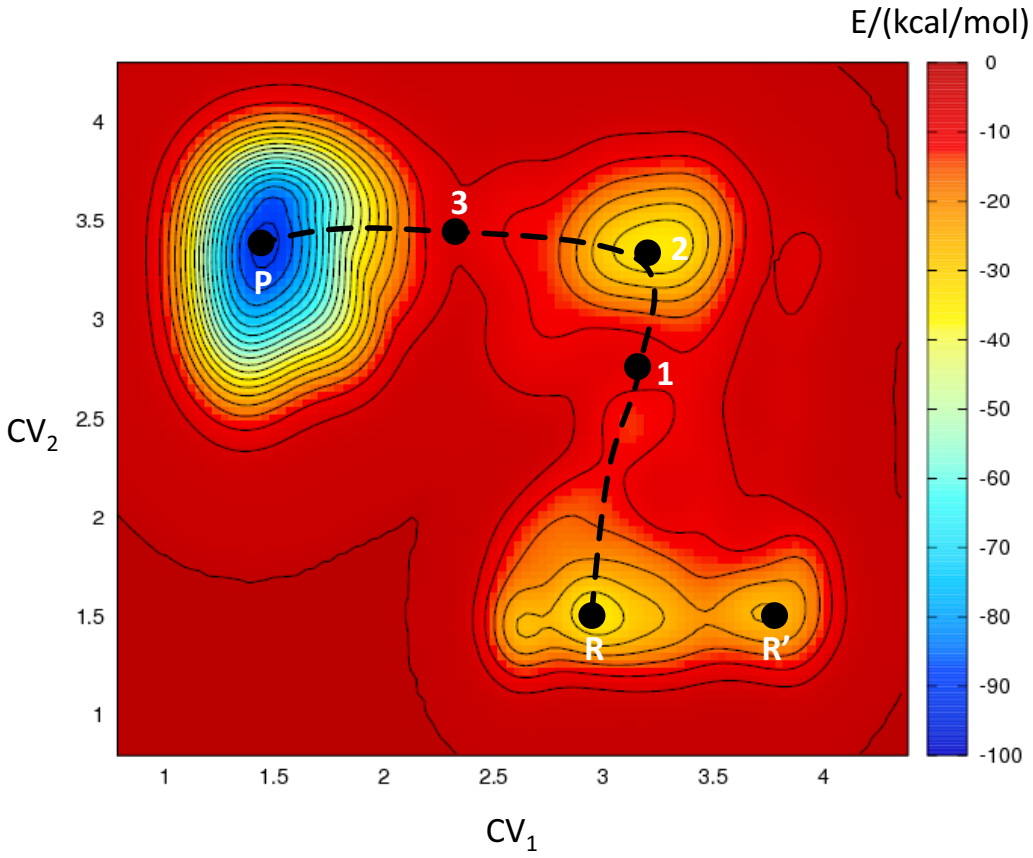

Figure VI - 8. Free energy landscape for the transglycosylation reaction (CV_1 and CV_2). The third variable (proton transfer, CV_3) has been integrated out in order to obtain a 2D contour plot. Contour lines represent 5 kcal·mol^{-1}.

Snapshots of the most relevant structures along the minimum free energy path are shown in Figure VI-9. The Michaelis complex (**R**) features a hydrogen bond between the hydroxymethyl group of the acceptor molecule and the leaving group of the donor unit (*p*-nitrophenol)(**Reactants** in Figure VI-9). This type of interaction was previously observed on the basis of CPMD/MM calculations for other retaining GTs studied in the group (Ardèvol, 2011)(Lira-Navarrete, 2014) as well in other groups (Gómez, 2012)(Gómez, 2014) and may be a common feature of enzymes operating via a front-face mechanism.

The reaction starts with the elongation of the C1-O_{PNP} bond of the donor molecule. The C-O distance increases more than 1 Å when going from the **Reactants** to the first transition state (TS)(**1**)(Figure VI-9 and Table VI-1). This bond is completely broken at **2** (C1-O_{PNP} = 3.4 Å, Figure VI-9 and Table VI-1). Remarkably, at this point of the reaction, the distance between the donor and the acceptor (C1···O6') is still long (~3 Å), indicating the formation of an oxocarbenium ion pair (**2** in Figure VI-9). Further evidence for the change in electronic configuration at the anomeric carbon atom is the shift from a tetrahedral geometry to a trigonal geometry, which is also associated with changes in the conformation of the pyranose ring along the reaction (Figure VI-10; discussed below). This change is accompanied by a decrease in the C1–O5 bond length (from 1.41 Å to 1.27 Å, Table VI-1) and an increase of the charge of the anomeric center (by 0.30 e^- when going from **R** to **2**).

The oxocarbenium ion pair formed during the reaction is stabilized by the O6-H···O_{PNP} hydrogen bond (**2** in Figure VI-9), which has also a role in orienting the acceptor sugar for the subsequent nucleophilic attack. Afterward, a slight displacement of the hydroxymethyl moiety coupled to a proton transfer, from the hydroxymethyl to the *p*NP leaving group, forms the sugar-sugar glycosidic bond (**3** and **P** in Figure VI-9).

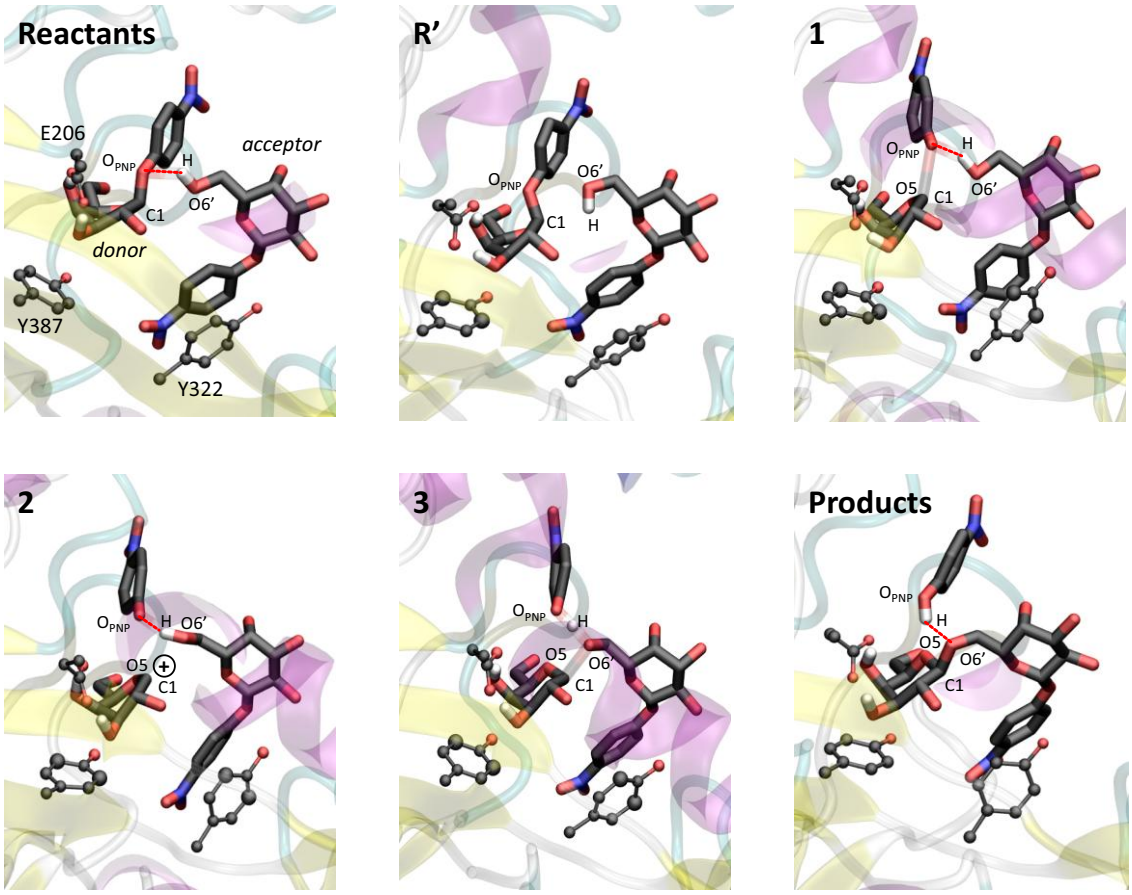

Figure VI - 9. Snapshots along the minimum free energy reaction path.

	C1-O_{PNP}	C1-O1	C1-O5	O1-H	O_{PNP}-H
Reactants	1.51(0.13)	2.65(0.12)	1.41(0.04)	0.98(0.02)	2.56(0.39)
R'	1.48(0.08)	3.76(0.09)	1.41(0.07)	1.00(0.03)	3.82(0.40)
1	2.69(0.13)	3.02(0.07)	1.28(0.02)	1.05(0.08)	2.01(0.27)
2	3.41(0.13)	3.15(0.11)	1.27(0.02)	1.01(0.08)	2.31(0.67)
3	3.48(0.06)	2.30(0.11)	1.39(0.10)	1.43(0.68)	1.47(0.46)
Products	3.50(0.02)	1.44(0.03)	1.42(0.01)	1.95(0.07)	0.97(0.07)

Table VI - 1. Evolution of the most relevant distances involving the donor and the acceptor along the reaction coordinate of Figure VI-9. Each distance corresponds to an average from all configurations falling into a small region around the corresponding point of the FEL.

Analysis of the -I subsite ring distortion along the reaction path (Figure VI-10) evidences that the conformational itinerary is $^1S_3$ (Reactants) – $^4H_3/E_3$ (Reaction Intermediate) – $^4C_1$ (Products). This conformational itinerary is the same followed by retaining β-D-gluco-active GH enzymes (Vocadlo, 2008) and identical to the one found by Biarnés *et al.* (Biarnés, 2011) for 1,3-1,4-β-Glucanase.

The $^1S_3$ conformation for the reactants structure, obtained by *ab initio* methods, differs from the one observed during the classical simulations (the predominant conformation for the -I subsite galactose molecule in the classical MD and metadynamics simulations for substrate binding was $^4C_1$; Figure VI-7B). This is probably the reason for the failure of the initial CPMD/MM metadynamics simulations, starting from the reactants structure obtained from the classical simulations (see Methods section of this chapter). Overall, this indicates that the GLYCAM force field does not describe well the sugar configuration inside the enzyme. The same was observed in previous studies of the group (Biarnés, 2006).

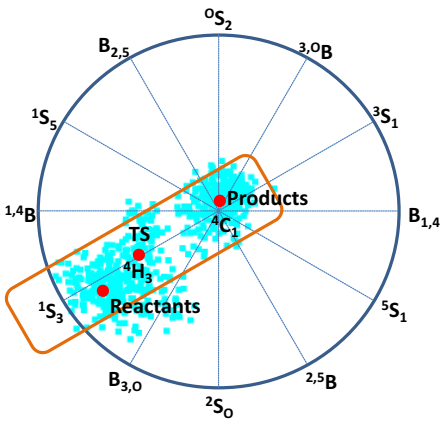

Figure VI – 10. Puckering coordinates of representative structures of the enzymatic reaction (red points) projected on the Stoddart's diagram. Blue dots represent snapshots along the metadynamics simulation.

It is nowadays well known that distorted conformations approach the substrate towards the TS of the reaction by increasing the C1-O1 bond distance (C1-O_{PNP} for pNPβGal molecules) and shortening the C1-O5 distance when compared with the $^4C_1$ conformation (Biarnés, 2006)(Biarnés, 2011). Moreover, the $^4C_1$ conformation does not comply with the ALPH theory (see the General Introduction; section 3.2). Specifically, charge transfer between the free electron pair of the pyranic oxygen with the antibonding σ* orbital of the glycosidic bond is not possible because of the absence of orbital overlap in this conformation. As a result, the glycosidic bond is not polarized and there is no partial double bond formation between C1 and O5 atoms. All these factors could explain why the metadynamics simulation starting with a donor substrate in a $^4C_1$ conformation did not evolve towards products.

It is important to mention that the CVs used in the CPMD/MM metadynamics simulations do not impose any particular conformational itinerary. For instance, the conformation of the donor molecule could have evolved to a $^4C_1$ conformation for the reactants (as observed in the classical simulations) if that was the most stable conformation along the reaction. As found in previous metadynamics studies of GH catalysis (Biarnés, 2011)(Petersen, 2010), the sugar ring adopts distinct conformations during the reaction pathway, which correspond to radial of peripherial routes on the Stoddart's diagram (Davies, 2012).

In summary, by means of a combination of classical MD, classical metadynamics and *ab initio* metadynamics, we have confirmed that an enzyme, whose function is to cleave sugar bonds, can be engineered to achieve synthesis of glycosidic bonds with retention of configuration.

4. Conclusions.

The main conclusions of the present chapter are:

1. The tranglycosylation reaction mechanism of E387Y SsβG mutant follows a front-face mechanism as observed for reatining glycosyl transferases. The energy barrier (~25 kcal·mol^{-1}) is similar to the ones reported for other enzymes studied with similar computational methodologies.

2. π···π stacking interactions, formed between tyrosine residues located at the active site of the enzyme (Y387 and Y322) and substrate molecules, explain why activated pNPβGal molecules are such a good substrates for SsβG E387Y.

3. The conformational itineray followed by the sugar donor in the SsβG E387Y mutant enzyme corresponds to a $^{1}S_{3} \rightarrow [^{4}H_{3}]^{\ddagger} \rightarrow {}^{4}C_{1}$, as observed for the hydrolysis of glycosidic bonds in retaining β-D-gluco active GH enzymes. Therefore, hydrolysis and synthesis of glycosidic bonds follow the same itinerary in this enzyme.

Chapter VII – Molecular mechanism of GalNAc-T2 glycosyl transferase.

Lira-Navarrete[*], E.; Iglesias-Fernández[*], J.; Zandberg, W. F.; Compañon, I.; Kong, Y.; Corzana, F.; Pinto, B. M.; Clausen, H.; Peregrina, J. M.; Vocadlo, D. J.; Rovira, C.; Hurtado-Guerrero, R.

Substrate-Guided Front-Face Reaction Revealed by Combined Structural Snapshots and Metadynamics for the Polypeptide N-Acetylgalactosaminyltransferase 2.

Angew. Chem. Int. Ed. **2014**, 53: 8206-8210.

[*]**Equal contribution**

Molecular mechanism of GalNAc-T2 glycosyl transferase.

1. Introduction.

Glycosyl transferases (GTs) catalyze the formation of glycosidic linkages by the transfer of a saccharide, typically a monosaccharide, from a sugar nucleotide donor to a variety of acceptor substrates: carbohydrates, proteins, lipids, DNA, etc. (Breton, 2012). Like glycosidases, GTs catalyze glycosyl transfer with either inversion or retention of the donor anomeric stereochemistry (Lairson, 2008). Whereas the mechanism of inverting GTs is clearly establish, an S_N2 reaction in a single displacement step, the mechanism of retaining GTs remains one of the most controversial issues in the field of glycobiology (Lairson, 2008)(see General Introduction; section 4).

By comparison with retaining GHs, the mechanism of retaining GTs was first suggested to be a double displacement mechanism with the formation of a covalent glycosyl-enzyme intermediate (Figure VII-1a). Even though a recent CPMD/MM metadynamics work (Rojas-Cervellera, 2013) demonstrates the double displacement mechanism for the enzyme alpha-3-galactosyl transferase (α3GalT; a family 6 GH), few other GTs have an appropriately positioned nucleophile, within the active site, to perform such a reaction (Lairson, 2008). On the contrary, crystal structures of other GT complexes (Persson, 2001)(Errey, 2010) showed that the position of the donor and acceptor substrates is in agreement with a frontal-face nucleophilic substitution. In this reaction mechanism, the nucleophilic hydroxyl group of the acceptor attacks the anomeric carbon from the same side from which the leaving group departs (Figure VII-1), thus leading to retention of stereochemistry (Sinnott, 1990). In addition, kinetic isotope effects and linear free energy relationships (Lee, 2011) and a CPMD/MM metadynamics study (Ardèvol, 2011) clearly support a front-face mechanism for the family 20 retaining GT trehalose-6-phosphate synthase, OtsA.

Subsequent theoretical studies of the reaction mechanism of retaining GTs include a static QM/MM of the family 8 GT lipopolysaccharyl-α-1,4-galactosyl transferase C (LgtC)(Gómez, 2012) and family 15 GT α-1,2-mannosyl transferase (Kre2p/Mnt1p)(Bobovská, 2014). Both studies, performed at the QM(DFT)/MM level of theory, support a front-face mechanism for glycosyl transfer (Figure VII-1b). However, whereas a step-wise mechanism was found for OtsA and Kre2p, a fully concerted mechanism (denoted as S_Ni by the authors) was found for LgtC. On the other hand, the last authors concluded that the enzyme α3GalT can follow either a double-displacement or a fully concerted front-face mechanism (Figure VII-1)(Gómez, 2012). Therefore, the fine details of the front-face mechanism in retaining GTs are still controversial.

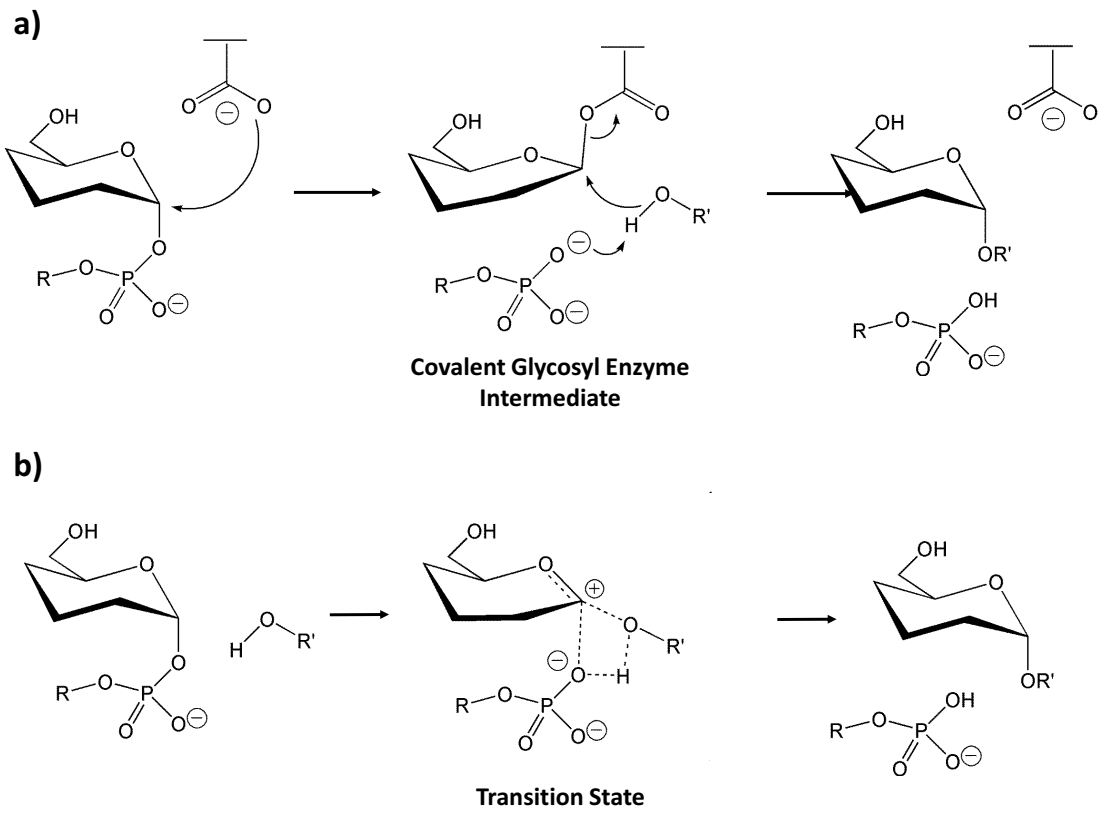

Figure VII - 1. Schematic representation of the possible reaction mechanisms for the enzymatic glycosyl transfer with retention of configuration. (a) Double displacement mechanism. (b) Front-face mechanism (the ion pair species can be either an oxocarbenium-ion transition state or a short-lived intermediate).

1.1 GalNAc-Transferase 2(GalNAc-T2).

N-acetylgalactosaminyl transferases (GalNAc-Ts, E.C 2.4.1.41) are GT-A fold enzymes that catalyze the transfer of GalNAc from a sugar donor UDP-GalNAc to the side chain of serine or threonine residues (R'=Ser or Thr in Figure VII-1) and, thereby, have a transcendental role in protein glycosylation (Bennett, 2012). They are classified as GT27 family members in the CAZy glycosyl transferases data base (Lombard, 2013).

GalNAc-Ts represent a large and evolutionary conserved family of enzymes with up to 20 gene entries identified in humans. These enzymes share a common type II membrane structure of Golgi GTs, having a C-terminal ricin-like lectin domain (~120 amino acids)(Hazes, 1996)(Imberty, 1997) in addition to a catalytic domain (GT-A motif, ~230 amino acids)(Hagen, 1999) formed by two interacting β-α-β Rossmann-like folds that bind the UDP-GalNAc donor (Figure VII-2). A DXH ion binding motif, conserved in all human isoforms, binds a Mn^{2+} ion and also interacts with the UDP diphosphate leaving group. There is also a loop (active loop in Figure VII-2) that undergoes large conformational changes in response to UDP-GalNAc binding and has an effect on the binding of the peptide (Bennett, 2012).

The lectin domains, which present poorly conserved sequences, are structures with high specificity for sugar molecules that have been classified as family 13 carbohydrate-binding modules in the CAZy database (Lombard, 2013). They have a function in modulating and improving the catalytic efficiency of GalNAc-Ts with partially GalNAc-glycosilated peptides (Bennett, 2012).

The catalytic and lectin domains are connected by a linker of 10-25 amino acids with variable sequence and flexibility, which may have a function in controlling the relative orientation between both domains. For example, GalNAc-T1 displays highly associated catalytic and lectin domains, with a large contact area (Fritz, 2004), whereas in GalNAc-T2 there is practically no interaction (Fritz, 2006). Therefore, it is expected that differences in linker sequence modify the specificities of GalNAc-Ts for partially glycosilated substrate peptides (Kubota, 2006).

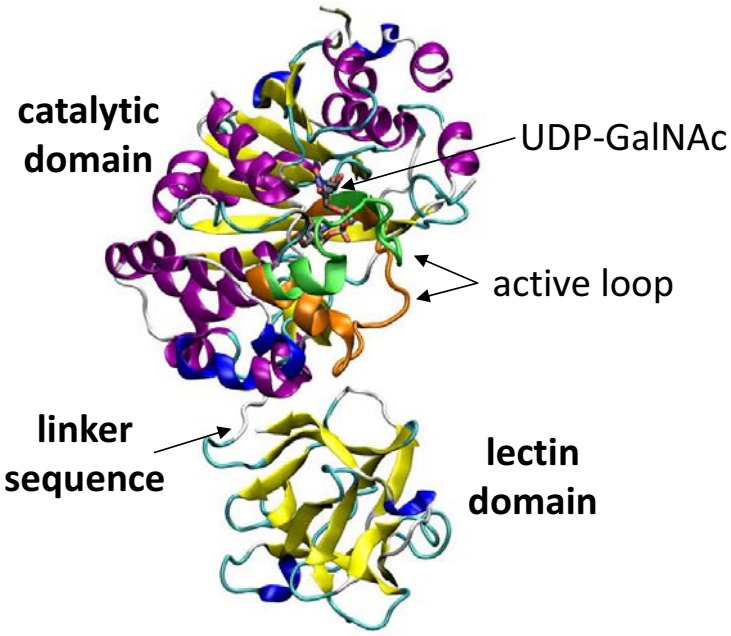

Figure VII - 2. Cartoon representation of GalNAc-T2 GT in complex with UDP-GalNAc. The flexible loop is shown in green and orange for the active (pdb code: 4D0T) and inactive conformations (data not published), respectively.

Recently, the group of Prof. Ramón Hurtado from the University of Zaragoza was able to solve the first X-ray structure of a GalNAc-T2 enzyme in complex with UDP-GalNAc and the acceptor peptide (Lira-Navarrete, 2014). The ternary complex (i. e. Michaelis complex; pdb code: 4D11) shows a threonine residue embedded within the active site and pointing towards the UDP-GalNAc donor sugar molecule in a conformation suitable for a front-face mechanism. Apart from being the first ternary complex of GalNAc-T2 with both the donor and the acceptor substrates, this structure is also the first one in which the ternary complex could be crystallized for the wild-type enzyme using natural substrates (i. e. without modification either on the enzyme or the substrates). In collaboration with the group of Prof. Hurtado, we used Car-Parrinello quantum mechanics/molecular mechanics (CPMD/MM)(Laio, 2002a) metadynamics simulations to decipher the reaction mechanism of GalNAc-T2.

2. Computational details.

2.1 Initial structure.

The initial structure for the calculations was taken from the complex of the enzyme GalNAc-T2 with UDP···Mn^{2+} and the N-acetylgalactosamine molecule covalently bound to a threonine residue of the EA2 peptide (PDB code: 4D0T), which structurally represents the products of the reaction. We selected this structure, and not the (also available) Michaelis complex (pdb code: 4D11) in view of its higher resolution. The protonation states and hydrogen atom positions of all ionizable amino acid residues were selected based on their hydrogen bond environment. Eight histidine residues were modeled in their neutral states and four in their protonated state. All the crystallographic water molecules were retained and extra water molecules were added to form a 10 Å water box around the protein surface. Nine chloride ions were also added to neutralize the enzyme charge.

2.2 Classical molecular dynamics simulations.

A classical molecular dynamics using Amber11 software (Case, 2010) was initially performed. The protein and the GalNAc molecule were modeled with the FF99SB (Hornak, 2003) and GLYCAM06 (Kirschner, 2008) force fields, respectively. Finally, all water molecules were described with the TIP3P force field (Jorgensen, 1983). The MD simulation was carried out in several steps. First, the system was minimized, holding the protein and substrate fixed. Then, the entire system was allowed to relax. To gradually reach the desired temperature of 300 K in the MD simulation, weak spatial constraints were initially added to the protein and substrate, while the water molecules and chloride ions were allowed to move freely. The constraints were then removed to reach the desired temperature and the system was subjected to 100 ps of constant pressure MD simulation to adjust the density of the water environment. After the equilibration process, 8 ns of constant volume MD simulation were performed. Analysis of the backbone RMSD showed that after this simulation time the protein was equilibrated to the desired conditions. A snapshot of the MD-equilibrated system was taken for the subsequent CPMD/MM and metadynamics simulations.

2.3 CPMD/MM metadynamics simulations.

CPMD/MM simulations were performed, as in the previous chapters, using the method developed by Laio *et al*., (Laio, 2002a) which combines Car-Parrinello MD (Car & Parrinello, 1985), based on DFT, with force field MD methodology. The QM region included the GalNAc sugar molecule, three amino acids of the EA2 acceptor peptide (backbone of Thr6, Thr7 and Pro8), the phosphate groups of the UDP leaving group, the Mn^{2+} cation and a close water molecule, as well as the side chain of Asp224, His226, His359 and Arg362 (113 QM atoms, 73400 MM atoms; Figure VII-3a). The side chains of Asp224, His226, His359 and Arg362 were capped at its C_β (histidines), C_α (aspartate) or C_γ (Arginine) atoms with a link-atom pseudopotential (Zang, 1999). The QM region was enclosed in an isolated 20.0 x 23.0 x 24.0 Å supercell. Kohn–Sham orbitals were expanded in a PW basis set with a kinetic energy cutoff of 70 Ry. Norm-conserving Troullier–Martins *ab initio* pseudopotentials were used for all elements (Troullier, 1991). The calculations were performed using the PBE generalized gradient-corrected approximation (Perdew, 1996). 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 0.12 fs. The simulation of the GalNAc-T2 complex with UDP, GalNAc and the EA2 peptide was equilibrated at 300 K for approximately 4 ps before starting the metadynamics run. The final structure of the CPMD/MM MD agrees with the initial X-ray structure (Figure VII-3b).

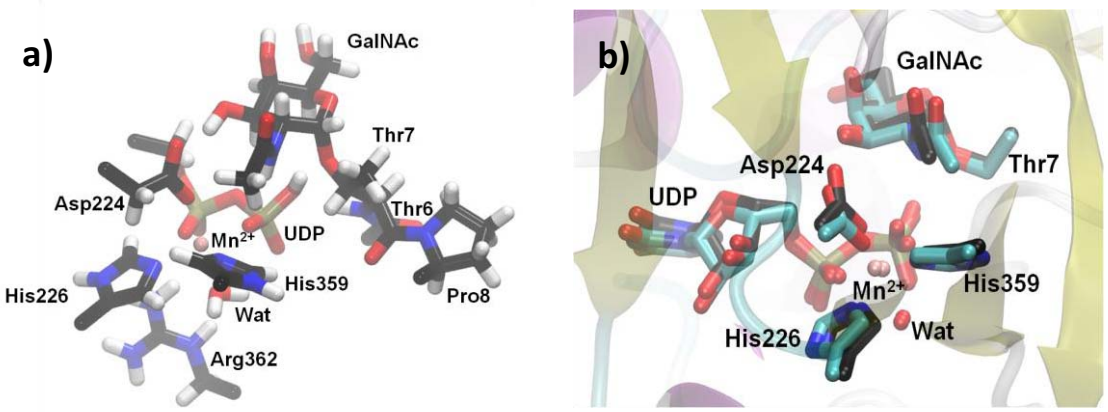

Figure VII - 3. a) *Ab initio* atoms selected for the simulations. b) Final structure of the MD CPMD/MM simulation of the reaction products (black) superimposed with the initial X-Ray structure (cyan) (PDB code 4D0T).

CPMD/MM metadynamics simulations, aimed at describing the reaction mechanism of GalNAc-T2, were performed on the extended Lagrangian formalism (Ianuzzi, 2003). Three collective variables (CVs) were used (Figure VII-4). The first one (CV_1, bond breaking) was taken as the distance between the anomeric carbon of the GalNAc molecule and the hydroxyl oxygen of threonine 7 (Thr7) from the EA2 peptide molecule ($CV_1 = dC_1\text{-}O_{Thr}$). The second collective variable (CV_2, bond formation) was taken as the distance between the anomeric carbon of the GalNAc molecule and the corresponding oxygen of the UDP phosphate ($CV_2 = dC_1\text{-}O_{UDP}$). Finally, the third collective variable (CV_3, proton transfer) includes all the atoms involved in the proton transfer process from the UDP phosphate to the threonine amino acid. This CV was taken as the difference in coordination number (CN) between the UDP phosphate oxygen with the hydrogen atom minus the oxygen of the acceptor threonine ($CV_3 = CN_{UDP\text{-}H} - CN_{Thr\text{-}H}$). The CN is defined using the following equation:

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

where d_{ij} is the internuclear distance of the atoms involved, d_0 is the threshold distance for bonding, and p and q are exponents that determine the steepness of CN_{ij} decay with respect to dij. CN values range from 0 (no bond) to 1 (a bond). The parameters used are p = 10, q = 16 and d_0 = 2.50 a. u. (1.32 Å).

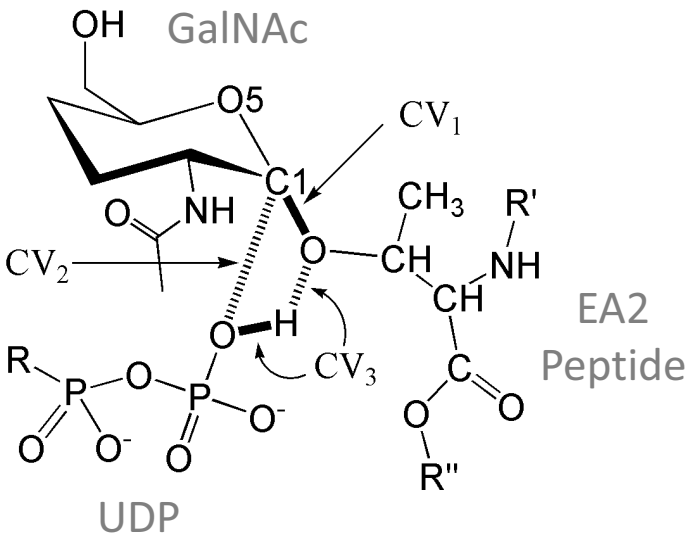

Figure VII - 4. Collective variables used in the CPMD/MM metadynamics .

The selected mass values (μ) and force constants (k) of the fictitious particles were 0.5 a. m. u./30 a. u. for both distances and 0.1 a. m. u./10 a. u. for the difference in coordination numbers. The height of the Gaussian used was of 1.0 kcal·mol^{-1}. The width of the Gaussian terms (0.25 a. u. for both distances and 0.05 arbitrary units for the difference in coordination numbers) were selected from the oscillations of the CVs in a free CPMD/MM simulation. A new Gaussian-like potential was added every 150 MD steps and the simulation was stopped after 6 ps once the intermediate and the reactants wells were sampled. A total of 350 Gaussian hills were deposited.

3. Results and discussion.

The free energy landscape (FEL) reconstructed from the metadynamics simulation is show in Figure VII-5. The two most stable minima are the products of the reaction (upper-left corner) and the reactants (lower-right corner). The reactants well is 13 kcal·mol^{-1} below the maximum point along the reaction pathway, which is in reasonable agreement with the free energy barrier of 17 kcal·mol^{-1} derived from the experimental k_{cat} value of 3.7 s^{-1} at 310 K (Fritz, 2006).

Two deep minima were found in the FEL, corresponding to the Michaelis complex (**Reactants**) and the **Products**. The other minimum on the FEL (**2** in Figure VII-5) corresponds to an oxocarbenium-phosphate ion pair (both, the bond between GalNAc-UDP and GalNAc-Thr7 were broke and the C1 carbon atom displays a positive charge development). A detailed description of the reaction can be obtained by following the minimum free energy pathway on the FEL (Figure VII-5) and examining snapshots along this pathway. Four structurally different species (**1** to **4** in Figure VII-5) were identified along the reaction path.

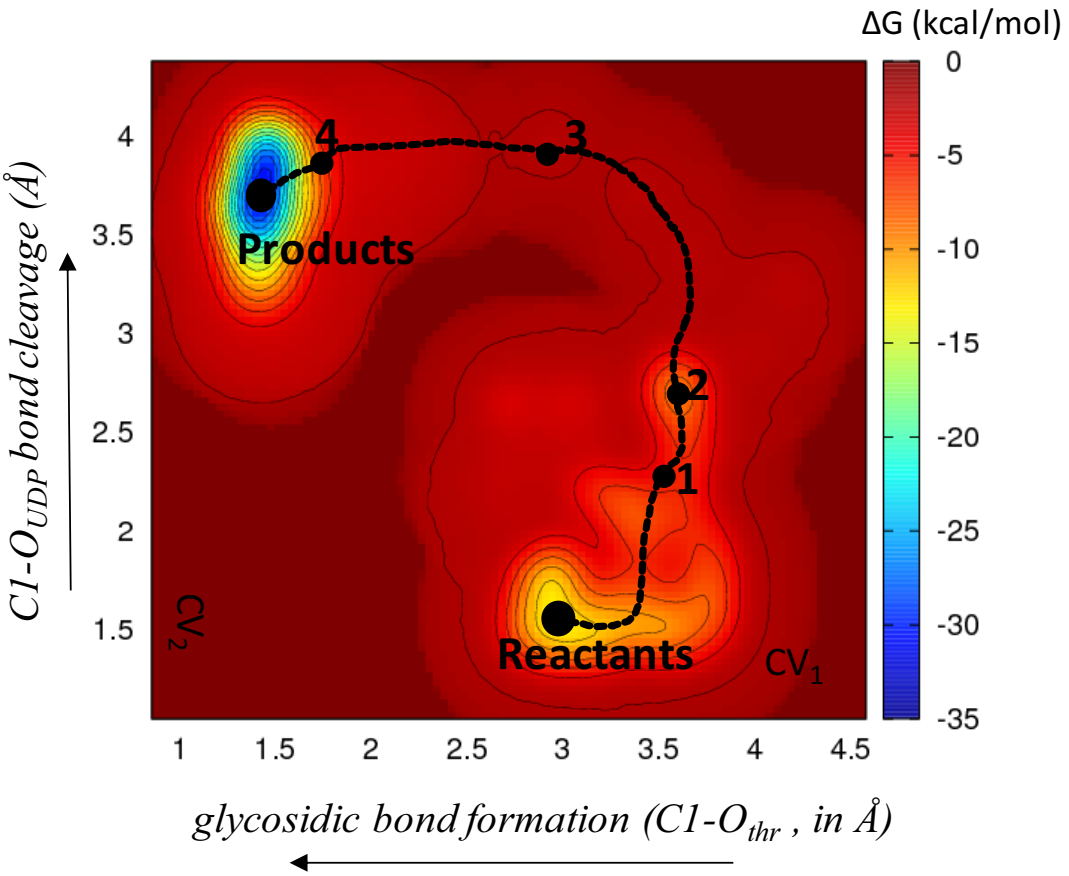

Figure VII - 5. Computed free energy landscape for the glycosyl transfer reaction in GalNAc-T2 with respect to two of the three collective variables described in the text (CV_1 and CV_2). The third variable (proton transfer) has been integrated out in order to obtain a 2D contour plot. The total free energy barrier accounts for 13 kcal·mol^{-1}.

Figure VII-6 shows representative snapshots of relevant configurations along the minimum energy pathway. The Michaelis complex (**Reactants** in Figure VII-6) features a hydrogen bond between the acceptor threonine oxygen atom and one of the pyrophosphate oxygen atoms. Within the literature it is often proposed, based on ternary complexes with substrate analogues (Persson, 2001), or structural superposition of complexes in which either the sugar moiety or the acceptor are absent (Fritz, 2006), that the acceptor hydroxyl group interacts with the leaving group oxygen atom of the pyrophosphate moiety. Here we note that this hydrogen bond was formed, not with the leaving group oxygen atom but, rather with one of the non-bridging pyrophosphate oxygen atoms. This interaction makes good chemical sense since hydrogen bonds are stronger when they are formed between atoms that would have more closely matched pK_a values. This type of configuration, previously proposed on the basis of CPMD/MM calculations of OtsA (Ardèvol, 2011), may be a common feature in GTs operating via front-face mechanism.

The reaction starts with the elongation of the C-O bond between the UDP and the GalNAc molecule of the donor (the C1-O distance increases almost 1 Å when going from the **Reactants** to **1**, Figure VII-6 and Table VII-1). Simultaneously, the O_{Thr}-H and the N_{acetyl}-H bonds change their hydrogen-bond partners, from the non-bridging phosphate oxygen to the leaving group oxygen, thus placing the hydrogen atoms at the proper position to assist the departure of the UDP leaving group (**1** in Figure VII-6). The N_{acetyl}-H···O_{UDP} and O_{Thr}-H···O_{UDP} hydrogen bonds not

only stabilize the negative charge being developed at the phosphate group, but they also appropriately position the acceptor molecule to favor the subsequent front-side nucleophilic attack. The stabilization role of the acetyl group was first described for the inverting O-GlcNAc transferase by a QM(DFT)/MM study (Tvaroska, 2012). Therefore, the first part of the reaction can be described as the cleavage of the sugar-phosphate bond, assisted by the sugar N-acetyl group, together with the formation of a hydrogen bond between the acceptor and the phosphate oxygen atom of the bond being broken.

The sugar-phosphate bond is completely broken at **2** ($C1\text{-}O_{UDP}$ = 2.7 Å, Figure VII-6 and Table VII-1). Remarkably, the distance between the donor and the acceptor ($C1\cdots O_{Thr}$) is still long (> 3 Å), indicating a discrete formation of an oxocarbenium–phosphate ion pair. Further evidence for the change in electronic configuration at the anomeric carbon atom is the shift from a tetrahedral geometry to a trigonal geometry that is also associated with changes in the conformation of the pyranose ring from a $^4C_1$ chair in the Michaelis complex to a $^4H_3$ half-chair conformation in which the C2, C1, O5, and C5 atoms are almost coplanar (**1**, **2**, and **3** in Figure VII-6). This change is accompanied by a decrease in the C1–O5 bond length (from 1.38 Å to 1.27 Å, Table VII-1) and an increase of the charge of the anomeric center (by 0.40 e^- when going from the **1** to **2**). These structural changes are in agreement with KIE experiments an inhibitor analysis on OtsA, which also follows a front-face reaction mechanism (Lee, 2011)(Errey, 2010), and concluded that species with substantial oxocarbenium ion character form along the reaction coordinate. Both the $O\text{-}H_{Thr}\cdots O_{UDP}$ and the $N_{GalNAc}\text{-}H\cdots O_{UDP}$ hydrogen bonds contribute to the stabilization of the phosphate–oxocarbenium ion pair intermediate (**2** in Figure VII-6), as well as orienting the sugar moiety for the subsequent nucleophilic attack.

Afterward, a slight displacement of the GalNAc donor and acceptor molecules takes place, facilitating the interaction between the anomeric carbon and the threonine oxygen atom (**3** in Figure VII-6). Finally, proton transfer to the phosphate group takes place (**4** in Figure VII-6), a process that is almost concomitant with the formation of the sugar–peptide glycosidic bond. A recent study using an static QM/MM approach (Gómez, 2014) found similar results for the changes in conformations of the donor sugar, but some sugar-enzyme interactions and the energy surface (potential energy in Gómez *et al.*) was found to be most consistent with a fully concerted mechanism than a step-wise reaction.

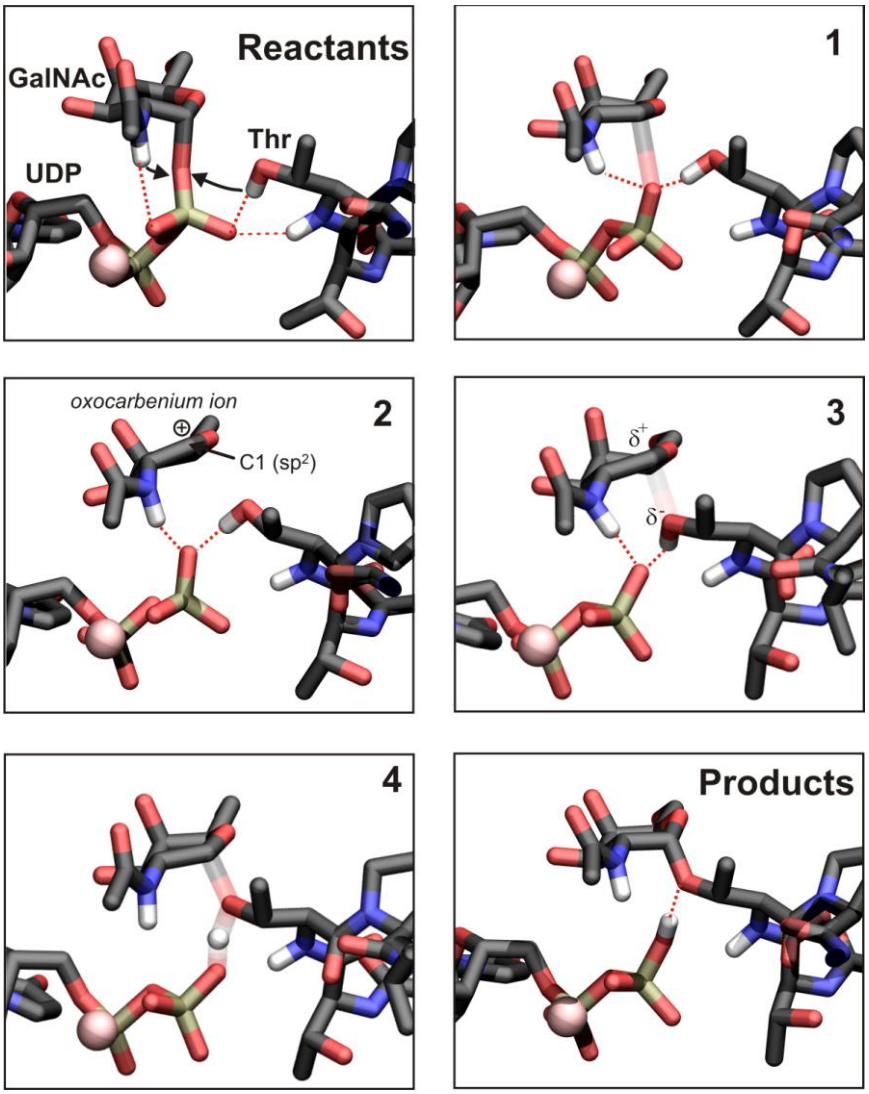

Figure VII - 6. Atomic rearrangements along the reaction pathway. Bonds being broken/formed are drawn as transparent (configurations 1, 3 and 4), whereas relevant hydrogen bonds are represented by a dotted red line. Hydrogen atoms have been omitted for clarity, except the acetamido NH group, and the NH and OH groups of the acceptor threonine. Only the substrates (UDP-GalNAc and the EA2 peptide) and the manganese cation are shown. The arrows in the reactants state indicate the switching motion of the Thr-OH and N-acetyl NH groups to facilitate cleavage of the sugar-phosphate bond.

	C1-O_{UDP}	C1-O_{Thr}	C1-O5	H-O_{Thr}	H-O_{UDP}
R	1.449(0.005)	3.023(0.035)	1.383(0.017)	0.998(0.047)	2.948(0.178)
1	2.213(0.093)	3.627(0.135)	1.277(0.013)	1.006(0.004)	2.943(0.137)
2	2.706(0.059)	3.693(0.021)	1.267(0.006)	1.039(0.047)	2.499(0.117)
3	3.660(0.028)	2.838(0.037)	1.269(0.013)	1.041(0.047)	1.477(0.115)
4	3.862(0.166)	1.907(0.138)	1.325(0.021)	1.438(0.164)	1.217(0.176)
P	3.726(0.197)	1.487(0.040)	1.421(0.028)	1.743(0.237)	1.010(0.021)

Table VII - 1. Relevant distances along the minimum free energy reaction pathway. Distances are in angstroms (standard deviations are given in parenthesis).

Therefore, the simulation shows that cleavage of the GalNAc-UDP bond and the formation of the GalNAc-peptide bond are entirely asynchronous and follow a stepwise pathway. The cleavage of the GalNAc-UDP bond precedes the formation of the GalNAc-peptide bond, and thus an intimate ion pair intermediate is formed, with assistance of coordinating hydrogen bonding interactions between the nucleophilic hydroxyl group, the pyrophosphate leaving group, and the *N*-acetyl NH moiety of the pyranose group. The relevance of the acetamido group in catalysis is further supported by the substrate specificity of this family of enzymes, which exclusively use UDP-GalNAc and rarely accept UDP-glucose (Wandall, 1997).

The overall reaction coordinate can be regarded as “a type of internal return”, as originally proposed by Sinnott and Jencks in their seminal work (Sinnott, 1980) on the solvolysis of glucose derivatives in mixtures of ethanol and trifluoroethanol, by analogy to the S_Ni mechanism for the decomposition of alkyl chlorosulfites (Lewis, 1952). For glycosyl transfer, the names *S_Ni-like* or *S_Ni-type* were later adopted, although the reaction has obvious differences with that of alkyl chlorosulfites (the main one being that for enzymes the nucleophile is not internal but external). The reaction can also be termed S_N1, since an intermediate (although short-lived) is formed. Independently of the terminology adopted, the reaction mechanism of GalNAc-T2 is consistent with a front-face attack, as previously found for OtsA (Ardèvol, 2011), LgtC (Gómez, 2012) and Kre2p/Mnt1p (Bobovská, 2014) as well as for glucosyl transfer in solution (Chan, 2012). Altogether, our results show that the glycosyl transfer with retention of the anomeric configuration can take place in the absence of a protein nucleophile, provided that hydrogen atoms of the donor an acceptor are properly oriented to assist the cleavage of the sugar-nucleotide bond.

4. Conclusions.

The main conclusions of the present chapter are:

1. The GalNAc transfer reaction from GalNAc-UDP to the acceptor peptide follows a very dissociative mechanism with a short-live ion pair intermediate. The cleavage of the GalNAc-UDP bond is assisted by the acceptor molecule as well as the acetyl group of the GalNAc molecule.

2. The oxocarbenium–phosphate ion pair is characterized by a positive charge development on the C1 anomeric carbon, which adopts a sp^2 hybridization.

3. The GalNAc transfer reaction in GalNAc-T2 enzyme can take place *via* a front-face mechanism in the absence of an enzymatic nucleophile.

Chapter VIII – Summary and conclusions.

Summary and conclusions.

In the present Thesis, *ab initio* molecular dynamics simulations have been used to study the conformational flexibility of isolated monosaccharides, with the aim to determine the relative stabilities of ring conformations and predict GH conformational itineraries. The same methodology has been applied to GH inhibitors and sugar cations (D-glucosyl, D-mannosyl, xylosyl and L-fucosyl cations), to gain insights into TS mimicry properties. Additionally, force field and hybrid CPMD/MM simulations were used to study substrate binding and reactivity in carbohydrate-active enzymes (one engineered GH and one GT). The adopted approach allows the study of reactive processes using models of the WT enzyme and its natural substrates, complementing experimental investigations using enzyme mutants or unreactive substrate mimics and providing a full molecular description of the enzymatic reaction.

The main conclusions reached in this Thesis are the following:

- Both sets of collective variables (CVs; Cartesian and Mercator) show identical results for the conformational free energy landscape (FEL) of cyclohexane. Therefore, both CVs can be used indistinctly in *ab initio* metadynamics simulations of sugar puckering.

- The conformational FEL of α-D-mannopyranose gives support to the two conformational itineraries proposed experimentally for the glycosylation reaction catalyzed by α-mannosidases ($^{1}S_{5} \leftrightarrow [B_{2,5}]^{\ddagger} \leftrightarrow {}^{O}S_{2}$ and $^{3}S_{1} \rightarrow [^{3}H_{4}]^{\ddagger} \rightarrow {}^{1}C_{4}$).

- The specific (preactivated) conformations that GHs use for catalysis can be predicted by *ab initio* MD simulations of their isolated sugar substrates.

- GHs further reshape the conformational energy landscape of substrates (i. e. they restrict the substrate accessible conformations) to stabilize those geometries that are on the *conformational catalytic itinerary*.

- One important factor that dictates the TS mimicry capabilities of an inhibitor is the resemblance of its intrinsic conformational properties (i. e. the FEL) to the ones of a pure sugar ion TS.

- The formation of the glycosidic bond with retention of the anomeric configuration, catalyzed by the enzyme mutant SsβG E387Y follows a front-face mechanism in which an oxocarbenium ion-like intermediate forms.

- The reaction catalyzed by GalNAc-T2 can be also described as a front-face mechanism with the formation of a short-lived oxocarbenium ion-like intermediate. The cleavage of the UDP-GalNAc bond is assisted by the peptide acceptor molecule and the N-acetamido donor substituent, with the phosphate leaving group acting as a general base that captures the acceptor proton.

Publications and presentations in congresses.

Publications.

The present Thesis has resulted in the following publications:

- Thompson, A. J.; Dabin, J.; Iglesias-Fernández, J.; Ardevol, A.; Dinev, Z.; Williams, S. J.; Bande, O.; Siriwardena, A.; Moreland, C.; Hu, T. C.; Smith, D. K.; Gilbert, H. J.; Rovira, C.; Davies, G. J. "The Reaction Coordinate of a Bacterial GH47 α-Mannosidase: A Combined Quantum Mechanical and Structural Approach." *Angew. Chem. Int. Ed.* **2012**, 51: 10997-11001.

- Williams*, R. J.; Iglesias-Fernández*, J.; Stepper, J.; Jackson, A.; Thompson, A. J.; Lowe, E. C.; White, J. M.; Gilbert, H. J.; Rovira, C.; Davies, G. J.; Williams, S. J. "Combined Inhibitor Free-energy Landscape and Structural Analysis Reports on the Mannosidase Conformational Coordinate." *Angew. Chem. Int. Ed.* **2014**, 53: 1087-1091. *Equal contribution.

- Lira-Navarrete*, E.; Iglesias-Fernández*, J.; Zandberg, W. F.; Compañon, I.; Kong, Y.; Corzana, F.; Pinto, B. M.; Clausen, H.; Peregrina, J. M.; Vocadlo, D. J.; Rovira, C.; Hurtado-Guerrero, R. "Substrate-Guided Front-Face Reaction Revealed by Combined Structural Snapshots and Metadynamics for the Polypeptide N-Acetylgalactosaminyltransferase 2." *Angew. Chem. Int. Ed.* **2014**, 53: 8206-8210. Equal contribution.

- Iglesias-Fernández, J.; Ardevol, A.; Raich, L.; Rovira, C. "The conformational free-energy landscape of cyclohexane and β-D-xylopyranose using both Cartesian and Mercator representations." *Submitted.*

Besides, the work during the PhD training has resulted in other publications not included in this Thesis:

- Biarnés, X.; Ardèvol, A.; Iglesias-Fernández, J.; 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**, 133, 20301-20309.

- Tankrathok, A.; Iglesias-Fernández, J.; Luang, S.; Robinson, R. C.; Kimura, A.; Rovira, C.; Hrmova, M.; Ketudat Cairns J. R. "Structural analysis and insights into the glycon specificity of the rice GH1 Os7BGlu26 β-D-mannosidase." *Acta Cryst. D* **2013**, 69, 2124-2135.

- Díaz-Rodríguez, A.; Iglesias-Fernández, J.; Rovira, C.; Gotor-Fernández, V. "Enantioselective Preparation of δ-Valerolactones with Horse Liver Alcohol Dehydrogenase." *ChemCatChem.* 2014, 6, 977-980.

Presentations in congresses:

The work in this Thesis has been presented in the following congresses:

- CECAM Tutorial: Hybrid Quantum Mechanics / Molecular Mechanics (QM/MM) Approaches to Biochemistry (and beyond). February 2011. Lausanne, Switzerland.
  Selected oral communication (15'): "Substrate distortion in the Michaelis Complex of Endoglucanase Cel5A for *Bacillus Agaradhaerens*. "

- CECAM Workshop: Innovative Approaches to Computational Drug Discovery. October 2013. Lausanne, Switzerland.
  Invited oral communication (25'): "Conformational free energy landscapes of glycoside hydrolase inhibitors."

- XXVIII annual meeting of the Theoretical Chemistry Network of Catalonia. June 2012. Barcelona, Spain.
  Selected short communication (20'): "Reaction coordinate of a GH47 α-mannosidase."

- CECAM Tutorial: Hybrid Quantum Mechanics / Molecular Mechanics (QM/MM) Approaches to Biochemistry (and beyond). February 2011. Lausanne, Switzerland. **Poster presentation**.

- Annual meeting of the Institute of Computational Chemistry of the University of Barcelona (IQTCUB). April 2011. Barcelona, Spain. **Poster presentation**.

- Carbohydrate Bioengineering Meeting. May 2011. Lisbon, Portugal. **Poster presentation.**

- CPMD2011: Extending the limits of *Ab initio* Molecular Dynamics Simulations for Chemistry, Materials Science and Biophysics. September 2011. Barcelona, Spain. **Poster presentation**.

- CECAM School: MolSim2012. January 2012. Amsterdam, Nederlands. **Poster presentation**.

- Frontiers in Dynamics Simulations of Biological Molecules. November 2013. Barcelona, Spain. **Poster presentation**.

Bibliography.

Bibliography.

Allinger, N. L. & Feriberg, K. A. (1960) *J. Am. Chem. Soc.* 82: 2393-2394.

Allinger, N. L. (1959). *J. Am. Chem. Soc.* 81: 5727-5733.

Allinger, N. L.; Sprague, J. T. (1972) *J. Am. Chem. Soc.* 94: 5734-5747.

Amat, L.; Carbó-Dorca, R. (2000) *J. Chem. Inf. Comput. Sci.* 40:1188-1198.

Amorim, L.; Marcelo, F.; Rousseau, C.; Nieto, L.; Jimenez-Barbero, J.; Marrot, J.; Rauter, A. P.; Sollogoub, M.; Bols, M.; Bleriot, Y. (2011) *Chem. Eur. J.* 17:7345-7356.

Anet, F. A. & Bourn, J. R. (1966) *J. Am. Chem. Soc.* 89: 760-768.

Ardèvol, A.; Biarnés, X.; Planas, A. & Rovira, C. (2010) *J. Am. Chem. Soc.*132: 16058-16065.

Ardèvol, A.; Rovira, C. (2011) *Angew. Chem. Int. Ed*. 50: 10897-10901.

Barducci, A.; Bonomi, M.; Parrinello, M. (2011) *WIREs Comput. Mol. Sci. 1*: 826-843.

Barducci, A.; Bussi, G. & Parrinello, M. (2008) *Phys. Rev. Lett*. 100: 020603.

Becke, A. D. (1984). *J. Chem. Phys.* 84: 4524-4529.

Becke, A. D. (1993). *J. Chem. Phys.* 98: 5648-5652.

Bennett, E. P.; Mandel, U.; Clausen, H.; Gerken, T. A.; Fritz, T. A.; Tabak, L. A. (2012) *Glycobiology*, 22: 736

Biarnés, X. (2007) *PhD Thesis: Mecanismes de preactivació de substrat en 1,3-1,4-β-glucanasa. Modelització mitjançant dinàmica molecular de primers principis.* Barcelona: Universitat de Barcelona. Departament de Química Física. http://www.tdx.cat/handle/10803/2752

Biarnés, X.; Ardèvol, A.; Iglesias-Fernández, J.; Planas, A. & Rovira, C. (2011) *J. Am. Chem. Soc.* 133: 20301-20309.

Biarnés, X.; Ardèvol, A.; Planas, A.; Rovira. C.; Laio, A. & Parrinello, M. (2007) *J. Am. Chem. Soc.* 129: 10686-10693.

Biarnés, X.; Nieto, J.; Planas, A. & Rovira, C. (2006) *J. Biol. Chem.* 281: 1432-41.

Bischoff, H. (1995) *Clin. Invest. Med.* 18: 303-311.

Blake, C. C. F.; Koenig, D. F.; Mair, G. A.; North, A. C. T.; Phillips, D. C. & Sarma, V. R. (1965) *Nature*, 206: 757-761.

Bobovská, A.; Tvaroska, I. & Kóna, J. (2014) *Org. Biomol. Chem*. 12: 4201-4210.

Breton, C.; Fournel-Gigleux, S.; Palcic, M. M. (2012) *Curr. Opin. Chem. Biol.* 22: 540-549.

Caines, M. E.; Hancock, S. M.; Tarling, C. A.; Wrodnigg, T. M.; Stick, R. V.; Stutz, A. E.; Vasella, A.; Withers, S. G. & Strynadka, N. C. (2007) *Angew. Int. Ed. Eng.* 46: 4474-4476.

Cantarel B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. (2009) *Nucleic Acids Res.* 37: D233-238.

Cantu, D.; Nerinckx, W.; Reilly, P. J. (2008) *Carbohydr. Res.* 343: 2235-2242.

Car, R. & Parrinello, M. (1985) *Phys. Rev. Lett.* 55: 2471-2474.

Carloni, P.; Rothlisberger, U. & Parrinello, M. (2002) *Acc. Chem. Res.*, 35 (6): 455–464.

Case, D. A; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; K.F. Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. -J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A. and Kollman, P. A. (2010), AMBER 11, University of California, San Francisco.

Chan, J.; Tang, A.; Bennet, A. J. (2012) *J. Am. Chem. Soc*. 134: 1212-1220.

Charnock, S. J. and Davies, G. J. (1999) *Biochemistry,* 38: 6380-6385.

Chiu, C. P.; Watts, A. G.; Lairson, L. L.; Gilbert, M.; Lim, D.; Wakarchuk, W. W.; Withers, S. G. and Strynadka, N. C. (2004) *Nature structural & molecular biology*, 11: 163-170.

Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. (1995) *J. Am. Chem. Soc. 117*: 5179-5197.

CPMD, Copyright IBM Corp 1990-2006, Copyright MPI für Festkörperforschung Stuttgart 1997-2001.

Cremer, D. & Pople, J. A. (1975) *J. Am. Chem. Soc.* 97: 1354-1358.

Crich, D. (2010) *Acc. Chem. Res.* 43: 1144-1153.

Davies, G. J. & Henrissat, B. (1995) *Structure 3*: 853-859.

Davies, G. J.; Dauter, M.; Brzozowski, A. M.; Bjrnvad, M. E.; Andersen, K. V. & Schlein, M. (1998) *Biochemistry* 37: 1926-1932.

Davies, G. J.; Gloster, T. M. & Henrissat, B. (2005). *Curr. Opin. Struct. Biol.* 15(6): 637–645.

Davies, G. J.; Mackenzie, L.; Varrot, A.; Dauter, M.; Brzozowski, A. M.; Schülein, M.; Withers, S. G. (1998) *Biochemistry*, 37: 11707-11713.

Davies, G. J.; Planas, A.; Rovira, C. (2012) *Acc. Chem. Res.* 45 (2): 308–316.

Davies, G. J.; Wilson, K. S. & Henrissat, B. (1997) *Biochem. J.* 321: 557-559.

Davis, B. G.; Hull, A.; Smith, C.; Nas, R. J.; Watson, A. A.; Winkler, D. A.; Griffiths, R. C.; Fleet, G. W. J. (1998) *Tetrahedron: Asymmetry* 9: 2947-2960.

Deslongchamps, P. (1975) *Tetrahedron* 31: 2463-2490.

Dixon, D. A. & Komornicki, A. (1990). *J. Phys. Chem.* 94: 5630-5636.

Dobson, C. M. (2004) *Seminars in Cell & Developmental Biology* 15: 3–16.

Dowd, M. K.; French, A. D.; Reilly, P. J. (1994) *Carbohydr. Res.* 264: 1-19.

Ducros, V.; Zechel, D. L.; Murshudov, G. N.; Gilbert, H. J.; Szabo, L.; Stoll, D.; Withers, S. G.; Davies, G. J. (2002) *Angew. Chem. Int. Ed*. 41: 2824-2827.

Ensing, B.; Laio, A.; Parrinello, M. & Klein, M. L. (2005) *J. Phys. Chem*. 109 : 6676-6687.

Ermert, P.; Vasella, A. (1993) *Carbohydr. Res.* 250: 113-128.

Errey, J. C.; Lee, S. S.; Gibson, R. P.; Fleites, C. M.; Barry, C. S.; Jung, P. M. J.; O'Sullivan, A. C.; Davis, B. G.; Davies, G. J. (2010) *Angew. Chem. Int. Ed*. 49:1234-1237.

Espinosa, J. F.; Montero, E.; Vian, A.; Garcia, J. L.; Dietrich, H.; Schmidt, R. R.; Martin-Lomas, M.; Imberty, A.; Canada, F. J.; Jimenez-Barbero, J. (1998) *J. Am. Chem. Soc. 120*: 1309-1318.

Fischer, E. (1894) *Ber. Dtsch. Chem. Ges*. 27: 2985–2993

Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R. (2004) GAUSSIAN 03, Gaussian Inc.: Wallingford, CT.

Fritz, T. A.; Hurley, J. H.; Trinh, L. B.; Shiloach, J.; Tabak, L. A. (2004) *Proc. Natl. Acad. Sci*. USA, 101: 15307-15312.

Fritz, T. A.; Raman, J.; Tabak, L. (2006) *J. Biol. Chem*. 281: 8613-8619.

Gamblin, S. J. & Skehel, J. J. (2010) *J. Biol. Chem.* 285: 28403-28409.

Garcia-Herrero, A.; Montero, E.; Munoz, J. L.; Espinosa, J. F.; Vian, A.; Garcia, J. L.; Asensio, J. L.; Canada, F. J.; Jimenez-Barbero, J. (2002) *J. Am. Chem. Soc.* 124: 4804-4810.

Gastinel, L. N.; Bignon, C.; Misra, A. K.; Hindsgaul, O.; Shaper, J. H. & Joziasse, D. H. (2001) *EMBO J.* 20: 638-649.

Gibson, R. P.; Turkenburg, J. P.; Charnock, S. J.; Lloyd, R. and Davies, G. J. (2002) *Chemistry & Biology*, 9: 1337-1347.

Gloster, T. M. & Davies, G. J. (2010) *Org. Biomol. Chem.* 8: 305-320.

Gloster, T. M. & Vocadlo D. J. (2012) *Nature Chemical Biology* 8: 683-694.

Gloster, T. M. (2012b) *Biochem. Soc. Trans.* 40: 913-928.

Gloster, T. M.; Meloncelli, P.; Stick, R. V.; Zechel, D. L.; Vasella, A. & Davies G. J. (2007) *J. Am. Chem. Soc.* 129: 2345-2354.

Gómez, H.; Lluch, J. M. & Masgrau, L. (2013) *J. Am. Chem. Soc.* 135: 7053-7063.

Gómez, H.; Polyak, I.; Thiel, W.; Lluch, J. M.; Masgrau, L. (2012) *J. Am. Chem. Soc.* 134: 4743-4752.

Gómez, H.; Rojas, R.; Patel, D.; Tabak, L. A.; Lluch, J. M.; Masgrau, L. (2014) *Org. Biomol. Chem.* 12: 2645-2655.

Griffiths, D. J. (1995) Introduction to Quantum Mechanics. Upper Saddle River, New Jersey: Prentice Hall. ISBN 0-13-124405-1.

Guerin, D. M. A.; Lascombe, M. B.; Costabel, M.; Souchon, H.; Lamzin, V.; Beguin, P.; Alzari, P. M. (2002) *J. Biol. Chem*. 316: 1061-1069.

Hagen, F. K.; Hazes, B.; Raffo, R.; deSa, D.; Tabak, L. A. (1999) *J. Biol. Chem*. 274: 6797-6803.

Hazes, B. (1996) *Protein Sci.* 5: 1490-1501.

He, Y.; Macauley, M. S.; Stubbs, K. A.; Vocadlo, D. J. & Davies, G. J. (2009) *J. Am. Chem. Soc.* 132: 1807-1809.

Heightman, T. D.; Vasella, A. T. (1999) *Angew. Chem. Int*. Ed. 38: 750-770.

Hendrickson, J. B. (1961) *J. Am. Chem. Soc.* 83: 4537–4547.

Hendrickson, J. B. (1967) *J. Am. Chem. Soc.* 89: 7047-7061.

Hockney, R. W. (1970) *Meth. Comp. Phys*., 9: 136‑211.

Hohenberg, P. & Kohn, W. (1964). *Phys. Rev. B,* 136 (3B): 864-871.

Hornak, V.; Simmerling, C. (2003) *Proteins* 51: 577-590.

Hrmova, M.; Varghese, J. N.; De Gori, R.; Smith, B. J.; Driguez, H. & Fincher, G. B. (2001) *Structure* 9: 1005-1016.

Hurtley, S. S. (2001) Carbohydrates and Glycobiology, Special Issue. *Science*, 2263-2503.

Iannuzzi, M.; Laio, A. & Parrinello, M. (2003) *Phys. Rev. Lett*. 90: 238302.

Imberty, A.; Piller, V.; Piller, F.; Breton, C. (1997) *Protein Eng.* 10: 1353-1356.

Ionescu, A. R.; Bérces, A.; Zgierski, M. Z.; Whitfield, D. M. & Nukada, T. (2005). *J. Phys. Chem. A* 109: 8096-8105.

Ireta, J.; Neugebauer, J.; Scheffler, M. (2004) *J. Phys. Chem. A*, 108: 5692-5698.

Isorna, P.; Polaina, J.; Latorre-Garcia, L.; Canada, F. J.; Gonzalez B. & Sanz-Aparicio, J. (2007) *J. Mol. Biol.* 371: 1204-1218.

IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). (1980). *Eur. J. Biochem.* 111: 295-298.

Jense, F. R.; Noyce, D. S.; Sederholm, C. H. & Berlin, A. J. (1960) *J. Am. Chem. Soc.* 84: 386-389.

Johnson, W. S.; Bauer, V. J.; Margrave, J. L.; Frisch, M. A.; Dreger, L. H. & Hubbard, W. N. (1961) *J. Am. Chem. Soc.* 83: 606-614.

Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey R. & Klein, M. (1983) *J. Chem. Phys.* 79: 926–935.

Kakhiani, K.; Lourderaj, U.; Hu, W.; Birney, D. & Hase, W. L. (2009). *J. Phys. Chem. A 113:* 4570-4580.

Karaveg, K.; Siriwardena, A.; Tempel, W.; Liu Z. J.; Glushka, J.; Wang, B. C.; Moremen, K. W. (2005) *J. Biol. Chem.* 280: 16197-16207.

Kirby, A. J. (1984). *Acc. Chem. Res.* 17: 305-311.

Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L. & Woods, R. J. (2008) *J. Comput. Chem.* 29: 622–655.

Koch, W., & Holthausen, M. (2000). *A Chemist's Guide de Density Functional Theory*. Weinheim, Germany: Wiley-VCH.

Kohn, W. & Sham, L. J. (1965). *Phys. Rev. A,* 140 (4A): 1133-1138.

Koshland, D. E. (1953) *Biol. Rev.* 28: 416-436.

Koshland, D. E. (1958) *Proc. Natl. Acad. Sci.* U S A. Feb 1958; 44(2): 98–104.

Kriegl, J. M.; Nienhaus, K.; Deng, P.; Fuchs, J. & Nienhaus, G. U. (2003). *Natl. Acad. Sci.* USA, 100: 7069-7074.

Kubota, T.; Shiva, T.; Sugioka, S.; Furukawa, S.; Sawaki, H.; Kato, R.; Wakatsuki, S.; Narimatsu, H. (2006) *J. Mol. Biol.* 359: 708-727.

Kuntz, D. A.; Liu, H.; Bols, M. & Rose, D. A. (2006) *Biocatalysis and Biotransformation* 24(1/2): 55-61.

Kuntz, D. A.; Tarling, C. A.; Withers, S. G.; Rose, D. R. (2008) *Biochemistry* 47: 10058-10068.

Laio, A. & Gervasio, F. L. (2008) *Reports on Progress in Physics*, 71: 126601.

Laio, A.; Parrinello, M. (2002a) *Proc. Natl. Acad. Sci. USA*, *99*: 12562-12566.

Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M. & Parrinello, M. (2005) *J. Phys. Chem. B,* 109: 6714-6721.

Laio, A.; VandeVondele, J. & Rothlisberger, U. (2002b) *J. Chem. Phys.* 116: 6941-6947.

Laio, A.; VandeVondele, J. & Rothlisberger, U. (2002c) *J. Phys. Chem. B,* 106: 7300-7307.

Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. (2008) *Annu. Rev. Biochem*. 77:521-555.

Lammerts van Bueren, A.; Ardèvol, A.; Fayers-Kerr, J.; Lou, Bo.; Zhang, Y.; Sollogoub, M.; Blériot, Y.; Rovira, C. & Davies G. J. (2010) *J. Am. Chem. Soc.* 132: 1804-1806.

Lau, K. S. & Dennis, J. W. (2008) *Glycobiology* 18:750-760.

Leach, A. R. (2001) Molecular Modeling. Principles and Applications. Prentice Hall. ISBN-10: 0582382106.

Lederkremer, G. Z. (2009) *Curr. Opin. Struct.* Biol. 19: 515-523.

Lee, C.; Yang, W. & Parr, R. G. (1988). *Phys. Rev. B,* 37: 785-789.

Lee, S. S.; Hong, S. Y.; Errey, J. C.; Izumi, A.; Davies, G. J.; Davis, B. G. (2011) *Nat. Chem. Biol*. 7: 631-638.

Leger, G. (1999) in *Iminosugars as glycosidase inhibitors.Nojirimycin and beyond*, ed. Stütz, A. E., Wiley-VCH, Weinheim: 31-67.

Lewis, E. S. & Boozer, C. E. (1952) *J. Am. Chem. Soc.* 74: 308-310.

Lin, I-C.; Coutinho-Neto, M. D.; Felsenheimer, C.; Anatole von Lilienfeld, O.; Tavernelli, I. and Rothlisberger, U. (2007) *Phys. Rev. B,* 75: 205131.

Lins, R. & Hünenberger, P. (2005) *J. Comput. Chem.* 26: 1400-1412.

Lira-Navarrete, E.; Iglesias-Fernández, J.; Zandberg, W. F.; Compañon, I.; Kong, Y.; Corzana, F.; Pinto, B. M.; Clausen, H.; Peregrina, J. M.; Vocadlo, D. J.; Rovira, C.; Hurtado-Guerrero, R. (2014) *Angew. Chem.* doi: 10.1002/ange.201402781.

Liu, H.; Liang, X.; Sohoel, H.; Bulow, A.; Bols, M., (2001) *J. Am. Chem. Soc.* 123: 5116-5117.

Lombard, V.; Ramulu, H. G.; Drula, E.; Coutinho, P. M. and Henrissat, B. (2014) *Nuc. Acids. Res.* 42: 490-495.

Lovering, A. L.; Lee, S. S.; Kim, Y. W.; Withers, S. G.; Strynadka, N. C. J. (2005) *J. Biol. Chem.* 280: 2105-2115.

Margrave, J. L.; Frisch, M. A.; Bautista, R. G.; Clarke, R. L. & Johnson, W. S. (1963) *J. Am. Chem. Soc.* 85: 546-548.

Marx, D. H. (2000) "Modern methods and algorithms of quantum chemistry". *NIC, FZ Jülich*.

McCarter, J. D. & Withers, S. G. (1994) *Curr. Opin. Struct. Biol.* 4: 885-892.

Molinari, M. (2007) *Nature Chemical Biology* 3: 313-320.

Monegal, A. & Planas, A. (2006) *J. Am. Chem. Soc.* 128: 16030-16031.

Money, V. A.; Smith, N. L.; Scaffidi, A.; Stick, R. V.; Gilbert, H. J.; Davies, G. J. (2006) *Angew. Chem. Int. Ed.* 45: 5136-5140.

Moremen, K. W. & Molinari M. (2006) *Curr. Opin. Struct. Biol.* 16: 592-599.

Murray, B. W.; Wittmann, V.; Burkart, M. D.; Hung, S. C. & C. H. Wong. (1997) *Biochemistry*, 36: 823-831.

Nosé, S. (1984) *Molecular Physics* 52: 255-268.

Notenboom, V.; Birsan, C.; Warren, R. A.; Withers, S. G.; Rose, D. R. (1998) *Biochemistry*, 37: 4751-4758.

Nukada, T.; Bérces, A.; Wang, L.; Zgierski, M. Z.; Whitfield, D. M. (2005) *Carbohydr. Res.* 340: 841-852.

Numao, S.; Kuntz, D. A.; Withers, S. G.; Rose, D. R. (2003) *J. Biol. Chem.* 278: 48074-48083.

Patenaude, S. I.; Seto, N. O. L.; Borisova, S. N.; Szpacenko, A.; Marcus, S. L.; Palcic, M. M. & Stephen V. E. (2002) *Nature Structural Biology,* 9: 685 – 690.

Pauling, L. (1946) *Chem. Eng. News* 24:1375-1377.

Pauling, L. (1948) *Nature,* 161: 707-709.

Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. (1995) *Comp. Phys. Commun.* 91: 141-168.

Perdew, J. P. (1986). *Phys. Rev. B,* 33: 8822-8824.

Perdew, J. P.; Burke, K. & Ernzerhof, M. (1996) *Phys. Rev. Lett.* 77: 3865-3868.

Persson, K.; Ly, H. D.; Dieckelmann, M.; Wakarchuk, W. W.; Withers, S. G.; Strynadka, N. C. (2001) *Nat. Struct. Biol.* 8: 166-175.

Petersen, L.; Ardevol, A.; Rovira, C.; Reilly, P. J. (2009) *J. Phys. Chem. B*, 113: 7331-7339.

Petersen, L.; Ardevol, A.; Rovira, C.; Reilly, P. J. (2010) *J. Am. Chem. Soc.* 132: 8291-8300.

Phillips, D. C. (1967). *Proc. Natl. Acad. Sci.* 57 (3): 484-495.

Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. (2005) *J. Comput. Chem. 26*: 1781-1802.

Rojas-Cervellera, V.; Ardèvol, A.; Boero, M.; Planas, A.; Rovira, C. (2013). *Chem. Eur. J.* 42: 14018–14023.

Rovira, C. (2005). The study of ligand-protein interaction by means of Density Functional Theory and first-principles molecular dynamics. En U. N. G., *Methods in molecular biology.* Humana Press, Inc.

Rovira, C. (2013) *WIREs Computational Molecular Science* ,3: 393-407.

Sabini, E.; Sulzenbacher, G.; Dauter, M.; Dauter, Z.; Jorgensen, P. L.; Schulein, M.; Dupont, C.; Davies, G. J.; Wilson, K. S. (1999) *Chem. Biol.* 6: 483-492.

Sachse, B. (1890) *Ber. Dtsch. Chem. 3:* 1.

Sandgren, M.; Berglund, G. I.; Shaw, A.; Stahlberg, J.; Kenne, L.; Desmet, T. & Mitchinson, C. (2004) *J. Mol. Biol.* 342: 1505-1517.

Schlick, T. (2000) Molecular Modeling and Simulation. An Interdisciplinary Guide: Springer. ISBN 978-0-387-95404-2.

Schramm, V. L. (2005) *Arch. Biochem. Biophys.* 433: 13-26.

Seen, H. M. & Thiel, W. (2009) *Angew. Chem.* 121: 1220-1254.

Sega, M.; Autieri, E. & Pederiva, F. (2009) *J. Chem. Phys*. 130: 225102(1-6).

Shah, N.; Kuntz, D. A. & Rose, D. R. (2003) *Biochemistry* 42: 13812-13816.

Sinnott, M. L. & Jencks, W. P. (1980) *J. Am. Chem. Soc*. 102: 2026-2032.

Soliman M. E.; Ruggiero, G. E.; Pernia, J. J.; Greig I. R. & Williams I. H. (2009) *Org. Biomol. Chem.* 7: 460-468.

Soya, N.; Fang, Y.; Palcic, M. M. & Klassen, J. (2011) *Glycobiology,* 21: 547-552.

Squiallacote, M.; Sheridan, R.; Chapman, O. L. & Anet, F. A. (1975) *J. Am. Chem. Soc.* 97:3244-3246.

Stortz, C. A. (2010) *J. Phys. Org. Chem*. 23: 1173-1186.

Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schülein, M.; Davies, G. J. (1996) *Biochemistry*, 35: 15280-15287.

Suzuki, R.; Fujimoto, Z.; Ito, S.; Kawahara, S.; Kaneko, S.; Taira, K.; Hasegawa, T.; Kuno, A. (2009) *J. Biochem.* 146:61-70.

Tailford, L. E.; Offen, W. A.; Smith, N.; Dumon, C.; Morland, C.; Gratien, J.; Heck, M. P.; Stick, R. V.; Bleriot, Y.; Vasella, A.; Gilbert, H. J.; Davies, G. J. (2008) *Nat. Chem. Biol.* 4: 306-312.

Taniguchi, N.; Honke, K. & Fukuda, M. (2002) *Handbook of Glycosyl-transferase and Related Genes.* Springer, Tokyo.

Taylor, E. J.; Goyal, A.; Guerreiro, C. I. P. D.; Prates, J. A. M.; Money, V. A.; Ferry, N.; Morland, C.; Planas, A.; Macdonald, J. A.; Stick, R. V.; Gilbert, H. J.; Fontes, C. M. G. A. & Davies, G. J. (2005) *J. Biol. Chem*. 280: 32761-32767.

Thompson, A. J.; Dabin, J.; Iglesias-Fernández, J.; Ardevol, A.; Dinev, Z.; Williams, S. J.; Bande, O.; Siriwardena, A.; Moreland, C.; Hu, T. C.; Smith, D. K.; Gilbert, H. J.; Rovira, C.; Davies, G. J. (2012) *Angew. Chem. Int. Ed.* 51: 10997-11001.

Tribello, G. A.; Bonomi, M.; Branduardi, D.; Caminolli, C.; Bussi, G. (2014) *Comp. Phys. Comm*. 185: 604-613.

Troullier, N. & Martins, J. L. (1991) *Condensed Matter* 43: 1993-2006.

Tvaroska, I. (2004) *Carbohydrate Research*, 339: 1007-1014.

Tvaroska, I.; Kozmon, S.; Wimmerová, M. & Koca, J. (2012) *J. Am. Chem. Soc.* 134: 15563-15571.

Vallée, F.; Lipari, F.; Yip, P.; Sleno, B.; Herscovics, A.; Howel P. L. (2000) *EMBO J*. 19: 581-588.

Verlet, L. (1967). *Physical Review*, 159: 98-103.

Vocadlo, D. J.; Davies, G. J. (2008) *Curr. Opin. Chem. Biol*. 12: 539-555.

von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U. & Sebastini, D. (2005) *J. Chem. Phys*. 122: 14113.

Vrielink, A.; Ruger, W.; Driessen, H. P. and Freemont, P. S. (1994) *EMBO J.* 13: 3413-3422.

Wandall, H. H.; Hassan, H.; Mirgorodskaya, E.; Kristensen, A. K.; Roepstorff, P.; Bennett, E. P.; Nielsen, P. A.; Hollingsworth, M. A.; Burchell, J.; Taylor-Papadimitriou, J.; Clausen, H. (1997) *The Journal of Biological chemistry*, 272: 23503-23514.

Warshel, A. & Levitt, M. (1976) *J. Mol. Biol.* 103: 227-249.

Weinan, E.; Weiqing, R.; Vanden-Eijnden, E. (2002) *Phys. Rev. B.* 66: 052301.

Whitfield, D. M. (2007) *Carbohydr. Res. 342*: 1726.

Williams, R. J.; Iglesias-Fernández, J.; Stepper, J.; Jackson, A.; Thompson, A. J.; Lowe, E. C.; White, J. M.; Gilbert, H. J.; Rovira, C.; Gideon, G. J. & Williams, S. J. (2014) *Angew. Chem. Int. Ed.* 53: 1087-1091.

Woods, R. J.; Andrews, C. W.; Bowen, J. P. (1992) *J. Am. Chem. Soc.* 114: 859-864.

Zechel, D. L. & Withers, S. G. (2000) *Acc. Chem. Res.*33: 11-18.

Zhang, Y. K.; Lee T. S. and Yang W. T. (1999) *J. Chem. Phys.* 110: 46-54.

Zhao, Y. & Truhlar, D. G. (2004) *J. Phys. Chem. A*, 108: 6908-6918.

Zhong, W.; Kuntz, D. A.; Ember, B; Singh, H.; Moremen, K. W.; Rose, D. R. & Boons G-J. (2008) *J. Am. Chem. Soc.* 130: 8975-8983.

Zhu, Y.; Suits, D. L.; Thompson, A. J.; Chavan, S.; Dinev, Z.; Dumon, C.; Smith, N.; Moreman, K.; Xiang, Y.; Siriwardena, A.; Williams, S. J.; Gilbert, H. J.; Davies, G. J. (2010) *Nat. Chem. Biol*. 6: 125-132.

Zou, J. -Y.; Kleywegt, G. J.; Stahlberg, J.; Driguez, H.; Nerinckx, W.; Claeyssens, M.; Koivula, A.; Teeri, T. T.; Jones, T. A. (1999) *Structure,* 7: 1035-1045.