Conformational and mechanistic analyses of α- and β-glycosidase substrates by *ab initio* QM/MM methods

Santiago Alonso Gil

UNIVERSITAT DE BARCELONA

FACULTAT DE QUÍMICA

PROGRAMA DE DOCTORAT EN QUÍMICA TEÒRICA I COMPUTACIONAL

Títol de la tesi

Conformational and mechanistic analyses of α- and β-glycosidase substrates by *ab initio* QM/MM methods

Memòria presentada per **Santiago Alonso Gil** per optar al títol de doctor per la Universitat de Barcelona.

Noms, cognoms i signatura del doctorand:

Santiago Alonso Gil

Noms, cognoms i afiliació de la Directora:

Dra. Carme Rovira Virgili

Departament de Química Inorgànica i Orgànica (Secció de Química Orgànica), Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona i Institució Catalana de Recerca i Estudis Avançats (ICREA)

Noms, cognoms i afiliació del tutor:

Dr. Jaime Rubio Martínez

Departament de Química Física i Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona

Acknowledgements

Parecía que no iba a llegar nunca este momento. Son tantas las emociones que me embargan ahora que estoy dando mis últimas pinceladas a la tesis y miro hacia atrás, asombrado. Podría escribir una saga thriller al más estilo Dan Brown sobre estos años, pero prefiero dedicarme a agradecer la fuerza que he recibido de mi entorno. Sí, no ha sido un camino de rosas, pero, nadie me dijo que fuera a serlo. Sobre todo, cuando las cosas no salen jajaja. Seguramente, alguien dirá que viene implícito en la responsabilidad del título que recibimos. Muchos pensaran que es el precio que debemos pagar. Sea como sea, he llegado a canasta, seguro de anotar.

Mi máximo apoyo, mi muleta cuando ninguna de mis dos piernas quería andar, mi abanico en pleno verano, mama, papa, muchas gracias por estar SIEMPRE a mi lado, confiando en mis posibilidades y apretándome cuando quería tirar la toalla. Mi abuela, mis tíos, mis primos, a mi Vera, doy gracias de tener una familia que me quiere tanto.

A mis amigos, ¿qué les puedo decir? Soy el hombre más afortunado por teneros. Alberto y David habéis estado por mí, en la facultad uno y el otro casi para todo lo demás. Habéis sido mis mayores confidentes y os habéis ganado el paraíso por ello. Mi círculo de físicos y otros frikis no sería nada sin Carlos, Oso, Raúl, Maneu, Óscar, Iria, Gibi, Adri, Sergi (mis refugees)… ¡La de patentes que se perderán! No me olvido de mis raíces en Vilafranca, Aitor, Josué, David, Xavi y Santi, siempre alegres y orgullosos de mí, no se puede tener mejores hermanos.

Agradecer la oportunidad y la tutela de Carme, tanto como la ayuda de Javi y Víctor cuando estaba dando mis primeros pasos con la amilosucrasa. Los días enteros en el despacho con Lluís, Alba, Joan, Mertxe y Binju.

La vida es pleno cambio y no me olvido de los que se han ido a un lugar mejor, Quico. Ni de las personas que me han dado su corazón y su apoyo, Irina y Paula. Madres y tías postizas, Graciela, Àngela y Ruth, como esa parte de Argentina que siempre piensa en mí.

Para amor incondicional, el que he recibido de Xochilt, Bea, Alba, Tània, Marc, Lorena, Heli, Vicky, Marta, Gisela, Fran, Quim, Pablo, Elsa, Dani, Pipeta, Bela, Jesús, Miriam, Sojo, Patxi, Javi, Mario… y de mis compañeros de máster (todos doctores jeje). No me olvido de mi gente del Tivoli, mis Blunts y Panxing. Y menos de mi peña de Torredarques, ¡este año arderá el pueblo! A Ramón (y el recuerdo de Isabel) y a Lluïs del Pau Boada. A mi maestro saxofonista, Àngel y a Antonio y mis compañeros de amor hacia la escritura. A mis queridos alumnos de los laboratorios de orgánica, de los que he aprendido mucho a disfrutar enseñando.

¡Muchas, muchas gracias a todos!

Collaborations

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

- **Dr. Isabelle André**
  Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, INSA, Toulouse, France.

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

- **Prof. Spencer J. Williams**
  School of Chemistry, University of Melbourne, Australia.

- **Prof. Maria Hrmova**
  Australian Center for Plant Functional Genomics, University of Adelaide, Australia.

- **Dr. Antoni Planas**
  Department of Bioengineering. IQS-Universitat Ramón Llull, Barcelona, Spain.

Fellowships

Ayuda a la Formación de Personal Investigador 2012 (FPI 2012), Ministerio de Educación y Ciencia, Gobierno de España.

Project reference: CTQ2011-25871

FPI 2012 reference: BES-2012-051782

Symbols and acronyms

6-membered rings	7-membered rings
B – boat	**B** – boat
C – chair	**BS** – boat-sofa
E – envelope	**C** – chair
H – half-chair	**S** – sofa
S – skew-boat	**TB** – twisted-boat
	TC – twisted-chair
	TS – twisted-sofa

-1 sugar molecule/ring/moiety – the sucar saccharide located in the -1 subsite

1PE – pentaethylene glycol

Å – Angstrom

α-D-gulosepta – α-D-glycero-D-guloseptanoside

α-D-idosepta – α-D-glycero-D-idooseptanoside

ALPH – Antiperiplanar Lone-Pair Hypothesis

a.m.u. – atomic mass units

AS – amylosucrase

a.u. – atomic units

β-D-gulosepta – β-D-glycero-D-guloseptanoside

β-L-idosepta – β-L-idoseptanoside

BOMD – Born-Oppenheimer molecular dynamics

BSSE – basis set superposition error

C_1 – anomeric carbon

C. O. M. – center of mass

CPU – central processing unit

CAZy – Carbohydrate Active enZymes database

CI (CI_1, CI_2) – covalent intermediate (see GEI)

CP – Car-Parrinello method

*Cp*GH125 – *Clostridium perfringens* GH125 exo-1,6-α-mannosidase

CPMD – Car-Parrinello molecular dynamics

CV – collective variable

DFT – density functional theory

ΔG – variation of free energy

ESP/RESP – (Restrained) electrostatic potential derived charges

FEL – free energy landscape

ff – force-field

g – gram

GEI – glycosil-enzyme intermediate

GGA – generalized gradient approximation

GH – glycoside hydrolase

GOL – glycerol

GROMOS – GROningen Molecular Simulation

GT – glycoside transferase

GTO – Gaussian-type orbitals

HIV – human immunodeficiency virus

HOMO/LUMO – highest occupied/lowest unoccupied molecular orbital

HvExoI – *Hordeum vulgare* GH3 exo-1,3-1,4-β-D-glucanase

IRC – intrinsic reaction coordinate

IUPAC – International Union of Pure and Applied Chemistry

K – kelvin

kcal/mol – kilocalories per mole

KIE – kinetic isotope effect

KS – Kohn-Sham

LDA – local density approximation

m (cm, nm) – meter (centimetre, nanometer)

MB-O – natural mannobiose

MB-S – thio-mannobiose

MC – Michaelis complex

MD – molecular dynamics

MFEP – minimum free energy pathway

MM – molecular mechanics

MTD - metadynamics

NEB – nudged elastic band

*Np*AS – *Neisseria polysaccharea* amylosucrase

NPT/NVT – thermodynamic statistics ensembles with number of atoms (N), pressure (P) or volume (V) and temperature (T) remaining constant.

O_5 / O_p – pyranic oxygen

O_g – glycosidic oxygen

OI (OI^+, OInp, OIp) – oxocarbenium ion (positively charged, non-protonated, protonated)

ONIOM – own N-layered integrated orbital and molecular mechanics

Oxoseptaion – oxocarbenium septanoside cation

P - products

PBE – Perdew, Becke and Ernzerhoff

PDB – protein data bank

PEL – potential energy landscape

pNP – para-nitrophenyl

PW – plane waves

QM – quantum mechanics

QM/MM – hybrid quantum mechanics/molecular mechanics method

R – reactants

RMSD – root-mean-square deviation

s (ns, ps, fs) – second (nanosecond, picosecond, femtosecond)

S_N2 – bimolecular nucleophilic substitution

TS – transition state

Well-T – well-tempered

WT – wild-type

Contents

Outline

Carbohydrates are one of the most important biochemical molecules due to their role as energetic resource for non-photosynthetic organisms, cell-cell recognition and adhesion, protection of cell membranes of bacteria and plants and ensure the proper functionality of several enzymes. Glycoproteins are important components of cell surfaces and the extracellular medium where molecular-cell interactions take place. Defects in the activity of glycoproteins are the cause of several human diseases, such as diabetes and lysosomal storage diseases.

Glycoside hydrolases (GHs) are one of the most relevant glycoproteins. They catalyse the cleavage of glycosidic bonds of oligosaccharides to generate smaller oligosaccharides or monosaccharides with relevant biological roles. To know the proper way to modulate the activity of these enzymes, it is crucial to understand their molecular mechanisms of action. Sugars are very flexible molecules, most of them formed by 6-membered rings whose conformation changes during the reaction catalysed by GHs. Capturing these conformations, in particular the one at the reaction transition state, is important in inhibitor design. Experiments aimed at characterizing (indirectly) the transition state conformation of GHs include crystallizing the Michaelis complex (MC), which usually is done using thioglycoside derivatives of the natural substrate (i.e. the glycosidic oxygen is substituted by sulfur) or enzyme mutants (e.g. the catalytic acid/base residue mutated to glutamine or asparagine). However, in some cases, the degree of resemblance of the sugar conformation in thioglycosides or enzyme mutants with respect to the natural/wild type counterparts unclear.

In last years, our group has demonstrated that the conformational free energy landscape (FEL) of natural sugars (e.g. β-glucose, α,β-mannose and β-xylose) can be used to predict catalytic pathways of GHs. The conformational FEL is explored using Cremer & Pople puckering coordinates as collective variables in the metadynamics method. A natural extension of these studies is to explore the conformational FEL of 7-membered sugar rings (septanosides), which are currently the focus of great interest as potential GH substrates.

In this Thesis, we applied Car-Parrinello molecular dynamics methods, within the QM/MM approach, to elucidate the catalytic mechanism of a family 13 retaining GH, (amylosucrase) and a family 125 inverting GH (exo-1,6-α-mannosidase). In both cases, prior structural information was available from enzyme mutants and/or thioderivative substrates. In addition, we have extended previous ring conformational analyses of pyranoses to septanosides in order to assess their potential as new substrates of selected GHs.

The Thesis is organised as follows:

Chapters I and II focuses on the conformational flexibility of carbohydrates, GH mechanisms and experimental techniques aimed to trap the MC, as well as the methodology used in this Thesis.

In Chapter III, we investigate the conformational landscape of α-glucose and the conformational itinerary that an α-glucoside (fructose) follows during catalysis by GH13 amylosucrase (both hydrolysis and polymerization have been investigated).

In Chapter IV, we present the results of conformational study of a thioglycoside *vs* the natural substrate in a GH125 α-mannosidase, including a mechanistic analysis that uncovered the conformational catalytic itinerary for family 125 GHs.

In Chapter V, we investigate four MC structures of a promiscuous GH3 β-glucohydrolase, able to cleave several types of glycosidic linkages, for which only structures with thioglycosides are available. The reconstruction of the structure with the natural substrate allowed us to predict the mechanism of action of these enzymes.

In Chapter VI, we apply the Cremer & Pople puckering coordinates for 7-membered rings as collective variable for metadynamics on several septanoside molecules (both isolated and "on-enzyme") and assess their potential as substrate/inhibitors of GHs.

Finally, Chapter VII contains the main conclusions of this work.

Chapter I – Introduction

Introduction

1. Carbohydrates. Classification and structure

Carbohydrates – or saccharides – are biomolecules formed basically by carbon, hydrogen and oxygen, being the most abundant biological structure on Earth, whose paper is not uniquely the obtaining of energy. Among their roles include their participation in cell-cell recognition and adhesion, the protection of the cell walls of bacteria and plants, tissues of animals, their attaching in the surface of an enzyme to ensure its proper activity and, finally, as mentioned previously, they are the main energy resource for non-photosynthetic cells.

These can be classified by their size: monosaccharides, oligosaccharides and polysaccharides. Monosaccharides are simple sugars which consist of a polyhydroxy aldehyde or ketone unit. The most abundant one in nature is the D-glucose, whose importance in the good behaviour of the human brain is well-known. Oligosaccharides are chains of less than 20 monosaccharides, whose joint between them is a characteristic linkage called glycosidic bond. One of the most known and sweet oligosaccharide is sucrose, a disaccharide of α-D-glucose and fructose. In the last position, larger polymers of monosaccharides are known as polysaccharides, commonly with a bad solubility in water. A good example of this last group is the cellulose present in the plant world, being a chain of β-D-glucose monosaccharides.

In a physiological point of view, sugars are in aqueous solution. In that conditions, monosaccharides with five or six carbon atoms tend to form cyclic structures (Figure 1.1). This process is done thanks to the reaction between the aldehyde or ketone unit with an alcohol group of the same molecule to form a hemiacetal or hemiketal, respectively. The stereochemistry of the attacked carbon (anomeric carbon) will distinguish the α from the β anomers, which are interchangeable in aqueous solution (mutarotation process, Figure 1.1). This equilibrium mixture generally do not contain equal quantities of both shapes.

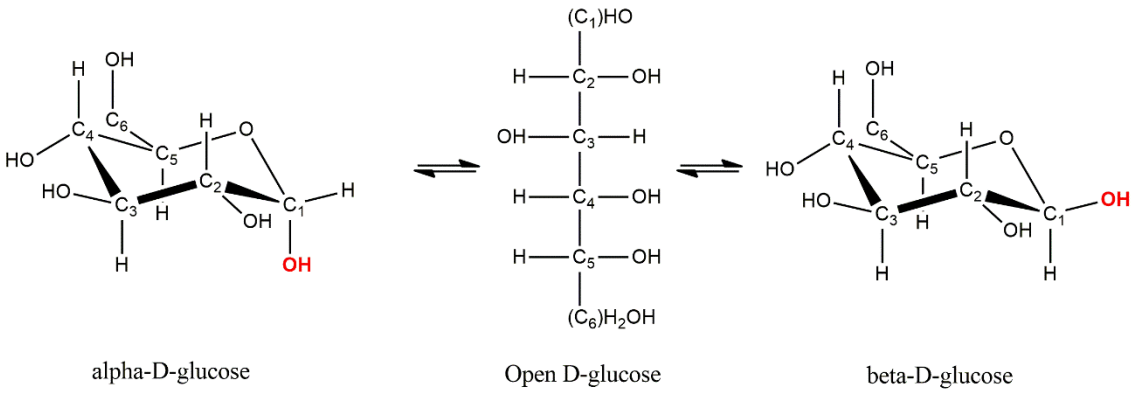


Figure 1.1. Mutarotation process: Formation of the two cyclic forms of D-glucose.

Apart from natural sugars, in the last decade, several synthetic strategies has been developed to expand the sugar rings. The resulting monosaccharides are known as septanosides (7-membered rings). In this field, the M. W. Peczuh research group is a guide to know the way is following the investigation in this synthesized sugars. In 2003, they reported the synthesis of a 7-membered ring from the D-glucose (Peczuh, 2003) and, in 2005, they presented a work to synthesize the α-D-glycero-D-idoseptanoside and the β-D-glycero-D-guloseptanoside (DeMatteo, 2005), which summarized process is shown in Figure 1.2.

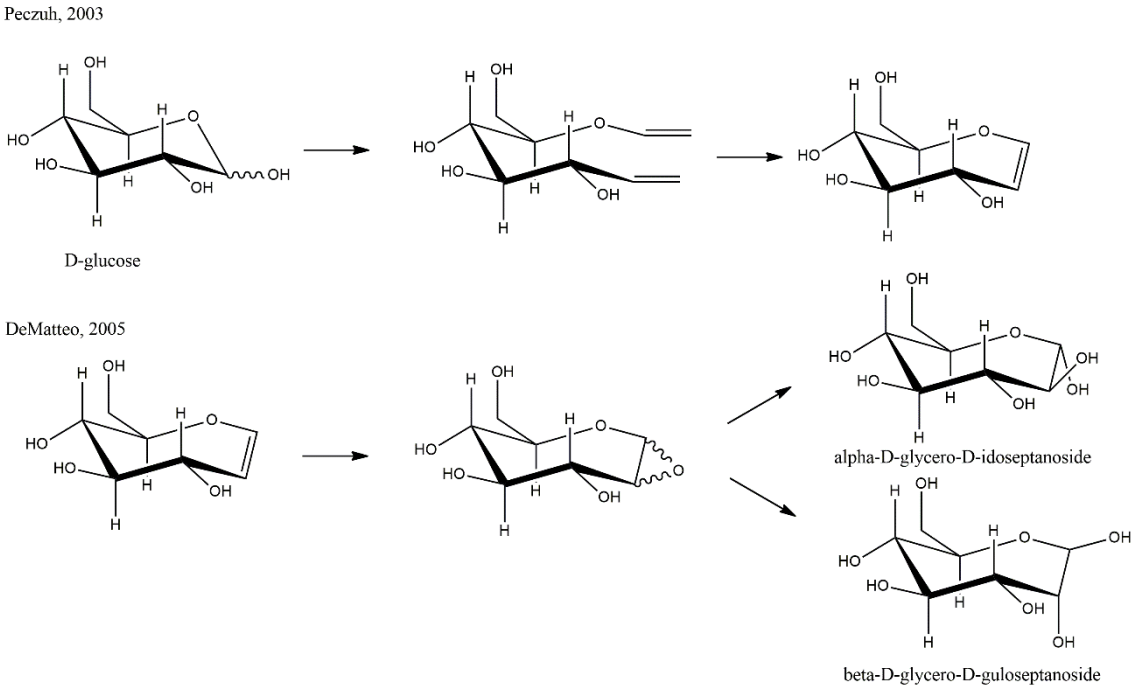


Figure 1.2. Summarized synthetic pathway from D-glucose to α-D-glycero-D-idoseptanoside and β-D-glycero-D-guloseptanoside (Peczuh, 2003) (DeMatteo, 2005).

The applicability of septanosides in biological procedures is already a work in progress. In 2005, the Peczuh group demonstrate the septanoside recognition of Concavalin A. In 2006, the kinetic activity of glucosidases to break para-nitrophenyl-L-septanosides was reported (Tauss, 2006). The activity decreased in two orders of magnitude respect the natural substrate. Nowadays, the question in the air is: Can be septanosides good substrates for glycoside hydrolases? Or are they just good inhibitors?

2. Conformations of pyranose rings

Pyranose rings are 6-membered very flexible rings consisting of five carbon atoms and one oxygen atom that can adopt a wide range of conformations.

In 1980, the IUPAC regulated the nomenclature to define the different conformations adopted by a 6-membered ring (IUPAC, 1980). There are 38 possible conformations grouped into five types: chair (C), boat (B), skew-boat (S), half-chair (H) and envelope (E) (Figure 1.3). The way to distinguish the different conformations inside a type is defining a reference plane formed by the maximum number of ring atoms (except envelope with five atoms, the rest contain four atoms in this plane). The numerals associated with each conformation (e.g. $^{4}C_{1}$) 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).

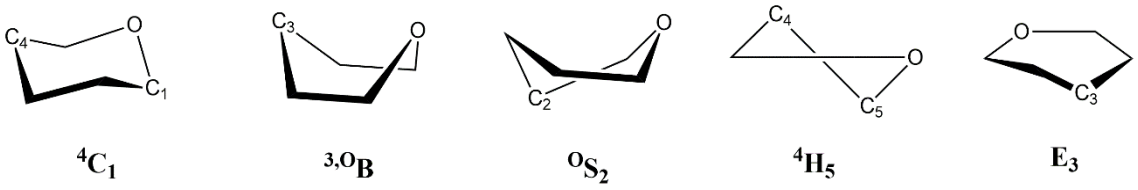


Figure 1.3. Nomenclature for a few different conformations of a pyranose ring.

- **Chair**: The reference plane is defined by two parallel ring sides, with the lowest numbered carbon atom located exoplanar. Possible conformations: $^{1}C_{4}$ and $^{4}C_{1}$.
- **Boat**: The reference plane is also defined by two paralle ring sides, but the two exoplanar atoms are on the same side of the reference plane. Possible conformations: $^{1,4}B$, $B_{1,4}$, $^{3,O}B$, $B_{3,O}$, $^{2,5}B$ and $B_{2,5}$.

- **Skew-boat**: The reference plane is formed by three adjacent atoms and one remaining non-adjacent atom separating the atoms above and below the plane. Possible conformations: $^{1}S_{3}$, $^{3}S_{1}$, $^{1}S_{5}$, $^{5}S_{1}$, $^{O}S_{2}$ and $^{2}S_{O}$.
- **Half-chair**: The reference plane is formed by four adjacent atoms. The two remaining atoms are located above and below the plane. Possible conformations: $^{O}H_{1}$, $^{1}H_{O}$, $^{1}H_{2}$, $^{2}H_{1}$, $^{2}H_{3}$, $^{3}H_{2}$, $^{3}H_{4}$, $^{4}H_{3}$, $^{4}H_{5}$, $^{5}H_{4}$, $^{5}H_{O}$ and $^{O}H_{5}$.
- **Envelope**: The reference plane is formed by five adjacent atoms. The remaining atom can be above or below the plane. Possible conformations: ^{O}E, E_{O}, ^{1}E, E_{1}, ^{2}E, E_{2}, ^{3}E, E_{3}, ^{4}E, E_{4}, ^{5}E and E_{5}.

2.1. Conformational diagrams

At time to differentiate mathematically the whole conformational space of 6-membered ring, we address to the Cremer and Pople spherical puckering coordinates (Cremer, 1975) to represent the 38 possible conformations. The chair conformations are represented at both poles of the sphere, boats and skew-boats in the equator and half-chairs and envelopes in the tropic regions (Figure 1.4, center). Due to the difficulty to represent and analyse a 3D sphere, two 2D projections are commonly used to smooth the complicated job to show the conformational behaviour of a 6-membered ring.

On the one hand, the Stoddart or Cartesian diagrams (Figure 1.4, left) are often used due to they easily inform of the possible conformational routes that might be followed by a pyranose ring. On the other hand, the Mercator projection (Figure 1.4, right), analogous to the cartographic maps of the Earth, may also be used.

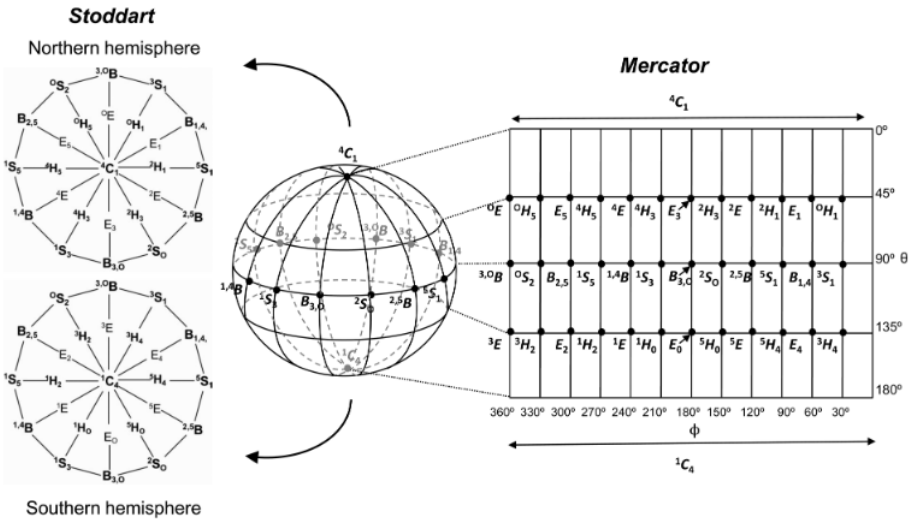


Figure 1.4. Main conformations on the Cremer and Pople sphere along with two of their most used representations: Stoddart diagram (Northern and Southern polar projections are shown) and Mercator projection. Figure taken from reference (Davies, 2012).

3. Conformations of septanose rings

Talking about complexity, the conformational space of a 6-membered ring seems a kid game compared with the shape worthiness of a septanose-like ring. Analogously to the hexose case, the septanose ring contains six carbon atoms and one oxygen atom in its structure. The nomenclature used in the present Thesis to define the different conformations of 7-membered rings is the same as followed in the reference (DeMatteo, 2005). There are 98 conformations grouped in seven types: twisted-chair (TC), chair (C), twisted-boat (TB), boat (B), sofa (S), twisted-sofa (TS) and boat-sofa (BS) (Figure 1.5). Analogously to the hexose nomenclature, a reference plane is defined by the maximum coplanar atoms in the ring (exception on TC and TB nomenclature, explained below and shown in Figure 1.6). The numerals associated with each conformation (e. g. $^{4,5}TC_{6,O}$) 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).

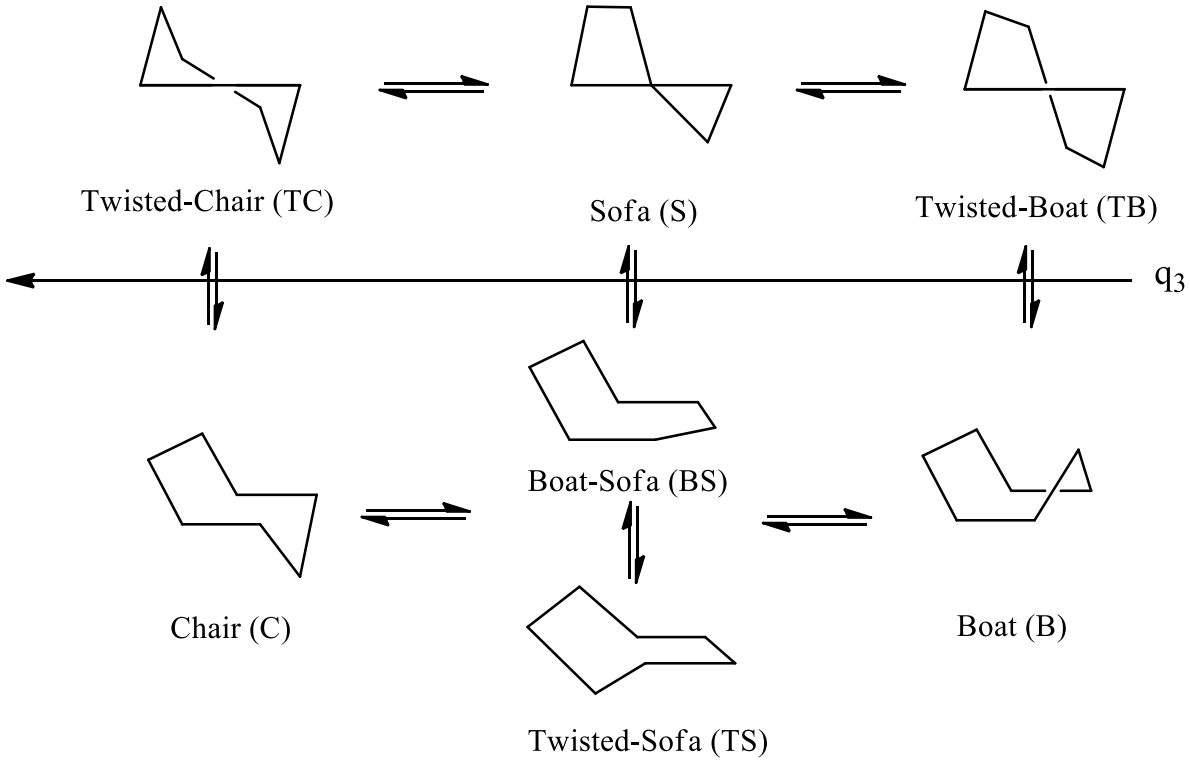


Figure 1.5. Nomenclature and equilibria of the different conformations of a 7-membered ring. Each column of conformations coexist in a conformational plane depending of the value of the puckering coordinate q_3.

- **Twisted-Chair**: The reference plane is formed by three adjacent atoms and the other four atoms form an S shape, two above and two below the plane (Figure 1.6, left). Possible conformations: $^{O,1}TC_{2,3}$; $^{3,2}TC_{1,O}$; $^{1,2}TC_{3,4}$; $^{4,3}TC_{2,1}$; $^{2,3}TC_{4,5}$; $^{5,4}TC_{3,2}$; $^{3,4}TC_{5,6}$; $^{6,5}TC_{4,3}$; $^{4,5}TC_{6,O}$; $^{O,6}TC_{4,5}$; $^{5,6}TC_{O,1}$; $^{1,O}TC_{6,5}$; $^{6,O}TC_{1,2}$ and $^{1,2}TC_{O,6}$.
- **Chair**: The reference plane is formed by two parallel ring sides. Two adjacent atoms and a third exoplanar atom can be above or below the plane, but not both groups in the same side. Possible conformations: $^{O,1}C_4$; $^{4}C_{O,1}$; $^{1,2}C_5$; $^{5}C_{1,2}$; $^{2,3}C_6$; $^{6}C_{2,3}$; $^{3,4}C_O$; $^{O}C_{3,4}$; $^{4,5}C_1$; $^{1}C_{4,5}$; $^{5,6}C_2$; $^{2}C_{5,6}$; $^{6,O}C_3$ and $^{3}C_{6,O}$.
- **Twisted-Boat**: As TC, the reference plane is formed by three adjacent atoms and the other four atoms form a Z shape, two above and two below the plane (Figure 1.6, right). Possible conformations: $^{O,1}TB_{2,3}$; $^{3,2}TB_{1,O}$; $^{1,2}TB_{3,4}$; $^{4,3}TB_{2,1}$; $^{2,3}TB_{4,5}$; $^{5,4}TB_{3,2}$; $^{3,4}TB_{5,6}$; $^{6,5}TB_{4,3}$; $^{4,5}TB_{6,O}$; $^{O,6}TB_{4,5}$; $^{5,6}TB_{O,1}$; $^{1,O}TB_{6,5}$; $^{6,O}TB_{1,2}$ and $^{1,2}TB_{O,6}$.
- **Boat**: As C, the reference plane is formed by two parallel ring sides. Two adjacent atoms and a third exoplanar atom can be above or below the plane, always both groups in the same side. Possible conformations: $^{(O,1),4}B$; $B_{(O,1),4}$; $^{(1,2),5}B$; $B_{(1,2),5}$; $^{(2,3),6}B$; $B_{(2,3),6}$; $^{O,(3,4)}B$; $B_{O,(3,4)}$; $^{1,(4,5)}B$; $B_{1,(4,5)}$; $^{2,(5,6)}B_2$; $B_{2,(5,6)}$; $^{3,(6,O)}B$ and $B_{3,(6,O)}$.
- **Sofa**: The reference plane is formed by three adjacent atoms and one remaining non-adjacent atom separating the atoms above and below the plane. Possible conformations: $^{O,1}S_3$; $^{3}S_{1,O}$; $^{1,O}S_5$; $^{5}S_{O,1}$; $^{O,6}S_4$; $^{4}S_{6,O}$; $^{6,O}S_2$; $^{2}S_{6,O}$; $^{6,5}S_3$; $^{3}S_{5,6}$; $^{4,3}S_1$; $^{1}S_{3,4}$; $^{5,4}S_2$ and $^{2}S_{4,5}$.
- **Twisted-Sofa**: Similar to the H conformation, the reference plane is formed by four adjacent atoms. Three adjacent atoms lie outside the plane, two adjacent atoms are above or below the plane and the third is in the opposite side. Possible conformations: $^{O,1}TS_2$; $^{2}TS_{1,O}$; $^{1,O}TS_6$; $^{6}TS_{O,1}$; $^{O,6}TS_5$; $^{5}TS_{6,O}$; $^{1,2}TS_3$; $^{3}TS_{2,1}$; $^{6,5}TS_4$; $^{4}TS_{5,6}$; $^{4,3}TS_2$; $^{2}TS_{3,4}$; $^{5,4}TS_3$ and $^{3}TS_{4,5}$.
- **Boat-Sofa**: Similar to the E conformation, the reference plane is formed by five adjacent atoms and the remaining two atoms are above or below the plane. Possible conformations: $^{O,1}BS$; $BS_{O,1}$; $^{1,2}BS$; $BS_{1,2}$; $^{2,3}BS$; $BS_{2,3}$; $^{3,4}BS$; $BS_{3,4}$; $^{4,5}BS$; $BS_{4,5}$; $^{5,6}BS$; $BS_{5,6}$; $^{6,O}BS$ and $BS_{6,O}$.

$^{3,4}TC_{5,6}$ $^{6,O}TB_{1,2}$

Figure 1.6. Nomenclature of the Twisted-Chair and Twisted-Boat septanose conformation used in the reference (DeMatteo, 2005).

3.1. Conformational diagrams

As can be supposed, the conformational space to represent 98 different conformation is a mess. There are needed four puckering coordinates (two radii and two angles) for differentiate the whole conformational behaviour of a 7-membered ring. Fortunately, the value of one of the radii depends on the value of the other (q_2 vs. q_3), and it is possible to visualize all the conformations in a 3D space. Furthermore, as mentioned in Figure 1.5 caption, groups of conformations are present in planes with the same value of q_3.

As shown in Figure 1.7, over the plane q_3 = 0.65, twisted-chair and chair conformations are coexisting being in an equilibrium TC – C – TC. Over the plane q_3 = 0.05, twisted-boat and boat conformations are coexisting being in an equilibrium TB – B – TB. Finally, over the plane q_3 = 0.35; sofa, twisted-sofa and boat-sofa conformations are coexisting in a complex equilibria net.

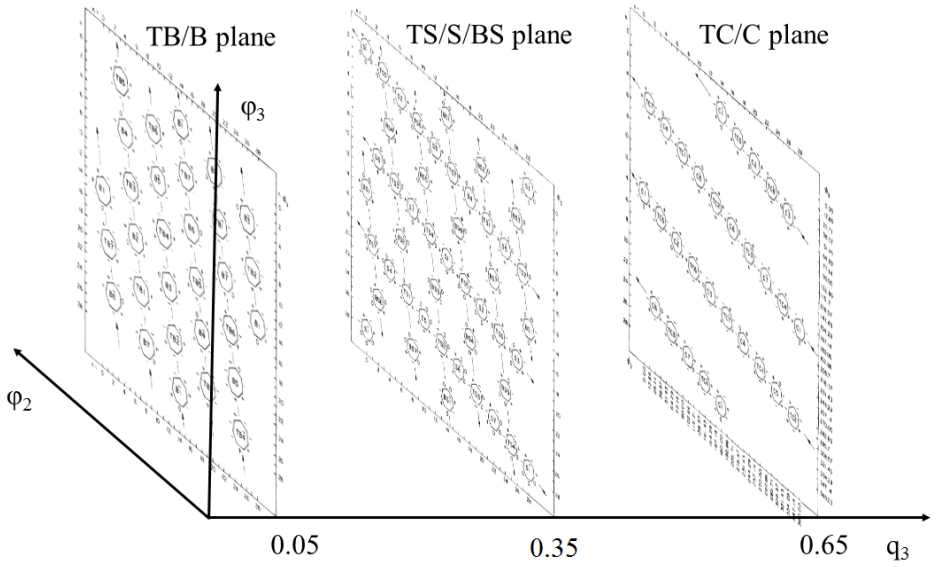


Figure 1.7. Schematic representation of the 3D conformational space of a 7-membered ring. The conformational planes along the q_3 puckering coordinate (Taken from the reference [Boessenkool, 1980]) represent the (φ_2, φ_3) subspaces of the TC/C, TB/B and TS/S/BS planes.

4. Glycoside Hydrolases (GHs) or glycosidases

Before going into matter, sugars and their interaction with glycoside hydrolases are very important to be studied with detail due to the essential role that play in a wide range of biological processes, such as control of protein folding and viral infections (Hurtley, 2001). For example, their metabolism is crucial in the understanding of diseases as diabetes type II (α-glucosidases [Bischoff, 1995] and α-mannosidases [Chui, 1997]) and viral infections, such as HIV or influenza (Gloster, 2007).

Glycoside hydrolases (GHs) represent a group of enzymes whose catalytic activity consists on the degradation or modification of polysaccharides and glycoconjugates throughout the cleavage of the glycosidic bond (Davies, 1995). Nowadays, 135 families of GHs are characterized according to their primary amino acid sequence identity in the Carbohydrate Acting enZymes database (www.cazy.org) (Henrissat, 1996) (Cantarel, 2009). Furthermore, some GHs families can be classified by their tertiary structure in clans (from GH-A to GH-N). Other classification can be by the mechanism of action. Enzymes that belong to the same family, usually have a similar tertiary structure and, as a consequence, a similar mechanism. According to their activity, GHs can also be classified into retaining and inverting GHs (mechanisms shown in Figure 1.9).

It is well-known that enzymes are amino acid chains whose folding has evolved during thousands of years in order to improve their catalytic function. This catalytic function, usually depends on the shape and the distribution of amino acids of an active site. This active site must be design to bind perfectly the substrate to be chemically changed and the catalytic residues and auxiliary molecules (metal ions, phosphates, etc.) must be distributed to attack properly the substrate. In this way, GHs can have three kinds of active site topology, due to the position of the glycosidic bond they have to cleave (Figure 1.8).

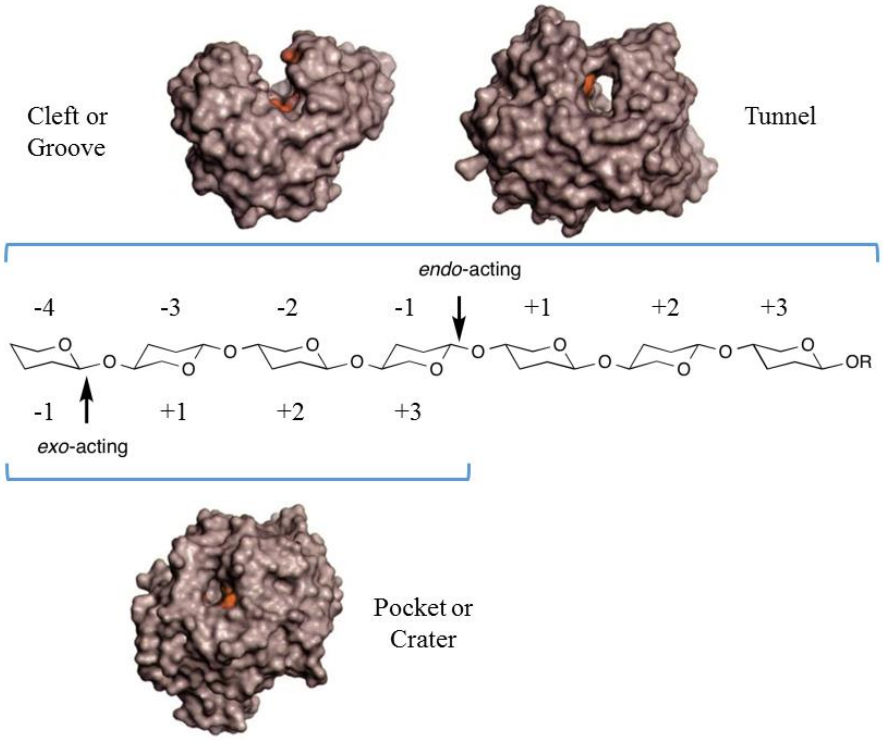


Figure 1.8. Dependence of the topological structure of the active site of GHs (Figures adapted from the reference [Davies, 1995]) and the position of the glycosidic bond to be cleave (Figure adapted from the reference [www.cazy.org]).

Supposing a hypothetical case of a seven sugars oligosaccharide (Figure 1.8, center), a typical strategy to obtain monosaccharides is a two steps reaction. First, a GH able to break glycosidic bonds in the middle of the chain, known as *endo*-GH, divide the chain in a 3 and a 4 sugar oligosaccharides. Then, a GH able to break glycosidic bonds in the top of the chain, known as *exo*-GH, obtain the appreciate monosaccharide. In endo-GHs, the shape of the active site needed is like a cleft or a tunnel (Figure 1.8, top). In exo-GHs, the shape of the active site needed is a pocket or crater in the protein surface (Figure I – H, bottom). In order to know the relative position of a sugar respect to the glycosidic bond broken, the region where is docked the sugar is labeled form –n to +m (where n and m are integers, Figure 1.8, center). The negative positions are the non-reducing sugars and the positive positions are the reducing ones (Davies, 1997). The sugar saccharide located in the -1 subsite is named as "-1 sugar molecule/ring".

4.1. GHs mechanisms

Polysaccharides are one of the most stable polymers in nature, which hydrolysis rate constant at room temperature is about $2 \cdot 10^{-15}$ s^{-1} (millions of years [Wolfenden, 1998]). However, GHs are capable of accelerating the spontaneous hydrolysis reaction of sugars in water by 17 orders of magnitude.

GHs catalyse the cleavage of the glycosidic bond with either inversion or retention of the configuration of the anomeric carbon, but in both cases, there are needed two catalytic residues: a general proton donor and a nucleophile or base agent.

- **Retaining mechanism** (Figure 1.9, up): implies two S_N2 (bimolecular nucleophilic substitution) displacement reactions and the presence of a covalently linked glycosyl-enzyme intermediate (GEI, product of the Glycosylation step). In the Glycosylation step, the acid/base residue transfers its proton to the glycosidic oxygen and the nucleophile oxygen attacks the anomeric carbon of the -1 sugar ring, forming the covalent intermediate specie. Normally, this is the rate limiting step of the hydrolysis reaction, which goes throughout an oxocarbenium ion transition state (TS). The second step is known as Deglycosylation step. The attack over the anomeric carbon is done by a catalytic water which is preactivated by the acid/base residue attracting one of the hydrogens of the water. The product is the monosaccharide with the same stereochemistry that had when was forming the oligo/poly-saccharide ($\alpha \rightarrow \alpha$).

- **Inverting mechanism** (Figure 1.9, down): the hydrolysis is performed in a single S_N2 displacement, in which the nucleophile agent is a water molecule, activated by a base agent (similar than the Deglycosylation step). The proton donor catalyses the departure of the leaving group by transferring its proton to the glycosidic bond oxygen. The base agent enhances the nucleophilicity of the attacking water by withdrawing one of its protons. The reaction goes throughout an oxocarbenium ion transition state (TS).

In both cases, the distance between the acid/base residue and the nucleophile agent is much conserved, being around 5-6 Å for retaining and 10-11 Å for inverting GHs.

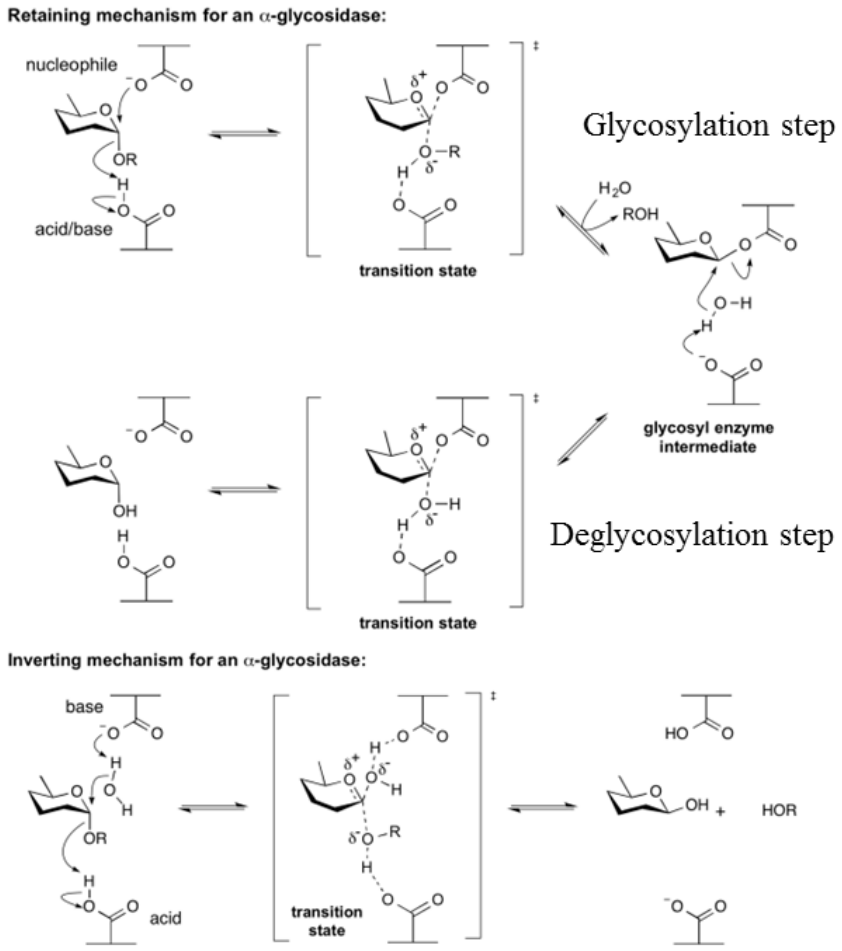


Figure 1.9. Reaction mechanism in retaining (top) and inverting (bottom) α-glycosidases. Figure adapted from the reference (www.cazy.org).

Finally, Kinetic Isotope Effects (KIE) studies reveal that both reaction mechanism of GHs involve a very similar sugar oxocarbenium ion TS (Zechel, 2000). This TS presents a highly dissociative character, with low bond orders between the anomeric carbon and the glycosidic oxygen as well as the nucleophile. Consequently, significant positive charge is developed at the anomeric carbon, which is partially compensated by an electron donation from the pyranic oxygen atom (O_p). This electron donation gives rise to a C_1-O_p partial double bond character (sp^2 character). This promote that the C_2-C_1-O_p-C_5 atoms adopt a planar geometry, This limitation affects to the possible conformations that can adopt the ring in the TS: two half-chair $^4H_3$, $^3H_4$, two boat $B_{2,5}$ and $^{2,5}B$ and four envelop E_3, ^3E, E_4 and ^4E conformations (see Chapter I – 2.1). Furthermore, conformational

studies of the most common sugar oxocarbenium ions were performed confirming this assumption (J. Iglesias-Fernández PhD Thesis, 2014).

4.2. 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 enzyme-substrate interactions (Blake, 1965). The results obtained were in agreement with the "induced fit theory" proposed in 1958 by Daniel Koshland (Koshland, 1958). 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.

In 1967, David Phillips observed the first evidence of substrate distortion in GHs during his studies on hen-egg white lysozyme (Phillips, 1967), proving that sugar conformations play an important role in GH catalysis. Furthermore, several studies (both experimental and computational) have shown that sugar molecules bind to GHs in a distorted shape, such as boat ($^{1,4}B$), skew-boat ($^{O}S_2$, $^{1}S_3$, $^{1}S_5$) and sometimes, the distortion is not needed and the sugar should be in the $^{4}C_1$ chair conformation (Davies, 2012). It is already stablished that substrate distortion has significant benefits on catalysis (Vocadlo, 2008).

4.2.1. Distortion preactivation: Steric and electronic effects

To quantify how much preactivated is a conformation of a sugar, there are several criteria to be analysed. This preactivation will be conditioned by the protean medium, the nature of the sugar and the interaction between them. One of the conditions will be the nucleophile does not found an impedance to attack the anomeric carbon (Steric hindrance). The second one is the effective charge of the sugar atoms that have an important catalytic role (Electronic similarities to the TS and Stereoelectronic theory). Of course, the stability of the conformation and other entropic effects must be taken into account.

- **Steric hindrance**: the nucleophilic attack at the anomeric carbon and the concomitant leaving group departure require an *in-line* approach of the nucleophile agent. Due to the typical distribution of this residue in α and β-glycosidases, the conformation that must adopt the sugar to be attacked have the

glycosidic bond in axial orientation. For example, in β-glucosidases, the β-glucose is in a $^{1}S_{3}$ conformation, where the glycosidic bond is directed to the axial position. Oppositely, in α-glucosidases, the α-glucose does not need to be distorted ($^{4}C_{1}$), and as in the other case, the glycosidic bond is in axial orientation (Figure 1.10).

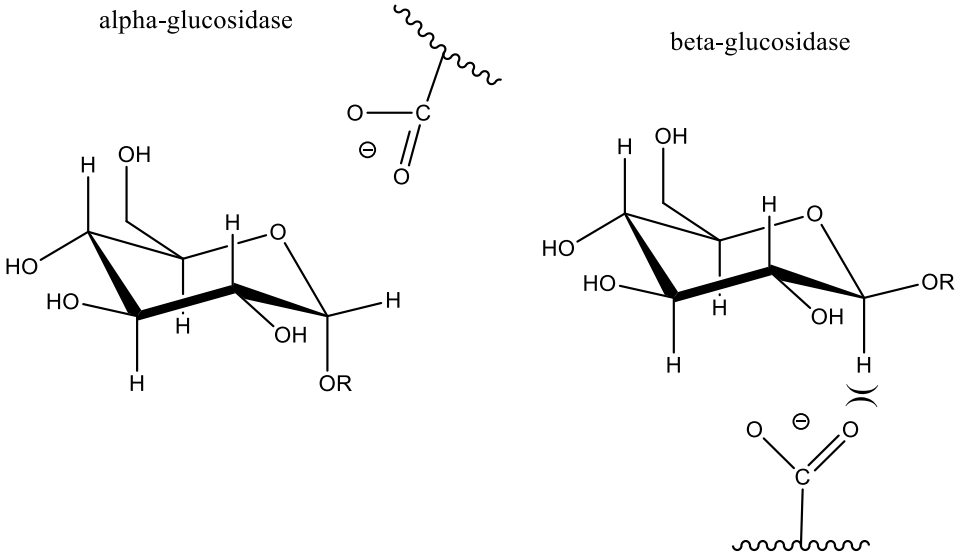


Figure 1.10. Steric effect of the chair conformation of the natural substrates in α and β-glucosidases.

- **Structural and electronic similarities to the TS**: As mentioned previously, the active site of enzymes tends to distort the substrate to a closer conformation of the TS of the enzymatic reaction. 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 C_1-O_1 distance (glycosidic bond) and a decrease of the C_1-O_p distance (partial double bond character). Suddenly, suggesting that substrate distortion resembles structurally and electronically the oxocarbenium TS of the reaction (Biarnés, 2006) (Biarnés, 2007).
- **Stereoelectronic theory**: As mentioned before, the TS of the reaction is characterized by a planar configuration of the -1 sugar molecule that increases electron delocalization between O_p and C_1. The key factor for the reactivity of orthoesters, acetal and other related compounds relapse in the ability of the pyranic oxygen to donate its nonbonding electrons (Kirby, 1984). The

stereoelectronic theory, derived from the ALPH (Antiperiplanar Lone-Pair Hypothesis) (Deslongchamps, 1975), proposes that such hydrolysis reactions are favoured by conformations in which the nonbonding (lone-pair) electrons of the intra-ring oxygen atom lie antiperiplanar to the leaving group bond. In this configuration, the overlap between the O_p 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 change hybridization into those of the oxocarbenium ion-like TS with minimal structural organization (Figure 1.11).

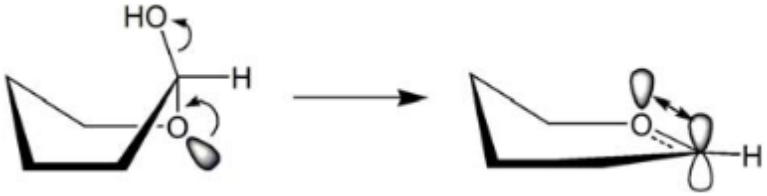


Figure 1.11. Schematic representation of a sugar hydrolysis according to the stereoelectronic theory. Figure taken from the reference (J. Iglesias Fernandez Thesis, 2014).

4.2.2. Conformational itineraries

As seen before, the substrate adopts a distorted conformation in the Michaelis complex (MC) of a big amount of GHs that approaches it to the TS of the reaction. Furthermore, reactants and products (inverting GHs) or covalent intermediates (retaining GHs) should adopt a conformation closer to one of the oxocarbenium-ion like TS shapes. Then, the conformational itineraries followed by the substrate of GHs can be classified according to the TS conformation adopted by the -1 sugar molecule (Davies, 2012).

The Cartesian or Stoddart diagram is used to systematically classify the existing information about the conformational itineraries of different GHs (Vocadlo, 2008) (Figure 1.12). As expected, GHs belonging to the same family tend to use the same conformational itinerary.

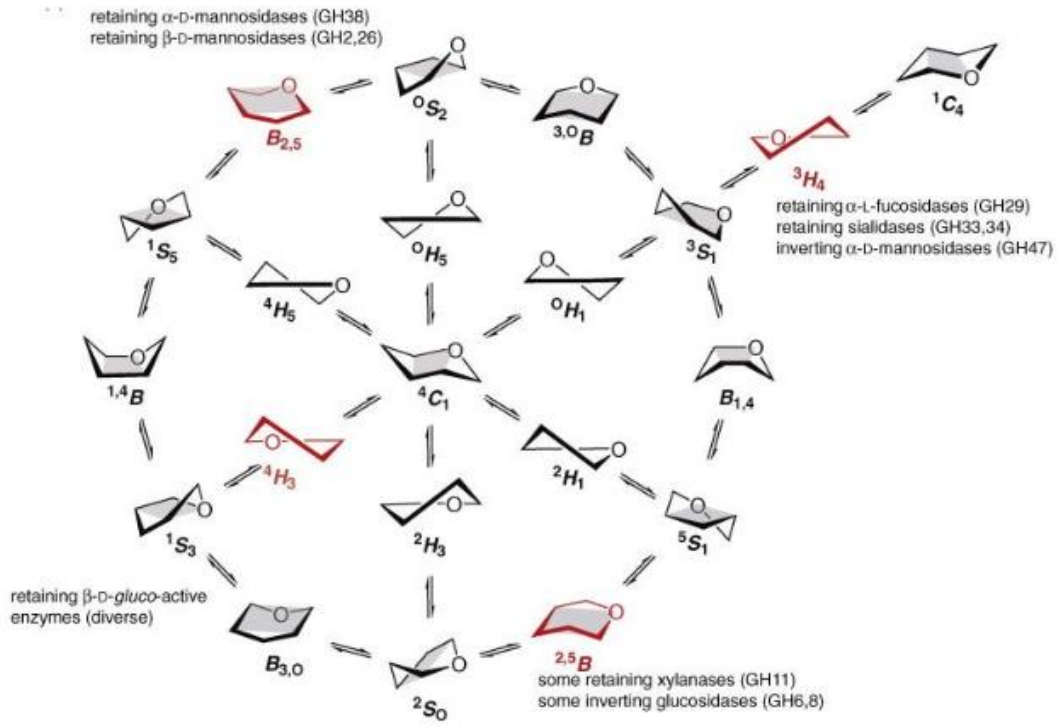


Figure 1.12. Cartesian or Stoddart diagram (Northern projection, including one itinerary that extends toward the Sourth Pole) of all possible conformations and their interconversions. In red are the TS conformations employed by different GH types and families. $^{4}H_{3}$ itinerary includes ^{4}E and E_{3} and $^{3}H_{4}$ includes ^{3}E and E_{4} TS conformations. Figure taken from reference (Vocadlo, 2008).

In this Thesis, members of two big families of enzymes have been studied extensively. On the one hand, GH13 and GH31 α-glucosidases, families of retaining GHs that hydrolyse α-glucose oligosaccharides. On the other hand, N-glycan processing α-mannosidases, where GH92 and GH125 conformational itineraries were not clear experimentally. Furthermore, both families of enzymes will be key to understand the interaction of septanosides with GHs.

4.3. α-glucosidases

The inhibition or activation of these kind of enzymes is a key factor in the control of the diabetes type II, where acarbose – α-glucosidase inhibitor – is one of the components of the medication for this disease. Concretely, the family 13 (GH13), which α-amylase is part from the family, is one of the most experimentally studied GH group. Furthermore,

this kind of enzymes have much industrial interest because, both GH13 and GH31, present some members that (some mutated and some naturally) get trans-glycosylation processes. The difference between GH13 and GH31 is in the acid/base residue: glutamic acid and aspartic acid, respectively. Both 3D structures are $(\beta/\alpha)_8$ barrels.

Trans-glycosylation is a process observed in retaining GHs occurred after the glycosylation step of the reaction. Instead of the deglycosylation process, where a water molecule hydrolysed the anomeric carbon, a new sugar molecule attacks the anomeric carbon, forming a new oligosaccharide. In the numerical nomenclature of enzymes, hydrolytic and trans-glycosylic GHs are different: 3.2.1.x and 2.4.1.x, respectively. For example, the well-known α-amylase code is 3.2.1.1 and for the amylosucrase studied in this Thesis the code is 2.4.1.4 (it is a GH that acts as a glycoside transferase (GT) at glucose high concentrations).

In the Table 1.1, a bibliographic research is presented to observe in perspective the most relevant crystallographic data available for α-glucosidases.

Table 1.1. Available substrate-enzyme complexes of retaining GH13 and GH31 α-glucosidases and the conformation adopted by the saccharide at -1 subsite.

Enzyme (Code)	PDB Code (Res. [Å])	Mutant	Subst. conform.	Type structure	Reference
α-amylase (3.2.1.1)	1BAG (2.50)	Glu208Gln (PD)	$^4C_1$	MC	(Fujimoto, 1998)
α-amylase (3.2.1.1)	1E3Z (1.93)		^2E / $^2H_3$	TS analogue	(Brzozowski, 2000)
α-amylase (3.2.1.1)	3BCD (2.20)		^2S_O	MC	(Tan, 2008)
α-amylase (3.2.1.1)	5TD4 (2.30)	Asp300Asn	^2E / $^2H_3$	MC	(Zhang, 2016)
Isomaltase (3.2.1.10)	3AXH (1.80)	Glu277Ala (PD)	$^4C_1$	MC	(Yamamoto, 2011)
α-glucosidase (3.2.1.20)	3WY4 (1.90)	Glu271Gln (PD)	$^4C_1$	MC	(Shen, 2015)

α-glucosidase (3.2.1.20)	3W38 (2.79)			WT	(Tagami, 2013)
maltase glucoamylase (3.2.1.4&20)	3TON (2.95)			WT	(Ren, 2011)
α-glucosidase II (3.2.1.20)	5DKX (1.40)			WT	(Satoh, 2016)
Dextran glucosidase (3.2.1.70)	2ZID (2.20)	Glu236Gln (PD)	$^4C_1$	MC	(Hondoh, 2008)
Pullulanase (3.2.1.41)	2FHF (1.65) 2FH6 (1.80)		$^4C_1$ $^4C_1$	MC Products	(Mikami, 2006)
Neopullulanase (3.2.1.135)	2D0H (2.10)	Glu396Gln (PD) Asp356Asn (Nuc)	$^4C_1$	MC	(Abe, 2006)
Malto-oligosyltrehalose trehalohydrolase (3.2.1.141)	3VGH (2.60)	Glu283Gln (PD)	$^2E/^4C_1$	MC	(Okazaki, 2011)
α-transglucosylase (2.4.1.161)	4BA0 (1.85)		$\mathbf{^1S_3}$	GEI	(Larsbrink, 2012)
Amylosucrase (2.4.1.4)	1JGI (2.00)	Glu328Gln (PD)	$\mathbf{^4C_1}$	MC	(Mirza, 2001)
Amylosucrase (2.4.1.4)	1S46 (2.20)	Glu328Gln (PD)	$\mathbf{^4C_1}$	GEI	(Jensen, 2004)
Sucrose Phosphorilase (2.4.1.7)	2GDU (2.1) 2GDV (2.0)	Glu232Gln (PD)	$^4C_1$ $\mathbf{^1S_3}$	MC GEI	(Mirza, 2006)

During the present Thesis, we have worked with 1JGI and 1S46 for amylosucrase studies and 3W38, 3TON and 3WY4 for septanoside docking in GHs.

4.4. N-glycan processing α-mannosidases

One of the most interesting biological tasks of sugars occurs in the Golgi's apparatus. The catalytic activity of an enzyme depends critically on its folding. To ensure the good folding of the enzymes, sugars are attached at the protein surface as a quality marker. Depending on the microorganism, the sugar chains are slightly different. In some cases, the sugars presents are α-mannose rings bonded between them in different hydroxyl position. The enzymes whose function is break the bond between mannoses are known as N-glycan processing α-mannosidases (Table 1.2). The lack of this kind of enzymes or the smooth loss of activity is an effect of a disease that shows an important role in the genetically acquired diabetes.

Table 1.2. Molecular properties of the N-glycan processing α-mannosidases and the conformational itinerary (both experimentally and computationally) followed by the saccharide in the -1 subsite.

Family	Mechanism	3D struc.	Metal active	Endo/ exo	Itinerary	Reference
GH38	Retaining	$(\beta/\alpha)_7$	Zn^{2+}	Exo	$^{O}S_2 \rightarrow B_{2,5} \rightarrow {}^{1}S_5$	(Petersen, 2010)
GH47	Inverting	$(\alpha/\alpha)_7$	Ca^{2+}	Exo	$^{3}S_1 \rightarrow {}^{3}H_4 \rightarrow {}^{1}C_4$	(Thompson, 2012a)
GH76	Retaining			Endo	$^{O}S_2 \rightarrow B_{2,5} \rightarrow {}^{1}S_5$	(Thompson, 2015)
GH92	Inverting	$(\alpha/\alpha)_6$	Ca^{2+}	Exo	Thio-derivative MC: **$^{4}C_1$** Mannoimidazole TS: **$^{1}S_5$ / $^{4}H_5$**	(Zhu, 2010)
GH99	Retaining	$(\beta/\alpha)_8$		Endo	Through epoxide TS	(Thompson, 2012b)
GH125	Inverting	$(\alpha/\alpha)_6$		Exo	Thio-derivative MC: **$^{4}C_1$**	(Gregg, 2011)

In the present work, the GH92 and GH125 inverting α-mannosidases have been studied computationally, due to the controversy between the results using thio-derivatives

and TS mimics and the expected conformational itineraries for this kind of enzymes. GH92 results are not presented due to the final conclusions are not clear yet.

Therefore, it is necessary to analyse which strategies are commonly used to freeze the enzyme-substrate complex in the reactants well.

4.5. Strategies to mimicking the natural substrate in GHs

One of the more crucial challenges in the research of catalytic pathways of GHs is the obtaining of a good structure of the Michaelis Complex (MC). The three common strategies to allocate the natural substrate or a mimic of it inside the enzyme are:

- **Mutation of one or both catalytic residues**: Returning to the point 4.1, we have seen that the reaction mechanism of GHs depends on a catalytic acid/base and a nucleophile agent. Mutating the carboxylic acid to a non-acid proton donor as an amide (glutamine or asparagine), the leaving group is not protonated and, then, the reaction is not activated, resulting the natural substrate inside the active site. This is the case of almost the whole cases of α-glucosidases.
- **Use a thio-derivative as mimic of the natural substrate**: Due to the acceptor should be a bad leaving group, it is necessary to be protonated for the catalysis. However, if the glycosidic oxygen is substituted by a less nucleophile atom, as sulphur, the acid/base residue does not transfer its proton to the leaving group and the reaction does not occur, resulting the thio-derivative inside the non-mutated active site. This is the case of GH92 and GH125 crystallised complexes.
- **Use a fluoro-derivative as mimic of the natural substrate**: The hydroxyl group in the position 2 (2-OH) is substituted by a Fluor atom. This atom increase the energy of the oxocarbenium ion due to the increase of negative charge near the anomeric carbon. The reaction barrier increase enough to stop the reaction.

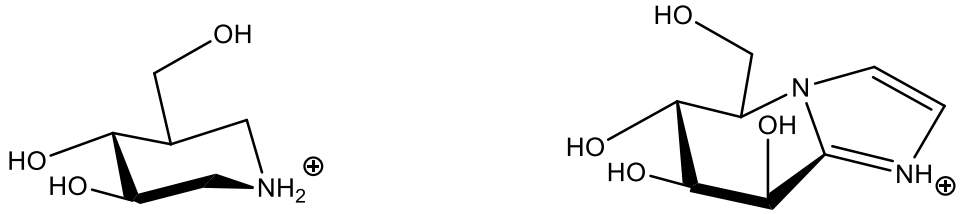
The point to be discussed is if the chemical changes in the substrates can affect to the conformational mimic potential.

4.6. Transition state mimics

As mentioned in Section 4.2.1., the active site of a GH is distributed to attach the sugar with a close conformation to the TS of the hydrolysis reaction. Then, we have

observed that the charge in the anomeric carbon in the TS will be the maximum positive value expected for the sugar ring. This two conditions are key to develop new inhibitors and transition state mimics for GHs. Inside this group, experimental and computational studies were done with isofagomine and mannoimidazole (Williams, 2014), to understand their inhibitor and TS mimic capacities.

On the one hand, known as isofagomine type inhibitors (Figure 1.13), which protonate inside the enzymatic catalytic site, resemble a glycosyl cation (e. g. charge is localized at C_1) but are poor mimics of the TS conformation. Most of the X-ray structures show isofagomine bound in a $^4C_1$ conformation, which is none of the eight possible TS conformations (^3E, ^4E, E_3, E_4, $^4H_3$, $^3H_4$, $B_{2,5}$ and $^{2,5}B$). On the other hand, known as mannoimidazole type inhibitors, they normally adopt one of these conformations, displaying a 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) (Thompson, 2012a). Furthermore, mannoimidazole inhibitors can mimic TS observed in the catalytic itineraries of α-mannosidases (both $B_{2,5}$ and $^3H_4$) (Tailford, 2008) (Thompson, 2012a). All these aspects were analysed in the reference (J. Iglesias Fernández Thesis).



Isofagomine

Mannoimidazole

Figure 1.13. Isofagomine and mannoimidazole TS mimic inhibitors.

Objectives

Objectives

The present Thesis is aimed at elucidating the molecular mechanisms of glycosidic bond hydrolysis by glycoside hydrolases (GHs) of families GH13 (amylosucrase, a retaining GH), GH125 (inverting α-mannosidases) and GH3 (retaining β-glucosidases), as well as assessing the potential of septanosides as novel GH substrates.

The following objectives have been pursued:

- To elucidate the catalytic mechanism and conformational itinerary of GH13 amylosucrase, a retainig GH, considering both hydrolysis and polymerization processes.

- To quantify the effect of substrate modifications (thio-derivatives) on the Michaelis complex substrate conformation of GH125 inverting α-mannosidases and predict their conformational catalytic itinerary.

- To quantify the effect of substrate modifications (thio-derivatives) on the Michaelis complex substrate conformations of GH3 retaining β-glucosidases and find out whether complexes with thio-derivative show different conformations than natural substrates, considering different types of glycosidic linkages.

- To obtain the conformational free energy landscapes of 7-membered ring sugars (septanosides) and asses their potential as substrates for glycoside hydrolases.

Chapter II – Methods

Methods

1. Philosophical introduction to theoretical chemistry

Theoretical chemistry is not just science. Sometimes, it is art. The theoretical chemist draws and gives life to an ideal of matter. The Laws of Physics are the ink, while, the artist places the nuclei and sculpts the electronic density, using his scientific intuition.

The principal reason to study chemical structures theoretically is to understand and to explain chemical events from the nanoscale point of view. However, our knowledge is conditioned by the complexity of the model, the mathematic expressions describing the model behaviour and the machine we are going to use to solve these expressions.

Then, the capability of ensure that the used model and methodology describe properly the experimental data becomes a necessary and sufficient condition to a theoretical chemist.

In the present chapter, several methodologies are going to be described. Adding to the complexity of the studied systems, the no-possibility of a human brain to calculate the mathematical equations of a chemical event during a human average life, it highlights the need to use faster and powerful supercomputers.

2. Quantum Mechanics (QM)

2.1. The molecular Schrödinger equation

At nanoscale, a molecule can be described as a combination of nuclei and electrons. The position of the nuclei (N_A), the electrons (n_i) and the energy (ε) are needed to define the state of the molecule. Applying quantum theory, the mathematical expression which create a relationship between the molecular wave-function ($\Psi(N_A, n_i)$) and the energy is the molecular Schrödinger equation (Szabo, 1982. Paniagua, 2000)

$$\widehat{H}\,\Psi(N_A, n_i) = \varepsilon\,\Psi(N_A, n_i) \qquad (\mathbf{Eq.\,II-1})$$

and the expression of the Hamiltonian operator (H) can be separated

$$\widehat{H} = \widehat{T_{N_A}} + \widehat{T_{n_i}} + V_{N_A} + V_{n_i} + V_{N_A - n_i} \qquad (\mathbf{Eq.\,II-2})$$

in kinetic (T) and potential (V) terms. This expression is a complex differential equation whose exact resolution is an utopian thinking as the size of the system to be studied increases.

Then, the objective will be simplify the problem, trying to separate the interactions in nuclei and electronic terms. However, the attractive term of the nuclei-electron interaction makes harder this separation.

2.2. Born – Oppenheimer Approximation

In 1927, Robert Oppenheimer and Max Born propose an approximation to solve the molecular Schrödinger equation. Initially, the position of the nuclei must be defined. Over this situation, the electronic Hamiltonian (H_{el}) is calculated using electronic wave-functions ($\psi_{el,A}$) which depends on the initial nuclear coordinates (A). The expression of the electronic Schrödinger equation in atomic units is

$$\left(\sum_{i=1}^{n} -\frac{1}{2}\nabla_i^{\,2} + \sum_{i=1}^{n-1}\sum_{j>i}^{n}\frac{1}{r_{i,j}} + \sum_{A=1}^{N}\sum_{i=1}^{n} -\frac{Z_A}{r_{i,A}}\right)\psi_{el,A}(i) = E_{el,A}\psi_{el,A}(i) \quad (\mathbf{Eq.\,II-3})$$

where ∇_i is the partial derivatives operator applied over the electronic coordinates of the electron *i*, the $r_{i,j}$ and $r_{i,A}$ are the distances between two different electrons *i* and *j*, and an electron *i* and the nucleus *A*, respectively, and the Z_A is the effective nuclear charge of the nucleus *A*.

Once the electronic energy ($E_{el,A}$) is calculated, it is used as component of the nuclear Hamiltonian (H_{nuc}) which is decomposed in the following expression.

$$\left(\sum_{A=1}^{N} -\frac{1}{2m_A}\nabla_A^{\,2} + \sum_{A=1}^{N-1}\sum_{B>A}^{N}\frac{Z_A Z_B}{R_{A,B}} + E_{el,A}\right)\psi_{nuc}(A) = E\psi_{nuc}(A) \quad (\mathbf{Eq.\,II-4})$$

The resulting eigenvalues (E) are also eigenvalues of the molecular Hamiltonian (Eq. II – 1).

In order to give a physical justification to this approximation, the unique term which depends on the nuclei mass is the nuclear kinetic energy. Knowing that the relation between the mass of the nuclei and electrons is around 1800, it is easy to see that the

electrons will move around nuclei as nuclei seems to be immobile. Furthermore, this slow motion of the nuclei makes them not feel the effect of the movement details of the electrons. Nuclei are affected by an electron cloud. This cloud acts counteracting the repulsion between nuclei in the stable states.

Now, the point is to choose a good strategy to define properly the wave-functions used to describe the electronic behaviour of the chemical systems. In this line, there is a group of methods which approximate the mathematical expression of these functions. In the present thesis, the methods based in Density Functional Density (DFT) were applied to develop the QM models.

2.3. Density Functional Theory (DFT)

In 1964, Pierre Hohenberg and Walter Kohn reported two theorems (Hohenberg, 1964) where the Hamiltonian of a polielectronic system is determined by the electronic density of its ground state and, the energy of this state (E_0) is a lower asymptotic value for the energy of the system from a testing density functional ($E[\rho]$).

In 1965, Kohn and Lu Jeu Sham developed a method to approximate the expression of the $E[\rho]$ functional, known as Kohn-Sham Method (Kohn, 1965). The central idea is to match a Hamiltonian from a hypothetical n-electronic system whose electrons do not interact between them with the same monoelectronic density that the ground state of the real system. This Hamiltonian (H^{KS}) has the following expression:

$$\widehat{H}^{KS} = \widehat{T_{el}} + V^{KS} = \sum_{i=1}^{n} \widehat{h^{KS}}(i) \qquad (\boldsymbol{Eq.II-5})$$

Every monoelectronic Hamiltonian (h^{KS}) applied to every spin-orbital (ψ_i^{KS}) forming the ground state wave-function (Ψ_0) result the following expression known as Kohn-Sham equation.

$$\widehat{h^{KS}}\psi_i^{KS} = \varepsilon_i^{KS}\psi_i^{KS} \qquad (\boldsymbol{Eq.II-6})$$

Then, the density function is defined as a combination of the *n* single-electron orbitals with the lowest energies:

$$\rho(r) = \sum_{i=1}^{occ} n_i \, |\psi_i(r)|^2 \qquad (\boldsymbol{Eq.II-7})$$

In conclusion, being the main objective of the DFT method to find the better energy that makes minimal the expression of the energy, the energy functional of our system can be expressed in terms of the density of the single-electron orbitals (ψ_i) for a given nuclear coordinates (N_A).

$$E^{DFT}[\rho] = \min_{\{\psi_i\}}(E^{KS}[\{\psi_i\},\{N_A\}]) \qquad (\boldsymbol{Eq.II-8})$$

Finally, developing each interaction term of the Kohn-Sham Hamiltonian, the E^{KS} expression results in a combination of the electronic kinetic energy ($T^{KS}[\rho]$), the Coulomb classical repulsion ($J[\rho]$), the exchange-correlation ($E_{xc}[\rho]$) and the electron-nucleus interaction ($V[\rho]$ density functionals. The density-dependent expression of the Kohn-Sham energy is shown in the Eq. II – 9.

$$E^{KS} = \sum_{i=1}^{oc} n_i \left\langle \psi_i \middle| -\frac{\nabla^2}{2}\psi_i \right\rangle + \frac{1}{2}\iint \frac{\rho(r_1)\rho(r_2)}{r_{1,2}} dr_1 dr_2 + E_{xc}[\rho] + \int V\rho dr \; (\boldsymbol{Eq.II-9})$$

Once seen the term decomposition of the Kohn-Sham energy, the kinetic energy, electron-nucleus interaction and Coulomb repulsion are exact and calculable according to the DFT. However, the exchange-correlation term is not provided by the theory, due to it depends on the kinetic energy and the potential repulsion of the real system.

$$E_{xc}[\rho] = (T[\rho] - T^{KS}[\rho]) + (V_{el}[\rho] - J[\rho]) \qquad (\boldsymbol{Eq.II-10})$$

The difference between two DFT functionals is in the way to approximate this exchange-correlation functional. The most common approximations used are:

- *Local Density Approximation (LDA).*

This approximation is based on the fact that: given a neutral charge and uniform density (as for nuclei or electrons), the exchange-correlation energy can be described as an integral of the monoelectronic density multiplied by an expression of the single-electron exchange-correlation, depending on the density ρ (Dirac, 1930. Vosko, 1980).

This expression gives a good approximation to the $E_{xc}[\rho]$ functional for systems with smoothly changing charge densities. However, one of the principal disadvantages of the

LDA is in the van der Waals interactions. This problems are insuperable in cases where the π-stacking has an important role.

- *Generalized Gradient Approximation (GGA).*

In the present thesis, the GGA functionals have been used extensively. The principal reason is in the equilibrium of precision and computational cost. The LDA is not good enough for biological systems, where the enzyme-ligand interactions must be well defined. Then, the application of GGA functionals presents an improvement in the possibility to study biochemical systems with an irrelevant increase in the computational cost.

This approach is based on the use of an exchange-correlation functional which depends on the density and the gradient of this density $\nabla\rho(r)$. This change allows to taking into account the non-homogeneity of the real electron density. The functional is separated in a term describing the exchange and another for the correlation. The most popular GGA functionals used in biochemical environments are the BP86 (exchange part by Becke [Becke, 1984] and correlation by Perdew [Perdew, 1986]), BLYP (exchange part by Becke and correlation part by Lee, Yang and Parr (Lee, 1988) and PBE of Perdew, Burke and Ernzerhoff (Perdew, 1996). Furthermore, hybrid functionals are extensively used. These functionals contain a variable contribution of the exact correlation energy. In that group, the most popular one is the B3LYP functional (formed by three parameters: Slater exchange, Hartree-Fock exchange and LYP correlation contributions), but the computational cost increase for the kind of simulations we have carried.

In this work, the PBE functional was chosen to describe the studied systems. The improvement of the bonding interactions and the good description of the hydrogen bond interactions (Ireta, 2004), compensate the bad description of the van der Waals contribution. This good relationship with biochemical system was proved previously in simulations of reactions inside glycoside hydrolases and transferases (Biarnés, 2006. Biarnés, 2011. Ardèvol, 2010. Rojas-Cervellera, 2013) and comparing the description of the conformational behaviour of sugar with hybrid functionals (Biarnés, 2007).

2.4. Basis functions

Returning to the Eq. II – 6, it is necessary a way to expand the KS single-electron orbitals to solve the KS equations. The two extensively used expressions are based in Gaussian functions (Taketa, 1966) or in Plane Waves (Ihm, 1979).

The Gaussian-Type Orbitals (GTO) are commonly used in the chemistry community, due to they are according to the chemical insight. This kind of expansion of the orbitals could be extended to infinity, but normally, it is truncated by a limited set of basis functions. The mathematical expression is shown in the Eq. II – 11.

$$\psi_i(r) = \sum_j c_j e^{-ajr^2} \qquad (\boldsymbol{Eq.II-11})$$

Inside the GTO basis set, the most common developed functions are the Pople basis set (Ditchfield, 1971) and the Correlation-Consistent basis sets (Dunning, 1989). Examples of both, respectively, are the 6-31G, 6-311G* (polarized), 6-311+G (diffuse functions), 6-311++G* (polarized and diffuse functions) and, cc-pVDZ (double-zeta), cc-pVTZ (triple-zeta). The notation of the Pople basis set depends on the number of functions representing every atomic angular momentum and the spread of the Gaussian function.

Otherwise, in the physics community, the most common used basis set is based in Plane Waves (Eq. II – 12).

$$\psi_i(r) = \frac{1}{\Omega^{1/2}} \sum_G^{Gmax} C_G e^{iGr} \qquad (\boldsymbol{Eq.II-12})$$

In this expression, the Ω is the volume of the cell and G is the plane wave momentum. Plane-Waves (PW) are denoted by a cut-off 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

$$N_{PW} \approx \frac{\Omega}{6\pi^2} E_{cut}^{\frac{3}{2}} \qquad (\boldsymbol{Eq.II-13})$$

Then, in which situations we are going to use one or the other. On the one hand, GTOs have to good points: they show results according to chemical insight and using small basis sets give already good results. However, they are not orthogonal, depend on atomic positions and present basis set superposition errors (BSSE). They are good for

quick optimization, vibrational calculation and molecular orbital construction of small molecules (<60 atoms, <16 CPUs). On the other hand, PWs present more good points. They are orthogonal, independent of atomic positions, no presence of BSSE, Fast-Fourier Transform can be used efficiently and are naturally periodic. However, there are needed many functions to implement a good calculation. In conclusion, PWs are the most efficient function to carry dynamic simulations of slightly higher molecules (60-100 atoms), but with a slightly higher computational resource (> 32 CPUs).

In this last case, normally, the core electrons are frozen into the KS equations, in order to decrease the computational cost (core levels are separated energetically from the valence electrons). The interactions of the nuclei and valence electrons are described by using pseudopotentials, which are derived from all electron atomic calculations. In this work, we have used norm-conserving pseudopotentials derived with the Martins-Troullier method (Troullier, 1991).

3. Introduction to Molecular Dynamics (MD)

Once introduced the methodology to construct and study mechanically a chemical system, it is time to go farther. Molecules are in movement. Translations, rotations, vibrations. All this events are important to understand conformational changes of sugars (little molecules) and enzymes (high molecules), during the substrate-enzyme binding, catalysis and the leaving of the substrate from the active site.

Then, some molecular dynamics (MD) techniques were developed to simulate the dynamic effects of the temperature (entropy) in all kind of structures. These techniques are based on the integration of Newton's laws of motion. Given the atomic positions, their velocities and specifying the time per integrated step, the system evolves showing a trajectory of the atoms, which is giving a very important amount of information concerning structural and energetic properties of the molecule.

In this thesis, three techniques were used to explore the free energy surface of the studied systems. Classical or Molecular Mechanics (MM) MD allow us to stabilize enzyme structure at physiological temperature with a low computational cost. The system is constructed as a charge-spring model where the electrons are not considered explicitly. Furthermore, MM MD simulations show the dynamic behaviour of large systems at a

long scale point of view. However, it is not possible to simulate chemical reactions and, of course, changes in the electronic configuration of the system. In that case, Carr-Parrinello Molecular Dynamics (CPMD) allow us to carry shorter simulations with a better description of the electronic degrees of freedom. This technique is divided in two methodologies depending on the way to describe the whole system. On the one hand, Quantum Mechanism MD combine MD and DFT methods to solving the Schödinger equation of the system meanwhile the system moves. On the other hand, in the case we wanted to study reactions and excitations inside enzymes, the hybrid of Quantum Mechanics and Molecular Mechanics (QM/MM) MD allow us to describe quantum mechanically the region where the reaction happens and classically (cheap computationally) the rest of the model (enzyme residues, solvent, counter-ions).

There are three ingredients which are necessary to implement a molecular dynamics: initial positions, initial velocities and an algorithm to make evolve the system. Initial position can be obtained from a crystallographic structure. Initial velocities can be obtained randomly from a Maxwell-Boltzmann distribution at a fixed temperature (Eq. II – 14). This 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.

$$p(v_{ix}) = \left(\frac{m_i}{2\pi kT}\right)^{1/2} e^{\left(-\frac{m_i v_{ix}^2}{2kT}\right)} \qquad (\boldsymbol{Eq.II-14})$$

The most used algorithm based on Newton's laws of motion is the Verlet algorithm (Verlet, 1967), which integrates the equations of motion numerically. Some variations applied to the algorithm were developed, deriving to the simplest one, called *leap-frog* algorithm (Hockney, 1970), whose expressions of the evolution of velocities (Eq. II – 15) and positions (Eq. II – 16) are implemented in the AMBER software used in our reported works.

$$v_i\left(t+\frac{1}{2}dt\right) = v_i\left(t-\frac{1}{2}dt\right) + a_i(t)dt^2 \qquad (\boldsymbol{Eq.II-15})$$

$$r_i(t+dt) = r_i(t) + v_i\left(t+\frac{1}{2}dt\right)dt \qquad (\boldsymbol{Eq.II-16})$$

3.1. Classical or Molecular Mechanics (MM) MD

The principal feature of the MM MD simulations is the simplicity of the molecular model. The electrons are not explicitly defined and the nuclei are dots with a charge and a mass previously parametrized. The bonds between atoms are defined as springs following the Hooke's law and all the interactions between bonded atoms are parametrized using calculated all-atom quantum simulations distances, angles, dihedrals, impropers and charges. Furthermore, the non-bond interactions are calculated taking into account expressions for the van der Walls and electrostatics terms.

All this information is needed to calculate the energy of the system, step by step. In MM MD, the defined function to calculate this energy is known as Force Field.

3.1.1. Force Fields

A force field is defined as an equation formed by a set of functions of the degrees of freedom of the system employed to calculate the different contributions to the energy, using previously calculated parameters. The transferability of this parameters from little systems to higher ones (e.g. parameters of a DFT calculation of a histidine in gas-phase, to define the behaviour of the histidines in the backbone of an enzyme) is key for a good force field.

The force field used in this Thesis is the ff99 (Cornell, 1995) implemented in AMBER software and its expression (E_{ff99}) is shown in the Eq. II – 17 as a function of the bond, angles, dihedrals, van der Walls and electrostatic contributions.

$$E_{ff99} = \sum_{bonds} k_b(r - r_{eq})^2 + \sum_{angles} k_\theta(\theta - \theta_{eq})^2 + \sum_{dihedral} \frac{V_n}{2}[1 + \cos(n\varphi - \alpha)] + \sum_{i<j} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \sum_{i<j} \frac{q_i q_j}{\varepsilon R_{ij}} \qquad (\boldsymbol{Eq.II-17})$$

Our force field defines the bond and angle interactions as harmonic potentials needing as parameters the equilibrium values of the bond radius and three consecutively bonded atoms angle (r_{eq} and θ_{eq}) and its vibrational constants (k_b and k_θ). 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).

The non-bonded interactions are present in the fourth and fifth terms, where the van der Walls interactions are described by a Lennard-Jones potential containing an attractive function for the dispersion ($1/R^6$) and a repulsive function for the exchange repulsion ($1/R^{12}$). The electrostatic term is defined by the Coulomb law applied to every two atoms-charges interaction.

3.1.2. Canonical ensemble

Once the system is parametrized (enzyme, substrate, counter-ions and solvent), the model is ready to being minimized, heated at the physiological temperature and stabilized. The definition of our simulation does not allow the matter exchange (number of atoms N is constant) and the temperature T is controlled using a thermostat (the energy of the system fluctuates). Then, in terms of thermodynamic statistics, we are working in the Canonical ensemble. Initially, during the heating of the system, the volume is fixed (NVT ensemble). Once the system is stabilized at the physiological temperature, the pressure is fixed at 1 atm (NPT ensemble) to allow the expansion of the system since the obtaining of a real density (~1.0 g/cm^3). Once the density is stabilized, the system is ready for the productive molecular dynamics, fixing the volume, returning to the NVT ensemble.

3.2. Car-Parrinello Molecular Dynamics (CPMD)

In 1985, Car and Parrinello (Car, 1985) developed a method to simulate the dynamical behaviour of a quantum mechanically described system where the nuclear and electronic degrees of freedom evolve simultaneously. This methodology that combines DFT and MD, usually employs PWs basis set and pseudopotentials to describe the electronic density.

The strategy of the CPMD method is to assign a fictitious electronic mass (μ) to the electronic Kohn-Sham orbitals $\{\psi_i(r)\}$. The electronic and nuclear degrees of freedom evolves according to a modified set of classical equations of motion (Eq. II -18).

$$\mu\ddot{\Psi}_i = \frac{\delta E_{KS}}{\delta \Psi_i^*} + \sum_j \Lambda_{ij}\Psi_i(r) \;\; (electrons) \qquad M_N\ddot{R}_N = \frac{\delta E_{KS}}{\delta R_N} \;\; (nuclei) \;\; (\boldsymbol{Eq.II-18})$$

The interaction description of a system treated with the CP method is described by a Lagrangian function (Eq. II – 19).

$$L = \frac{1}{2}\left(\sum_{N}^{Nuc} M_N \dot{R}_N^2 + \mu \sum_{i}^{el} \int dr\, |\psi_i(r,t)|^2\right) - E^{KS}$$
$$+ \sum_{ij} \Lambda_{ij}\left(\int dr \psi_i^* \psi_j^* - \delta_{ij}\right) \qquad (\boldsymbol{Eq.II-19})$$

where the first term is the nuclear and electronic kinetic energies and the last term is a orthogonality constraint.

CPMD simulations provide the evolution of the nuclear trajectories and, the difference with the Born-Oppenheimer approximation, the evolution of the KS orbitals. The electrons evolve around the Born-Oppenheimer surface. In Figure 2.1, the procedure for a CPMD methodology is shown.

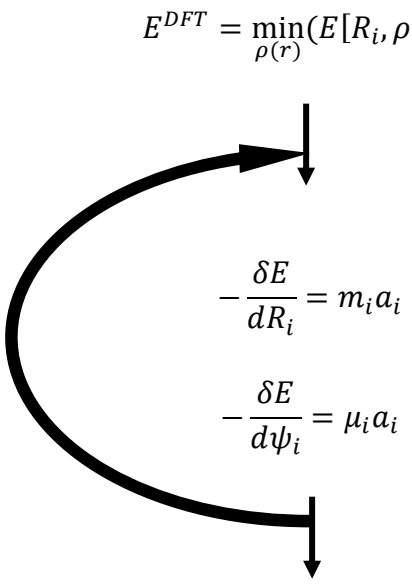


Figure 2.1. Schematic diagram of a QM CP molecular dynamics. Adapted from (Rovira, 2005).

To perform a good CPMD simulation, the adiabaticity between nuclei and electrons is the most important condition, due to make evolve both simultaneously is not free. The objective is to separate electronic and nuclear vibrational frequencies, normally around 10000 cm^{-1} and 3500 cm^{-1} (stretching C-H). The value of the chosen fictitious electronic

mass (E_{mass}) and the time per step (τ_{step}) is crucial in that way. By contrast with BOMD (~1 fs), the time step in CPMD will be shorter (<0.2 fs), in order to describe properly the faster movement of the electrons. The minimum electronic frequency depends on the electronic mass, where $\nu_{el}^{min} = ((HOMO\text{-}LUMO)_{gap}/\mu)^{1/2}$, then, small values of E_{mass} give a better adiabaticity of the system. If this energy exchange is small, the trajectory of a CPMD and a BOMD simulations are identical (Marx, 2000).

Inside the CPMD methodology, full QM or hybrid QM/MM methods can be applied, depending on the size of the chemical system.

3.2.1. Quantum Mechanics (QM) MD

Molecular models can be extremely complex (membrane protein, membrane phospholipids, counter-ions, substrate, solvent) or can be extremely simplistic (sugar molecule in gas phase). Sometimes, as a first step to study the mechanistic behaviour of an enzyme over a little substrate, the conformational and structural information given by a molecular dynamics of the molecule in gas phase is crucial to understand the following results obtained with complex models. In that case, where a single sugar ring is the object of study, being molecules with 20-40 atoms, the CPMD method can be applied over a full QM description of the atoms.

The parameters that have to be taken into account are the size of the box where PWs are expanded, the E_{mass}, the time step and the electronic kinetic energy expected value. The size of the box is the maximum size occupied by the atoms in the Cartesian axis plus 3.5 Å per side, so 7 Å at x, y and z box components. The E_{mass} and time step will be tested running MD at the work temperature and chosen the higher E_{mass} and time-step where the system is still adiabatic (~<0.5 kcal/mol/ps·atom). Once chosen the pair E_{mass}- τ_{step}, the mean value of the electronic kinetic energy obtained in that simulation is chosen as a restraint value for a Nosé-Hoover thermostat (Nosé, 1984).

3.2.2. Hybrid QM/MM MD

In the MM section, we have exposed the impossibility to study reactions, excitations or other abrupt changes in the electronic density using force field methods, when our system is large and complex. Furthermore, full QM studies are expensive and, nowadays, impossible to be applied on system with more than 200 atoms.

For these reasons, in 1976, M. Karplus, M. Levitt and A. Warshel developed a hybrid method that mix the QM good electronic description of the region where the reactions happen and the MM cheaper and good van der Walls description of the structural region that maintain in a catalytic conformation the active region (Warshell, 1976). This hybrid method is known as Quantum Mechanics/ Molecular Mechanics (QM/MM) MD.

In summary, the model is separated in two differentiable regions: normally, the substrate of the protein and the side chains of the catalytic residues are introduced in the QM region and the rest of the model (protein, counter-ions, solvent) is described by a force field in the MM region.

The Hamiltonian for a QM/MM system is shown in the Eq. II – 20, where is separated in a QM, a MM and a QM/MM components.

$$H_{hybrid} = H_{QM} + H_{MM} + H_{QM/MM} \qquad (\boldsymbol{Eq.II-20})$$

The higher challenge of the QM/MM approach is in the definition of the $H_{QM/MM}$, so that, the way to define the interactions involving both the classical and the quantum parts of the system.

Using the CPMD software, the QM/MM models of this Thesis were parametrized following the methodology developed by Laio, Vandevondele and Röthlisberger (Laio, 2002b) which combines classical GROMOS force fields (with AMBER parameters) and Car-Parrinello MD. The degrees of freedom of the QM/MM interactions can be grouped in bonded and non-bonded (Laio, 2002b).

- *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 were developed to address this problem, such as the link-atom approach or the hydrogen capping methodology.

In this Thesis, normally, describing a C_α - C_β bond of a aspartate or glutamate amino acid, we have used the link-atom approximation through a monovalent pseudopotential

that saturate the QM region. The increase of computational cost due to the increase of QM atoms compensate the fact to add hydrogen capping unphysical atoms.

- *Non-bonded interactions:*

As in the AMBER force field, the non-bonded interactions are divided into electrostatics and van der Walls (Eq. II – 21).

$$H_{QM/MM}^{non-bonded} = \sum_{i\in MM} q_i \int \frac{\rho(r)}{|r-r_i|} dr + \sum_{i\in MM, j\epsilon QM} V_{vdW}(r_{ij}) \qquad (\boldsymbol{Eq.\, II-21})$$

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 a Lennard-Jones potential. The electrostatic interactions between the QM electronic density and the MM point charges is calculated using the Coulomb's law, where one of the charges is described by the probability of the electron be at a distance of the MM atom.

One of the problems of the hybrid approximation is the presence of positive MM charges in the QM box. This effect is called the *electron spill-out phenomenon*. This effect consist in the anomalous rearrangement of the QM electronic density localized near positive MM charges. Due to the free delocalization of the density when PWs are used, this effect is very important in CPMD. Then, the solution is in a modification of the Coulomb potential in the core region. The electrostatic potential is multiplied by a factor $v_i(r) = (r_{ci}^4 - r^4)/(r_{ci}^5 - r^5)$, that for values of r shorter than the covalent radius of atoms, the Coulomb potential goes to a finite value.

The second problem is the computational cost related to the huge number of operations that would be required for the direct evaluation of the electrostatic energy in a PW scheme. In the used QM/MM approach, the electrostatic interactions are taken into account within a multilayer approach, in which MM layer is treated in different degrees of simplification (Laio, 2002c).

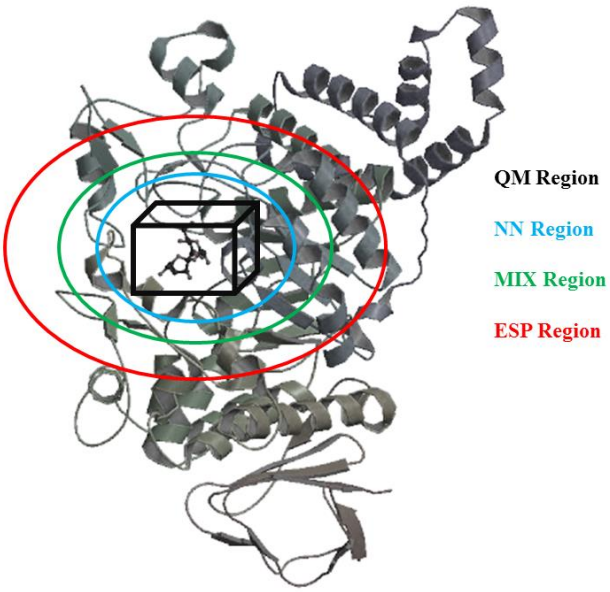


Figure 2.2. Schematic representation of the NN, MIX and ESP regions of electrostatic interactions used in the multilayer approach.

Observing the Figure 2.2, there are three layers around the QM box, named NN, MIX and ESP regions. For the classical atoms present in the first internal layer (NN region), the electrostatic interactions between each classical point charge and the electron density are explicitly calculated. In the second region (MIX), this interaction is only calculated for atoms with point charges higher than 0.1 e^- (in absolute value). In the external layer (ESP), atoms present there and MIX atoms with charge < 0.1 e^- are equally treated with respect of the electron density. These interactions are calculated between the MM partial charges, obtained for the force field, and the dynamically generated RESP charges (D-RESP. Laio, 2002c) of the QM system. Outside these last region, the interactions are simply described through multipole expansions in order to reduce the computational cost.

Then, QM/MM MD simulations need a test for the parameters named in the QM MD description (3.2.1) and the NN, MIX and ESP radii.

4. Free Energy Landscape (FEL) Sampling Methods

In the previous MD part, we have introduced the interest of the computational chemistry in the simulation of reaction mechanisms, conformational changes, and electronic excitations. However, we have shown the impossibility of simulate this events

using classical MD. In the case of *ab initio* QM MD, the hardness comes from a statistical reason.

Imagine a reaction mechanism which pathway goes through a transition state 10 kcal/mol higher in energy. Knowing that the relationship of the population of two states is proportional to the difference in energy of both states, through the Boltzmann equation (Eq. II – 22), this relationship for our hypothetic case at 300 K is $5.18 \cdot 10^{-8}$. So that, the system will explore one time the transition state every 20 million steps of simulation. For a CPMD simulation, choosing a time-step of 0.12 fs, it is equal to a 2.4 ns simulation (when, nowadays, we are able to do hundreds of picoseconds).

$$\frac{N_i}{N_j} = e^{(E_j - E_i)/k_b T} \qquad (\boldsymbol{Eq. II - 22})$$

Then, meanwhile the computer power went improving, in the last years several methodologies were developed to accelerate rare events. The final objective of this methods is the reconstruction of the energy pathway of this events. This pathway is related with a set of internal coordinates known as Internal Reaction Coordinate (IRC) which is the combination of the minimum set of degrees of freedom of the molecular system that evolve quantitatively along the reaction process. So that, a reaction is a process which can be described knowing the evolution of the IRC set and the energy of the system. The combination of this information is known as energy landscape. When the reaction is studied through minimized structures along the reaction pathway, so the velocities and temperature tend to zero, the resulting energy landscape is known as Potential Energy Landscape (PEL) and it is important to know if a rare event is kinetic and thermodynamically possible. However, there are conformational and reaction events that are possible thanks to the temperature. So, the effect of the entropy is crucial to understand the processes. In those cases, adding the dynamic effects in a simulation, the resulting landscape is known as Mean Force Potential or Free Energy Landscape (FEL).

Once introduced the concepts, the mostly used methods are Thermodynamic Integration, Replica Exchange, Umbrella Sampling, Nudged Elastic Band (NEB), Steered MD, String Method, Metadynamics (MTD), Well-Tempered Metadynamics (Well-T MTD), etc.

In this Thesis, the FEL reconstructions were done using the MTD and Well-T MTD methods, being methods that allow the system to explore its FEL activating collective variables (CV) throughout repulsive Gaussian potentials.

4.1. Metadynamics (MTD)

In 2002, Alessandro Laio and Michelle Parrinello reported a methodology to help chemical system escape from free energy minimas (Laio, 2002a). Metadynamics (MTD) became a technique aimed at enhancing the conformational sampling of the phase space and the estimation of the FEL (Laio, 2008). The challenge of the algorithm consists in a dimensional reduction of the FEL throughout a well selection of a set of CVs that allow the system to explore non-equilibrium regions.

During the last decade, several approximations were applied to improve the efficiency of the method. Inside this approaches, three methods were used extensively: Lagrangian MTD, Direct MTD and Well-T MTD. In this Thesis, we have applied the last two methods.

Considering the definition of a set of collective variable *s* which is an explicit function of the coordinates *x* and allowing a good description of the chemical changes we want to study (e.g. distance describing the cleavage/formation of a bond). The equilibrium behaviour of these variables is defined by a probability distribution (Eq. II – 23).

$$P(s) = \frac{e^{-\left(\frac{1}{T}\right)F(s)}}{\int e^{-\left(\frac{1}{T}\right)F(s)}\, ds} \qquad (\boldsymbol{Eq.\, II - 23})$$

In a continuous CV space of a dynamic system, the probability to find the system in a given point tends to zero, however, knowing that the thermal fluctuations ($1/2 \cdot k_b \cdot T$) allow the system evolve around a minima position, the probability to find the system in this region is higher.

The s-dependent free energy F(s) expression (Eq. II – 24) is given by a function of *s* (value of the CV), S(x) (the function defining the CV respect the system coordinates) and V(x) (external potential).

$$F(s) = -Tln\left[\int e^{-\frac{1}{T}V(x)}\delta\big(s - S(x)\big)dx\right] \qquad (\boldsymbol{Eq.\, II - 24})$$

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

$$f_i^V = -\frac{\delta V}{\delta r_i} \qquad (\boldsymbol{Eq.II-25})$$

This methodology consists in the adding of repulsive Gaussian potentials every τ_d (deposition time) MD steps, whose shape depends on two more parameters: the Gaussian height *w* and the Gaussian width δs to construct the V_G potential (Eq. II – 26) biased by a history dependent potential as $V_T = V + V_G$.

$$V_G(S(x),t) = \sum_{t'=n\tau_d} we^{\left[-\frac{(S(x)-s(t'))^2}{2(\delta s)^2}\right]} \; for\; n = 1, \dots, \#\, Gaussian \; (\boldsymbol{Eq.II-26})$$

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

$$f_i^{V_T} = f_i^V + f_i^{V_G} = -\frac{\delta V}{\delta r_i} - \frac{\delta V_G}{\delta r_i} \qquad (\boldsymbol{Eq.II-27})$$

Then, the repulsive additive V_G and the relaxation of the system during the steps between Gaussian addition promotes the system evolve to non-equilibrium region and transition states between two minima. Once the system have enough energy to leave the reactants region, it evolves to the products well. This process have to occur equally to go from products to reactants. At infinite time (Eq. II – 28), it was demonstrated that the bias potential will converge to the negative of the free energy and then, since $V_G(s,t) = -F(s,t)$, it will oscillate around the space of reaction interest. At this point, the motion of the CVs becomes diffusive and the simulation must be stopped (the over excitation of the system can promote secondary reactions or the system collapse).

$$\lim_{t\to\infty} V_G(s,t) = \sum we^{\left[-\frac{(s(t)-s(t'))^2}{2(\delta s)^2}\right]} \approx F(s) \qquad (\boldsymbol{Eq.II-28})$$

The problem of this method is in the error due to the height of the Gaussian (*w*) and the deposition time. Clearly, in an extreme case, we can add a Gaussian per step (deposition time = 1 MD step), but the system have not enough time to sample neighbour regions. Then, which is the good equilibrium between computational cost and deposition

time? What about a *w* parameter dependent of the energy added in each point of the CV space?

The first question depends on the dynamic behaviour of the system and has to be tested for everyone. The second question is answered with the Well-tempered approach.

4.2. Well-tempered Metadynamics (Well-T MTD)

One of the more important decisions to take during a MTD simulation is when the FEL is converged, and then, the simulation can be stopped. In fact, a standard MTD result not converge in a value but oscillates around the correct result. This leads to an average error that is proportional to the square root of the deposition time τ_d (Laio, 2005). Added to the computational cost reason, the increase of the deposition time have the risk of the promotion of the system to high energy regions of the FEL with no physical relevance.

Well-T MTD (Barducci, 2008) provides a framework to compute free energies that are converged to the exact result at long times and control the regions of the FEL that are meaningful to explore. As it was mentioned before, the height of the Gaussian potentials added during the simulation is modified according to the expression:

$$w^{Well-T} = we^{-[V(s,t)/\Delta T]}t' \qquad (\boldsymbol{Eq.II-29})$$

Where V(s,t) is an estimation of the free energy value at the current CV position 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 + ΔT. In this method, the bias potential V_G does not fully compensate for the underlying free energy landscape; rather, we have:

$$G(s) + V_G(s,t) = G(s) - \frac{\Delta T}{T+\Delta T}G(s) = \frac{T}{T+\Delta T}G(s) \quad (\boldsymbol{Eq.II-30})$$

Following, the convergence of Direct MTD and Well-T MTD methods are going to be analysed from a mathematical point of view.

4.3. FEL convergence

In 2015, Pratyush Tiwary and Michelle Parrinello reported a time-independent free energy estimator for MTD (Tiwary, 2015). This method presents a way to reconstruct the

evolution of the FEL making comparable one frame with the others, and then, to make possible the calculation of the standard deviation associated with each point of the landscape at the end of the simulation. The expression of the FEL estimator is presented in the Eq. II – 31.

$$\frac{F(s)}{k_b T} = -\frac{\gamma V(s,t)}{k_b \Delta T} + ln \int e^{\frac{\gamma V(s,t)}{k_b \Delta T}} ds \ \ where\, \gamma = \frac{T + \Delta T}{T} \quad (\boldsymbol{Eq.II - 31})$$

Then, the Well-T bias factor (γ) is normally chosen when it is known the activation barrier of an event. The value of this parameter for the standard MTD tends to infinite, then γ/ΔT tends to 1/T.

This method was applied for the same simulation using both Direct and Well-T MTD and the results shown that in Direct MTD the error is lower in minima than in the high energy regions, but it is randomly distributed along the surface. On the contrary, in Well-T simulations, the error is lower in minima than in maxima in a homogeneous way (Tiwary, 2015).

4.4. Collective variables (CVs)

The critical point that differentiate a good and a bad MTD simulation is the selection of the set of CVs. The CVs are functions of the microscopic coordinates of the system. To guarantee an effective application of the biasing potential, the CVs must fulfil the following requirements (Ensing, 2005) (Iannuzzi, 2003) (Laio, 2008).

- They must be a function of the microscopic coordinates of the system and the function must have a continuous derivative.
- They should distinguish between the initial and final states and describe all the relevant intermediates of the process under study.
- 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.
- They should be limited in number.

In this Thesis, the CVs chosen for each simulation of the systems are described in the respective chapters, being distances, angles, dihedrals or combination of them and puckering coordinates.

Chapter III – Conformational itinerary and catalytic mechanism of GH13 amylosucrase

Conformational itinerary and catalytic mechanism of GH13 amylosucrase

1. Introduction

Amylosucrase (AS, sucrose-1,4-α-glucan glucosyltransferase) is an exo-acting α-retaining transglycosylase from the glycoside-hydrolase family 13 (GH13), which gathers mainly hydrolytic enzymes involved in starch and amylose degradation. This enzyme is thus quite unique in the GH13 family as it is the only polymerase. Amylosucrase follows a double-displacement mechanism, involving two steps, common to all enzymes of the GH13 family. From sucrose substrate, an abundant agro-resource, this enzyme catalyses mainly the synthesis of a range of soluble oligosaccharides (called maltooligosaccharides) and unsoluble polysaccharide (amylose) α-1,4 linked glucopyranosyl residues. These reactions are accompanied by minor production of byproducts: fructose and glucose resulting from hydrolysis and sucrose isomers (turanose, trehalulose) formed by transfer of the glucosyl moiety onto fructose released during reaction (André, 2010). In addition to all these natural reactions, it has also been shown that amylosucrase has a large promiscuity toward non-natural acceptor substrates introduced in the reaction media, which could be glucosylated. The different reactions catalysed by this enzyme are summarized in Figure 3.1.

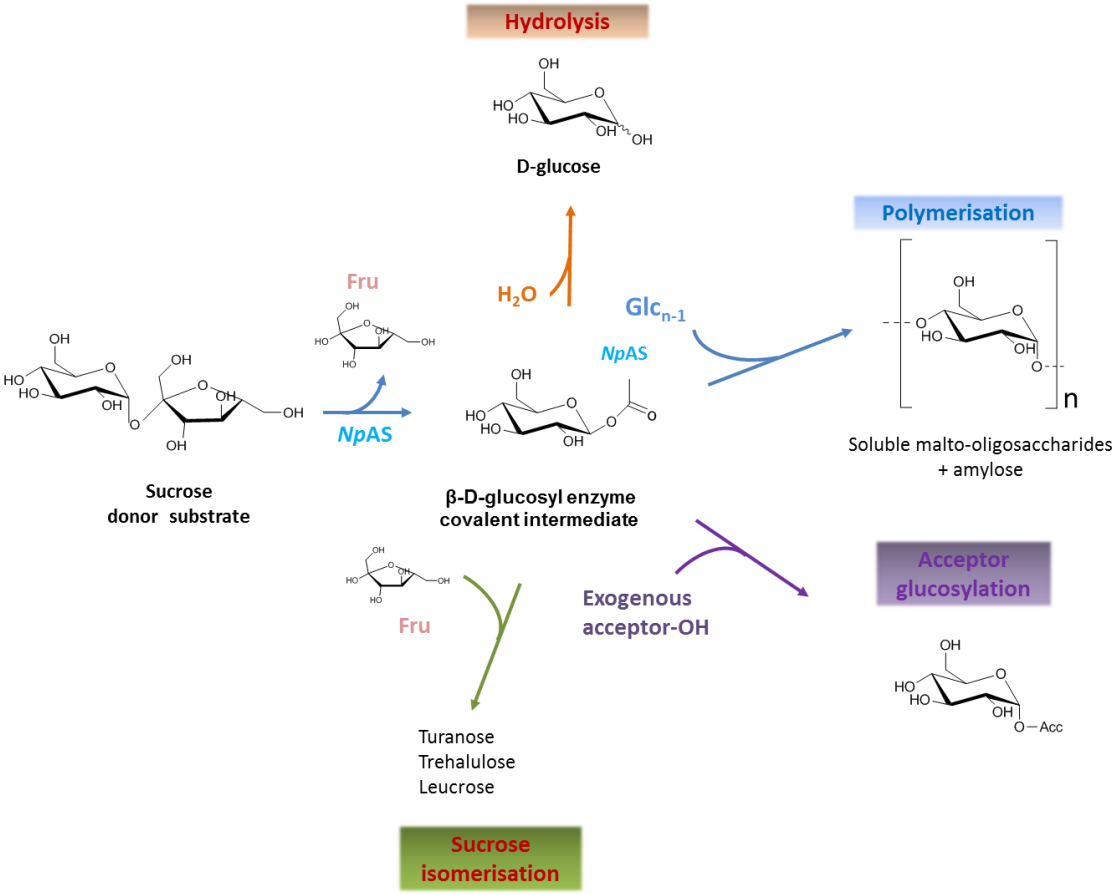


Figure 3.1. Reactions catalysed by amylosucrase.

The amylosucrase activity has been first reported in several species of *Neisseria* (*N. polysaccharea, N. canis, N. dentrificans, N. sicca, N. suflava)*. These bacteria belong to the microbiota of the buccal cavity and maybe involved in the formation of dental carries. The biological importance of amylosucrase in humans was demonstrated in 1983 (Riou, 1983), when the strain *Neisseria polysaccharea* isolated from the throat of European and African healthy children revealed the presence of extracellular amylosucrase. Since then, this enzyme has been found in other species of *Deinococcus* (*D. radiodurans, D. geothermalis, D. radiopugnans*), *Arthobacter* and *Alteromonas* (Moulis, 2016).

Among glucansucrase, amylosucrase is certainly by far the most studied enzyme at both structural and biochemical levels. It has also been widely exploited to access novel carbohydrate-based molecules for biotechnological applications and synthetic chemistry (Daudé, 2014). Additionally, molecular engineering of amylosucrase has been largely

reported to overcome some limitations of the enzyme in terms of selectivity, stability and catalytic efficiency and increase thus its utilization in biotechnologies.

Directed evolution and rational engineering strategies, relying on a better comprehension of structure-function relations and molecular modelling studies, have been applied to amylosucrase and led to several achievements such as enhanced thermal stability, improved catalytic activity on sucrose, controlled production of maltooligosaccharides or unsoluble amylose and glucosylation of non-natural substrates (polyols, flavonoids, carbohydrates, …), sometimes not recognized by parental wild-type enzyme (Daudé, 2014).

With the aim of improving our knowledge on the mechanism of action of AS, we used in this chapter QM/MM metadynamics to get a detailed description of the reaction pathway, the intermediates and transition states formed during the reaction as well as the energetics. This work has been done in collaboration with the group of Dr. Isabelle André (Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, INSA Toulouse, France), who works actively on amylosucrases, their engineering, characterization and applications. The study will be focused more particularly on the amylosucrase from *N. polysaccharea* (*Np*AS).

Available structures of amylosucrase from *Neisseria polysaccharea*

The cloning and characterization of amylosucrase from *Neisseria polysaccharea* was first reported in 1999 (Potocki de Montalk, 1999). Several three-dimensional structures of *Np*AS have been solved either in free form or in complex with various ligands (including sucrose, glucose, covalently bound glucosyl, maltoheptaose and turanose complexes). They are gathered in Table 3.1.

Table 3.1. List of X-ray crystallographic structures available for *Np*AS.

PDB code (ref.)	Type of structure	Resolution (Å)	Mutations	Substrate / -1 sugar conformation
1G5A (Skov, 2001)	apo	1.40	Gly1Ser, Leu2Pro, Ile3Asn, Leu4Ser, Gly537Asp	-
1JGI (Mirza, 2001)	MC	2.00	Glu328Gln	Sucrose / $^4C_1$
1S46 (Jensen, 2004)	GEI	2.20	Glu328Gln	D-glucosyl / $^4C_1$
1MW0 (Skov, 2002)	MC	2.01	Glu328Gln	Maltoheptose / $^4C_1$

In 2001, structure of the wild-type *Np*AS was reported (PDB 1G5A, [Skov, 2001]). Simultaneously, a structure of the Glu328Gln mutant (Glu328 is the acid/base residue) of *Np*AS in complex with the sucrose natural substrate came out (PDB 1JGI [Mirza, 2001]). The α-glucose at the -1 subsite was found not to be distorted ($^4C_1$ conformation, Figure 3.2 [a]). In 2002, the same acid/base mutant was crystallized with maltoheptaose oligosaccharide (PDB 1MW0, [Skov, 2002]), evidencing a pocket active site formed by 6-7 sugar subsites (Figure 3.2 [c]), none of them bearing a distorted sugar. In 2004, the same mutant was crystallized with a D-glucosyl bound to the nucleophile residue, Asp286, forming the glycosyl-enzyme reaction intermediate (GEI, see Figure 1.9), also named the covalent intermediate (PDB 1S46 [Jensen, 2004]). Again, the -1 sugar ring was found not to be distorted ($^4C_1$, Figure 3.2 [b]).

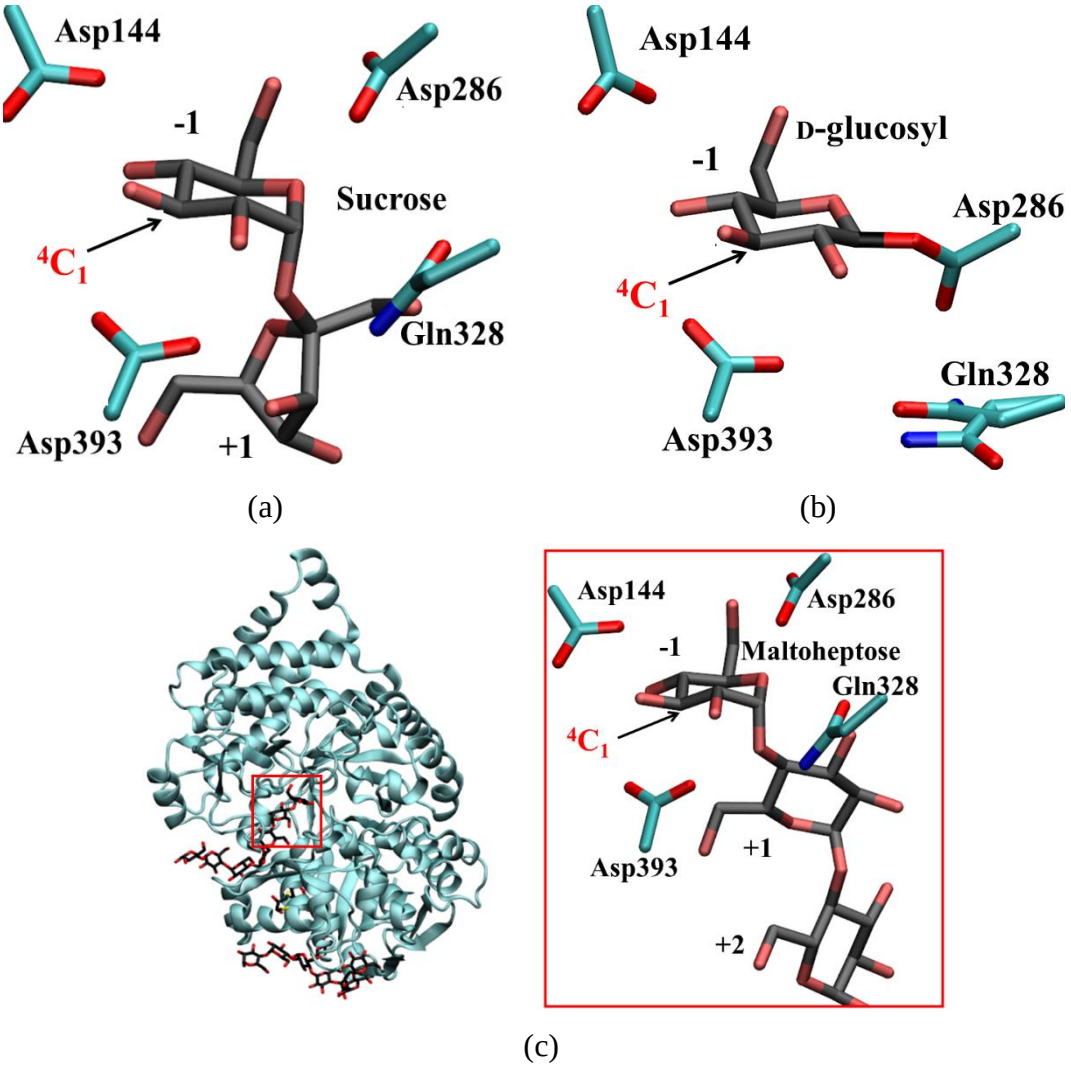


Figure 3.2. (a) Close view of the active site of the Michaelis complex structure of *Np*AS with sucrose (Mirza, 2001). (b) Active site structure of the GEI of *Np*AS with α-glucose (Jensen, 2004). (c) Overall enzyme structure of the Michaelis complex of *Np*AS with maltoheptose and a zoom of its active site structure (Skov, 2002). This structure has an additional maltoheptose molecule docked in the surface of the enzyme, interacting with hydrophilic aminoacids.

Altogether these structural information revealed that *Np*AS is formed by five domains of which three are common to all GH13 enzymes, namely the A, B and C domains (Figure 3.3). The catalytic A domain is formed by a $(\beta/\alpha)_8$- barrel, the B-domain is an extension of the loop3 from the catalytic barrel, the C-ter domain is formed by an eight-stranded β-sheet Greek key. Two additional domains (N and B') are specific to

amylosucrases. The N-terminal domain is formed of α-helices, and the B' domain is an extension of the loop 7 from catalytic barrel and it is responsible for the pocket-like topology of the active site. It is composed of two helices followed by a β-strand and is terminated by a last short α-helix.

The catalytic site is located at the bottom of a narrow and 17 Ang. Deep pocket, closed at the bottom by a salt bridge between residues located respectively in the second and eight loops of the $(\beta/\alpha)_8$-barrel (Skov, 2001).

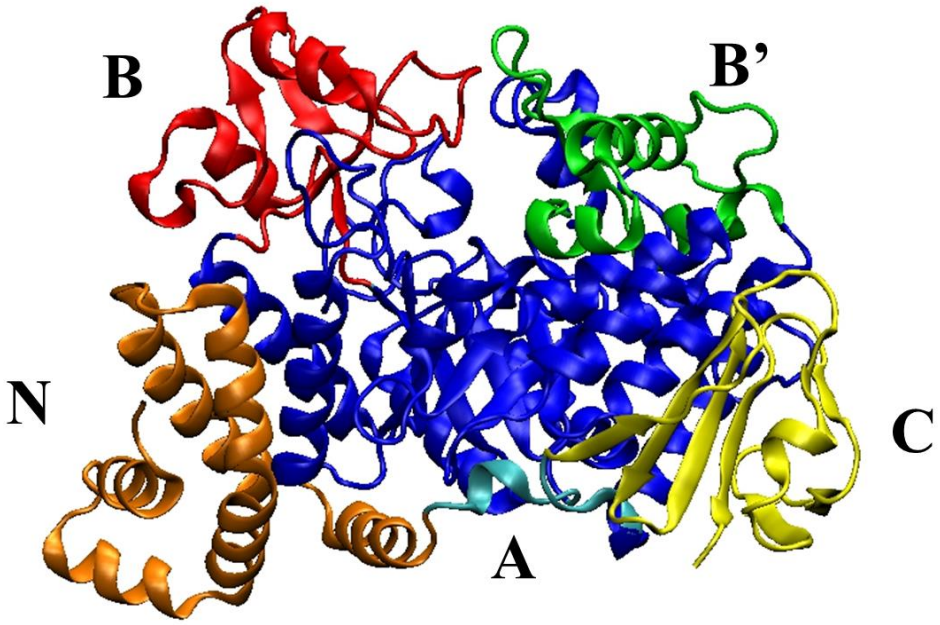


Figure 3.3. Overall enzyme structure of the apo form of *Np*AS and colour classification for the five domains forming the enzyme (Cyan fragments are unclassified chains or connectors between domains).

The polar residues forming the -1 subsite of the enzyme are shown in the Figure 3.2. Glu328 is the acid/base residue (mutated to Gln in the experimental structure). It acts as a proton donor during the retaining mechanism of hydrolysis. Asp286 acts as a nucleophile, which forms a covalent bond with the sugar ring (Figure 3.2b). Aspartates 144 and 393 help to maintain the sugar in the chair conformation (Asp393 interacts with the 2-OH and 3-OH glucosyl sugar substituents and Asp144 interacts with 4-OH).

The available structures for family GH13 members, in addition to that of amylosucrase, have been discussed in Chapter I (see Table 1.1, Section I – 4.3). Notably,

there are two GEI structures available, which differ on the conformation of the -1 sugar. On the one hand, the GEI of amylosucrase exhibits a $^4C_1$ conformation. Taking into account that all – except α-amylase, discussed below – MC structures of GH13 members display a $^4C_1$ conformation, this suggests that the glycosylation conformational itinerary is cyclic ($^4C_1 \rightarrow TS \rightarrow {}^4C_1$). On the other hand, the GEI of *Bifidobacterium adolescentis* sucrose phosphorylase shows the -1 sugar distorted in a $^1S_3$ conformation, suggesting a completely different conformational pathway ($^4C_1 \rightarrow TS \rightarrow {}^1S_3$) (Mirza, 2006). Being both enzymes from the same family and acting over the same substrate, this discrepancy is puzzling.

Regarding α-amylase, there are three available MC structures, whose -1 sugar rings show different conformations: an enzyme complex with α-cyclodextrin (^2S_O) (Tan, 2008), a Glu208Gln mutant (residue 208 is the acid/base) in complex with maltopentaose ($^4C_1$) (Fujimoto, 1998) and a Asp300Asn mutant (residue 300 is equivalent to 393 in amylosucrase and interacts with the 2- and 3-OH) in complex with maltooctaose ($^2E/^2H_3$). Furthermore, two experimental works tried to mimic the GEI using as substrates a fluorocompound of the α-glucose ($^4C_1$) (Zhang, 2009) and a glucosyl epi-cyclophellitol ($^4C_1$, the nucleophile shows two different orientation states) (Caner, 2016). However, these results were not conclusive on the sugar conformation because the limited mimicking properties of the sugar-like molecules employed.

Recently, a computational QM/MM work of the hydrolysis mechanism of GH13 α-amylase was reported (Pinto, 2015), using the ONIOM scheme (Svensson, 1996) over a model reconstructed from the α-amylase – acarbose complex (PDB 1CPU, resolution 2.0 Å) (Brayer, 2000). A potential energy surface (PES) with respect to two reaction coordinates, followed by calculation of the reaction pathway following the intrinsic reaction coordinate (IRC) (Kui, 1981) was reported. Even though, the question of the conformational itinerary of the -1 sugar was not addressed, it is easy to infer the evolution of sugar conformation from one of the manuscript figures, reproduced in Figure 3.4.

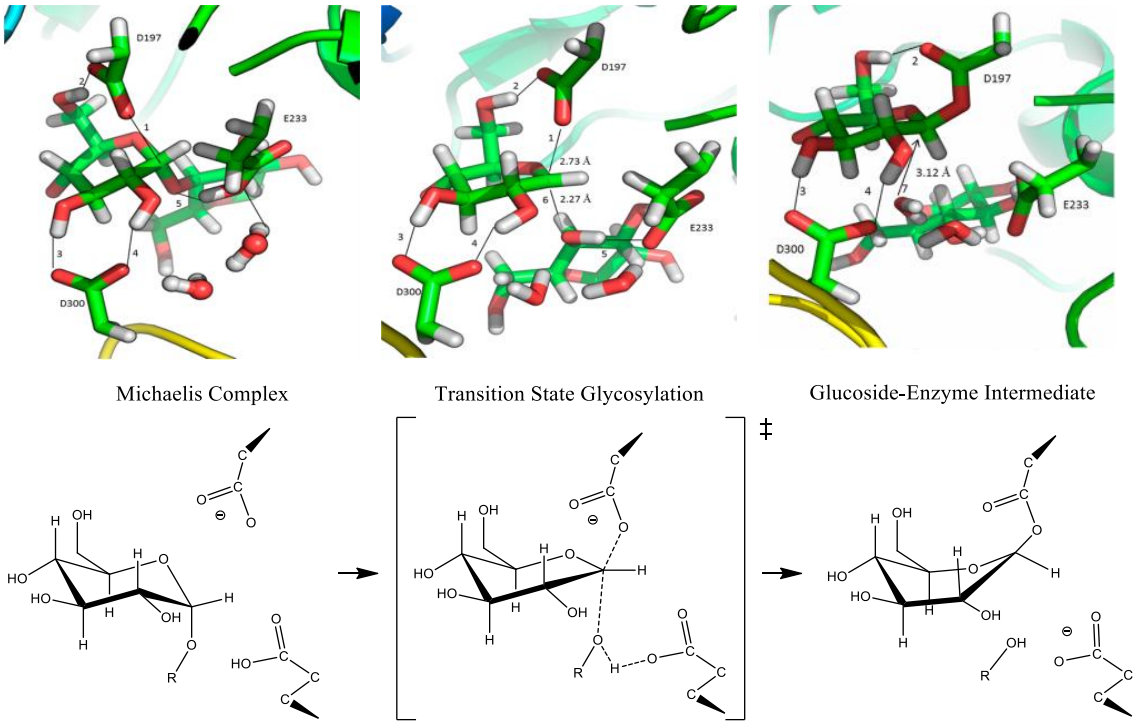


Figure 3.4. (Top) Three-dimensional structures of the reactants, transition state and covalent intermediate obtained by Pinto et al. during the simulation of the glycosylation step of GH13 α-amylase. (Bottom) Mechanism obtained computationally. Figure adapted from reference (Pinto, 2015).

According to the structures shown in the top panel of Figure 3.4, the substrate of α-amylase follows a cyclic conformational itinerary, i.e. the -1 sugar starts as $^{4}C_{1}$, evolves towards an envelope E_{3} and returns to the initial $^{4}C_{1}$ chair conformation once the covalent intermediate is formed. In contrast, the schematic representation (Figure 3.4, bottom) surprisingly shows that the -1 sugar is distorted ($^{1,4}B$) in the GEI, which is also what is mentioned in the manuscript text. This implies that the -1 sugar follows a $^{4}C_{1} \rightarrow {}^{4}E \rightarrow {}^{1,4}B$ itinerary, which is far away from what the optimized structures show. Thus it remains unclear what is the conformational itinerary of GH13 α-amylase and how does it compares with AS. Taking into account that AS is a transglycosylase, whereas α-amylase and sucrose phosphorylase are hydrolases, it is reasonable to think that they can follow different itineraries. However, we do not see any reason why the GEIs of both hydrolases (α-amylase and sucrose phosphorylase) should adopt different conformations.

Concerning kinetics, the catalytic constant (k_{cat}) of the hydrolysis reaction of *Np*AS is 33 min^{-1} (Potocki de Montalk, 2000), which, according to transition sate theory, corresponds to an activation barrier of ≈ 17.9 kcal/mol at 300 K. The identity of the rate limiting step for *Np*AS is not known, but, taking into account the kinetic results presented by Pinto et al. (Pinto, 2015), we can assume it might be the glycosylation step. The aforementioned QM/MM analysis of α-amylase reported a potential energy profile with a rate-limiting glycosylation step (Pinto, 2015).

Understanding better the reaction mechanism of AS at electronic level is crucial to guide the design of AS inhibitors as well as enzyme mutants to develop efficient biocatalysts for biotechnological applications. In this work, we uncover the conformational itinerary of the α-glucosyl unit during catalysis, considering both the hydrolysis and transglycosylation reaction (Figure 3.5).

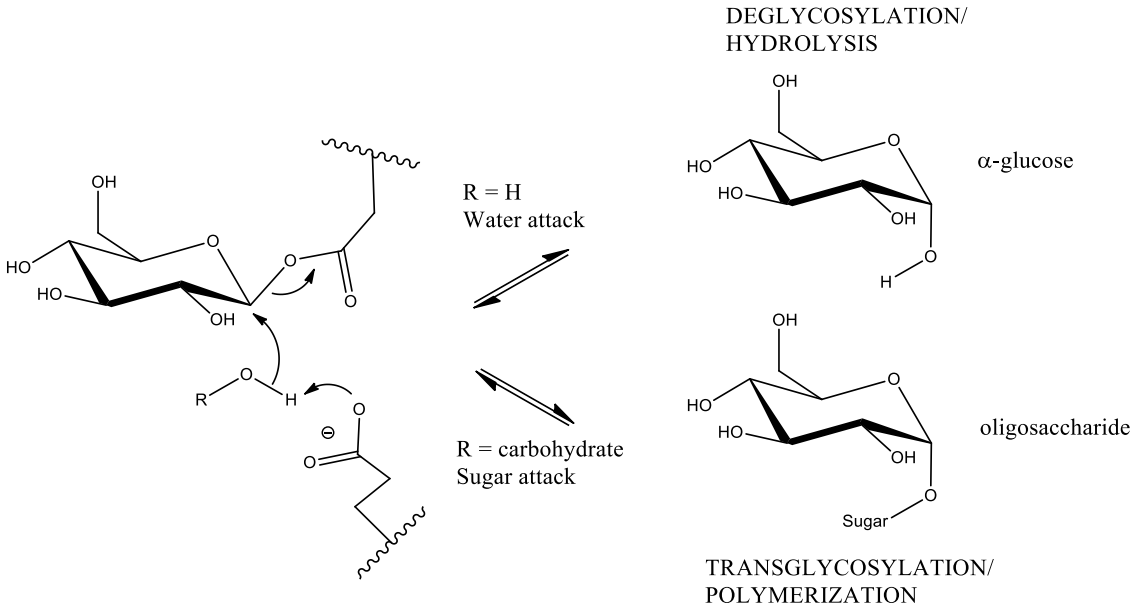


Figure 3.5. A schematic representation comparing the molecular changes in the enzyme deglycosylation step of hydrolysis and the transglycosylation step of polymerization process.

2. Objectives

The main objectives of the present chapter are:

1) To determine the most stable conformation of the α-glucosyl ring in the -1 subsite of amylosucrase.
2) To model the hydrolysis reaction (enzyme glycosylation and deglycosylation steps) of the natural substrate, sucrose, in *Np*AS, deciphering the conformational pathway of the -1 sugar ring during catalysis.
3) To model the transglycosylation reaction in *Np*AS, leading to the formation of α-1,4-maltose disaccharide.

3. Computational details

Model building

The initial structure of the AS-sucrose complex was taken from the Michaelis complex structure of the Asp328Asn mutant (Mirza, 2001, PDB 1JGI, Table 3.1). Two MD simulations were performed, one simulation of the MC complex of the mutant enzyme at the temperature of the X-ray structure determination (100 K), to check that the force field used reproduces the experimental structure, and another simulation of the MC complex of the wild-type (WT) enzyme at 300 K (the mutation was manually reverted). In both cases, the enzyme was parametrized with AMBER 11 FF99 (Cornell, 1995), whereas the Glycam 06 (Kirshner, 2008) force-field was used to parametrize the sugars. The protonation states and hydrogen atom positions of all histidine residues were selected based on their hydrogen bond network (Appendix B). All Asp and Glu residues were taken as deprotonated (i.e. negative charge) except Glu328 (the acid-base residue). Both systems were solvated with 20,725 TIP3P water molecules (Jorgensen, 1983) and 16 sodium ions were added to achieve neutrality of the protein structure, forming a parallelepiped box with dimensions of 89.1 Å x 106.4 Å x 81.1 Å. The results of the 100 K simulation are described in appendix B, whereas in the following we refer to the simulations of the WT enzyme at 300 K.

Simulation protocol

The following simulation protocol was used. First of all, the system was equilibrated at 300 K with AMBER (see details below). Then, a representative structure was chosen to construct a QM/MM model, which was equilibrated further. A snapshot of the last simulation was used to investigate (1) the enzyme glycosylation reaction (Figure 1.9); (2) the approach of one water molecule to the active center ("water approach simulation"); (3) the enzyme deglycosylation reaction; (4) the transglycosylation reaction.

Classical MD simulations

The AS-sucrose complex was initially subjected to classical molecular dynamics (MD) using the AMBER11 software. 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, spatial constraints were initially added to the interactions between the protein and the substrate, while water molecules and sodium ions were allowed to move freely. The constraints were then removed and the whole system was allowed to reach the desired temperature. During all the process, the acid/base residue (Glu328) showed an erratic behaviour (the proton position evolved from cis to trans conformations and vice versa), thus its relative position with respect to the sugar was controlled with wall restraints (force constant of 5 kcal/mol/Å^2). The simulation was pursued for 100 ps at constant pressure, allowing the cell volume to evolve, until the density stabilized (~1.06 g/cm^3). The MD simulation was extended to 15 ns at constant volume until the system had reached equilibrium. A timestep of 1 fs was used, increasing it to 2 fs during the last 14 ns. The –1 sugar ring conformation evolved from $^4C_1$ to $B_{3,O}$ and returned to the original $^4C_1$ conformation during the last 7 ns of the simulation. Analysis of the trajectories was carried out by using standard tools of AMBER and VMD. The root-mean-square-deviation (RMSD) of the protein backbone atoms with respect to the crystal structure was found to oscillate around at 1.4 Å in the equilibrated structures (see Appendix B). One snapshot from the last 500 ps of simulation was taken as starting structure for the subsequent enzyme-substrate QM/MM MD simulations.

QM/MM simulations

QM/MM simulations were performed with the CPMD software (Car & Parrinello, 1985). The following QM regions were used for each simulation (see Figure 3.3):

(i) Conformational free energy landscape (FEL) of the -1 sugar (α-glucosyl): the QM region included the whole sucrose molecule. The collective variables used to reconstruct the conformational FEL were the Cremer and Pople (Cremer, 1975) puckering coordinates (ϕ, θ), resulting in a Mercator representation.

(ii) Glycosylation step: the QM region includes the sugar, the acid/base residue (Glu328, capped at the C_β) and the nucleophile residue (Asp286, capped at the C_α).

(iii) Water approach simulation: the QM box includes the two sugars, the two catalytic residues (see above) and a neighbouring water molecule.

(iv) Deglycosylation step: the QM box includes the same atoms as in iii.

(v) Transglycosylation reaction: to model this reaction, the fructose leaving group was manually substituted by an α-glucose. The QM region includes the catalytic residues, the α-glucosyl group covalently bound to the enzyme and the new α-glucose.

In all cases, the frontier atoms between QM and MM region were described using pseudopotential carbon link atoms. Following test of different E_{mass} and $\tau_{step,}$ values of 600 a.m.u and 0.12 fs, respectively, were taken for all simulations. Further details are shown in the Appendix F.

QM/MM metadynamics simulations

QM/MM metadynamics simulations were performed to characterize the conformational FEL of the -1 α-glucosyl ring in the active site of *Np*AS and to simulate the different steps of the enzymatic reaction. The following collective variables were used:

(i) Conformational FEL: the Cremer-Pople puckering coordinates (Cremer & Pople, 1975) phi and theta (ϕ, θ) of the -1 sugar ring were used as collective variables, following the methodology previously developed in our group to rationalize and predict catalytic itineraries in several GH families (Biarnés et al. 2007; Ardèvol & Rovira, 2015).

(ii) Glycosylation step: three collective variables representing the proton transfer (CV1), the nucleophilic attack (CV2) and the glycosidic bond cleavage (CV3) were used (Figure 3.6a). Unless specified, CVs were taken as interatomic distances or distance differences.

(iii) Water approach: two collective variables, representing the interaction between the nucleophile and the 6-OH arm (CV1) and the water – C_1 distance (CV2), were used (Figure 3.6b).

(iv) Deglycosylation step: three collective variables representing the covalent enzyme-substrate interaction (CV1), the water attack (CV2) and the proton transfer (CV3) were used (Figure 3.6c).

(v) Transglycosylation reaction: as mentioned above, the system was manually built from the structure of the glycosylation products. The fructose molecule was deleted and substituted by an α-glucose with the 4-OH in a nucleophilic attack orientation. The corresponding structure was equilibrated by classical MD and, subsequently, by QM/MM (see details in appendix F). Three collective variables representing the covalent enzyme-substrate interaction (CV1), the 4-OH α-glucose attack (CV2) and the proton transfer (CV3) were taken, as shown in Figure 3.6d.

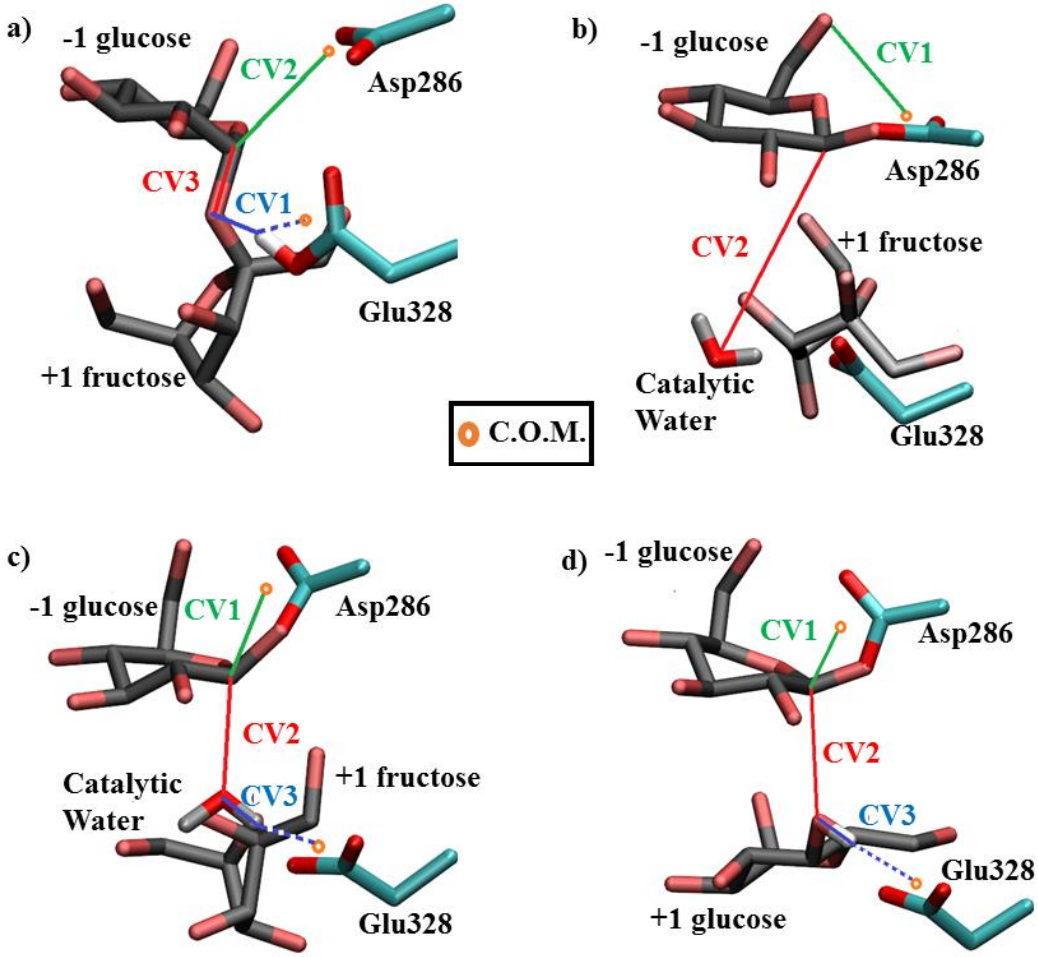


Figure 3.6. QM regions and chosen CVs to simulate: (a) glycosylation step, (b) water approach process, (c) deglycosylation step and (d) transglycosylation in *Np*AS. C.O.M. is the center of mass of the aspartate/glutamate oxygens. C. O. M. is the center of mass of the two oxygen atoms of a given carboxylate side chain.

Metadynamics simulations full details and parameters are presented in Appendix F. In case of chemical reactions, all simulations were considered converged once the number of recrossings around the transition state was two (the system evolved from reactants to products and returned to the initial state).

4. Results

4.1. Classical MD simulations

The results of the classical MD of the Glu328Gln mutant and WT amylosucrase models are presented in the Appendix B. The RMSD of the mutant backbone was stabilized at 0.6 Å at 100 K. The conformation of the active site and the sugar remained as in the crystallographic structure, which validates the method employed. In the case of the WT system, the conformation of the α-glucosyl ring at the -1 subsite (hereafter referred as *-1 sugar*) changes from $^4C_1$ to $B_{3,O}$ during the first 6 ns of NVT production dynamics, due to an increase in the distance between the nucleophile and the anomeric carbon, and returns to $^4C_1$ during the last 8 ns of MD simulation. This conformation is maintained during the QM/MM MD equilibration. The RMSD of the backbone stabilizes at 1.4 Å.

4.2. Conformational study of the α-glucosyl ring at the-1 subsite

The conformational FEL of the α-glucosyl ring at the -1 subsite, analysed by QM/MM metadynamics, is shown in Figure 3.7. As previously observed during the classical MD, the -1 sugar adopts two conformations: the most stable $^4C_1$ chair and the $B_{3,O}$ distorted conformation, which is 4.7 kcal/mol less stable (the statistical error of the metadynamics simulation is ≈ 0.8 kcal/mol, Appendix C) (Tiwary, 2015).

The main enzyme-sugar interactions for each conformer are shown in Figure 3.8 and their respective values are listed in Table 3.2. The glycosidic bond distances (C_1-O_g) are 1.44 and 1.42 Å, for $^4C_1$ and $B_{3,O}$, respectively. The internal ring bond (C_1-O_p) is

complementary to the C_1-O_g distance. (1.42 and 1.45 Å, respectively). Notably, the distance between of the nucleophile oxygen ($O_{393\text{-}1}$) and the anomeric carbon (C_1) is higher when the sugar is distorted, with values of 3.10 Å ($^4C_1$) and 3.30 Å ($B_{3,O}$), indicating steric hindrance between the hydrogen and the nucleophilic oxygen of the Asp286.

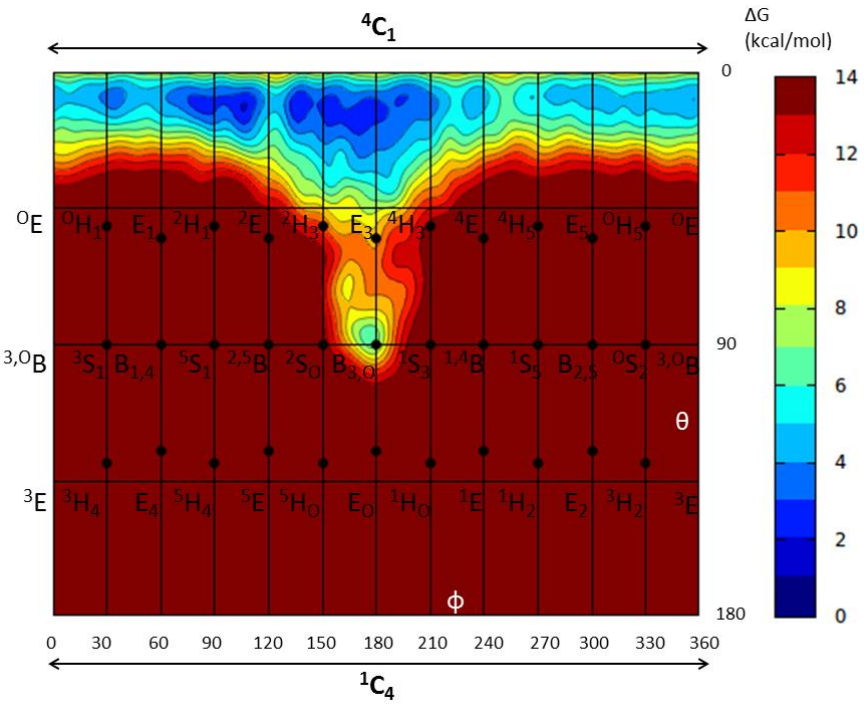


Figure 3.7. Mercator projection of the conformational FEL of the α-glucosyl ring at the -1 subsite of *Np*AS. Isolines at 1 kcal/mol.

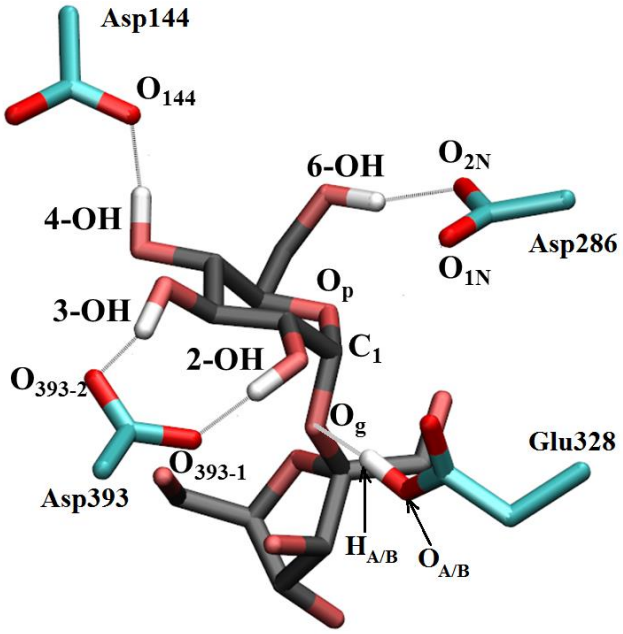


Figure 3.8. Representation of the most important hydrogen bond interactions between the -1 sugar and the enzyme active site residues.

Table 3.2. Calculated values of the most important AS-glucose interactions and its standard deviations over the FEL minima (energies in kcal/mol, distances in Å and puckering coordinates in degrees).

Conformation	**$^4C_1$**	**$B_{3,O}$**
ϕ, θ	136º, 10º	176º, 87º
ΔG	0.0	4.7
d(C_1 – O_g)	1.44 ± 0.05	1.42 ± 0.04
d(C_1 – O_p)	1.42 ± 0.04	1.45 ± 0.04
d(2-OH – O_{393-1})	1.51 ± 0.10	1.54 ± 0.09
d(3-OH – O_{393-2})	1.60 ± 0.16	1.66 ± 0.16
d(4-OH – O_{144})	1.49 ± 0.09	1.50 ± 0.09
d(6-OH – O_{2N})	1.63 ± 0.24	1.66 ± 0.16
d(C_1 – O_{1N})	3.10 ± 0.12	3.30 ± 0.16

4.3. Simulation of the first reaction step (enzyme glycosylation)

The simulation of the enzyme glycosylation reaction was initiated from a snapshot of the global minimum of the conformational FEL ($^4C_1$ conformation). The values of the collective variables, relevant distances and puckering coordinates of the -1 ring during the metadynamics simulation are provided in Appendix B.

The glycosylation FEL reconstructed from the metadynamics simulation, along with its 2D projections, are shown in Figure 3.6. The FEL exhibits two important minima in opposite regions: reactants (R) and products (P), being the second 1.46 kcal/mol more stable. During the simulation, the system evolved from R to P, and it returned to the initial state, passing through a transition state (TS), with a free energy barrier of 17.25 kcal/mol. The computed free energy barrier is in very good agreement with experimental value (17.9 kcal/mol) estimated from the rate constant at room temperature.

The minimum free energy pathway (MFEP) was computed according to the intrinsic reaction coordinate (IRC) (Fukui, 1981). To analyse the evolution of the system along the MFEP, we performed a statistical analysis along the MFEP. All structures in a "box" defined by CV1 ± 0.2, CV2 ± 0.2, CV3 ± 0.2 Å were extracted and several structural parameters (CVs, distances/bonds and puckering coordinates) were analysed for each "box" along the reaction pathway. Six characteristic points along the MFEP were chosen for analysis: MC, A, TS_1, B, C, GEI (see Figure 3.10).

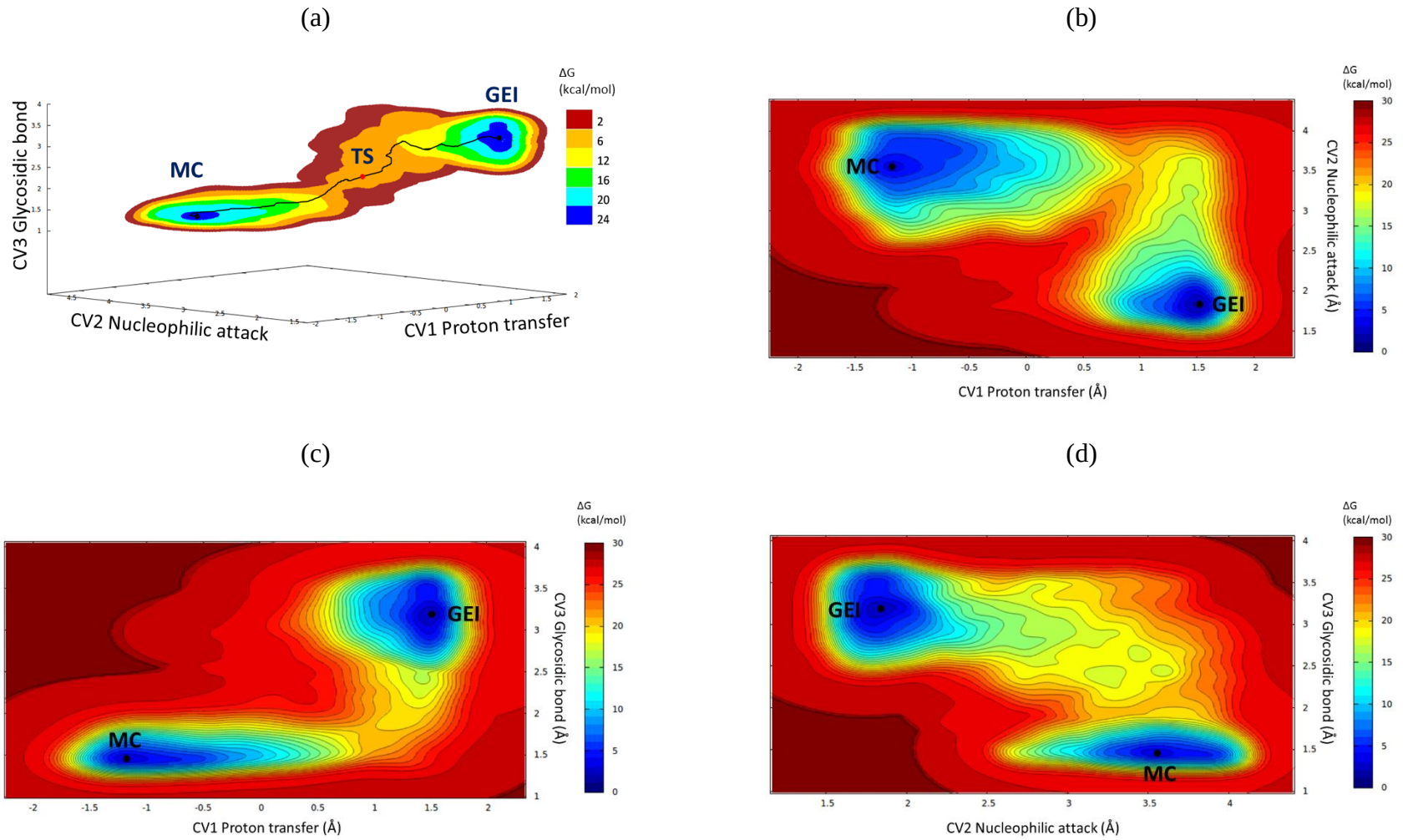


Figure 3.9. (a) Three dimensional (3D) FEL of the glycosylation reaction of *Np*AS. The bold black line indicates the MFEP. (b, c, d) Two dimensional (2D) projections on each pair of collective variables. Isolines at 1 kcal/mol.

Figure 3.10 shows the evolution of the main catalytic distances and the distortion of the glucose ring during the glycosylation reaction, which can be divided into five phases:

(I) *Approach of Glu328 to the glycosidic oxygen*

During this phase, the distance between the Glu328 acid/base proton ($H_{A/B}$) and the glycosidic oxygen (O_g) decrease from 2.5 to 1.5 Å. The conformation of the -1 sugar remains at $^4C_1$.

(II) *Proton transfer and formation of a hydrogen bond between Glu328 and fructose*

The d($H_{A/B}$ – O_g) evolves from 1.5 to 1.0 Å (bond formation) and the distance between the proton $H_{A/B}$ and the Glu328 oxygen ($O_{A/B}$) evolves from 1.0 to 1.5 Å (bond cleavage). The conformation of the -1 sugar changes from $^4C_1$ to $^4H_3/^4C_1$.

(III) *Breaking of the glycosidic bond*

The glycosidic bond (C_1 – O_g) enlarges from 1.9 to 3.1 Å (bond cleavage). The conformation of the -1 sugar changes from $^4H_3/^4C_1$ to E_3.

(IV) *Formation of the covalent glycosyl enzyme intermediate*

In this phase, the distance between the nucleophilic oxygen of Asp286 (O_{1N}) and the anomeric carbon (C_1) decreases from 3.0 to 1.5 Å (glycosyl-enzyme bond formation). The conformation of the sugar is a mixture of $^4H_3/^4E$ and $^4C_1$ conformations (we cannot differentiate the two along the simulation).

(V) *Reorientation of Asp286 and "unexpected" conformational change of the -1 sugar*

Finally, the hydrogen bond interaction between the second oxygen of Asp286 (O_{2N}) and the 6-OH breaks and Asp283 rotates around its covalent bond with the -1 sugar. The -1 sugar collapses to an undistorted $^4C_1$ conformation.

Representative structures of the active site along the reaction are presented in Figure 3.11 and Table 3.3 lists the most important structural parameters. The breaking of the glycosidic bond (phase III) happens before the formation of the GEI (phase IV). Likewise, the TS features a glycosidic bond partially broken (C_1-O_g = 1.98 ± 0.10 Å, Table 3.3) whereas the glycosyl-enzyme bond distance (2.98 ± 0.11 Å) is similar to that of the MC (3.13 ± 0.09 Å). Therefore, the reaction follows a dissociative mechanism and can be described as D_NA_N (Guthrie, 1989) (Schramm, 2001).

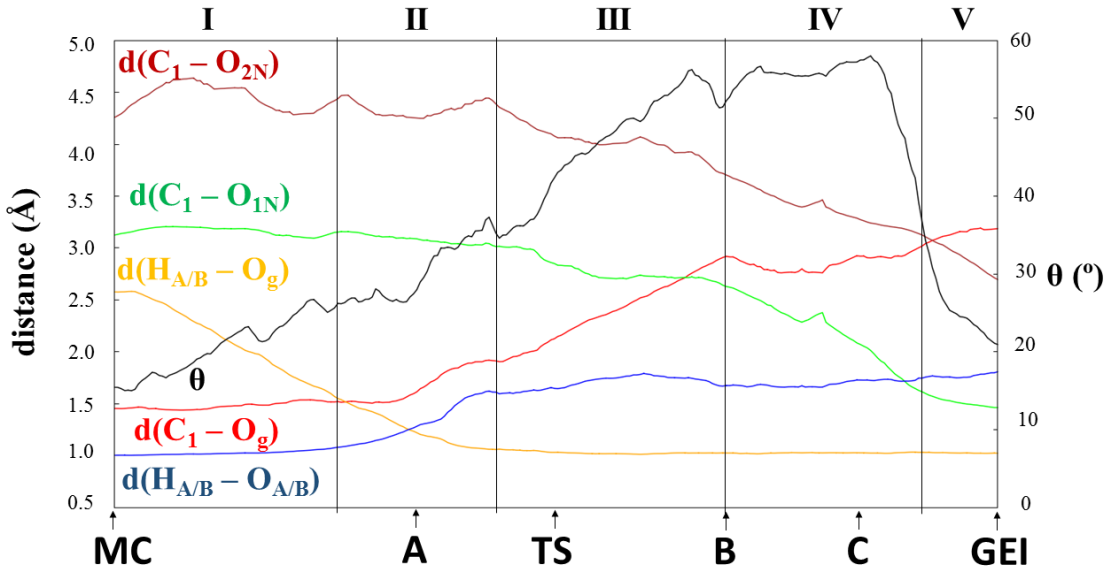


Figure 3.10. Evolution of the main distances and the conformation of the -1 ring during the glycosylation reaction of *Np*AS. See Figure 3.8 for the atomic labelling.

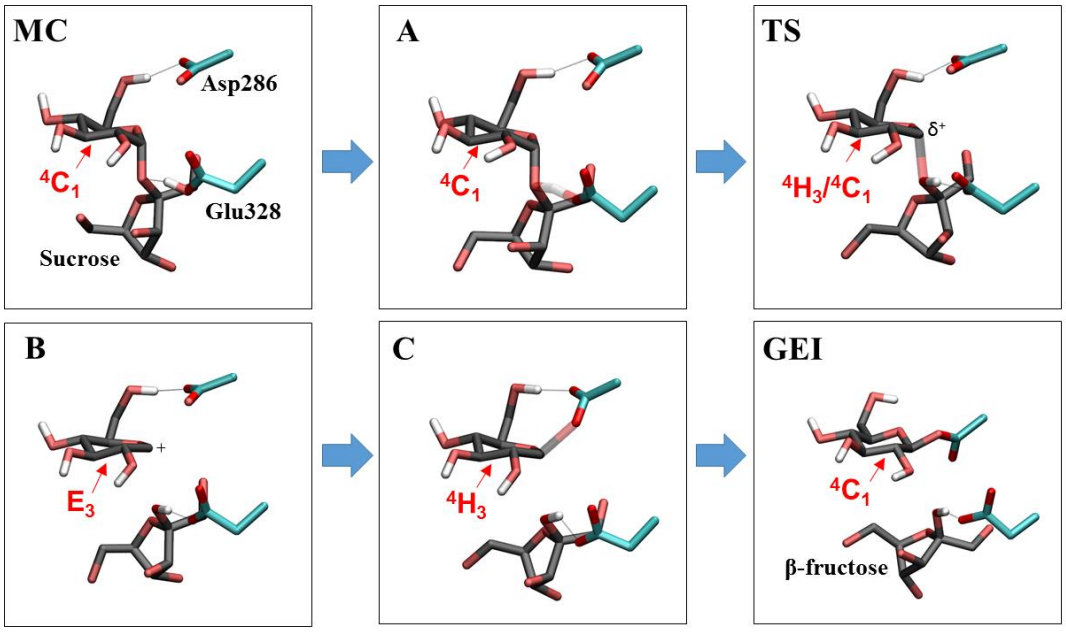


Figure 3.11. Representative structures along the reaction coordinate of the glycosylation reaction.

Table 3.3. Calculated values of the most relevant catalytic distances (in Å) and puckering coordinates (in degrees) and its standard deviations along the glycosylation MFEP.

	ΔG	d(C_1-O_{1N})	C_1-O_g	d($H_{A/B}$-O_g)	d($H_{A/B}$-$O_{A/B}$)	ϕ	θ	Conformation
MC	0.0	3.13 ± 0.09	1.46 ± 0.11	2.58 ± 0.07	1.00 ± 0.03	188 ± 30	16 ± 7	$^4C_1$
A	14.2	3.08 ± 0.10	1.58 ± 0.12	1.25 ± 0.08	1.25 ± 0.08	203 ± 17	27 ± 8	$^4C_1$
TS	17.3	2.98 ± 0.11	1.98 ± 0.10	1.05 ± 0.05	1.61 ± 0.12	204 ± 14	37 ± 8	$^4H_3/^4C_1$
B	14.4	2.72 ± 0.14	2.79 ± 0.12	1.02 ± 0.04	1.72 ± 0.13	190 ± 11	55 ± 8	E_3
C	11.7	2.14 ± 0.17	2.89 ± 0.12	1.02 ± 0.04	1.71 ± 0.12	206 ± 14	57 ± 8	$^4H_3$
GEI	-1.5	1.46 ± 0.05	3.18 ± 0.11	1.02 ± 0.04	1.78 ± 0.13	296 ± 29	21 ± 8	$^4C_1$

It is interesting to analyse the evolution of the conformation of the -1 sugar during the glycosylation reaction, defined by the θ puckering coordinate (Figure 3.12). The conformational itinerary is cyclic on the puckering surface, i.e. it starts and ends up in the same conformation ($^4C_1 \rightarrow {}^4H_3/E_3 \rightarrow {}^4C_1$). From the MC to the TS, the conformation evolves from $^4C_1$ to $^4H_3/^4C_1$. Afterwards, the sugar adopts an E_3 envelope conformation (point B), followed by a $^4H_3$ half-chair (point C). At this point, a sudden conformational change towards $^4C_1$ takes place, concomitant with rotation of Asp286. However, the covalent intermediate is already formed before the rotation (GEI' in Figure 3.12). Between C and GEI' the system maintains the 6-OH – O_{2N} interaction and the -1 sugar exhibits a $^4H_3/^4E$ conformation (θ ~ 58º). Afterwards, the system does not maintain the 6-OH – O_{2N} interaction and the -1 sugar exhibits a $^4C_1$ conformation (θ ~ 20º). This is the reason why the puckering coordinate value in the GEI' point is an average value (θ ~ 40º). Thus, we conclude that are two different GEI structures that are isoenergetic. Interestingly, only the $^4H_3/^4E$ conformation of the GEI is preactivated for nucleophilic attack by a water molecule or a sugar (deglycosylation/transglycosylation processes). As we will see later on 6-OH – O_{2N} interaction has an important role in the subsequent reaction steps.

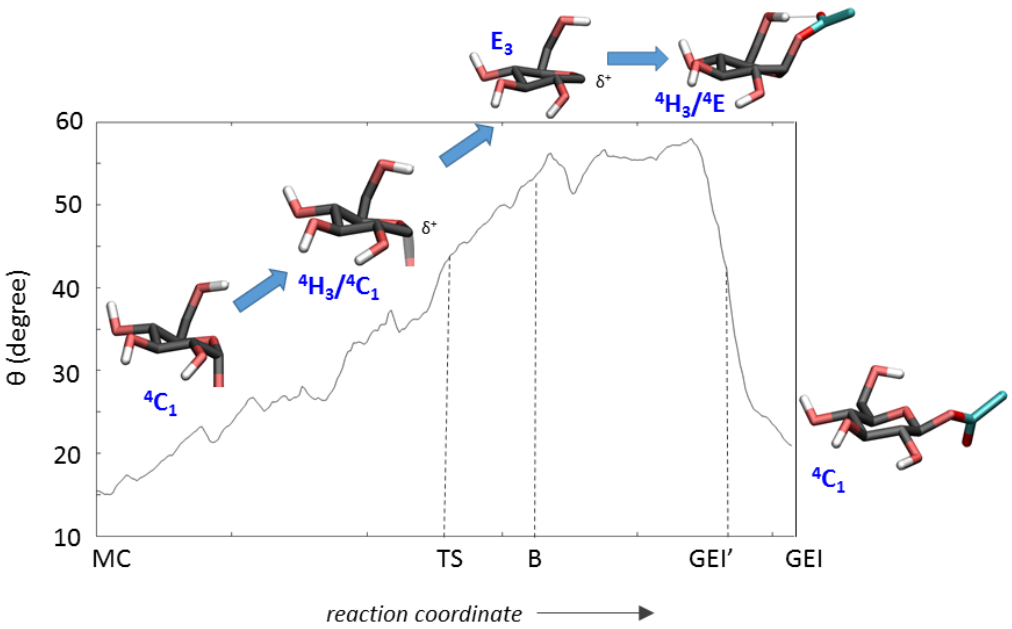


Figure 3.12. Conformational pathway followed by the sugar during the enzyme glycosylation step.

4.4. Water approach simulation

The simulations of the glycosylation reaction show that one water molecule remains at ~ 6 Å from the anomeric carbon (Figure 3.13). Its position is stabilized by hydrogen bond interactions with the acid/base residue (Glu328) and the auxiliary aspartate (Asp393). This water molecule is the best candidate to act as a nucleophile in the deglycosylation process, but it is still too far and not well oriented to attack the anomeric carbon (Figure 3.13). Most likely, there is an energetic barrier to bring the water molecule to a reactive configuration.

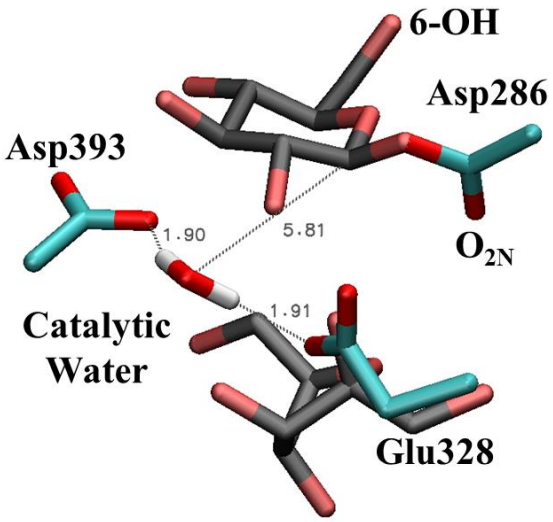


Figure 3.13. Active site structure after the glycosylation reaction. The orientation of the catalytic water is determined by two hydrogen bond interactions with Asp393 and Glu328.

To obtain the energy barrier associated to the approach of the water molecule to the anomeric carbon, we designed a new metadynamics simulation, using two collective variables: the separation between Asp286 and the 6-OH substituent (CV1) and the distance between the oxygen atom of the water molecule and the anomeric carbon (CV2). This CV is necessary to differentiate the two configurations of the GEI previously identified (GEI amd GEI', see Figure 3.12), which can be interconverted via the 6-OH$\cdots O_{2N}$ interaction. CV2 was defined as the distance between the center of mass of the two carboxylate oxygens of Asp286 and the H atom of the 6-OH. The evolution of the collective variables during the metadynamics simulation is provided in Appendix B.

The FEL reconstructed from the metadynamics simulation is shown in Figure 3.14. The FEL shows two visible regions, north and south, separated by a saddle point (TS_w). GEI_w, in which the catalytic water is at 5.3 Å, is the most stable minima, being 3.5 kcal/mol more stable than the GEI found in the previous step (Figure 3.11). This state corresponds to the lowest-energy form of the glycosyl-enzyme intermediate.

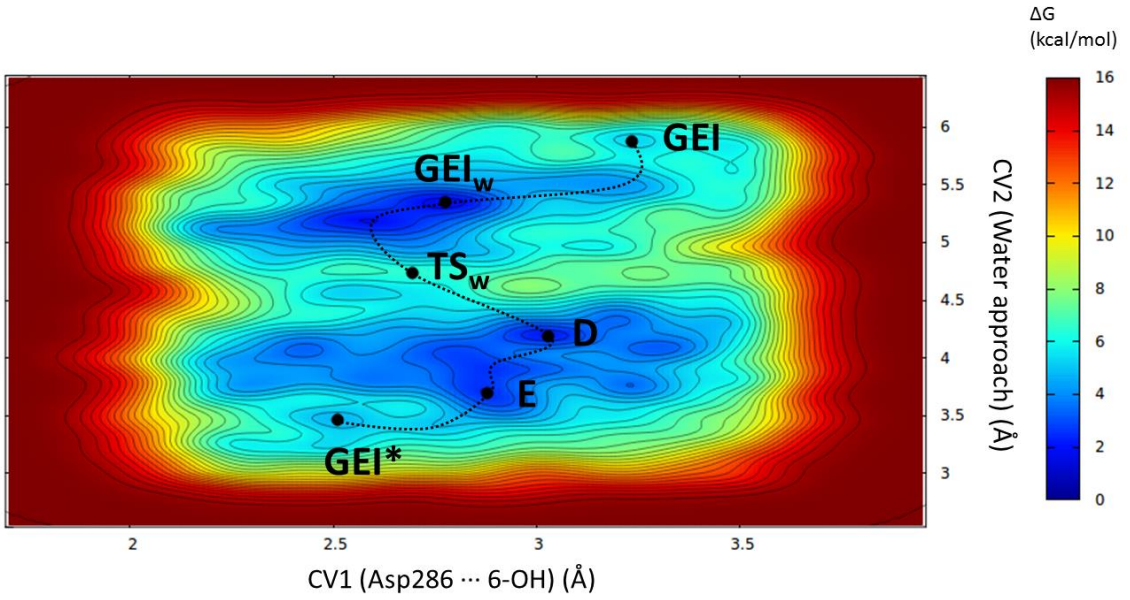


Figure 3.14. Free energy landscape of the water approach simulation. Each isoline is separated by 0.5 kcal/mol. Dotted line represents the MFEP from GEI to GEI* specie.

The state labelled as GEI* is the minimum in which the catalytic molecule is closest to the anomeric carbon (≈ 3.4 Å). Here, the Asp286···6-OH interaction is formed, thus this state is similar to the GEI' state previously found, the main difference is the presence of the water molecule at a catalytic distances. Remarkably, at this state the -1 sugar is activated for catalysis, as it exhibits a conformation ($E_3/^4C_1$) with a pseudo-axial orientation of the leaving group. Moreover, the water molecule is well oriented for catalysis, as it has the oxygen atom lone pairs pointing towards the anomeric carbon. This state is the one we chose to start the simulations of the deglycosylation and transglycosylation reactions.

It is easy to see that here is a zig-zag pathway connecting the GEI and the GEI* states. In this pathway, the Asp286 – 6-OH interaction is forming while the catalytic water approaches the anomeric carbon. The GEI* state is isoenergetic with the GEI (i.e. the

product of the glycosylation reaction, Figure 3.11), being 5 kcal/mol above GEI_w. The whole process (from GEI to GEI*) has an energy barrier of 4.7 kcal/mol. This is the energetic cost of bringing a water molecule to a catalytic orientation (see Figure 3.16). D and E species are intermediate stable species where the Asp286···6-OH interaction is broken and the -1 sugar is not distorted and the water is near the C_1 (3.5 – 4.1 Å). Due to the Asp286···6-OH interaction is broken, it was not appropriate to start the deglycosylation reaction from this region of the FEL (the number of collective variables needed will be higher than 3).

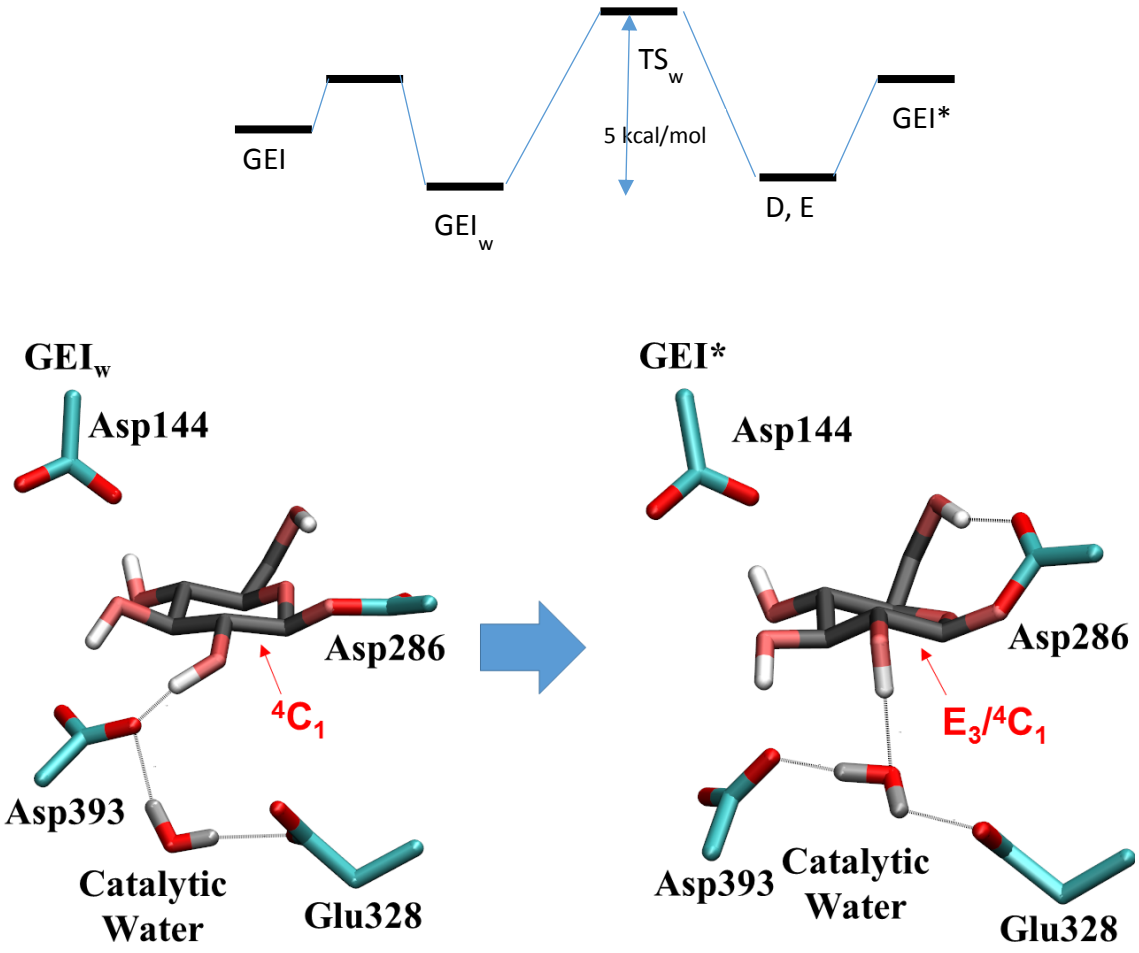


Figure 3.15. Representative structures along the reaction coordinate of the water approach simulation.

The process of approaching the catalytic water to the active center can be summarized as follows. Initially, the water molecule is far from the catalytic region, there is no Asp286

– 6-OH interaction and the -1 sugar is not distorted. Finally, the water molecule rotates and approaches to the catalytic zone, the Asp286 – 6-OH interaction forms and the conformation of the -1 sugar changes to $E_3/^4C_1$, being prepared for the deglycosylation reaction.

Returning to the two intermediate species problem, the results of an statistical analysis performed over the FEL (CV1 ± 0.1, CV2 ± 0.1 Å box) and the Mercator puckering coordinates space ($\varphi \pm 0.1$, $\theta \pm 0.1$ rad. box) are shown in Figure 3.15. On the one hand, in Figures 3.15a and 3.15b, there are represented the expected values of the θ puckering coordinate (-1 sugar) and the CV2 used in the glycosylation simulation (Glyco. CV2) over the FEL of the present simulation. On the other hand, classifying the structures along the metadynamics MD by the (ϕ, θ) puckering coordinates set (-1 sugar), the expected value of the present CV1 (Asp286··6-OH interaction) and the Glyco. CV2 (Asp286···C_1 interaction) variables were calculated to understand why we are able to differentiate both GEI forms using CV1 instead of using the Glyco. CV2.

There are four important observations to be mentioned:

1) In Figure 3.15a, meanwhile the range of CV1 is (2.00 – 3.60) Å, the Glyco. CV2 evolves in a more stretched range (1.70 – 2.30) Å and there is no correlation between both.
2) In Figure 3.15b, a good correlation between CV1 and the distortion of the sugar is observed. When the CV1 (Asp286···6-OH interaction) is broken (> 2.75 Å), the sugar is not distorted. When the CV1 interaction is formed (< 2.75 Å), the sugar is slightly distorted ($\theta > 30º$). The role of the water is important, then, the sugar is in E_3 conformation ($\theta \sim 60º$), when the water arrives to the active zone (~3.2 Å).
3) In Figures 3.15c/d, the correlation of CV1 and the conformation is fully observed, meanwhile, it is not the case for Glyco. CV2.
4) The selected CVs make the conformational pathway follows the trajectory of the simulation (Deglycosylation initial state in an $E_3/^4C_1$ conformation).

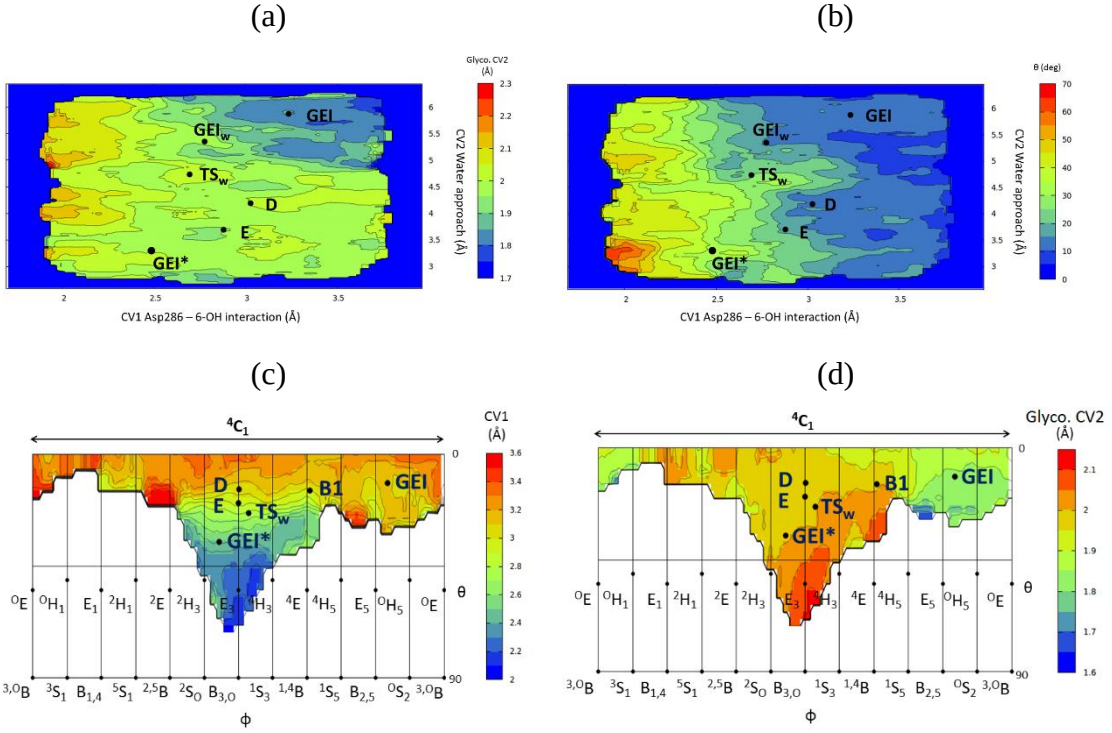


Figure 3.16. Representation of the expected value surfaces of: (a) the Glyco. CV2 (in Å, isoline every 0.05 Å) and (b) the θ puckering coordinate (in radians, isolines every 5 degrees) over the water approach simulation CVs space. Representation of the expected value of: (c) the CV1 (in Å, isoline every 0.1 Å) and (d) the Glyco. CV2 (in Å, isoline every 0.05 Å) over the Mercator puckering coordinate projection (ϕ, θ).

4.5. Simulation of the second reaction step (enzyme deglycosylation)

As mentioned before, the initial structure for the enzyme deglycosylation step is the GEI* structure of the previous simulation, where the water molecule is at 3.3 Å from the anomeric carbon and its orientation is optimum for catalysis. The position of the water molecule is stabilized by three hydrogen bond interactions (Figure 3.16 [GEI*]): the 2-OH of the glucose with the water oxygen, one of the hydrogens of the water with the Asp393 and the other hydrogen (acid) with the Glu328 catalytic residue.

During the metadynamics simulation of the enzyme deglycosylation reaction, the collective variable values and the puckering coordinates of the -1 sugar were monitored (Appendix B) to analyse the structural changes along the reaction.

The resulting deglycosylation FEL and its 2D projections are shown in Figure 3.17. The surface shows two important minima in opposite regions: the activated glycosyl-enzyme intermediate (GEI*) and hydrolysis products (P_h), which are 7.6 kcal/mol more stable. The transition state (TS_h) from GEI* to P_h, is 13.3 kcal/mol higher in energy.

To analyse the evolution of the conformation of the -1 sugar along the reaction pathway, we performed a statistical analysis, choosing the structures in a box of (CV1 ± 0.2, CV2 ± 0.2, CV3 ± 0.2 Å) along the MFEP. Two states before the transition state (F and G) and one state after TS_h (H) were chosen to better characterize all changes occurring along the reaction coordinate. Figure 3.18 displays the evolution of the main distances and the puckering coordinate (θ) of the glucose ring along the MFEP.

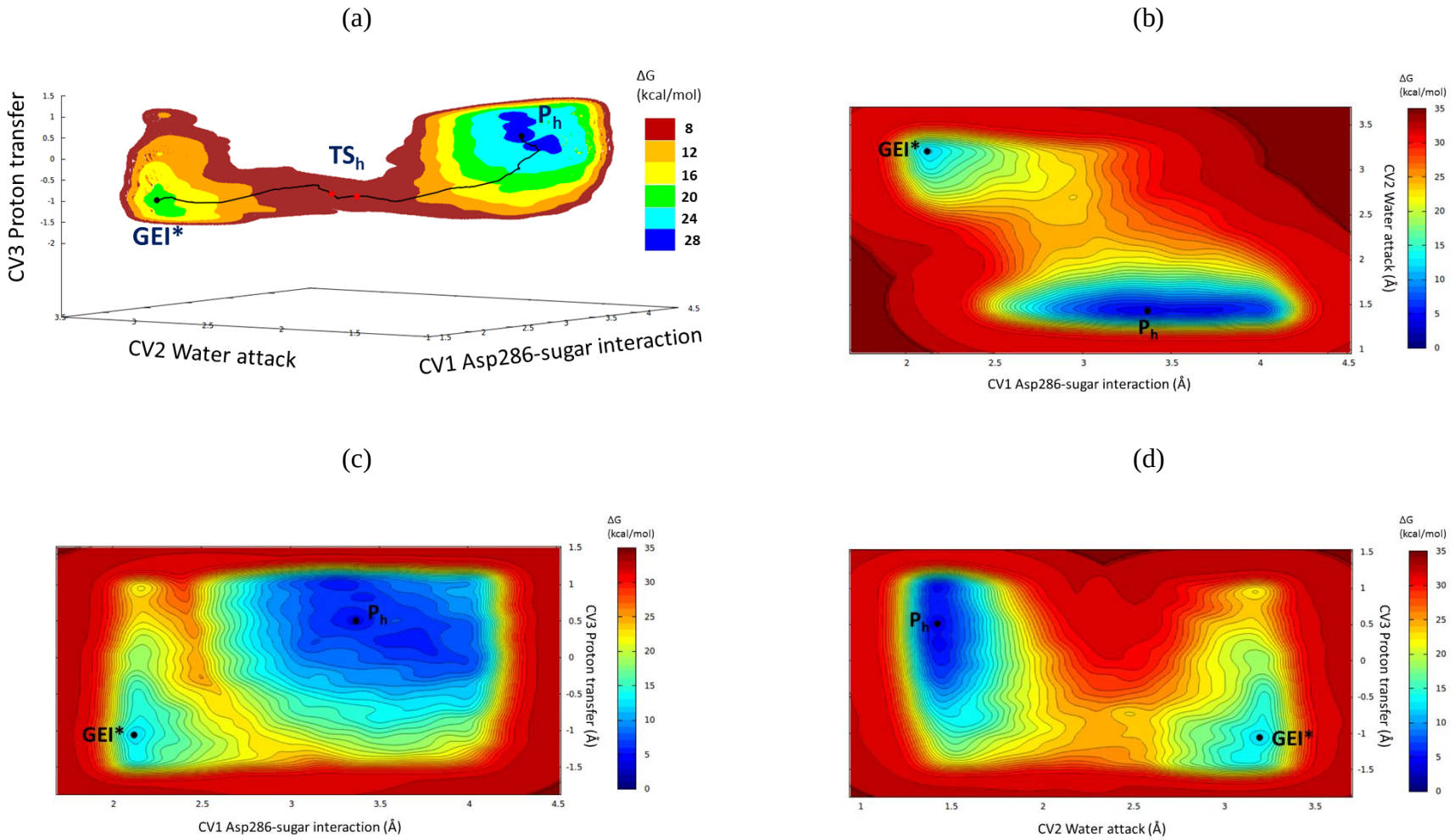


Figure 3.17. (a) Three dimensional (3D) FEL of the deglycosylation reaction of *Np*AS. The bold black line indicates the MFEP. (b, c, d) Two dimensional (2D) projections on each pair of collective variables. Isolines at 1 kcal/mol.

Figure 3.18 shows the evolution of the most relevant catalytic distances along the MFEP and Figure 3.19 shows the change of puckering coordinates in a Mercator representation. The structure of the active site along the MFEP is shown in Figure 3.20 and Table 3.4 lists the most important distances and puckering coordinates. The mechanism of the deglycosylation reaction can be separated in four phases:

(I) *Cleavage of the covalent glycosyl-enzyme bond*

In this initial phase, the CV1 value evolves from 2.2 to 2.8 Å, it means that the covalent bond between the glucose and the Asp286 starts its cleavage.

(II) *Water attack to the anomeric carbon*

In this phase, the CV2 (distance between the oxygen of the catalytic water and the anomeric carbon) evolves from 3.0 to 2.0 Å. The system overcomes the transition state and the conformation of the -1 sugar changes to an E_3 envelope.

(III) *Glucosyl-water bond formation and total cleavage of the enzyme-sugar bond*

In this intermediate phase, the CV1 becomes greater than 3.0 Å and the CV2 value collapses at 1.5 Å, it means the formation of a protonated α-glucose molecule not covalently bonded to the enzyme.

(IV) *Proton transfer from the protonated glucose to the Glu328*

In this last phase, the CV3 evolves from negative to positive values, then, the base Glu328 attracts the acid proton of the α-glucose. This glucose adopts a $^4C_1$ conformation.

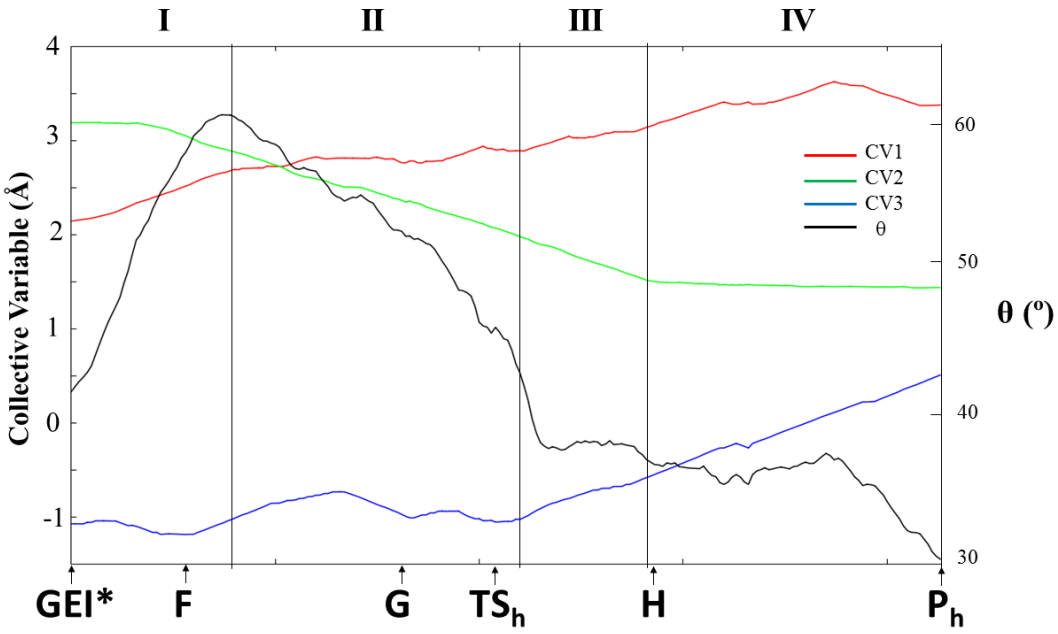


Figure 3.18. Evolution of the collective variables and the θ puckering coordinate of the -1 sugar during the deglycosylation reaction of *Np*AS (MFEP).

In Figure 3.19, the evolution of the conformation of the -1 glucosyl over the puckering surface during the deglycosylation simulation is reported. The pathway shows a conformational itinerary going throw the $E_3/^4C_1 \rightarrow E_3 \rightarrow {}^4C_1$ region. As observed in the water approach simulation, the initial conformation of the sugar is a quasi-distorted $E_3/^4C_1$ conformation (GEI*). The pathway evolves vertically along the region where φ is 180º to the E_3 subspace. Finally, the system goes to the products where the formed glucose is undistorted.

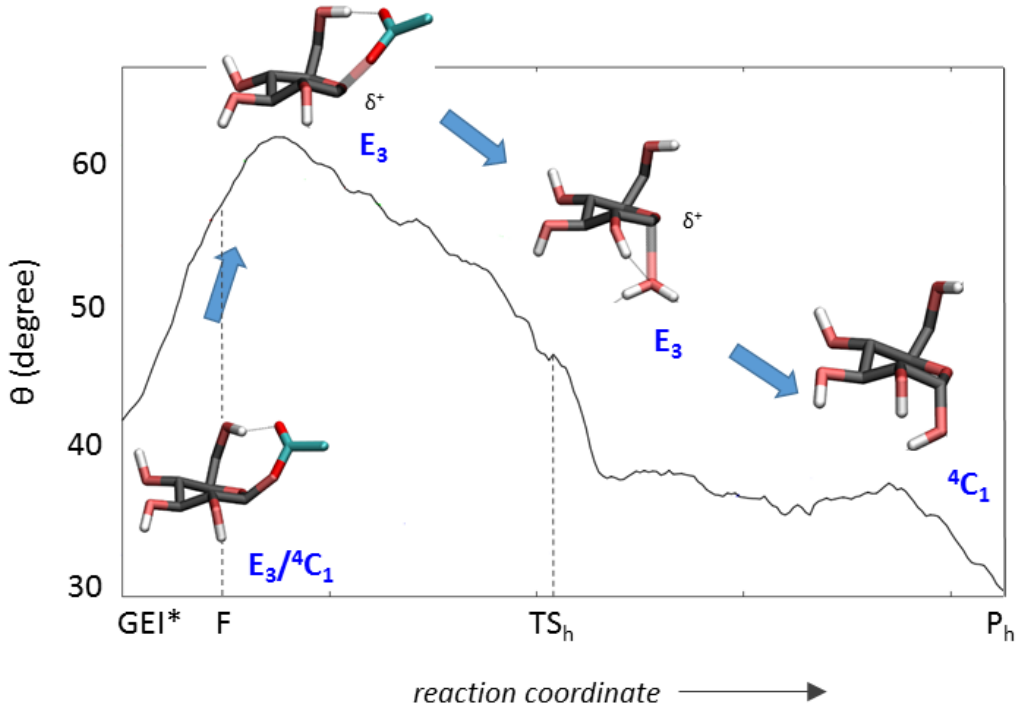


Figure 3.19. Conformational pathway followed by the sugar during the enzyme deglycosylation step.

In summary, following the information available in the Figure 3.20, the hydrolysis reaction ends with the formation of the α-glucose inside the amylosucrase and the catalytic acid/base residue (Glu328) is newly protonated, as expected. The conformation of the sugar during the first steps of the deglycosylation process is an E_3 envelope, and then, at the end of the reaction, the sugar returns to the initial $^4C_1$ conformation. In the first phase (from GEI* to F), the Asp286 is the main character, trying to go far from the sugar, feeling the presence of the water molecule. The second and third phases shown the water attack and the enzyme-substrate bond cleavage, and finally, in the last phase, the Glu328 acts as base deprotonating the new formed α-glucose.

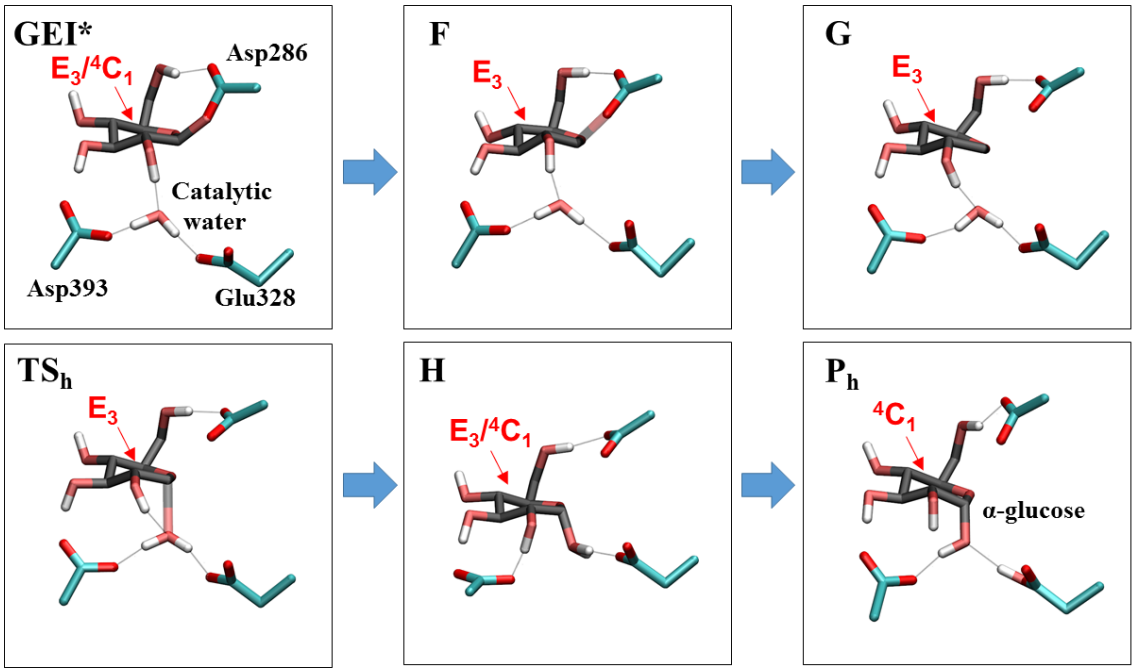


Figure 3.20. Representative structures along the reaction coordinate of the enzyme deglycosylation reaction.

Table 3.4. Calculated expected values of the most relevant catalytic distances (in Å) and puckering coordinates (in degrees) and its standard deviations along the deglycosylation MFEP (ΔG in kcal/mol).

	ΔG	**CV1**	**CV2**	**CV3**	**φ**	**θ**	**Conformation**
GEI*	0.0	2.14 ± 0.09	3.19 ± 0.11	-1.08 ± 0.11	184 ± 9	42 ± 11	$E_3/^4C_1$
F	3.7	2.33 ± 0.11	3.18 ± 0.10	-1.10 ± 0.08	181 ± 8	55 ± 6	E_3
G	11.0	2.81 ± 0.12	2.50 ± 0.12	-0.74 ± 0.11	183 ± 11	52 ± 9	E_3
TS_h	13.3	2.81 ± 0.12	2.23 ± 0.12	-0.94 ± 0.12	187 ± 10	48 ± 7	E_3
H	3.2	3.11 ± 0.12	1.55 ± 0.12	-0.62 ± 0.11	191 ± 14	38 ± 12	$E_3/^4C_1$
P_h	-7.6	3.38 ± 0.12	1.44 ± 0.10	0.51 ± 0.12	189 ± 24	30 ± 15	$^4C_1$

4.6. Simulation of the transglycosylation step

The initial structure for the transglycosylation step is obtained from a structure stabilization of a substituted model of the deglycosylation step, where the fructose ring is changed by a glucose ring, whose 4-OH substituent is stabilized around 3.5 Å respect the anomeric carbon. We assumed that the initial state of transglycosylation is similar to the GEI_w of the water approach simulation, where the water is around 5 Å to the anomeric carbon and the fructose is at 3.3 – 3.5 Å.

During the transglycosylation simulation, the collective variable values and the puckering coordinates of the -1 sugar were monitored (Appendix B) to analyse the structural changes observed along the MFEP.

The resulting transglycosylation FEL and its 2D projections are shown in the Figure 3.21. The surface shows three important minima in opposite regions: reactants (GEI_t' and GEI_t) and products (P_t), being the GEI_t the most stable (1.2 kcal/mol with respect to GEI_t' and 4.6 kcal/mol with respect to P_t). The transition state (TS_t) is 22.1 kcal/mol less stable than the GEI_t minimum.

To analyse the evolution of the conformation of the sugar along the reaction pathway, we performed a statistical analysis on the metadynamics trajectory, identifying structures in a box of (CV1 ± 0.2, CV2 ± 0.2, CV3 ± 0.2 Å) along the MFEP.

Apart from the reactants (GEI'_t), transition state (TS) and products (P_t), two more states (one before the TS_t, named J, and one after the TS_t, named K) were chosen to better characterize the changes occurring along the reaction coordinate. It is to be noted that this reaction is the inverse process of the glycosylation step if one replaces fructose by glucose. The energy barrier of the inverse reaction (from P_h to GEI_t), it amounts to 17.4 kcal/mol, which is very similar to the 17.2 kcal/mol previously computed for the glycosylation step.

Returning to the statistical analysis, we represent the evolution of the CVs (Figure 3.22) and the puckering coordinate (θ) of the glucose ring along the internal reaction coordinate (Figure 3.23).

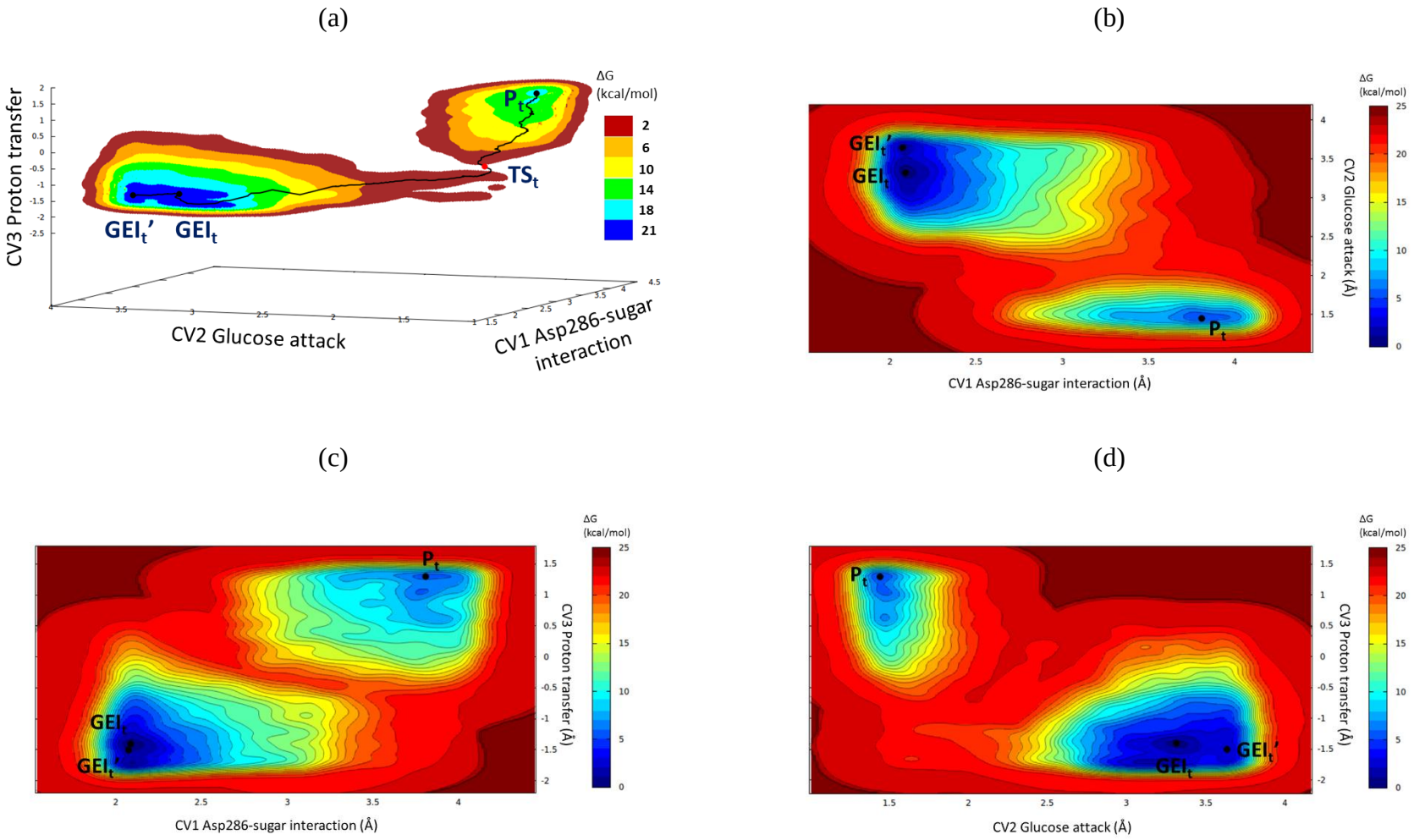


Figure 3.21. (a) Three dimensional (3D) FEL of the transglycosylation reaction of *Np*AS. The bold black line indicates the MFEP. (b, c, d) Two dimensional (2D) projections on each pair of collective variables. Isolines at 1 kcal/mol.

Figure 3.22 shows the evolution of the collective variables along the reaction pathway. The mechanism of the transglycosylation reaction can be divided in three phases:

(I) *Glucose approach to the anomeric carbon*

The first phase is similar to the mechanism observed during the water approach simulation. The acceptor α-glucose is at 3.6 Å and there are structures at the initial state of the FEL that are undistorted ($^4C_1$), whereas others are slightly distorted ($^4H_3/^4C_1$). At the end of this phase, the CV2 evolves to 3.0 Å and the -1 sugar adopts an intermediate $^4H_3/^4C_1$ conformation.

(II) *Glycosidic bond formation and glycoside-enzyme bond cleavage*

In this phase, CV1 and CV2 evolve complementarily: the glycosyl-enzyme bond increases from 2.1 to 3.4 Å and the C_1 – 4-OH distance (CV2) decreases from 3.0 to 1.7 Å (glycosidic bond formed). The conformation of the -1 sugar evolves to an E_3 conformation.

(III) *Proton transfer from the protonated maltose to the Glu328*

Finally, CV3 evolves from negative to positive values, as observed in the deglycosylation reaction, resulting in the protonation of the acid/base residue (Glu328). The -1 sugar does not distort.

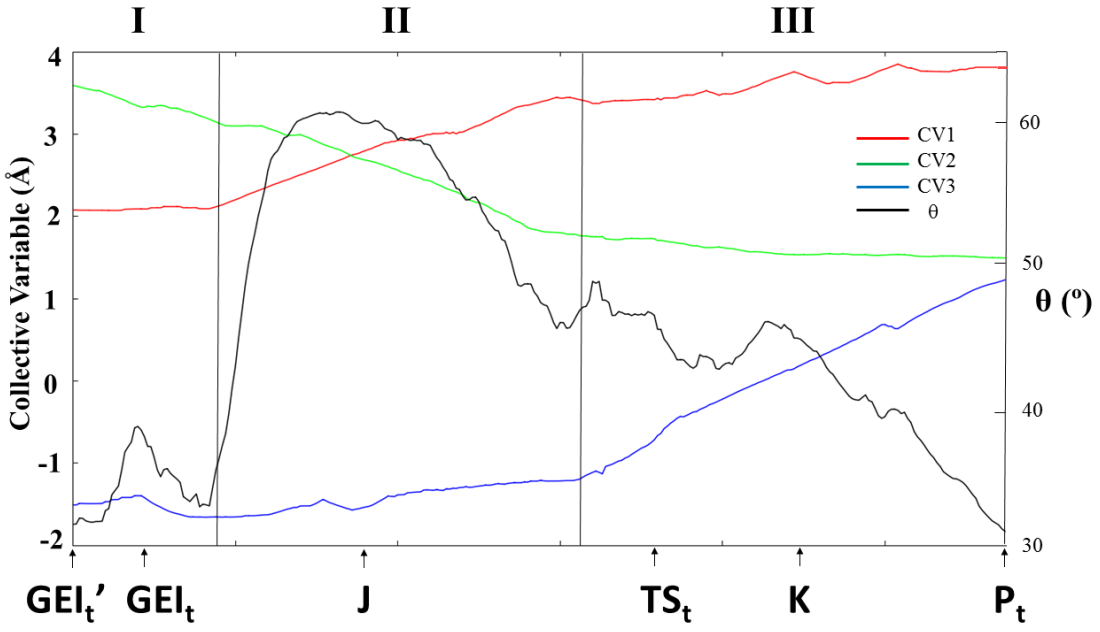


Figure 3.22. Evolution of the CVs and the θ puckering coordinate (-1 sugar) along the transglycosylation reaction coordinate.

Figure 3.23 shows the evolution of the conformation of the -1 glucose over the puckering surface (Mercator representation) during the transglycosylation simulation. The reaction follows a cyclic conformational itinerary: $^4C_1 \rightarrow E_3 \rightarrow {}^4C_1$. As observed in the water approach simulation, when the acceptor is at more than 3.4 Å to the anomeric carbon, the -1 ring is either not distorted ($^4C_1$) or in a mixture of $^4C_1$ and E_3 conformations. As observed in Figure 3.23, the GEI_t' can be described as a mixture of undistorted and slightly distorted structures. However, when the acceptor glucose attacks the -1 sugar, the conformation is a well-defined intermediate $^4H_3/^4C_1$ conformation, being similar to the one observed during the deglycosylation reaction. Again, the pathway evolves vertically over the distorted E_3 conformational space (J and TS_t species show $E_3/B_{3,O}$ and E_3 conformations, respectively). Finally, the system evolves to the products well where the maltose disaccharide is formed by two undistorted glucose rings.

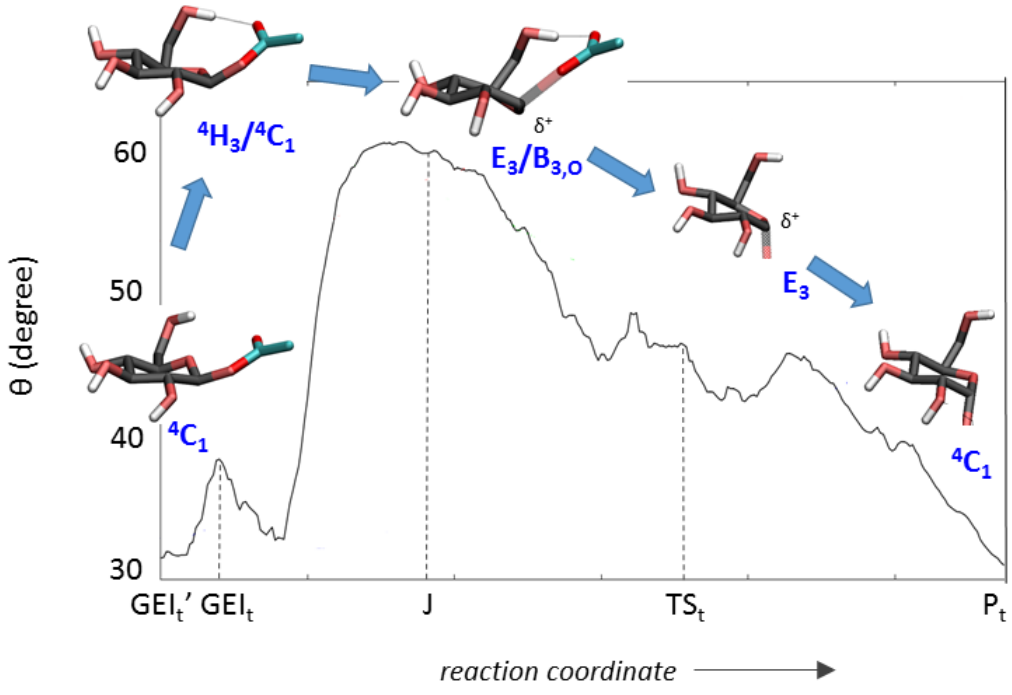


Figure 3.23. Conformational pathway followed by the -1 sugar during the transglycosylation step.

In summary, analysing the Figure 3.24, the transglycosylation reaction ends with the formation of a maltose molecule inside the amylosucrase and the catalytic acid/base residue (Glu328) is protonated newly, as expected. The conformation of the sugar during the first steps of the transglycosylation process is a mixture between $^4C_1$ and $^4H_3$, and then, while the reaction is visiting the transition state, the sugar become a pure E_3 to finalize the reaction as a chair.

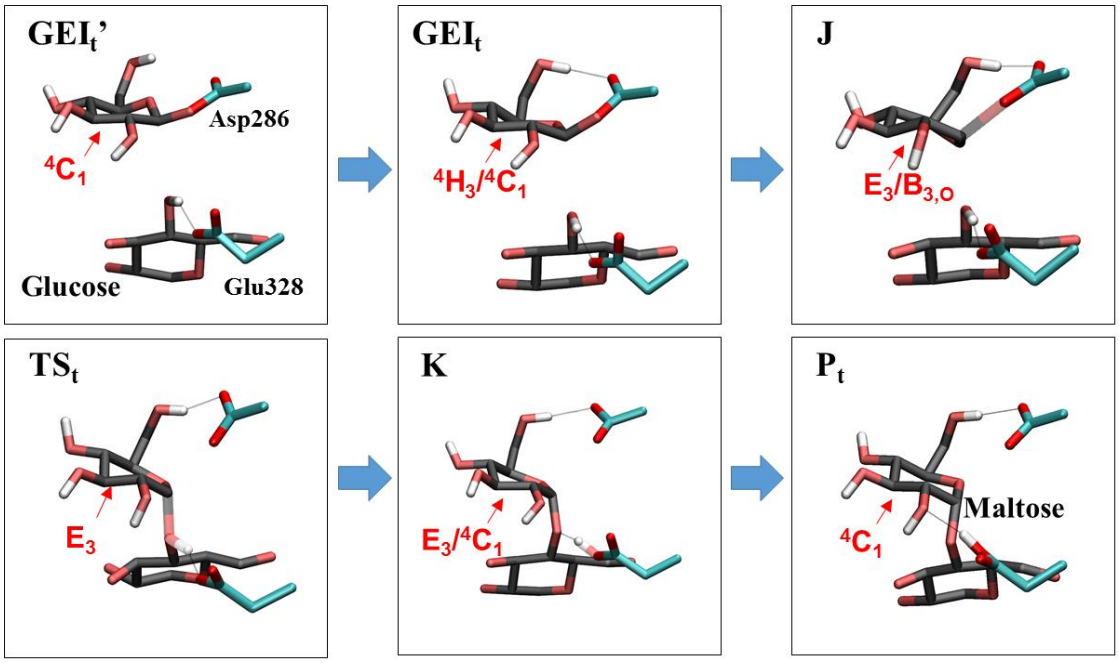


Figure 3.24. Representative structures along the reaction coordinate of the transglycosylation reaction.

Table 3.5. Calculated expected values of the most relevant catalytic distances (in Å) and puckering coordinates (in degrees) and its standard deviations along the transglycosylation MFEP (ΔG in kcal/mol).

	ΔG	**CV1**	**CV2**	**CV3**	**φ**	**θ**	**Conformation**
GEI$_t$'	1.2	2.08 ± 0.08	3.59 ± 0.14	-1.51 ± 0.18	228 ± 47	32 ± 21	$^4C_1$
GEI$_t$	0.0	2.09 ± 0.09	3.33 ± 0.17	-1.40 ± 0.17	206 ± 35	38 ± 20	$^4H_3/^4C_1$
J	11.9	2.74 ± 0.09	2.75 ± 0.17	-1.56 ± 0.16	179 ± 8	60 ± 5	$E_3/B_{3,O}$
TS$_t$	22.1	3.41 ± 0.08	1.73 ± 0.12	-0.89 ± 0.16	186 ± 11	46 ± 9	E_3
K	12.3	3.75 ± 0.09	1.54 ± 0.12	0.13 ± 0.17	176 ± 9	45 ± 12	$E_3/^4C_1$
P$_t$	4.6	3.81 ± 0.09	1.50 ± 0.12	1.23 ± 0.14	168 ± 20	31 ± 15	$^4C_1$

5. Discussion

Starting from the conformational study of the glucose inside the enzyme, the results are in good agreement with the experimental data (Mirza, 2001) (in both cases, the sugar ring is in a chair conformation) and the length of the glycosidic bond and the ring $C_1 - O_p$ bond show a better preactivation for the chair than for the distorted shape. Furthermore, the steric hindrance of the hydrogen in the anomeric carbon when the sugar is distorted, increases the distance between the nucleophile and the anomeric carbon.

The glycosylation simulation shows a cyclic conformational itinerary ($^4C_1 \rightarrow {}^4H_3/E_3 \rightarrow {}^4C_1$), resulting in a covalent enzyme-substrate intermediate whose conformation is not activated for the deglycosylation step. The final structure obtained from the simulation is in good agreement with the crystallographic structure of the covalent intermediate (Jensen, 2004).

During the water approach simulation, we have demonstrated that the sugar slightly changes shape ($E_3/{}^4C_1$) when the catalytic water is at ~3.2 Å from the anomeric carbon, ready to deglycosylate the substrate.

The deglycosylation simulation shows a conformational itinerary around the Mercator meridian defined by $\phi = 180^o$ ($E_3/{}^4C_1 \rightarrow E_3 \rightarrow {}^4C_1$). As expected from a retaining mechanism, α-glucose is the final product, adopting the chair conformation.

The -1 sugar in the transglycosylation simulation starts with a similar conformational itinerary as in the water approach simulation: the -1 glucose distorts with the approach of the 4-OH nucleophile oxygen at distances 3 – 3.2 Å. The complete conformational itinerary is similar to the one observed for deglycosylation ($E_3/{}^4C_1 \rightarrow E_3 \rightarrow {}^4C_1$). Furthermore, the energy barrier observed is in agreement with the inverse reaction of the glycosylation step, but here the system evolves from the covalent intermediate to the disaccharide product.

In summary (Figure 3.25), the hydrolysis reaction in *Np*AS shows that the glycosylation step is the rate limiting step (17.3 kcal/mol), the approaching of the catalytic water is a quasi-diffusive step (4.7 kcal/mol) and the deglycosylation step (12.3 kcal/mol) leads to the products state, which is 8.6 kcal/mol more stable than the initial reactants. These results are in very good agreement with the kinetic experimental data (k_{cat} = 33 min^{-1}; ΔG = 17.9 kcal/mol).

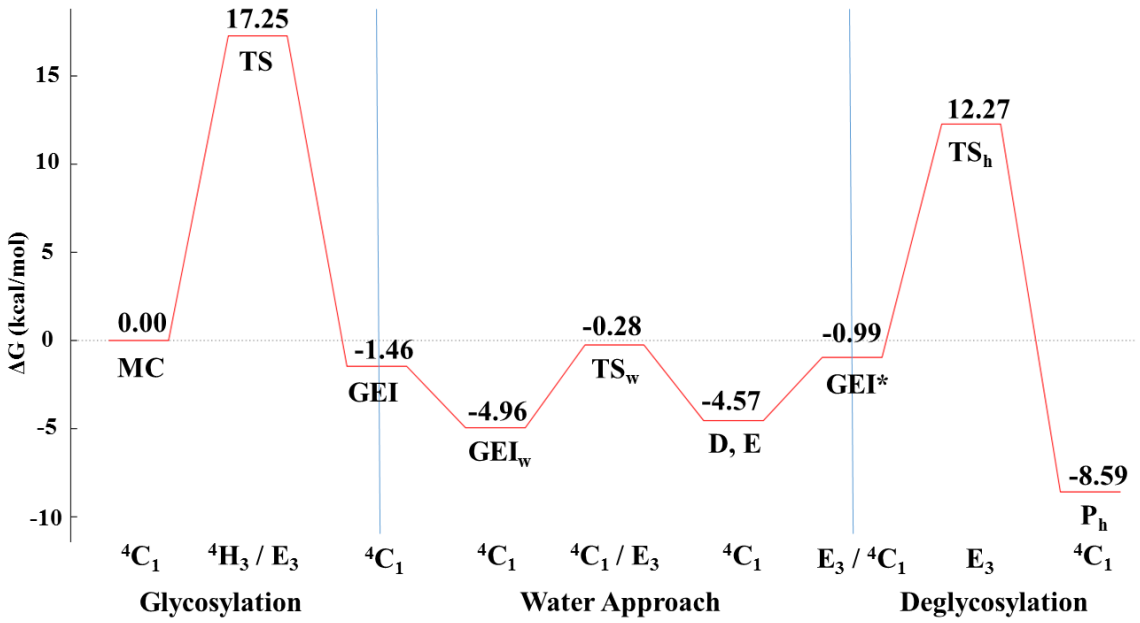


Figure 3.25. Schematic representation of the complete hydrolysis pathway of sucrose by *Np*AS.

In the case of the transglycosylation step of the reaction (Enz-G + G → Enz + G_2), the process is symmetrical to the glycosylation step, then, the transition state of the glycosylation and the polymerization are energetically similar. So, the Enz···GF and Enz···G_2 Michaelis Complexes are energetically undistinguishable.

Experimentally, in conditions of low concentration of glucose and without the presence of large glucose oligosaccharides, the activity rate ratio of the AS for hydrolysis/transglycosylation is 2. In our simulations, we have found the reason why it is so. Taking into account that the Enz-G + G → Enz + G_2 is an inverse reaction of the Enz-GF → Enz-G + F, the energy profile is quite symmetrical. The hydrolysis reaction is favoured because deglycosylation step goes throughout a lower energy barrier than the polymerization step. Then, in the simulation conditions, the hydrolysis is kinetically and thermodynamically favoured with respect to the single glucose transglycosylation.

Future goals of this project will aim at simulating the reaction using maltooligotetrasaccharide acceptors, for which the transglycosylation activity is known to be higher due to the presence of strong anchoring interactions at subsite +4 involving salt bridge interactions between amylosucrase and the oligosaccharide (Albenne, 2002). Also, introduction of mutations in amylosucrase active site have shown significant effects

on the catalytic properties of the enzyme as well as on the products formed (Cambon 2014, Vergès 2017). We will thus also attempt understanding at electronic level the effect of these mutations on the enzyme reaction.

6. Conclusions

The main conclusions of the present chapter are the following:

1. The conformational study of the α-glucosyl group in the -1 subsite of the AS confirms the experimental chair conformation as the most stable shape of the -1 sugar in the Michaelis complex.

2. The simulation of the hydrolysis reaction (glycosylation and deglycosylation steps) of sucrose by amylosucrase confirms that the glycosylation is the rate limiting step of the reaction. The resulting activation energy (17.3 kcal/mol) is in good agreement with the experiments (17.9 kcal/mol).
3. The conformational itinerary of the glycosylation step is the $^{4}C_{1} \rightarrow {}^{4}H_{3}/E_{3} \rightarrow {}^{4}C_{1}$ cyclic pathway, being the covalent intermediate not distorted, a result in good agreement with the crystallographic structure reported in the reference (Jensen, 2004).
4. The -1 sugar becomes slightly distorted ($E_{3}/{}^{4}C_{1}$) when the catalytic water/glucose approaches the anomeric carbon (3.0-3.3 Å), being well preactivated for the deglycosylation/transglycosylation processes.
5. In the simulation conditions (absence of large glucose oligosaccharides and low glucose concentrations), the deglycosylation step (12.3 kcal/mol) is more energetically favourable than the transglycosylation (17.1 kcal/mol). This result is in good agreement with the experimental data. Both conformational itineraries evolve throughout an E_{3} transition state.

Chapter IV – Conformational and catalytic study of GH125 α-mannosidases

This work has appeared in:

S. Alonso-Gil, A. Males, P. Fernandes, S. J. Williams, G. J. Davies, C. Rovira. *J. Am. Chem. Soc.* **2017**, 139 (3), 1085 – 1088.

Conformational and catalytic study of GH125 α-mannosidases

1. Introduction

In Section I – 4.4, we introduced a family of enzymes whose function is N-glycan processing. In the present chapter, we focus on one of them, GH125 exo-1,6-α-mannosidase, an inverting exo-acting metal-independent α-mannosidase (Table 1.2). This enzyme is expected to follow the classical mechanism of inverting GHs (Figure 1.1, bottom).

Added to the previous studies shown in the present Thesis, the importance of the conformational behaviour of the sugars along the reaction coordinate in the glycoside hydrolases is crucial to understand the action of some microbial pathogens and symbionts. Exo-1,6-α-mannosidases from family GH125 is found in a variety of fungi and human gut symbionts (Gregg, 2011). These enzymes take part of a large group of retaining and inverting α-mannosidases that exhibit N-glycan processing activity, including families 38, 47, 76, 92 and 99 (Table 1.2).

In the retaining enzymes group, exo-GH38 (Petersen, 2010), being Zn^{2+} dependent, and endo-GH76 (Thompson, 2015) follow a ${}^{O}S_2 \rightarrow B_{2,5} \rightarrow {}^{1}S_5$ conformational pathway. Endo-GH99 enzymes constitute a special case as they do not follow the classical inverting GH mechanism (Figure 1.1, bottom) but involve an epoxide intermediate (Thompson, 2012b) and a ${}^{4}C_1 \rightarrow {}^{4}E \rightarrow {}^{4}H_5$ itinerary (Petrisevic, 2017).

In the inverting enzymes group, exo-GH47, being Ca^{2+}-dependent, follows a ${}^{3}S_1 \rightarrow {}^{3}H_4 \rightarrow {}^{1}C_4$ (Thompson, 2012a). The mechanism for Ca^{2+}-dependent GH92 and exo-GH125 enzymes is not known. In both cases, the Michaelis complex with a mannobiose thio-derivative shows a non-distorted (${}^{4}C_1$, Figure 4.1) sugar ring in the -1 subsite (Zhu, 2010) (Gregg, 2011). However, this is not consistent with the expected behaviour of inverting α-mannosidases. Furthermore, a complex of GH92 with a mannoimidazole inhibitor shows a ring conformation in a ${}^{1}S_5/{}^{4}H_5$ distorted shape). The fact that the thioglycoside proposes a ${}^{4}C_1 \rightarrow {}^{4}H_3/E_3/{}^{4}E \rightarrow {}^{1}S_3$ itinerary, but the mannoimidazole proposes a ${}^{1}S_5/{}^{4}H_5$ transition state makes wonder about the good mimicking properties of the thio-compound in α-mannosidases.

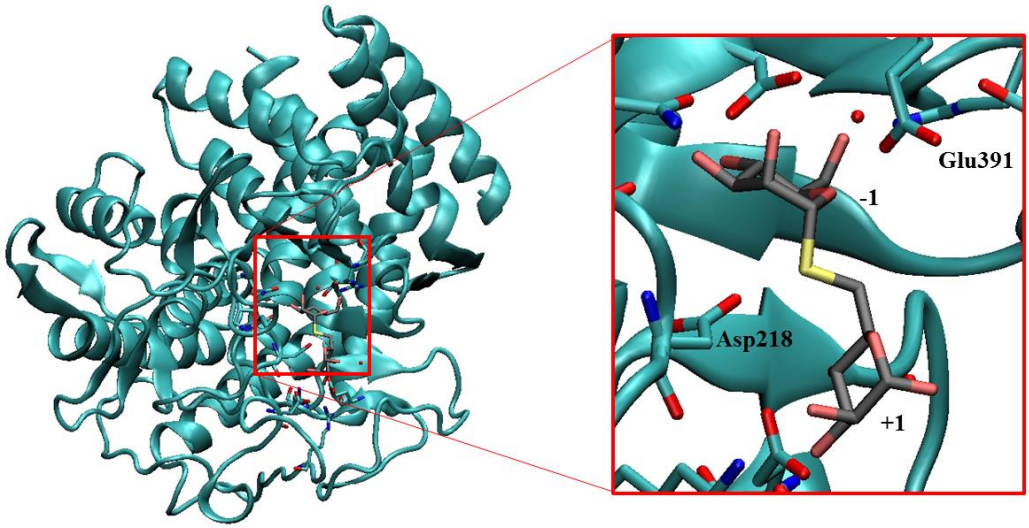


Figure 4.1. A view of the secondary structure and a zoom of its active site pocket of the crystallographic structure of the *Cp*GH125 – thio-mannobiose complex (PDB 3QT9) (Gregg, 2011). The conformation of the α–mannosyl (-1 subsite) is $^4C_1$.

The principal aim of this work is to elucidate the most stable conformation of the natural substrate in GH125 α-mannosidases. In case the the natural substrate adopts a distorted conformation, this would lead to the natural question: why is the thio-compound not distorted? Are the thio-compounds always good mimics of the natural substrates of α-mannosidases? Another crucial point to be studied is the effect of the sulphur atom in the conformational behaviour of the α-mannose in gas phase. Then, the conformational FEL of the 1-thio substituted α-mannose was studied to compare with the natural mannose conformational FEL, reported by our group (Thompson, 2012a).

To answer the above questions, we will first analyse the effect of introducing a sulphur atom in the conformational properties of α-mannose in gas phase. The conformational FEL of the 1-thio substituted α-mannose will be compare with those of natural mannose conformational FEL, previously reported by our group (Thompson, 2012a).

Finally, the conformational FELs of the 1-thio-α-mannosyl residue of the complex of GH125 with 1,6-α-thiomannobiose will be analysed, in comparison with the FEL of the natural substrate. The most stable conformation of the latter will be used to simulate the hydrolysis reaction using QM/MM metadynamics. The free energy barrier obtained in the simulation will be compared with the available kinetic results on GH125. In particular, the synthesis of 4-methylumberlliferyl α-D-mannopyranosyl-(1,6)-β-D-mannopyranoside

(4-MU-α-manno-1,6-β-mannoside) using GH125 exo-α-1,6-mannosidase controlled by a fluorescent assay were reported in (Deng, 2013). The k_{cat} of the hydrolysis of the 4-MU-α-manno-1,6-β-mannoside catalysed by the *Clostridium perfringens* GH125 enzyme is 294 min^{-1} ($\Delta G^{\ddagger}$ = 16.6 kcal/mol).

This work motivated crystallographic experiments carried out in the group of Professor Gideon J. Davies (University of York). An acid/base mutant (Asp218Asn) of *Cp*GH125 was crystallized in complex with the natural substrate 1,6-α-mannobiose. Our results and their crystal structure turned to be in good agreement and both works have been published recently (S. Alonso-Gil, 2017).

2. Objectives

The main objectives of the present chapter are:

1) To compare the conformational behaviour of α-D-mannopyranose molecule and the 1-thio-α-D-mannopyranose in gas phase.
2) To reconstruct the conformational FELs of the thio-compound and the natural mannobiose inside the *Cp*GH125 exo-1,6-α-mannosidase.
3) To elucidate the catalytic itinerary of the -1 sugar ring during the hydrolysis reaction.

3. Computational details

Isolated molecule metadynamics

The free energy surface of the α-mannose molecule, obtained by *ab initio* metadynamics, was taken from a previous study by our group (Thompson, 2012a). The same methodology was used to reconstruct the FEL of the 1-thio-α-mannopyranoside (1-thio-α-mannose) molecule. The geometry of the 1-thio-α-mannose molecule (QM box: 12.0 x 12.3 x 13.2 $Å^3$) was optimized until the maximum component of the forces over the atoms was lower than $5 \cdot 10^{-5}$ a.u. using the CPMD software. The electronic mass and the time-step chosen were 850 a.m.u. and 5.0 a.u., respectively. The parameters of this metadynamics simulation are presented in the Appendix F.

Model building

The modelling of the enzyme-substrate systems was done from the X-ray diffraction data reported in the reference (Gregg, 2011, PDB 3QT9 Res. 2.05 Å) of the complex GH125 α-mannosidase – thio-mannobiose crystal. Two systems were modelled: GH125 α-mannosidase – thio-mannobiose complex (GH125-MB-S) and GH125 α-mannosidase – natural mannobiose complex (GH125-MB-O). In both cases, the crystallographic waters outside the active center and the 1,2-ethanediol molecules were removed.

Structure preparation was done with the AMBER 9 software. The parameters used for the sugars were extracted of the Glycam 06 (Kirshner, 2008) library. The final systems were modelled with the tleap software. The protonation of histidines is provided in the Appendix B. Our analysis showed that His348 needs to be protonated in the delta nitrogen

to describe properly the active site interactions. All Asp and Glu residues were taken as deprotonated (i.e. negative charge) except the Asp218 (the acid-base residue). To compensate the final total charge of -18.0 a.u., eighteen sodium cations were added. Finally, 15039 TIP3P (Jorgensen, 1983) water molecules were added generating a parallelepiped box of 81.7 x 91.6 x 84.5 $Å^3$ dimensions. The final structures have 424 residues of the protein, one substrate residue (thio-mannobiose MB-S / mannobiose MB-O), eighteen sodium cations and 15039 water molecules. The force field FF99 (Cornell, 1995) was used to define the potential energy of the system.

Classical MD simulations

Geometry optimization, a progressive heating to 200 K during 400 ps in steps of 50 K, a progressive heating to 300 K during 400 ps in steps of 25 K, density equilibration using a barostat centred in 1 atm (NPT collectivity) and molecular dynamics production using a thermostat at 300 K (NVT collectivity) of the systems were performed with the AMBER 11 software. The time-step chosen during the preparation was 1 fs and it was increased during the production to 2 fs. The aim of this first treatment is to guarantee that the force field allows a good description of the real system with our model equilibrated at 300K. At the beginning of the geometry optimization, a favourable restraint was applied over the distance between the proton of the Asp218 (acid/base residue) and the glycosidic sulphur/oxygen of the substrate to promote a 1.6-2.0 Å interaction between them. This interaction is maintained without restraints for the rest of the simulation. The RMSD of both structures (shown in the Appendix B) respect the crystallographic structure stabilizes quickly around 1.1 Å after 10 ns of production dynamics, indicating the rigidity of the enzyme and the reliability of the experimental structure. During both simulations, the -1 sugar remained in the initial $^4C_1$ conformation.

QM/MM MD simulations

Once the structures of GH125-MB-S and GH125-MB-O were equilibrated, two snap-shots (Figure 4.2), respectively, were taken to start the QM/MM simulations. For the conformational study, each QM box was chosen as composed by the whole sugar, whereas a larger QM box, including the catalytic residues, was chosen to model the hydrolysis reaction. In all simulations, the electronic density was defined using the PBE functional (Perdew, 1996), the cut-off of the plane waves was 70.0 Ry and norm-

conserving Martins-Troullier pseudopotentials (Troullier, 1991) were used to describe the core electrons.

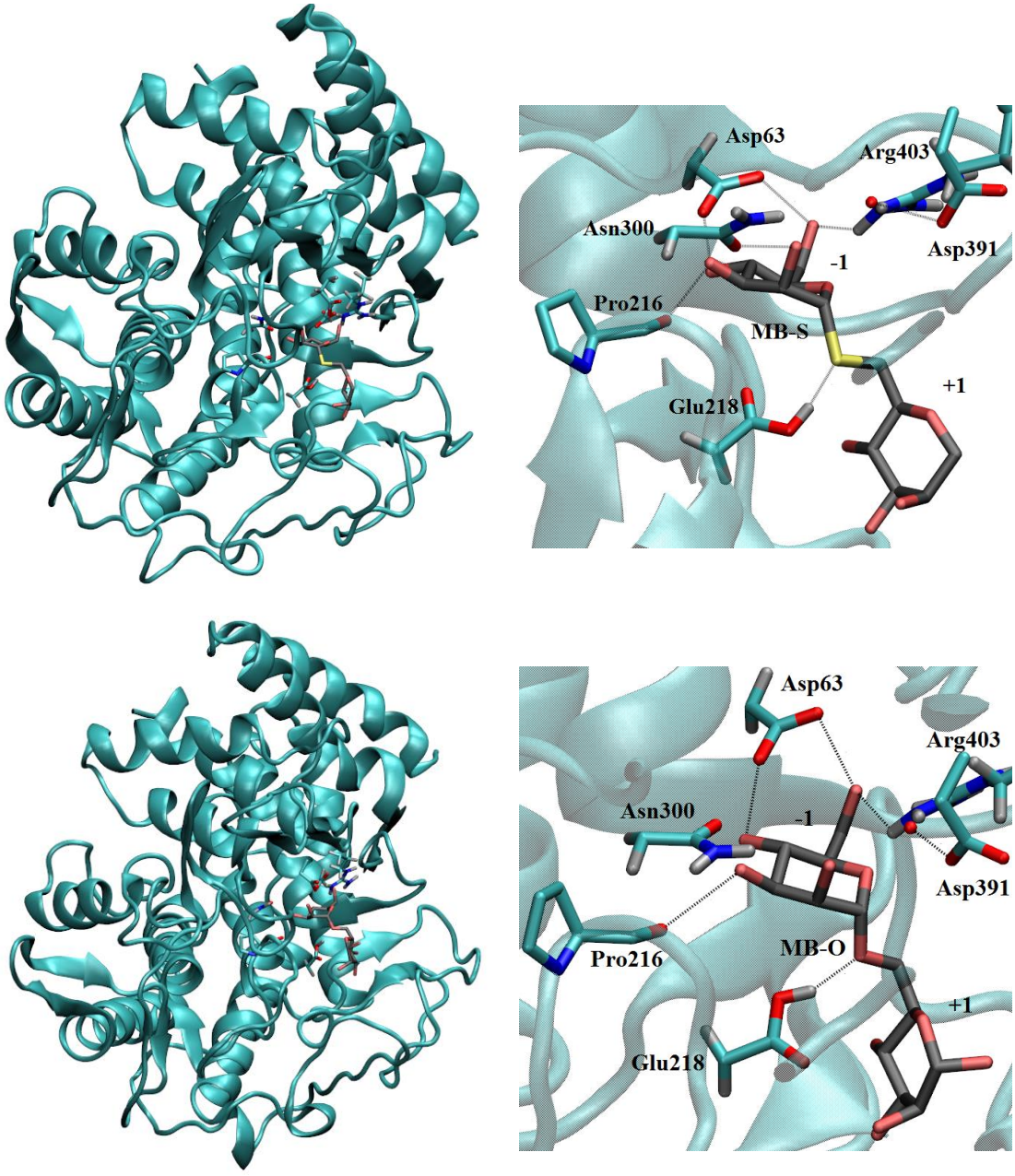


Figure 4.2. (left) Secondary structure and (right) active center of the GH125-MB-S (top) and the GH125-MB-O (bottom) complexes obtained after the classical molecular dynamics simulation. The QM box of the QM/MM models are formed by the sugar molecules (black carbon atoms). The RMSD of the backbone between both structures is 0.93 Å.

QM/MM metadynamics simulations

The conformational free energy landscape (FEL) of the -1 sugar (both α-mannosyl 1-thio-α-mannosyl) ring in the active site of the GH125 α-mannosidase was computed by QM/MM metadynamics. The Cremer-Pople puckering coordinates (Cremer, 1975) phi and theta (ϕ, θ) of the -1 sugar ring were used as collective variables, following the methodology was previously used in the group and in chapter III. The parameters used in the well-tempered MTD are listed in Appendix F. The Lagrangian electronic mass and the time step were taken as 600.0 a.m.u. and 5.0 a.u., respectively. Before QM/MM MD stabilization, an annealing optimization was done until the maximum force component over the QM atoms was lower than 10^{-4} a.u.

The modeling of the hydrolysis reaction was initiated from a representative structure of the most stable conformation of MB-O in the enzyme active site. The two catalytic residues and the nucleophile water molecule were also included in the QM region. A preliminary QM/MM MD simulation (data not shown) revealed the formation of a water proton channel formed by 2 water molecules. Thus, these water molecules were added to the QM region (Figure 4.3) shows the 69 atoms forming the QM region. QM/MM frontier carbons in the catalytic residues are the C_α of the Asp218 and the C_β of the Glu391.

Four collective variables were chosen to simulate the reaction pathway of the hydrolysis reaction: CV1 accounts for the proton transfer of the acid/base residue (Asp218) to the glycosidic oxygen. CV2 accounts for the cleavage of the glycosidic bond of the sugar (CV2). CV3 represents the nucleophilic attack of the catalytic water over the anomeric carbon. Finally, CV4 accounts for the water proton channel activation where the Glu391 is protonated.

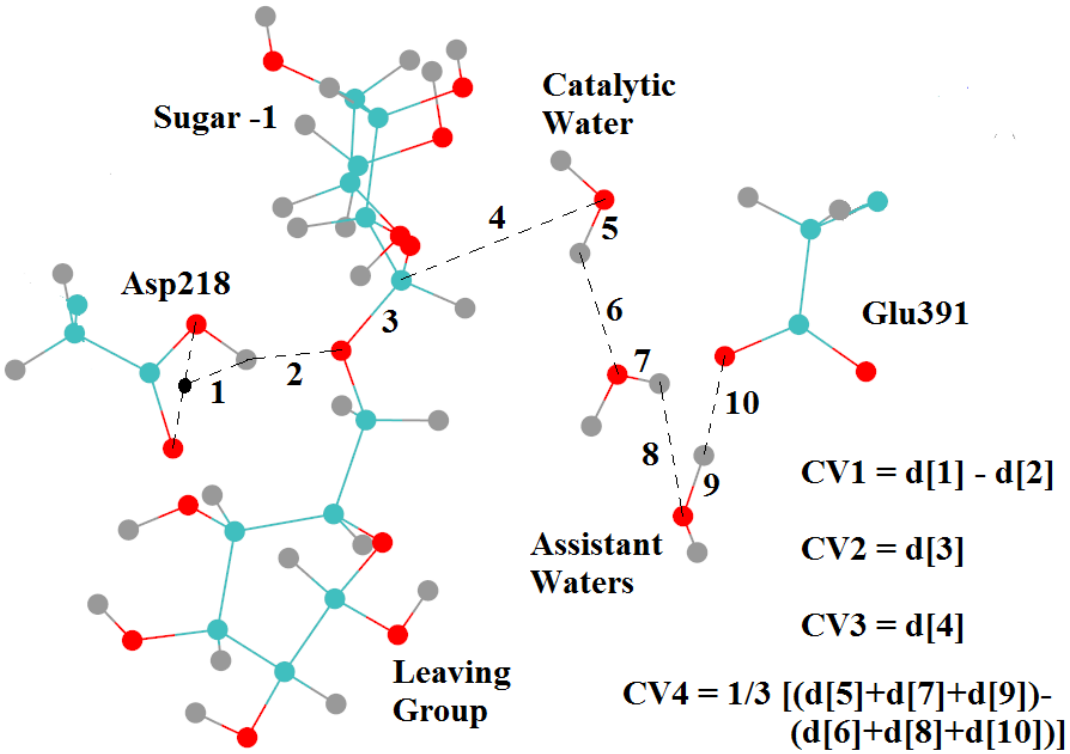


Figure 4.3. QM region (formed by 69 atoms) used in the simulation of the hydrolysis reaction in GH125-MB-O. The conformation of the α–mannosyl (-1 subsite) is $^{O}S_{2}$. QM/MM frontier carbons in the catalytic residues are the C_{α} of the Asp218 and the C_{β} of the Glu391. Atoms legend: hydrogen (grey), carbon (blue), oxygen (red).

4. Results

4.1. Conformational study of the 1S-thio-α-D-mannopyranose in gas phase. Thioglycoside vs. natural substrate comparison

The conformational FEL of α-D-mannopyranose (Thompson, 2012a) is shown in Figure 4.4a. The reconstruction of the conformational FEL of 1-thio-α-D-mannopyranose (thioglycoside) has given as a result the surface shown in the Figure 4.4b.

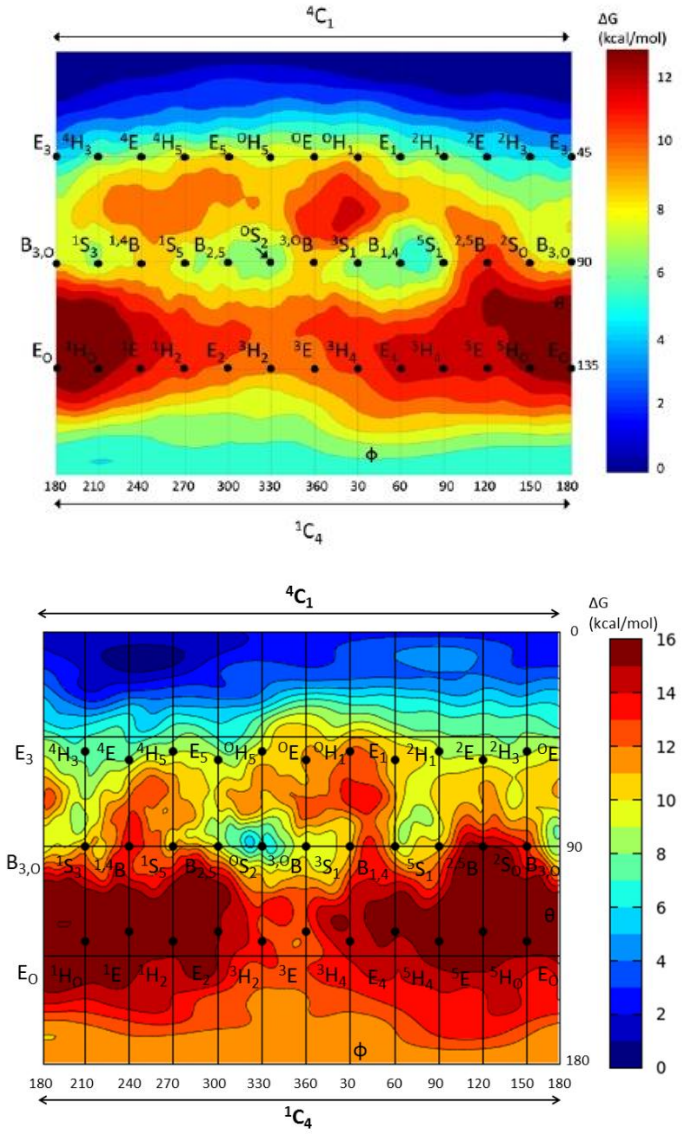


Figure 4.4. (a) Conformational FEL of isolated α-D-mannopyranose (Adapted from the reference [Thompson, 2012a]); contoured at 1 kcal/mol. (b) Conformational FEL of isolated 1S-thio-α-D-mannopyranose; contoured at 1 kcal/mol.

The principal minima allocated over the thio-compound conformational space are in the $^{4}C_{1}$ (most stable conformation), $^{O}S_{2}$ (5.2 kcal/mol), $^{2}S_{O}/B_{3,O}$ (7.3 kcal/mol), $^{1}S_{3}$ (8.1 kcal/mol), $B_{1,4}/^{5}S_{1}$ (8.2 kcal/mol), $B_{2,5}$ (8.2 kcal/mol), $^{3,O}B$ (8.6 kcal/mol), $B_{3,0}$ (8.7 kcal/mol), $^{3}S_{1}/^{3,O}B$ (8.8 kcal/mol) and $^{1}C_{4}$ (10.5 kcal/mol) regions.

Comparing both FELs, the most relevant change is observed in the relative stability of the $^{4}C_{1}$ versus the $^{1}C_{4}$ conformation (10.5 [thio-compound] and 4.0 kcal/mol [natural mannose]). The presence of a sulphur atom in the position 1 of the sugar improves the stability of the $^{4}C_{1}$ and the $^{O}S_{2}$ conformation to the detriment of the $^{5}S_{1}$, $^{1}S_{3}$, $B_{3,O}$ and $B_{2,5}$ conformations. The Boltzmann contribution of the $^{1}C_{4}$ shape decrease dramatically.

4.2. Conformational FEL of the thio-compound α-mannosyl ring at the -1 subsite

The reconstructed free energy landscape of the GH125-MB-S is shown in the Figure 4.5. Each point of the surface corresponds to a different conformation of the α-mannosyl ring in the -1 subsite of the enzyme (-1 sugar). Two important minima are observed: the most stable region is where the sugar adopts a $^{4}C_{1}$ conformation, being 7.27 kcal/mol more stable than the $B_{3,O}$ conformation region. This is indicative of a $^{4}C_{1} \rightarrow {}^{4}H_{3}/E_{3} \rightarrow {}^{1}S_{3}/B_{3,O}$ catalytic itinerary.

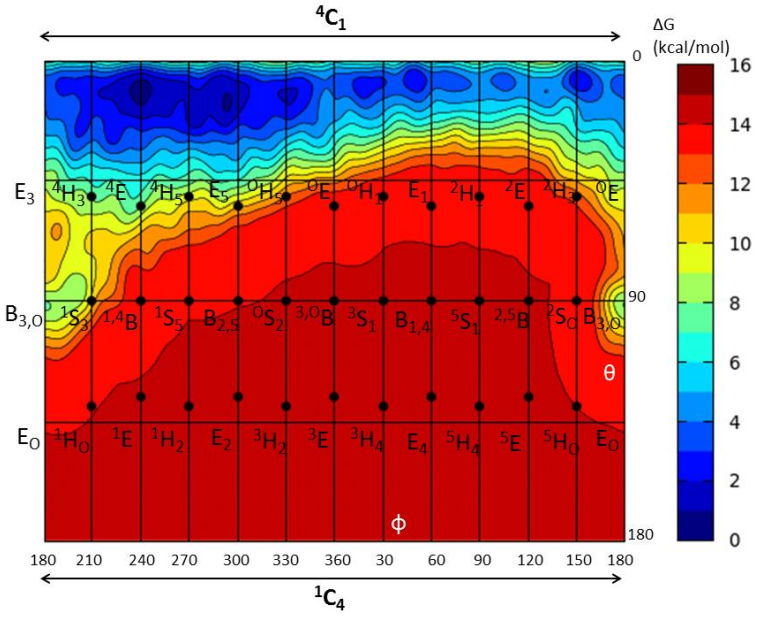


Figure 4.5. Conformational free energy map (Mercator's projection) of the 1-thio-α-mannosyl in the exo-α-1,6-mannosidase. Isolines at 1 kcal/mol.

4.3. Conformational FEL of the natural substrate α-mannosyl ring at the -1 subsite

The reconstructed free energy landscape of GH125-MB-O is shown in Figure 4.6. Five relevant minima can be observed. The most stable minimum is over the $^{O}S_2$ region. We also observe a region around the $^{4}C_1$ conformation, which is 3.70 kcal/mol less stable. As previously observed in the thio-compound, a $B_{3,O}$ minimum is found, with a relative energy of 8.61 kcal/mol. With a similar stability, a $^{1}S_5/B_{2,5}$ conformation is located at the ϕ = 284º, θ = 100º. Finally, a small minimum in the $^{1}C_4$ region (at 11.5 kcal/mol) is observed.

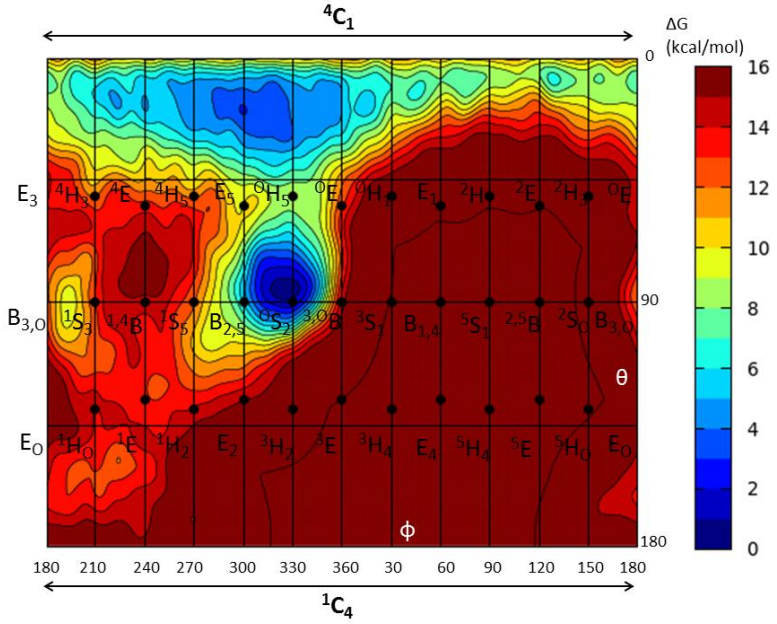


Figure 4.6. Conformational free energy map (Mercator's projection) of the natural α-mannosyl sugar in the -1 subsite of the exo-α-1,6-mannosidase. Isolines at 1 kcal/mol.

The results obtained suggest two possible reaction pathways for GH125 enzymes: the most favoured energetically is the $^{O}S_2 \rightarrow B_{2,5} \rightarrow {}^{1}S_5$ pathway and the less likely one is $^{4}C_1 \rightarrow {}^{4}H_3 \rightarrow {}^{1}S_3/B_{3,O}$.

4.4. Conformational analysis

A statistical analysis of the free energy surfaces was performed. The structures around the minima whose puckering coordinates fall inside a square $\phi_{min} \pm 0.1$, $\theta_{min} \pm 0.1$ box (in radians) were chosen in order to calculate the average value of relevant distances for catalysis and the enzyme-sugar ring interactions.

As seen in Table 4.1, on the one hand, the expected values of the thio-glycosidic bond, d(C_1-S), of the $^4C_1$ and $B_{3,O}$ minima are 1.89 ± 0.07 Å and 1.84 ± 0.04 Å, respectively. On the other hand, the glycosidic bond, d(C_1-O_g), takes values of 1.48 ± 0.05 and 1.49 ± 0.05 Å for $^OS_2$ and $^4C_1$, respectively, decreasing to values of 1.43 ± 0.04 and 1.44 ± 0.04 Å for the $B_{3,O}$ and $^1C_4$ conformations. In terms of catalytic efficiency, the conformation with a larger glycosidic bond shows a shorter value of the distance between the anomeric carbon and the pyranic oxygen, d(C_1-O_p). In the case of the thio-derivative is easy to see that the most populated conformation ($^4C_1$) is the most activated one for catalysis. However, in the case of the natural substrate, two conformations are similarly activated ($^OS_2$ and $^4C_1$), but, which is the reason of the stability difference between them?

Table 4.1. Main distances around the anomeric carbon (C_1) of the saccharide unit at the -1 subsite corresponding to the local minima of the conformational FELs (distances (d) are in Å and energies in kcal/mol).

System	Conformation	ΔG	d(C_1-O_g / S)	d(C_1-O_p)
MB-S	$^4C_1$	0.00	1.89 ± 0.08	1.41 ± 0.05
MB-S	$^1S_3/B_{3,O}$	7.27	1.84 ± 0.05	1.44 ± 0.04
MB-O	$^OS_2$	0.00	1.48 ± 0.05	1.41 ± 0.04
MB-O	$^4C_1$	3.70	1.49 ± 0.05	1.41 ± 0.05
MB-O	$^1S_5/B_{2,5}$	8.58	1.48 ± 0.05	1.42 ± 0.03
MB-O	$^1S_3/B_{3,O}$	8.61	1.43 ± 0.04	1.43 ± 0.04
MB-O	$^1C_4$	11.44	1.44 ± 0.04	1.44 ± 0.03

Figure 4.7 shows the most important interactions between the enzyme and the -1 sugar. During the metadynamics simulation, we have seen a relevant change in the 2-OH and 3-OH hydroxyls interactions when the GH125-MB-O system goes from the $^4C_1/B_{3,O}$ region to the $^OS_2/^1S_5/^1C_4$ region. The expected values of these distances and their standard deviations are listed in Table 4.2.

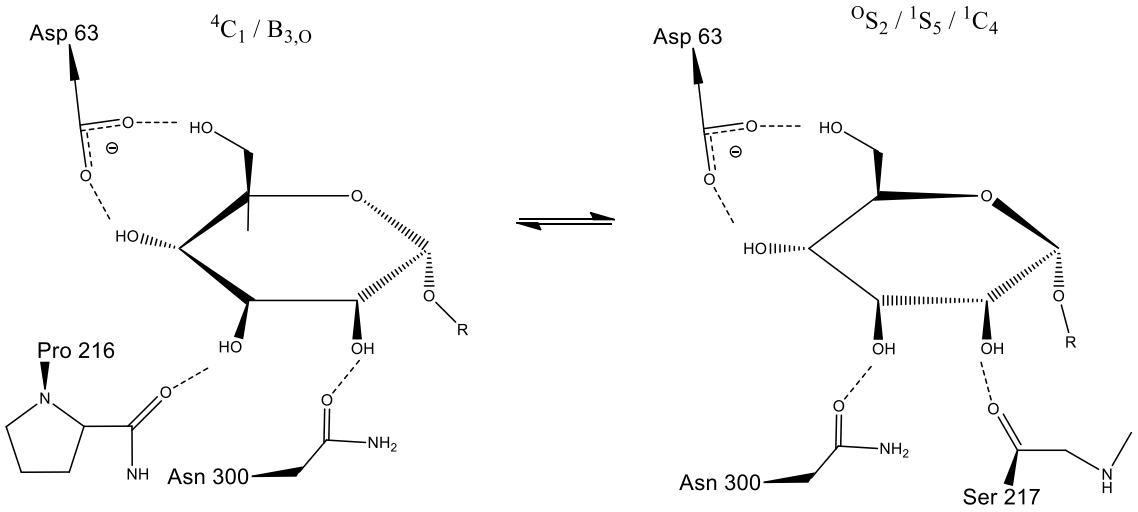


Figure 4.7. Schematic representation of the most important -1 sugar – enzyme interactions in the GH125 complexes, depending on the conformation adopted by the -1 sugar ring.

The most important change in the hydrogen bond pattern of the -1 sugar ring when its conformation goes from $^4C_1$ to the $^OS_2$ region is the reorientation of the 2-OH hydroxyl from the carbonyl group of the Asn300, d(H_2-O_{Asn300}), to the carbonyl group of the Ser217, d(H_2-O_{Ser217}). Furthermore, the 3-OH hydroxyl takes profit to the 2-OH movement to approach its position to the available Asn300, d(H_3-O_{Asn300}).

There are two points to be extracted from Table 4.2. On one side, the interaction of the 4-OH and 6-OH hydroxyls with the enzymatic media do not change over both conformational FEL (their values fluctuate between 1.4 and 1.6 Å). On the other side, the principal interactions showing higher standard deviations (2-OH and 3-OH hydroxyls) are the transitory interactions, so that, 2-OH and 3-OH substituents change its interaction as described in Figure 4.6.

Table 4.2. Main interactions around the saccharide unit at the -1 subsite with the active site residues of the *Cp*GH125 exo-α-1,6-mannosidase corresponding to the local minima of the FEL (Distances are in Å).

Conf.	System	d(H_2-O_{Asn300})	d(H_3-O_{Pro216})	d(H_4-O_{Asp63})	d(H_6-O_{Asp63})	d(H_2-O_{Ser217})	d(H_3-O_{Asn300})
$^{O}S_2$	MB-O	4.55 ± 0.46	4.13 ± 0.27	1.57 ± 0.10	1.48 ± 0.10	**2.19 ± 0.33**	1.83 ± 0.30
$^{4}C_1$	MB-O	**2.75 ± 0.26**	1.63 ± 0.18	1.52 ± 0.11	1.54 ± 0.10	**4.38 ± 0.26**	3.25 ± 0.60
	MB-S	**1.86 ± 0.20**	1.57 ± 0.09	1.47 ± 0.10	1.52 ± 0.13	-	-
$^{1}S_3/B_{3,O}$	MB-O	3.39 ± 0.34	1.63 ± 0.08	1.50 ± 0.11	1.52 ± 0.09	-	-
	MB-S	2.00 ± 0.40	1.55 ± 0.07	1.43 ± 0.06	1.55 ± 0.13	-	-
$^{1}S_5/B_{2,5}$	MB-O	4.61 ± 0.28	4.01 ± 0.22	1.56 ± 0.11	1.56 ± 0.15	2.06 ± 0.41	1.73 ± 0.24
$^{1}C_4$	MB-O	4.72 ± 0.23	4.10 ± 0.19	1.54 ± 0.10	1.45 ± 0.10	2.19 ± 0.38	1.72 ± 0.30

Comparing the chair conformation of MB-S and MB-O systems, the principal change is observed in the 2-OH substituent interaction, then, the thioglycoside is interacting with the Asn300 (1.86 Å), as long as, the natural substrate is not interacting with anything (2.75 Å), being the reason why both conformational surfaces are different, so that, the 2-OH – Asn300 interaction of the MB-S cannot be broken and the system does not explore the $^{O}S_2$ / $^{1}S_5$ / $^{1}C_4$ region.

Comparing the $^{O}S_2$ and the chair conformation of the MB-O system, the 2-OH interaction of the distorted sugar with the Ser217 (2.19 Å) is stronger than the 2-OH interaction of the chair sugar with the Asn300 (2.75 Å), being the reason of the relative stability between both conformations.

4.5. Hydrolysis reaction simulation

The initial structure for the QM/MM simulation was selected from the trajectory of the conformational metadynamics of the GH125-MB-O system. A snap-shot corresponding to the global minimum of the conformational FEL (Figure 4.6) was taken ($^{O}S_2$ conformation).

The four collective variables, the six distances that take part in the fourth variable (CV4) and the puckering coordinates of the -1 mannosyl group were monitored every 10 MD steps in order to see their evolution, and once calculated the MFEP, to study their expected values over that pathway. Their representations are shown in the Appendix B and the evolution of the conformation will be presented in the following pages.

Due to the complexity to represent a 4D free energy hyper-surface, the topology of the reduced (CV1, CV2, CV3) free energy landscape is shown in Figure 4.8, obtained from the integration of the full FEL over the CV4 variable.

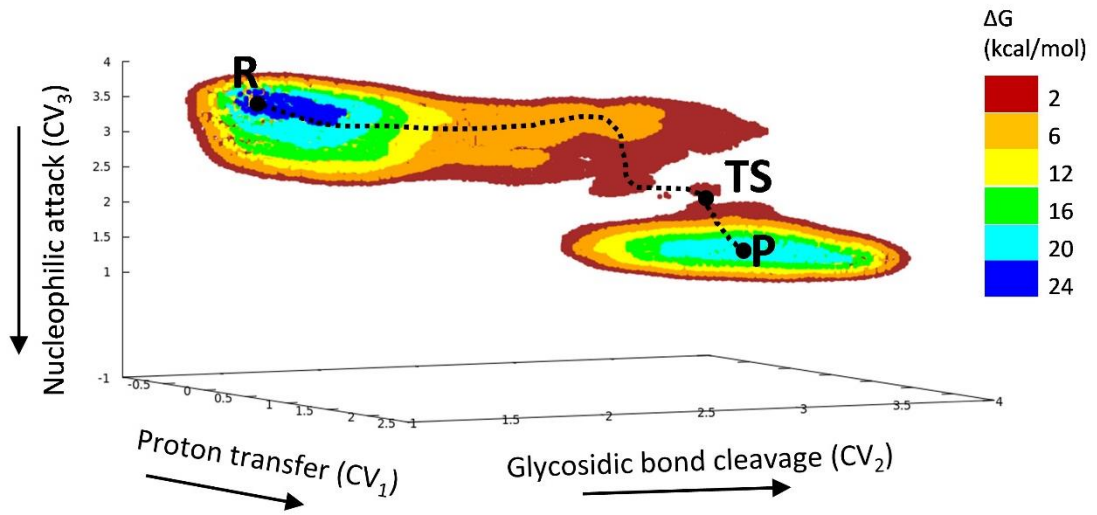


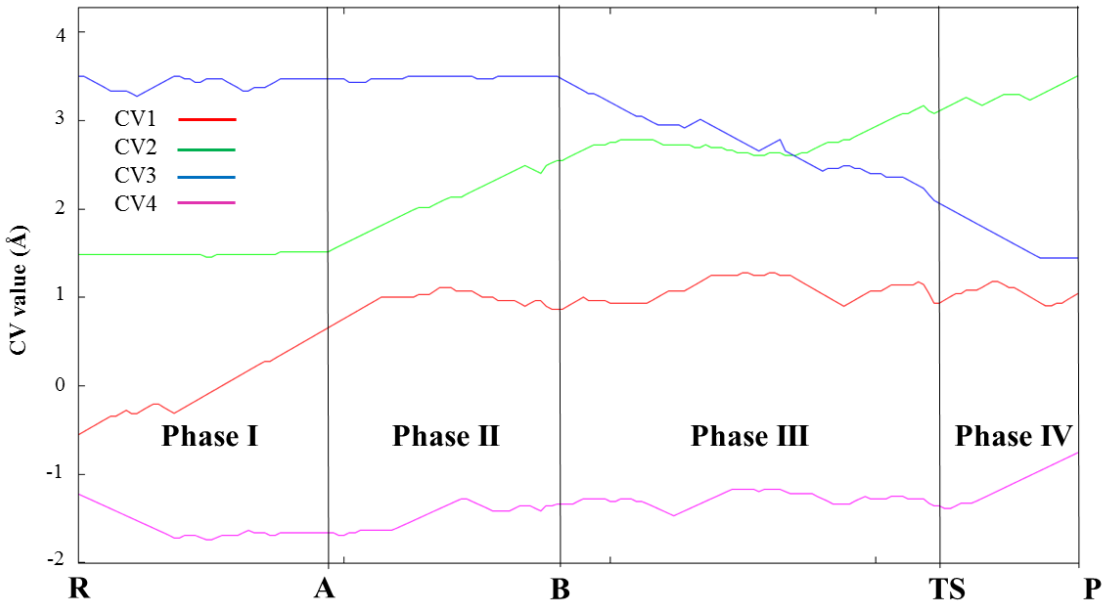
Figure 4.8. 3D projection of the 4D free energy surface reconstructed from the metadynamics simulation. The colours represent isosurfaces with the same free energy value. The TS was taken as the highest point along the minimum free energy pathway, obtained from intrinsic reaction coordinate analysis (Kui, 1981). Collective variables are given in angstrom.

We consider that a metadynamics simulation of a process is converged when the system evolves from the initial well (reactants) to the final state (products) and returns to reactants throughout the same transition state. Previous studies in the group on enzymatic reactions show that the height of the free energy barrier does not change significantly upon multiple recrossings. In this case, due to the complexity of CV4 (formed by several water-sugar, water-water and water-enzyme interactions) and the mobility of the system, the system was not able to return to reactants because it tends to explore non relevant regions of the products region. As a consequence, the energy difference between reactants and products is not well-described. However, the energy barrier and, more importantly, the conformational itinerary, is reliable.

Analysing the topology of the free energy hyper-surface, looking for minima and saddle points, the energy profile of the minimum free energy pathway (MFEP) was reconstructed (Figure 4.8, black dotted line). The energy profile shows three differentiable regions: the reactants valley, the transition state region and the products valley. The minimum free energy point in the reactants valley is the point R and the

maximum free energy point in the transition region is the point TS, whose energy is 22.3 kcal/mol greater than R. The minimum free energy point in the products valley is the point P (not converged), being 3.2 kcal/mol less stable than R. Thus, the calculated activation energy of the reaction is $\Delta G^{\ddagger}$ = 22.3 kcal/mol.

Once the MFEP was known, a statistical study over this pathway is reported using the evolution of the variables followed along the metadynamics trajectory. In Figure 4.9, the average values of the collective variables, the six distances taking part in the CV4 and the puckering coordinates over the MFEP are represented. Analysing the evolution of the collective variables of the system, we can see that the reaction follows a four phase mechanism with a unique transition state. We wanted to show the points where starts and finalize each phase of the reaction, for this reason, we have chosen the point A as the initial point of the Phase II and the B point as the initial point of the Phase III of the reaction.



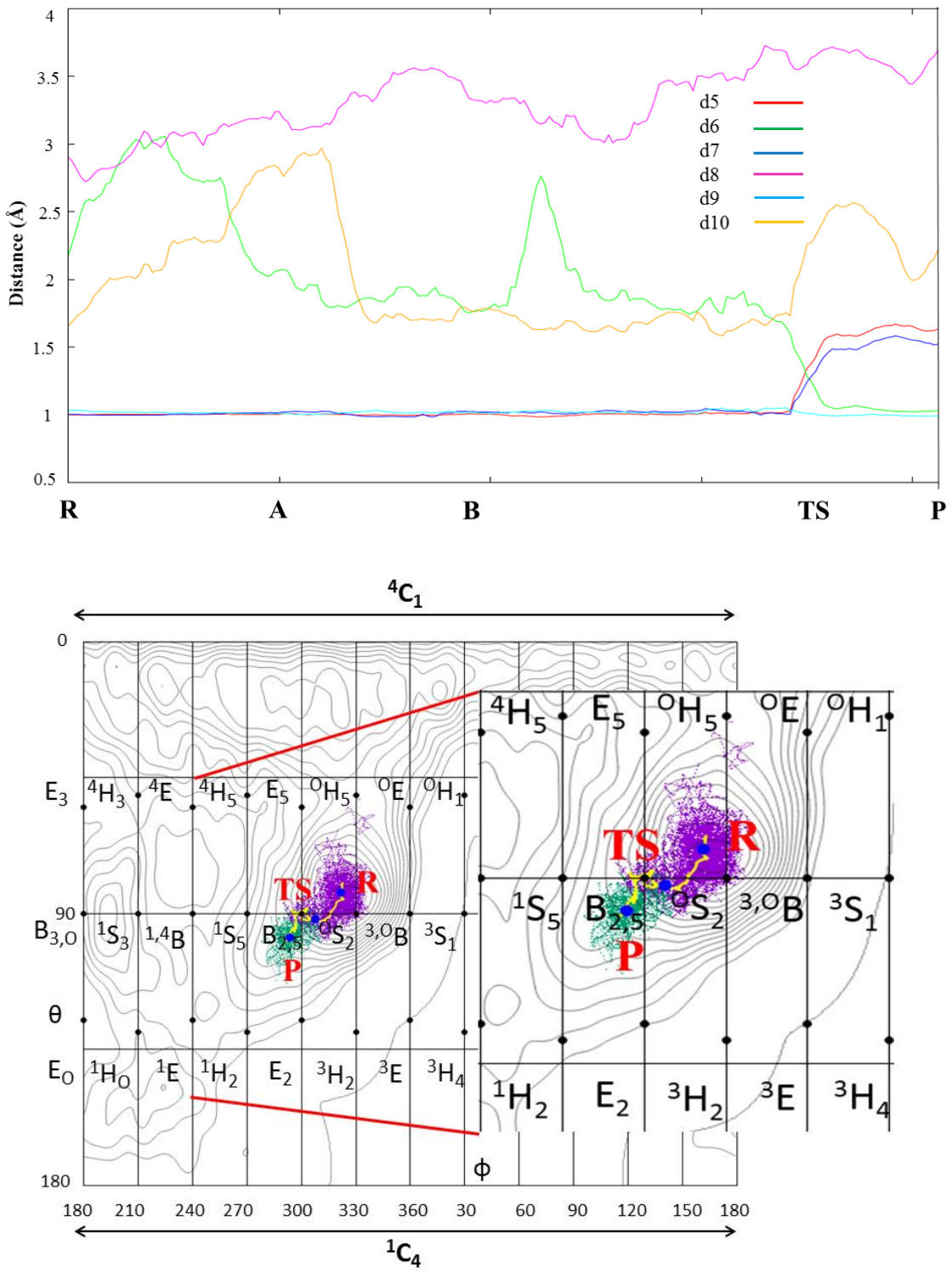


Figure 4.9. Representation of the mean values obtained after a statistical analysis over the MFEP of: the value of the four collective variables (top panel), the value of the six distances (defined in Figure 4.3) involved in the CV4 collective variable (middle panel) and the puckering coordinates of the sugar ring in the -1 subsite (bottom panel). Blue points represent the reactants, transition state and products conformations, the yellow line

shows the evolution of the conformation along the MFEP and the purple and cyan dots show the evolution of the sugar puckering before and after the system crossed the transition state. Black lines are the contour lines of the conformational FEL of the natural substrate in the GH125 enzyme.

The evolution of the system in each phase is the following:

Phase I (from R to A): CV1 evolves from -0.57 to 0.68 Å and the CV4 evolves from -1.22 to -1.66 Å, meanwhile CV2 and CV3 remain constant. The CV1 evolution involves the proton transfer from Asp218 to the glycosidic oxygen. The CV4 evolution implicate a non-relevant movement of the auxiliary waters in the water proton channel.

Phase II (from A to B): CV2 evolves from 1.53 to 2.54 Å and the CV4 increase from -1.66 to -1.33 Å, meanwhile CV1 increases slightly from 0.68 to 0.85 Å and CV3 maintains its value. The CV2 evolution shows the cleavage of the glycosidic bond due to the previous activation of the glycosidic oxygen. The CV1 evolution shows the final formation of a hydrogen bond between the transferred proton and the Asp218. The CV4 evolution shows the formation of a hydrogen bond between the catalytic water and its neighbour water.

Phase III (from B to TS): CV3 evolves from 3.48 to 2.15 Å and the CV2 from 2.54 to 3.22 Å. Then, in this step, the catalytic water starts the nucleophilic attack to the positively charged anomeric carbon, while the protonated leaving group separates definitely from the sugar in the -1 subsite. Furthermore, in spite of the CV4 does not change, a bridge between the catalytic water, the first auxiliary water and the Glu391 is formed.

Phase IV (from TS to P): CV3 evolves from 2.15 to 1.47 Å and the CV4 from -1.34 to -0.92 Å. Structurally, the catalytic water finishes its nucleophilic attack over the anomeric carbon, and immediately, that water transfers one proton to the neighbour water (d5 cleavage and d6 bond formation) and that water transfers one proton to the Glu391 residue (d7 cleavage). The formation of the proton – Glu391 distance is not quantified due to the *a priori* assumption that the third water would have a paper in the proton channel.

From the conformational point of view, the evolution of the ϕ and θ puckering coordinates along the MFEP evolves throughout the $^{O}S_2 \rightarrow B_{2,5} \rightarrow {}^{1}S_5$ region. During Phase I, the sugar cross the $^{O}S_2$ region (322-323, 83-85). During the Phase II and the transition Phase III, the sugar goes to the $B_{2,5}$ region, over the point (308-296, 92-96). Finally, in the last step of the simulation, the ring changes its form to an intermediate conformation $^{1}S_5/B_{2,5}$ in the surface (293, 98).

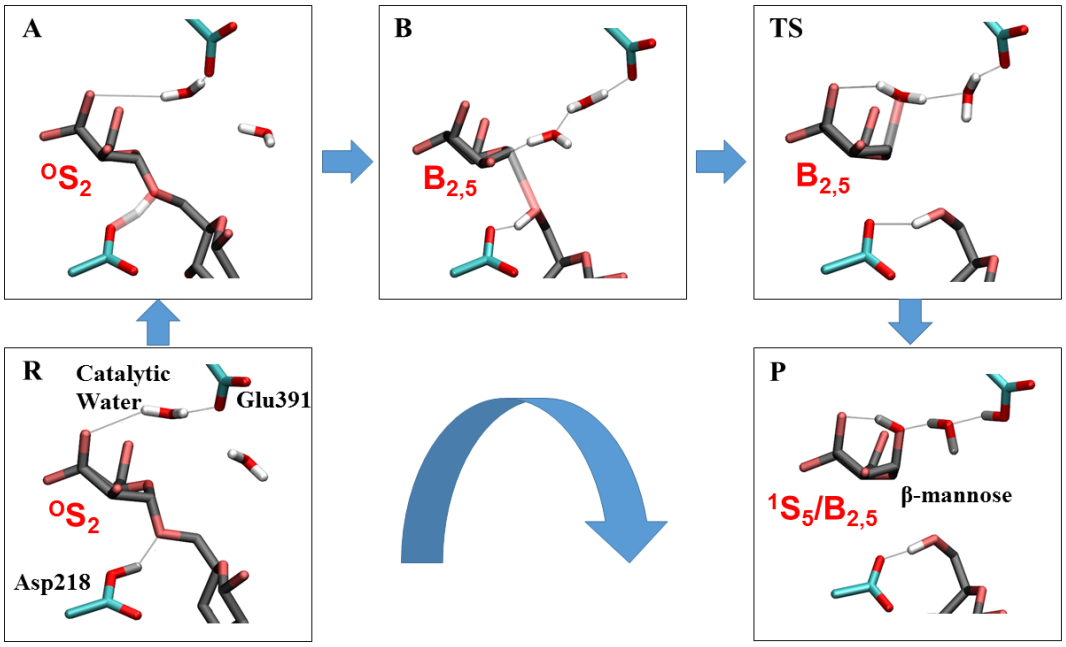


Figure 4.10. Representative structures along the reaction coordinate of the hydrolysis reaction in the.

Table 4.3. Calculated expected values of the collective variables (in Å) and puckering coordinates (in degrees) and its standard deviations along the hydrolysis MFEP (ΔG in kcal/mol).

	ΔG	CV1	CV2	CV3	CV4	ϕ	θ	Conf.
R	0.0	-0.57 ± 0.09	1.47 ± 0.08	3.50 ± 0.08	-1.22 ± 0.09	322 ± 7	83 ± 5	$^{O}S_2$
A	6.1	0.68 ± 0.09	1.53 ± 0.09	3.49 ± 0.09	-1.66 ± 0.09	323 ± 6	85 ± 4	$^{O}S_2$
B	18.4	0.85 ± 0.08	2.54 ± 0.09	3.48 ± 0.07	-1.33 ± 0.09	308 ± 4	92 ± 3	$B_{2,5}$
TS	22.3	1.19 ± 0.09	3.22 ± 0.09	2.15 ± 0.10	-1.34 ± 0.04	296 ± 3	96 ± 3	$B_{2,5}$
P	3.2*	0.88 ± 0.09	3.34 ± 0.08	1.47 ± 0.07	-0.92 ± 0.09	293 ± 7	98 ± 3	$^{1}S_5/B_{2,5}$

* not converged.

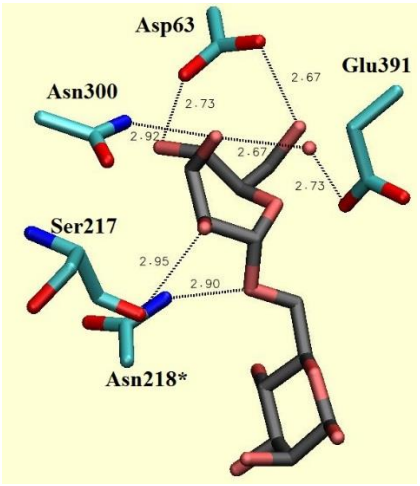
5. Possible errors affecting the results

The results obtained are in reasonable agreement with kinetical experiments with the 4-MU-α-manno-1,6-β-mannoside (Deng, 2013), a better leaving group, whose activation barrier is 16.6 kcal/mol. Most likely the effect of the methylumbelliferyl group decreases the activation barrier. It could also decrease the K_m of the substrate, making the enzyme more efficient. Our computed free energy value (22.3 kcal/mol) is in reasonable agreement with experiments, taking into account that zero-point energy corrections, which decrease the energy barrier by 1-2 kcal/mol, are not included, and the expected error of DFT (GGA functionals) in energy barriers (1-4 kcal/mol) (Fox, 2014) (Catlow, 2016). Previous studies of the group using a similar methodology obtained similar energetic results (Petersen, 2010). Instead, a much higher energy barrier was obtained in GH8 inverting endoglucanase (Petersen, 2009). In this case, the computational resources did not allow using more than two collective variables (the proton donor attack, which is similar to CV1 here, and the antisymmetric combination of the glycosidic bond cleavage, our CV2, and the catalytic water attack, our CV3). The water proton channel (our CV4) was assumed to be spontaneous and thus not included in the CV set. Here, we have separated their second collective variables in two independent coordinates and we have activated properly the water proton channel. As a consequence, our energy barrier is 13 kcal/mol lower, although the mechanism is essentially the same. Recently, a reactivity study of an inverting GH134 β-mannanase performed by our group (Yin, 2016) presented an activation barrier of 17 kcal/mol using 3 collective variables, demonstrating the good results of our methodology.

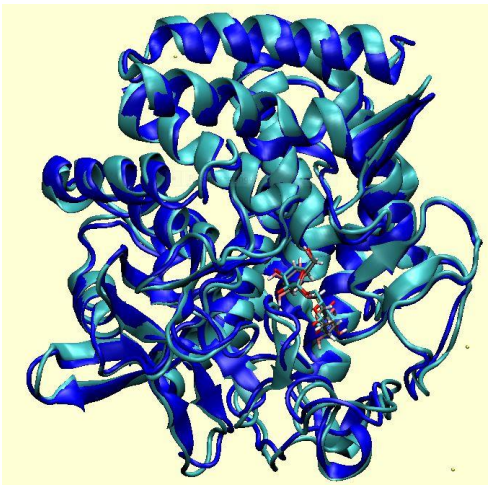
6. Experimental X-ray results

As mentioned in the introduction, our results motivated new experiments of the conformational pathway of GH125 α-mannosidases. There are three experimental strategies to mimic the Michaelis Complex (MC) of an enzyme: mutating the catalytic residues and using the natural substrate, complexing the Wild Type enzyme with a thioglycoside and complexing it with a fluoroderivative (Chapter I – Section 4.5). Due to the strategy of the thioglycoside shown by Gregg et al. does not mimic properly the conformation of the natural sugar, our collaborators (G. J. Davies group) were motivated by our results to change to the mutation strategy.

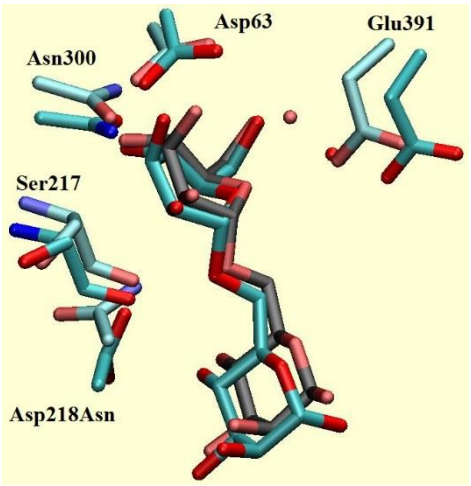
Concretely, as done with the amylosucrase MC, the mannobiose natural substrate was complexed in a proton donor (Asp218Asn) mutant of the *Cp*GH125 α-mannosidase (PDB 5M7Y, resolution 1.55 Å). The resulting structure shows a mannobiose -1 sugar in an $^{O}S_2$ distorted conformation (Figure 4.11a). A representative structure from the conformational metadynamics with the -1 sugar ring in the $^{O}S_2$ conformation were extracted to be superposed and compared with the experimental crystal structure (Figure 4.11b, RMSD 3.6 Å and Figure 4.11c).



(a)



(b)



(c)

Figure 4.11. (a) -1 sugar cavity structure and enzyme-glucoside interactions of the Asp218Asn *Cp*GH125 MC. Superposition of the (b) mutant (blue) crystal structure and (cyan) a representative structure of the conformational metadynamics backbones and (c) a zoom over the -1 sugar cavity (black sugar and brilliant residues corresponds to crystal).

7. General discussion

The conformational free energy landscapes of α-D-mannopyranose and 1-thio-α-D-mannopyranose show that the replacement of oxygen by sulphur does not affect significantly the molecular conformational properties. However, an important change in the conformational free energy landscapes of the 1-thio-α-(1,6)-D-mannobiose and α-(1,6)-D-mannobiose is observed. If the chemical effect of the sulphur atom does not have an important role, which is the reason of this difference?

The results of the conformational study of the thio-compound in GH125 α-1,6-mannosidase are in agreement with the experimental data available, where the most stable conformation of the α-mannosyl in the -1 subsite is the non-distorted $^4C_1$. However, the free energy landscape of the natural substrate in GH125 α-1,6-mannosidase shows a different conformational behaviour, whose most stable minima are $^OS_2$ and $^4C_1$. Most importantly, the distorted minimum ($^OS_2$) is 3.7 kcal/mol more stable than the chair one. Both conformations are preactivated for catalysis, as the glycosidic bond is axial in both bases. Why the distorted conformation is most stable? Analysing the binding of the sugar inside the active site of the enzyme evidences two important features. In first place, the $^OS_2$-distorted sugar interacts better than the $^4C_1$ one with its neighbour residues. In second place, comparing the interactions of both $^4C_1$ conformation of the thio-compound and the natural substrate, there is a visible difference in the hydrogen bond pattern of the hydroxyl 2-OH with Asn300 (1.8 vs. 2.7 Å). The reason that the thio-derivative does not explore the skew-boat region is because the 2-OH hydroxyl interacts strongly with Asn300 and it needs a lot of energy to break this interaction, redirecting itself to Serine 217, a key factor for the ring to become distorted. The difference in the conformational behaviour is given by the enzyme-glycoside interactions of both molecules, whose difference is the length of the sulphur – carbon distance respect to the oxygen – carbon one (1.9 vs 1.5 Å). The bond stretching promotes a better interaction of the $^4C_1$ conformation with the protein active site.

The hydrolysis reaction of the substrate inside the enzyme was simulated using QM/MM metadynamics. The results show a minimum free energy pathway for the hydrolysis reaction with an activation barrier of 22.3 kcal/mol. The reaction can be decomposed into four phases. In the first phase, the acid/base residue (Glu218) protonates the glycosidic oxygen, converting the α-mannosyl at the +1 subsite in a good leaving group. In the second phase, the glycosidic bond increased by 1 Å. In the third phase, the

catalytic water attacks the anomeric carbon, while the leaving group goes far from the -1 sugar. Finally, the catalytic water transfers its activated proton to its neighbour water that makes the same with Glu391, taking part in a two-water proton channel. During this process, the conformation of the -1 sugar evolves through the $^{O}S_2 \rightarrow B_{2,5} \rightarrow {}^{1}S_5$ region, consisting with the proposed catalytic itinerary for inverting α-mannosidases.

Our interaction with experimental colleagues gave as a result the empiric confirmation of the results obtained in these computational study of GH125 α-mannosidases. Their crystal structure showed the natural substrate in an $^{O}S_2$ conformation and a similar hydrogen bond pattern as observed during the simulations.

8. Conclusions

The main conclusions of the present chapter are the following:

1. In the gas phase, the 1-thio-α-mannose and the α-mannose do not present relevant conformational differences. Uniquely, the $^{1}C_4$ conformation is less stable for 1-thio-α-mannose than for α-mannose, relative to the most stable $^{4}C_1$ conformation.
2. The conformational free energy landscape of the 1-thio-α-mannosyl group in the -1 subsite of GH125 α-mannosidase confirms that the most stable conformation is $^{4}C_1$.
3. The conformational free energy landscape of the α-mannosyl group in the -1 subsite of GH125 enzyme α-mannosidase reveals that the most stable conformation is the $^{O}S_2$ skew-boat. Thus, substitution of oxygen by sulphur changes completely the conformation of the sugar in the Michaelis complex.
4. The conformational differences between both sugars can be explained in terms of different hydrogen bond interactions in the active site.
5. The simulation of the hydrolysis process in GH125 α-mannosidase reveals that the -1 sugar follows a $^{O}S_2 \rightarrow B_{2,5} \rightarrow {}^{1}S_5/B_{2,5}$ catalytic itinerary.

Chapter V - Structural study of a promiscuous GH3 β-D-glucan glucohydrolase

Structural study of a promiscuous GH3 β-D-glucan glucohydrolase

1. Introduction

Cell walls of the barley (*Hordeum vulgare*) are rich in (1,3)- and (1,4)-β-D-glucans. One of the key processes in plant development is the (1,3)-, (1,4)- and (1,3;1,4)-β-D-glucans hydrolysis during the wall degradation and wall loosening in elongating coleoptiles (Varghese, 1999). The enzymes, which carry out this step are endo- and exo-acting glucanases (Hrmova, 2001a).

Concretely, a member of the family 3 of the Glycoside Hydrolases (GH3) known as GH3 β-d-glucan glucohydrolase was exhaustively studied due to its important role in barley development. This work will be focused in the exo-1,3-1,4-β-D-glucanase, designated HvExoI, whose experimental evidences are reported in CAZy and PDB databases under our collaborators name (22 crystal structures; Professor M. Hrmova, The University of Adelaide).

In 1999, the first crystal structure of the HvExoI enzyme was obtained by X-ray diffraction (Varghese, 1999), being to date, the only GH3 structure available from plant sources. In Figure 5.1, the representation of the secondary structure of the HvExoI enzyme is shown (PDB 1EX1, resolution 2.2 Å). This barley structure consists in two domains connected by a helix-like strand of 16 residues. The first domain (358 residues) forms an $(\alpha/\beta)_8$ barrel and the second domain (186 residues) forms an $(\alpha/\beta)_6$ sheet, consisting of five parallel β strands and one antiparallel strand, with three α-helices on either side of the structure (Harvey, 2000).

In this structure, a glucose molecule is located in the -1 subsite, where the aspartate 285 is well oriented to the anomeric carbon, being the nucleophilic agent, while the glutamate 491 is positioned as an acid/base residue. The distribution of the catalytic amino acid residues confirms the retaining hydrolytic mechanism of the enzyme.

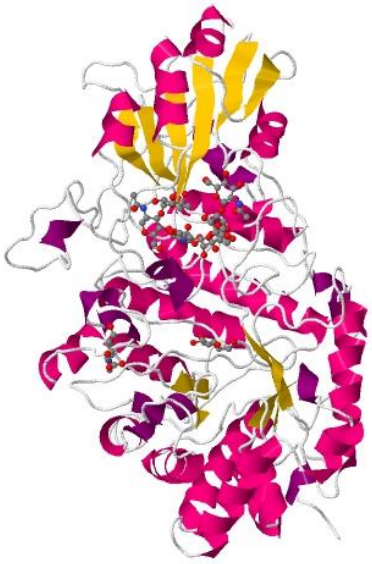


Figure 5.1. A representation of the secondary structures of the HvExoI crystal structure (Varghese, 1999, PDB 1EX1).

This presumption was confirmed experimentally in 2001 (Hrmova et al., 2001b), when the native HvExoI enzyme was crystallized (Figure 5.2d, PDB 1IEQ [resolution 2.70 Å]), 2-deoxy-2-fluoro-β-D-glucose (a covalent intermediate mimic, Figure 5.2c, PDB 1IEW [resolution 2.55 Å]) and 4^{I},4^{III},4^{V}-S-trithiocellohexaose (a substrate mimic, Figure 5.2a, PDB 1IEX [resolution 2.20 Å]). Instead it is not shown in the figure, in 2002, the enzyme was crystallized too in complex with 4'-nitrophenyl 3^{I}-thiolaminaritrioside (a substrate mimic, PDB 1J8V [resolution 2.40 Å]) (Hrmova, 2002). In all structures, the sugar or sugar moiety in the -1 subsite was found to be in a chair conformation.

In 2005, two transition state mimics (glucoimidazole-like compounds) complexes were crystallized (Hrmova, 2004 and 2005). In Figure 5.2b, the inhibitor is bound in the -1 subsite, and its conformation is an envelope (^{4}E), suggesting the shape of the transition state of the glycosylation step during the retaining mechanism catalysed by the HvExoI enzyme.

Thus, the resulting conformational pathway of the catalytic mechanism of this GH3 enzyme will be a cyclic itinerary, where the sugar starts in a chair conformation, which later crosses the ^{4}E region by forming an undistorted covalent intermediate. However, typically, the Michaelis complex of the GH3 β-glucosidases/glucanases should be distorted ($^{1}S_{3}$) to avoid the steric hindrance to the nucleophilic attack and directing the glycosidic bond to an axial position.

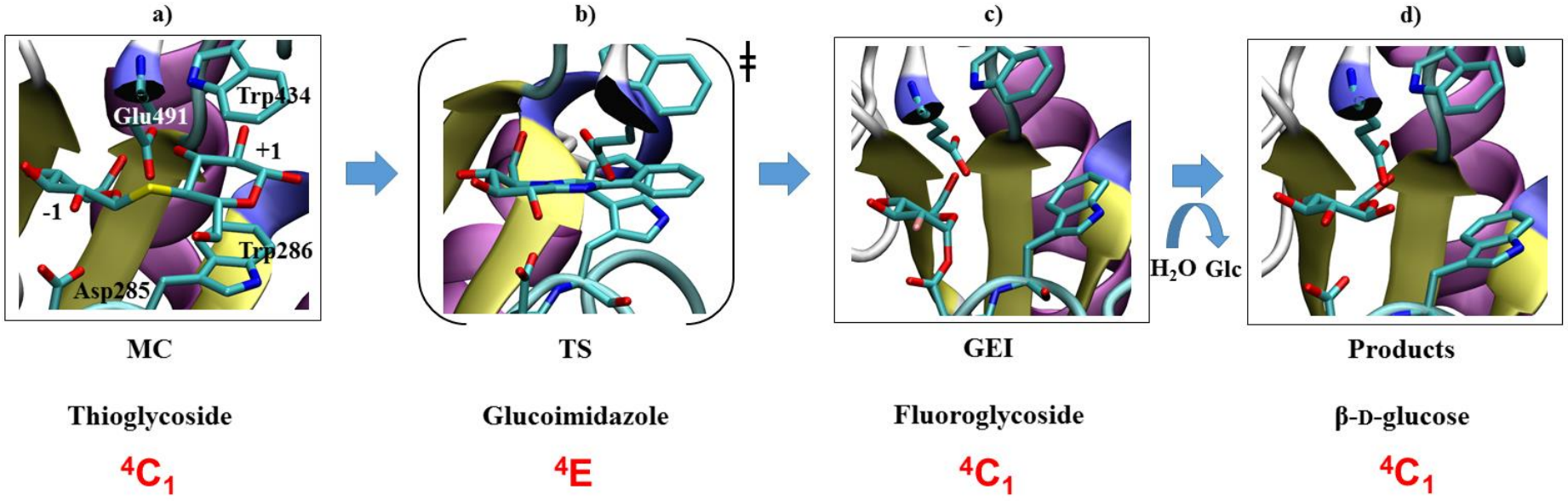


Figure 5.2. Proposed reaction mechanism of HvExoI using the crystallographic structures available (Hrmova, 2001) (Hrmova, 2005) ([a], PDB 1IEX; [b], PDB 1X38; [c], PDB 1IEW; [d], PDB 1IEQ).

The promiscuity of this enzyme is well-known: it exhibits hydrolytic activity of sugar chains with different stereochemistry. In 1998, it was observed that HvExoI hydrolyses 1,3-, 1,4- and 1,6- glycosidic bonds (Hrmova, 1998). Furthermore, the enzyme was crystallised using thio-sugar analogues with 1,3- and 1,4- stereochemistry as substrates (Hrmova, 2001b) (Hrmova, 2002), showing both -1 sugars in the $^4C_1$ conformation. Finally, kinetic studies demonstrated the capability of the HvExoI enzyme to hydrolyse 1,2-, 1,3-, 1,4- and 1,6- linked sugars (k_{cat} ~ 2-12 s^{-1}) (Hrmova, 2002). This promiscuity is due to the fact that the +1 subsite sugar is bound through a π-stacking sandwich interactions of the Trp286 and Trp434, and is not surrounded by hydrophilic residues that constraint the stereochemistry of the substrate.

In this chapter, we present simulations from four crystallographic structures of the HvExoI enzyme in complex with thio-sugar analogues with different stereochemistry, obtained from the Professor Hrmova's research group: the HvExoI/methyl 2-S-β-D-glucopyranosyl-2-thio-β-D-glucopyranoside (methyl-O-β-thio-sophoroside) complex (**G2S complex**); the HvExoI/4-nitrophenyl *S*-(β-D-glucopyranosyl)-(1,3-β-D-glucopyranoside (4-nitrophenyl-O-β-thio-laminaribioside) complex (**G3S complex**); the HvExoI/4^I, 4^{III}, 4^V-S-trithiocellohexaose complex (**G4S complex**); and the HvExoI/methyl 6-S-β-D-glucopyranosyl-6-thio-β-D-glucopyranoside (methyl-O-β-thio-gentiobioside) complex (**G6S complex**). The -1 sugar conformation for each system is represented in Figure 5.3.

The calculated puckering coordinates of the -1 sugar of the four complexes show important differences. Whereas G3S and G4S thio-sugars exhibit ^4E and $^4H_3$ conformations, respectively, G2S and G6S thio-sugars are not distorted ($^4C_1$). Taking into account the previously analysed thioglycoside substrates in GH125, in which the thio-sugar did not mimic the conformation of the natural substrate, the conformational behaviour of the thio-derivative substrates in HvExoI is surprising, and thus we decided to analyse this aspect in detail, in comparison with its natural substrates. Furthermore, we will identify critical residues at the -1 and +1 subsites that play a role in the acceptance of several substrates with different stereochemistries and the capability of hydrolyse them.

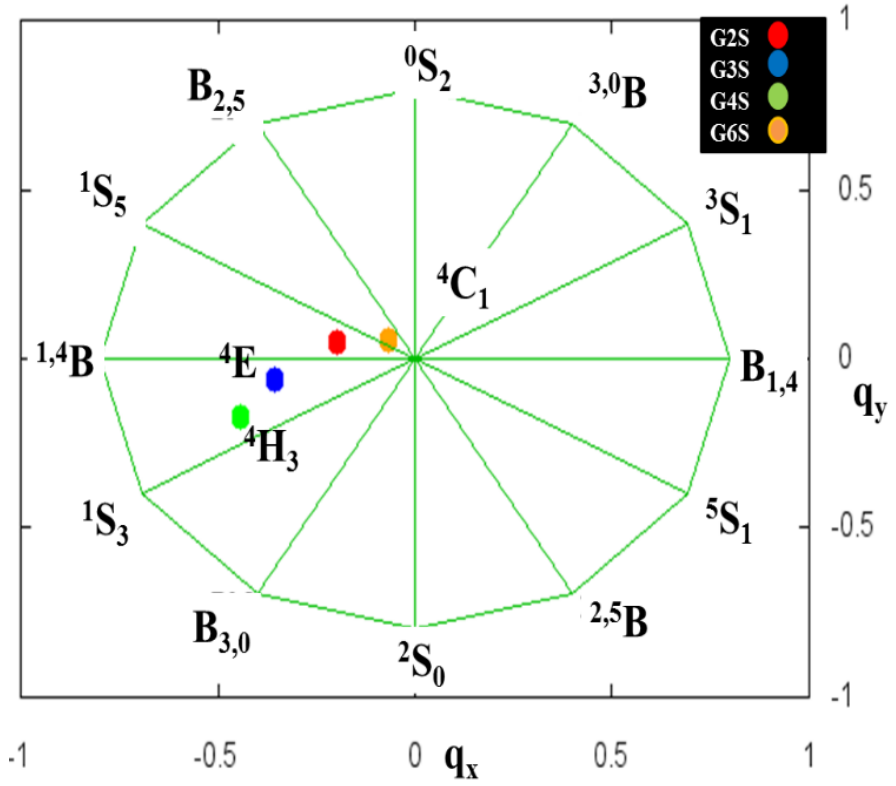


Figure 5.3. Representation of the puckering coordinates (Stoddart projection) of the –1 sugar rings of the thio-sugar analogues in the crystallographic complexes of HvExoI.

2. Objectives

The main objectives of the chapter are:

1) To characterize the experimental GH3 crystallographic structures and analyse the conformational behaviour of the -1 subsite thio-glycoside moieties.
2) To reconstruct and characterize the GH3 – natural substrates models and analyse the conformational behaviour of the -1 subsite sugar moieties.
3) To reconstruct the conformational free energy landscape of the sugar at the -1 subsite in the G3O model.

3. Computational details

G2S and G2O models

The structure of the HvExoI (Trp292Phe/G2S complex) is formed by 602 protein residues and the methyl-O-β-thio-sophoroside ligand, three N-acetyl β-D-glucosamine (NAG) molecules, four glycerol (GOL) molecules, two SO_4^{2-} anions and 430 crystallographic water molecules. The resolution of this structure is at 2.16 Å and it has an R-free factor of 0.21. The B-factor of the sugar in the -1 subsite is between 26.3 and 38.4 $Å^2$.

To generate the wild-type HvExoI/G2x models, the mutation was reverted (Phe286 converted to Trp286), and the 1-OMe group of the sugar moiety at the +1 subsite, the sulphate anions, and the NAG and GOL molecules were removed. Using the tleap program of the AMBER 11 code, the -3.0 total charge of the systems was compensated with three sodium cations and the models were solvated with 25348 TIP3P water molecules (Jorgensen, 1983), forming a parallelepiped box with the dimensions of 103.1 Å x 92.1 Å x 106.7 Å.

G3S and G3O models

The structure of the HvExoI/G3S complex is formed by 602 protein residues and the 4-nitrophenyl-O-β-thio-laminaribioside substrate, one pentaethylene glycol (1PE) molecule, one N-acetyl glucosamine (NAG) molecule, 8 SO42- anions and 352 crystallographic water molecules. The resolution of this structure is at 2.00 Å and it has an R-free factor of 0.23. The B-factor of the sugar in the -1 subsite is between 34.3 and 52.3 $Å^2$.

To generate the wild-type HvExoI/G3x models, the sulphate anions, the NAG, 1PE molecules and the 4-nitro-phenyl group were removed. Using the same computational strategy that in the previous model, the -3.0 total charge of the systems was compensated with three sodium cations and the models were solvated with 25080 TIP3P water molecules, forming a parallelepiped box with the dimensions of 103.0 Å x 92.1 Å x 106.4 Å.

G4S and G4O models

The structure of the HvExoI/G4S complex is formed by 602 residues and the 4^{I}, 4^{III}, 4^{V}-S-trithiocellohexaose substrate, three glycerol (GOL) molecules, three NAG molecules, three SO_4^{2-} anions, two 1PE molecules and 545 crystallographic water molecules. The resolution of this structure is at 1.78 Å and it has an R-free factor of 0.17. The B-factor of the sugar in the -1 subsite is between 23.2 and 33.5 $Å^2$.

To generate the wild-type HvExoI/G4x models, the sulphate anions, the GOL, NAG and 1PE molecules were removed. Using the same computational strategy that in the previous model, the -3.0 total charge of the systems was compensated with three sodium cations and the models were solvated with 26087 TIP3P water molecules, forming a parallelepiped box with dimensions of 102.9 Å x 91.7 Å x 107.0 Å.

G6S and G6O models

The structure of the G6S complex is formed by 602 protein residues and the methyl-O-β-thio-gentiobioside, ten GOL molecules, one NAG molecule, two SO_4^{2-} anions and 852 crystallographic water moleculess. The resolution of this structure is at 1.57 Å and it has an R-free factor of 0.18. The B-factor of the sugar in the -1 subsite is between 13.3 and 16.3 $Å^2$.

To generate the wild-type HvExoI/G6x models, the sulphate anions, the GOL and NAG molecules were removed. Using the same computational strategy that in the previous model, the -3.0 total charge of the systems was compensated with three sodium cations and the models were solvated with 25834 TIP3P waters, forming a parallelepiped box with dimensions of 106.5 Å x 91.7 Å x 108.2 Å.

Classical molecular dynamics

All models were parametrized using the FF99 (Cornell, 1995) force field of AMBER 11, the Glycam 06 libraries (Kirschner, 2008) for the sugars and the TIP3P three charges description for water molecules.

The process of stabilization of models at 300 K consists of 40000 steps of solvent and ions minimization, 40000 steps of all system minimization, a progressive heating to 200 K during 300 ps in steps of 50 K, a progressive heating to 300 K during 400 ps in steps of 25 K, 50 ps of NVT stabilization, 100 ps of NPT simulation (P = 1 atm) to ensure a physiological density (1.06 g/cm^3) (time-step 1 fs) and a final NVT simulation (time-

step 2 fs) at 300 K for 8-11 ns, until the RMSD (Root Mean Square Deviation) value of the protein backbone was stabilized.

As an exception, during a previous simulation we observed that the conformation of the -1 sugar in the **G4S model** oscillates frequently from $^4C_1$ to $^1S_3$. Thus, we decided to start the stabilization process through restraining the -1 sugar conformation during heating and density arrangements and through 10 ns of NVT molecular dynamics. Subsequently, the restriction was released and the simulation was extended by10 additional nanoseconds.

QM/MM molecular dynamics (MD) and well-tempered metadynamics

As all GxO (x means 1,2-; 1,3; 1,4 and 1,6-linked sugar residues) models exhibited the same conformational pattern, we decided to study one of the models **(G3O model)** using a more accurate method to validate the conformational behaviour of the -1 glucosyl moiety observed with classical MD. Thus, we opted to improve the description of the sugar molecule from the classical to quantum level of theory. Our QM/MM model takes used quantum mechanics to describe the substrate and classical MD to describe the rest of the enzyme. The simulation was performed using the CPMD software patched with the Plumed 2 libraries. The QM box is formed by all substrate atoms (a disaccharide β-1,3-thio-laminaribioside) and its (it this correct) electron density is described by a 70 Ry plane waves package in a 13.5 Å x 19.3 Å x 14.8 Å box. The NN, MIX and ESP interaction radii are 5.3 Å, 5.3 Å and 9.0 Å. Initially, taking an equilibrated structure from classical MD, an energy minimization using annealing velocities was performed until the maximum nuclear gradient was under 5·10-4 a.u. using a Lagrangian electronic mass of 600 a.u. and a time step of 5 a.u. After we ensured the adiabaticity between the nuclei and the electrons, the system was equilibrated at 300 K using a Nosé-Hoover (Nosé, 1984) thermostat to control the nuclei temperature (300 K, 3500 cm-1) and the electronic kinetic energy (0.009 a.u., 10000 cm-1) during a 8.5 ps MD simulation. Here, Root Mean Square Deviation (RMSD) value stabilization occurred below 0.5 Å (Figure 5.4).

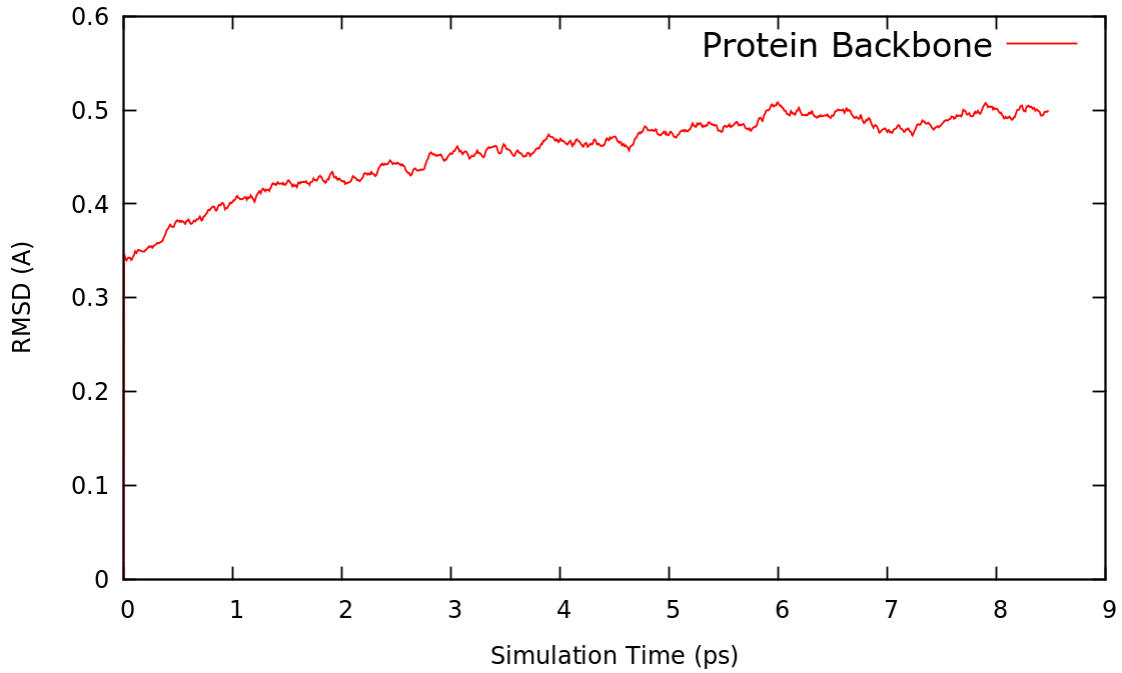


Figure 5.4. Evolution of the RMSD value of the protein backbone of the QM/MM model of the G3O complex during CP MD prior to well-tempered metadynamics simulation.

Once the system was equilibrated, a conformational study of the -1 sugar ring was performed using the Cremer and Pople (Cremer, 1975) puckering coordinates (φ and θ, Mercator projection) as collective variables within the well-tempered metadynamics (Barducci, 2008) algorithm. The following simulation paratamers were used: 842 hills of 0.1 rad x 0.1 rad (well-tempered height of 0.75-0.15 kcal/mol, bias-factor of 12 kcal/mol, temperature 300 K) added along a 40.5 ps molecular dynamics. Deposition time: 400 steps x 5.0 a.u.

4. Results

The HvExoI enzyme is able to bind at least three sugar moieties in the active site pocket. The -1 subsite is formed by Leu54, Asp95, Phe144, Lys206, His207 (δ-protonated), Asp285 (a catalytic nucleophile in a dissociated form), Glu220 and Trp430 (proton donors), and Glu491 (a catalytic acid/base in a protonated form). The +1 subsite is formed by Gly57, Tyr253 and **two Trp286 and Trp434 in the nearly parallel orientation**. These two residues form sandwich π-stacking interactions with the +1 sugar moiety. The +2 subsite is partially solvent exposed and it is formed by Arg291 and Glu287. The RMSD value of the active site residues was calculated from the position of these amino acid residues, as well as Trp286 and Trp434 at the +1 subsite.

This section is organized as follows: first, we present the analysis of G3S and G3O classical simulations. This is illustrated by the representation of the RMSD value evolution of the protein backbone, the active site residues, Trp286 and Trp434, (a sugar substrate along the NVT stabilization, the superposition of the crystal structure (blue), the mean structure of the last 5 ns of classical MD simulation (cyan) and the comparison of both simulations and initial coordinates. For the G4x, G6x and G2x, the RMSD value evolution and the superposition Figures will be presented in the Appendix B. Secondly, the conformational behaviour and the positions of substrates is presented. Finally, the results of the QM/MM study will provide a more precise description of the conformational free energy landscape of the -1 sugars in the active site of HvExoI.

4.1. Classical MD of the HvExoI/laminaribiose (1,3-linked disaccharide) and HvExoI/thio-laminaribiose complexes

In Figure 5.5, the RMSD value evolution of the protein backbone, active site residues, sugar molecule and the +1 subsite Trp286 and Trp434, along the NVT MD simulations applied over the G3x models demonstrate a convergent behaviour for both systems (10 ns for G3S and 11 ns for G3O were required).

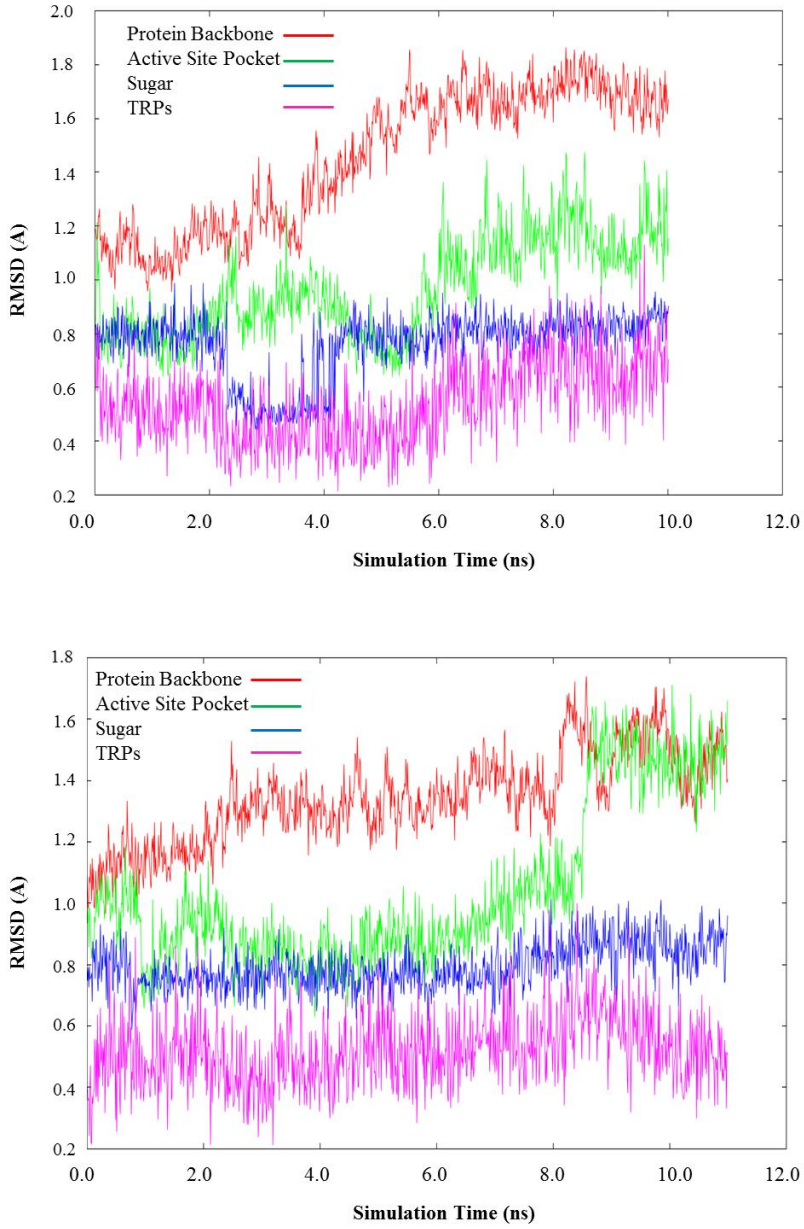
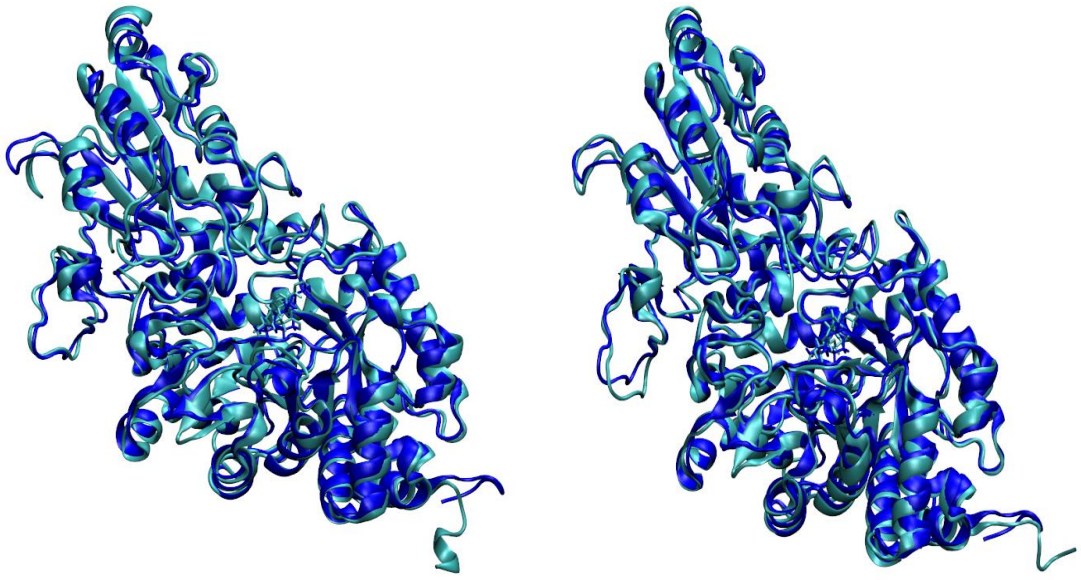


Figure 5.5. Evolution of the RMSD value of: the protein backbone (red), the active site pocket (green), the sugar (blue) and the tryptophan (Trp286 and Trp434) residues at the +1 subsite (purple) of the G3S (top) and G3O (bottom) models during the classical MD.

After 10 ns of NVT simulation (G3S model), the RMSD value of the protein backbone was stabilized around 1.70 Å. The superposition of the crystallographic data and the mean structure of the last 5 ns show the RMSD value of 1.71 Å (Figure 5.6 [a]).

The active site pocket, the G3S sugar, and Trp286 and Trp434 (delineating the +1 subsite) converged around 1.10 Å, 0.80 Å and 0.65 Å, respectively.

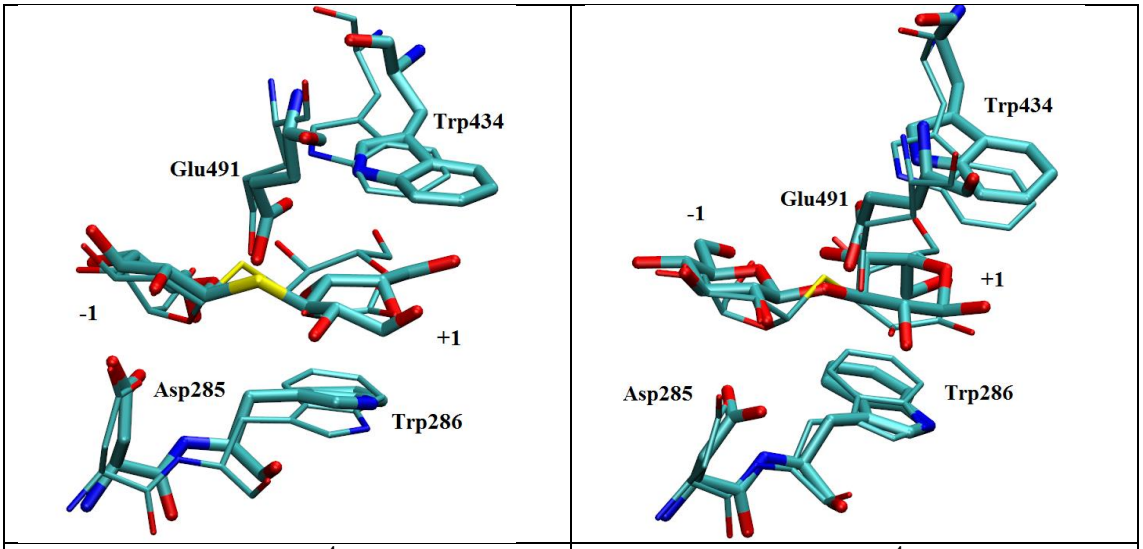
After 11 ns of NVT simulation (G3O model), the RMSD value of the protein backbone was stabilized around 1.50 Å. The superposition of the crystallographic data and the mean structure of the last 5 ns show a RMSD value of 1.45 Å (Figure 5.6 [b]). The active site pocket, the G3O sugar, and Trp286 and Trp434 converged around 1.50 Å, 0.80 Å and 0.50 Å, respectively. The difference between G3S and G3O models was in the abrupt change of the RMSD of active site residues during the G3O simulation. This change was due to a rotation of the Tyr253, which however did not affect significantly the active site pocket topology.



(a) (b)

Figure 5.6. Superposition of the crystallographic structure (blue) and the average structure of the last 5 ns of classical MD (a, G3S model, cyan, and b, G3O model, cyan).

When focusing on the active site pocket, in particular on residues involved in binding sugars (Figure 5.7), the superposition of both models showed that the resulting conformation of the -1 sugar was $^{4}C_{1}$. Furthermore, the Asp285 was well-oriented for the nucleophilic attack despite the ring not being distorted. The Glu491 interacted directly with the glycosidic oxygen as the acid/base agent. In both cases, Trp286 and Trp434formed a sandwich around the +1 sugar. The +1 sugar in the G3S model was distorted ($^{2}S_{O}$), something that we think is an artefact of the force-field.



-1 Sugar $^4C_1$ +1 Sugar ^2S_O (artificially distorted)	-1 Sugar $^4C_1$ +1 Sugar $^4C_1$

*The thin representation corresponds to the crystal structure.

Figure 5.7. Superposition of the substrate, catalytic residues and +1 tryptophan residues between the crystal structure (G3S) and those of the average structures of G3S (left), and G3O (right) models.

The most relevant enzyme-sugar interactions are listed in Table 5.1. It is worth mentioning the internal interaction between the pyranic oxygen (O_p) of the -1 sugar with the 4-OH hydroxyl of the +1 sugar and the mobility of the 6-OH arm to interact with Asp95, 4-OH of the +1 sugar and a neighbouring crystallographic water.

Table 5.1. Principal interactions between the -1 and +1 sugars and the enzymatic medium (HvExoI) in the G3x simulations.

-1 Sugar		+1 Sugar	
2-OH	O_2-Asp285	Ring	π-stacking Trp434 π-stacking Trp286
3-OH	N-ε-His207	4-OH	O_p (-1 Sugar)
4-OH	O_2-Asp95 6-OH	1-OH, 2-OH, 6-OH	Solvent
6-OH	4-OH (+1 Sugar) Internal H_2O		

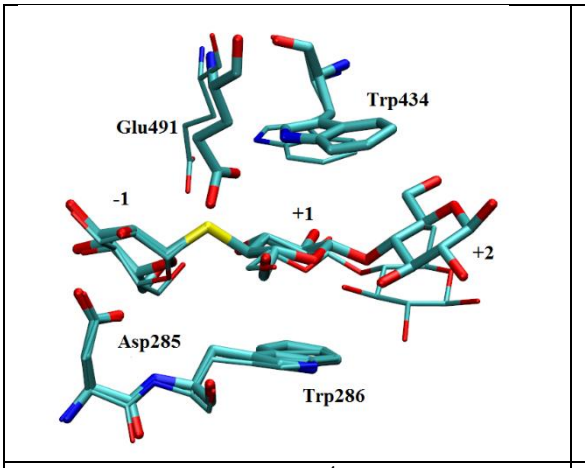
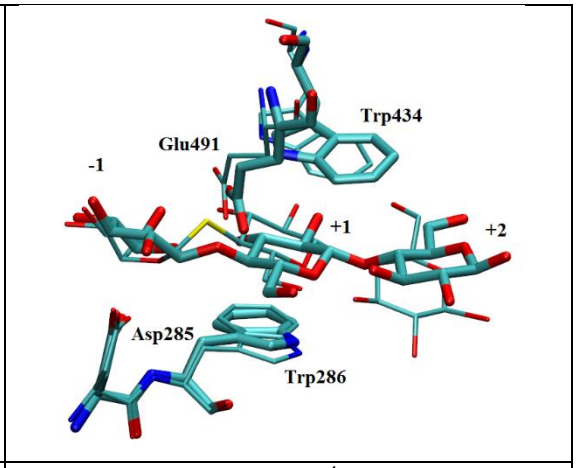
4.2. Classical MD of the HvExoI/cellotriose (1,4-linked disaccharide) and HvExoI/thio-cellobiose complexes

In the case of the G4S model, the NVT simulation was performed in three steps: the first 10 ns with the restriction of the -1 sugar ring, the following 5 ns where the structure was adapted to the conformation of the substrate. The global RMSD value stabilized during the following 5 ns. In the case of the G4O system, 11 ns were necessary to observe a stable RMSD value.

After 20 ns of NVT simulation (G4S model), the RMSD value of the protein backbone was stabilized around 1.60 Å. The superposition of the crystallographic data and the mean structure of the last 5 ns show a RMSD value of 1.71 Å (Appendix B). The active site pocket, sugar, and Trp286 and Trp434 (delineating the +1 subsite) converged around 0.90 Å, 0.80 Å and 0.50 Å, respectively.

After 11 ns of NVT simulation (G4O model), the RMSD of the protein backbone was stabilized around 1.45 Å. The superposition of the crystallographic data and the mean structure of the last 5 ns show a RMSD value of 1.52 Å (Appendix B). The active site pocket, sugar, and Trp286 and Trp434 converge around 0.70 Å, 0.70 Å and 0.40 Å, respectively.

When focusing on the active site pocket, in particular on residues involved in binding sugars (Figure 5.8), the superposition of both models showed that the resulting conformation of the -1 sugar was $^4C_1$. Moreover, the Asp285 is well-oriented for the nucleophilic attack despite the ring was not distorted. The Glu491 was interacting directly with the glycosidic oxygen as an acid/base agent. In both cases, Trp286 and Trp434 were forming a sandwich around the +1 sugar. In this case, both the +1 and +2 sugar rings remained undistorted.

	
-1 Sugar $^4C_1$ +1 Sugar $^4C_1$ +2 Sugar $^4C_1$	-1 Sugar $^4C_1$ +1 Sugar $^4C_1$ +2 Sugar $^4C_1$

*The thin representation is the active site and sugar of the crystal structure (numbering follows that of crystal structures).

Figure 5.8. Superposition of the substrate, catalytic residues and +1 tryptophan residues (Trp286 and Trp434) between the crystal structure (G4S) and the average structures of G4S (left) and G4O (right) models.

Key enzyme-sugar interactions are listed in Table 5.2. It is worth mentioning that the internal interaction was occurring between the pyranic oxygen (O_p) of the -1 sugar and the 3-OH hydroxyl of the +1 sugar. The +1 and +2 sugars are partially and completely solvent exposed, respectively.

Table 5.2. Key interactions between the -1 and +1 sugars and HvExoI during G4x simulations.

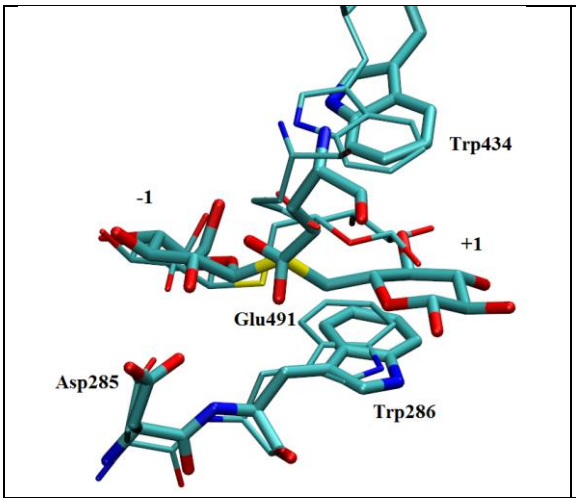
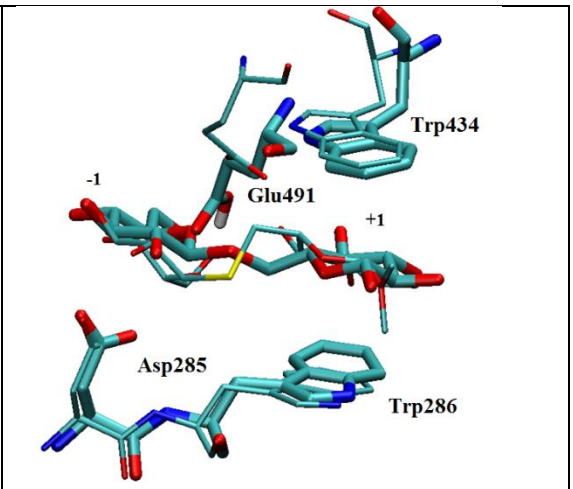
-1 Sugar		+1 Sugar	
2-OH	O_2-Asp285	Ring	π-stacking Trp 434 π-stacking Trp 286
3-OH	N-ε-His207	3-OH	O_p (-1 Sugar)
4-OH	O_1-Asp95 6-OH	1-OH, 2-OH, 6-OH	Solvent
6-OH	O_2-Asp95		

4.3. Classical MD of the HvExoI/gentiobiose (1,6-linked disaccharide) and HvExoI/thio-gentiobiose complexes

After 10 ns of NVT simulation (G6S model), the RMSD of the protein backbone is stabilized around 1.40 Å. The superposition of the crystallographic data and the mean structure of the last 5 ns shows a RMSD between them of 1.33 Å (Appendix B). That of the active site pocket, sugar and Trp residues delineating the +1 subsite converge around 1.20 Å, 0.30 Å and 0.90 Å. This system was the first to show a non-expected behaviour. The approach of the acid proton to the glycosidic position allows the rotation of the 6-S-CH_2- between the -1 and the +1 sugars. Despite this fact, the conformation of the +1 sugar did not change. (Figure 5.9 [left]).

After 10 ns of NVT simulation, the RMSD of the protein backbone stabilized around 1.30 Å. The superposition of the crystallographic data and the mean structure of the last 5 ns shows a RMSD between them of 1.31 Å (Appendix B). The active site pocket, sugar and TRPs converge around 1.30 Å, 1.00 Å and 0.80 Å. As seen before in the G6S model, the RMSD of the delimiting TRPs of the +1 subsite is higher than in the G3 and G4 simulations.

Focusing on the active site pocket, in Figure 5.9, the superposition of both systems show that the resulting conformation for the -1 sugar is $^4C_1$. Notably, Asp285 is well-oriented for nucleophilic attack despite the ring is not distorted. Once Glu491, the acid/base catalytic residue, is interacting directly with the glycosidic oxygen, the -1,6-arm rotates to expose the lone pair electrons of the glycosidic oxygen. In both cases, TRPs are forming a sandwich around the +1 sugar. +1 Sugars are not distorted.

	
-1 Sugar $^4C_1$ +1 Sugar $^4C_1$	-1 Sugar $^4C_1$ +1 Sugar $^4C_1$

*The thin representation is the active site and sugar of the crystallographic structure (numbering follows that of crystal structures).

Figure 5.9. Superposition of the substrate, catalytic residues and +1 tryptophan residues between the crystallographic structure (G6S) and the average structure of the G6S (left) and the G6O (right) models.

The principal enzyme-sugar interactions are listed in Table 5.3. It is worth mentioning the internal interaction between the pyranic oxygen (O_p) of the -1 sugar with the 4-OH hydroxyl of the +1 sugar. The main difference between both simulations is the interaction of the acid/base with the 2-OH (G6S) and the glycosidic oxygen (G6O).

Table 5.3. Principal interactions between the -1 and +1 sugars and the enzymatic medium (HvExoI) in the G6x simulations.

-1 Sugar		+1 Sugar	
2-OH	O_2-Asp285 $H_{A/B}$-Glu491 (thio)	Ring	π-stacking Trp 434 π-stacking Trp 286
3-OH	N-ε-His207	4-OH	O_p (-1 Sugar)
4-OH	O_1-Asp95	1-OH, 2-OH, 3-OH	Solvent
6-OH	Internal H_2O		

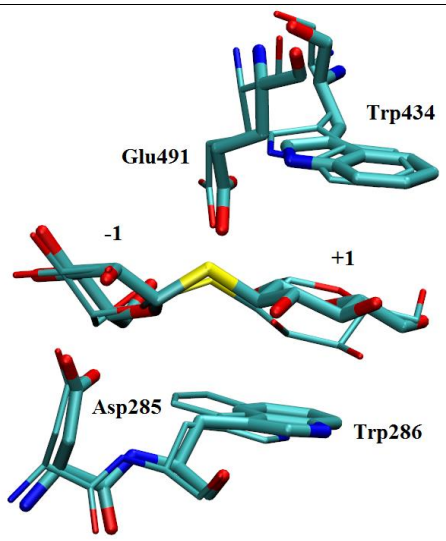
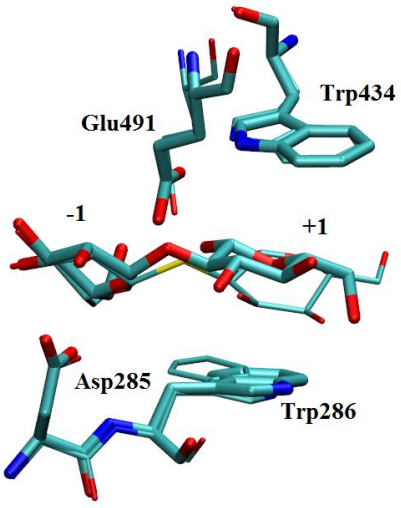
4.4. Classical MD of the GH3 – β-1,2-glucobiose and thio-compound complexes

The crystallographic structure of this system was obtained from the sole mutation of Trp286 to a phenylalanine. The mutation was reversed manually in the simulations.

After 10 ns of NVT simulation (G2S model), the protein backbone is stabilized around 1.40 Å. The superposition of the crystallographic data and the mean structure of the last 5 ns shows a RMSD between them of 1.34 Å (Appendix B). The active site pocket, sugar and TRPs (delineating the +1 subsite) converge around 1.10 Å, 0.70 Å and 0.80 Å.

After 8 ns of NVT simulation (G2O model), the protein backbone is stabilized around 1.50 Å. The superposition of the crystallographic data and the mean structure of the last 5 ns shows a RMSD between them of 1.46 Å (Figure below). The active site pocket, sugar and TRPs converge around 1.10 Å, 0.75 Å and 0.75 Å.

Focusing on the active site pocket, in Figure 5.10, the superposition of both systems show that the resulting conformation for the -1 sugar is $^4C_1$. Notably, Asp285 is well-oriented for the nucleophilic attack despite the ring is not distorted. The Glu491 is interacting directly with the glycosidic oxygen as acid/base agent. In both cases, TRPs are forming a sandwich around the +1 sugar. This glycoside ring remains undistorted and moves its position due to the presence of the mutated Trp286.

	
-1 Sugar $^4C_1$ +1 Sugar $^4C_1$	-1 Sugar $^4C_1$ +1 Sugar $^4C_1$

*The thin representation is the active site and sugar of the crystallographic structure (numbering follows that of crystal structures).

Figure 5.10. Superposition of the substrate, catalytic residues and +1 tryptophan residues between the crystallographic structure (G2S) and the average structure of the G2S (left) and the G2O (right) models.

The principal enzyme-sugar interactions are related in the Table 5.4. It is worth mentioning the internal interaction between the pyranic oxygen (O_p) of the -1 sugar with the 4-OH hydroxyl of the +1 sugar. There are not important differences between both simulations.

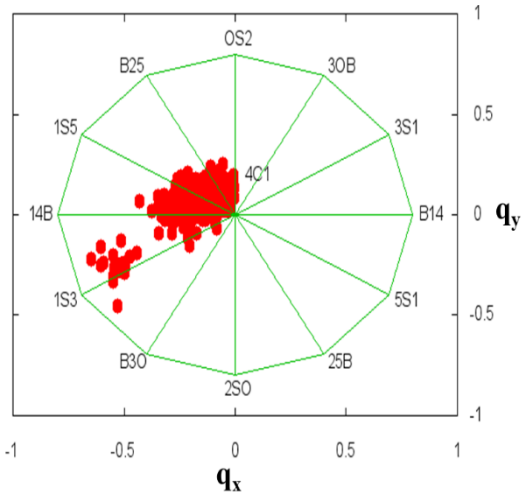
Table 5.4. Principal interactions between the -1 and +1 sugars and the HvExoI enzyme in the G2x simulations.

-1 Sugar		+1 Sugar	
2-OH	O_2-Asp285	Ring	π-stacking Trp434 π-stacking Trp286
3-OH	N-ε-His207	4-OH	O_p (-1 Sugar)
4-OH	O_2-Asp95 6-OH	1-OH, 3-OH, 6-OH	Solvent
6-OH	O_1-Asp95		

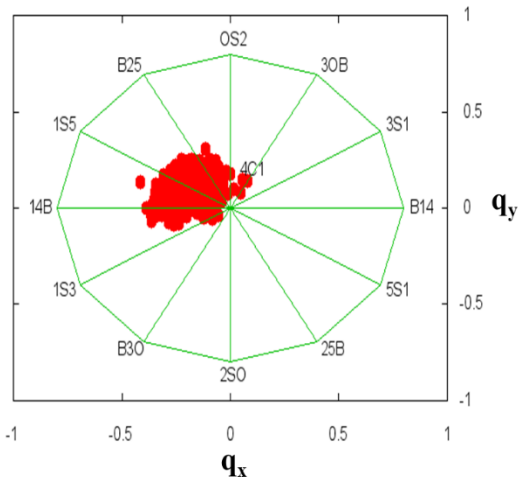
4.5. General conformational observations

The HvExoI simulations showed two differential conformational patterns: the thioglycoside systems adopt conformations in the $^4C_1$ – $^4E/^4H_3$ – $^1S_3$ region while the natural substrate systems adopt conformations close to $^4C_1$ (Figure 5.11). However, in all cases the most populated region is the $^4C_1$ conformation.

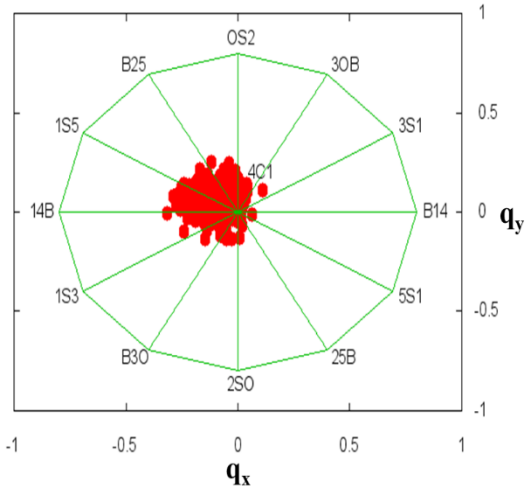
G2S



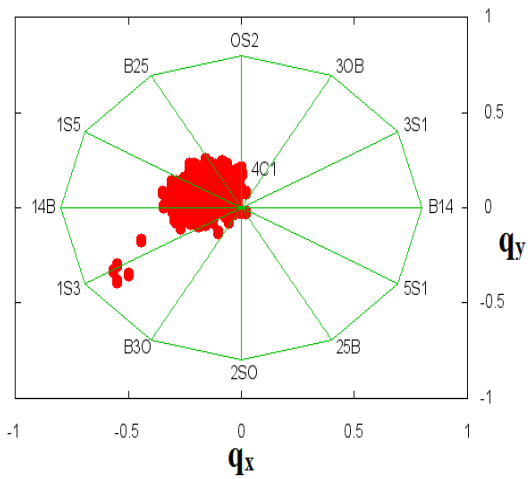
G2O



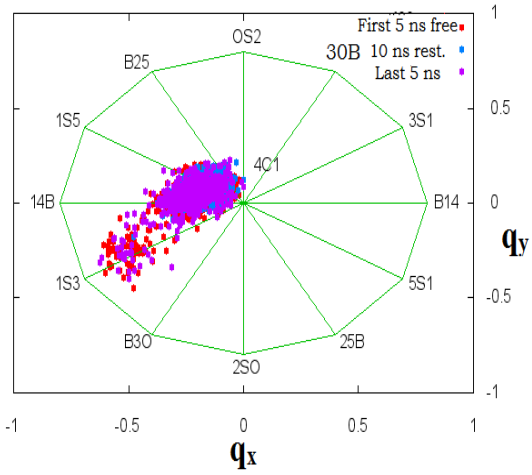
G3S



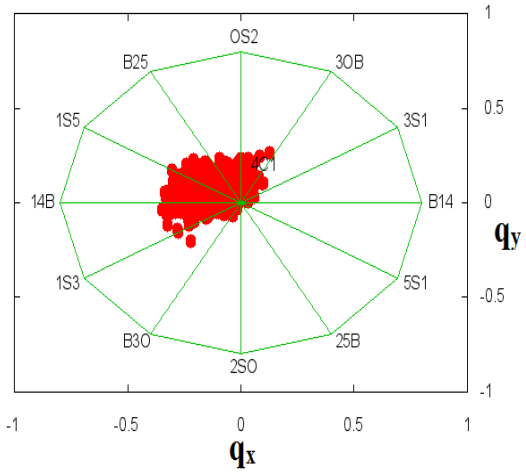
G3O



G4S



G4O



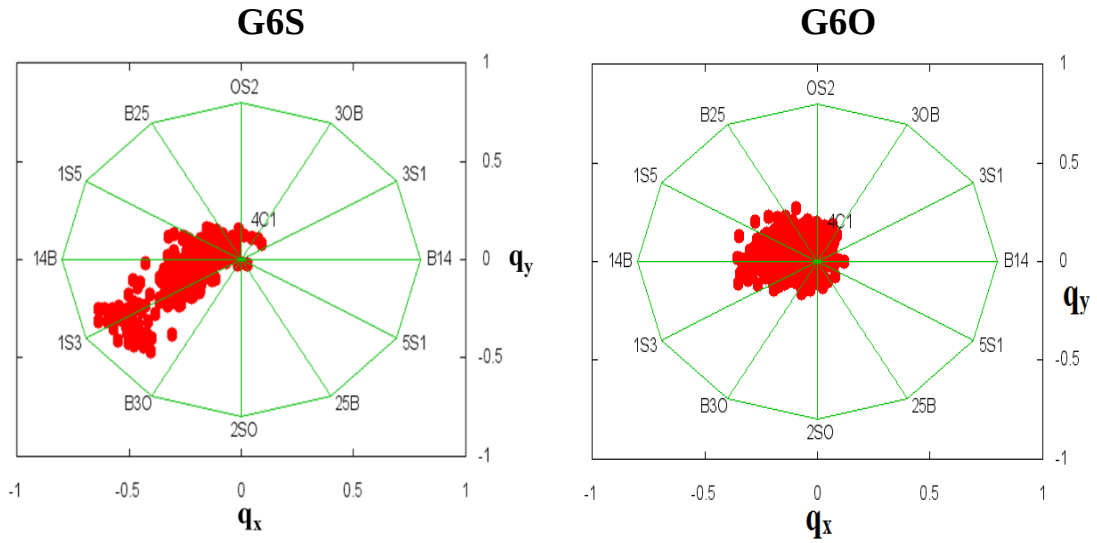


Figure 5.11. Representations of the conformational behaviour (Stoddart diagram) of the -1 subsite sugars during classical MD of each model once the RMSD value was converged.

The discrepancies between both conformational behaviours can be attributed to the presence of the sulphur atom between the -1 and the +1 sugars. Analysing the trajectory of the simulations, we observed that the glycosidic bond distance is larger for the thio-sugars (1.87 Å) than for the O-linked sugars (1.46 Å) and the angle of the C_1-S-C_N (N = 2, 3, 4, 6) is wider (104º) than in their oxygen counterparts (109º). This causes the 6-OH of the -1 sugar to change its orientation to interact with the near hydroxyl of the +1 sugar, which in turn promotes the conformational change from the $^4C_1$ chair to a skew-boat (Figure 5.12).

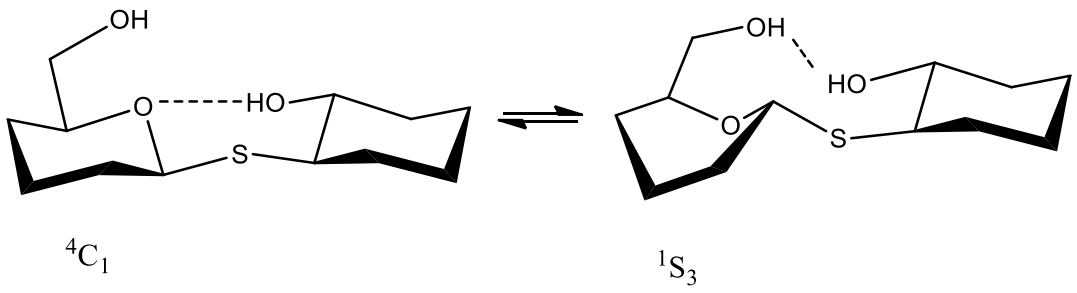


Figure 5.12. Schematic representation of hydrogen bonds between the -1 and the +1 subsite sugars in S- and O-linked disaccharide. These interactions depend on the conformation that is adopted by the -1 subsite sugar.

4.6. Conformational study of the HvExoI/laminaribiose (1,3-linked disaccharide) complex

The resulting conformational free energy landscape of the β-glucosyl ring in the -1 subsite of the G3O model is shown in Figure 5.13. The results show a unique stable conformation around θ = 270º, ϕ = 20º, whose conformation was the expected $^{4}C_{1}$. The system evolved to the distorted $^{4}H_{3}$ / E_{3} / $^{1}S_{3}$ region but, there is no stable minimum in this region. The error in the explored region is lower than 1.2 kcal/mol (Tiwary, 2015, Appendix C). Due to the fact that normally, sugars explore two stable minima inside active site pockets, we tried to simulate a hypothetical case in which the -1 sugar was in the $^{1}S_{3}$ conformation. However, the system returned quickly to the undistorted $^{4}C_{1}$ conformation.

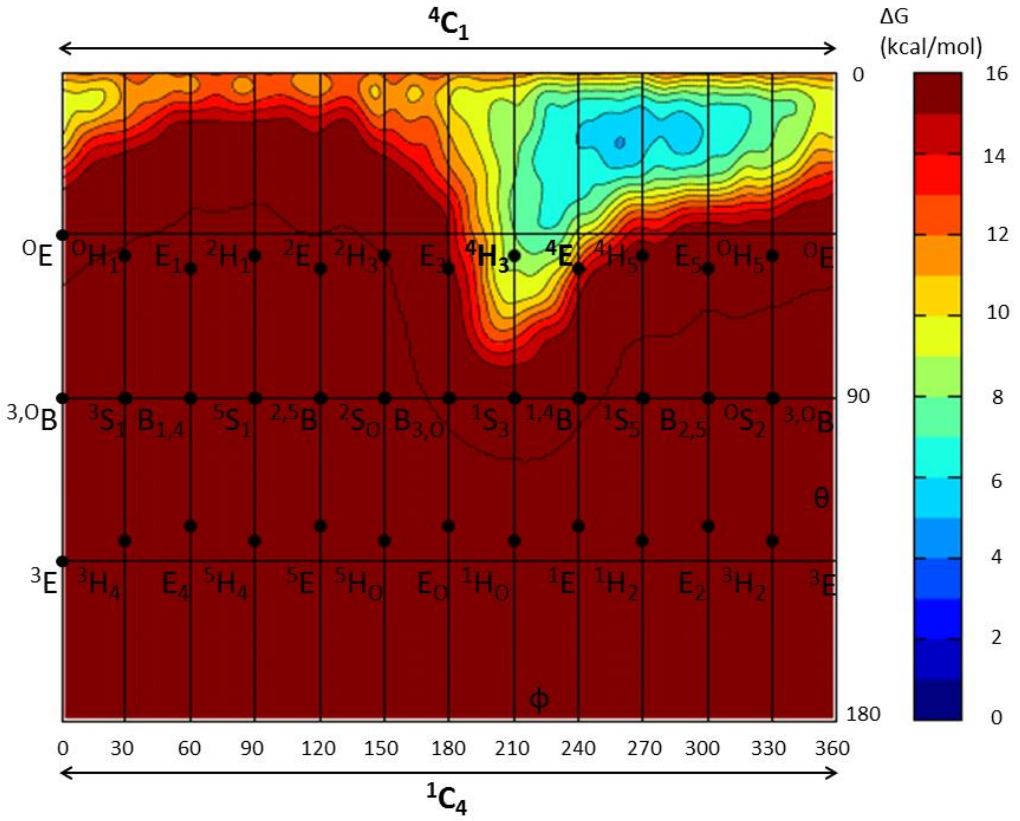


Figure 5.13. Conformational free energy landscape (FEL) of the -1 sugar ring in the G3O complex. Contour lines at 1 kcal/mol.

5. Discussion

Two types of conformational behaviour were observed when comparing simulations with GxS and GxO structures. On one hand, the thio-compound simulations showed that the sugar in the -1 subsite tends to be in the $^4C_1$ chair conformation, but it also explores the $^1S_3$ and the ^4E / $^4H_3$ distorted regions. On the other hand, the natural substrate simulations showed that the -1 subsite sugar samples uniquely the $^4C_1$ conformation. However, retaining exo-β-glucosidases typically distort the sugar to be hydrolysed (Davies, 2012). Why the -1 subsite sugar was not distorted in this case?

In all cases, we observed an interaction between the pyranic oxygen of the -1 subsite sugar and a hydroxyl group of the leaving group that stabilizes the $^4C_1$ chair conformation. Furthermore, due to sugar is not in a pure chair conformation ($\theta = 0.0$, $\phi = 0.0$), the slightly distorted $^4C_1$ conformation ($\theta = 270.0$, $\phi = 20.0$) of the -1 sugar is enough to avoid the steric effect of the hydrogen of the anomeric carbon over the nucleophilic attack. Assuming that the substrate is not distorted in the Michaelis complex, the conformational itinerary of HvExoI would be cyclic ($^4C_1 \rightarrow {}^4E \rightarrow {}^4C_1$), as observed for amylosucrase in Chapter III of the present Thesis.

In the case of the thio-compounds, the sulphur glycosidic linkages have longer bond distances (1.87 Å versus 1.46 Å) and more closed angles (104º versus 109º) than those in the oxygen counterparts. This increases the distance between the sugar rings at the -1 and the +1 subsites, and leads to weaker interactions with the active site residues that no longer stabilize the $^4C_1$ chair conformation of natural substrates. For this reason, the -1 subsite sugar of thio-derivatives could be distorted, unlike that of natural sugars.

Specific structural features of the residues delineating the -1 and +1 subsites are crucial to understand the stereochemical promiscuity of HvExoI (it is able to hydrolyse 1,2-, 1,3-, 1,4- and 1,6- linked sugars). On one hand, Glu491 and Asp285, acting as nucleophile and acid/base catalysts, respectively, ensure strong interactions with the 2-OH group. Other residues, such as His207, interact with the 3-OH, and Asp95 makes contacts with the 4-OH and the 6-OH groups of -1 subsite sugars. On the other hand, Trp286 and 434 make a "sandwich" type of interactions with the +1 subsite sugar, due to π-stacking interaction between the indolyl rings of tryptophan residues and the α- and β-faces of a glucosyl moiety. This allows an optimum spatial docking of β-glucoses with different stereochemistry, which underlie a broad substrate specificity of HvExoI.

6. Conclusions

The main conclusions of the present chapter are the following:

1. Classical MD simulation methods reproduce the experimental structure of complexes of HvExoI with thio-derivatives. It is thus expected that the same method can be applied over HvExoI complexes with natural substrates.
2. Both thio-sugar compounds and natural sugar substrates exhibit a $^4C_1$ chair conformation as the most populated one, but only the thio-compounds are able to distort their rings in the -1 subsite due to geometrical features of the thio-glycosidic bonds. This explains the conformations observed for the G3S and the G4S thio-sugars in the crystallographic structures (^4E and $^4H_3$, respectively).
3. The conformational study of the β-glucosyl group in the -1 subsite of the HvExoI confirmed that the $^4C_1$ chair conformation is the most stable, suggesting a cyclic itinerary for the hydrolysis reaction ($^4C_1 \rightarrow {}^4E \rightarrow {}^4C_1$).

Chapter VI – Computational insights into the conformational dynamics and reactivity of septanosides in glycoside hydrolase

Computational insights into the conformational dynamics and reactivity of septanosides in glycoside hydrolase

1. Introduction

The development of structural mimics of sugars is one of the most common strategies for the design for glycoside hydrolase inhibitors (Gruner, 2002) (Ganem, 1996) (Asano, 2009). As mentioned in Chapter I, Section 4.6, isofagomine and mannoimidazole are examples of 6-membered ring molecules used as mimics of the transition state of the reaction catalysed by GHs. In the last years, a new set of artificial sugars was synthesized (Bozó, 2000) (DeMatteo, 2005) (Saha, 2011) (Vannam, 2013, 2016) in order to test their capabilities as substrates/inhibitors of GHs (Tauss, 2006) by expanding the ring size (Peczuh, 2003). The ring-expanded versions of hexoses are the so-called septanoses, i.e. 7-membered ring molecules.

There are several computational works analysing the conformational richness of these structures. The simplest septanose-ring molecule – cycloheptane C_7H_{14} – was studied using a parametrized potential function (Bocian, 1975) and a coupled-cluster method (Freeman, 2008). On the one hand, Bocian et al. found a symmetrical potential energy landscape with 14 minima over the TC/C conformational space, where the TC was the most stable conformation and the C was 1.30 kcal/mol less stable. The barrier between TC/C plane and TB/B plane was 9.67 kcal/mol and TB and B conformations were 3.39 and 3.42 kcal/mol less stable than TC, respectively. On the other hand, Freeman et al. found the same relative stability between TC and TB (3.39 kcal/mol), and the pseudorotational barrier was 8.55 kcal/mol.

Furthermore, Peczuh et al. synthesized and analysed the conformational behaviour of the methyl α-D-glycero-D-idoseptanoside and the methyl β-D-glycero-D-guloseptanoside using potential energy calculations and MM MD methods (DeMatteo, 2005). Later on, a recognition study of a septanoside by the Concavalin A (Castro, 2005) and a binding analysis (Duff, 2011) were reported, followed by a kinetic study of the hydrolytic activity of α- and β-glucosidases acting over α- and β-L-idoseptanosides (Tauss, 2006). These works show that, in the case of α-glucosidase, the catalytic constant

(k_{cat}) with β-L-idoseptanosides decreases two orders of magnitude with respect to the natural substrate and one order of magnitude in terms of enzyme-substrate affinity (K_M).

In this chapter, we consider the Cremer and Pople puckering coordinates (Cremer, 1975) for 7-membered rings (φ_2, φ_3, q_2, q_3, see Appendix A-2) to describe the conformational space of septanoside molecules as potential substrates of GHs. The four puckering coordinates were programmed (Appendix A-3) for the Well-T MTD code (Barducci, 2008) in the Plumed 2 software.

To validate the Cremer-Pople implementation, we took the simplest 7-membered ring, cycloheptane, and compared the computed conformational free energy landscape with the experimental and theoretical information available. In a second step, the method will be applied to reconstruct the conformational free energy landscapes of a group of septanoside molecules: α-D-glycero-D-idoseptanoside (α-D-idosepta), β-D-glycero-D-guloseptanoside (β-D-gulosepta), β-L-idoseptanoside (β-L-idosepta) and the oxocarbenium septanoside cation ($[C_6H_{11}O]^+$) (Figure 6.1). Our results for α-D-idosepta and β-D-gulosepta will be compared to those of DeMatteo et al. (DeMatteo, 2005). The conformational FEL of the cation will provide the possible conformations adopted by a septanoside in the TS of the reaction. In addition, analysis of the simulations will allow us to determine the degree of preactivation of every conformation for catalysis. Our simulations are aimed at answering the following two questions: Why L-septanosides are not good substrates for GHs? Could D-septanosides be suitable substrates for GHs?

To answer the above questions, we obtained the conformational FEL of the pNP – β-L-idosepta molecule in a *Beta vulgaris* GH31 α-glucosidase (PDB 3W38 [Tagami, 2013]). This enzyme structure was selected because it has been obtained without mutating any of the catalytic residues, thus we minimize the possible errors of docking the septanoside substrates. It is a well-studied monomeric GH, with an intermediate resolution [2.79 Å], which shows a well-oriented pattern of catalytic residues. At the time of stating this study (2014) GH31 α-glucosidase seemed a very good candidate to test the potential of septanosides as GH substrates. Nowadays, together with the GH13 α-glucosidase, GH31 enzyme remains being a good structure for MD simulations. Using a combination of docking, classical and QM/MM MD and metadynamics, we obtained the conformational FEL of the β-L-idosepta ring in the -1 subsite and modeled the glycosylation reaction.

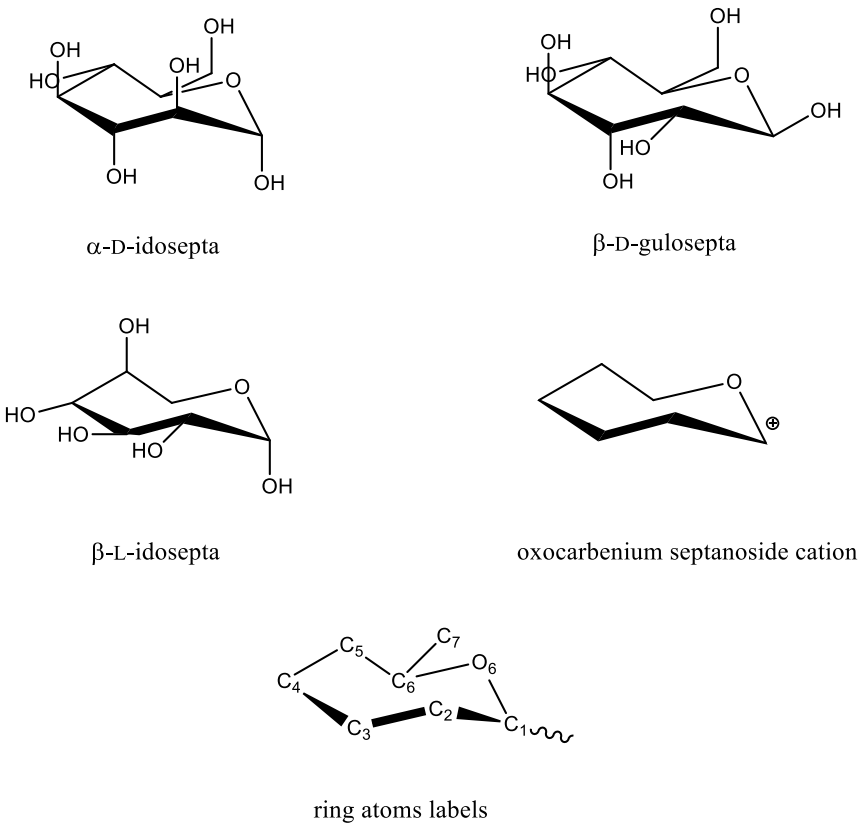


Figure 6.1. Representation of the α-D-glycero-D-idoseptanoside (α-D-idosepta), β-D-glycero-D-guloseptanoside (β-D-gulosepta), β-L-idoseptanoside (β-L-idosepta), oxocarbenium septanoside cation structures and the labelling of the ring atoms.

In a second step, we tried to find a pattern between 6-membered and 7-membered ring conformations using the preactivated shapes found for α-D-idosepta, β-D-gulosepta and β-L-idosepta molecules to foresee their possible catalytic activity in glucosidases, mannosidases and xylanases. Molecular engineering was applied to reconstruct GH-septanoside complexes in α- (*Halomonas sp. H11* GH13 PDB 3WY4 [Shen, 2015]) and β-glucosidases (*Oryza sativa* GH1 PDB 3GNP [Seshadri, 2009]). Finally, the glycosylation reaction was simulated in the GH13 – α-D-gulosepta-1,4-α-D-glucopyranose complex (GH13-α-gulosepta system) to decipher the molecular mechanism and elucidate the conformational itinerary of the sugar during the reaction.

2. Objectives

The main objectives of the present chapter are:

1) To obtain the conformational FEL of cycloheptane in the gas phase, comparing the results with previously reported studies based on potential energy analyses.
2) To obtain the conformational FEL of α-D-idosepta, β-D-gulosepta and β-L-idosepta in the gas phase and calculate the preactivation index for each conformation/system.
3) To obtain the conformational FEL of an isolated oxocarbenium septanoside cation and predict possible conformational itineraries of a 7-membered septanoside substrate hydrolysed by a GH.
4) To analyse the interactions between 1-para-nitrophenyl – β-L-idosepta in the active site of GH31 α-glucosidase, obtaining the conformational FEL of the sugar at the -1 subsite.
5) To simulate the glycosylation step of the hydrolysis reaction of the pNP - β-L-idosepta in GH31 α-glucosidase.
6) To reconstruct GH-septanoside complexes with appropriated sugar conformations to test the behaviour of septanosides as substrates of glucosidases and mannosidases.
7) To simulate the glycosylation step of the hydrolysis reaction of α-D-gulosepta-1,4-α-D-glucopyranose in GH13 α-glucosidase.

3. Computational details

Isolated molecules calculations

Previous calculations of the Intrinsic Reaction Coordinate (IRC) (Kui, 1981) of the pseudorotational change of cycloheptane using PBE/6-31G*, PBE/cc-PVDZ, PBE/cc-PVTZ were run with Gaussian 09 software to observe the effect of the basis set. These optimized structures were recomputed with PBE/PW (70 Ry) and the CPMD software to calculate the entropic contribution in the conformational behaviour of cycloheptane.

Gas phase ab initio metadynamics simulations of cycloheptane, α-D-idosepta, β-D-gulosepta, β-L-idosepta and oxocarbenium septanoside cation, at 300 K, were performed using the Car-Parrinello method (Car, 1985), as implemented in the CPMD 3.15.1

software patched with the Plumed 2 libraries (Barducci, 2008). The summarized details are shown in the Appendix F. The common aspects of all the simulations are the following: the electronic structure was computed within the DFT, using the PBE exchange-correlation functional (Perdew, 1996). Kohn-Sham orbitals were expanded in a PW basis set with a kinetic energy cutoff of 70 Ry and a convergence criteria of 10^{-7} a.u. in the electronic gradient. Ab initio pseudopotentials were employed, generated in the Martins-Troullier scheme (Troullier, 1991). The fictitious electronic mass (E_{mass}) and the time step (τ_{step}) chosen for each system were 850 a.m.u and 0.12 fs, respectively. The convergence criteria for the structure optimization was a maximum component of the nuclear forces of $5 \cdot 10^{-5}$ a.u. The collective variables were the four 7-membered ring puckering coordinates (φ_2, φ_3, q_2, q_3). In Appendix D, the effect of the dimensionality reduction in the conformational study of the cycloheptane in gas phase is studied. Our calculations using 4 CVs show better accuracy and less computational cost than the simulations with a 3CVs reduced space or a toroidal coordinates set.

Enzyme-substrate modelling and classical MD simulations

The model of the GH31 α-glucosidase and the pNP-β-L-idosepta was constructed from a docking simulation of the substrate in a $^{4,3}TC_{2,1}$ conformation using the AutoDock Vina software (Trott, 2010). All the rotatable bonds of the ligand were activated. The grid box were centred in the point (-2.0, -12.0, -24.0) with a cubic shape of 30.0 x 30.0 x 30.0 $(Å)^3$. The best structure obtained had a predicted binding energy of -6.67 kcal/mol. Comparing the conformation with its hexose analogous, we can see that it is very similar to the $^4C_1$ chair, adopted by the natural substrate (α-glucose) in α-glucosidases.

Both enzyme-substrate models were solvated in a TIP3P water molecules box (Jorgensen, 1983), seven (GH31) and thirty-one (GH13) sodium cations added to compensate the initial charge and the Asp511 (GH31) and the Glu271 (GH13, acid/base residue) were protonated. Both substrates were parametrized using a HF (Hartree, 1928) (Fock, 1930)/6-31G* (Ditchfield, 1971) optimization and RESP charges (calculated using the antechamber software) (Case, 2006). The rest of the system was parametrized with the AMBER 11 constants for the FF99 force-field (Cornell, 1995). First, the waters and sugar were minimized during 40000 simulation steps and then, the whole system was minimized during 70000 simulation steps. Afterwards, the system was smoothly heated to 200 K increasing the temperature by 50 K in 100 ps simulations and to 300 K increasing the temperature by 25 K in 100 ps simulations. The density is optimized to a realistic

value of 1.06 g/cm^3 in a 100 ps NPT simulation. To stabilize the RMSD of the backbone, 10 ns of NVT simulation at 300 K for the GH31-L-septa and 2 ns for the GH13-agulo system were needed.

QM/MM MD and MTD simulations

From a stabilized structure of the GH31 complex, CPMD QM/MM simulations were performed. In the first simulation, the QM box is formed by the whole sugar (pNP – β-L-idosepta, 36 atoms). The objective is to calculate the conformational FEL of the septanoside in the -1 subsite to find the most stable conformation to simulate the glycosylation reaction in the enzyme. Once obtained this conformation, a frame of a representative structure of this conformation is chosen to start the second MTD simulation. The new QM box is formed by the whole sugar, the Asp511 (acid/base) cut in the C_α and Asp412 (nucleophile) cut in the C_α using frontier carbon pseudopotentials (50 atoms, Figure 6.2).

The collective variables of the conformational study (Well-T. MTD) inside the GH31 enzyme are the four puckering coordinates (φ_2, φ_3, q_2, q_3). The collective variables of the glycosylation step (Direct MTD) inside the enzyme are two differences of distances (Figure 6.2) mimicking the proton transfer (CV2) and the nucleophilic substitution (CV1). Computational details in Appendix F.

From a stabilized structure of the GH13 complex (where the septanoside is in a $^{5,6}TC_{O,1}$ conformation), we have simulated the glycosylation step. The QM box is formed by the whole sugar (septanoside [-1] and glucose [+1]), the Glu271 (acid/base) cut in the C_β and Asp202 (nucleophile) cut in the C_α using frontier carbon pseudopotentials (64 atoms, Figure 6.3). Both QM/MM models are described with the same GGA functional, the same convergence criteria, the same PW cutoff and pseudopotentials (except frontier atoms) that gas phase calculations. E_{mass} and τ_{step} are 600 a.m.u and 0.12 fs.

The collective variables of the glycosylation step (Direct MTD) are two distances (CV1 and CV2) and a difference of distances (CV3 equivalent to the CV2 of the previous simulation). The first is the distance between the anomeric carbon (C_1) and the center of mass of both nucleophile oxygens. The second is the glycosidic bond distance.

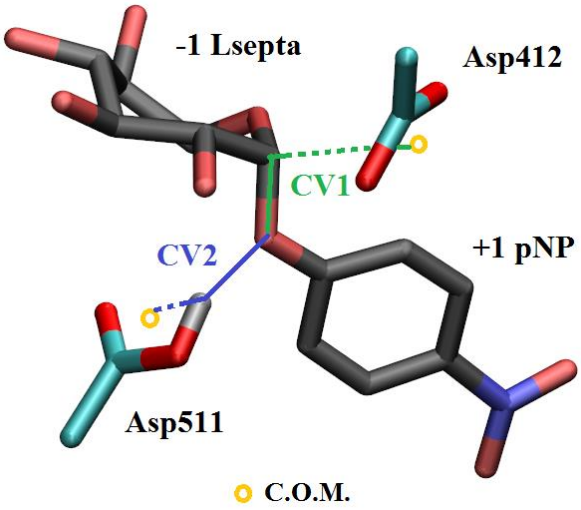


Figure 6.2. Atoms included in the QM box of the glycosylation step simulation in the GH31-L-septa complex. The color lines indicate the collective variable used in the MTD calculation.

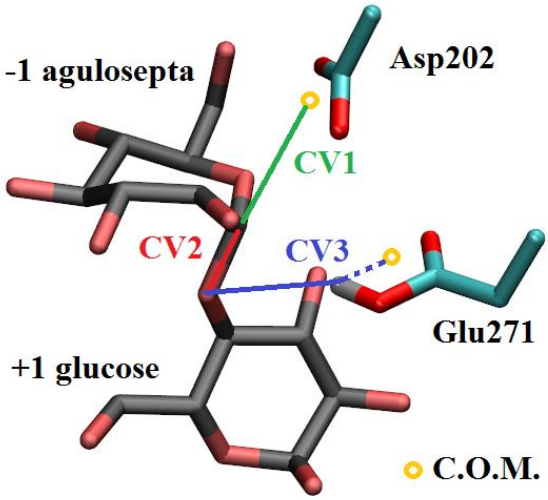


Figure 6.3. Atoms included in the QM box of the glycosylation step simulation in the GH13-α-gulosepta complex. The color lines indicate the collective variable used in the MTD calculation.

4. Results

All conformational studies were analysed with a 5D matrix analysis program (Appendix A – 4), from which minima and saddle points in the FELs were located. The projections (φ_2, φ_3) of the three conformational planes of 7-membered rings, the atomic structures of the local minima obtained, the collective variables evolution of each simulation, the preactivation parameters and the RMSD of the enzyme-substrate systems are presented in Appendix E.

4.1. Conformational study of cycloheptane in the gas phase

The conformational FEL of cycloheptane, reconstructed from the Well-T. MTD simulation using φ_2, φ_3, q_2, q_3 as CVs was projected in planes for the sake of visualization clarity. In Figure 6.5, both (φ_2, φ_3) projections at plane q_3 = 0.65 (TC/C plane) and q_3 = 0.05 (TB/B plane) are presented, respectively. The first surface is quite symmetric, showing the most stable 14 minima corresponding with the 14 possible TC (Figure 6.4a) conformations adopted by a 7-membered ring (Table 6.1). The free energy describing these states has a standard deviation of 0.46 kcal/mol. The transition states between minima correspond with the 14 possible chair (C) conformations, which are 1.12 ± 0.38 kcal/mol less stable than twisted chair (TC) conformers.

The second projection shows an irregular surface where the TB (Figure 6.4b) and B conformations are energetically indistinguishable, being 3.43 ± 0.25 kcal/mol less stable than TC conformations. The energy differences between conformations in the surface are below the error of the method (~0.6 kcal/mol). The average energy to go from TC plane to TB plane is – <ΔG(TC)> of the TC minima, being 4.66 kcal/mol.

a) b)

Figure 6.4. A representative structure of: a) the TC conformation and b) the TB conformation of cycloheptane (without hydrogens).

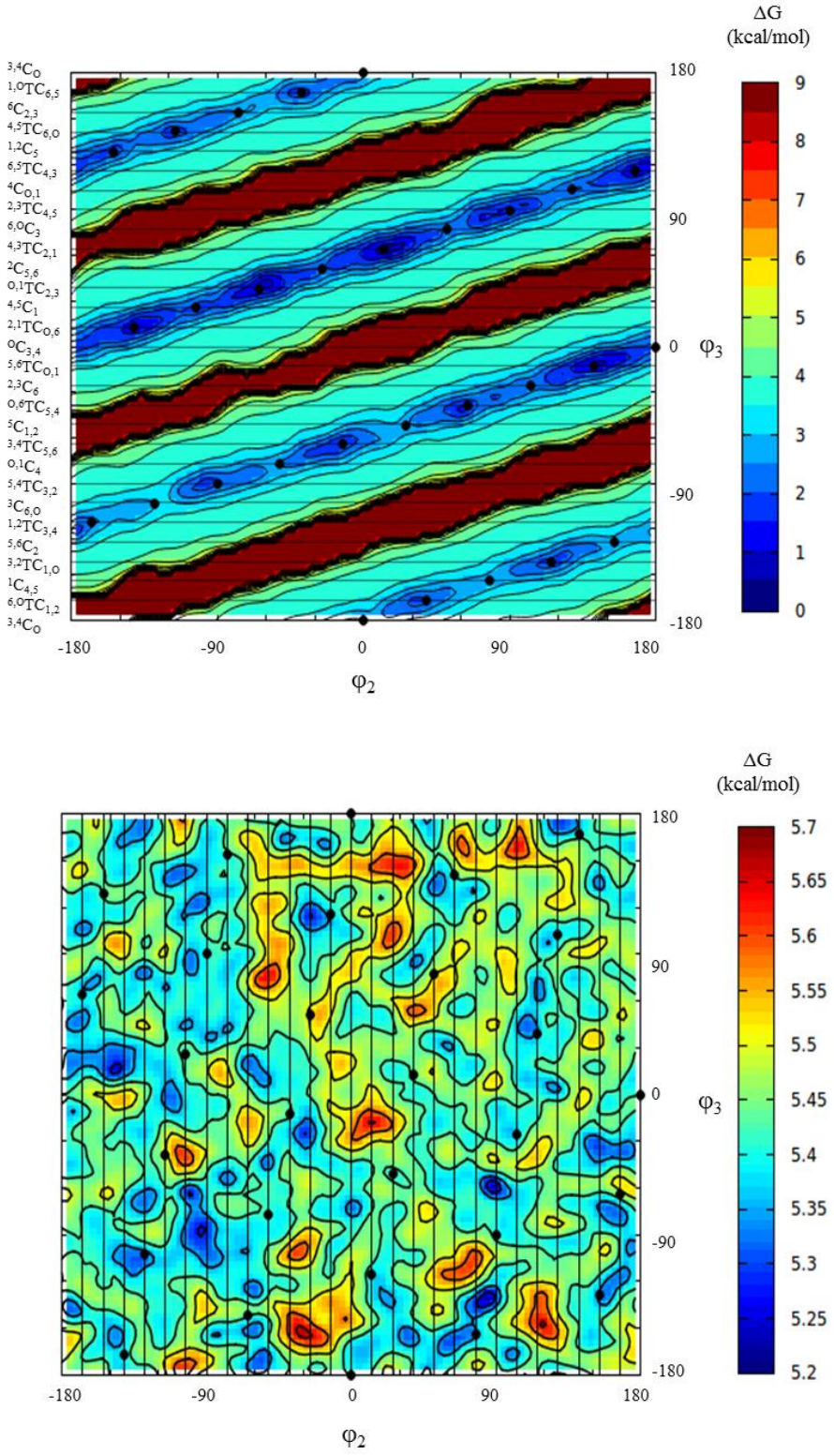


Figure 6.5. (φ_2, φ_3) projections of (top) TC/C plane (Isoline per 0.5 kcal/mol) and (bottom) TB/B plane (isolines at 0.1 kcal/mol) of the conformational FEL of cycloheptane.

Table 6.1. Analysis of the minima corresponding to the twisted-chair regions of the FEL of the cycloheptane molecule in the gas phase (φ in degrees, q in Å and energies in kcal/mol).

Minimum	φ_2	φ_3	q_2	q_3	MTD ΔG	Rel. ΔG	Conf.
1	33	-169	0.55	0.63	-4.11	1.33	$^{6,0}TC_{1,2}$
2	123	-140	0.53	0.63	-4.14	1.31	$^{3,2}TC_{1,0}$
3	-176	-123	0.50	0.65	-4.31	1.13	$^{1,2}TC_{3,4}$
4	-97	-90	0.53	0.63	-4.47	0.98	$^{5,4}TC_{3,2}$
5	-18	-65	0.55	0.63	-4.52	0.93	$^{3,4}TC_{5,6}$
6	61	-40	0.55	0.63	-4.73	0.72	$^{0,6}TC_{5,4}$
7	148	-11	0.51	0.63	-5.14	0.31	$^{5,6}TC_{0,1}$
8	-137	14	0.51	0.63	-5.14	0.30	$^{2,1}TC_{0,6}$
9	-61	43	0.53	0.63	-5.30	0.14	$^{0,1}TC_{2,3}$
10	14	65	0.55	0.63	-5.44	0.00	$^{4,3}TC_{2,1}$
11	76	87	0.51	0.63	-4.65	0.79	$^{2,3}TC_{4,5}$
12	166	115	0.53	0.65	-4.83	0.62	$^{6,5}TC_{4,3}$
13	-119	140	0.55	0.62	-4.30	1.15	$^{4,5}TC_{6,O}$
14	-36	166	0.55	0.63	-4.10	1.35	$^{1,0}TC_{6,5}$
				<ΔG(TC)>=	-4.66	0.79	
				St. Dev. =	0.46	0.46	

Comparing our results with the published by Bocian et al. (Bocian, 1975) and Freeman et al. (Freeman, 2008), the presence of 14 symmetrical TC conformers and the relative stability between TC, C, TB and B conformers are in good agreement with their results. Furthermore, the kinetic barrier between TC and TB (4.66 kcal/mol) is lower than in Bocian results (9.67 kcal/mol), fact due to the temperature effect (entropy), so, thermal vibrations make easier the conformational transition.

4.2. Conformational study of α-D-glycero-D-idoseptanoside in the gas phase

The conformational FEL of α-D-idosepta molecule in the gas phase shows 14 relevant minima in both TC/C and TB/B planes (Table 6.2). The most stable conformation is the $^{3,4}TC_{5,6}$, followed by $^{1,2}TC_{3,4}$ (2.25-2.27 kcal/mol), $^{1,O}TB_{6,5}/B_{2,(5,6)}$ (5.52 kcal/mol), $B_{2,(5,6)}$(5.70-7.03 kcal/mol), $^{5,6}TC_{O,1}$ (6.50 kcal/mol), $^{O,1}TC_{2,3}$ (distorted **[d]**, 8.21 kcal/mol), $^{3,4}C_O$ (d, 8.33 kcal/mol), $^{4,3}TC_{2,1}$ (9.08 kcal/mol) and $^{2,1}TC_{O,6}$ (planar **[p]**, 9.32

kcal/mol). The system also explored the transition regions corresponding with the $^{O,1}S_3$, $^{3}S_{5,6}$, $^{O,6}S_4$, $^{1}S_{3,4}$, $^{5}S_{O,1}$ and $^{6,O}S_2$ conformations (projection $q_3 = 0.35$).

Table 6.2. Analysis of the most important minima over the FEL of the α-D-idosepta molecule in gas phase (φ in degrees, q in Å, energies in kcal/mol and N_i/N is the population of the *i*th minimum).

Min.	φ_2	φ_3	q_2	q_3	MTD ΔG	Rel. ΔG	N_i/N (%)	Conformation
1	-11	-65	0.55	0.67	-11.14	0.00	95.66	$^{3,4}TC_{5,6}$
2	-158	-115	0.57	0.65	-8.88	2.25	2.18	$^{1,2}TC_{3,4}$
3	-176	-119	0.55	0.65	-8.87	2.27	2.14	$^{1,2}TC_{3,4}$
4	-29	-47	1.14	0.08	-5.61	5.52	9.06E-03	$^{1,O}TB_{6,5}/B_{2,(5,6)}$
5	-29	11	1.14	0.08	-5.44	5.70	6.76E-03	$B_{2,(5,6)}$
6	-25	108	1.16	0.07	-4.66	6.48	1.83E-03	$B_{2,(5,6)}$
7	144	-11	0.49	0.65	-4.63	6.50	1.75E-03	$^{5,6}TC_{O,1}$
8	-25	169	1.16	0.07	-4.19	6.94	8.37E-04	$B_{2,(5,6)}$
9	-29	-101	1.16	0.05	-4.17	6.96	8.07E-04	$B_{2,(5,6)}$
10	-25	-148	1.14	0.05	-4.11	7.03	7.24E-04	$B_{2,(5,6)}$
11	-47	43	0.57	0.61	-2.93	8.21	1.00E-04	$^{O,1}TC_{2,3}$ (d)
12	22	-166	0.38	0.68	-2.81	8.33	8.19E-05	$^{3,4}C_O$ (d)
13	7	61	0.49	0.61	-2.05	9.08	2.31E-05	$^{4,3}TC_{2,1}$
14	-130	7	0.43	0.65	-1.82	9.32	1.56E-05	$^{2,1}TC_{O,6}$ (p)

To know the predisposition of a conformation to be a good candidate for catalysis in GHs (*conformational preactivation*), the most relevant structural and electronic parameters during the simulation were calculated. As shown in Section 4.2.1, the $C_1 - O_1$ distance ($d_{C1\text{-}O1}$) and its orientation (axial/equatorial, $\Omega_{C1\text{-}O1}$), the $C_1 - O_p$ distance ($d_{C1\text{-}O6}$) and the effective charge over these three atoms (q_{O1}, q_{C1}, q_{O6}) are key to know if a conformation is catalytically preactivated. Stability, given by the relative energy (ΔG_{rel}), is another parameter to be considered. As discussed in previous works of the group (Ardèvol, 2010), conclusions cannot be based on one parameter separately; all of them need to be considered together.

The average values of each parameter were calculated for every conformation from the structures in a in a box defined by $\varphi_2 \pm 0.1$, $\varphi_3 \pm 0.1$, $q_2 \pm 0.025$, $q_3 \pm 0.025$. For the

calculation of the preactivation index (ζ, Eq. VI – 3) coring functions that normalize the score of each parameter depending on positive factors (Eq. VI – 1) or negative factors (Eq. VI – 2) are needed s. We define an index ζ_j as the average of the scores for the *n* parameters (n = 7 in our case) for a given conformation *j*.

$$score(x_{i,j}) = \frac{x_{i,j} - x_{i,j}^{min}}{x_{i,j}^{max} - x_{i,j}^{min}} \cdot 100 \; for \; x_i = d_{C1-O1}, q_{C1}, q_{O6}, \Omega_{C1-O1} \quad (\boldsymbol{Eq.VI-1})$$

$$score(x_{i,j}) = \frac{x_{i,j}^{max} - x_{i,j}}{x_{i,j}^{max} - x_{i,j}^{min}} \cdot 100 \; for \; x_i = d_{C1-O6}, q_{O1}, q_{O6}, \Delta G_{rel} \quad (\boldsymbol{Eq.VI-2})$$

$$\zeta_j = \sum_j \frac{\boldsymbol{score(x_{i,j})}}{\boldsymbol{n}} \quad (\boldsymbol{Eq.VI-3})$$

Figure 6.6 represents the change of ζ_j with ring distortion. The most preactivated conformation is the more stable one, $^{3,4}TC_{5,6}$, with an index of 75 (Figure 6.6a). The $B_{2,(5,6)}$ (ζ = 59) and $^{5,6}TC_{O,1}$ (ζ = 57) conformations constitute a second group of interest. Meanwhile, all parameters in $^{3,4}TC_{5,6}$ have important contributions. $^{5,6}TC_{O,1}$ (Figure 6.7b) displays a short $C_1 - O_1$ bond and the charge in O_6 is too negative so the preactivation index is lower (57). In $B_{2,(5,6)}$, the orientation of the $C_1 - O_1$ bond (see Figure 6.7c) is equatorial and the charge of C_1 and O_1 are not good enough for catalysis.

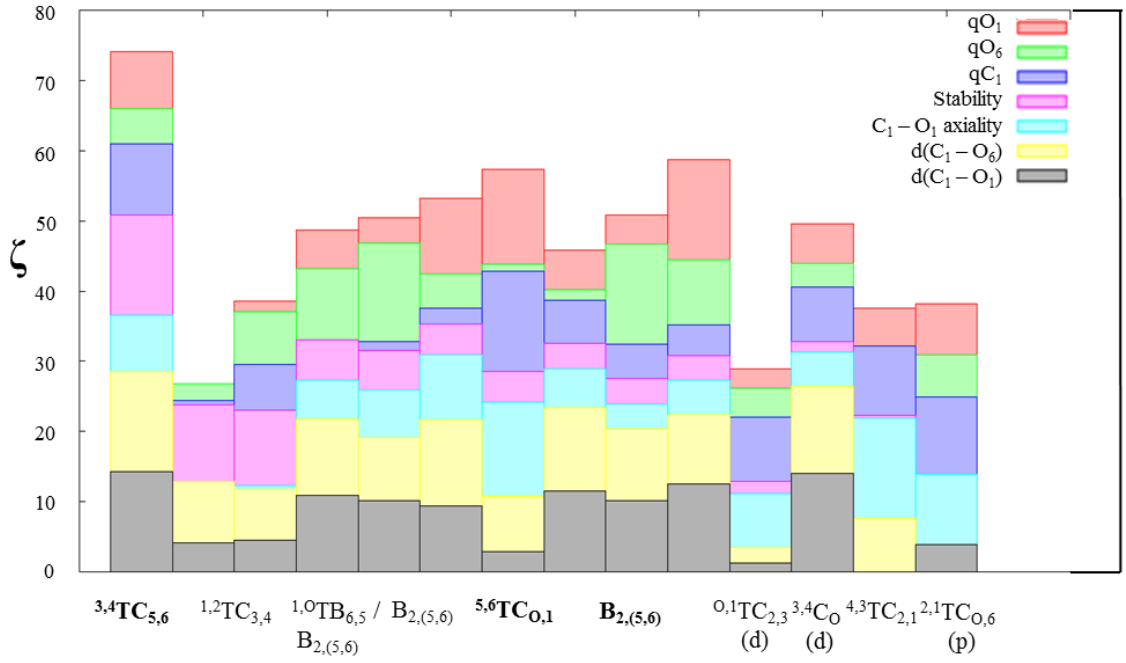


Figure 6.6. Colour decomposition of the score contribution of each preactivation parameter for every α-D-idosepta stable conformation.

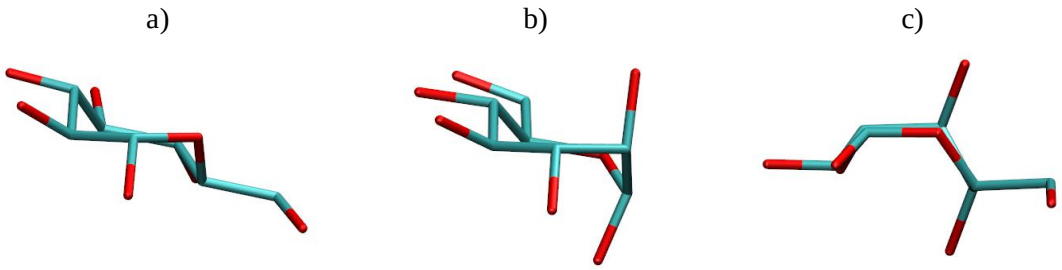


Figure 6.7. A representative structure of: a) the $^{3,4}TC_{5,6}$, b) the $^{5,6}TC_{O,1}$ and c) the $B_{2,(5,6)}$ conformations of α-D-idosepta (without hydrogens).

4.3. Conformational study of β-D-glycero-D-guloseptanoside in the gas phase

The conformational FEL of β-D-gulosepta molecule in the gas phase shows 15 relevant minima for both TC/C and TB/B planes (Table 6.3). The most stable conformation is $^{O,6}TC_{5,4}$ (Figure 6.9a), followed by $^{2}C_{5,6}$ (2.04-2.55 [d] kcal/mol), $^{1,2}TC_{3,4}$ (2.72 kcal/mol), $^{5,6}TC_{O,1}$ (2.74-2.75 [d] kcal/mol), $^{3,4}TC_{5,6}$ (d, 2.82 kcal/mol), $^{2,(5,6)}B$ (4.05-4.06 kcal/mol), $^{5,4}TC_{3,2}$ (4.06 kcal/mol), $^{O}C_{3,4}$ (4.77 [d] – 5.04 kcal/mol), $^{1,2}TB_{3,4}$ (6.81 kcal/mol), $^{O,6}TB_{5,4}$ (8.17 kcal/mol) and $^{3,4}TB_{5,6}$ (9.26 kcal/mol). The system has explored the transition regions corresponding with the $^{O,1}S_3$, $^{3}S_{5,6}$, $^{O,6}S_4$, $^{1}S_{3,4}$ and $^{5}S_{O,1}$ (Projection $q_3 = 0.35$).

Table 6.3. Analysis of the most important minima over the FEL of the β-D-gulosepta molecule in gas phase (φ in degrees, q in Å, energies in kcal/mol and N_i/N is the population of the *i*th minimum).

Min.	φ_2	φ_3	q_2	q_3	MTD ΔG	Rel. ΔG	N_i/N (%)	Conf.
1	58	-40	0.54	0.66	-10.27	0.00	91.78	$^{O,6}TC_{5,4}$
2	-29	43	0.48	0.64	-8.23	2.04	2.99	$^{2}C_{5,6}$
3	0	54	0.52	0.63	-7.72	2.55	1.27	$^{2}C_{5,6}$ (d)
4	-158	-112	0.48	0.66	-7.55	2.72	0.96	$^{1,2}TC_{3,4}$
5	140	-11	0.48	0.69	-7.53	2.74	0.92	$^{5,6}TC_{O,1}$
6	126	-14	0.52	0.69	-7.53	2.75	0.92	$^{5,6}TC_{O,1}$ (d)
7	-7	-58	0.52	0.66	-7.46	2.82	0.81	$^{3,4}TC_{5,6}$ (d)
8	155	-40	1.12	0.11	-6.22	4.05	0.10	$^{2,(5,6)}B$
9	-93	-87	0.48	0.63	-6.21	4.06	0.10	$^{5,4}TC_{3,2}$
10	155	-18	1.14	0.12	-6.21	4.06	0.10	$^{2,(5,6)}B$

11	-155	3	0.38	0.70	-5.50	4.77	0.03	$^{O}C_{3,4}$ (d)
12	-169	0	0.34	0.71	-5.25	5.03	0.02	$^{O}C_{3,4}$
13	-162	33	1.14	0.08	-3.47	6.81	1.01E-03	$^{1,2}TB_{3,4}$
14	47	-22	1.06	0.19	-2.10	8.17	1.03E-04	$^{O,6}TB_{5,4}$
15	-3	-72	1.04	0.18	-1.02	9.26	1.66E-05	$^{3,4}TB_{5,6}$

As in the α-D-idosepta case, the average value of each parameter was calculated for every conformation from the structures in a box defined by $\varphi_2 \pm 0.1$, $\varphi_3 \pm 0.1$, $q_2 \pm 0.025$, $q_3 \pm 0.025$. The preactivation index (ζ) for every conformation is shown in Figure 6.8. In this case, the most activated conformation is not the most stable. Two conformations share the same preactivation index: $^{2,(5,6)}B$ ($\zeta = 69$-66) and $^{5,4}TC_{3,2}$ ($\zeta = 68$). As previously found for α-D-idosepta, boat conformations (Figure 6.9c) show an equatorial orientation of the $C_1 - O_1$ bond. In this case, the "best" conformation – all the parameters contribute maximally to the preactivation index– is $^{5,4}TC_{3,2}$ (Figure 6.9b). It is worth mentioning that $^{1,2}TC_{3,4}$ as the better axial behaviour of the $C_1 - O_1$ bond, but the length of this bond is practically inactivated and $C_1 - O_6$ is too long, causing a significant decrease of the preactivation index (53).

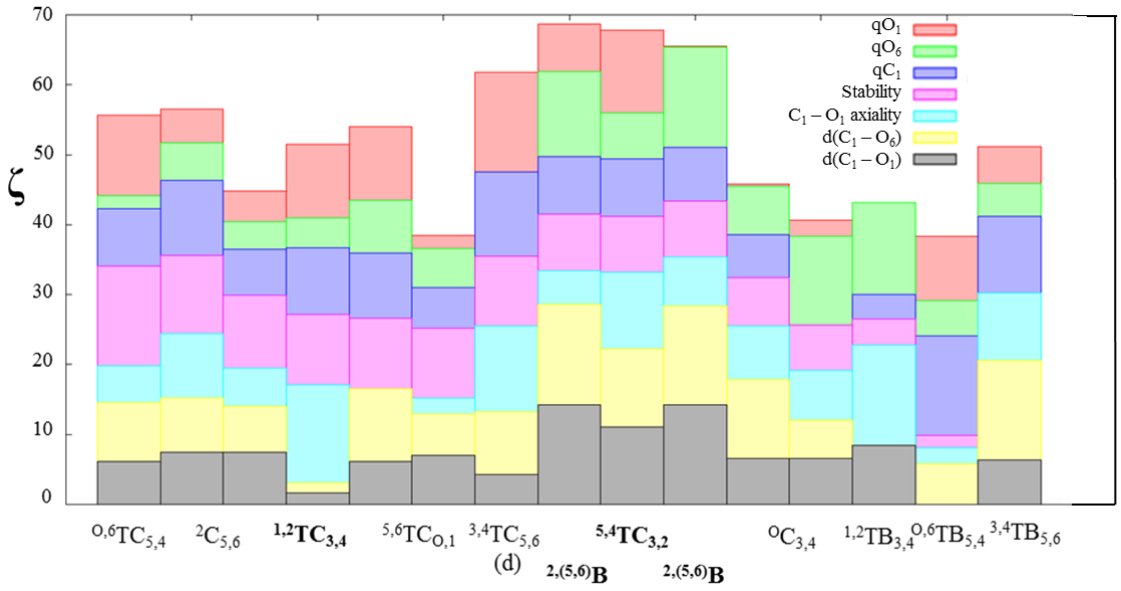


Figure 6.8. Colour decomposition of the score contribution of each preactivation parameter for every β-D-gulosepta stable conformation.

a) b) c)

Figure 6.9. A representative structure of: a) the $^{O,6}TC_{5,4}$, b) the $^{5,4}TC_{3,2}$ and c) the $^{2,(5,6)}B$ conformations of β-D-gulosepta (without hydrogens).

4.4. Conformational study of the oxocarbenium septanoside cation in the gas phase

The conformational FEL of oxocarbenium septanoside cation (oxoseptaion) in gas phase shows 17 relevant minima in TC/C, TS/S/BS and TB/B planes (Table 6.4). The TC/C plane is quite symmetric with respect to both most stable conformations ($^{4}C_{O,1}$ and $^{O,1}C_{4}$, Figure 6.10a), but the first is 1.17 kcal/mol more stable than the second. Following in order of stability are $^{3}TS_{4,5}$ (1.73 kcal/mol), $^{5,4}TS_{3}$ (3.10-3.31 kcal/mol), $^{3,2}TB_{1,O}$ (3.47 kcal/mol), $^{O,1}S_{3}$ (4.13 kcal/mol, , Figure 6.10b), $^{O,1}TB_{2,3}$ (4.30 kcal/mol), $^{3,4}BS$ (5.03 kcal/mol) and $BS_{3,4}$ (5.28 kcal/mol, Figure 6.10c). The system also explored the $^{2}S_{O,6}$, $^{O,6}S_{2}$, $^{3}S_{O,1}$, $^{4,5}BS$ and $BS_{4,5}$ conformations.

Table 6.4. Analysis of the most important minima of the FEL of the oxocarbenium septanoside cation molecule in gas phase (φ in degrees, q in Å, energies in kcal/mol and N_i/N is the population of the *i*th minimum).

Min.	φ_2	φ_3	q_2	q_3	MTD ΔG	Rel. ΔG	N_i/N (%)	Conf.
1	140	104	0.46	0.63	-6.38	0.00	56.52	$^{4}C_{O,1}$
2	119	104	0.31	0.62	-5.98	0.40	28.71	$^{4}C_{O,1}$
3	-58	-76	0.41	0.62	-5.21	1.17	7.96	$^{O,1}C_{4}$
4	58	-162	0.79	0.34	-4.65	1.73	3.10	$^{3}TS_{4,5}$
5	29	112	0.14	0.56	-4.19	2.19	1.43	$^{4}C_{O,1}$ (d)
6	0	108	0.20	0.54	-3.79	2.59	0.73	$^{4}C_{O,1}$ (d)
7	-137	14	0.84	0.34	-3.28	3.10	0.31	$^{5,4}TS_{3}$
8	-119	18	0.84	0.36	-3.07	3.31	0.22	$^{5,4}TS_{3}$
9	-101	108	0.10	0.59	-3.06	3.32	0.22	$^{4}C_{O,1}$ (d)
10	-137	-72	0.20	0.54	-3.04	3.33	0.21	$^{O,1}C_{4}$ (d)
11	115	-151	1.05	0.17	-2.91	3.47	0.17	$^{3,2}TB_{1,O}$
12	-123	108	0.10	0.56	-2.91	3.47	0.17	$^{4}C_{O,1}$

13	-166	-68	0.20	0.55	-2.80	3.58	0.14	$^{O,1}C_4$ (d)
14	-76	33	1.02	0.24	-2.25	4.13	0.06	$^{O,1}S_3$
15	-68	14	1.03	0.17	-2.08	4.30	0.04	$^{O,1}TB_{2,3}$
16	-18	123	0.52	0.43	-1.35	5.03	0.01	$^{3,4}BS$
17	162	-54	0.67	0.37	-1.10	5.28	8.09E-03	$BS_{3,4}$

Remarkably, this is the first case in which we can see that the distortion capability (d) and planarity (p) of a specific conformation is able to change the value of φ_2, maintaining the value of φ_3 (the TC/C conformations depend on the φ_3 coordinate) constant.

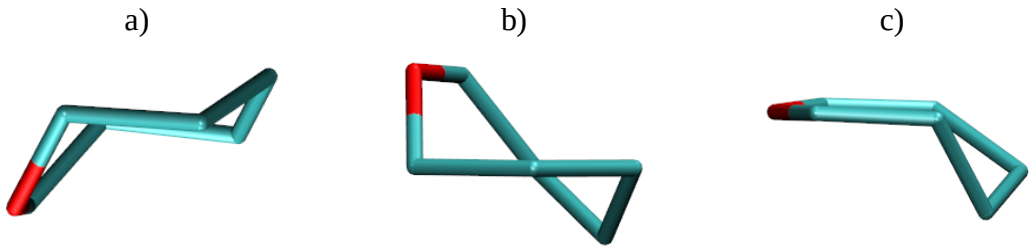


Figure 6.10. A representative structure of: a) the $^4C_{O,1}$, b) the $^{O,1}S_3$ and c) the $BS_{3,4}$ conformations of oxoseptaion (without hydrogens).

4.5. Conformational study of β-L-idoseptanoside in the gas phase

The conformational FEL of β-L-idosepta in the gas phase shows 18 relevant minima in both TC/C and TB/B planes (Table 7.5). The most stable conformation is $^{4,5}TC_{6,O}$, stability followed by $^{6,5}TC_{4,3}$ (d, 0.37 kcal/mol), $^{O,1}TC_{2,3}$ (0.38 kcal/mol), $^{3,4}TC_{5,6}$ (0.52 kcal/mol), $^{4,3}TC_{2,1}$ (1.20 kcal/mol), $^{4,5}C_1$ (1,45 kcal/mol), $^{2,3}TC_{4,5}$ (1.66 kcal/mol), $^6C_{2,3}$ (1.70 kcal/mol), $^4C_{O,1}$ (2.02 kcal/mol), $^{4,3}TC_{2,1}/^{6,O}C_3$ (2.26 kcal/mol), $^{5,6}C_2$ (2.80 kcal/mol), $^{1,2}TC_{3,4}$ (d, 3.18 kcal/mol), $^{O,1}TB_{2,3}$ (3.56 kcal/mol), $^{6,O}TC_{1,2}/^1C_{4,5}$ (p/d, 3.95 kcal/mol), $^{1,O}TB_{6,5}$ (4.42 kcal/mol), $^{5,6}TC_{O,1}$ (D, 5.12 kcal/mol), $^{(1,2),5}B$ (5.29 kcal/mol) and $^{2,1}TB_{O,6}$ (5.36 kcal/mol). The system has explored the transition regions corresponding with the $^{O,1}S_3$, $^4S_{6,O}$, $^2S_{O,6}$, $^3S_{5,6}$ and $^{6,O}S_2$ (Projection $q_3 = 0.35$).

Table 6.5. Analysis of the most important minima over the FEL of the β-L-idosepta molecule in gas phase (φ in degrees, q in Å, energies in kcal/mol and N_i/N is the population of the *i*th minimum).

Min.	φ_2	φ_3	q_2	q_3	MTD ΔG	Rel. ΔG	N_i/N (%)	Conf.
1	-123	140	0.56	0.65	-7.79	0.00	34.49	$^{4,5}TC_{6,O}$
2	175	117	0.53	0.66	-7.42	0.37	18.65	$^{6,5}TC_{4,3}$ (d)
3	-68	40	0.60	0.62	-7.41	0.38	18.14	$^{O,1}TC_{2,3}$
4	-22	-61	0.53	0.62	-7.27	0.52	14.35	$^{3,4}TC_{5,6}$
5	3	65	0.53	0.63	-6.59	1.20	4.60	$^{4,3}TC_{2,1}$
6	-112	22	0.53	0.65	-6.34	1.45	3.05	$^{4,5}C_1$
7	87	90	0.47	0.63	-6.13	1.66	2.12	$^{2,3}TC_{4,5}$
8	-61	162	0.43	0.69	-6.09	1.70	2.00	$^{6}C_{2,3}$
9	112	97	0.47	0.65	-5.77	2.02	1.17	$^{4}C_{O,1}$
10	29	68	0.47	0.63	-5.53	2.26	0.78	$^{4,3}TC_{2,1}/^{6,O}C_3$
11	155	-133	0.47	0.63	-4.99	2.80	0.32	$^{5,6}C_2$
12	-166	-123	0.56	0.61	-4.61	3.18	0.17	$^{1,2}TC_{3,4}$ (d)
13	-54	36	1.11	0.12	-4.23	3.56	0.09	$^{O,1}TB_{2,3}$
14	54	-169	0.41	0.68	-3.84	3.95	0.05	$^{6,O}TC_{1,2}$ (p) / $^{1}C_{4,5}$ (d)
15	-29	126	1.11	0.09	-3.37	4.42	0.02	$^{1,O}TB_{6,5}$
16	151	-7	0.49	0.61	-2.67	5.12	6.42E-03	$^{5,6}TC_{O,1}$ (d)
17	-158	108	1.13	0.10	-2.50	5.29	4.87E-03	$^{(1,2),5}B$
18	-148	162	1.08	0.10	-2.43	5.36	4.31E-03	$^{2,1}TB_{O,6}$

As in the D-septanoside simulations, the average value of each parameter was calculated for every conformation from the structures in a box defined by $\varphi_2 \pm 0.1$, $\varphi_3 \pm 0.1$, $q_2 \pm 0.025$, $q_3 \pm 0.025$. The preactivation index (ζ) for every conformation is shown in Figure 6.11 (the bold conformations in this figure are the most preactivated ones). Again, the most stable conformation ($^{4,5}TC_{6,O}$) is the less preactivated one (ζ = 33, Figure 6.12a). The most preactivated conformation is the $^{4}C_{O,1}$ (ζ = 78, Figure 6.12b), which shows high indexes in almost every parameter (thus this is the most "complete" conformation). In second place in the ranking we found $^{2,3}TC_{4,5}$. This conformation has a similar preactivation index (ζ = 75, Figure 6.12c), but the charge at O_6 is low (not suitable for catalysis) and this lowers the preactivation index. In third place, $^{4,5}C_1$ (ζ = 68) and $^{4,3}TC_{2,1}$ (ζ = 66) are two conformations to take into account to understand the

conformational behaviour of this molecule in enzymatic medium. In forth place, we found a distorted conformation $^{1,O}TB_{6,5}$ with a moderate preactivation index ($\zeta = 62$) due to the unfavourable orientation of the $C_1 - O_1$ bond and low C_1 charge. Finally, the $^{O,1}TC_{2,3}$ ($\zeta = 58$), whose role in GHs will be discussed in the following sections, shows the lowest negative charge at O_1 and the smallest (positive) charge at C_1 (not favourable for catalysis). In the case of $^{6,5}TC_{4,3}$ ($\zeta = 56$), distances are not activated and the charge at the intraring oxygen (O_6) is quite negative.

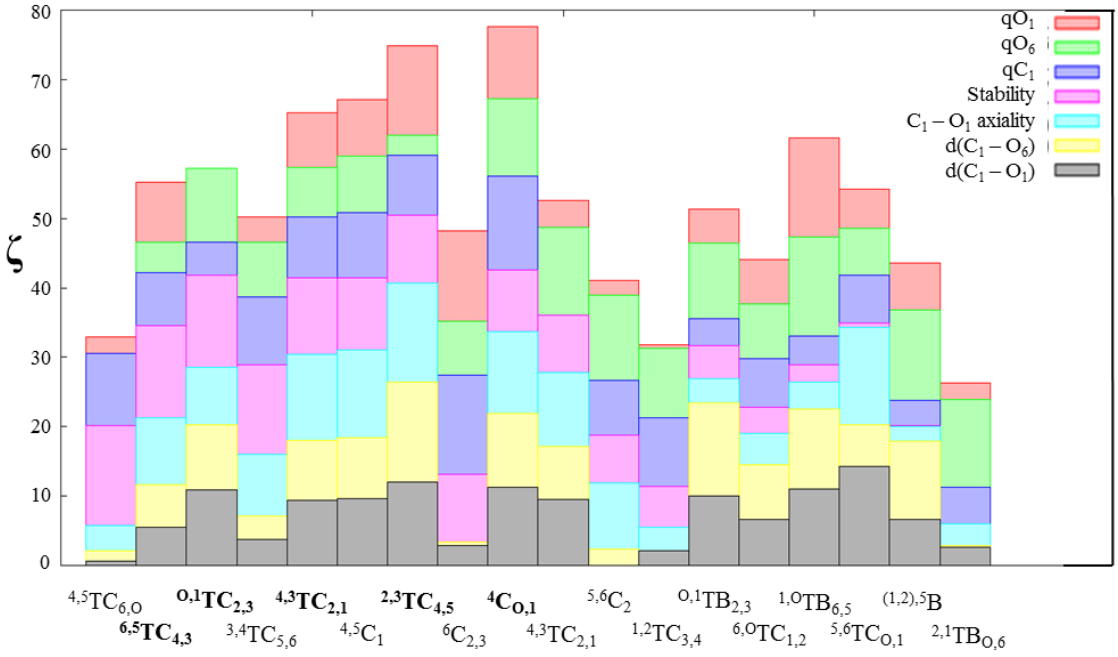


Figure 6.11. Colour decomposition of the score contribution of each preactivation parameter for every β-L-idosepta stable conformation.

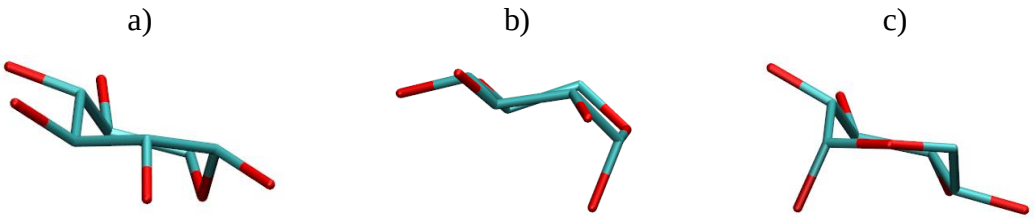


Figure 6.12. A representative structure of: a) the $^{4,5}TC_{6,O}$, b) the $^{4}C_{O,1}$ and c) the $^{2,3}TC_{4,5}$ conformations of β-L-idosepta (without hydrogens).

4.6. Docking and classical MD simulations of the complex of GH31 α-glucosidase with the pNP-β-L-idoseptanoside substrate

Among all possible conformations of β-L-idosepta, the best preactivated one and, at the same time, similar to the $^{4}C_{1}$ (Figure 6.13a) conformation of a hexose (which is the natural substrate in α-glucosidases), are $^{4,3}TC_{2,1}$ (ζ = 66, Figure 6.13b) and $^{4,5}C_{1}$ (ζ = 68, Figure 6.13c). Docking of pNP-β-L-idosepta with the sugar in a $^{4,3}TC_{2,1}$ conformation resulted in a pose with an optimum position and orientation of the catalytic residues (Figure 6.14 [left]). The nucleophile (Asp 412) was at 3.68 Å of C_1 and the acid/base (Asp511) oxygen was at 3.68 Å (2.68 Å from the hydroxyl H atom) of the O_g and its lone oxygen was interacting with the 2-OH substituent. Furthermore, the auxiliary Asp300 interacts well with the 4-OH and 5-OH substituents. The 3-OH interacted with the 4-OH substituent (auto-interaction).

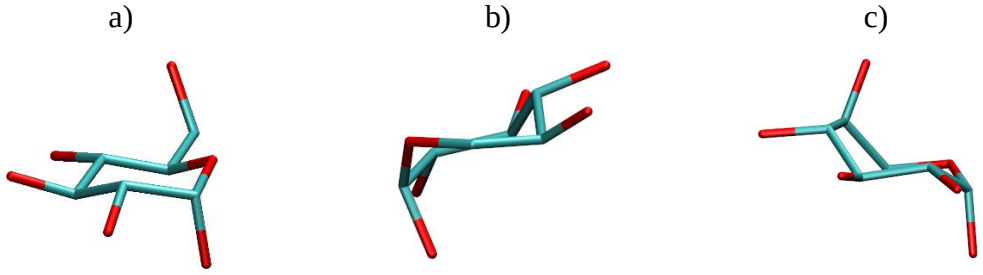


Figure 6.13. A representative structure of: a) the $^{4}C_{1}$ of α-D-glucose, b) the $^{4,3}TC_{2,1}$ and c) the $^{4,5}C_{1}$ conformations of β-L-idosepta (without hydrogens).

The sugar undergoes some conformational changes during the classical MD simulation. First, the conformation evolves from $^{4,3}TC_{2,1}$ (ζ = 66, Figure 6.14 [left]) to $^{6,5}TC_{4,3}$ (ζ = 56) during the minimization and remains during the heating (Figure 6.14 [center]). Second, the conformation evolves towards $^{O,1}TC_{2,3}$ (ζ = 58, Figure 6.14 [right]) during the equilibration and remains during the production phase at 300 K (NVT simulation). Simultaneously, the position of the nucleophile changes during the production phase due to the absence of the sugar 7-CH_2-OH "arm". In α-glucosidases, this group interacts with the nucleophile, keeping it in an appropriate position to attack the anomeric carbon. During the simulation, the nucleophilic oxygen remained near the anomeric carbon (~3.3 Å), but the other carboxylic oxygen moved to interact with the 2-OH sugar substituent. The RMSD of the protein backbone, the active site residues and sugar were stabilized around 2.2, 2.4 and 0.7 Å.

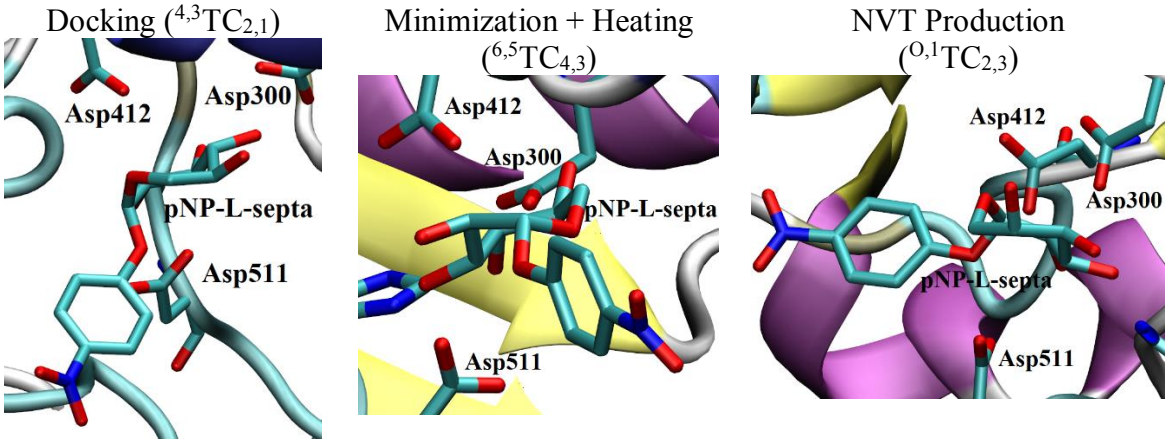


Figure 6.14. Atomic representations of the active site of the GH31 – β-L-idosepta model during the classical MD calculation, from the docking structure to the last frame of the stabilization at 300 K

4.7. Conformational study of the β-L-idoseptanosyl ring at the -1 subsite of GH31 α-glucosidase

The conformational FEL of the β-L-idoseptanoside at the -1 subsite of GH31 α-glucosidase shows 8 relevant minima in the TC/C, TS/S/BS and TB/B planes (Table 6.6). The most stable conformation is the $^{O,1}TC_{2,3}$, followed by $^{4,5}C_1$ (1.37 kcal/mol), $^{2,1}TB_{O,6}$ (5.79 kcal/mol), $^{1,2}C_5$ (8.87 kcal/mol), $^{(1,2),5}B$ (9.75 kcal/mol), $^{5,4}TS_3/^{4,5}BS$ (10.18 kcal/mol), $B_{3(6,O)}/^{5,4}TS_3$ (12.08 kcal/mol) and $^{5,4}TS_3$ (12.20 kcal/mol). The system also sampled the $^{O,1}S_3$ and $^{1,2}TS_3$ conformational space.

Table 6.6. Analysis of the most important minima over the FEL of the β-L-idosepta molecule in complex with GH31 α-glucosidase (φ in degrees, q in Å, energies in kcal/mol and N_i/N is the population of the *i*th minimum).

Min.	φ_2	φ_3	q_2	q_3	MTD ΔG	Rel. ΔG	N_i/N (%)	Conf.
1	-79	33	0.64	0.57	-14.75	0.00	90.84	$^{O,1}TC_{2,3}$
2	-101	29	0.63	0.59	-13.38	1.37	9.16	$^{4,5}C_1$
3	-144	144	1.04	0.15	-8.95	5.79	5.47E-03	$^{2,1}TB_{O,6}$
4	-148	133	0.64	0.56	-5.88	8.87	3.13E-05	$^{1,2}C_5$
5	-148	-144	1.04	0.06	-5.00	9.75	7.19E-06	$^{(1,2),5}B$
6	-115	7	0.84	0.29	-4.57	10.18	3.47E-06	$^{5,4}TS_3/^{4,5}BS$
7	-123	43	0.96	0.23	-2.67	12.08	1.44E-07	$B_{3(6,O)}/^{5,4}TS_3$
8	-112	47	0.95	0.27	-2.55	12.20	1.18E-07	$^{5,4}TS_3$

The above results show two possible catalytic pathways. The most probable is $^{O,1}TC_{2,3}/^{4,5}C_1 \rightarrow {}^{5,4}TS_3 \rightarrow B_{3(6,O)}$, since $^{O,1}TC_{2,3}$ is the most stable conformation. The less likely is $^{1,2}C_5 \rightarrow {}^{1,2}TS_3 \rightarrow {}^{(1,2),5}B$, where the $^{1,2}C_5$ is almost 9 kcal/mol less stable than the twisted-chair.

A priori, the -1 sugar adopts a $^{O,1}TC_{2,3}$ conformation ($\zeta = 58$, according to the FEL of an isolated β-L-idoseptanoside, see section 4.5). We hypothesized that it needs to change towards a more preactivated conformation such as $^{4,5}C_1$ ($\zeta = 68$). In the next section, we will show if this hypothesis is true or false.

4.8. Simulation of the glycosylation step in the GH31 – pNP-β-L-idoseptanoside complex

The glycosylation step FEL resulting from the MTD simulation is reported in Figure 6.15. The most relevant information extracted from the surface is the position of the reactants (R) and a oxocarbenium ion specie (OI) minima, as well as the transition state (TS). During the simulation, the system went beyond the OI region, when the leaving group went far from the anomeric carbon (C_1) of the septanoside in the -1 subsite. Then, four minima appeared, two (OIp) in the region where the proton was not transferred (the para-nitrophenyl is a good leaving group) and two other minima (OInp) in which the proton was transferred, leading to para-nitrophenol. At this point, we decided to stop the simulation, because the system was not going to return to the reactants minima.

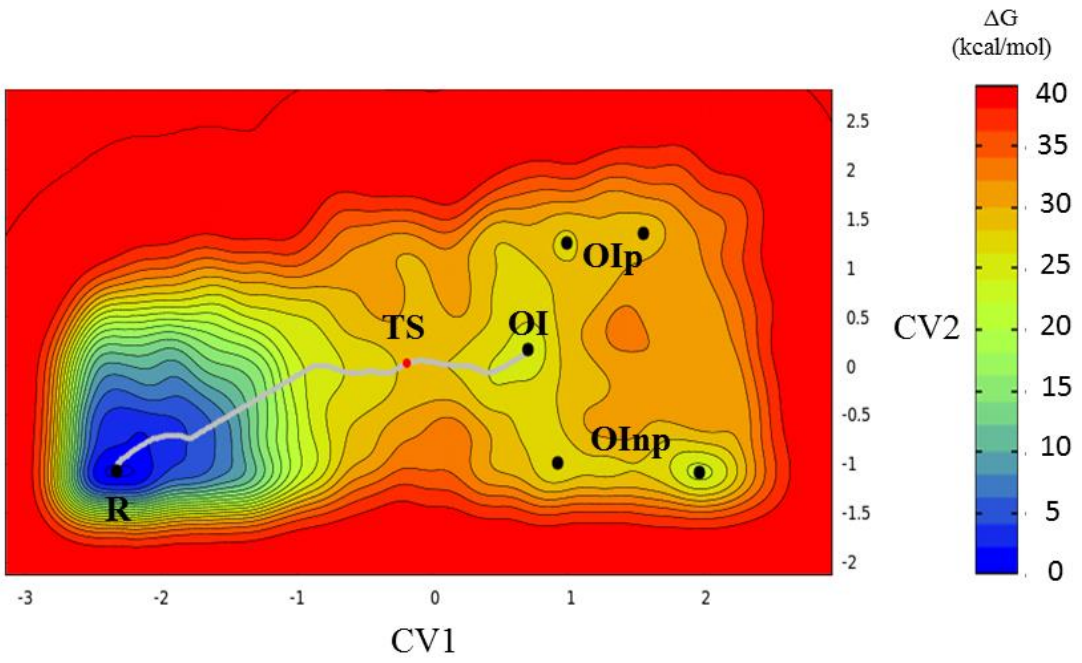


Figure 6.15. Representation of the FEL of the glycosylation step in the GH31 – pNP-β-L-septa complex (isolines every 2 kcal/mol). CV1 represents the nucleophilic attack and CV2 represents the acid/base proton transfer. The grey line represents the MFEP of the process.

The resulting activation energy ($\Delta G^{\ddagger}$) is 28.62 kcal/mol and the free energy barrier from reactants (R) to products (OI) is 24.99 kcal/mol. The nucleophile (Asp412) does not form a covalent intermediate with the septanoside, a high energy oxocarbenium ion state is explored over the FEL.

Table 6.7 lists the puckering, structural and energetic parameters calculated at every important point of the FEL, choosing frames in a box defined by CV1±0.1 Å and CV2±0.1. The results show a $^{O,1}TC_{2,3} \rightarrow {}^{O,1}S_3/{}^{5,4}TS_3 \rightarrow {}^{O,1}S_3/B_{(2,3),6}$ conformational itinerary. Representative atomic configurations of the R, TS, OI, OIp and OInp points of the FEL are presented in Figure 6.16.

Table 6.7. Analysis of the most important minima over the FEL of the glycosylation step of the pNP – β-L-idosepta in the GH31 enzyme (φ in degrees, q in Å, Y(C_2-C_1-O_6-C_6) in degrees and energies in kcal/mol).

FEL Point	φ_2	φ_3	q_2	q_3	Y(C_2-C_1-O_6-C_6)	MTD ΔG	Rel. ΔG	Conf.
R	-74	32	0.65	0.57	-64	-40.29	0.00	$^{O,1}TC_{2,3}$
TS	-91	37	0.81	0.38	-17	-11.67	28.62	$^{O,1}S_3/^{5,4}TS_3$
OI	-89	38	1.00	0.24	13	-15.3	24.99	$^{O,1}S_3$
OInp**	-86	44	1.01	0.25	10	-16.57	23.72	$^{O,1}S_3/B_{(2,3),6}$
OIp**	-93	42	0.95	0.29	10	-12.38	27.91	$^{O,1}S_3/^{4,5}TB_{6,O}$

** Not converged.

At the reactants state (R), the nucleophile is at 3.3 Å from the anomeric carbon and the acid/base is at 2.0 Å from the glycosidic oxygen. In the first step of the reaction, the system evolves to the transition state (TS) and the glycosidic bond is partially cleaved (2.2 Å), while the proton of the acid/base is shared between the acid/base and the glycosidic oxygen. The approach of the nucleophile is almost inappreciable (2.9 Å), thus it is a dissociative TS. The conformation evolved from $^{O,1}TC_{2,3}$ to $^{O,1}S_3/^{5,4}TS_3$. In the next step (OI), the glycosidic bond is already broken (2.9 Å), but the nucleophile has returned to the initial position (3.3 Å). The oxoseptaion adopts a $^{O,1}S_3$ conformation and seems to be stable and prepared for the approach of a water molecule (deglycosylation step). In the next steps, independently whether the leaving group is protonated (OIp) or not (OInp), there is a shy approach of the nucleophile to the anomeric carbon (2.8 Å) and the conformation of the oxoseptaion evolves to a $^{O,1}S_3/B_{(2,3),6}$ shape.

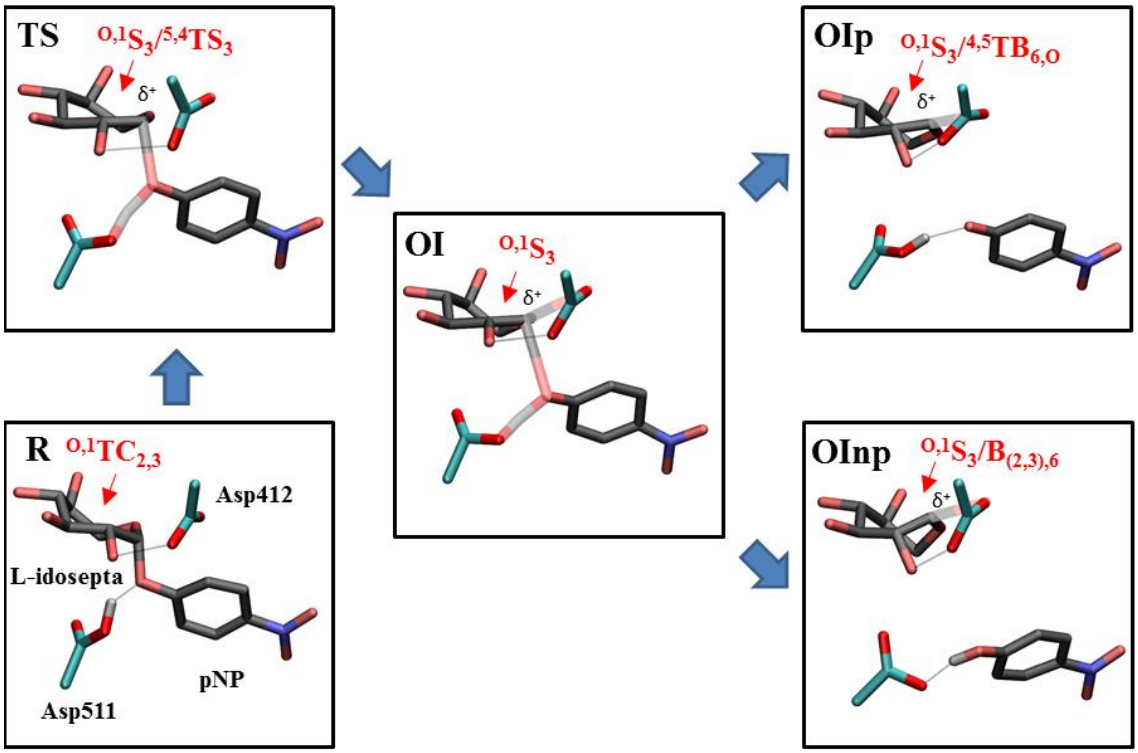


Figure 6.16. Representative structures of the minima and transition state corresponding to the glycosylation reaction of pNP – β-L-idosepta in the GH31 enzyme.

4.9. Docking and minimization of GH – D-septanoside complexes

In this section, we present the structures of septanoside oligosaccharides in complex with two selected glucosidases, obtained by docking and energy minimization (Figure 6.17):

(1) α-D-idosepta-(1,4)-α-glucose and α-D-gulosepta-(1,4)-α-glucose (septanose ring in the $^{5,6}TC_{O,1}$ [ζ = 55] conformation) in complex with GH13 α-glucosidase.

(2) β-D-gulosepta-(1,3)-β-glucose (septanoside ring in the $^{1,2}TC_{3,4}$ [ζ = 52] conformation) in complex with GH1 β-glucosidase.

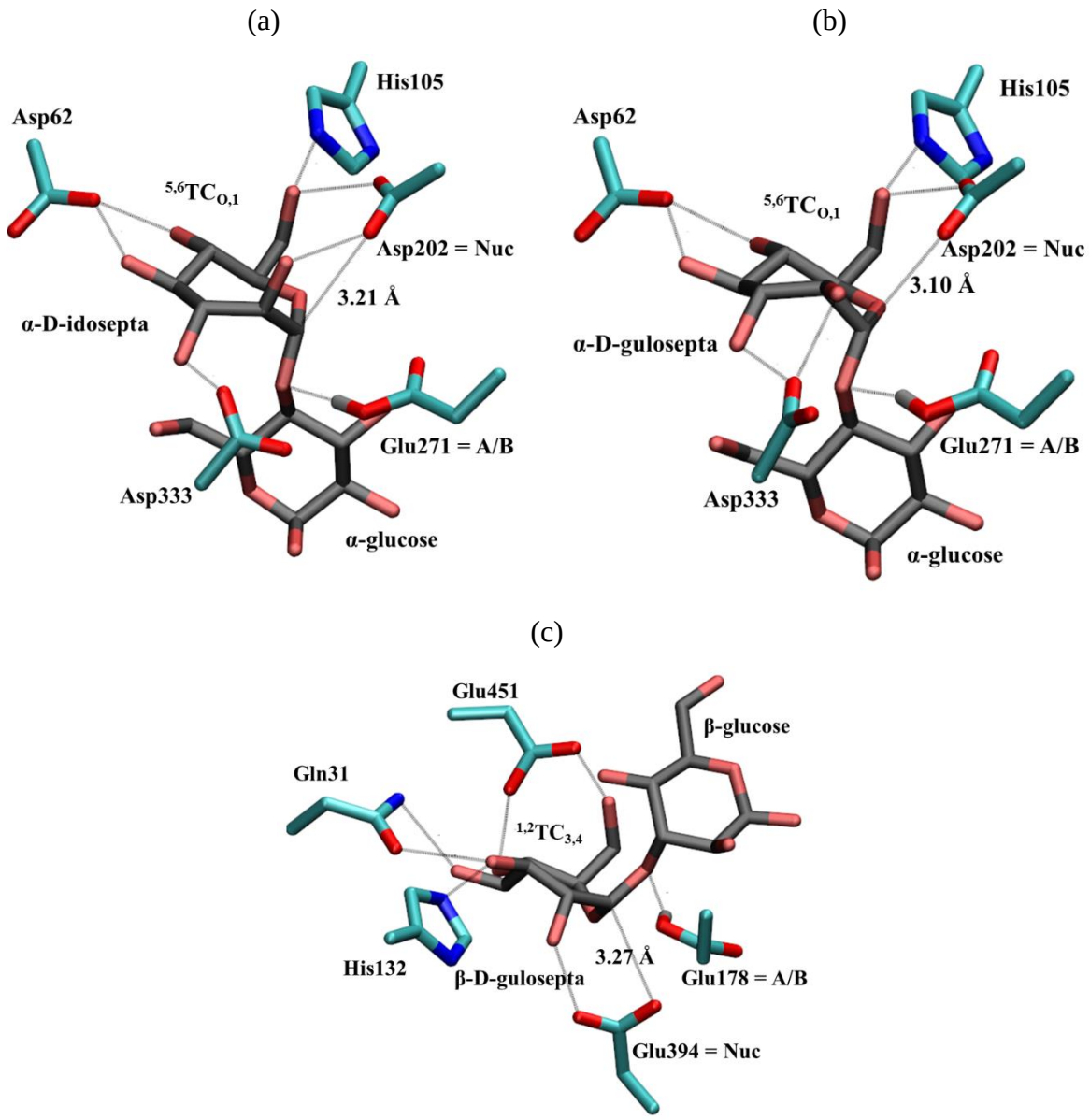


Figure 6.17. Minimized structures of: (a) α-D-idosepta-(1,4)-α-glucose and (b) α-D-gulosepta-(1,4)-α-glucose in GH13 α-glucosidase and (c) β-D-gulosepta-(1,3)-β-glucose in GH1 β-glucosidase.

4.10. Simulation of the glycosylation step of the GH13 – α-D-gulosepta-1,4-α-glucose complex

To analyse the structural changes observed along the Minimum Free Energy Pathway (MFEP), collective variable values, relevant catalytic distances and bonds and the puckering coordinates of the guloseptanosyl group were monitored during the simulation (Appendix E).

The system sampled three important regions: initially, in the reactants well, then, the variables have changed to the oxocarbenium ion region and finally, the covalent intermediate was formed. The convergence was accepted once the molecule returned from the enzyme-substrate intermediate to the reactants configuration.

The resulting glycosylation FEL and its 2D projections are shown in Figure 6.18. The surface presents four important minima in opposite regions: Reactants (R), Oxocarbenium Ion (OI^+) and two Covalent Intermediates (CI_1 and CI_2), being the reactants the most stable region by 3.35, 5.12 and 5.70 kcal/mol, respectively. During the simulation, the system evolved from R to CI, and then returned to the initial state, passing through a transition zone, with a TS being 12.20 kcal/mol less stable than the R minimum.

To describe the evolution of the conformation of the sugar along the reaction pathway, we performed a statistical analysis along the trajectory of the simulation choosing the structures in a box defined by CV1 ± 0.2, CV2 ± 0.2, CV3 ± 0.2 Å, along the minimum free energy profile. We represent the evolution of the CVs (Figure 6.19) and the puckering coordinate (φ_2, φ_3, q_3) of the septanoside ring along the internal reaction coordinate in Figure 6.20.

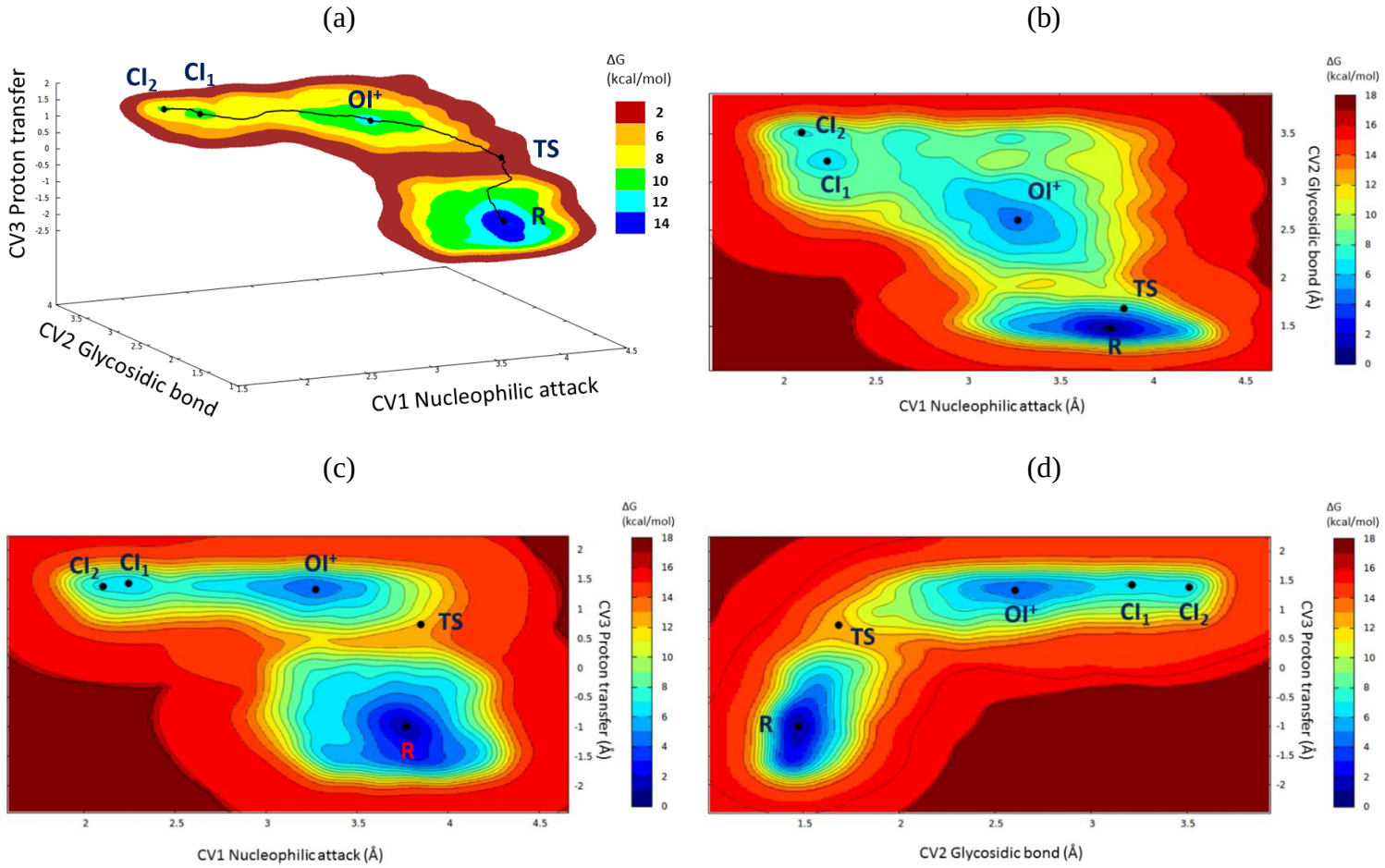


Figure 6.18. (a) Glycosylation FEL and (b, c, d) its 2D projections of the α-gulosepta within the GH31 α-glucosidase. Bold white line indicates the MFEP. Isolines at 1 kcal/mol.

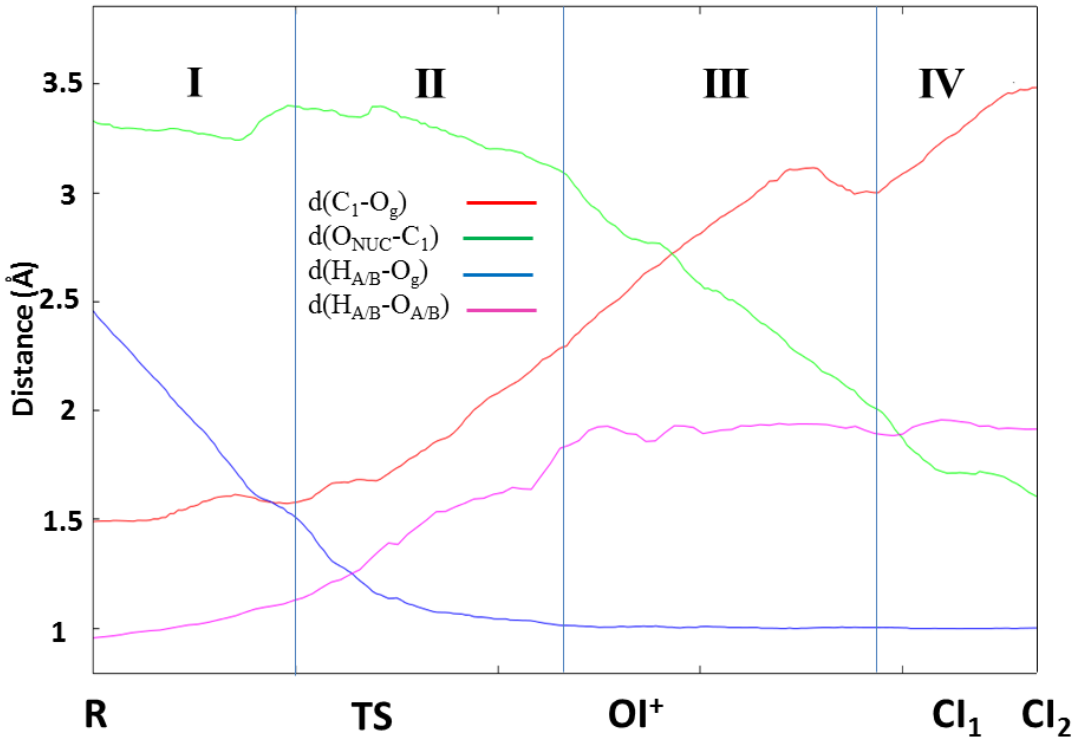


Figure 6.19. Representation of the evolution of the main catalytic distances along the glycosylation reaction coordinate.

Figure 6.19 shows the evolution of the main catalytic distances along the MFEP. The mechanism of the glycosylation reaction of the disaccharide can be split in four phases:

(I) *Approach of the Glu271 to the glycosidic oxygen*

In this initial phase, the distance between the proton of the Glu271 ($H_{A/B}$) and the glycosidic oxygen (O_g) evolves from 2.5 to 1.5 Å.

(II) *Proton transfer of the Glu271 to the O_g and formation of the oxocarbenium ion*

In this second phase, the d($H_{A/B}$-O_g) evolves from 1.5 to 1.0 Å (bond formation) and the d($H_{A/B}$-$O_{A/B}$) evolves (in a complementary way) from 1.0 to 1.9 Å (hydrogen bond interaction). The glycosidic bond evolves from 1.7 to 2.4 Å. The conformation of the sugar changes from the TC plane to a $BS_{3,4}$ boat-sofa.

(III) *Nucleophilic attack of the Asp202 and total cleavage of the glycosidic bond*

The glycosidic bond distance evolves from 2.4 to 3.1 Å (broken bond). The distance between the nucleophilic oxygen (O_{NUC}) of the Asp202 and the anomeric carbon (C_1) evolves from 3.2 to 2.1 Å.

(IV) *Formation of the covalent glycoside-enzyme intermediate (CI_1 and CI_2)*

Finally, the d(O_{NUC}-C_1) value evolves from 2.1 to 1.6-1.7 Å (bond formation). The conformation of the -1 septanoside collapses in the $BS_{3,4}/^{1}S_{3,4}/^{5,6}TB_{O,1}$ region.

Taking into account that the breaking of the glycosidic bond (Phase III) occurs before of the formation of the covalent glycosyl enzyme intermediate (Phase IV), the reaction follows a **dissociative** mechanism, as previously observed for GH13 amylosucrase (Chapter III).

Figure 6.20 shows the evolution of the conformation of the guloseptanosyl ring at the -1 subsite over the puckering reduced hyperspace (φ_2, φ_3, q_3). The ring follows a linear conformational itinerary, going through the $^{5,6}TC_{O,1} \rightarrow {}^{2}TS_{3,4}/BS_{3,4} \rightarrow BS_{3,4}/^{1}S_{3,4}/^{5,6}TB_{O,1}$ region. The conformational change seems very complex due to the nomenclature of the conformations. However, comparing these conformations with their 6-membered ring analogs, the resulting pathway is similar to the $^{4}C_1 \rightarrow {}^{4}H_3/E_3 \rightarrow E_3/^{1}S_3$, which is the expected linear pathway for α-glucosidases.

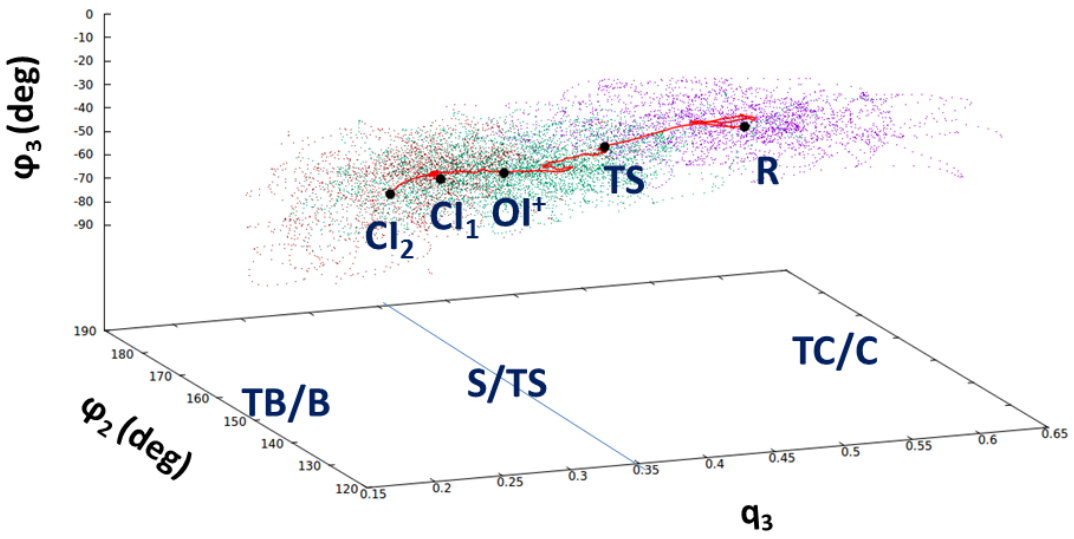


Figure 6.20. Representation of the conformational behaviour of the -1 septanosyl ring along the metadynamics reaction (purple, cyan and brown points) and the mean value of the reduced space puckering coordinates (φ_2, φ_3, q_3) over the minimum free energy pathway (MFEP – red line).

To understand the different conformational behaviour between this process and the cyclic itinerary of amylosucrase (both members of GH family 13), it is crucial to analyse the role of Thr203. While the nucleophile of AS interacts uniquely with the 6-OH of the glucose ring, the lone oxygen of Asp202 in GH13 α-glucosidase is restricted between the 7-OH of the sugar and the alcohol group of Thr203. This makes impossible the rotation of the nucleophile once the covalent intermediate is formed (Figure 6.21), precluding the return of the -1 sugar to the undistorted initial configuration.

Finally, Table 6.8 lists the mean values of the catalytic distances and puckering coordinates of the five relevant states over the FEL.

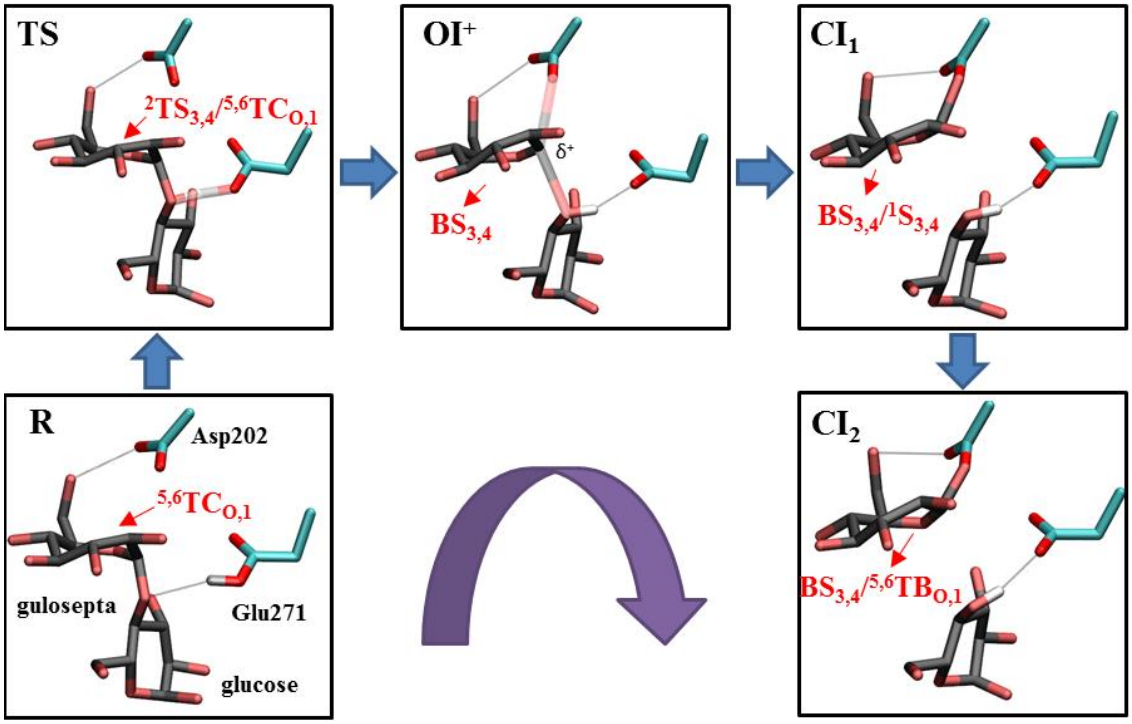


Figure 6.21. Atomic representation of the most relevant points over the glycosylation pathway.

Table 6.8. Analysis of the most important minima and transition state over the glycosylation FEL of the α-D-gulosepta-1,4-α-glucose in the GH13 enzyme (φ in degrees, q and distances in Å and energies in kcal/mol).

	ΔG	d(C_1-Onuc)	Glyc. Bond	d($H_{A/B}$-O_g)	
R	0.0	3.33 ± 0.13	1.49 ± 0.09	2.46 ± 0.12	
TS	12.2	3.38 ± 0.13	1.72 ± 0.11	1.14 ± 0.05	
***OI*$^+$**	3.4	2.77 ± 0.16	2.63 ± 0.12	1.01 ± 0.03	
***CI*$_1$**	5.1	1.72 ± 0.14	3.22 ± 0.12	1.00 ± 0.03	
***CI*$_2$**	5.7	1.61 ± 0.09	3.48 ± 0.09	1.00 ± 0.03	
	φ_2	**φ_3**	**q_2**	**q_3**	**Conformation**
R	145 ± 7	-22 ± 6	0.62 ± 0.066	0.50 ± 0.07	$^{5,6}TC_{O,1}$
TS	149 ± 5	-30 ± 8	0.63 ± 0.080	0.40 ± 0.08	$^{2}TS_{3,4}/^{5,6}TC_{O,1}$
***OI*$^+$**	155 ± 7	-44 ± 7	0.67 ± 0.062	0.35 ± 0.05	$BS_{3,4}$
***CI*$_1$**	166 ± 5	-57 ± 9	0.74 ± 0.10	0.33 ± 0.06	$BS_{3,4}/^{1}S_{3,4}$
***CI*$_2$**	167 ± 6	-62 ± 7	0.74 ± 0.10	0.29 ± 0.05	$BS_{3,4}/^{5,6}TB_{O,1}$

There are two important points to remark from Table 6.8 in order to understand the low kinetic barrier and the instability of the covalent intermediate with respect to the reactants. Taking a look to the mean value of the glycosidic bond in the reactants well, this value is 1.49 Å higher than the expected value of this bond in natural hexose oligosaccharides (~1.44 – 1.45 Å). This means that the enzyme environment activates the cleavage of this bond when the oligosaccharide is formed from our septanoside and the glucose leaving group. This could explain the increase of the catalytic activity of the enzyme.

Alternatively, the mean value of the enzyme-substrate bond formed in the CIs minima (around 1.6-1.7 Å) can be explained by the attraction of the Thr203 proton over the lone oxygen of the nucleophile. Returning to the AS case, the enzyme-glucose bond formed is around 1.46 Å. Then, the length of the septanoside-aspartate bond is shorter than the hexose-aspartate bond in GH13 amylosucrase. This could be the reason why the covalent intermediate species is higher in energy than the reactants.

Finally, as it was observed in the GH31-β-L-idosepta complex, the oxocarbenium ion is more stable than the covalent intermediate species. This could be an interesting point to be studied in future works to know what changes from 6 to 7-membered ring sugars.

5. Discussion

This section is separated in three main parts: Cycloheptane and general observations, why L-septanosides are not good substrates for GHs? And, can D-septanosides be good substrates for GHs?

5.1. Cycloheptane

Comparison with previous work

The first and the most important point of our study is to ensure the good behaviour of our methodology at time to explore the conformational space of 7-membered rings. For this reason, the comparison of our results with cycloheptane with other computational works is key to continue our investigation. In the Table 7.9, we present the relative energies of our MTD simulation, compared with the work of references (Bocian, 1975) and (Freeman, 2008), in addition to our Gaussian 09 IRC simulation of the pseudorotational change in cycloheptane with different basis sets.

Table 6.9. Comparison of different cycloheptane conformational studies: our results of the conformational MTD with PBE and PW at 300 K (MTD/PBE PW), results of references (Bocian, 1975) (Bocian FF) and (Freeman, 2008) (CCSD/cc-PVDZ), our results of Gaussian 09 calculations with PBE and different basis set and our results of potential energy with CPMD and PBE (PBE/PW). Energies in kcal/mol.

Conf.	MTD/PBE/ PW	Bocian FF	CCSD/ cc-PVDZ	PBE/ cc-PVDZ	PBE/ cc-PVTZ	PBE/ PW
TC	0.00 ± 0.46	0.00	0.00	0.00	0.00	0.00
C	1.12 ± 0.38	1.30	-	-	-	-
TB	3.43 ± 0.25	3.39	3.39	3.43	3.25	3.35
B	3.43 ± 0.25	3.42	-	-	-	-
TS	4.66 ± 0.46	9.67	8.55	8.47	7.93	8.03

In all cases, the TC conformation is the most stable, followed by the C conformation. Bocian et al. found an energy difference of 1.30 kcal/mol, while ours is 1.12 ± 0.38 kcal/mol. Then, while other authors find a difference between TB and TC in the range of {3.25, 3.43}, our result is 3.43 ± 0.25 kcal/mol. Here, our results are in very good agreement with other works. It is also worth to discuss the role of entropy, as our results

are free energy whole other authors report potential energy. This could have affected the transition state conformation between the TC/C and TB/B planes. While the value of the potential energy barrier is around {7.93-9.67} kcal/mol, our results at 300 K show an activation barrier of 4.66 kcal/mol. Taking into account the value obtained with the PBE/PW method (8.03 kcal/mol), the entropy of the conformational change in the cycloheptane at 300 K is 11.23 cal/mol·K. In 1956, Finke et al. measured experimentally the standard entropy of cycloheptane (g) (?) finding a value of 81.82 cal/mol·K (Finke, 1956). It is reasonable to think that the conformational degrees of freedom can be one eighth of the whole entropic contribution of the cycloheptane in gas phase.

Furthermore, to be sure that the results obtained for cycloheptane are extrapolated to septanoside sugars, the work of the reference (DeMatteo, 2005) with methyl septanosides can be compared with our simulations with hydrolysed septanosides (section XXX). In this work, the more stable conformation for the methyl α-D-septa and the methyl β-D-septa are, respectively, $^{3,4}TC_{5,6}$ (0.00 kcal/mol) and $^{1,2}TC_{3,4}$ (3.20 kcal/mol), on the one side, and $^{O,6}TC_{5,4}$ (0.00 kcal/mol), $^{5}C_{1,2}$ (2.02 kcal/mol), $^{5,6}TC_{O,1}$ (2.31 kcal/mol), $^{1,2}TC_{3,4}$ (3.46 kcal/mol) and $^{3,4}TC_{5,6}$ (4.47 kcal/mol), on the other side. Changing the methyl of the glycosydic oxygen for a proton, the conformational behaviour should be similar to our simulations. In the case of the α-D-septa molecule, the most important conformation is the $^{3,4}TC_{5,6}$ (0.00 kcal/mol) and the second is the $^{1,2}TC_{3,4}$ (2.25 kcal/mol). In the case of the β-D-septa, for the most populated conformational region, $^{O,6}TC_{5,4}$ (0.00 kcal/mol), we found a planar region around $^{5}C_{1,2}$ (2.50 kcal/mol) and minima around $^{5,6}TC_{O,1}$ (2.75 kcal/mol), $^{1,2}TC_{3,4}$ (2.72 kcal/mol) and $^{3,4}TC_{5,6}$ (2.82 kcal/mol). In both cases, our simulations are in agreement with the work of DeMatteo. Therefore we conclude that our method gives meaningful information.

General considerations

The amount of information for the conformations of cycloheptane is very large, thus we have tried to extract the most important features in a single graph, which we call a "conformational clock" of 28 hours/conformations per ring (Figure 6.16). The outer ring shows the conformations over the TC/C plane, with respect to their φ_3 values. The inner circle show the conformations over the TB/B plane, with respect to their φ_2 values (boat conformations are not shown for the sake of clarity). The central circle shows the S conformations and the TS/BS conformations are in the middle of the S points. All circles are connected between them by the pathway $^{O,1}TC_{2,3} \rightarrow {}^{O,1}S_3 \rightarrow {}^{O,1}TB_{2,3}$ (top-right). This

is because this is practically the reaction pathway found in the glycosylation simulation and it is the conformational region more frequented by all the simulated molecules.

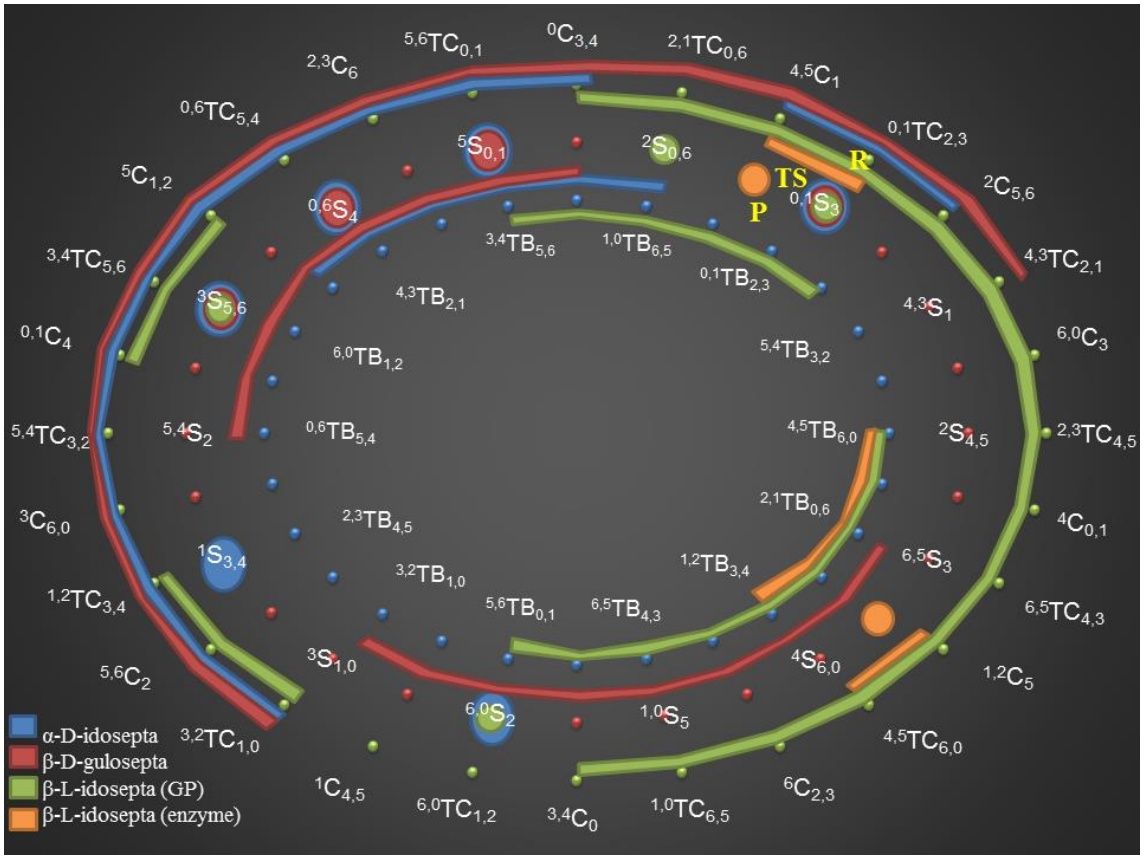


Figure 6.16. Conformational clock for 7-membered rings. Coloured lines indicate the most sampled regions for the studied septanoside sugars.

In this figure, two different patterns are observed, represented as two semicircles. On the one hand, the α-D-idosepta (blue) and the β-D-gulosepta (red) adopt conformations of the left-side semicircle. On the other hand, the β-L-idosepta (green) adopts conformations of the right-side semicircle. This behaviour is analogous to the hexose rings, where D-sugars prefers the skew-boats from the left-side of the Stoddart diagram. However, the β-L-idosepta is able to adopt conformations of the left-side (e.g. $^{3,4}TC_{5,6}$). This could be due to the absence of the steric hindrance of the 7-CH_2-OH arm, which restricts the conformational space available for the studied D-sugars.

Finally, returning to the simulation of the oxoseptaion conformations, we have uncovered all the possibilities available for the transition state of a septanose in a hydrolysis reaction. In Figure 6.17, all these possibilities and the logical/linear pathway

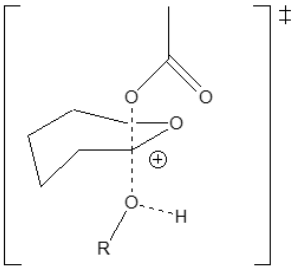
are shown. We have separated these conformations in three groups: Low Energy TS, Distorted TS and Complex Pathway TS. In hexoses, all the possible conformations of the oxocarbenium ion are distorted, however, in septanosides, the most stable conformations of the oxoseptaion are in the TC/C plane, the $^{4}C_{O,1}$ and the $^{O,1}C_{4}$ chairs, forming the first group. Chairs are transition states between two Twisted-Chair conformations. The second group are formed by Twisted-Sofa and Sofa conformations, knowing that the first one is transition state between a Chair and a Boat, and the second one is a transition state between a Twisted-Chair and a Twisted-Boat. Finally, Boat-Sofa conformations can be transition states between complex combinations of conformations from the contiguous regions, then, they are forming the third group.

Oxocarbenium ion like transition state conformations

Low Energy Transition States

$^{3,4}TC_{5,6} \rightleftharpoons [^{O,1}C_4]^{\ddagger} \rightleftharpoons {}^{5,4}TC_{3,2}$

$^{6,5}TC_{4,3} \rightleftharpoons [^{4}C_{O,1}]^{\ddagger} \rightleftharpoons {}^{2,3}TC_{4,5}$



Distorted Transition States

$^{1}C_{4,5} \rightleftharpoons [^{3}TS_{4,5}]^{\ddagger} \rightleftharpoons B_{1,(4,5)}$ $\quad$ $^{O,1}TC_{2,3} \rightleftharpoons [^{O,1}S_3]^{\ddagger} \rightleftharpoons {}^{O,1}TB_{2,3}$

$^{4,5}C_{1} \rightleftharpoons [^{5,4}TS_3]^{\ddagger} \rightleftharpoons {}^{1,(4,5)}B$ $\quad$ $^{3,2}TC_{1,O} \rightleftharpoons [^{3}S_{1,O}]^{\ddagger} \rightleftharpoons {}^{3,2}TB_{1,O}$

$^{2,1}TC_{O,6} \rightleftharpoons [^{2}S_{O,6}]^{\ddagger} \rightleftharpoons {}^{2,1}TB_{O,6}$ $\quad$ $B_{(2,3),6} \rightleftharpoons [^{O,1}TB_{2,3}]^{\ddagger} \rightleftharpoons {}^{(O,1),4}B$

$^{6,O}TC_{1,2} \rightleftharpoons [^{6,O}S_2]^{\ddagger} \rightleftharpoons {}^{6,O}TB_{1,2}$ $\quad$ $^{(2,3),6}B \rightleftharpoons [^{3,2}TB_{1,O}]^{\ddagger} \rightleftharpoons B_{(O,1),4}$

Complex Pathway Transition States

$[^{4,5}BS]^{\ddagger}$ $\quad$ $[BS_{4,5}]^{\ddagger}$ $\quad$ $[^{3,4}BS]^{\ddagger}$ $\quad$ $[BS_{3,4}]^{\ddagger}$

Figure 6.17. Scheme of the possible conformations that can be adopted by an oxocarbenium septanoside cation and the catalytic pathway followed by the substrate in a GH.

The existence of the Low Energy TS is interesting. As observed for β-L-idosepta, these conformations are specially preactivated for GH catalysis, thus if we are able to allocate the septanosides in conformations whose TS is one of this conformations, such septanoside could be a good substrate for GHs.

5.2. Why L-septanosides are not good substrates for GHs?

Tauss et al. (Tauss, 2006) observed a decrease of k_{cat} (i.e. increase of activation energy) and an increase of K_M (i.e. decrease of enzyme-substrate affinity) for pNP – L-idosepta in the α-glucosidase enzyme. During our classical MD simulation at 300 K, the nucleophile moves its position, due to the absence of 6-CH_2-OH arm of the β-L-idosepta ring. Following, the conformation of the sugar changes from an intermediate preactivated $^{4,3}TC_{2,1}$ conformation (ζ = 66) to a non-preactivated $^{O,1}TC_{2,3}$ conformation (ζ = 58). Because the nucleophile catalytic residue needs to change its position to stabilize its action over the sugar, the enzyme-substrate affinity decreases. Furthermore, considering the decrease on the preactivation parameter during the conformational change, the decrease of the k_{cat} could be conditioned by that fact.

Our simulation of the glycosylation step of the hydrolysis reaction in the GH31 overrides the predictions from analysis of the preactivation indexes. The calculated activation barrier is 28.92 kcal/mol and the mechanism goes through a high energy transition state – oxocarbenium septanoside cation is metastable – where the nucleophile does not form a covalent intermediate with the sugar. Normally, the simulation of a glycosylation step in a retaining GH has a barrier of {17-23} kcal/mol (Section 4.2. Glycosylation step in GH13 amylosucrase). Thus, the decrease of the catalytic activity for the septanoside substrate is significant.

However, there is hope for these "disarmed" L-septanosides. β-xylose is the paradigm of hexoses without the 5-CH_2-OH arm. In 2015, J. Iglesias-Fernández et al. found the GH11 β-xylanases catalytic itinerary (Iglesias-Fernández, 2015): $^5S_1 \rightarrow {}^{2,5}B \rightarrow {}^2S_O$. Which relationship can be between this itinerary and the preactivated conformations of the β-L-septanoside? These conformations are $^4C_{O,1}$ (ζ = 78) and $^{2,3}TC_{4,5}$ (ζ = 75). Interstingly, there is a possible catalytic itinerary with these two conformations: $^{2,3}TC_{4,5}$ (ζ = 75) → $^4C_{O,1}$ (ζ = 78) → $^{6,5}TC_{4,3}$ (ζ = 56). In Figure 6.18, we compare both β-L-septanoside and β-xylose pathways. Furthermore, these two sugars have the same stereochemistry in the carbons 1, 2 and 3, but the superposition of the 4-hexose with the 5-septanoside carbons is unfavourable. Changing the stereochemistry of our L-idosepta 5th carbon ring, the resulting sugar will be the D-glucoseptanoside ("disarmed"). Therefore, we propose to use L-septanosides or D-glucoseptanoside as substrate of β-xylanases.

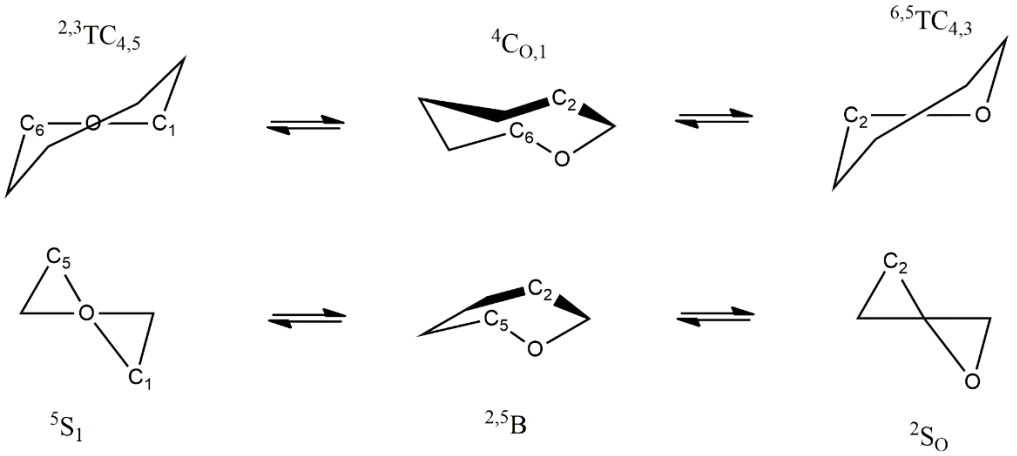


Figure 6.18. Comparison of similar catalytic pathways for L-septanosides and β-xylanases.

5.3. Can D-septanosides be good substrates for GHs?

Taking into account the mimic tendency observed in the previous Figure 6.18, knowing that some mannosidases follow a $^{O}S_2 \rightarrow B_{2,5} \rightarrow {}^{1}S_5$ itinerary, a good 7-membered ring substrate has to follow a $^{3,4}TC_{5,6} \rightarrow {}^{O,1}C_4 \rightarrow {}^{5,4}TC_{3,2}$ itinerary.

Returning to the "armed" α-D-idosepta and β-D-gulosepta gas phase simulations, the most preactivated conformations for both sugars are $^{3,4}TC_{5,6}$ ($\zeta = 75$) and $^{5,4}TC_{3,2}$ ($\zeta = 68$), respectively. In Figure 6.19, the structures of these conformers are compared with the crystallographic structures of the GH76 endo-α-mannosidase complex (PDB 5AGD [Thompson, 2015]) and the GH2 β-mannosidase (PDB 2WBK [Offen, 2009]) complex whose sugars are in $^{O}S_2$ and $^{1}S_5$ conformations.

In the first case, the stereochemistry of both sugar carbons are the same and the carbon 3 of the septanoside is in equatorial orientation. Thus, the overlap seems acceptable. In the second case, the stereochemistry of the carbon 2 of the β-D-gulosepta sugar is opposite to the α-mannose inside the enzyme. Newly, the carbon 3 of the septanoside is in equatorial orientation. Then, the overlap will be good changing the configuration of the carbon 2, then, the good candidate as β-mannosidase substrate is the β-D-idosepta.

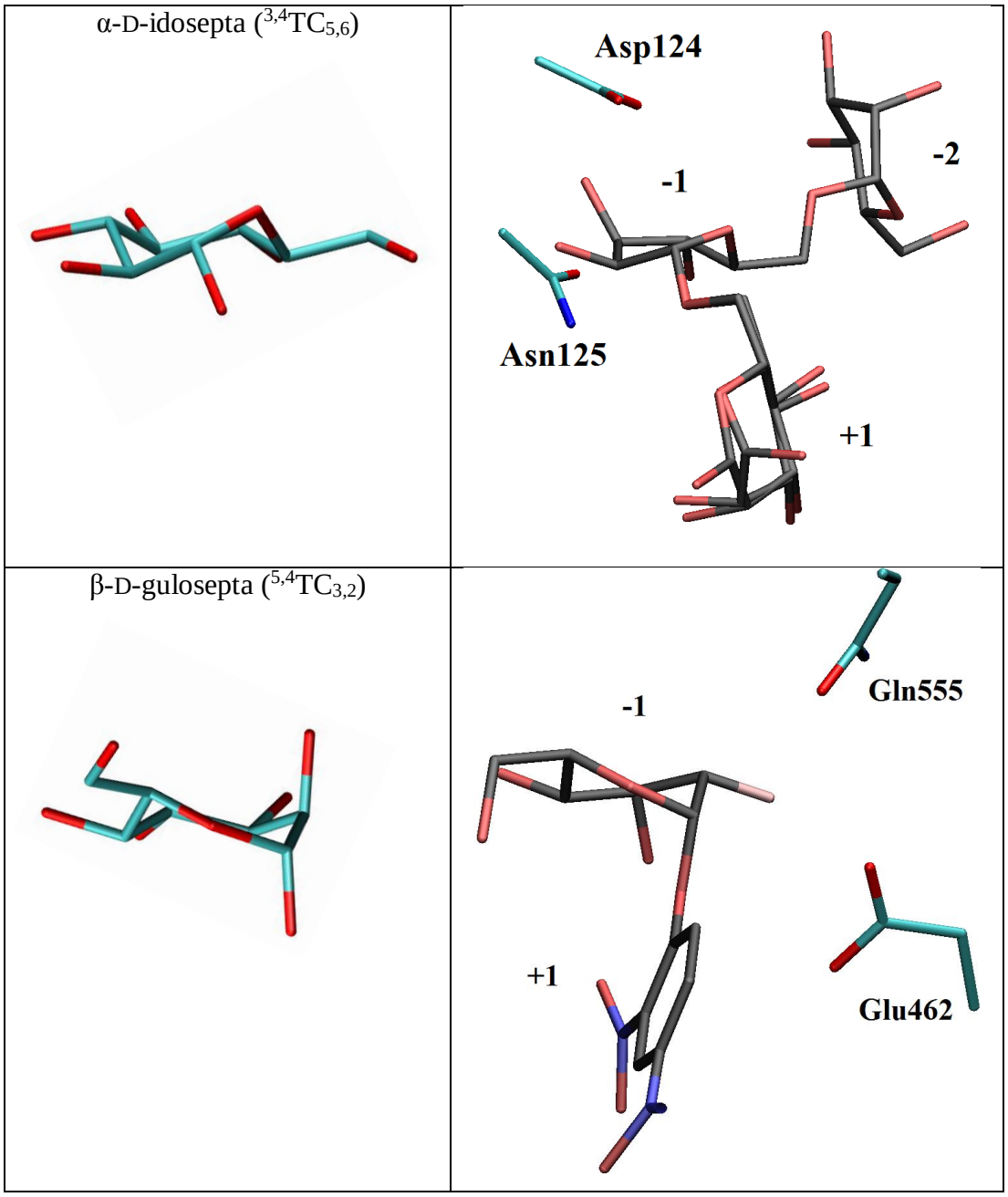


Figure 6.19. Comparison of D-septanoside preactivated configuration with crystallographic structures of α- (top) and β-mannosidase (bottom) complexes.

Finally, we have presented three reconstructed glucosidase-septanoside complexes. In the case of α-glucosidase, the configuration of the carbon 2 of α-D-idosepta difficults the nucleophilic attack, then, changing to α-D-gulosepta, this problem is solved. Furthermore, we have simulated the glycosylation step (rate limiting step) in this last complex and the result is positive, and it shows a very good catalytic activity of the α-

glucosidases with the α-D-gulosepta inside the -1 subsite. In the case of β-D-gulosepta, the $^{1,2}TC_{3,4}$ seems to dock well in the β-glucosidase, but this shape is not well preactivated ($\zeta = 52$).

In summary, our proposal is to use the α-D-idosepta as substrate of α-mannosidases, the α-D-gulosepta as substrate of α-glucosidases, the β-D-idosepta as substrate of β-mannosidases and to continue uncovering the secrets of the GH-septanoside activity.

6. Conclusions

The main conclusions of the present chapter are the following:

1. The use of the Cremer and Pople puckering coordinates as collective variables, within the well-tempered metadynamics approach, is suitable to reconstruct the conformational free energy landscape of 7-membered rings, both gas phase and enzyme-sugar complexes.
2. The conformational barrier ($TC \rightarrow [S]^{\ddagger} \rightarrow TB$) of the cycloheptane at 300 K (4.66 kcal/mol) is 3-4 kcal/mol lower than the potential energy barrier (8-9 kcal/mol), resulting an entropic contribution of 11 cal/mol·K.
3. The most stable conformation adopted by the α-D-idoseptanoside is $^{3,4}TC_{5,6}$ and it is the best preactivated conformation for GH catalysis ($\zeta = 75$). The most stable conformation adopted by the β-D-guloseptanoside is $^{O,6}TC_{5,4}$, but the best preactivated conformation is $^{5,4}TC_{3,2}$ ($\zeta = 68$).
4. Analogously to hexoses, the oxocarbenium septanoside cation can adopt a limited number of conformations (14), due to the planarity of the C_2-C_1-O_6-C_6 atoms (sp^2 hybridity). In this case, it can adopt two chair conformations ($^4C_{O,1}$ and $^{O,1}C_4$).
5. There is a correlation between the conformational behaviour of β-L-idoseptanoside in the gas phase and in complex with the GH31 enzyme. The most stable conformation in gas phase is $^{4,5}TC_{6,O}$ and the most preactivated ones are $^4C_{O,1}$ ($\zeta = 78$) and $^{2,3}TC_{4,5}$ ($\zeta = 75$). The sugar inside the enzyme adopts a $^{O,1}TC_{2,3}$ ($\zeta = 58$) conformation.
6. The glycosylation reaction of the pNP – β-L-idoseptanoside within the GH31 involves an stable oxocarbenium ion species and the conformational itinerary is $^{O,1}TC_{2,3} \rightarrow {}^{O,1}S_3/{}^{5,4}TS_3 \rightarrow {}^{O,1}S_3/B_{(2,3),6}$. However, the reaction is not feasible, due to the high activation barrier (29 kcal/mol).

7. We are able to construct glucosidase-septanoside structures, using docking and molecular engineering methods.
8. The glycosylation reaction of the α-D-gulosepta-1,4-α-glucose within the GH13 enzyme shows a low activation barrier of 12 kcal/mol. The reaction involves a stable oxocarbenium ion (3.5 kcal/mol above the reactants) and a covalent intermediate (5 kcal/mol above the reactants). The conformational itinerary followed by the -1 sugar is the $^{5,6}TC_{O,1} \rightarrow {}^{2}TS_{3,4}/BS_{3,4} \rightarrow BS_{3,4}/{}^{1}S_{3,4}/{}^{5,6}TB_{O,1}$, similar to the expected pathway of an α-glucose inside an α-glucosidase.
9. Taking into account the results of preactivation parameters and reaction modelling, we consider the use of the α-D-idosepta as substrate of α-mannosidases, the α-D-gulosepta as substrate of α-glucosidases, the β-D-idosepta as substrate of β-mannosidase and "disarmed" L-idosepta or D-glucoseptanosides as substrate of xylanases.

Chapter VII – Summary and conclusions

Summary and conclusions

In the present Thesis, Car-Parrinello molecular dynamics methods have been applied to study the conformational landscape of 6-membered and 7-membered ring saccharides, both in the gas phase (via a full QM description) and in the enzymatic environment (via the QM/MM approximation). The use of the metadynamics approach has allowed us to reconstruct conformational and reaction free energy landscapes for retaining and inverting GHs catalysing the hydrolysis of diverse oligosaccharides.

The main conclusions achieved in this Thesis are the following:

- The reaction mechanism catalysed by amylosucrase is characterized by a cyclic conformational itinerary ($^4C_1 \rightarrow {}^4H_3/E_3 \rightarrow {}^4C_1$) of the -1 sugar. The first step of the double-displacement reaction is dissociative (glycosylation) and rate-limiting. In the absence of large glucose oligosaccharides and under low glucose concentrations, the deglycosylation step is energetically more favourable than the transglycosylation step.

- The conformational itinerary followed by the natural mannobiose in GH125 exo-1,6-α-mannosidases is $^OS_2 \rightarrow B_{2,5} \rightarrow {}^1S_5$. In this case, thio-glycosides are not suitable conformational mimics of the natural substrate, as they feature a $^4C_1$ conformation of the -1 sugar at the MC instead of $^OS_2$. Our prediction has been verified experimentally.

- The $^4C_1$ chair conformation is the most stable conformation of the β-glucosyl group in the -1 subsite of GH3 HvExoI, suggesting a cyclic itinerary for the hydrolysis reaction ($^4C_1 \rightarrow {}^4E \rightarrow {}^4C_1$). The fact that the β-configured substrates do not need to be distorted to facilitate cleavage of the glycosidic bond is probably due to the high conformational freedom of the leaving group in exo-GHs such as HvExoI. Substitution of the glycosidic oxygen by sulphur changes the pattern of intramolecular interactions, which slightly affects to substrate conformation.

- The use of the Cremer and Pople puckering coordinates as collective variables in the metadynamics approach allows to reconstruct the conformational free

energy landscape of 7-membered rings, both in the gas phase and enzyme-sugar complexes.

- Septanoside substrates are potential substrates for GHs. The conformational itinerary of their reaction with GHs is very similar to the corresponding hexoses. However, septanoside oxocarbenium cation species appear as stable minima not transition states. We propose the use of α-D-idosepta as substrate for α-mannosidases, α-D-gulosepta for α-glucosidases, β-D-idosepta for β-mannosidase and "disarmed" L-idosepta and D-glucoseptanosides for xylanases.

Resumen en castellano

En las últimas décadas, las mejoras en términos tecnológicos han propiciado un aumento en la capacidad de cálculo y almacenamiento de las supercomputadoras. Este hecho, añadido al desarrollo de nuevos métodos y aproximaciones matemáticas que facilitan el tratamiento de sistemas moleculares complejos, ha empujado a la comunidad científica a interesarse en estudiar computacionalmente sistemas de naturaleza biomolecular.

Concretamente, en la presente Tesis Doctoral, el objeto de estudio han sido sistemas enzima-sustrato encargados de regular el metabolismo de carbohidratos, en este caso, las enzimas estudiadas son conocidas como Glicosil Hidrolasas (GH) y los carbohidratos hidrolizados por estas, generalmente, han sido cadenas de azúcares tales como glucosa o manosa (anillos de 6 átomos), aunque, se ha querido estudiar el comportamiento como sustrato de una nueva clase de carbohidratos artificiales conocidos como septanósidos (dado que forman un anillo de 7 átomos).

Antes de entrar en materia, la importancia de estudiar los azúcares y sus interacciones con las Glicosil Hidrolasas en detalle es crucial a la hora de entender un amplio rango de procesos biológicos, tales como el empaquetamiento proteico y las infecciones víricas (Hurley, 2001). Por ejemplo, el metabolismo de los azúcares es esencial para entender enfermedades como la diabetes de tipo II (α-glucosidasas [Bischoff, 1995] y α-manosidasas [Chui, 1997]) y enfermedades víricas, como el VIH o la gripe (Gloster, 2007).

Las GHs representan un grupo de enzimas cuya actividad catalítica consiste en la degradación y modificación de polisacáridos y glucoconjugados a través de la ruptura del enlace glicosídico (Davies, 1995). Hoy en día, 135 familias de GHs han sido caracterizadas acorde con sus secuencias primarias de aminoácidos en la base de datos *Carbohydrate Acting enZymes* (www.cazy.org) (Henrissat, 1996) (Cantarel, 2009). Además, algunas de estas familias pueden ser clasificadas en clanes según su estructura terciaria (desde GH-A hasta GH-N). Otra clasificación viene dada por el mecanismo de acción de la enzima, si este se da a través de retención o inversión de la configuración del centro anomérico del azúcar en el subsitio -1 de la proteína (Figura R1).

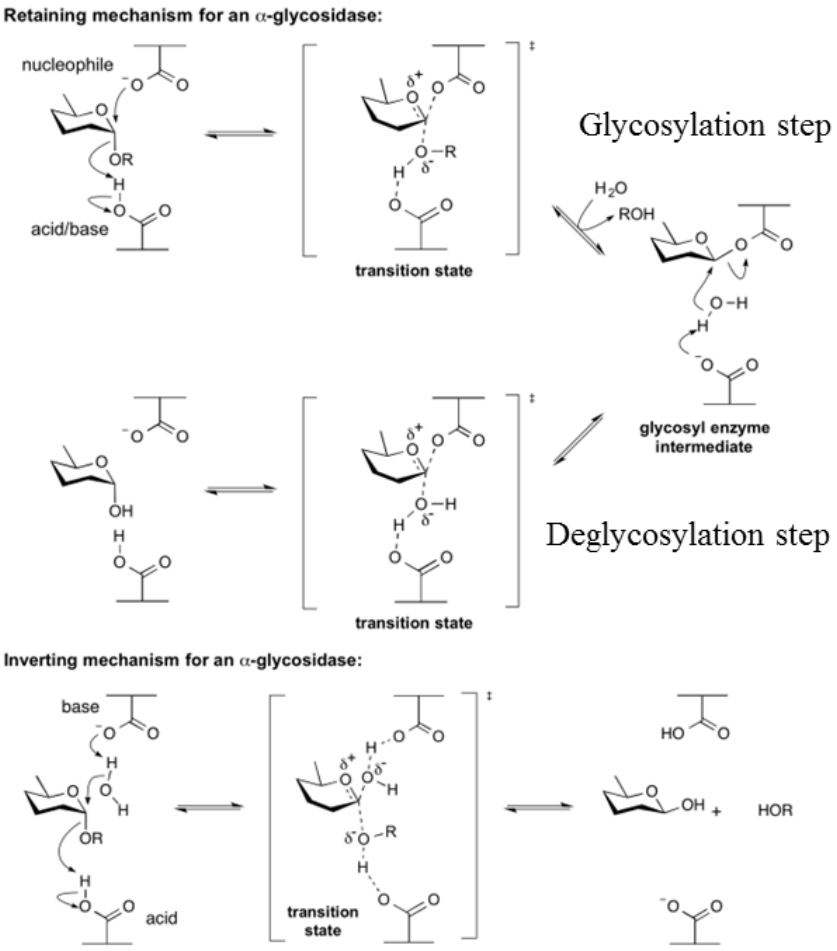


Figura R1. Mecanismo de reacción por retención (arriba) e inversión (abajo) de la configuración del centro anomérico en una α-glicosidasa. Figura adaptada de la referencia (www.cazy.org).

Entrando en materia mecanística, estudios de Efectos Cinéticos Isotópicos (KIE) revelaron que ambos mecanismos envuelven un estado de transición (TS) de tipo ión oxocarbenio (Zechel, 2000). Este TS presenta un carácter disociativo muy grande, con bajos órdenes de enlace entre el carbono anomérico y el oxígeno glicosídico como con el nucleófilo. Consecuentemente, se desarrolla una significante carga positiva en el carbono, que se compensa gradualmente por una transferencia electrónica del oxígeno piránico. Esta donación electrónica da pie al carácter parcial de doble enlace entre el carbono y el oxígeno piránico (C_1-O_p) (carácter sp^2). Esto hace que los átomos del anillo C_2-C_1-O_p-C_5 adopten una geometría plana y, por lo tanto, limita el número de conformaciones que puede adoptar el azúcar en el TS de la reacción: dos medias-sillas (H) $^4H_3$ y $^3H_4$, dos barcas (B) $B_{2,5}$ y $^{2,5}B$ y cuatro sobres (E) E_3, ^3E, E_4 y ^4E. Esa presunción fue confirmada

computacionalmente a través del estudio conformacional de varios iones oxocarbenios (J. Iglesias-Fernández PhD Thesis, 2014).

Así pues, entra en juego una variable como es la conformación de los azúcares que ha resultado ser crucial para entender y caracterizar el mecanismo de las GHs. Pero, no solo es importante la conformación del TS de la reacción, sino que se ha observado que las GHs tienden a distorsionar el azúcar en el momento que este entra en el centro activo para facilitar la hidrólisis.

En 1967, David Phillips observó la primera evidencia de distorsión de un sustrato en una GH a través de sus estudios de la lisozima blanca *hen-egg* (Phillips, 1967), demostrando el papel importante de la conformación del azúcar en la catálisis. Además, muchos estudios, tanto cristalográficos como computacionales, muestran como los azúcares se unen a GH en una forma distorsionada, como barcas ($^{1,4}B$), barcas distorsionadas ($^{O}S_2$, $^{1}S_3$, $^{1}S_5$) y, algunas veces, la distorsión no es necesaria y el azúcar se mantiene en la clásica conformación en forma de silla ($^{4}C_1$) (Davies, 2012) (Vocadlo, 2008).

Para entender el efecto de la enzima sobre la conformación del azúcar y viceversa, debe analizarse el grado de preactivación que muestra una conformación dada de un azúcar para ser hidrolizado. Varios factores son clave para cuantificar dicha preactivación que pueden ser explicados a nivel teórico a través de efectos estéricos entre el nucleófilo y los átomos que rodean el carbono asimétrico, la activación del enlace glicosídico, las atracciones electroestáticas entre los átomos envueltos en la catálisis, el grado de axialidad del enlace glicosídico o cuan parecida es la conformación inicial al estado de transición de la reacción. Así pues, para predecir la adecuación de un azúcar o una de sus conformaciones como posible sustrato de una GH, calcularemos la carga del carbono anomérico, del oxígeno piránico y del oxígeno glicosídico, mediremos la elongación de los enlaces C_1-O_1 y C_1-O_p y cuantificaremos el grado de axialidad del enlace glicosídico.

De cara a estudiar y diferenciar matemáticamente las diferentes conformaciones de un anillo de N átomos (6 y 7 son los valores que nos conciernen en este trabajo), en 1975, Cremer y Pople desarrollaron el desarrollo matemático para transformar las coordenadas cartesianas de un sistema de 3N coordenadas en unas variables que llamaron coordenadas de empaquetamiento (Cremer, 1975). En su trabajo demostraron que utilizando N-3 coordenadas eran capaces de diferenciar todas las conformaciones posibles de un anillo

formado por N elementos. En el caso de un anillo de 6 (como el de la glucosa), son necesarias 3 coordenadas (Q, ϕ, θ ó q_x, q_y, q_z) para diferenciar las 38 conformaciones estándar y el espacio continuo de conformaciones intermedias entre ellas (Figura R2). No obstante, estudios recientes de nuestro grupo de investigación han observado que la variable Q actúa como un modo rápido cuando se usa como variable colectiva en metadinámica (Iglésias-Fernández, 2015) y, nuestros estudios conformacionales de azúcares formados por anillos de 6 átomos han podido ser llevados a cabo utilizando dos variables (ϕ, θ – Proyección de Mercator).

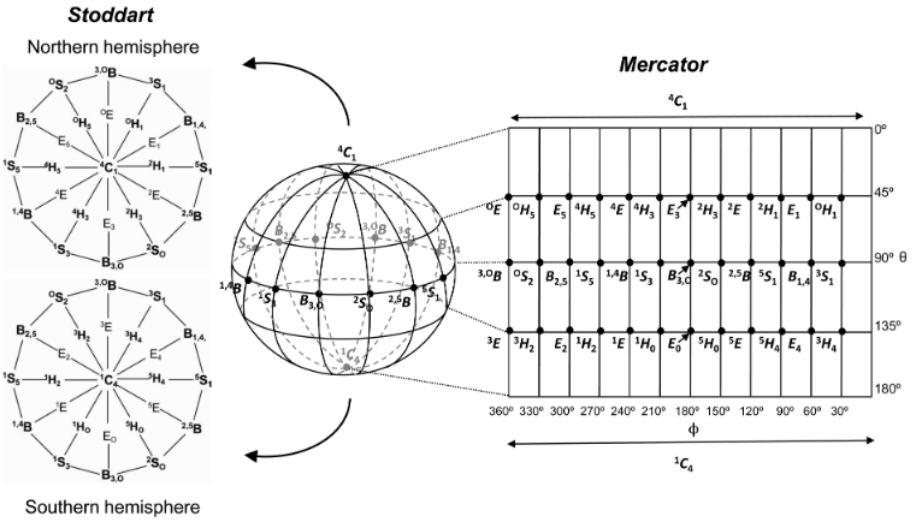


Figura R2. Conformaciones principales sobre la esfera de Cremer y Pople a lo largo de dos de sus representaciones más usadas: diagrama de Stoddart (se muestran las proyecciones desde el Polo Norte y el Sud) y la proyección de Mercator. Figura tomada de la referencia (Davies, 2012).

No obstante, para llevar a cabo el estudio de moléculas formadas por anillos de 7 átomos, tanto el cicloheptano como posibles septanósidos que actúen como sustrato de GHs, la complejidad matemática y el número de conformaciones a contemplar aumenta considerablemente. Para este caso, son necesarias cuatros variables de Cremer y Pople (φ_2, φ_3, q_2, q_3) para poder diferenciar las 98 conformaciones estándar y el espacio continuo que las separa. Encontramos siete grupos de catorce conformaciones en cada uno: silla distorsionada (TC), silla (C), sofá (S), sofá distorsionado (TS), sofá-barca (BS), barca (B) y barca distorsionada (TB) (Boessenkool, 1980).

Por lo tanto, aprovechando la experiencia de los antiguos miembros de nuestro grupo (Dr. X. Biarnés, Dr. A. Ardèvol, Dr. J. Iglésias-Fernández, Dr. Rovira) en estudios

conformacionales de azúcares, tanto aislados como en el interior de GHs, en la presente tesis se han llevado a cabo estudios conformacionales y de reactividad en sistemas GH-hexosa (GH3, GH13 y GH125) y, se ha aplicado la misma metodología como estrategia para cuantificar la capacidad de los septanósidos (anillos de 7 átomos) como posibles sustratos y/o inhibidores del tipo de enzimas que nos ocupa esta tesis.

Capítulo III. Itinerario conformacional y mecanismo catalítico de la GH13 amilosucrasa.

La amilosucrasa (AS) es una enzima presente en la vida microbiana que pertenece a la familia 13 de las GHs, familia formada por enzimas que se dedican a la transformación del almidón. No obstante, el papel de la AS no es degradar almidón, sino sintetizar polímeros amiloideos (α-1,4 glucanos), a partir de la degradación del azúcar común (sacarosa o sucrosa) (Emond, 2008b).

La primera señal de amilosucrasa en humanas fue observada en 1975 (Riou, 1983), donde fue aislada en la especie *Neisseria polysaccharea* de la garganata de niños saludables de Europa y África. El ejemplo más rutinario de la acción de esta enzima en nuestro cuerpo lo experimentamos al despertar algunos días, cuando notamos una materia insoluble en la garganta. Durante la noche, las bacterias presentes en nuestras gargantas digieren la sacarosa (glucosa y fructosa). La amilosucrasa (*Np*AS) hidroliza el azúcar y acumula glucose hasta que llega a una concentración límite que permite a la enzima formar cadenas de anillos de glucosa unidos por las posiciones 1,4, formando polímeros insolubles. En resumen, bajo ciertas condiciones de concentración de sacarosa, la enzima actúa como hidrolasa y bajo otras, esta actúa como polimerasa (Potocki de Montalk, 2000). En estos experimentos, en condiciones donde la concentración de sacarosa era inferior a 20 mM, el ratio hidrólisis/polimerización era 20 min^{-1}/10 min^{-1}, mientras que con concentraciones mayores, este ratio se invertía a 33 min^{-1}/113 min^{-1}.

En este estudio, el objetivo principal es caracterizar el itinerario conformacional de la glucosa del subsitio -1 de la *Np*AS y simular el mecanismo de retención de la hidrólisis y la polimerización dentro de la enzima.

Dicho propósito se ha llevado a cabo en colaboración con el grupo de la Prof. Isabelle André (Institut National des Sciences Appliquées in Toulouse), cuya contribución en el ámbito del estudio de la amilosucrasa es extensivamente conocido (Remaud-Simeon,

1995) (Potocki de Montalk, 1999) (Potocki de Montalk, 2000) (Emond, 2008a) (Daudé, 2013) (Cambon, 2014) (Moulis, 2016).

El código genético de la enzima en *Neisseria polysaccharea* fue analizado en 1999 (Potocki de Montalk, 1999) y algunas estructuras cristalográficas fuero obtenidas a partir de entonces (Tabla R1).

Table R1. Lista de estructuras cristalográficas obtenidas por rayos X disponibles de NpAS.

Código PDB	Resolución	Mutaciones	Sustrato	Conformación	Referencía
1G5A	1.40 Å	G1S, L2P, I3N, L4S, G537D	-	-	(Skov, 2001)
1JGI (MC)	2.00 Å	E328Q	Sacarosa	$^4C_1$	(Mirza, 2001)
1S46 (CI)	2.20 Å	E328Q	D-glucosilo	$^4C_1$	(Jensen, 2004)
1MW0	2.01 Å	E328Q	Maltoheptosa	$^4C_1$	(Skov, 2002)

Dado que la amilosucrasa es considerada como una α-glucosidasa, el itinerario conformacional esperado para un miembro de su familia sería el que atraviesa la región $^4C_1$ – $^4H_3$ – $^1S_3$. No obstante, experimentalmente, las estructuras del Complejo de Michaelis (MC) y el intermedio covalente (CI), muestran un camino cíclico que empieza y acaba en la región $^4C_1$ no distorsionada. En ambos casos, se tuvo que mutar el residuo catalítico ácido/base E328Q de modo que la reacción no se llevara a cabo (Mirza, 2001) (Jensen, 2004). De cara a entender el papel como polimerasa de la amilosucrasa, en 2002 se cristalizó el mismo mutante formando un complejo con maltoheptosa (Skov, 2002), lo que evidenció un centro activo en forma de bolsillo con siete subsitios ocupables por un oligosacárido.

Volviendo al tema de la conformación del CI de la reacción, se conocen dos estructuras de este tipo de intermedio en la familia 13, la nombrada amilosucrasa que no muestra distorsión y la sucrosa fosforilasa (Mirza, 2006) que muestra la esperada conformación $^1S_3$, que confirmaría el camino conformacional lineal esperado en α-glucosidasas. Siendo ambas enzimas de la misma familia y actuando sobre el mismo sustrato, dicha discrepancia es inesperada, no obstante, podría explicarse entendiendo que la amilosucrasa puede actuar como hidrolasa y polimerasa, mientras que la sucrosa

fosforilasa puede actuar como hidrolasa y fosforilasa, a la vez. La diferencia estructural observada entre ambas enzimas reside en la interacción del oxígeno libre del nucleófilo Asp286, una vez formado el intermedio. En el caso de la sucrosa fosforilasa, este oxígeno interacciona con el hidroxilo 6-OH de la glucosa, mientras que, en la amilosucrasa el nucleófilo experimenta una rotación y el oxígeno libre está dirigido hacia el subsitio +1.

Por lo tanto, dadas las discrepancias observadas dentro de la misma familia, el objetivo de comprender el mecanismo de acción de la *Np*AS a lo largo de la hidrólisis es crucial, de cara a diseñar inh]ibidores para dicha enzima como crear nuevos mutantes para desarrollar biocatalizadores más eficientes para aplicaciones biotecnológicas. En este trabajo, mostraremos los resultados obtenidos de modelizar las reacciones de hidrólisis y polimerización (transglicosilación) en el interior de la *Np*AS (la transglicosilación es equivalente a la hidrólisis, únicamente cambiando la molécula de agua [H-O-H] del paso de deglicosilación por un grupo aceptor [H-O-R]).

Capítulo III – Objetivos

1) Determinar la conformación más estable del anillo de glucosa en el subsitio -1 de la amilosucrasa.
2) Modelizar la reacción de hidrólisis (pasos de glicosilación y deglicosilación) del sustrato natural en la *Np*AS, descifrando el itinerario conformacional del anillo de azúcar en la posición -1.
3) Modelizar la reacción de transglicosilación en la *Np*AS, favoreciendo la formación de un disacárido de 1,4-maltosa.

Capítulo III – Métodos, resultados y discusión

A partir de la estructura cristalográfica publicada en la referencia (Mirza, 2001), se han construido los modelos solvatados en agua y contrarrestadas las cargas con contraiones (Na^+) del mutante (E328Q) y la enzima salvaje (WT) de la *Np*AS. Una vez demostrado el buen funcionamiento del campo de fuerzas de AMBER 11 y Glycam 06 para parametrizar y estudiar el mutante, todas las posteriores simulaciones se llevaron a cabo sobre el modelo WT *Np*AS.

Se ha optimizado, aumentado la temperatura del sistema paulatinamente hasta llegar a 300 K, se ha estabilizado la densidad del sistema y se ha llevado a cabo dinámica molecular clásica a dicha temperatura hasta que la estructura de la proteína se ha

estabilizado. Una vez hecho eso, se ha escogido una geometría representativa para empezar una dinámica molecular híbrida de Mecánica Cuántica/ Mecánica Molecular (QM/MM), estabilizar la estructura nuevamente a 300 K y, utilizando metadinámica "bien-temperada" (Well-T MTD) y las coordenadas de empaquetamiento como variables colectivas, se ha reconstruido la superficie de energía libre (FEL) conformacional del azúcar dentro de la enzima (Figura R3). En ella se observa, en buen acuerdo con el experimento, que la conformación más estable es la silla $^4C_1$, siendo unas 5 kcal/mol más estable que el mínimo en la región barca $B_{3,O}$.

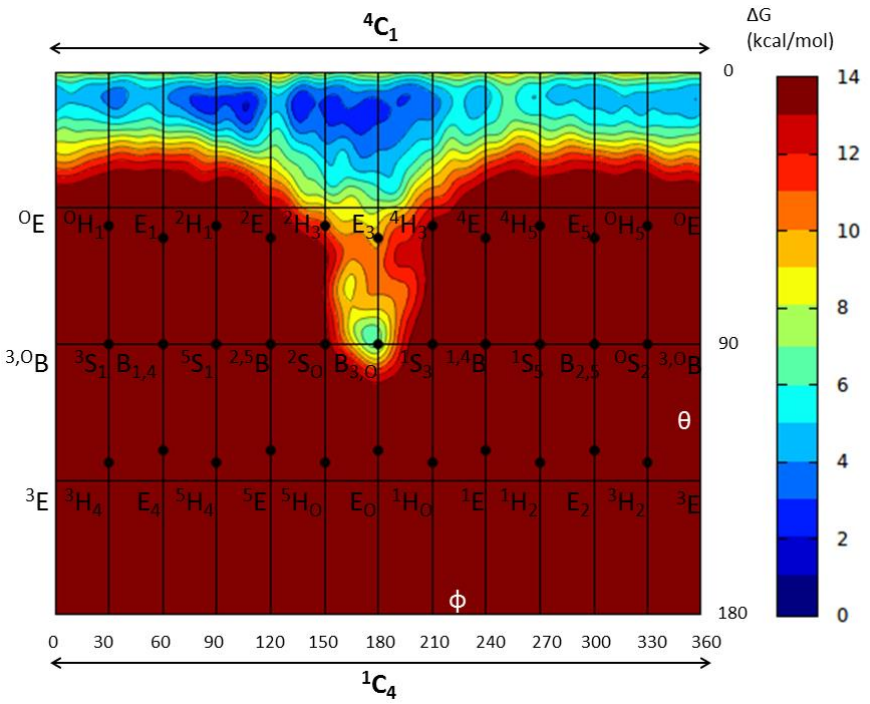


Figure R3. Proyección de Mercator de la FEL conformacional del anillo α-glucosil del subsitio -1 de la amilosucrasa. Cada isolínea está separada por 1 kcal/mol.

A partir de una estructura representativa del mínimo no distorsionado, se llevó a cabo la simulación del paso de glicosilación utilizando metadinámica activando tres variables colectivas que correspondían al paso de protón del Glu328 al oxígeno glicosídico, la ruptura del enlace glicosídico y el ataque nucleofílico del Asp286 al carbono asimétrico. Tras esperar a que el sistema fuera del pozo de reactivos al de productos y volviera al estado inicial, la superficie de energía libre mostraba un camino de reacción disociativo, donde la ruptura del enlace glicosídico se daba antes que el ataque nucleofílico del Asp286. La barrera cinética y termodinámica del proceso eran de $\Delta G^{\ddagger}$ = 17.3 kcal/mol y ΔG = -1.5 kcal/mol, respectivamente. El camino conformacional seguido por el anillo de

glucosa durante el proceso fue cíclico, como se había observado experimentalmente, pasando por un estado de transición de la forma $^{4}H_{3}/E_{3}$.

No obstante, para que se llevara a cabo el proceso de deglicosilación era necesario acercar una molécula de agua cerca de la región catalítica. Al terminar la glicosilación, la mejor candidata a ese papel se encontraba a 6 Å del carbono asimétrico. Por ello, se llevó a cabo una simulación intermedia (ataque del agua) donde se activaban dos variables, de cara a hacer accesible el carbono anomérico al agua. Por un lado, se activó la distancia agua-carbono y, por el otro, encontrando una correlación entre la interacción 6-OH – oxígeno solitario del Asp286 con el cambio conformacional necesario para el ataque del agua (la silla muestra impedimento estérico para el ataque), se activó dicha interacción 6-OH – Asp286. Los resultados fueron prometedores, pues, a medida que el agua se acercaba a la región catalítica, el azúcar mostraba mejor capacidad para cambiar su forma a una conformación $E_{3}/^{4}C_{1}$ que permitía el proceso de deglicosilación. Durante el proceso, se encontró un mínimo correspondiente al CI que era 5 kcal/mol más estable que los reactivos iniciales (MC). El proceso de acercamiento del agua implicaba una barrera cinética de 4.6 kcal/mol y llevaba a un estado inicial para la deglicosilación siendo 1 kcal/mol más estable que el estado MC.

Partiendo de dicha estructura favorable para la deglicosilación donde el azúcar estaba en una conformación $E_{3}/^{4}C_{1}$, se llevó a cabo la simulación del segundo paso de la hidrólisis usando tres variables colectivas activando la ruptura del enlace Asp286-azúcar, la formación del enlace entre el agua catalítica y el carbono anomérico y la transferencia del protón del agua al Glu328 desprotonado. La FEL resultante muestra dos mínimos separado por un estado de transición, cuyas barreras cinética y termodinámica son $\Delta G^{\ddagger}$ = 13.3 kcal/mol y ΔG = -7.6 kcal/mol, respectivamente. La conformación de la glucosa -1 ha seguido un itinerario $E_{3}/^{4}C_{1} – E_{3} – {}^{4}C_{1}$. El perfil resultante de la simulación del proceso de hidrólisis se muestra en la Figura R4, donde se observa que la barrera cinética del proceso es 17.3 kcal/mol (en acuerdo con los resultados experimentales, 17.9 kcal/mol era la barrera esperada, dada una k_{cat} de 33 min^{-1}), dando como paso limitante el proceso de glicosilación.

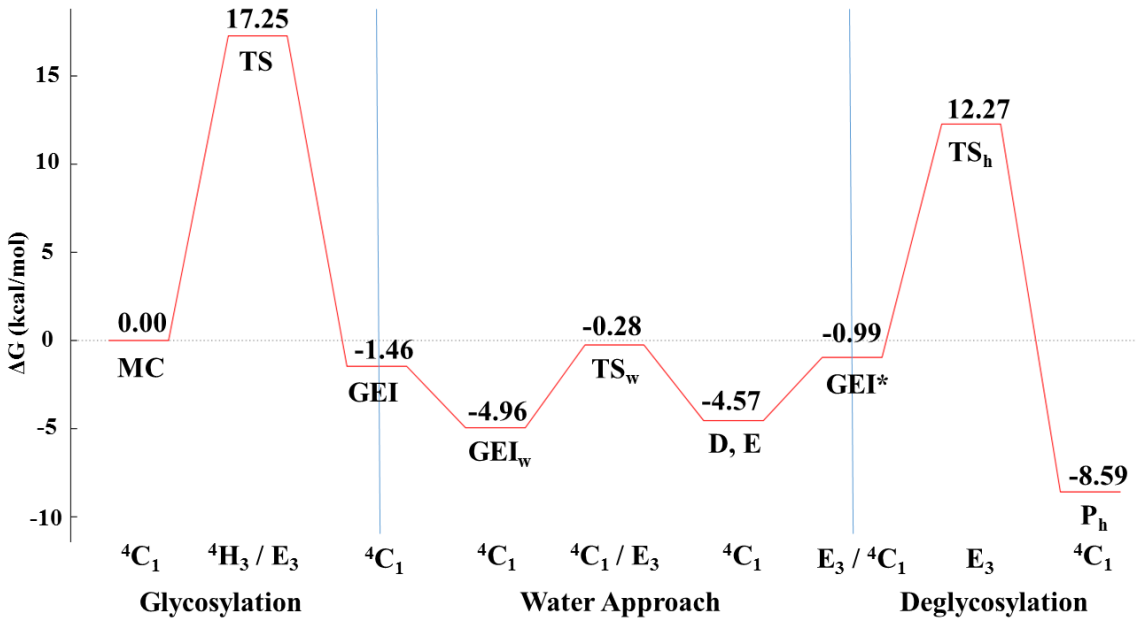


Figure R4. Representación esquemática del camino complete de la hidrólisis de la sucrosa en el interior de la GH13 amilosucrasa.

Por último, se reconstruyó un modelo sustituyendo el agua catalítica y la fructosa (grupo saliente) por una molécula de α-glucosa preparada para atacar el carbono anomérico con el 4-OH y llevar a cabo la transglicosilación. Una vez el sistema estuvo preparado para la dinámica molecular QM/MM, se estabilizó la estructura a 300 K y se simuló el proceso de transglicosilación con metadinámica, activando nuevamente tres variables colectivas que activaban la ruptura del enlace Asp286-glucosa, el ataque del 4-OH de la glucosa sobre el carbono anomérico y la transferencia del protón del 4-OH al Glu328 desprotonado. Como en el caso de la deglicosilación, el proceso pasó a través de un estado de transición de tipo E_3 y finalizó en $^4C_1$. No obstante, las barreras de activación y termodinámica fueron distintas: $\Delta G^{\ddagger}$ = 22.1 kcal/mol y ΔG = +4.6 kcal/mol, respectivamente, lo que implicaba un perfil simétrico entre glicosilación y polimerización, donde ambos estados de transición estaban a energías parecidas (17.3 y 17.1 kcal/mol) y el producto de polimerización era prácticamente igual de estable que el reactivo MC. Este hecho se comprende al contemplar la formación del disacárido maltosa como la reacción inversa de la ruptura del disacárido sucrosa.

Por lo tanto, a partir de nuestras simulaciones, llevadas a cabo en condiciones de bajas concentraciones de sucrosa, observamos que la hidrólisis está favorecida energéticamente sobre la polimerización.

Capítulo III – Conclusiones

1. El estudio conformacional del grupo α-glucosil en el subsitio -1 de la *Np*AS confirma la conformación silla observada experimentalmente como la más estable en el Complejo de Michaelis.
2. La simulación de la reacción de hidrólisis (glicosilación y deglicosilación) de la sucrosa dentro de la amilosucrasa confirma que la glicosilación es el paso limitante de la reacción. La barrera de activación resultante (17.3 kcal/mol) está en buen acuerdo con los experimentos (17.9 kcal/mol).
3. El itinerario conformacional del paso de glicosilación es el $^{4}C_{1} \rightarrow {}^{4}H_{3}/E_{3} \rightarrow {}^{4}C_{1}$ camino cíclico, mostrando un intermedio covalente no distorsionado, resultado nuevamente en acuerdo con la estructura cristalográfica publicada en la referencia (Jensen, 2004).
4. Debido al hecho que la conformación silla no está preactivada para el proceso de deglicosilación/transglicosilación, el azúcar -1 se distorsiona ligeramente ($E_{3}/{}^{4}C_{1}$) cuando el agua/glucosa catalítica se acerca al carbono anomérico (3-3.2 Å).
5. En las condiciones de la simulación (no presencia de grandes oligosacáridos y bajas concentraciones de glucosa/sucrosa), el proceso de deglicosilación (12.3 kcal/mol) está más favorecido energéticamente que la transglicosilación (17.1 kcal/mol). Este resultado está en buen acuerdo con la información experimental. Ambos itinerarios conformaciones evolucionan a través de un estado de transición de tipo E_{3}.

Capítulo IV - Itinerario conformacional y mecanismo catalítico de la GH125 α-manosidasa

Dentro del grupo de enzimas que muestran procesado de N-glicanos, la familia que nos ocupa en este capítulo es la GH125 exo-1,6-α-manosidasa. Esta enzima actúa a través de un mecanismo de inversión de la configuración (Figura R1 [abajo]) y es una manosidasa que rompe enlaces exo- con independencia de un catión metálico.

Añadido a los estudios previos mostrados en la presente Tesis, la importancia de conocer el comportamiento conformacional de los azúcares a lo largo de la coordenada de reacción en GHs es crucial para entender la acción de algunos patógenos microbianos. En este caso, la GH125 exo-1,6-α-manosidasa fue encontrada en una variedad de hongos

y bacterias del colon humano (Gregg, 2011). Dicha enzima forma parte de un gran grupo de manosidasas que muestran procesado de N-glicanos en las familias 38, 47, 76, 92 y 99. Algunos de ellos actúan por retención (38, 76 y 99) y otros por inversión (47, 92 y 125) de la configuración del azúcar a hidrolizar.

En las enzimas por retención, la exo-GH38 (Petersen, 2010), siendo Zn^{2+}-dependiente, y la endo-GH76 (Thompson, 2015) siguen el itinerario conformacional ${}^{O}S_2 \rightarrow B_{2,5} \rightarrow {}^{1}S_5$, mientras que, para la endo-GH99, el mecanismo propuesto va a través de un intermedio epóxido (Thompson, 2012b).

En el caso de las enzimas por inversión, la exo-GH47, siendo Ca^{2+}-dependiente, sigue el itinerario ${}^{3}S_1 \rightarrow {}^{3}H_4 \rightarrow {}^{1}C_4$ (Thompson, 2012a). La Ca^{2+}-dependiente GH92 y nuestra exo-GH125 son casos problemáticos. En ambos casos, se han cristalizado complejadas con tioderivados de la manobiosa, mostrando un anillo de manosa no distorsionado (${}^{4}C_1$) en el subsitio -1 (Zhu, 2010) (Gregg, 2011), hecho que no está de acuerdo con el comportamiento esperado para α-manosidasas por inversión. Además, la GH92 fue cristalizada junto a un inhibidor manoimidazol cuya conformación era una forma distorsionada ${}^{1}S_5/{}^{4}H_5$. Este hecho nos hizo dudar de la buena mimetización de los tioderivados respecto a los sustratos naturales, pues, un Complejo de Michaelis no distorsionado implicaría un estado de transición de la forma ${}^{4}H_3/E_3/{}^{4}E$.

En definitiva, el objetivo principal de este trabajo es encontrar la conformación más estable del sustrato natural dentro de la GH125 α-manosidasa. En segundo lugar, nos gustaría poder responder a la pregunta de si la conformación del sustrato natural no es silla, por qué el tioderivado nos muestra dicha conformación. ¿No son siempre este tipo de mímicos un buen aliado para comprender el comportamiento de los sustratos naturales? Por lo tanto, un estudio crucial para responder a dichas preguntas sería cuantificar el efecto del átomo de azufre en el comportamiento conformacional de la α-manosa en fase gas. Finalmente, partiendo de la conformación del sustrato natural más estable, tenemos como objetivo simular la reacción de hidrólisis dentro de la enzima GH125 y confirmar el itinerario conformacional de la manosa en su interior, durante todo el proceso.

Este trabajo ha motivado experimentos cristalográficos llevados a cabo por el grupo del Prof. Gideon J. Davies (Royal Society Ken Murray Research Professor en el Departamento de Química de la Universidad de York) donde un mutante ácido/base

(Gln218Glu) de la *Clostridium perfringens* GH125 ha sido cristalizada junto al sustrato natural. Nuestros resultados y su estructura cristalográfica están de acuerdo entre ellos y ambos trabajos han estado publicados recientemente (S. Alonso-Gil, 2017).

Capítulo IV – Objetivos

1) Comparar el comportamiento conformacional de las moléculas de α-D-manopiranosa y de 1-tio-α-D-manopiranosa en fase gas.
2) Reconstruir las FELs conformacionales del tioderivado y del sustrato natural dentro de la GH125 exo-1,6-α-manosidasa.
3) Modelizar el mecanismo por inversión de configuración sobre la conformación más estable del sustrato natural en el interior de la GH125 exo-1,6-α-manosidasa.

Capítulo IV – Métodos, resultados y discusión

Como cálculo previo, se reconstruyó la FEL conformacional de la 1-tio-α-D-manopiranosa aislada utilizando metadinámica con las coordenadas de empaquetamiento de la proyección de Mercator como variables colectivas. Comparando dicha FEL con la obtenida en la referencia (Thompson, 2012a) para la α-D-manopiranosa natural, no se observan grandes variaciones de estabilidad en la región ecuatorial respecto a la conformación silla $^4C_1$ que sigue siendo la más estable. La única diferencia considerable es la estabilidad relativa de la silla inversa $^1C_4$, que se ve disminuida debido a la presencia del azufre en la posición del carbono asimétrico. Por lo tanto, podemos afirmar que la sustitución del oxígeno glicosídico por azufre no debería afectar al comportamiento conformacional y, por lo tanto, a la capacidad mímica de los tioderivados.

A partir de la estructura cristalográfica de la referencia (Gregg, 2011), se construyeron dos modelos de la enzima GH125 en complejo con el tioderivado y el sustrato manobiosa natural (cambio del azufre en la posición glicosídica por oxígeno), se solvataron y se compensaron las cargas con contraiones (Na^+). Se llevó a cabo el mismo proceso que en el capítulo anterior: minimización, calentamiento a 300 K, estabilización de la densidad y de la estructura proteica a dicha temperatura. En ambos casos, la conformación del azúcar se mantuvo en la silla $^4C_1$ inicial.

Como se hizo con la amilosucrasa, ambos modelos de la GH125 fueron tratados con metodología QM/MM y se estudió el comportamiento conformacional del sustrato natural y de su análogo con azufre, utilizando metadinámica Well-T. En la Figura R5 se muestran

ambas FEL conformacionales, donde se pueden evidenciar las diferencias conformacionales entre ambos azúcares.

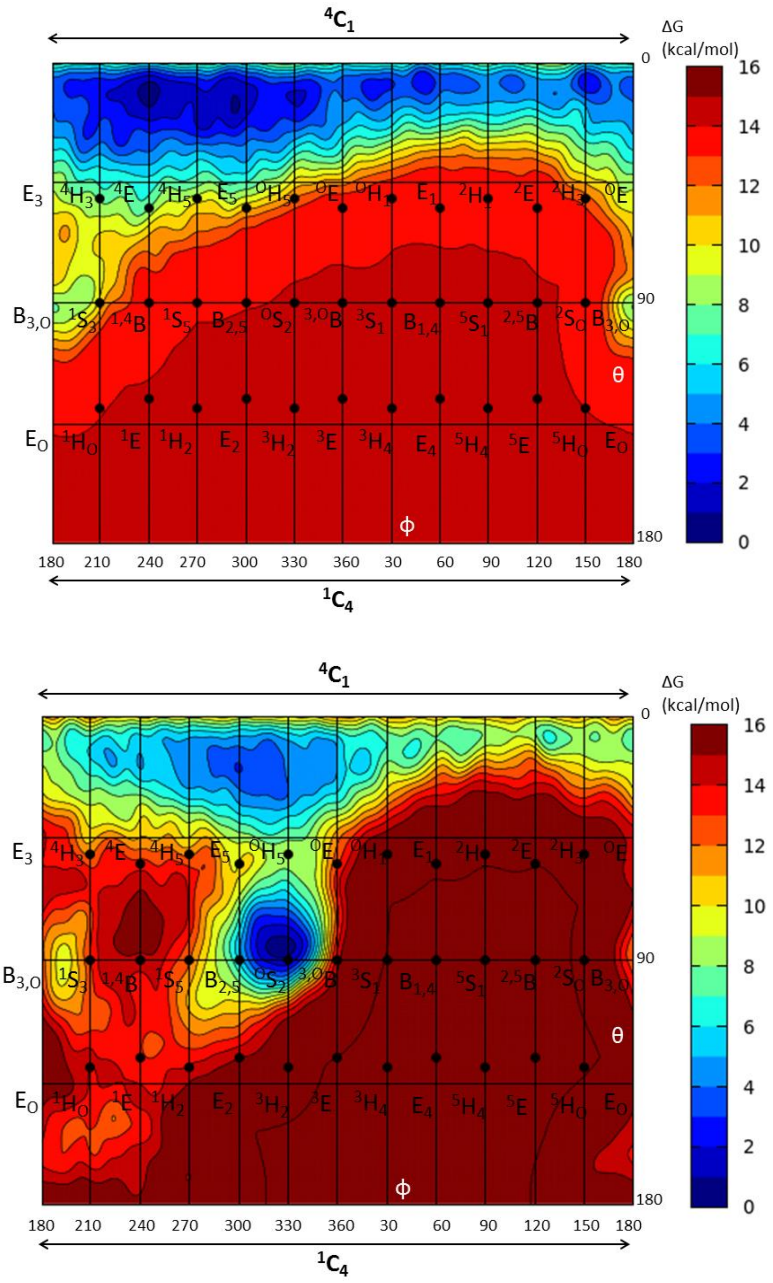


Figura R5. Superficie de energia libre conformacional (proyección de Mercator) del anillo de (arriba) 1-thio-α-manosil y de (abajo) α-manosil en el subsitio -1 de la GH125 exo-1,6-α-manosidasa. Cada isolínea separada por 1 kcal/mol.

Como se puede apreciar en la Figura R5 (arriba), los resultados del estudio conformacional del tioderivado muestran como conformación más estable la silla $^4C_1$, confirmando el resultado experimental. No obstante, observando la Figura R5 (abajo), el comportamiento conformacional del sustrato natural dentro de la enzima es

completamente distinto. En este caso, la conformación más estable es la barca distorsionada $^{O}S_2$. Analizando la trayectoria de ambas simulaciones y calculando las interacciones entre los residuos polares del centro activo y los grupos OH del anillo de azúcar, encontramos la razón de la divergencia en el comportamiento conformacional de ambas moléculas. Cuando se da el cambio conformacional de silla $^{4}C_1$ a $^{O}S_2$, hay dos grupos alcohol que cambian sus orientaciones y, por lo tanto, encuentran nuevos aminoácidos con los que interaccionar. La clave para entender la diferencia de estabilidad entre ambas conformaciones está en la manera de encajar el azúcar dentro del centro activo. Cuando la manobiosa está distorsionada, el alcohol 2-OH interacciona fuertemente (~2.1 Å) con la Ser217, mientras que ese mismo alcohol cuando el azúcar está en silla, interacciona débilmente (o no interacciona) con la Asn300 (~2.7 Å). En cambio, la interacción del 2-OH del tioderivado con dicha asparagina es fuerte (~1.8 Å) y eso se explica a través del cambio de tamaño del tiocompuesto respecto al sustrato natural. Un enlace C-S es más largo que un enlace C-O y el ángulo formado por C-S-C es más cerrado que el C-O-C. Esto promociona una interacción enzima-sustrato distinta en el caso del tioderivado, respecto a lo observado con el sustrato natural y, esa es la razón por la cual, en este caso, el tiocompuesto no mimetiza correctamente el comportamiento conformacional de la manobiosa natural.

A partir de una estructura representativa del mínimo sobre la región $^{O}S_2$ del mapa conformacional, se ha llevado a cabo la simulación de la reacción de hidrólisis de la manobiosa en el interior de la enzima GH125. Utilizando metadinámica, se han activado cuatro variables colectivas que representan la transferencia de protón del ácido/base Asp218 al oxígeno glicosídico, la ruptura del enlace glicosídico, el ataque del agua catalítica sobre el carbono anomérico y una transferencia de protones desde el agua catalítica hasta el residuo auxiliar Glu391, a través de dos aguas adicionales. Debido a la complejidad de simular una reacción con cuatro grados de libertad, la reacción se paró una vez había contemplado toda la región de productos, a pesar de que el sistema no volvió al pozo de reactivos, pasando por el estado de transición. La barrera de activación que mostró la FEL era de $\Delta G^{\ddagger}$ = 22.3 kcal/mol y el resultado final (sin convergir) mostraba una endotermicidad del proceso de unas 3 kcal/mol. El cálculo muestra un camino de mínima energía libre a través de un mecanismo disociativo, donde la rotura del enlace glicosídico ocurre antes que el ataque nucleofílico del agua catalítica. La conformación del grupo manosil del subsitio -1 evoluciona a través de la región $^{O}S_2$ – $B_{2,5}$ – $^{1}S_5/B_{2,5}$

confirmando uno de los itinerarios conformacionales esperados para las α-manosidasas que actúan por inversión de la configuración del centro asimétrico.

Capítulo IV – Conclusiones

1. En fase gas, la 1S-tio-α-manosa y la α-manosa no muestran diferencias relevantes en sus FEL conformacionales. Únicamente, disminuye la estabilidad relativa de la silla $^{1}C_{4}$ en presencia de un átomo de azufre.
2. El estudio conformacional del grupo 1S-tio-α-manosil en el subsitio -1 de la enzima GH125 confirma la conformación experimental silla $^{4}C_{1}$ como la más estable.
3. El estudio conformacional del grupo α-manosil en el subsitio -1 de la enzima GH125 muestra que la conformación barca distorsionada $^{O}S_{2}$ como la forma más estable del sustrato natural en el Complejo de Michaelis.
4. Las diferencias conformacionales entre ambos azúcares dentro de la enzima son debidas a el patrón de interacciones de puente de hidrógeno entre dichos azúcares y los aminoácidos polares del centro activo de la proteína.
5. La simulación de la reacción de hidrólisis dentro de la GH125 α-manosidasa muestra un itinerario catalítico a través de la región $^{O}S_{2} \rightarrow B_{2,5} \rightarrow {}^{1}S_{5}/B_{2,5}$. La reacción sigue un proceso de cuatro paso: el ácido/base protona el oxígeno glicosídico, empieza la rotura del enlace glicosídico, el agua catalítica ataca el carbono anomérico y finalmente, la α-manosa protonada (cargada positivamente) transfiere el protón sobrante al residuo base asistente a través de un túnel con dos aguas implicadas.

Capítulo V – Caracterización de la estructura cristalográfica de una GH3 β-D-glucan glucohidrolasa promiscua

Las paredes celulares de la cebada (*Hordeum vulgare*) son ricas en (1,3)- y (1,4)-β-D-glucanos. Uno de los procesos clave en el desarrollo de la planta es la hidrólisis de (1,3)-, (1,4)- y (1,3;1,4)-β-D-glucanos durante la degradación de la membrana y la debilitación de dicha pared en la elongación de coleóptilos (Varghese, 1999). Las enzimas encargadas de esa función son las endo- y exo-glucanasas (Hrmova, 2001a).

Concretamente, un miembro de la familia 3 de las GHs conocida como GH3 β-D-glucan glucohidrolasa ha sido exhaustivamente estudiada debido a su rol importante en el desarrollo de la cebada. Este trabajo está centrado en la forma exo-1,3-1,4-β-D-glucanasa designada HvExoI, cuyas evidencias experimentales han sido publicadas en CAZy y en el *Protein DataBase (PDB)* bajo el nombre de nuestros colaboradores (22 estructuras cristalinas; Prof. Maria Hrmova, Universidad de Adelaida).

Experimentalmente, hay dos detalles de esta enzima que llaman la atención por encima de detalles estructurales: por un lado, el itinerario conformacional deducido a partir de mímicos del MC (tioderivados, [Hrmova, 2001b y 2002]), GEI (fluoroderivado, [Hrmova, 2001b]) y productos (β-glucose, [Hrmova, 2001b]) contempla un camino cíclico, partiendo de una inesperada forma no distorsionada ($^4C_1$, cuando se esperaría la $^1S_3$) a través de la región $^4C_1$ - ^4E – $^4C_1$. Por el otro lado, esta enzima muestra promiscuidad a la hora de romper oligosacáridos con múltiples estereoquímicas (1,2)- , (1,3)-, (1,4)- y (1,6)-, lo que fue demostrado por estudios cinéticos (k_{cat} ~ 2-12 s^{-1}) en la referencia (Hrmova, 2002). Dicha promiscuidad es debida al hecho que en subsitio +1 está formado por dos triptófanos (Trp286 y Trp434) que estabilizan la posición del anillo +1 a partir de un sándwich de interacciones *π-stacking* y no hay residuos hidrofílicos que restrinjan la posición exacta de los alcoholes del azúcar y, por lo tanto, no diferencian entre las posibles estereoquímicas.

En este capítulo, presentamos simulacions a partir de cuatro nuevas estructuras cristalográficas de la enzima HvExoI en complejo con tioderivados de los sustratos naturales con las estereoquímicas anteriormente numeradas (complejos G2S [G = glucanasa, 2 = (1,2)-, S = tioderivado], G3S, G4S y G6S). La conformación del azúcar -1 de cada sistema es la siguiente: para los complejos G2S y G6S, la conformación es silla $^4C_1$, en el complejo G3S se encuentra en la forma ^4E y en el complejo G4S en la forma $^4H_3$. Dada la desavenencia en los resultados y las dudas que, a veces, pueden generar los tioderivados a la hora de mimetizar conformacionalmente al sustrato natural, el objetivo será reconstruir los complejos G2O, G3O, G4O y G6O (sustituyendo el azufre glicosídico por oxígeno), comparar el comportamiento de ambos azúcares y, finalmente, reconstruir la FEL conformacional del sustrato natural de cara a confirmar la conformación más estable para el anillo β-glucosil -1 dentro de la enzima HvExoI.

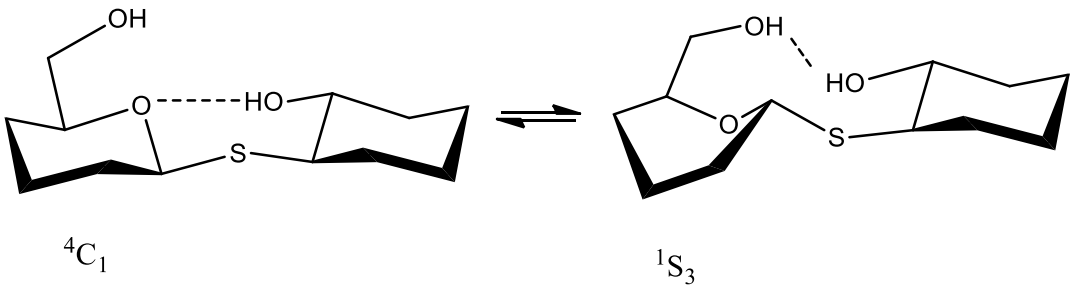
Capítulo V – Objetivos

1) Caracterizar las estructuras cristalográficas experimentales a 300 K, utilizando dinámica molecular clásica. Analizar el comportamiento conformacional de los tioglicósidos en el subsitio -1.
2) Reconstruir y caracterizar los modelos GH3 – sustrato natural a 300 K, utilizando dinámica molecular clásica. Analizar el comportamiento conformacional del azúcar en el subsitio -1.
3) Reconstruir la superficie de energía libre (FEL) conformacional del azúcar en el subsitio -1 en el model G30, utilizando metadinámica Well-T.

Capítulo V – Métodos, resultados y discusión

Llevando la misma estrategia mostrada en los trabajos anteriores, se reconstruyeron ocho modelos (G2S, G2O, G3S, G3O, G4S, G4O, G6S y G6O) solvatados y contrarrestadas las cargas superficiales con contraiones (Na^+). Se llevó a cabo la minimización, calentamiento a 300 K, estabilización de densidad y estructura a dicha temperatura. Una vez estabilizada la proteína, se alargó la simulación para seguir el comportamiento conformacional del azúcar una vez su entorno en el centro activo se mostraba inmutable.

Una vez analizadas las trayectorias de los cálculos, se observaron discrepancias entre el comportamiento conformacional de los tioderivados respecto al sustrato natural, aunque, en todos los casos la conformación más poblada siempre era la silla $^4C_1$. Dichas discrepancias se debían a la presencia del átomo de azufre entre los azúcares -1 y +1. En los tioderivados, la distancia (1.87 Å) entre los anillos unidos glicosídicamente era más larga que en los azúcares naturales (1.46 Å) y el ángulo C_1-S-C_N (N = 2, 3, 4, 6) era más cerrado (104º) que en los análogos con oxígeno (109º). Este hecho promocionaba que el alcohol 6-OH del azúcar -1 cambiara su orientación para interactuar con el alcohol cercano del azúcar +1, este cambio estaba correlacionado con el cambio conformacional de la silla $^4C_1$ a la barca distorsionada (Figura R6).



$^4C_1$ $^1S_3$

Figura R6. Representación esquemática de las interacciones puente de hidrógeno entre los azúcares -1 y +1 en disacáridos S- y O-enlazados. Schematic representation of hydrogen bonds between the -1 and the +1 subsite sugars in S- and O-linked disaccharide. Dichas interacciones dependen de la conformación adoptada por el azúcar -1.

Como todos los modelos GxO mostraban el mismo patrón conformacional, decidimos estudiar uno de dichos modelos (G3O) utilizando un método más preciso (metadinámica sobre un model QM/MM) para reconstruir la FEL conformacional del anillo glucosil -1 dentro de la enzima HvExoI.

La FEL conformacional resultante del sustrato natural (G3O) de la enzima GH3 se muestra en la Figura R8. Los resultados muestran una única conformación estable alrededor del punto (270º, 20º), región asociada a la conformación esperada $^4C_1$. El sistema evolucionó hacia la región distorsionada $^4H_3$ / E_3 / $^1S_3$ pero, el azúcar no encontró la estabilidad en dicha forma. Dado que normalmente, los azúcares suelen explorar dos mínimos estables (como mínimo) dentro de los centros activos, tratamos de simular un caso hipotético donde el azúcar -1 estuviera en la conformación $^1S_3$. No obstante, el sistema volvía rápidamente al mínimo de la conformación no distorsionada.

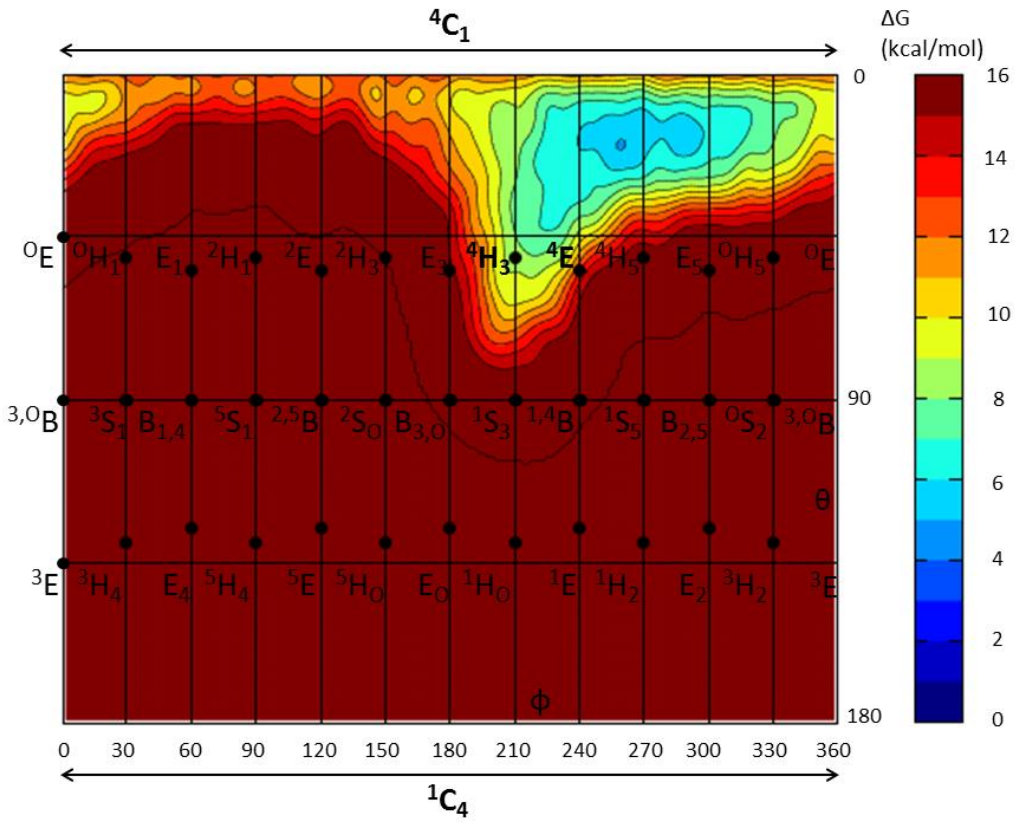


Figure R7. FEL conformacional (Proyección de Mercator) del anillo de β-glucosil en el subsitio -1 del complejo G3O de la enzima HvExoI. Cada isolínea está separada por 1 kcal/mol.

Por lo tanto, confirmamos que, al contrario de lo que se espera para una β-glucosidasa, pero, en acuerdo con los resultados experimentales de las referencias (Hrmova, 2001b y 2002), la conformación del azúcar -1 en el MC con la enzima HvExoI es la silla $^4C_1$, y el itinerario conformacional se presume cíclico (como en el caso de la amilosucrasa).

Capítulo V – Conclusiones

1. Hemos caracterizado las estructuras cristalográficas experimentales utilizando MD clásica y simulado los complejos de la enzima HvExoI con los sustratos naturales.
2. Los compuestos tioderivados y los azúcares naturales mostraron la conformación $^4C_1$ como la región más poblada durante las simulaciones de MD clásica, pero, los tiocompuestos eran capaces de distorsionar sus anillos -1. Este hecho explica por

qué las conformaciones de las estructuras G3S y G4S están en las conformaciones distorsionadas ^4E y $^4H_3$, respectivamente.

3. El estudio conformacional del grupo β-glucosil -1 dentro de la enzima HvExoI confirmó que la conformación silla era la más estable, fortaleciendo la hipótesis del itinerario cíclico observado experimentalmente para la hidrólisis ($^4C_1$ - ^4E - $^4C_1$).

Capítulo VI – Observaciones computacionales de la reactividad y la naturaleza conformacional de septanósidos en glicosil hidrolasas

En el diseño de inhibidores para GHs, el desarrollo de mímicos estructurales de azúcares es una de las estrategias más comunes utilizadas experimentalmente (Gruner, 2002) (Ganem, 1996) (Asano, 2009). En los últimos años, un nuevo grupo de azúcares artificiales fueron sintetizados (Bozó, 2000) (DeMatteo, 2005) (Saha, 2011) (Vannam, 2013) para probar sus capacidades como sustratos/inhibidores de GHs (Tauss, 2006). El método utilizado para mimetizar carbohidratos naturales es a través de la expansión del tamaño de sus anillos (Peczuh, 2003). Las versiones de las hexosas con anillos expandidos se conocen como septanosas o septanósidos. Como el nombre sugiere, estas moléculas están formadas por anillos de 7 átomos.

Dado que la molécula más simple que está formada por un anillo de dichas dimensiones es el cicloheptano, hemos decidido estudiar y utilizar como test el comportamiento conformacional de dicha molécula, a la hora de comprobar que nuestro método (metadinámica Well-T con cuatro coordenadas de empaquetamiento a 300 K) describe apropiadamente la riqueza conformacional de anillos de 7 átomos. Una vez confirmado dicho propósito, se han estudiado tres de los septanósidos más utilizados y que se conoce bien su método de obtención: α-D-glicero-D-idoseptanósido (α-D-idosepta), β-D-glicero-D-guloseptanósido (β-D-gulosepta) y β-L-idoseptanósido (β-L-idosepta). De cara a conocer las posibles conformaciones que puede adoptar el anillo en el estado de transición dentro de una GH, se ha reconstruido también la FEL conformacional del catión oxocarbenio septanósido (oxoseptaion, $[C_6H_{11}O]^+$). Una vez conocida la preactivación de las conformaciones observadas en los azúcares sintéticos aislados, se han construido cuatro modelos GH-septanósido: GH31 α-glucosidasa en complejo con pNP-β-L-septa (emulando los experimentos de la referencia [Tauss, 2006],

donde se observaba que la k_{cat} disminuía dos órdenes de magnitud y la K_M empeoraba en un orden de magnitud), GH13 α-glucosidasa en complejo con α-D-idosepta-(1,4)-α-glucosa y α-D-gulosepta-(1,4)-α-glucosa y GH1 β-glucosidasa en complejo con β-D-gulosepta-(1,3)-β-glucosa. Para finalizar, se estudió el comportamiento conformacional del pNP-β-L-septa en la GH31 y se modelizó la reacción de glicosilación en dicho sistema y en el GH13- α-D-gulosepta-(1,4)-α-glucosa que se preveía, pudiera ser un caso positivo de actividad enzima-sustrato.

Capítulo VI – Objetivos

1) Reconstruir la FEL conformacional del cicloheptano en fase gas. Cuantificar el efecto del método y la base con la que se describe la función de onda, comparando nuestros resultados con cálculos Gaussian 09 y estudios conformacionales previos.
2) Reconstruir la FEL conformacional de las moléculas en fase gas de α-D-idosepta, β-D-gulosepta y β-L-idosepta. Calcular el índice de preactivación para cada conformación estable de cada sistema.
3) Reconstruir la FEL conformacional del catión oxocarbenio septanósido en fase gas. Descubrir los posibles itinerarios conformacionales de un anillo de 7 miembros cuando es hidrolizado por una GH.
4) Simular dinámicamente las interacciones entre el 1-para-nitrophenyl – β-L-idosepta y una GH31 α-glucosidasa. Reconstruir la FEL conformacional del anillo de 7 átomos dentro del subsitio -1.
5) Simular la etapa de glicosilación de la reacción de hidrólisis del pNP - β-L-idosepta dentro de la GH31 α-glucosidasa.
6) Modelizar complejos GH-septanósidos con conformaciones del azúcar apropiadas para prever un comportamiento bueno o mal de los septanósidos como sustratos de glucosidasas y manosidasas.
7) Simular la etapa de glicosilación de la reacción de hidrólisis del the α-D-gulosepta-1,4-α-D-glucopiranosa dentro de la GH13 α-glucosidasa.

Capítulo VI – Métodos, resultados y discusión

Los cálculos previos para simular el cambio pseudorotacional del cicloheptano usando diferentes bases (PBE/6-31G*, PBE/cc-PVDZ, PBE/cc-PVTZ) fueron llevados a

cabo con el método IRC implementado en el programa Gaussian 09. A partir de esas estructuras optimizadas pudimos calcular las contribuciones energéticas usando ondas planas (PBE/PW) de 70 Ry con el programa CPMD.

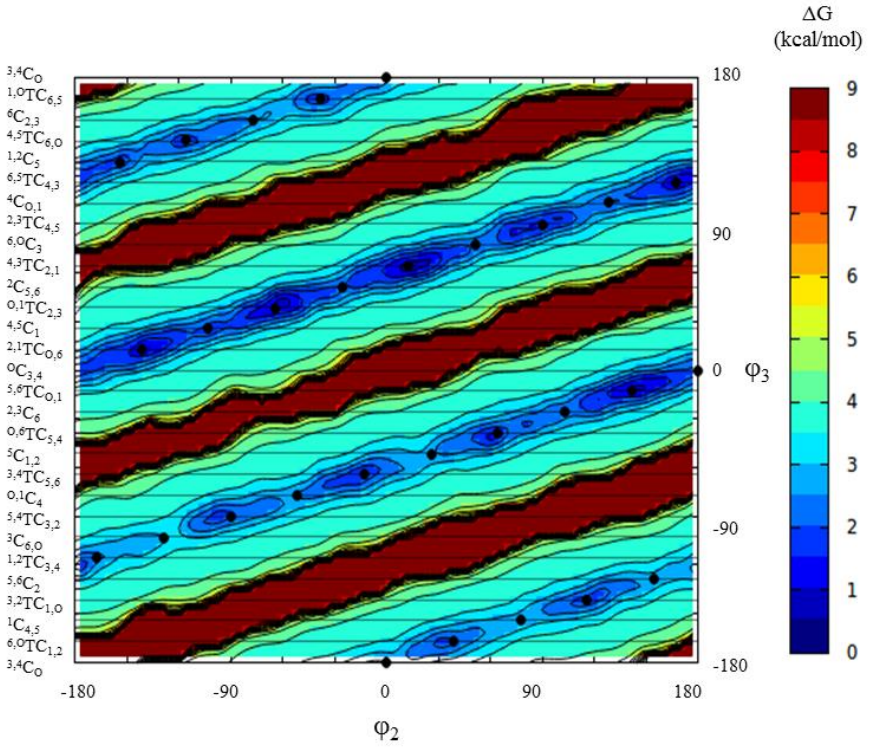
Las simulaciones para reconstruir las FEL conformacionales del of cicloheptano, α-D-idosepta, β-D-gulosepta, β-L-idosepta y del catión oxocarbenio septanósido, a 300 K, fueron llevadas a cabo con CPMD 3.15.1 parcheado con librerías de Plumed 2 y una versión propia de las coordenadas de empaquetamiento como variables colectivas para anillos de 7 miembros.

Debido a la ausencia de estructuras cristalográficas de complejos GH-septanósido, el modelo del complejo GH31 - pNP-β-L-idosepta fue construido a partir de una simulación *docking* utilizando el programa AutoDock Vina. En cambio, dado los malos resultados de este cálculo con el complejo GH13 - α-D-gulosepta-1,4-α-D-glucopiranosa, éste y los complejos con la GH1 fueron reconstruidos manualmente a partir de la edición de las coordenadas de los PDBs disponibles de los MC con sus sustratos naturales. Es decir, a partir del complejo GH13 – maltosa, se expandió el anillo del azúcar -1 para que fuera un α-D-guloseptanósido.

Una vez construidos los modelos enzima-sustrato, se llevó a cabo la misma metodología seguida en los capítulos anteriores: minimización y estabilización a 300 K usando MM MD, dinámica QM/MM con CPMD y Well-T. MTD para simular las FEL conformacionales (usando 4 coordenadas de empaquetamiento de Cremer y Pople) y de la reacción de glicosilación de los azúcares dentro de las enzimas.

Debido a la simetría que presenta la molécula de cicloheptano, la superficie conformacional de esta molécula debería mostrarnos catorce mínimos equivalentes en las regiones que representan las conformaciones *twisted-chair* (TC). Una vez convergida nuestra simulación Well-T. MTD utilizando las cuatro coordenadas de empaquetamiento de Cremer y Pople (φ_2, φ_3, q_2, q_3), la FEL resultante se muestra en la Figura R8, donde se observan las proyecciones (φ_2, φ_3) en los planos $q_3 = 0.65$ (TC/C) y $q_3 = 0.05$ (TB/B). En la primera proyección se pueden distinguir los 14 mínimos sobre las regiones TC, mientras que en el plano TB/B cuesta distinguir energéticamente las formas TB de las B, ya que, a nivel de energía potencial, ya se preveían diferencias energéticas por debajo de 1 kcal/mol. La desviación estándar entre las 14 conformaciones TC es de 0.46 kcal/mol, la diferencia de energía entre los mínimos TC y los estados de transición C es de 1.12 ±

0.38 kcal/mol y entre TC y el plano TB/B es de 3.43 ± 0.25 kcal/mol. La barrera cinética observada para pasar de TC a TB es de 4.66 ± 0.46 kcal/mol.



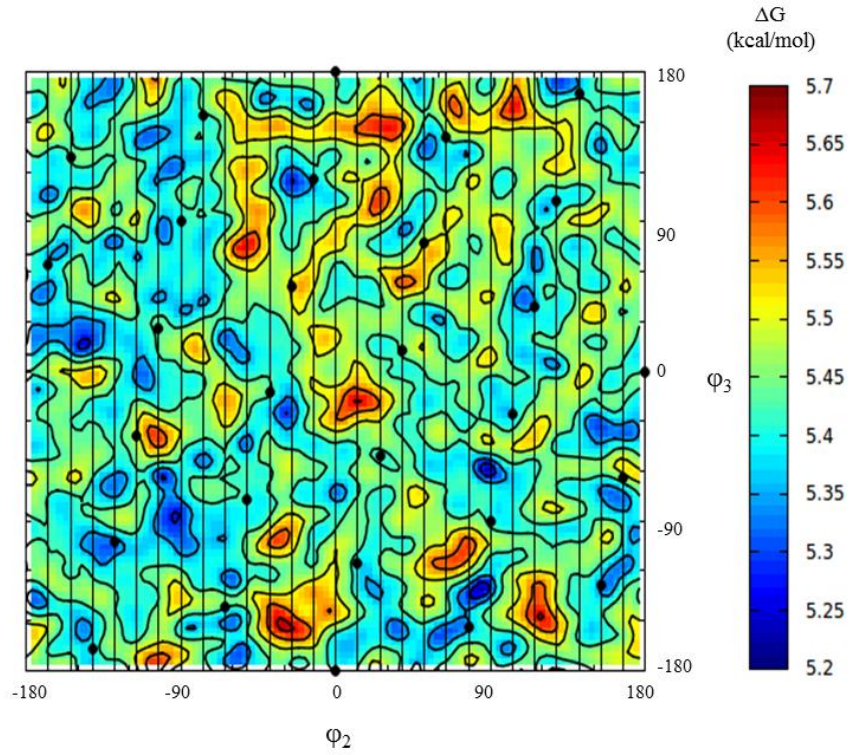


Figure R8. Proyecciones (φ_2, φ_3) del (arriba) plano TC/C (Isolíneas cada 0.5 kcal/mol) y (abajo) el plano TB/B (Isolíneas cada 0.1 kcal/mol) de la FEL conformacional del cicloheptano.

El primer y principal punto a demostrar en este estudio es el buen comportamiento de nuestra metodología a la hora de explorar el espacio conformacional de anillos de 7 miembros. La comparación de nuestros resultados obtenidos con Well-T MTD y los cálculos de la superficie de energía potencial, tanto los nuestros como los trabajos de (Bocian, 1975) y (Freeman, 2008), se muestran en la Tabla R2.

Table R2. Comparación de diferentes estudios conformacionales de la molécula de cicloheptano en fase gas. Energies in kcal/mol.

Conf.	MTD/ PBE	Bocian FF	CCSD/ cc-PVDZ (Freeman)	PBE/ cc-PVDZ	PBE/ cc-PVTZ	PBE/ PW
TC	0.00 ± 0.46	0.00	0.00	0.00	0.00	0.00
C	1.12 ± 0.38	1.30	-	-	-	-
TB	3.43 ± 0.25	3.39	3.39	3.43	3.25	3.35
B	3.43 ± 0.25	3.42	-	-	-	-
TS	4.66 ± 0.46	9.67	8.55	8.47	7.93	8.03

Analizando la tabla podemos observar la semejanza energética de nuestros resultados con los obtenidos en la referencia (Bocian, 1975), mientras que, donde se observa una discrepancia mayor entre los cálculos de energía potencial y energía libre es en la barrera cinética para pasar de TC a TB, pasando por TS (5 vs. 7-10 kcal/mol). Teniendo en cuenta el valor obtenido utilizando el método PBE/PW (8.03 kcal/mol), la entropía del cambio conformacional en el cicloheptano a 300 K es 11.23 cal/mol·K. En 1956, Finke y col. observaron experimentalmente que la entropía estándar del cicloheptano (g) era de 81.82 cal/mol·K (Finke, 1956). Así pues, parece un resultado razonable que una octava parte de la contribución entrópica del cicloheptano proceda de los grados de libertad conformacionales.

Una vez comprobada la validez del método en un sistema sencillo, la reconstrucción de las FEL conformacionales de los septanósidos α-D-idosepta y β-D-gulosepta nos

servirán tanto como test (pues son comparables a los estudios conformacionales de sus análogos metil α-D-idosepta y metil β-D-gulosepta de la referencia [DeMatteo, 2005]) como para conocer sus conformaciones más estables y las más preactivadas de cara a la reactividad en GHs. Por un lado, el metil α-D-idosepta muestra dos mínimos importantes en la regiones $^{3,4}TC_{5,6}$ (0.00 kcal/mol) y $^{1,2}TC_{3,4}$ (3.20 kcal/mol). Nuestra simulación muestra que dichas conformaciones son las más importantes para la molécula hidrolizada, $^{3,4}TC_{5,6}$ (0.00 kcal/mol) y $^{1,2}TC_{3,4}$ (2.25 kcal/mol). Por el otro lado, el metil β-D-gulosepta muestra cinco conformaciones relevantes: $^{O,6}TC_{5,4}$ (0.00 kcal/mol), $^{5}C_{1,2}$ (2.02 kcal/mol), $^{5,6}TC_{O,1}$ (2.31 kcal/mol), $^{1,2}TC_{3,4}$ (3.46 kcal/mol) y $^{3,4}TC_{5,6}$ (4.47 kcal/mol). Nuestro análogo hidrolizado muestra el mismo mínimo de máxima estabilidad ($^{O,6}TC_{5,4}$), una región plana en el espacio $^{5}C_{1,2}$ (2.50 kcal/mol) y mínimos en $^{5,6}TC_{O,1}$ (2.75 kcal/mol), $^{1,2}TC_{3,4}$ (2.72 kcal/mol) y $^{3,4}TC_{5,6}$ (2.82 kcal/mol). Esta comparación da mayor viabilidad y consistencia a nuestras simulaciones. Una vez calculados los índices de preactivación se observó que para la molécula de α-D-idosepta, la conformación más preactivada coincide con la más estable ($^{3,4}TC_{5,6}$, $\zeta = 75$) y para la molécula de β-D-gulosepta, las conformaciones más preactivadas son la $^{2,(5,6)}B$ ($\zeta = 69$-66) y la $^{5,4}TC_{3,2}$ ($\zeta = 68$).

De cara a contemplar la posible correlación conformacional entre los septanósidos en disolución (aislados) o en medio enzimático, se estudió la molécula de β-L-idosepta, tanto en fase gas como en forma de pNP - β-L-idosepta, dentro de la GH31 α-glucosidasa (de cara a comprender los experimentos de Tauss y col., 2006). En fase gas, la molécula mostraba 18 mínimos sobre la FEL conformacional, siendo la conformación $^{4,5}TC_{6,O}$, la más estable, seguida por la $^{6,5}TC_{4,3}$ (0.37 kcal/mol), $^{O,1}TC_{2,3}$ (0.38 kcal/mol) y la $^{3,4}TC_{5,6}$ (0.52 kcal/mol). No obstante, de esos 18 mínimos, los dos que muestran mayor índice de preactivación son la forma $^{4}C_{O,1}$ ($\zeta = 78$) y la $^{2,3}TC_{4,5}$ ($\zeta = 75$). Para llevar a cabo el docking, se escogió una conformación para el septanósido parecida a la $^{4}C_{1}$ de la glucosa, esta fue la $^{4,3}TC_{2,1}$ y se añadió el grupo para-nitrofenil manualmente. Tras la estabilización a 300 K con MM MD, la conformación del septanósido evolucionó hacia la forma $^{O,1}TC_{2,3}$ ($\zeta = 58$). Al calcular la FEL conformacional del anillo de β-L-idosepta en el interior de la enzima, se confirma que esa conformación es la más estable. A pesar de ser la tercera conformación más estable de la molécula en fase gas, ésta no es de las más preactivadas, no obstante, la correlación entre molécula aislada y en medio enzimático se sigue observando para azúcares de 7 átomos en su anillo. Analizando la FEL, el itinerario

conformacional previsto sería el que transcurre en la región $^{O,1}TC_{2,3}/^{4,5}C_1 \rightarrow {}^{5,4}TS_3 \rightarrow B_{3(6,O)}$.

Una vez encontrada la conformación de la que partirá la reacción, simulamos la etapa de glicosilación en la enzima GH31. La barrera de activación resultante al analizar la FEL de la reacción era de 28.6 kcal/mol, lo que nos indicaba que la reacción no se daría cuantitativamente a temperatura ambiente, tal y como se observaba experimentalmente (Tauss, 2006), donde la constante catalítica disminuía dos órdenes de magnitud respecto a la maltosa (sustrato natural de la enzima). Además, a pesar de la energía añadida al sistema, el nucleófilo no lograba formar un enlace covalente con el carbono anomérico del azúcar, por lo que, únicamente vimos la formación de intermedio catión oxocarbenio. Durante la dinámica molecular, observamos que la posición del nucleófilo cambiaba, para compensar la ausencia del brazo $–CH_2$-OH del azúcar y pasaba a interaccionar con el sustituyente 2-OH. Esta sería una razón importante por la cual la afinidad enzima-septanósido disminuye frente a el par enzima-maltosa (K_M). El itinerario conformacional seguido por el septanósido durante la reacción fue $^{O,1}TC_{2,3} \rightarrow {}^{O,1}S_3/^{5,4}TS_3 \rightarrow {}^{O,1}S_3/B_{(2,3),6}$.

Para comprender mejor el comportamiento conformacional del estado de transición en una reacción enzimática GH-septanósido, se simuló la FEL conformacional de la forma más simple de catión oxocarbenio septanósido. Las formas más estables encontradas fueron sillas $^4C_{O,1}$ y $^{O,1}C_4$. Seguidas en estabilidad por las conformaciones $^3TS_{4,5}$ (1.73 kcal/mol), $^{5,4}TS_3$ (3.10-3.31 kcal/mol), $^{3,2}TB_{1,O}$ (3.47 kcal/mol), $^{O,1}S_3$ (4.13 kcal/mol), $^{O,1}TB_{2,3}$ (4.30 kcal/mol), $^{3,4}BS$ (5.03 kcal/mol) y $BS_{3,4}$ (5.28 kcal/mol). El sistema evolucionó también por las regions $^2S_{O,6}$, $^{O,6}S_2$, $^3S_{O,1}$, $^{4,5}BS$ y $BS_{4,5}$.

Por lo tanto, analizando la información obtenida por las diferentes simulaciones presentadas, construimos un reloj conformacional esquemático para tener una visión global del comportamiento de las moléculas aisladas que estudiamos (Figura R9). Tal y como se observa en azúcares D y L de anillos de 6 miembros, observamos que se forman dos hemiciclos distinguidos entre los D-septanósidos y los L-septanósidos. No obstante, observamos que el L-septanósido también muestra población alrededor de las conformaciones $^{3,4}TC_{5,6}$ y $^{5,6}C_2$. Este hecho es debido a la ausencia de brazo $–CH_2$-OH de esta molécula, pues, el impedimento estérico que ejerce sobre los sustituyentes cercanos no existe en dicha molécula y aumenta la capacidad conformacional de ésta.

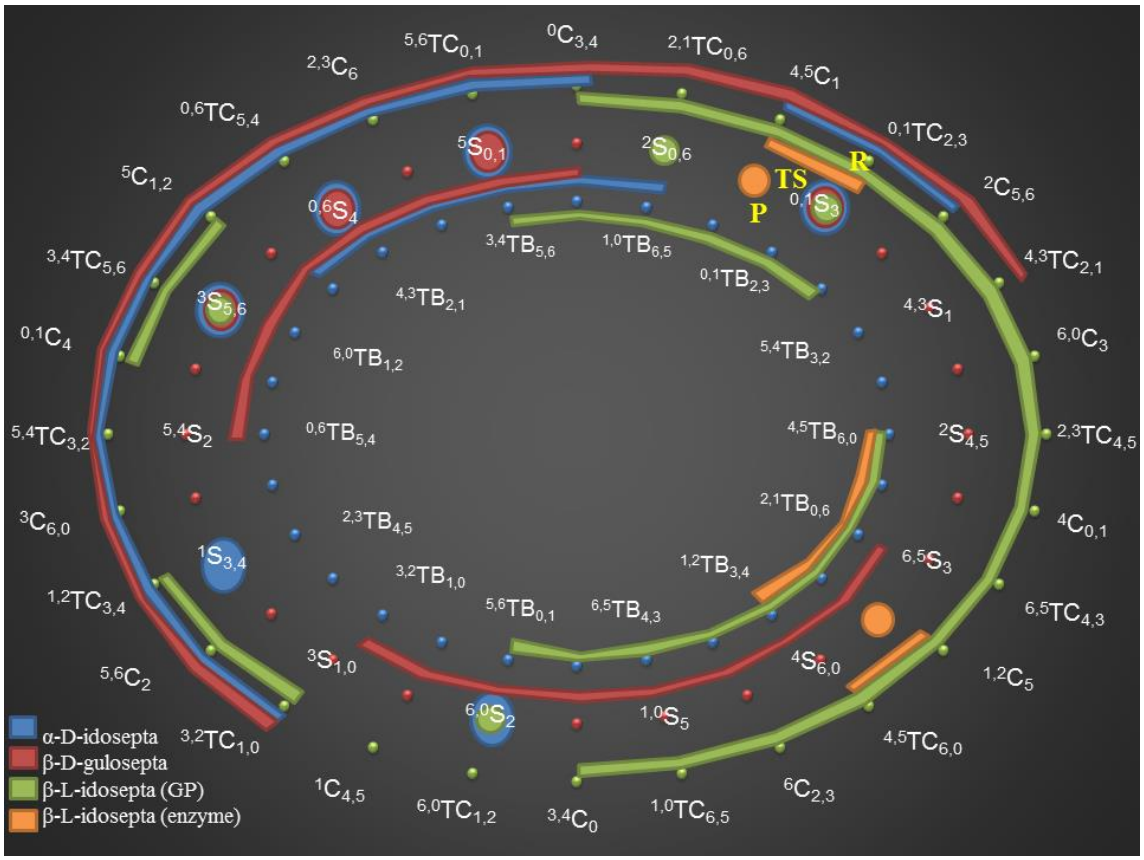


Figure R9. Reloj conformacional para anillos de 7 miembros. Las líneas coloreadas indicant las regiones más pobladas para cada uno de los septanósidos estudiados.

Para acabar, se llevaron a cabo la reconstrucción de tres modelos GH-septanósido. Una vez minimizados con el algoritmo del programa AMBER 11 se escogió el sistema que pudiera mostrar mejor capacidad catalítica. Debido a las interacciones enzima-sustrato y el buen posicionamiento de los residuos catalíticos se decidió simular la reacción de glicosilación para el sistema GH13 - α-D-gulosepta-1,4-α-D-glucopiranosa. El procedimiento seguido con este sistema fue el mismo que el llevado a cabo para la amilosucrasa (aprovechando que son enzimas de la misma familia). Una vez simulada la reacción dentro de la enzima observamos varios puntos interesantes y catalíticamente relevantes. En primer lugar, la barrera cinética era de 12 kcal/mol, es decir que la reacción no solo era cuantitativamente accesible a temperatura ambiente, sino que, la barrera es inferior a la encontrada, por ejemplo, en la reacción dentro de la amilosucrasa (17 kcal/mol). En la FEL de la reacción se encontraron mínimos estables en tres regiones: reactivos (Michaelis complex), un intermedio estable catiónico (catión oxocarbenio, 3.5 kcal/mol) y un intermedio covalente enzima-septanósido (GEI, 5.1-5.7 kcal/mol). A

diferencia de la mayoría de reacciones enzimáticas con azúcares naturales, la reacción con el septanósido muestra un intermedio catiónico estable, cuya formación está más favorecida que el intermedio covalente. La etapa de glicosilación del septanósido sigue por el itinerario conformacional $^{5,6}TC_{O,1} \rightarrow {}^{2}TS_{3,4}/BS_{3,4} \rightarrow BS_{3,4}/{}^{1}S_{3,4}/{}^{5,6}TB_{O,1}$, camino que muestra mucha similaridad al observado en α-glucosidasas con sus sustratos naturales ($^{4}C_1 \rightarrow E_3 \rightarrow {}^{1}S_3$).

Finalmente, en forma de propuesta para futuros trabajos experimentales, analizando las diferentes formas preactivadas para cada septanósido y viendo la semejanza de conformaciones entre septanósidos y el equivalente de 6 átomos, observamos posibles capacidades como buenos sustratos de: la molécula de α-D-idosepta como sustrato de α-manosidasas, la molécula de β-D-idosepta como sustrato de β-manosidasas y los "sin brazo" L-septanósidos o D-glucoseptanósidos como sustratos de β-xilanasas.

Capítulo VI – Conclusiones

1. La metadinámica Well-T usando las coordenadas de empaquetamiento de Cremer y Pople es un buen método para reconstruir las superficies de energía libre conformacional de anillos de 7 miembros, tanto en fase gas como en sistemas enzima-azúcar.
2. La barrera conformacional ($TC \rightarrow [S]^{\ddagger} \rightarrow TB$) del cicloheptano a 300 K (4.66 kcal/mol) es 3-4 kcal/mol más pequeña que la barrera de energía potencial (8-9 kcal/mol), resultando una contribución entrópica de 11 cal/mol·K.
3. La conformación más estable adoptada por la molécula de α-D-idoseptanósido es la $^{3,4}TC_{5,6}$ y, a su vez, la más preactivada ($\zeta = 75$). La conformación más estable adoptada por la molécula de β-D-guloseptanósido es la $^{O,6}TC_{5,4}$, pero la más preactivada es la $^{5,4}TC_{3,2}$ ($\zeta = 68$).
4. Analógamente a las hexosas, el catión oxocarbenio septanósido puede adoptar un número fijo de conformacions (14), debido a la planaridad de los átomos C_2-C_1-O_6-C_6 (hibridación sp^2). En este caso, puede adoptar dos conformaciones silla ($^{4}C_{O,1}$ and $^{O,1}C_4$).
5. Existe una correlación entre el comportamiento conformacional del β-L-idoseptanósido en fase gas y complejado con la enzima GH31. La forma más estable en fase gas es la $^{4,5}TC_{6,O}$ y la más preactivada la $^{4}C_{O,1}$ ($\zeta = 78$) y la $^{2,3}TC_{4,5}$ ($\zeta = 75$). El azúcar dentro de la enzima adopta la conformación $^{O,1}TC_{2,3}$ ($\zeta = 58$).

6. La etapa de glicosilación del pNP – β-L-idoseptanósido dentro de la enzima GH31 muestra una barrera alta de activación (29 kcal/mol), donde la especie catiónica (oxocarbenio) es un mínimo estable en la FEL. El itinerario conformacional es el $^{O,1}TC_{2,3} \rightarrow {}^{O,1}S_3/{}^{5,4}TS_3 \rightarrow {}^{O,1}S_3/B_{(2,3),6.}$
7. Hemos sido capaces de construir estructuras estables glucosidasa-septanósido, utilizando docking y métodos de ingeniería molecular.
8. La etapa de glicosilación de la α-D-gulosepta-1,4-α-glucosa dentro de la enzima GH13 muestra una barrera baja de activación (12 kcal/mol), donde el catión oxocarbenio es una especie estable (3.5 kcal/mol) y el intermedio covalente se forma con un gasto energético de 5 kcal/mol. El itinerario conformacional seguido por el azúcar -1 es el $^{5,6}TC_{O,1} \rightarrow {}^{2}TS_{3,4}/BS_{3,4} \rightarrow BS_{3,4}/{}^{1}S_{3,4}/{}^{5,6}TB_{O,1}$, similar al esperado para una α-glucosa dentro de una α-glucosidasa.
9. Teniendo en cuenta los resultados de parámetros de preactivación y modelización de reacciones, consideramos la utilización de α-D-idosepta como sustrato de α-manosidasas, de α-D-gulosepta como sustrato de α-glucosidasas, de β-D-idosepta como sustrato de β-manosidasas y L-idosepta o D-glucoseptanosides “sin brazo” como sustrato de xilanasas.

Publications and presentations in congresses

The present Thesis has led to the following publications:

(1) S. Alonso-Gil, A. Males, P. Fernandes, S. J. Williams, G. J. Davies, C. Rovira. Computational design-of-experiment unveils the conformational reaction coordinate of GH125 α-mannosidases. *J. Am. Chem. Soc.* **2017**, 139 (3), 1085 – 1088.

(2) S. Alonso-Gil, J. Iglesias-Fernández, M. Remaud-Simeon, I. André, C. Rovira. Conformational itinerary and catalytic mechanism of GH13 amylosucrases. In preparation.

(3) S. Luang, V. A. Streltsov, J. K. Cairns, S. Fort, S. Alonso-Gil, M. Alfonso-Prieto, C. Rovira, M. Hrmova. Tryptophan residues in a plant exohydrolase control flexible docking of sugars with different stereo-chemistries that underlie substrate specificity. In preparation.

(4) S. Alonso-Gil, A. Planas, C. Rovira. Computational insights into the conformational dynamics and reactivity of septanosides in glycoside hydrolase. In preparation.

Some works presented in this Thesis were presented in the following congresses:

(1) C. Rovira, S. Alonso-Gil. "Reaction pathway and conformational study in GH125 exo-1,6-α-mannosidase: Is thio-mannobiose lying?" XXXI XRQTC Meeting 2015, Girona, June 25th-26th, 2015.

(2) S. Alonso-Gil, J. Iglesias-Fernández, M. Remaud-Simeon, I. André, C. Rovira. "Simulation of glycosylation process within the GH13 amylosucrase by QM/MM ab initio metadynamics." V IQTCUB Symposium, May 30th, 2012 (Poster session).

(3) C. Rovira, S. Alonso-Gil. "Reactivity insights inside the amylosucrase. When occurs the polymerization?" IQTCUB internal seminar, October 14th, 2016.

Bibliography

Theory and methodology

A. D. Becke, *J. Chem. Phys.* **1984**, 84, 4524 – 4529.

A. Barducci, G. Bussi, M. Parrinello, *Phys. Rev. Lett.* **2008**, 100, 020603.

R. Car, M. Parrinello, *Phys. Rev. Lett.* **1985**, 55, 2471 – 2474.

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

D. Cremer, J. A. Pople, *J. Am. Chem. Soc.* **1975**, 97, 1354 – 1358.

P. Deslongchamps, *Tetrahedron.* **1975**, 31, 2463 – 2490.

R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1971**, 54, 724.

T. H. Dunning, *J. Chem. Phys.* **1989,** 90 (2), 1007 – 1023.

P. A. M. Dirac, *Proc. Cambridge Phil. Roy. Soc.* **1930**, 26 (3), 376 – 385.

B. Ensing, A. Laio, M. Parrinello, M. L. Klein, *J. Phys. Chem.* **2005**, 109, 6676 – 6687.

V. Fock, *Z. Phys.* **1930**, 61, 126 – 148.

K. Fui, *Acc. Chem. Res.* **1981**, 14, 363 – 368.

R. D. Guthrie, W. P. Jencks, *Acc. Chem. Res.* **1989**, 22, 10, 343 – 349.

D. R. Hartree, *Math. Proc. Camb. Philos. Soc.* **1928**, 24, 89–132, 426 – 437.

R. W. Hockney, *Meth. Comp. Phys.* **1970**, 9, 136 – 211.

P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, 136, 864 – 871.

M. Iannuzzi, A. Laio, M. Parrinello, *Phys. Rev. Lett.* **2003**, 90, 238302.

J. Ihm, A. Zunger, M. L. Cohen, *J. Phys. C.* **1979**, 12, 4409.

W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, 79, 926 – 935.

A. J. Kirby, *Acc. Chem. Res.* **1984**, 17, 305 – 311.

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

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

A. Laio, F. L. Gervasio, *Reports on Progress in Physics*. **2008**, 47, 126601.

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

A. Laio, A. Rodríguez-Fortea, F. L. Gervasio, M. Ceccarelli, M. Parrinello, *J. Phys. Chem. B*. **2005**, 109, 6714 – 6721.

A. Laio, J. VandeVondele, U. Röthlisberger, *J. Chem. Phys.* **2002**, 116, 6941 – 6947. (b)

A. Laio, J. VandeVondele, U. Röthlisberger, *J. Phys. Chem.* **2002**, 106, 7300 – 7307. (c)

C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B.* **1988**, 37, 785 – 789.

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

S. Nosé, *Molecular Physics*. **1984**, 52, 255 – 268.

J. C. Paniagua, P. Alemany, Química Quàntica (Vol. 2), **2000**, University of Barcelona.

J. P. Perdew, Phys. Rev. B. **1986**, 33, 8822 – 8824.

J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, 77, 3865 – 3868.

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

V. L. Schramm, W. Shi, *Current Opinion in Structural Biology*. **2001**, 11, 657 – 665.

M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, *J. Phys, Chem.* **1996**, 100, 19357 – 19363.

A. Szabo, N. S. Ostlund, Modern Quantum Chemistry, Introduction to Advanced Electronic Structure Theory, **1989**, 1st ed., rev. New York: McGraw-Hill.

P. Tiwary, M. Parrinello, *J. Phys. Chem. B*. **2015**, 119 (3), 736 – 742

N. Troullier, J. L. Martins, *Condensed Matter*, **1991**, 43, 1993 – 2006.

L. Verlet, *Physical Review.* **1967**, 159, 98 – 103.

S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, 58 (8), 1200.

A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, 103, 227 – 249.

Amylosucrase

C. Albenne, L. K. Skov, O. Mirza, M. Gajhede, G. Potocki-Véronèse, P. Monsan, M. Remaud-Simeon, *FEBS Lett.* **2002**, 527, 67 – 70.

I. André, G. Potocki-Veronese, S. Morel, P. Monsan, M. Remaud-Simeon, *Top Curr. Chem.* **2010**, 294, 25 – 48.

E. Cambon, S. Barbe, S. Pizzut-Serin, M. Remaud-Simeon, I. André, *Biotechnology and Bioengineering.* **2014**, 111, 1719 – 1728.

D. Daudé, I. André, P. Monsan, M. Remaud-Simeon. Chapter 8: Successes in engineering glucansucrases to enhance glycodiversification. *Carbohydrate chemistry*. **2014**, Vol. 40.

S. Emond, S. Mondeil, K. Jaziri, I. André, P. Monsan, M. Remaud-Simeon, G. Potocki-Véronèse. *FEMS Microbiol. Let.* **2008**, 285, 25 – 32. (b)

C. Moulis, I. André, M. Remaud-Simeon, *Cellular and Molecular Life Sciences*, **2016**, 1 – 19.

G. Potocki de Montalk, M. Remaud-Simeon, R.-M. Willemot, P. Sarçabal, V. Planchot, . Monsan, *FEBS Let.* **2000**, 471, 219 – 223.

M. Remaud-Simeon, F. Albaret, B. Canard, I. Varlet, P. Colonna, R. M. Willemot, P. Monsan, Studies on a recombinant amylosucrase. In: S. B. Petersen, B. Svensson, S. Pedersen editors. *Carbohydrate bioengineering.* Amsterdam, The Netherlands: Elsevier Science B.V. **1995**, 313 – 320.

J. Y. Riou, M. Guibourdenche , M.Y. Popoff, *Ann. Microbiol. (Inst. Pasteur)* **1983**, 134B, 257 – 267.

A. Verges, E. Cambon, S. Barbe, C. Moulis, M. Remaud-Simeon, I. André, *Protein Science.* **2017**, 26, 566 – 577.

Reactivity of enzymes and sugars

A. Ardèvol, X. Biarnés, A. Planas, C. Rovira, *J. Am. Chem. Soc.* **2010**, 132, 16058 – 16065.

N. Asano, *Cell. Mol. Life Sci.* **2009**, 66, 1479 – 1492.

X. Biarnés, J. Nieto, A. Planas, C. Rovira, *J. Biol. Chem.* **2006**, 281, 1432 – 1441.

X. Biarnés, A. Ardèvol, J. Iglesias-Fernández, A. Planas, C. Rovira, *J. Am. Chem. Soc.* **2011**, 133, 20301 – 20309.

C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips, V. R. Sarma, *Nature*, **1965**, 206, 757 – 761.

H. Bischoff, "The mechanism of alpha-glucosidase inhibition in the management of diabetes*". Clin. Invest. Med.* **1995**, 18 (4), 303 – 311.

B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, *Nucleic acids research*. **2009**, 37, D233 – 8.

C. R. A. Catlow, *Angew. Chem. Int. Ed.* **2016**, 55, 9132 – 9133.

D. Chui, M. Oh-Eda, Y-F. Liao, K. Panneerselvam, K. W. Marek, H. H. Freeze, K. W. Moremen, M. N. Fukuda, J. D. Marth, "Alpha-Mannosidase-II deficiency results in Dyserythropoiesis and unveils an alternate pathway in oligosaccharide biosynthesis". *Cell*, **1997**, 90, 157 – 167.

G. J. Davies, B. Henrissat, *Structure.* **1995**, 3, 853 – 859.

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

G. J. Davies, K. S. Wilson, B. Henrissat, *Biochem. J.* **1997**, 321, 557 – 559.

L. Deng, P. Tsybina, K. J. Gregg, R. Mosi, W. F. Zandberg, A. B. Boraston, D. J. Vocadlo, *Bioorganic & Medicinal Chemistry*. **2013**, 21, 4839 – 4845.

S. J. Fox, J. Dziedzic, T. Fox, C. S. Tautermann, C.-K. Skylaris, *Proteins*. **2014**, 82, 3335 – 3346.

B. Ganem, *Acc. Chem. Res.* **1996**, 29, 7, 340 – 347.

T. M. Gloster, P. Melloncelli, R. V. Stick, D. L. Zechel, A. Vasella, G. J. Davies, *J. Am. Chem. Soc.* **2007**, 129, 2345 – 2354.

S. A. W. Gruner, E. Locardi, E. Lohof, H. Kessler, *Chem. Rev.* **2002**, 102, 2, 491 – 514.

T. D. Heightman, A. T. Vasella, *Angew. Chem. Int. Ed.* **1999**, 38, 750 – 770.

B. Henrissat, A. Bairoch, *Biochem. J*. **1996**, 316 (2), 695 – 696.

S. S. Hurtley, *Carbohydrates and Glycobiology, Special Issue. Science.* **2001**, 2263 – 2503.

J. Iglesias-Fernández, L. Raich, A. Ardèvol, C. Rovira, *Chem. Sci.* **2015**, 6, 1167 – 1177.

D. E. Koshland, Proc. *Natl. Acad. Sci. USA*, **1958**, 44 (2), 98 – 104.

R. E. Ley, D. A. Peterson, J. I. Gordon, *Cell*. **2006**, 124, 837 – 848.

L. Petersen, A. Ardèvol, C. Rovira, P. J. Reilly, *J. Phys. Chem. B.* **2009**, 113 (20), 7331 – 7339.

L. Petersen, A. Ardèvol, C. Rovira, P. J. Reilly, *J. Am. Chem. Soc.* **2010**, 132 (24), 8291 – 8300.

G. P. Pinto, N. F. Brás, M. A. S. Perez, P. A. Fernandes, N. Russo, M. J. Ramos, M. Toscano, *J. Chem. Theory Comput.* **2015**, 11, 2508 – 2516.

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

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

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

D. J. Vocadlo, G. J. Davies, *Curr. Opin. Chem. Biol.* **2008**, 12, 539 – 555.

R. Wolfenden, X. D. Lu, G. Young, *J. Am. Chem. Soc*. **1998**, 120, 6814 – 6815.

D. L. Zechel, S. G. Withers, Acc. Chem. Res. 2000, 33, 11 – 18.

Conformational studies of sugars

A. Ardèvol, C. Rovira, *J. Am. Chem. Soc*. **2015**, 127 (34), 7528 – 7547.

X. Biarnés, A. Ardèvol, A. Planas, C. Rovira, A. Laio, M. Parrinello, *J. Am. Chem. Soc.* **2007**, 129, 10686 – 10693.

J. Iglesias Fernández, PhD Thesis: *Elucidating catalytic mechanisms of glycoside hydrolases and transferases by means of ab initio molecular dynamics simulations.* Barcelona: Universitat de Barcelona. Departament de Química Orgànica. http://www.tdx.cat/handle/10803/283551

IUPAC-IUB Joint Comission on Biochemical Nomenclature (JCBN). *Eur. J. Biochem.* **1980**, 111, 295 – 298.

M. Petrisevic, L. F. Sobala, P. Z. Fernandes, L. Raich, A. J. Thompson, G. Bernardo-Seisdedos, O. Millet, S. Zhu, M. Sollogoub, J. Jiménez-Barbero, C. Rovira, G. J. Davies, S. J. Williams, *J. Am. Chem. Soc.* **2017**, 139, 1089 – 1097.

V. Spiwok, B. Králová, I. Tvaroška, *Carbohydrate Research*. **2010**, 335, 530 – 537.

A. J. Thompson, J. Dabin, J. Iglesias-Fernandez, A.Ardevol, Z. Dinev, S. J. Williams, O. Bande, A. Siriwardena, C. Moreland, T.Hu, D. K. Smith, H. J. Gilbert, C. Rovira, G. J. Davies, *Angew. Chem. Int. Ed.* **2012**, 51, 10997 – 11000. (a)

A. J. Thompson, G. Speciale, J. Iglesias-Fernández, Z. Hakki, T. Belz, A.Cartmell, R. J. Spears, E. Chandler, M. J. Temple, J. Stepper, H. J. Gilbert, C. Rovira, S. J. Williams, G. J. Davies, *Angew. Chem. Int. Ed.* **2015**, 54, 5378 – 5382.

A. J. Thompson, R. J. Williams, Z. Hakki, D. S. Alonzi, T. Wennekes, T. M. Gloster, K. Songsrirote, J. E. Thomas-Oates, T. M. Wrodnigg, J. Spreitz, A. E. Stütz, T. D. Butters, S. J. Williams, G. J. Davies, *Proc. Natl. Acad. Sci.* **2012**, 109, 3, 781 – 786. (b)

R. J. Williams, J. Iglesias-Fernández, J. Stepper, A. Jackson, A. J. Thompson, E. C. Lowe, J. M. Lowe, J. M. White, J. Gilbert, C. Rovira, G. J. Davies, S. J. Williams, *Angew. Chem. Int. Ed. Engl.* **2014**, 53, 4, 1087 – 1091.

Y. Yin, M. Petricevic, A. John, L. Raich, H. Jenkins, L. P. De Souza, F. Cuskin, H. J. Gilbert, C. Rovira, E. D. Goddard-Borger, S. J. Williams, G. J. Davies, *ACS Central Science*. **2016**, 2, 896 – 903.

Septanosides.

D. F. Bocian, H. M. Pickett, T. C. Rounds, H. L. Strauss, *J. Am. Chem. Soc.* **1975**, 97, 4, 687 – 695.

I. K. Boessenkool, J. C. A. Boeyens, *J. Crys. Mol. Struc.* **1980**, 10, 11 – 18.

É. Bozó, S. Boros, L. Párkányi, J. Kuszmann, *Carbohydrate Res.* **2000**, 329, 269 – 286.

S. Castro, M. Duff, N. Snyder, M. Morton, C. V. Kumar, M. W. Peczuh, *Org. Biomol. Chem.* **2005**, 3, 3869 – 3872.

M. DeMatteo, N. L. Snyder, M. Morton, D. M. Baldisseri, C. M. Hadad, M. W. Peczuh, *J. Org. Chem.* **2005**, 70, 24 – 38.

M. R. Duff, W. F. Fyvie, S. D. Markad, A. E. Frankel, C. V. Kumar, J. A. Gascón, M. W. Peczuh, *Org. Biomol. Chem.* **2011**, 9, 154 – 164.

H. L. Finke, D. W. Scott, M. E. Gross, J. F. Messerly, G. Waddington, *J. Am. Chem. Soc.* **1956**, 78, 5469 – 5476.

F. Freeman, J. H. Hwang, E. H. Hunge, P. D. Parmar, Z. Renz, J. Trinh, *Int. J. Quantum Chem.* **2008**, 108, 339 – 350.

M. W. Peczuh, N. L. Snyder, *Tetrahedron Lett*. **2003**, 44, 4057 – 4061.

J. Saha, M. W. Peczuh, *Chem. Eur. J.* **2011**, 17, 7357 – 7365.

A. Tauss, A. J. Steiner, A. E. Stütz, C. A. Tarling, S. G. Withers, T. M. Wrodnigg, *Tetrahedron: Asymmetry*. **2006**, 17, 234 – 239.

R. Vannam, M. W. Peczuh, *Org. Lett.* **2013**, 15, 16, 4122 – 4125.

R. Vannam, M. W. Peczuh, *Eur. J. Org. Chem.* **2016**, 1800 – 1812.

Crystallographic structures.

A. Abe, H. Yoshida, T. Tonozuka, Y. Sakano, S. Kamitori, *Febs J.* **2005**, 272, 6145 – 6153.

G. D. Brayer, G. Sidhu, R. Maurus, E. H. Rydberg, C. Braun, Y. Wang, N. T. Nguyen, C. M. Overall, S. G. Withers. *Biochemistry.* **2000**, 39, 4778 – 4791.

A. M. Brzozowski, D. M. Lawson, J. P. Turkenburg, H. Bisgaard-Frantzen, A. Svendsen, T. V. Borchert, Z. Dauter, K. S. Wilson, G. J. Davies. *Biochemistry.* **2000**, 39, 9099 – 9107.

S. Caner, X. Zhang, J. Jiang, H. M. Cheng, N. T. Nguyen, H. Overkleeft, G. D. Brayer, S. G. Withers. *FEBS Lett.* **2016**, 590, 1143 – 1151.

K. J. Gregg, W. F. Zandberg, J. H. Hehemann, G. E. Whitworth, L. Deng, D. J. Vocadlo, A. B. Boraston, *J. Biol. Chem.* **2011**, 286, 15586 – 15596.

H. Hondoh, W. Saburi, H. Mori, M. Okuyama, T. Nakada, Y. Matsuura, A. Kimura, *J. Mol. Biol.* **2008**, 378, 911 – 920.

M. H. Jensen, O. Mirza, C. Albenne, M. Remaud-Simeon, P. Monsan, M. Gajhede, L. K. Skov, *Biochemistry.* **2004**, 43, 3104 – 3110.

J. Larsbrink, A. Izumi, G. R. Hemsworth, G. J. Davies, H. Brumer, *J. Biol. Chem.* **2012**, 287, 43288.

B. Mikami, H. Iwamoto, D. Malle, H.-J. Yoon, E. Demirkan-Sarikaya, Y. Mezaki, Y., Katsuya, *J. Mol. Biol.* **2006**, 359, 690 – 707.

O. Mirza, L. K. Skov, M. Remaud-Simeon, G. Potocki de Montalk, C. Albenne, P. Monsan, M. Gajhede, *Biochemistry.* **2001**, 40, 9032 – 9039.

O. Mirza, L. K. Skov, D. Sprogoe, L. A. M. van den Broek, G. Beldman, J. S. Kastrup, M. Gajhede, *J.Biol.Chem.* **2006**, 281, 35576 – 35584.

W. Moremen, Y. Xiang, A. Siriwardena, S. J. Williams, H. J. Gilbert, G. J. Davies, *Nature Chem. Biol.* **2010**, 6, 125 – 132.

W. A. Offen, D.L. Zechel, S. G. Withers, H. J. Gilbert, G. J. Davies, *Cell (Cambridge, Mass.)* **2009**, 18, 2484.

N. Okazaki, T. Tamada, M. D. Feese, M. Kato, Y. Miura, T. Komeda, K. Kobayashi, K. Kondo, M. Blaber, R. Kuroki, *Protein Sci.* **2012**, 21, 539 – 552.

L. M. Ren, X. H. Qin, X. F. Cao, L. L. Wang, F. Bai, G. Bai, Y. Shen, *Protein Cell.* **2011**, 2, 827 – 836.

T. Satoh, T. Toshimori, G. Yan, T. Yamaguchi, K. Kato, *Sci Rep.* **2016**, 6, 20575 – 20575.

S. Seshadri, T. Akiyama, R. Opassiri, B. Kuaprasert, J. R. K. Cairns, *Plant Physiol.* **2009**, 151, 47 – 58.

X. Shen, W. Saburi, Z. Gai, K. Kato, T. Ojima-Kato, J. Yu, K. Komoda, Y. Kido, H. Matsui, H. Mori, M. Yao, *Acta Crystallogr. Sect.D.* **2015**, 71, 1382 – 1391.

L. K. Skov, O. Mirza, A. Henriksen, G. P. De Montalk, M. Remaud-Simeon, P. Sarcabal, R. M. Willemot, P. Monsan, M. Gajhede, *J. Biol. Chem.* **2001**, 276, 25273 – 25278.

L. K. Skov, O. Mirza, D. Sprogoe, I. Dar, M. Remaud-Simeon, C. Albenne, P. Monsan, M. Gajhede, *J. Biol. Chem.* **2002**, 277, 47741 – 47747.

T. Tagami, K. Yamashita, M. Okuyama, H. Mori, M. Yao, A. Kimura, *J. Biol. Chem.* **2013**, 288, 19296 – 19303.

T.-C. Tan, B. N. Mijts, K. Swaminathan, B. K. C. Patel, C. Divne, *J. Mol. Biol.* **2008**, 378, 850 – 868.

K. Yamamoto, H. Miyake, M. Kusunoki, S. Osaki, *J. Biosci. Bioeng.* **2011**, 112, 545 – 550.

X. Zhang, S. Caner, E. Kwan, C. Li, G. D. Brayer, S. G. Withers. *Biochemistry.* **2016**, 55, 6000 – 6009.

R. Zhang, C. Li, L. K. Williams, B. P. Rempel, G. D. Brayer, S. G. Withers. *Biochemistry.* **2009**, 48, 10752 – 10764.

Y. Zhu, M. D. L. Suits, A. J. Thompson, S. Chavan, Z. Dinev, C. Dumon, N. Smith, K. *Nat. Chem. Biol.* **2010**, 6 (2), 125 – 132.

GH3 exo-β-glucanase.

A. J. Harvey, M. Hrmova, R. De Gori, J. N. Varghese, G. B. Fincher, *Proteins*, **2000**, 41, 257 – 269.

M. Hrmova, R. De Gori, B. J. Smith, J. K. Fairweather, H. Driguez, J. N. Varghese, G. B. Fincher, *The Plant Cell*. **2002**, 14, 1033 – 1052.

M. Hrmova, G. B. Fincher, *Carb. Res.* **2007**, 342, 1613 – 1623.

M. Hrmova, G. B. Fincher, *Plant Mol. Biol.* **2001**, 47, 73 – 91. (a)

M. Hrmova, A. J. Harvey, J. Wang, N. J. Shirley, G. P. Jones, B. A. Stone, P. B. Hoj, G. B. Fincher, *J. Biol. Chem.* **1996**, 271, 5277 – 5286.

M. Hrmova, V. A. Streltsov, B. J. Smith, A. Vasella, J. N. Varghese, G. B. Fincher, *Biochemistry*. **2005**, 44, 16529 – 16539.

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

M. Hrmova, J. N. Varghese, P. B. Hoj, G. B. Fincher, *Acta Cryst.* **1998**, 54, 687 – 689.

J. N. Varghese, M. Hrmova, G. B. Fincher, *Structure*. **1999**, 7, 179 – 190.

Software

Classical Molecular Dynamics

D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Shafmeister, W. S. Ross, P. A. Kollman, AMBER 9, **2006**, University of California, San Francisco.

D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, AMBER 11, **2011**, University of California, San Francisco.

Car-Parrinello Molecular Dynamics

R. Car, M. Parrinello, *Phys. Rev. Lett.* **1985**, 55, 2471 – 2474.

CPMD, http://www.cpmd.org/, Copyright IBM Corp 1990-2017, Copyright MPI für Festkörperforschung Stuttgart 1997-2001.

Metadynamics

G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi. **Plumed 2**: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, 185 (2), 604 – 613.

Programming language

ISO/IEC. (2014). *ISO International Standard ISO/IEC 14882:2014(E) – Programming Language C++*. [Working draft]. Geneva, Switzerland: International Organization for Standardization (ISO).

ANSI X3.198-1992 (R1997) / ISO/IEC 1539:1991. American National Standard – Programming Language Fortran Extended. American National Standards Institute / ISO/IEC. Informally known as Fortran 90.

Docking

O. Trott, A. J. Olson, **AutoDock Vina**: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *Journal of Computational Chemistry*. 31 (**2010**) 455 – 461.

Appendix A

In 1975, Cremer and Pople published the mathematical development of a general definition of ring puckering coordinates (Cremer, 1975). In their definition, the conformational space of an N-membered ring can be described using N-3 puckering coordinates.

Cremer and Pople puckering coordinates for 6-membered rings.

For 6-membered rings, there are needed 3 puckering coordinates to describe their whole conformational space: a radius Q and two phase angles ϕ and θ. The Q coordinate is the sum of the perpendicular displacement of each ring atom (j) to the ring average plane ($Q = \Sigma_j^6 z_j$). The ϕ and θ coordinates are obtained by solving the following system of equations.

$$Q\ sin\theta \cos\phi = \sqrt{\frac{1}{3}} \sum_{j=1}^{6} z_j \cos\left[\frac{2\pi}{6} 2(j-1)\right]$$

$$Q\ sin\theta \sin\phi = \sqrt{\frac{1}{3}} \sum_{j=1}^{6} z_j \sin\left[\frac{2\pi}{6} 2(j-1)\right]$$

$$Q \cos\theta = \sqrt{\frac{1}{6}} \sum_{j=1}^{6} (-1)^{j-1} z_j$$

In the present Thesis, all conformational studies has been performed over the Mercator surface (ϕ, θ). However, the Stoddart representation is available, transforming the polar coordinates to Cartesian ones. The $q_x = Q \sin\theta \cos\phi$, $q_y = Q \sin\theta \sin\phi$ and $q_z = Q \cos\theta$ are possible collective variables to reconstruct conformational FELs of 6-membered sugars.

Cremer and Pople puckering coordinates for 7-membered rings.

In the present Thesis, the principal goal is sampling the complete conformational hyper-surface of 7-membered ring molecules. The expressions of the four puckering coordinates (φ_2, φ_3, q_2, q_3) for this structures are defined by

$$q_2 \cos\varphi_2 = \left(\frac{2}{7}\right)^{1/2} \sum_{j=1}^{7} z_j \cos[4\pi(j-1)/7] = \left(\frac{2}{7}\right)^{1/2} A_2 \quad \text{(I)}$$

$$q_2 \sin\varphi_2 = -\left(\frac{2}{7}\right)^{1/2} \sum_{j=1}^{7} z_j \sin[4\pi(j-1)/7] = \left(\frac{2}{7}\right)^{1/2} B_2 \quad \text{(II)}$$

$$q_3 \cos\varphi_3 = \left(\frac{2}{7}\right)^{1/2} \sum_{j=1}^{7} z_j \cos[6\pi(j-1)/7] = \left(\frac{2}{7}\right)^{1/2} A_3 \quad \text{(III)}$$

$$q_3 \sin\varphi_3 = -\left(\frac{2}{7}\right)^{1/2} \sum_{j=1}^{7} z_j \sin[6\pi(j-1)/7] = \left(\frac{2}{7}\right)^{1/2} B_3 \quad \text{(IV)}$$

where z_j is the displacement of the atom *j* respect of the mean geometrical plane formed for the seven ring atoms. Applying (V) = (I)^2 + (II)^2, (VI) = (III)^2 + (IV)^2, (VII) = (II) / (I) and (VIII) = (IV) / (III), we obtain the final expressions of

$$q_2 = \left(\frac{2}{7}\right)^{1/2} \left(A_2^{\,2} + B_2^{\,2}\right)^{1/2} \quad \text{(V)}$$

$$q_3 = \left(\frac{2}{7}\right)^{1/2} \left(A_3^{\,2} + B_3^{\,2}\right)^{1/2} \quad \text{(VI)}$$

$$\varphi_2 = \arctan\left(\frac{B_2}{A_2}\right) \quad \text{(VII)}$$

$$\varphi_3 = \arctan\left(\frac{B_3}{A_3}\right) \quad \text{(VIII)}$$

whose derivatives expressions are defined by

$$\frac{dq_2}{dR} = \left(\frac{2}{7}\right)^{1/2} \left(A_2^{\,2} + B_2^{\,2}\right)^{-1/2} \left(A_2 \cdot \frac{dA_2}{dR} + B_2 \cdot \frac{dB_2}{dR}\right) \quad \text{(IX)}$$

$$\frac{dq_3}{dR} = \left(\frac{2}{7}\right)^{1/2} \left(A_3{}^2 + B_3{}^2\right)^{-1/2} \left(A_3 \cdot \frac{dA_3}{dR} + B_3 \cdot \frac{dB_3}{dR} \right) \quad \text{(X)}$$

$$\frac{d\varphi_2}{dR} = \left(A_2{}^2 + B_2{}^2\right)^{-1} \left(B_2 \cdot \frac{dA_2}{dR} - A_2 \cdot \frac{dB_2}{dR} \right) \quad \text{(XI)}$$

$$\frac{d\varphi_3}{dR} = \left(A_3{}^2 + B_3{}^2\right)^{-1} \left(B_3 \cdot \frac{dA_3}{dR} - A_3 \cdot \frac{dB_3}{dR} \right) \quad \text{(XII)}$$

All these expressions (V – XII) has been implemented in the Plumed 2 software as a set of collective variables.

Programming 1 – Implementation of the Cremer and Pople puckering coordinates for 7-membered rings as collective variables in the Plumed 2 software (C++).

In this code two sets of collective variables are presented: the previously named "polar" puckering coordinates (4 CVs, phase, phase2, amplitude and amplitude2) and the Cartesian transformation of the 4 polar coordinates into a toroidal space (3 CVs, q_x, q_y and q_z).

```
class Puckering7N : public Colvar {
  bool cartesian;

public:
  Puckering7N(const ActionOptions&);
// active methods:
  virtual void calculate();
  static void registerKeywords(Keywords& keys);
};

PLUMED_REGISTER_ACTION(Puckering7N,"PUCKERING7N")

void Puckering7N::registerKeywords(Keywords& keys){
   Colvar::registerKeywords( keys );
   keys.add("atoms","ATOMS","the seven atoms involved in the
puckering");
   keys.add("compulsory","TYPE","POLAR","Should be POLAR or
CARTESIAN.");
}
```

```
Puckering7N::Puckering7N(const ActionOptions&ao):
PLUMED_COLVAR_INIT(ao),
cartesian(false)
{
  vector<AtomNumber> atoms;
  parseAtomList("ATOMS",atoms);
  if(atoms.size()!=7) error("only for 7-membered rings");
  string type;
  parse("TYPE",type);
  if(type=="POLAR"){
    cartesian=false;
  } else if(type=="CARTESIAN"){
    cartesian=true;
  } else error("Type should be POLAR or CARTESIAN");

  checkRead();

  if(cartesian){
    addComponentWithDerivatives("qx");
componentIsNotPeriodic("qx");
    addComponentWithDerivatives("qy");
componentIsNotPeriodic("qy");
    addComponentWithDerivatives("qz");
componentIsNotPeriodic("qz");
  }

  addComponentWithDerivatives("phase");
componentIsPeriodic("phase","0","2pi");
  addComponentWithDerivatives("phase2");
componentIsPeriodic("phase2","0","2pi");
  addComponentWithDerivatives("amplitude");
componentIsNotPeriodic("amplitude");
  addComponentWithDerivatives("amplitude2");
componentIsNotPeriodic("amplitude2");


  requestAtoms(atoms);
}

// calculator
void Puckering7N::calculate(){

  vector<Vector> r(7);
  r[0]=getPosition(0);
  for(unsigned i=1;i<7;i++){
    r[i]=r[i-1]+pbcDistance(getPosition(i-1),getPosition(i));
  }

  vector<Vector> R(7);
  Vector center;
  for(unsigned j=0;j<7;j++) center+=r[j]/7.0;
  for(unsigned j=0;j<7;j++) R[j]=(r[j]-center);

  Vector Rp,Rpp;
  for(unsigned j=0;j<7;j++) Rp +=R[j]*sin(2.0/7.0*pi*j);
  for(unsigned j=0;j<7;j++) Rpp+=R[j]*cos(2.0/7.0*pi*j);
```

```
  Vector n=crossProduct(Rp,Rpp);
  Vector nhat=n/modulo(n);

  Tensor dn_dRp=dcrossDv1(Rp,Rpp);
  Tensor dn_dRpp=dcrossDv2(Rp,Rpp);

  Tensor dnhat_dn= 1.0/modulo(n)*( Tensor::identity() -
extProduct(nhat,nhat));
  Tensor dnhat_dRp=matmul(dnhat_dn,dn_dRp);
  Tensor dnhat_dRpp=matmul(dnhat_dn,dn_dRpp);

  vector<double> z(7);
  for(unsigned j=0;j<7;j++) z[j]=dotProduct(R[j],nhat);

  vector<vector<Vector> > dz_dR(7);
  for(unsigned j=0;j<7;j++) dz_dR[j].resize(7);

  for(unsigned i=0;i<7;i++) for(unsigned j=0;j<7;j++){
    if(i==j) dz_dR[i][j]+=nhat;
    dz_dR[i][j]+=matmul(R[i],dnhat_dRp)*sin(2.0/7.0*pi*j);
    dz_dR[i][j]+=matmul(R[i],dnhat_dRpp)*cos(2.0/7.0*pi*j);
  }

  double B=0.0;
  for(unsigned j=0;j<7;j++) B+=z[j]*cos(4.0/7.0*pi*j);

  vector<Vector> dB_dR(7);
  for(unsigned i=0;i<7;i++) for(unsigned j=0;j<7;j++){
    dB_dR[i]+=dz_dR[j][i]*cos(4.0/7.0*pi*j);
  }
  Vector Bsum;
  for(unsigned j=0;j<7;j++) Bsum+=dB_dR[j];
  for(unsigned j=0;j<7;j++) dB_dR[j]-=Bsum/7.0;;

  double A=0.0;
  for(unsigned j=0;j<7;j++) A+=z[j]*sin(4.0/7.0*pi*j);

  vector<Vector> dA_dR(7);
  for(unsigned i=0;i<7;i++) for(unsigned j=0;j<7;j++){
    dA_dR[i]+=dz_dR[j][i]*sin(4.0/7.0*pi*j);
  }
  Vector Asum;
  for(unsigned j=0;j<7;j++) Asum+=dA_dR[j];
  for(unsigned j=0;j<7;j++) dA_dR[j]-=Asum/7.0;;

  double D=0.0;
  for(unsigned j=0;j<7;j++) D+=z[j]*cos(6.0/7.0*pi*j);

  vector<Vector> dD_dR(7);
  for(unsigned i=0;i<7;i++) for(unsigned j=0;j<7;j++){
    dD_dR[i]+=dz_dR[j][i]*cos(6.0/7.0*pi*j);
  }
  Vector Dsum;
  for(unsigned j=0;j<7;j++) Dsum+=dD_dR[j];
  for(unsigned j=0;j<7;j++) dD_dR[j]-=Dsum/7.0;;
```

```
  double C=0.0;
  for(unsigned j=0;j<7;j++) C+=z[j]*sin(6.0/7.0*pi*j);

  vector<Vector> dC_dR(7);
  for(unsigned i=0;i<7;i++) for(unsigned j=0;j<7;j++){
    dC_dR[i]+=dz_dR[j][i]*sin(6.0/7.0*pi*j);
  }
  Vector Csum;
  for(unsigned j=0;j<7;j++) Csum+=dC_dR[j];
  for(unsigned j=0;j<7;j++) dC_dR[j]-=Csum/7.0;;


// PHASE DERIVATIVES
    vector<Vector> dphase_dR(7);
    for(unsigned j=0;j<7;j++){
      dphase_dR[j]=1.0/(A*A+B*B) * (-B*dA_dR[j] + A*dB_dR[j]);
    }

    Value* vphase=getPntrToComponent("phase");
    vphase->set(phase);
    setAtomsDerivatives (vphase,0, dphase_dR[0] );
    setAtomsDerivatives (vphase,1, dphase_dR[1] );
    setAtomsDerivatives (vphase,2, dphase_dR[2] );
    setAtomsDerivatives (vphase,3, dphase_dR[3] );
    setAtomsDerivatives (vphase,4, dphase_dR[4] );
    setAtomsDerivatives (vphase,5, dphase_dR[5] );
    setAtomsDerivatives (vphase,6, dphase_dR[6] );
    Tensor phase_virial;
    for(unsigned j=0;j<7;j++){
      phase_virial-=extProduct(r[j],dphase_dR[j]);
    }
    setBoxDerivatives (vphase,phase_virial);

// PHASE2 DERIVATIVES
   vector<Vector> dphase2_dR(7);
   for(unsigned j=0;j<7;j++){
     dphase2_dR[j]=1.0/(C*C+D*D) * (-D*dC_dR[j] + C*dD_dR[j]);
   }
   Value* vphase2=getPntrToComponent("phase2");
   vphase2->set(phase2);
   setAtomsDerivatives (vphase2,0, dphase2_dR[0] );
   setAtomsDerivatives (vphase2,1, dphase2_dR[1] );
   setAtomsDerivatives (vphase2,2, dphase2_dR[2] );
   setAtomsDerivatives (vphase2,3, dphase2_dR[3] );
   setAtomsDerivatives (vphase2,4, dphase2_dR[4] );
   setAtomsDerivatives (vphase2,5, dphase2_dR[5] );
   setAtomsDerivatives (vphase2,6, dphase2_dR[6] );
   Tensor phase2_virial;
   for(unsigned j=0;j<7;j++){
     phase2_virial-=extProduct(r[j],dphase2_dR[j]);
   }
   setBoxDerivatives (vphase2,phase2_virial);

//  AMPLITUDE
    double amplitude=sqrt((2.*(A*A+B*B)/7.));
```

```
//  AMPLITUDE DERIVATIES
    vector<Vector> damplitude_dR(7);
    for (unsigned j=0;j<7;j++){
      damplitude_dR[j]=0.5*sqrt(2.0/7.0)/(sqrt(A*A+B*B)) *
(2*A*dA_dR[j] + 2*B*dB_dR[j]);
    }

    Value* vamplitude=getPntrToComponent("amplitude");
    vamplitude->set(amplitude);
    setAtomsDerivatives (vamplitude,0, damplitude_dR[0] );
    setAtomsDerivatives (vamplitude,1, damplitude_dR[1] );
    setAtomsDerivatives (vamplitude,2, damplitude_dR[2] );
    setAtomsDerivatives (vamplitude,3, damplitude_dR[3] );
    setAtomsDerivatives (vamplitude,4, damplitude_dR[4] );
    setAtomsDerivatives (vamplitude,5, damplitude_dR[5] );
    setAtomsDerivatives (vamplitude,6, damplitude_dR[6] );
    Tensor amplitude_virial;
    for(unsigned j=0;j<7;j++){
      amplitude_virial-=extProduct(r[j],damplitude_dR[j]);
    }
    setBoxDerivatives (vamplitude,amplitude_virial);

//  AMPLITUDE2
    double amplitude2=sqrt((2.*(C*C+D*D)/7.));

//  AMPLITUDE2 DERIVATIES
    vector<Vector> damplitude2_dR(7);
    for (unsigned j=0;j<7;j++){
      damplitude2_dR[j]=0.5*sqrt(2.0/7.0)/(sqrt(C*C+D*D)) *
(2*C*dC_dR[j] + 2*D*dD_dR[j]);
    }

    Value* vamplitude2=getPntrToComponent("amplitude2");
    vamplitude2->set(amplitude2);
    setAtomsDerivatives (vamplitude2,0, damplitude2_dR[0] );
    setAtomsDerivatives (vamplitude2,1, damplitude2_dR[1] );
    setAtomsDerivatives (vamplitude2,2, damplitude2_dR[2] );
    setAtomsDerivatives (vamplitude2,3, damplitude2_dR[3] );
    setAtomsDerivatives (vamplitude2,4, damplitude2_dR[4] );
    setAtomsDerivatives (vamplitude2,5, damplitude2_dR[5] );
    setAtomsDerivatives (vamplitude2,6, damplitude2_dR[6] );
    Tensor amplitude2_virial;
    for(unsigned j=0;j<7;j++){
      amplitude2_virial-=extProduct(r[j],damplitude2_dR[j]);
    }
    setBoxDerivatives (vamplitude2,amplitude2_virial);


  if(cartesian){

    double QX1=cos(amplitude);
    double
QX2=(amplitude*sin(phase)+amplitude*sin(phase)*cos(phase2));
    double QX3=cos(phase)*cos(phase2);
    double QX4=amplitude2*sin(phase2)*cos(phase);
```

```
    double QY1=sin(amplitude);
    double
QY2=(amplitude*cos(phase)+amplitude2*cos(phase)*cos(phase2));
    double QY3=sin(phase)*cos(phase2);
    double QY4=amplitude2*sin(phase2)*sin(phase);

    double QZ1=sin(phase2);
    double QZ2=amplitude2*cos(phase2);


// qx
    double qx=(amplitude+amplitude2*cos(phase2))*cos(phase);

// qx derivaties
    vector<Vector> dqx_dR(7);
    for(unsigned j=0;j<7;j++){
     dqx_dR[j]=QX1*damplitude_dR[j]-
QX2*dphase_dR[j]+QX3*damplitude2_dR[j]-QX4*dphase2_dR[j];
    }

    Value* vqx=getPntrToComponent("qx");
    vqx->set(qx);
    setAtomsDerivatives (vqx,0, dqx_dR[0] );
    setAtomsDerivatives (vqx,1, dqx_dR[1] );
    setAtomsDerivatives (vqx,2, dqx_dR[2] );
    setAtomsDerivatives (vqx,3, dqx_dR[3] );
    setAtomsDerivatives (vqx,4, dqx_dR[4] );
    setAtomsDerivatives (vqx,5, dqx_dR[5] );
    setAtomsDerivatives (vqx,6, dqx_dR[6] );
    Tensor qx_virial;
    for(unsigned j=0;j<7;j++){
      qx_virial-=extProduct(r[j],dqx_dR[j]);
    }
    setBoxDerivatives (vqx,qx_virial);

// qy
   double qy=(amplitude+amplitude2*cos(phase2))*sin(phase);

// qy derivatives
    vector<Vector> dqy_dR(7);
    for(unsigned j=0;j<7;j++){

dqy_dR[j]=QY1*damplitude_dR[j]+QX2*dphase_dR[j]+QX3*damplitude2_
dR[j]-QX4*dphase2_dR[j];
    }

    Value* vqy=getPntrToComponent("qy");
    vqy->set(qy);
    setAtomsDerivatives (vqy,0, dqy_dR[0] );
    setAtomsDerivatives (vqy,1, dqy_dR[1] );
    setAtomsDerivatives (vqy,2, dqy_dR[2] );
    setAtomsDerivatives (vqy,3, dqy_dR[3] );
    setAtomsDerivatives (vqy,4, dqy_dR[4] );
    setAtomsDerivatives (vqy,5, dqy_dR[5] );
    setAtomsDerivatives (vqy,6, dqy_dR[6] );
```

```
    Tensor qy_virial;
    for(unsigned j=0;j<7;j++){
      qy_virial-=extProduct(r[j],dqy_dR[j]);
    }
    setBoxDerivatives (vqy,qy_virial);


// qz
   double qz=amplitude2*sin(phase2);

// qz derivatives
    vector<Vector> dqz_dR(7);
    for(unsigned j=0;j<7;j++){
      dqz_dR[j]=QZ1*damplitude2_dR[j]+QZ2*dphase2_dR[j];
    }

    Value* vqz=getPntrToComponent("qz");
    vqz->set(qz);
    setAtomsDerivatives (vqz,0, dqz_dR[0] );
    setAtomsDerivatives (vqz,1, dqz_dR[1] );
    setAtomsDerivatives (vqz,2, dqz_dR[2] );
    setAtomsDerivatives (vqz,3, dqz_dR[3] );
    setAtomsDerivatives (vqz,4, dqz_dR[4] );
    setAtomsDerivatives (vqz,5, dqz_dR[5] );
    setAtomsDerivatives (vqz,6, dqz_dR[6] );
    Tensor qz_virial;
    for(unsigned j=0;j<7;j++){
      qz_virial-=extProduct(r[j],dqz_dR[j]);
    }
    setBoxDerivatives (vqz,qz_virial);

  }


}

}
}
```

Programming 2 – 3 and 4 collective variable free energy landscapes analysis. Minima and saddle point search in a grid box (Fortran 90)

Analysis of a discretized (100x100x100x100) 4D landscape

Using the sum_hills code of the Plumed program, we reconstruct a free energy landscape from a HILLS file. In this case, the metadynamics simulation is performed using 4 collective variables (Conformational Puckering Coordinates Analysis of 7-membered ring or complex reaction simulation as an inverting mechanism throughout of a catalytic water nucleophylic attack). Once the fes.dat file is written with the whole

energy landscape information, there is no way to represent an intelligible form of our 5D matrix (4 variables + energy). However, we need to extract information from it. The most important data to be identified is the minima, the transition states and the minimum energy pathway over the landscape.

Minima finder (minima4d.f90)

The code reads the whole fes.dat file, saving the x, y, z, w arrays in CV1, CV2, CV3 and CV4.dat files. Then, the program reads these files to construct the xf, yf, zf, wf vectors. To allocate a minimum in the surface, the program compare one point with the 11 adjacent points. If the energy of all the 11 points is higher of the energy in the central point, the coordinates and energy of this point is saved in the Min4D.dat file.

```
program minima4d
implicit none
integer :: i,j,k,l,ii,jj,kk,ll,s
real*8,dimension(100,100,100,100) :: FES
real*8 :: x,y,z,w
real*8, dimension(100) :: xf,yf,zf,wf

open(10,file="fes.dat",access="sequential",status="unknown")
open(12,file="CV1.dat",access="sequential",status="unknown")
open(13,file="CV2.dat",access="sequential",status="unknown")
open(14,file="CV3.dat",access="sequential",status="unknown")
open(15,file="CV4.dat",access="sequential",status="unknown")

do i=1,100
   do j=1,100
      do k=1,100
         do l=1,100
           read(10,*) x,y,z,w,FES(i,j,k,l)
           if(i==1.and.j==1.and.k==1) then
             write(15,*) w
           endif
         enddo
         read(10,*)
```

```
          if(i==1.and.j==1) then
            write(14,*) z
          endif
       enddo
       read(10,*)
       if(i==1) then
         write(13,*) y
       endif
    enddo
    read(10,*)
    write(12,*) x
enddo

do i=10,15
  close(i)
enddo

open(12,file="CV1.dat",access="sequential",status="unknown")

do i=1,100
   read(12,*) xf(i)
enddo

close(12)

open(13,file="CV2.dat",access="sequential",status="unknown")

do i=1,100
   read(13,*) yf(i)
enddo

close(13)

open(14,file="CV3.dat",access="sequential",status="unknown")
```

```
do i=1,100
   read(14,*) zf(i)
enddo

close(14)

open(15,file="CV4.dat",access="sequential",status="unknown")

do i=1,100
   read(15,*) wf(i)
enddo

close(15)

open(11,file="Min4D.dat",access="sequential",status="unknown")

do i=2,99
   do j=2,99
      do k=2,99
         do l=2,99
            s=0
            do ii=-1,1
               do jj=-1,1
                  do kk=-1,1
                     do ll=-1,1

if(FES(i+ii,j+jj,k+kk,l+ll)<FES(i,j,k,l)) then
                           s=s+1
                        endif
                     enddo
                  enddo
               enddo
            enddo
            if (s==0.and.FES(i,j,k,l)<0.d0) then
```

```
                write(11,*) xf(i),yf(j),zf(k),wf(l),FES(i,j,k,l)
             endif
          enddo
       enddo
    enddo
enddo

close(11)

end program
```

TS Finder (TSfinder4D.f90)

Taking into account the definition of the transition state of a reaction, we must find a saddle point of order 1 over the surface, then, we need to allocate a point that it is a maximum over a direction and a minimum respect N-1 orthogonal directions. For this reason, it is very important to know how many sets of orthogonal direction (using 1, 0, -1 directions, due to the discretization of the space) exist in a 4D space. Working with the whole set of directions, we found 13 sets of orthogonal vectors. Then, 52 (13 x 4) kinds of transition state could exist over a 4D discretized surface ((M,m,m,m), (m,M,m,m), (m, m, M, m) and (m, m, m, M) where m=minimum, M=maximum).

13 sets of orthogonal vectors:

1st: unitary vectors (Written in the TSxyzt.dat file)

(1,0,0,0)
(0,1,0,0)
(0,0,1,0)
(0,0,0,1)

2nd: diagonal-unitary vectors (Written in the TSd.dat file)

(1,0,0,0)
(0,1,0,0)
(0,0,1,-1)
(0,0,1,1)

(1,0,0,0)

(0,1,-1,0)

(0,1,1,0)

(0,0,0,1)

(1,0,-1,0)

(1,0,1,0)

(0,1,0,0)

(0,0,0,1)

(1,0,0,-1)

(1,0,0,1)

(0,1,0,0)

(0,0,1,0)

(1,0,0,0)

(0,1,0,-1)

(0,1,0,1)

(0,0,1,0)

3rd double -1 direction (ex. (-1,-1,1,1) or (1,-1,-1,1)) (Written in the TS-2.dat file)

(-1,-1,1,1)

(1,1,1,1)

(1,-1,0,0)

(0,0,1,-1)

(-1,1,-1,1)

(1,1,1,1)

(0,-1,0,1)

(1,0,-1,0)

(1,-1,-1,1)

(1,1,1,1)

(1,0,0,-1)

(0,1,-1,0)

4rd single -1 direction (ex. (1,-1,1,1) or (-1,1,1,1)) (Written in the TS-1.dat file)

(1,1,1,-1)

(-1,-1,1,-1)

(0,0,1,1)

(1,-1,0,0)

(1,1,-1,1)

(-1,1,-1,-1)

(0,1,1,0)

(1,0,0,-1)

(1,-1,1,1)

(1,-1,-1,-1)

(1,1,0,0)

(0,0,1,-1)

(-1,1,1,1)

(-1,-1,-1,1)

(1,0,0,1)

(0,1,-1,0)

During the TS scan, when the program confirms a point as a TS, the code goes to a subroutine that reconstruct the pathway from the TS to the two minima through the minimum energy pathway. The code choose the neighbor point with the higher negative gradient. Finally, all the points of the pathways are written in the Pathway.dat file.

```
program TSfinder4d
implicit none
real*8,dimension(100,100,100,100) :: FES
real*8,dimension(200) :: x,y,z,t
integer :: i,j,k,l,p,q,r,s

open(55,file="fes.dat",access="sequential",status="unknown")

do i=1,100
   do j=1,100
      do k=1,100
         do l=1,100
            read(55,*) x(i),y(j),z(k),t(l),FES(i,j,k,l)
         enddo
         read(55,*)
      enddo
      read(55,*)
   enddo
   read(55,*)
enddo

open(56,file="TSxyzt.dat",access="sequential",status="unknown")
open(57,file="TSd.dat",access="sequential",status="unknown")
open(58,file="TS-2.dat",access="sequential",status="unknown")
open(59,file="TS-1.dat",access="sequential",status="unknown")
open(60,file="Pathway.dat",access="sequential",status="unknown")

do i=2,99
   do j=2,99
      do k=2,99
         do l=2,99
         !TSxyzt

if(FES(i,j,k,l)>=(FES(i+1,j,k,l)).and.FES(i,j,k,l)>FES(i-
1,j,k,l).and.&
```

```
FES(i,j,k,l)<=(FES(i,j+1,k,l)).and.FES(i,j,k,l)<FES(i,j-
1,k,l).and.&

FES(i,j,k,l)<=(FES(i,j,k+1,l)).and.FES(i,j,k,l)<FES(i,j,k-
1,l).and.&

FES(i,j,k,l)<=(FES(i,j,k,l+1)).and.FES(i,j,k,l)<FES(i,j,k,l-1))
then


            write(56,*) x(i),y(j),z(k),t(l),FES(i,j,k,l),"1  0  0
0"
            write(60,*) x(i),y(j),z(k),t(l),FES(i,j,k,l)
            p=i+1
            q=j
            r=k
            s=l
            call path(p,q,r,s,x,y,z,t,FES)
            write(60,*) x(i),y(j),z(k),t(l),FES(i,j,k,l)
            p=i-1
            q=j
            r=k
            s=l
            call path(p,q,r,s,x,y,z,t,FES)
            go to 888

elseif(FES(i,j,k,l)<=(FES(i+1,j,k,l)).and.FES(i,j,k,l)<FES(i-
1,j,k,l).and.&

FES(i,j,k,l)>=(FES(i,j+1,k,l)).and.FES(i,j,k,l)>FES(i,j-
1,k,l).and.&

FES(i,j,k,l)<=(FES(i,j,k+1,l)).and.FES(i,j,k,l)<FES(i,j,k-
1,l).and.&

FES(i,j,k,l)<=(FES(i,j,k,l+1)).and.FES(i,j,k,l)<FES(i,j,k,l-1))
then


(Repiting elseif with every TS option)
```

```
subroutine path(i,j,k,p,x,y,z,t,FES)
implicit none
integer :: i,j,k,p
integer :: l,m,n,o,s,movx,movy,movz,movt
real*8, dimension(100) :: x,y,z,t
real*8, dimension(100,100,100,100) :: FES
real*8 :: ENERGY

s=1

write(60,*) x(i),y(j),z(k),t(p),FES(i,j,k,p)

do while(s.ne.0)
ENERGY=0.d0
s=0
   do l=-1,1
      do m=-1,1
         do n=-1,1
            do o=-1,1

if(FES(i+l,j+m,k+n,p+o)<FES(i,j,k,p).and.FES(i+l,j+m,k+n,p+o)-
FES(i,j,k,p)<ENERGY)then
            s=s+1
            movx=l
            movy=m
            movz=n
            movt=o
            ENERGY=FES(i+l,j+m,k+n,p+o)-FES(i,j,k,p)
            endif
            enddo
         enddo
      enddo
   enddo
i=i+movx
j=j+movy
```

```
k=k+movz
p=p+movt
  if(s.ne.0)then
  write(60,*) x(i),y(j),z(k),t(p),FES(i,j,k,p)
  endif
enddo

end subroutine
```

Appendix B

Histidine protonation of the biological models

In the present work, we are presenting simulations from different "based on" crystallographic data models. Concretely, six different enzymatic species has been studied: GH13 amylosucrase, GH125 exo-α-1,6-mannosidase, GH31 α-glucosidase, GH13 α-glucosidase and GH3 HvExoI. To be sure of the structural stability of them, the protonation of the histidines is crucial. Following, we present the chosen histidine position where the protons has been positioned (HID = δ-nitrogen, HIE = ε-nitrogen, HIP (+) = both nitrogens).

GH13 amylosucrase			
HIE 39	HID 306	HIE 392	HID 565
HIE 173	HIE 332	HIE 414	HID 591
HIE 187	HIE 370	HIE 463	HIP (+) 601
HID 192	HIP (+) 377	HIE 512	
HIE 233	HIP (+) 382	HIE 540	
GH125 exo-α-1,6-mannosidase			
HIE 70	HID 197	HIE 356	HID 415
HIE 143	HID 271	HIE 390	
HIP (+) 176	HID 348	HIE 394	
GH31 α-glucosidase			
HID 88	HID 165	HIP (+) 457	HID 696
HID 90	HIE 189	HIP (+) 467	HIE 698
HIE 99	HIE 205	HIP (+) 507	HID 721
HID 113	HIE 268	HID 569	HID 729
HID 149	HID 316	HIE 576	
HIE 156	HIE 330	HIE 692	
GH13 α-glucosidase			
HID 88	HID 173	HID 289	HID 427
HIE 102	HID 174	HID 301	HID 456
HID 107	HID 206	HIE 329	HID 512
HID 161	HID 237	HID 345	

GH3 HvExoI			
HID 1	HID 115	HID 266	HID 423
HID 2	HID 211	HID 335	HID 491
HID 102	HID 238	HID 381	HID 590

Classical MD simulation results of amylosucrase and GH125 α-mannosidases

In the present section, we present the analysis of the classical MD performed over the amylosucrase and GH125 models: RMSD evolution during the structure stabilization and the relevant conformational information.

Figures of the 1.5 ns NVT simulation at 100 K of the Glu328Gln mutant of the *Np*AS model: RMSD evolution of the backbone (Figure B – 1), RMSF of every aminoacid (Figure B – 2) and structure superposition with the crystallographic structure (Figure B – 3).

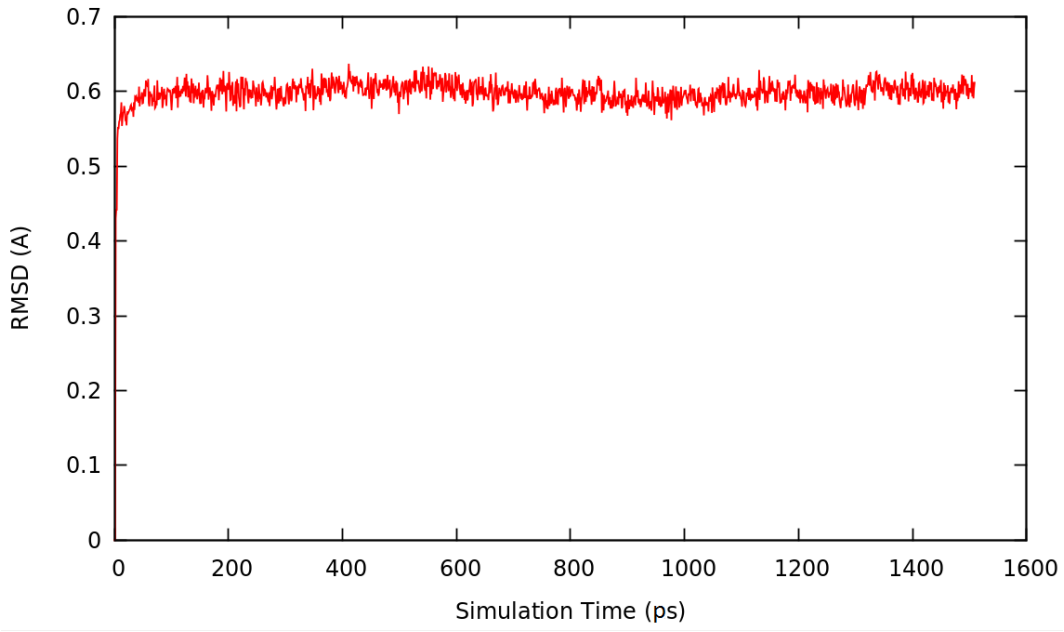


Figure B – 1. Representation of the Root Mean Square Deviation of protein backbone (Red) (Units = Å) during the NVT molecular dynamics simulation of the Glu328Gln *Np*AS model at 100 K.

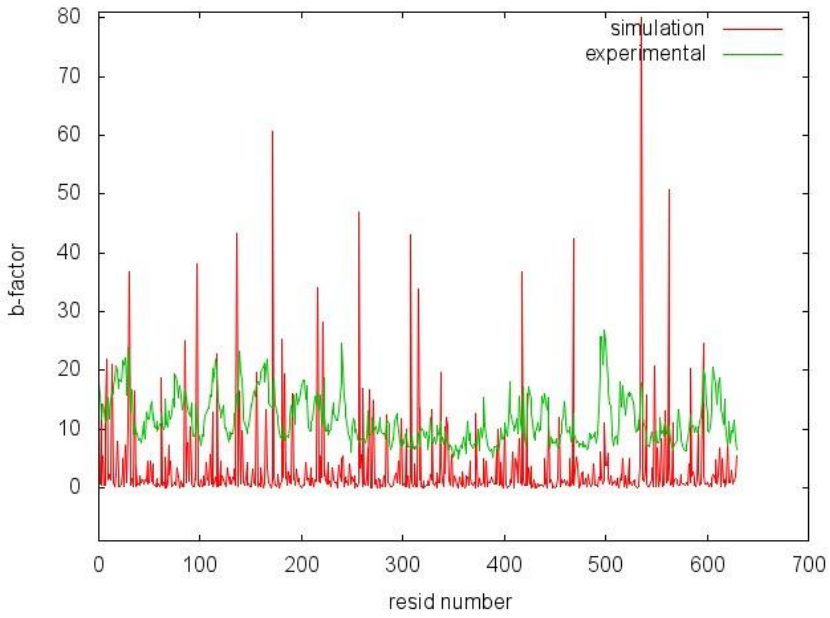


Figure B – 2. Comparison of the b-factor for the simulated (mutant) and experimental structures.

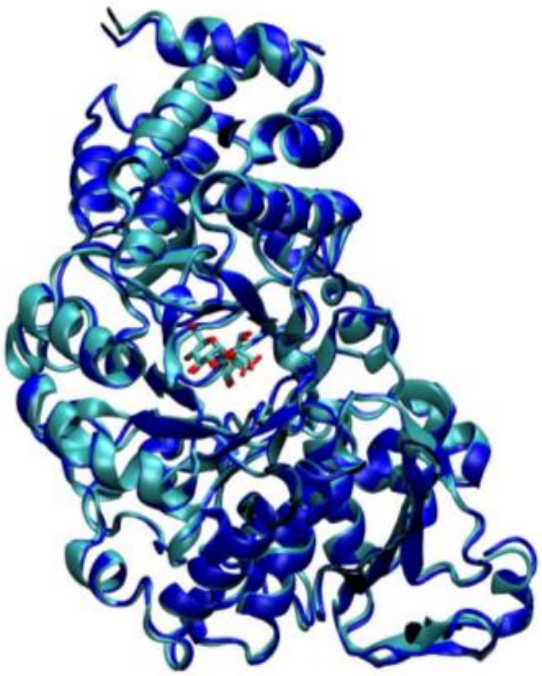


Figure B – 3. Superposition of the simulated and experimental structures (mutant).

During the 15 ns NVT simulation at 300 K, we have observed a correlation between the relative position of the nucleophile (Asp286) and the anomeric carbon with the conformation adopted by the -1 glucose sugar (Figure B – 4). The conformation of the sugar when the structure is stabilized (the RMSD of the protein backbone is constant,

Figure B – 5) collapses to the $^4C_1$ chair conformation (Y (C_2-C_1-O_5-C_5) < 0). The final jump of the RMSD of the sugar is due to a rotation of the fructose ring (not affecting the -1 subsite).

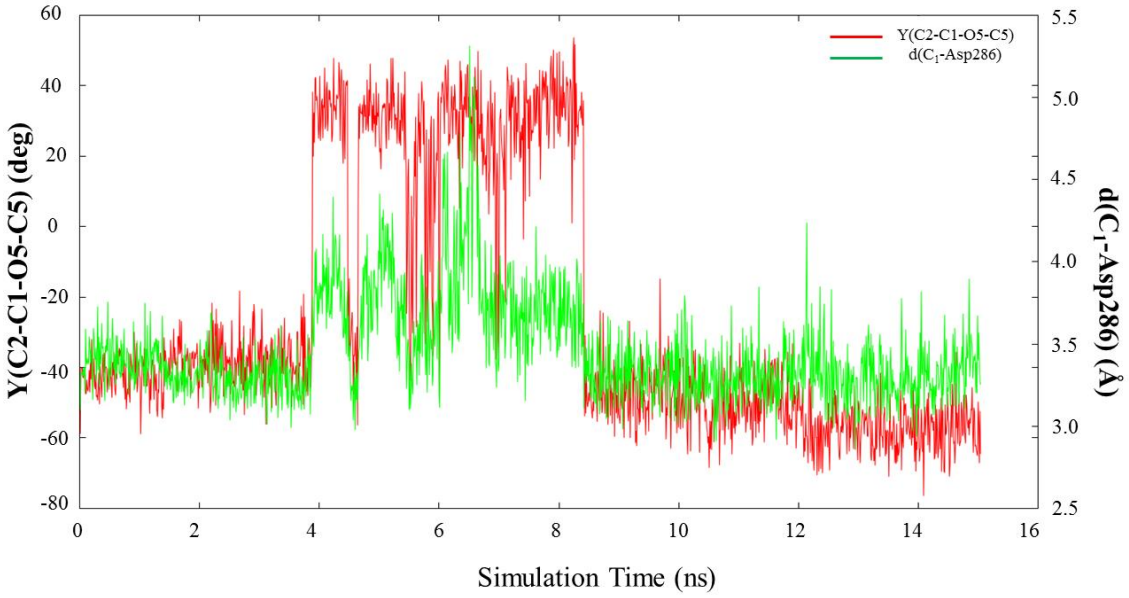


Figure B – 4. Correlation between the distance (Anomeric Carbon – Asp286) (Green) and the conformation of the sugar (dihedral angle C2-C1-O5-C5 ring atoms) (Red) found during the heating of the AS. Negative values of the dihedral angle are found in chair conformation. Positive values are found in boat/skew boat conformation.

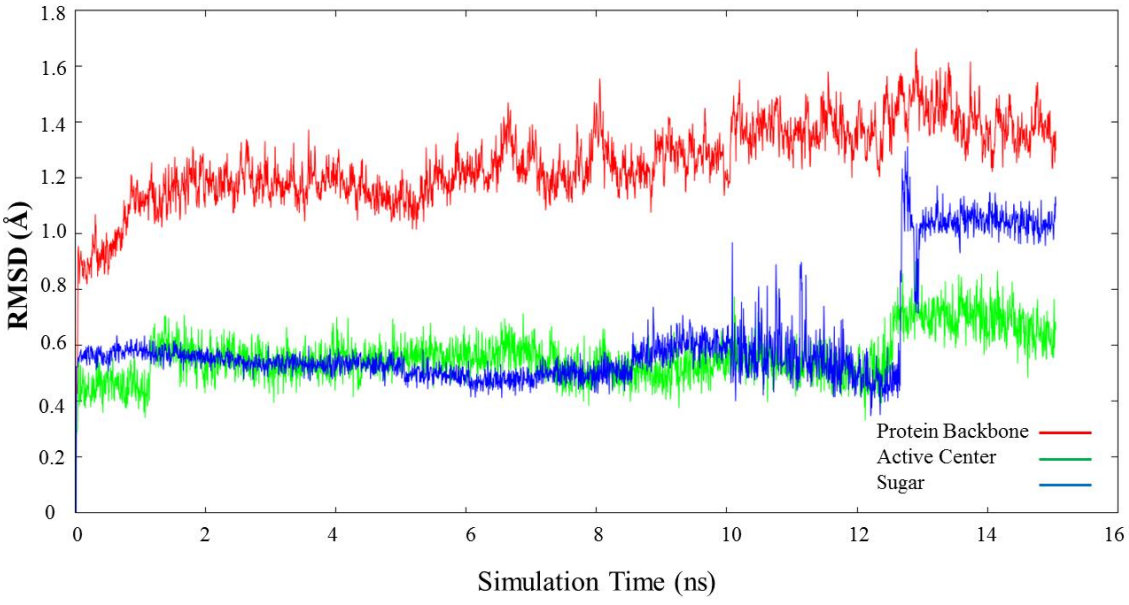


Figure B – 5. Representation of the Root Mean Square Deviation of: Protein backbone (Red), active center residues (Green) and substrate (Blue) (Units = Å) during the NVT molecular dynamics simulation of the AS model.

Chapter IV – GH125 α-mannosidase

In both models (thio-derivative and natural substrate), a 10 ns NVT MD simulation was needed to stabilize the structures (Figure B – 6 and 7). During all the process, both sugars maintain the initial chair conformation.

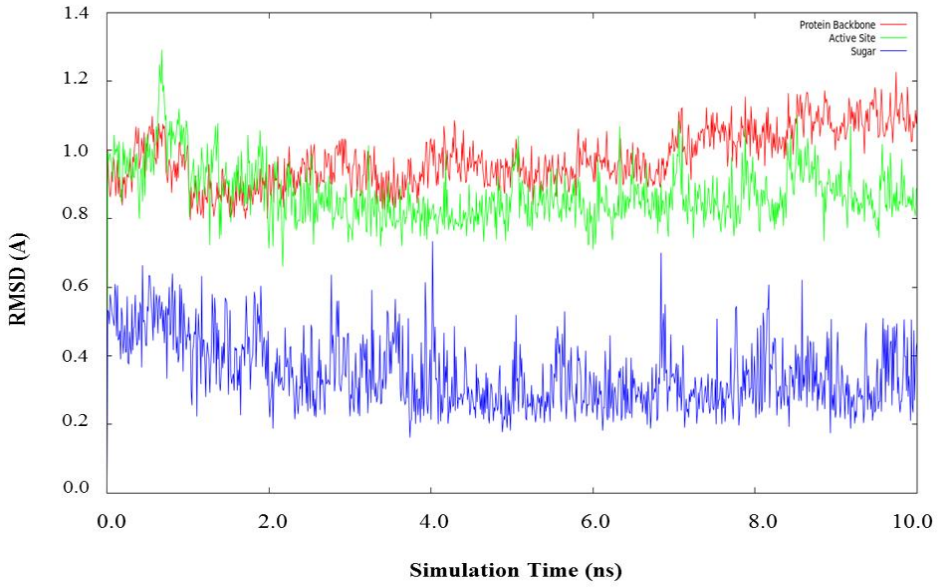


Figure B – 6. Representation of the RMSD of: Protein backbone (Red), active center residues (Green) and substrate (Blue) (Units = Å) during the NVT molecular dynamics simulation of the GH125-thio-derivative model.

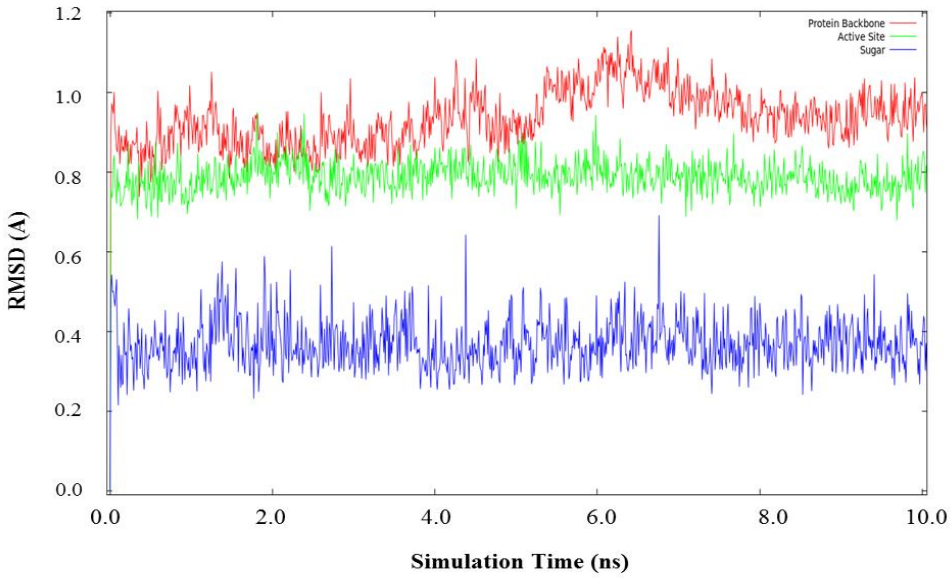


Figure B – 7. Representation of the RMSD of: Protein backbone (Red), active center residues (Green) and substrate (Blue) (Units = Å) during the NVT molecular dynamics simulation of the GH125-natural substrate model.

Chapter VI – GH3 HvExoI

The results of the classical MD simulations of the G2x, G3x, G4x and G6x models are presented in the subsequent figures:

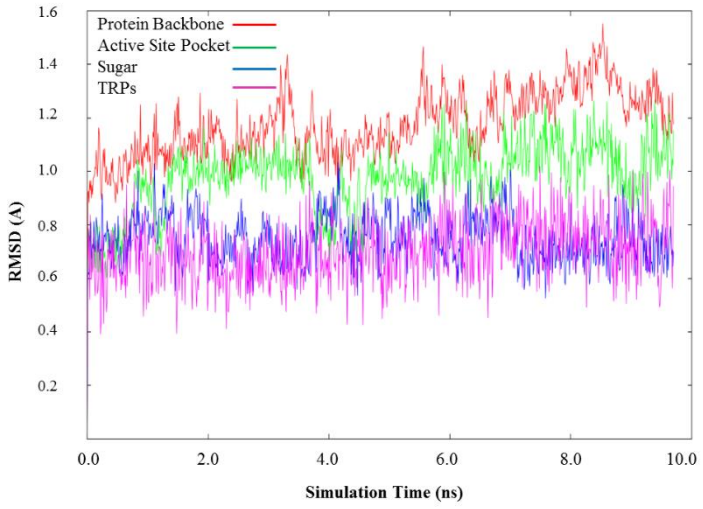


Figure B – 8. Representation of the RMSD of: Protein backbone (Red), active center residues (Green), substrate (Blue) and TRPs (Purple, Trp286 and Trp438) (Units = Å) during the NVT molecular dynamics simulation of the G2S model.

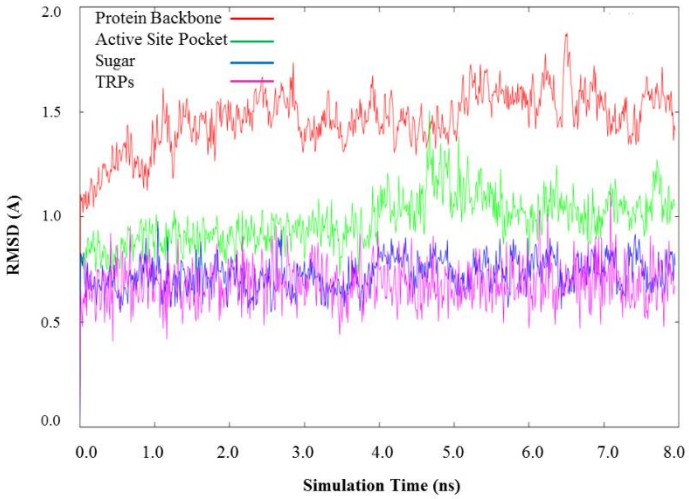


Figure B – 9. Representation of the RMSD of: Protein backbone (Red), active center residues (Green), substrate (Blue) and TRPs (Purple, Trp286 and Trp438) (Units = Å) during the NVT molecular dynamics simulation of the G2S model.

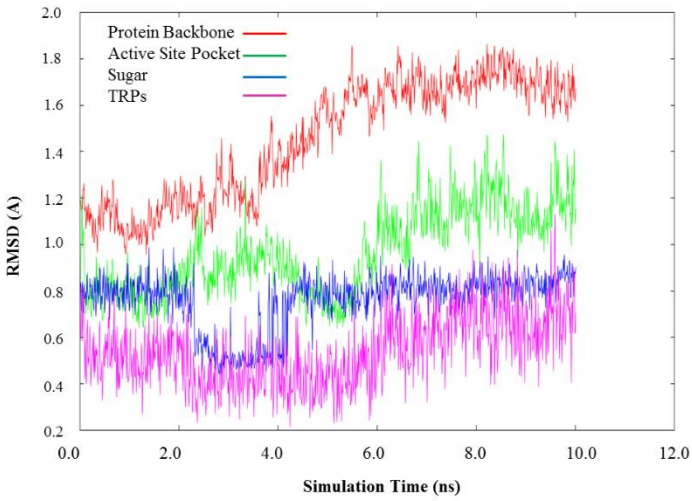


Figure B – 10. Representation of the RMSD of: Protein backbone (Red), active center residues (Green), substrate (Blue) and TRPs (Purple, Trp286 and Trp438) (Units = Å) during the NVT molecular dynamics simulation of the G3S model.

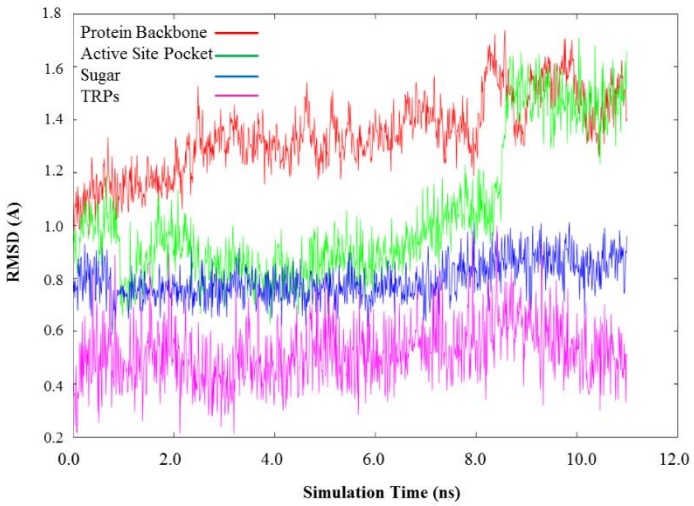


Figure B – 11. Representation of the RMSD of: Protein backbone (Red), active center residues (Green), substrate (Blue) and TRPs (Purple, Trp286 and Trp438) (Units = Å) during the NVT molecular dynamics simulation of the G3O model.

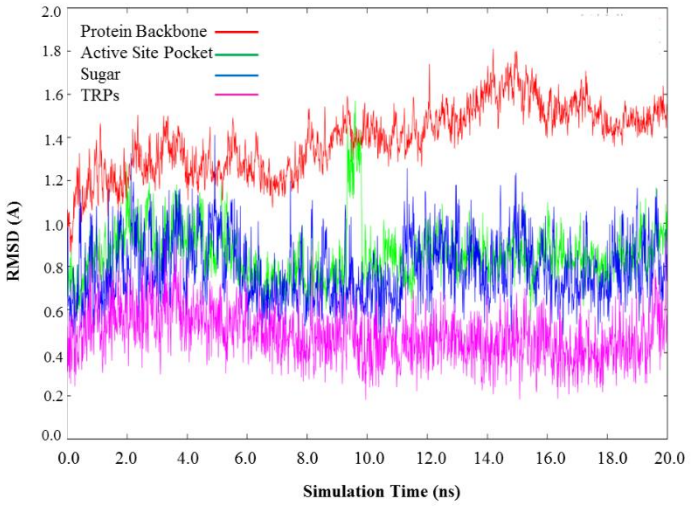


Figure B – 12. Representation of the RMSD of: Protein backbone (Red), active center residues (Green), substrate (Blue) and TRPs (Purple, Trp286 and Trp438) (Units = Å) during the NVT molecular dynamics simulation of the G4S model.

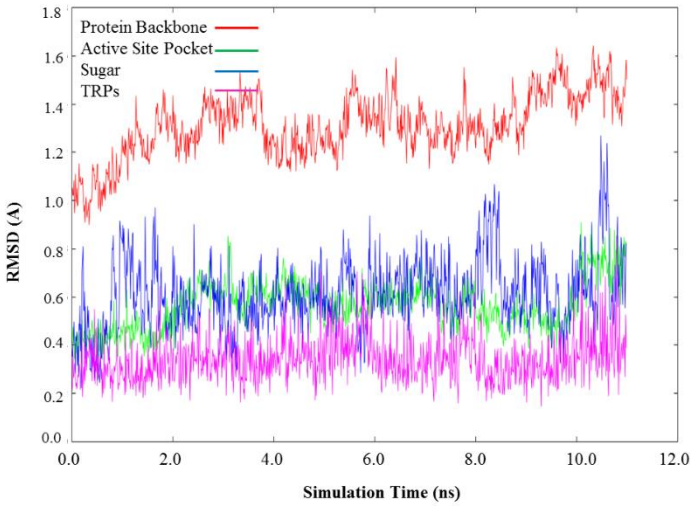


Figure B – 13. Representation of the RMSD of: Protein backbone (Red), active center residues (Green), substrate (Blue) and TRPs (Purple, Trp286 and Trp438) (Units = Å) during the NVT molecular dynamics simulation of the G4O model.

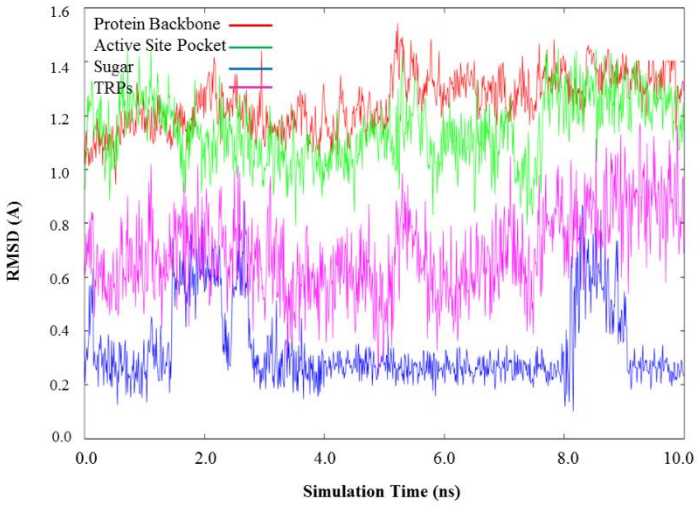


Figure B – 14. Representation of the RMSD of: Protein backbone (Red), active center residues (Green), substrate (Blue) and TRPs (Purple, Trp286 and Trp438) (Units = Å) during the NVT molecular dynamics simulation of the G6S model.

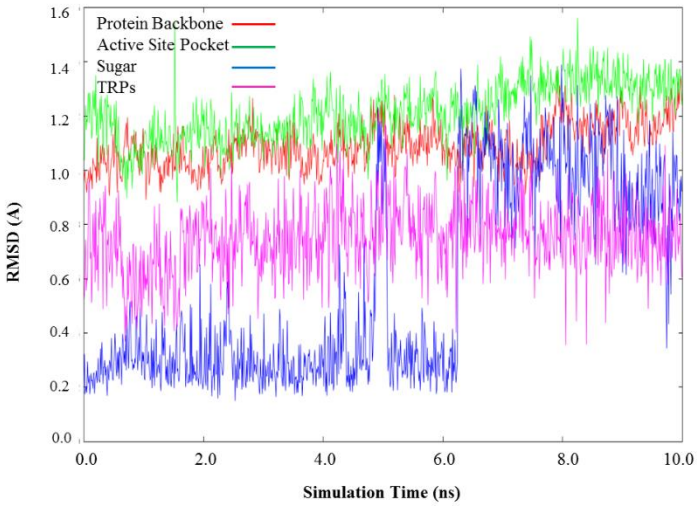


Figure B – 15. Representation of the RMSD of: Protein backbone (Red), active center residues (Green), substrate (Blue) and TRPs (Purple, Trp286 and Trp438) (Units = Å) during the NVT molecular dynamics simulation of the G6O model.

Collective variable evolution of the MTD simulations of amylosucrase and GH125 α-mannosidases.

Chapter III – Amylosucrase. Glycosylation step simulation

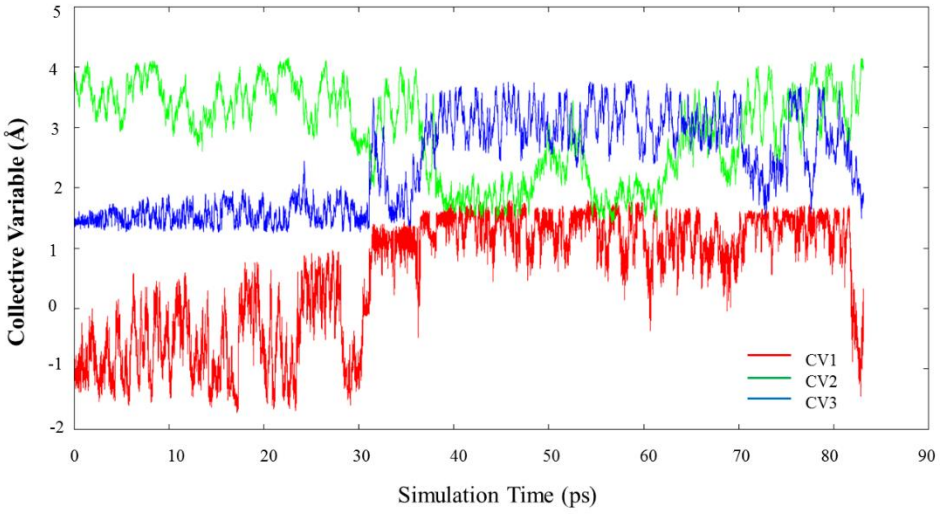


Figure B – 16. Evolution of the three chosen CVs to model the glycosylation reaction in the GH13 amylosucrase.

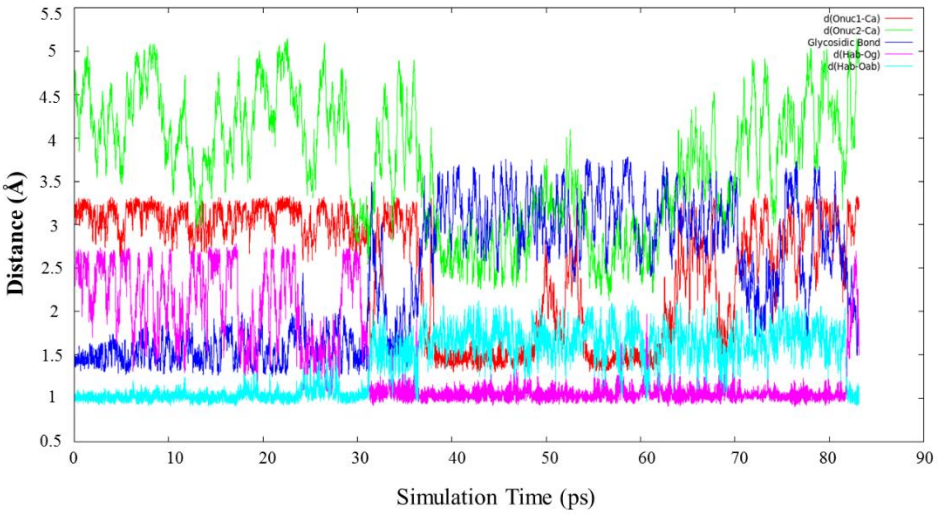


Figure B – 17. Evolution of the most relevant distances for the catalysis during the glycosylation metadynamics simulation.

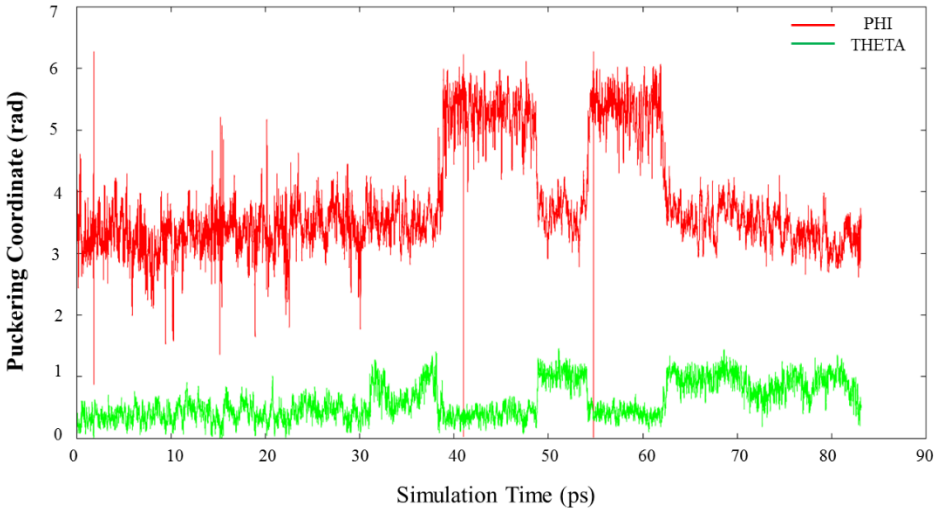


Figure B – 18. Evolution of the puckering coordinates of the -1 sugar during the glycosylation metadynamics simulation.

Water Approach simulation

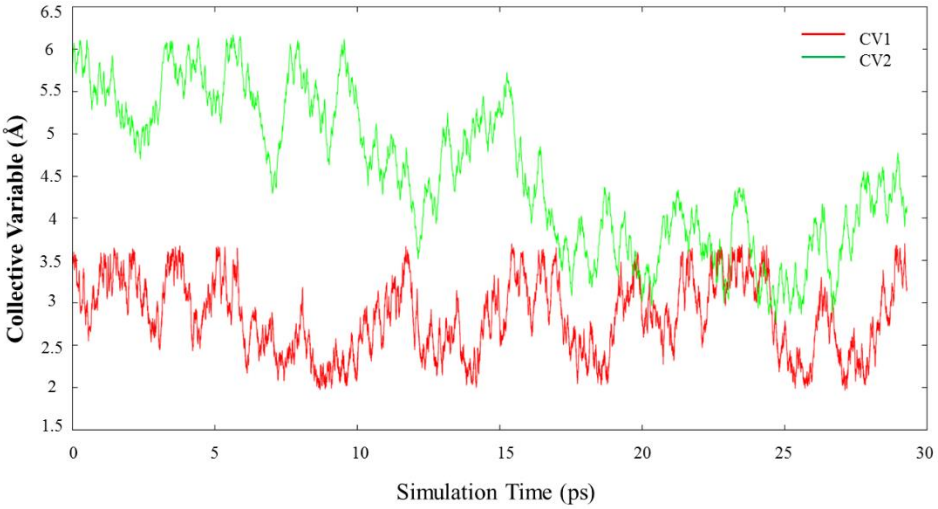


Figure B – 19. Evolution of the two chosen CVs to model the water approach intermediate step in the GH13 amylosucrase.

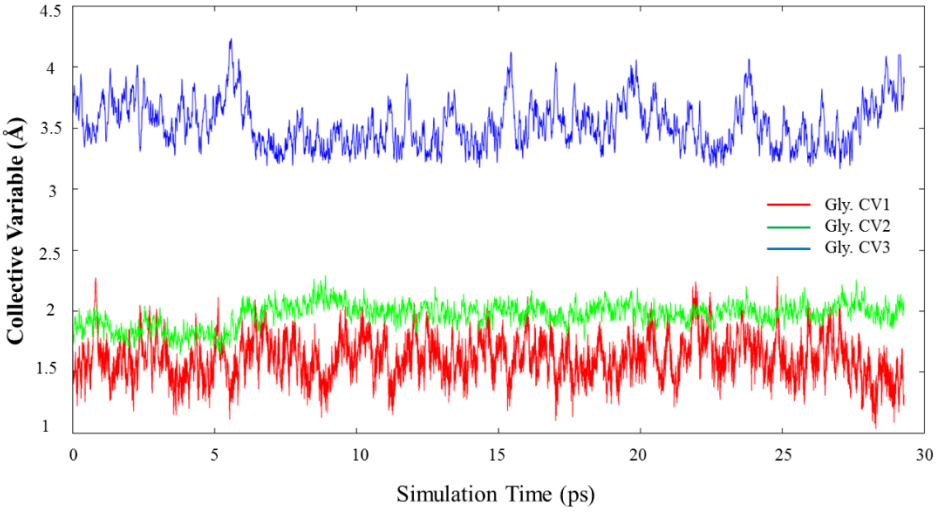


Figure B – 20. Evolution of the glycosylation reaction CVs during the water approach metadynamics simulation.

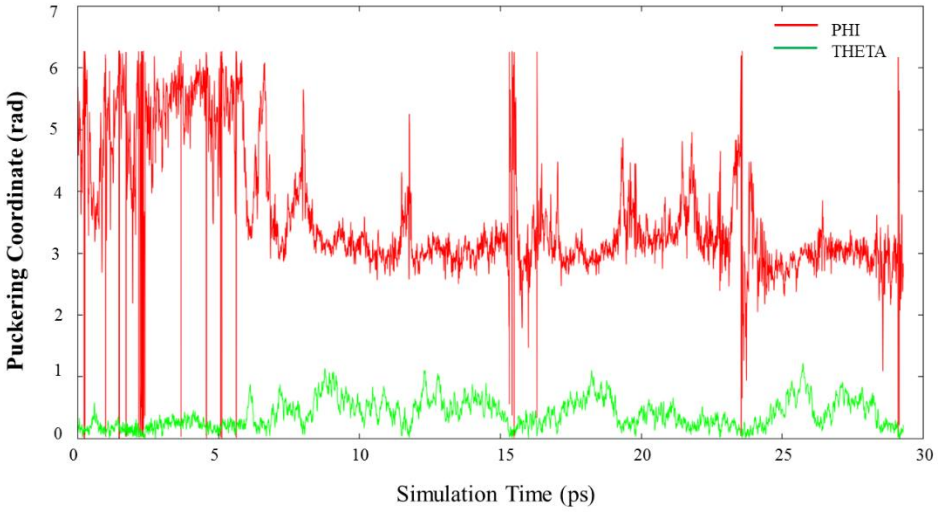


Figure B – 21. Evolution of the puckering coordinates of the -1 sugar during the water approach metadynamics simulation.

Deglycosylation step simulation

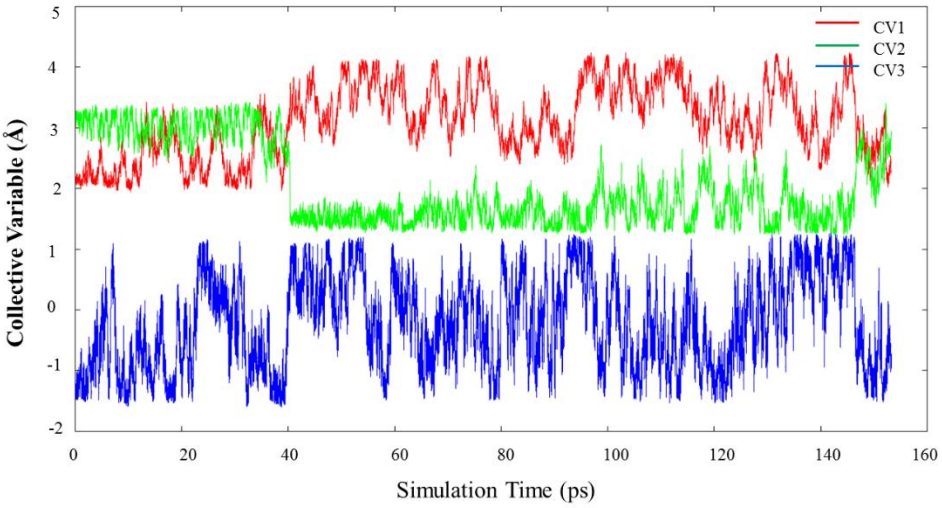


Figure B – 22. Evolution of the three chosen CVs to model the deglycosylation reaction in the GH13 amylosucrase.

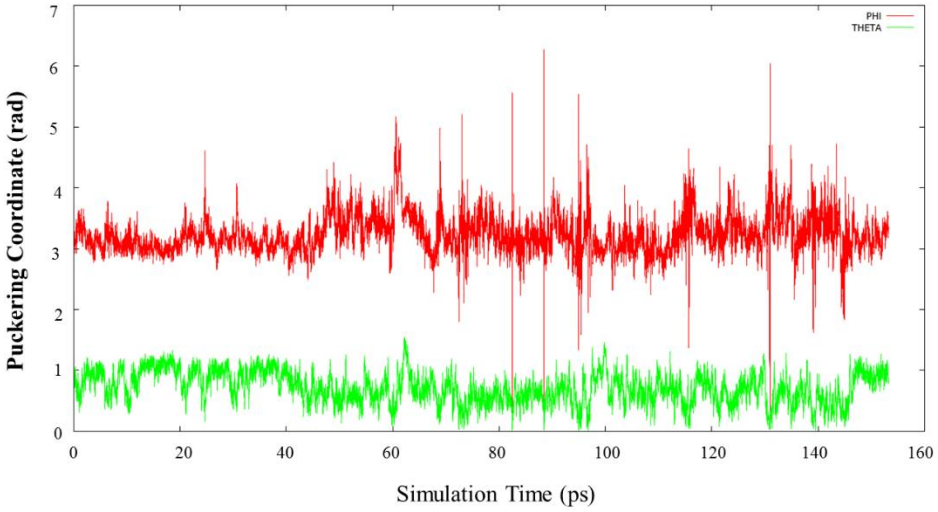


Figure B – 23. Evolution of the puckering coordinates of the -1 sugar during the deglycosylation metadynamics simulation.

Transglycosylation reaction simulation

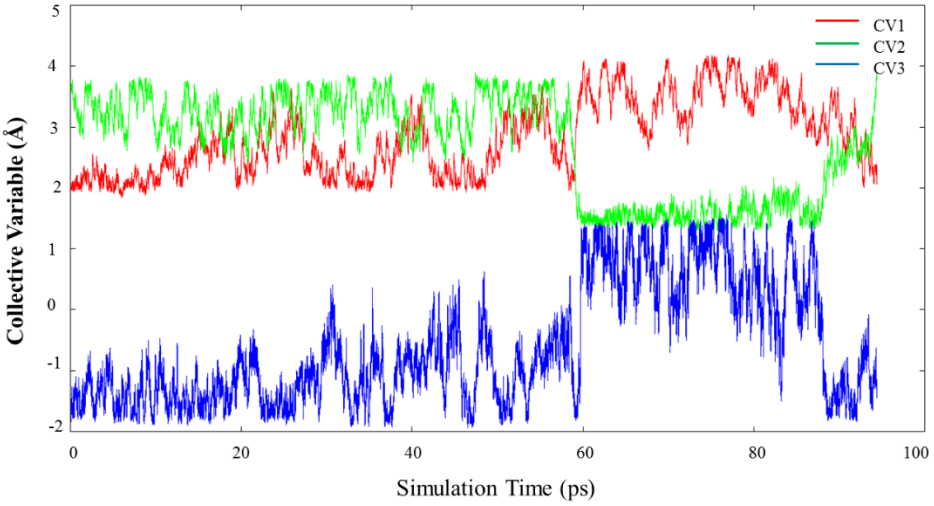


Figure B – 24. Evolution of the three chosen CVs to model the transglycosylation reaction in the GH13 amylosucrase.

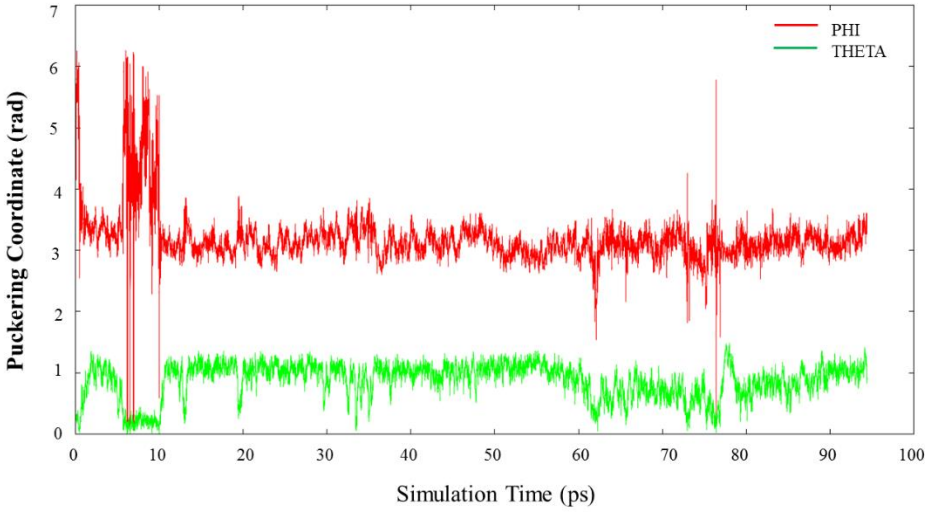


Figure B – 25. Evolution of the puckering coordinates of the -1 sugar during the deglycosylation metadynamics simulation.

Chapter IV – GH125 α-mannosidase. Hydrolysis reaction simulation

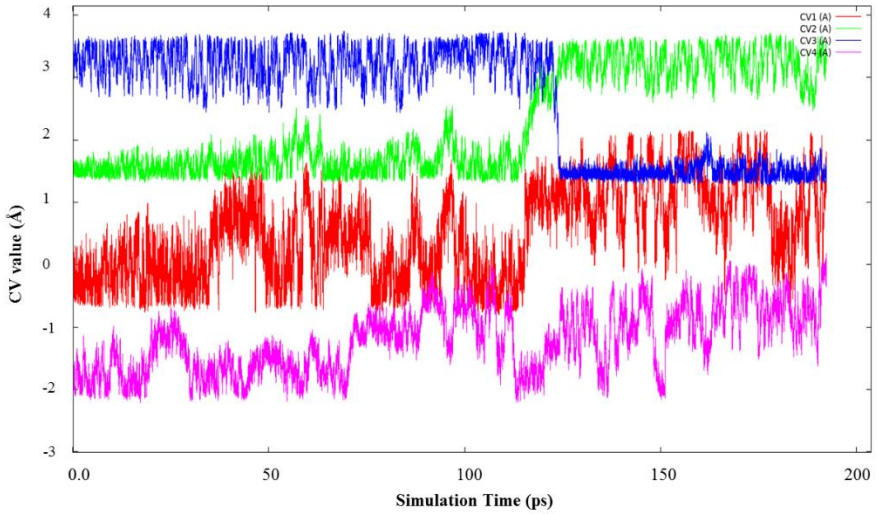


Figure B – 26. Evolution of the four chosen CVs to model the transglycosylation reaction in the GH125 α-mannosidase.

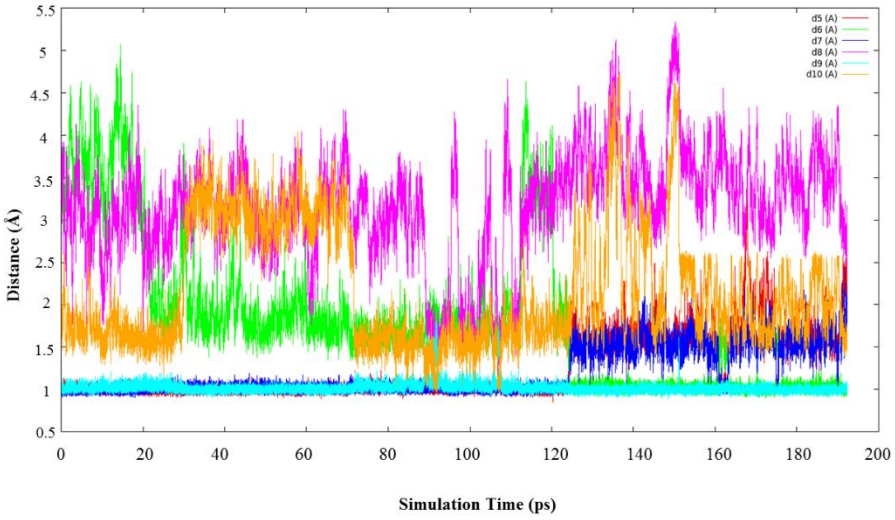


Figure B – 27. Evolution of the most relevant distances taking part in the CV4 during the hydrolysis metadynamics simulation.

Appendix C

Free energy standard deviation landscapes of the conformational studies

Based on the FEL convergence criteria of the reference (Tiwary, 2015) (Section II – 4.3), we present the standard deviation landscapes of the conformational puckering coordinates simulations reported in the present Thesis.

Section III – 4.1 Amylosucrase

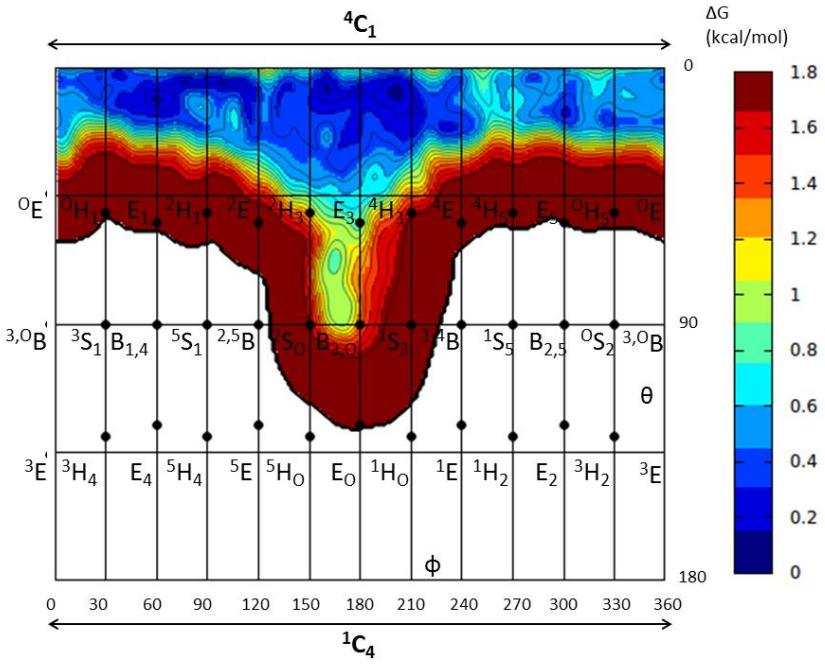


Figure C – 1. Standard deviation error associated to the explored conformational FEL of the AS – sucrose system. An isoline every 0.1 kcal/mol.

Section IV – 4.2 GH125 α-mannosidase and thio-derivative

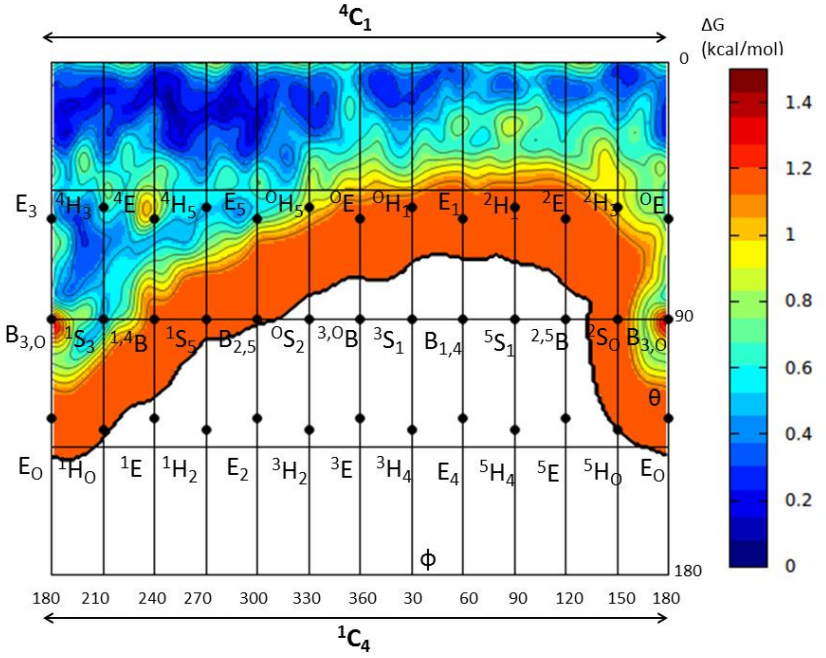


Figure C – 4. Standard deviation error associated to the explored conformational FEL of the GH125 – MB-S system. An isoline every 0.1 kcal/mol.

Section IV – 4.3 GH125 α-mannosidase and natural substrate

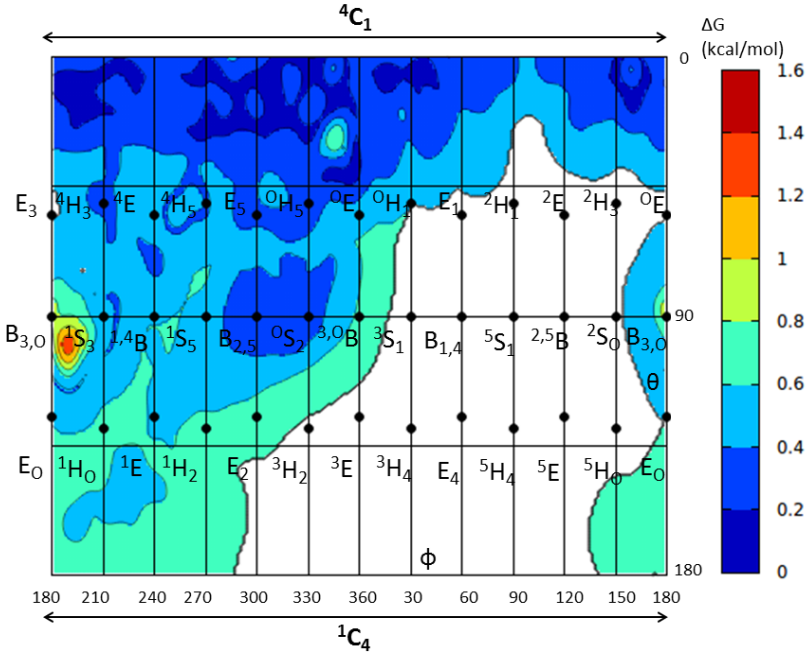


Figure C – 4. Standard deviation error associated to the explored conformational FEL of the GH125 – MB-O system. An isoline every 0.1 kcal/mol.

Section VI – 4.5 GH3 HvExoI (G3O model)

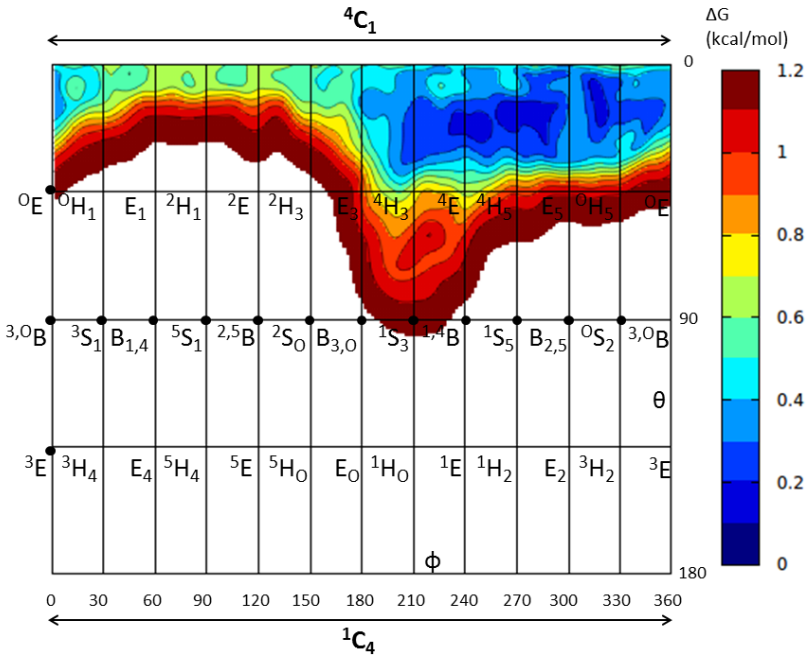


Figure C – 7. Standard deviation error associated to the explored conformational FEL of the GH3 – G3O complex. An isoline every 0.1 kcal/mol.

Appendix D

Effect of the dimensionality reduction in the conformational study of the cycloheptane in gas phase

In Chapter VI Section 4.1, the results of a conformational study of the cycloheptane using a complete set of puckering coordinates (φ_2, φ_3, q_2, q_3). The details of that simulation are presented in the subsequent Table VI – D1.

Table VI – D1. Conformational study of the isolated cycloheptane molecule using a complete set of puckering coordinates.

Cycloheptane						**# Gaussians**	**5208**
Nosé T Elec.	0.00282	10000	**MD$_0$ τ**	4.70	**Box**	21	atoms
Nosé T Nuc.	300	3500	**MTD τ**	**315**	10.5	12.3	11.9
w	0.3	**τ_D**	500	**γ**	6.0	**δs**	0.1

Due to the fact that it is possible to represent all known conformations without the q_2, a new conformational study of the cycloheptane using a reduced conformational space (φ_2, φ_3, ~~q_2~~, q_3) was performed. The computational details are presented in the Table VI – D2.

Table VI – D2. Conformational study of the isolated cycloheptane molecule using a reduced set of puckering coordinates.

Cycloheptane						**# Gaussians**	**6721**
Nosé T Elec.	0.00282	10000	**MD$_0$ τ**	4.70	**Box**	21	atoms
Nosé T Nuc.	300	3500	**MTD τ**	**407**	10.5	12.3	11.9
w	0.3	**τ_D**	500	**γ**	6.0	**δs**	0.1

In both cases, the simulation was stopped once the system crossed four times the TS/S plane (double recrossing). The resulting 2D projections of the reduced FEL over the TC/C and TB/B planes are presented in Figure D1.

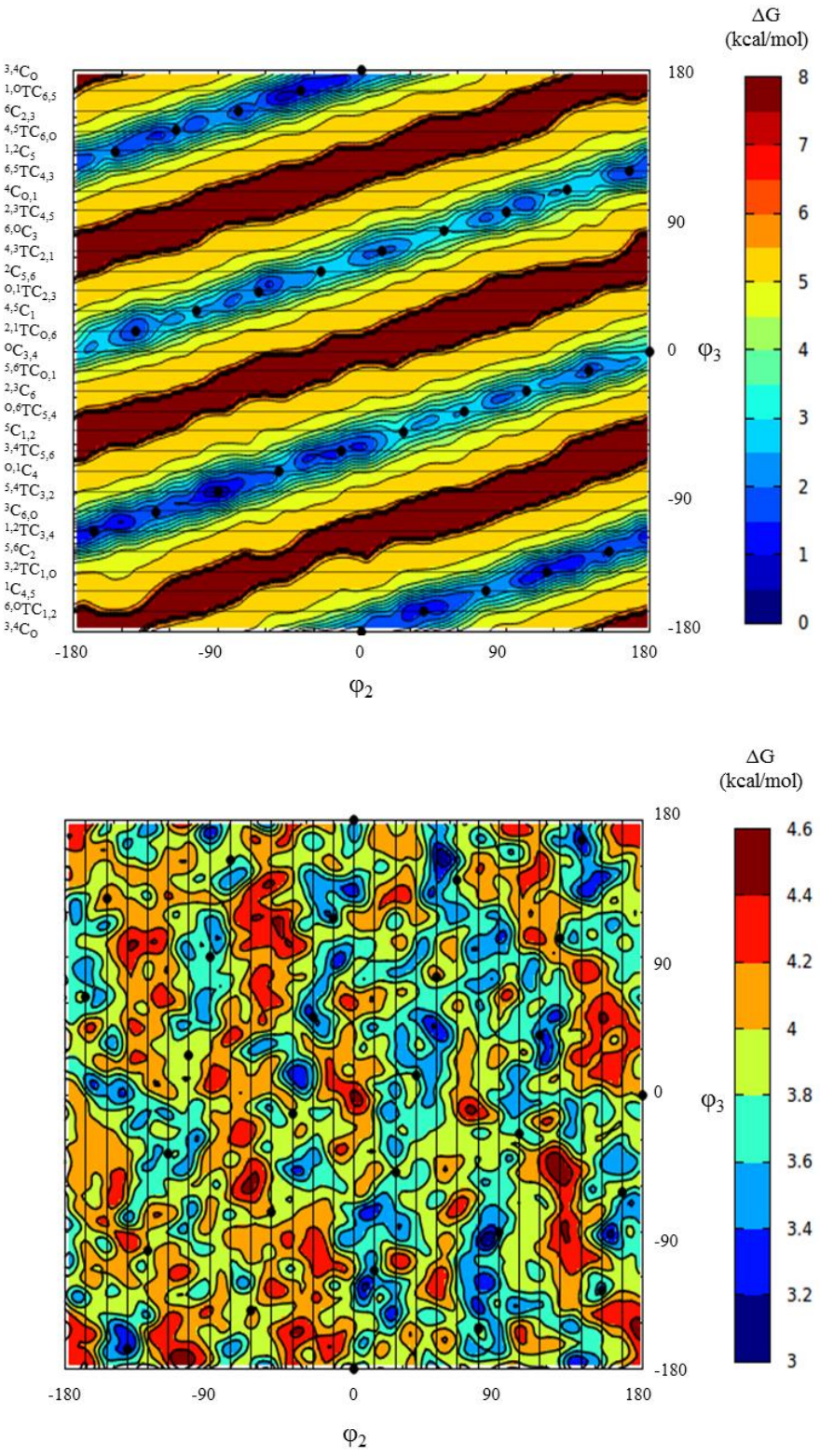


Figure D1. (φ_2, φ_3) projections of (up) TC/C plane (Isoline per 0.5 kcal/mol) and (down) TB/B plane (Isoline per 0.2 kcal/mol) of the reduced conformational FEL of cycloheptane.

In Table VI – D3, the fourteen principal minima over the TC/C plane are presented, showing an average TC → [S] barrier of 4.97 ± 0.52 kcal/mol. The TB and B conformations are energetically indistinguishable and the thermodynamic TC → TB barrier is 3.07 ± 0.50 kcal/mol.

Table VI – D3. Analysis of the minima corresponding to the twisted-chair regions of the Reduced FEL of the cycloheptane molecule in gas-phase. The collective variables are in c.v.u. and energy in kcal/mol[a].

Min.	φ_2	φ_3	q_3	Meta. ΔG^a	Rel. ΔG^a	Conformation
1	0.5271	-2.9307	0.6289	-5.35	0.30	$^{6,0}TC_{1,2}$
2	2.2139	-2.4247	0.6384	-5.47	0.18	$^{3,2}TC_{1,0}$
3	-3.0994	-2.0874	0.6384	-5.30	0.35	$^{1,2}TC_{3,4}$
4	-1.5392	-1.5813	0.6289	-5.65	0.00	$^{5,4}TC_{3,2}$
5	-0.3584	-1.1596	0.6289	-5.05	0.60	$^{3,4}TC_{5,6}$
6	1.0331	-0.6958	0.6195	-3.81	1.85	$^{0,6}TC_{5,4}$
7	2.4669	-0.2319	0.6289	-4.53	1.12	$^{5,6}TC_{0,1}$
8	-2.4669	0.2319	0.6289	-4.82	0.83	$^{2,1}TC_{0,6}$
9	-0.9910	0.7380	0.6384	-5.00	0.66	$^{0,1}TC_{2,3}$
10	0.2741	1.1175	0.6384	-4.43	1.22	$^{4,3}TC_{2,1}$
11	1.3705	1.4970	0.6384	-4.54	1.11	$^{2,3}TC_{4,5}$
12	2.8042	1.9609	0.6289	-4.84	0.81	$^{6,5}TC_{4,3}$
13	-2.2139	2.4247	0.6384	-5.20	0.46	$^{4,5}TC_{6,O}$
14	-0.5693	2.9729	0.6289	-5.59	0.06	$^{1,0}TC_{6,5}$
			<En>=	**-4.97**	0.68	
			St. Dev. =	**0.52**	0.52	

In Table VI – D4, the results of both simulations are summarized. There are two points to take into account to make a proper comparison and a correct rationalization to choose the best set of puckering coordinates.

On the one hand, the simulation using the complete set of CVs was 315 ps long (5208 Gaussian potentials), meanwhile, paradoxically, using the reduced set the simulation, we need to calculate 407 ps (6721 Gaussian potentials).

Table VI – D4. Comparison of the results of the cycloheptane complete and reduced conformational simulations. The energies are in kcal/mol.

Conformation	CPMD-4CVs	CPMD-3CVs
TC	0.00 ± 0.46	0.00 ± 0.52
C	1.12 ± 0.38	1.14 ± 0.32
TB	3.43 ± 0.25	3.07 ± 0.50
B	3.43 ± 0.25	3.07 ± 0.50
TS	4.66 ± 0.46	4.97 ± 0.52

On the other hand, the numerical results show similar standard deviations in the TC/C and TS energies, but in the case of the TB/B plane, it is observed a rougher surface where the standard deviation in the reduced set simulation is the double than the complete set one. Furthermore, the TC → [S] barrier is overestimated 0.31 kcal/mol (7%) and the thermodynamic TC → TB barrier is underestimated 0.36 kcal/mol (11%).

Thus, due to the overestimation of the kinetical barrier, the simulation needs more time to go from the TC/C plane to the TB/B plane and the computational resources are not optimized using a lower number of collective variables. Furthermore, the accuracy is importantly decreased in the description of the TB/B plane, instead of a similar behaviour of the TC/C and TS conformations.

Appendix E

Chapter VI. Projections of the calculated FELs. Collective variable evolution. Atomic structures of the found minima. RMSD evolution of enzyme-substrate simulations

Section VII – 4.1 Cycloheptane.

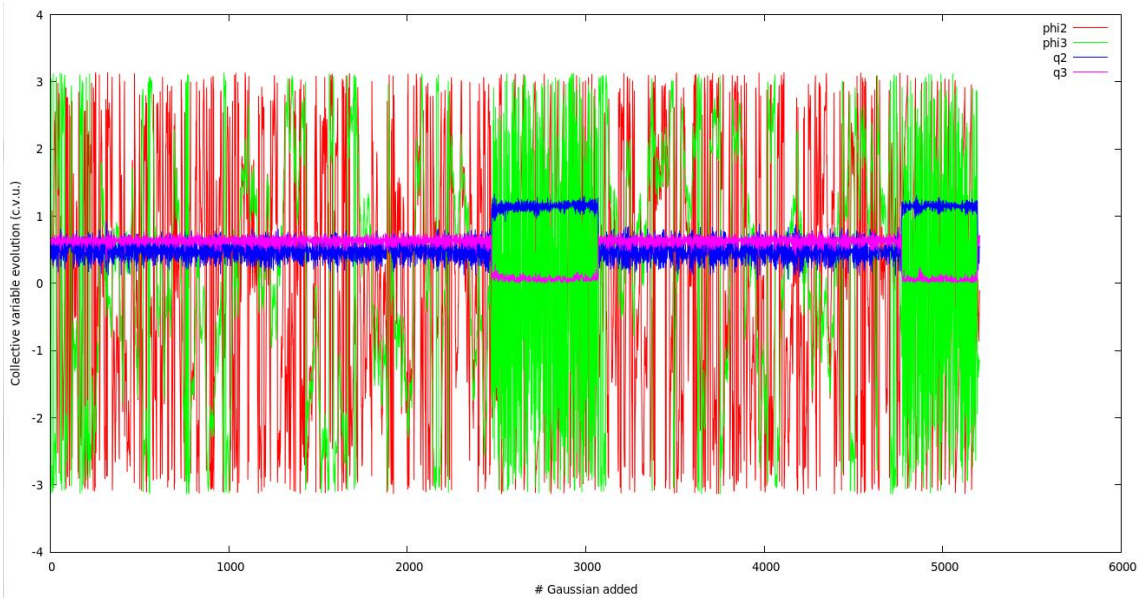


Figure E – 1. Collective variable evolution during the cycloheptane conformational study using the **complete** set of puckering coordinates.

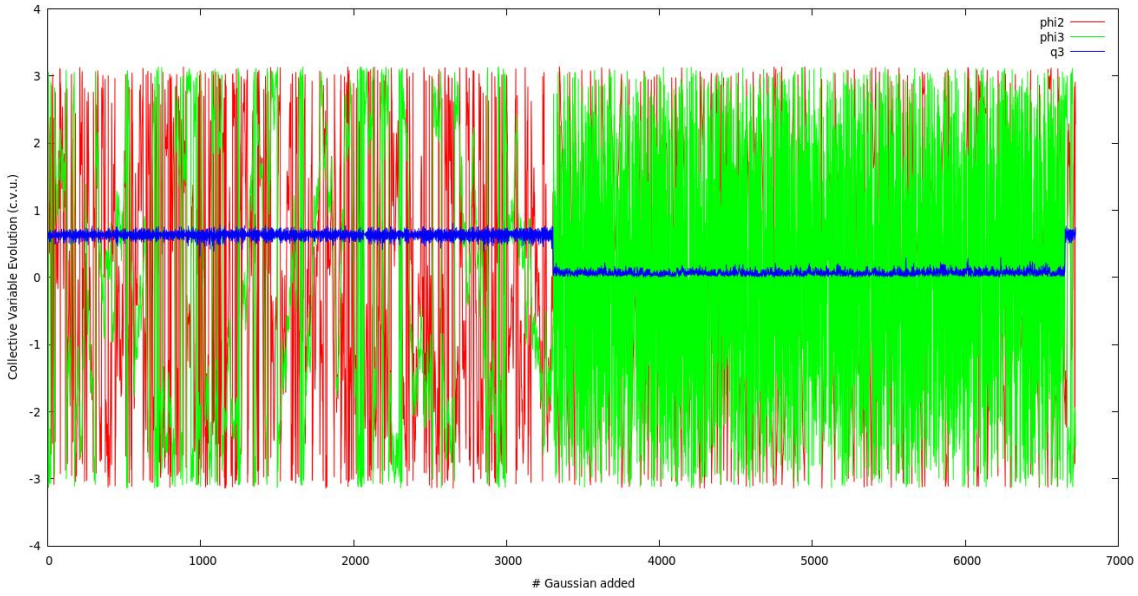


Figure E – 2. Collective variable evolution during the cycloheptane conformational study using the **reduced** set of puckering coordinates.

Section VII – α-D-glycero-D-idoseptanoside.

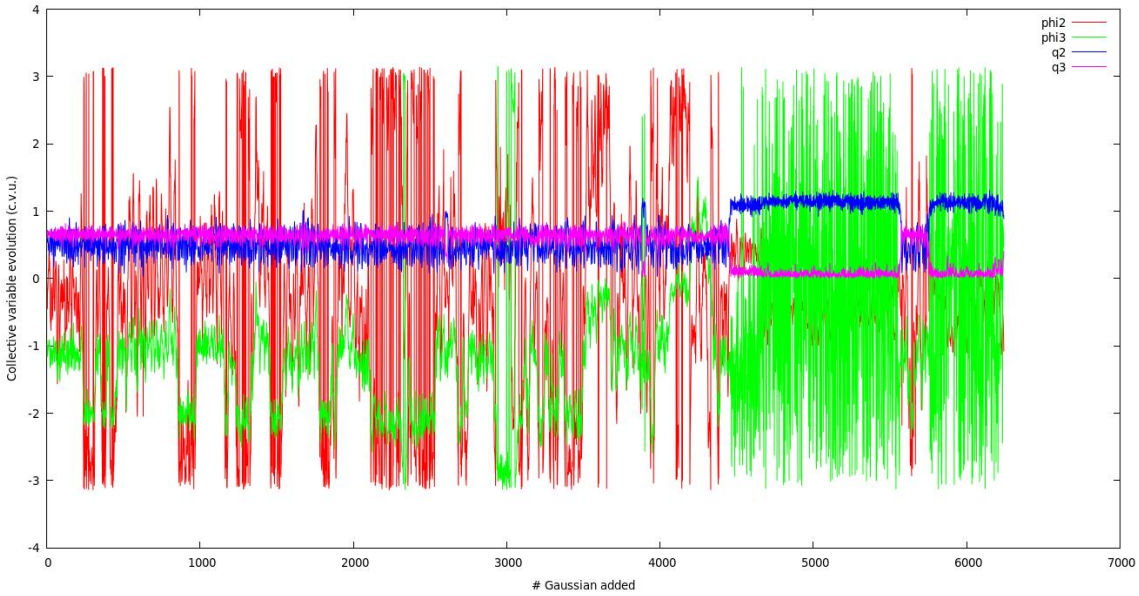
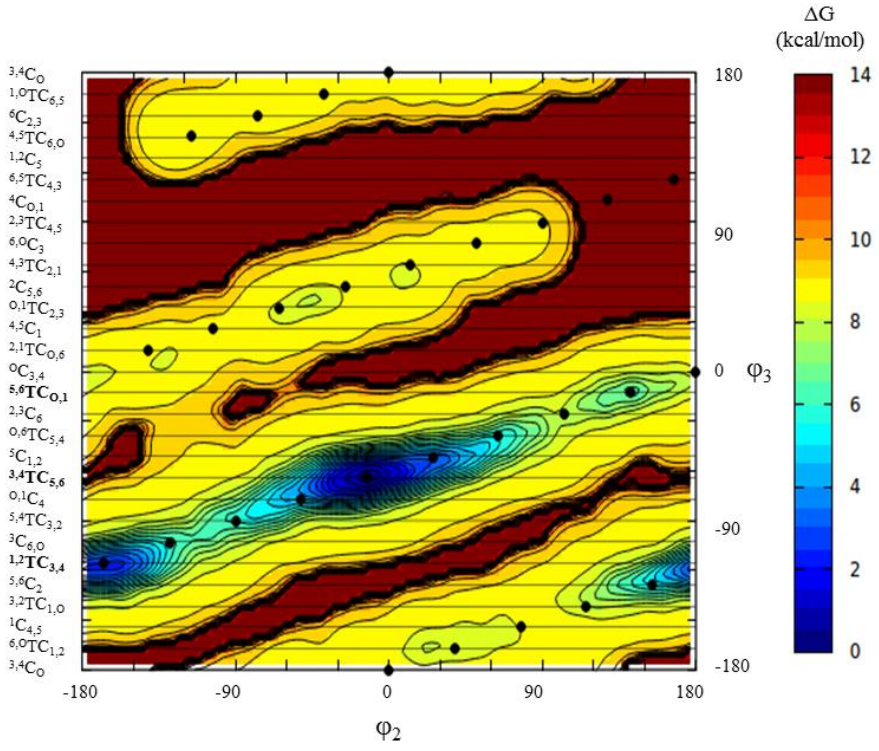


Figure E – 3. Collective variable evolution during the α-D-glycero-D-idoseptanoside conformational study using the complete set of puckering coordinates.



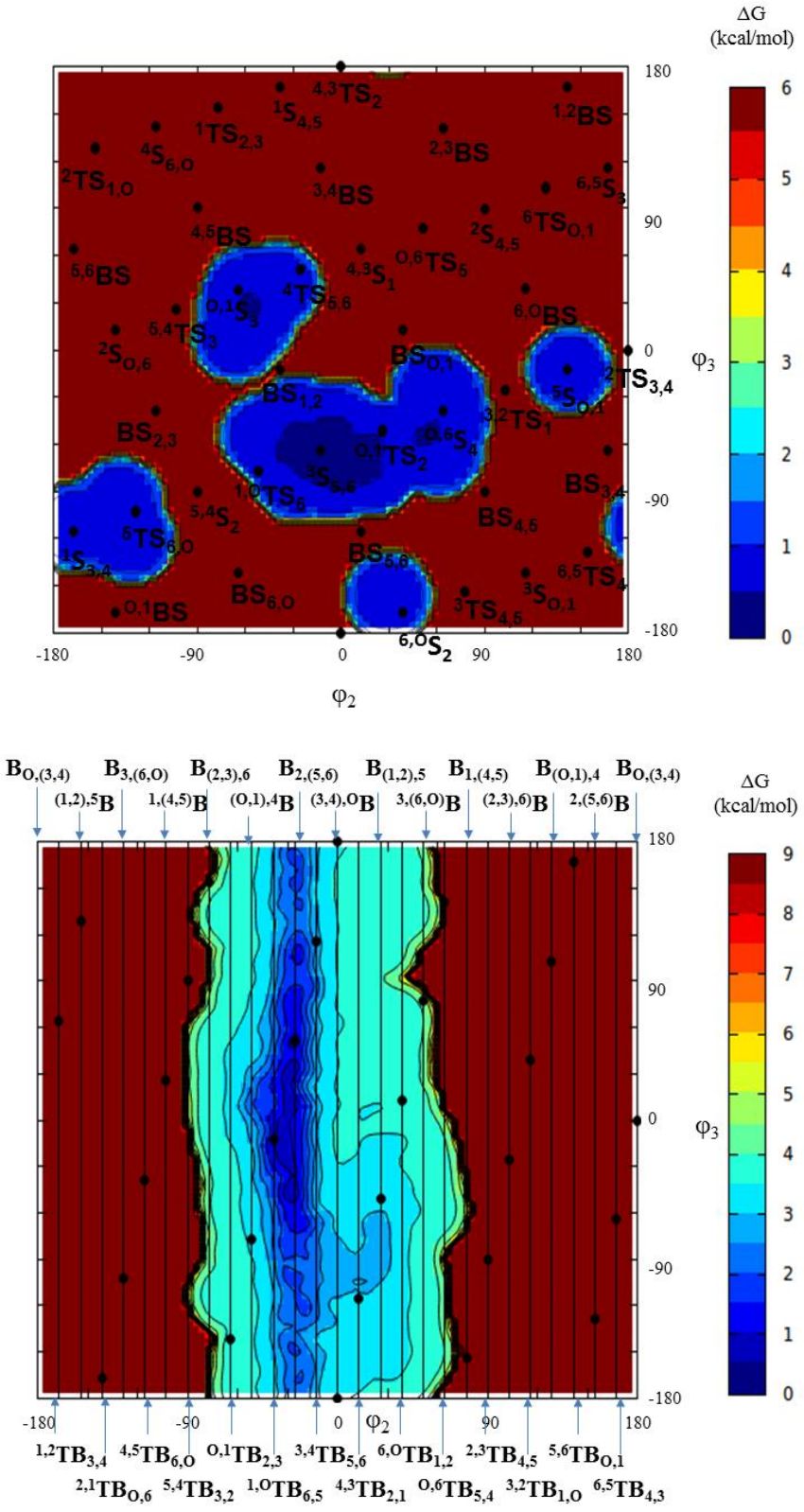


Figure E – 4. (φ_2, φ_3) projections of (up) TC/C plane (center) S/TS plane and (down) TB/B plane of the conformational FEL of α-D-glycero-D-idoseptanoside (One isoline each 0.5 kcal/mol).

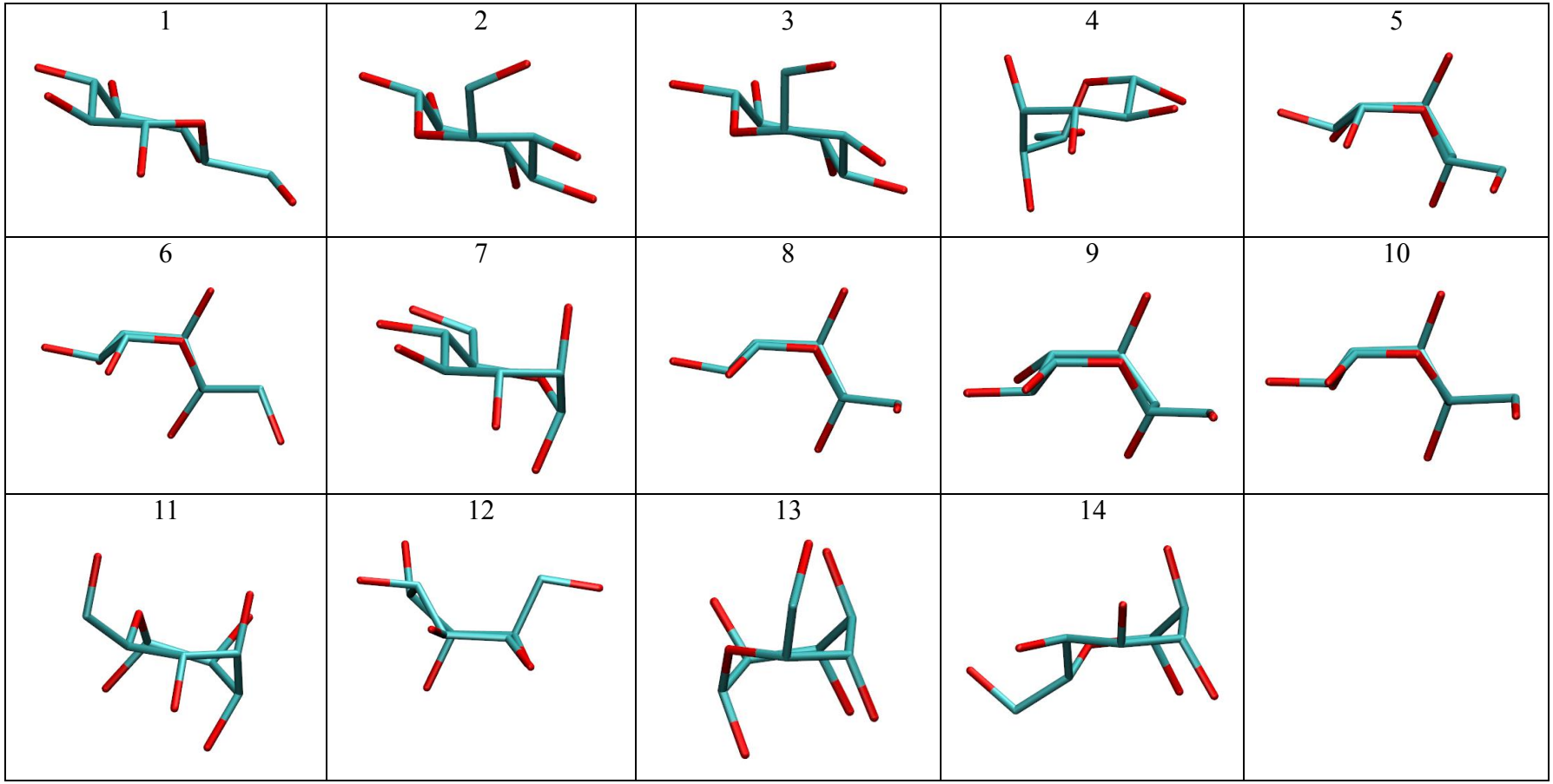
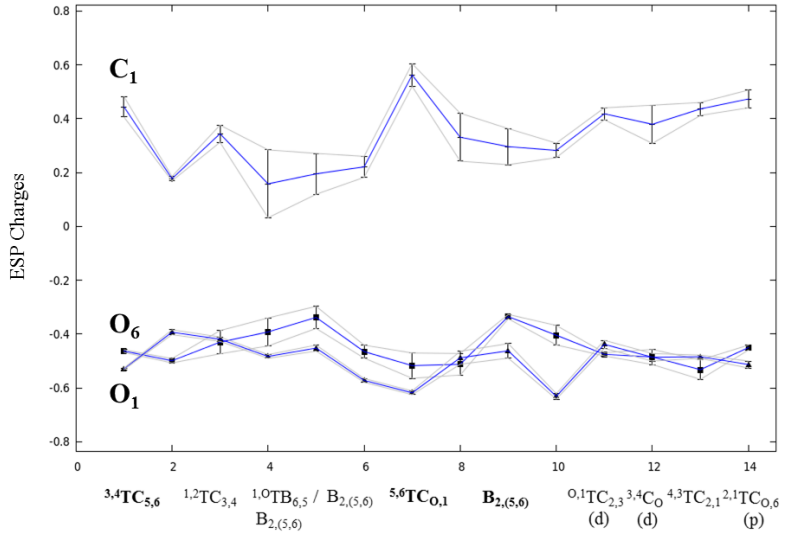
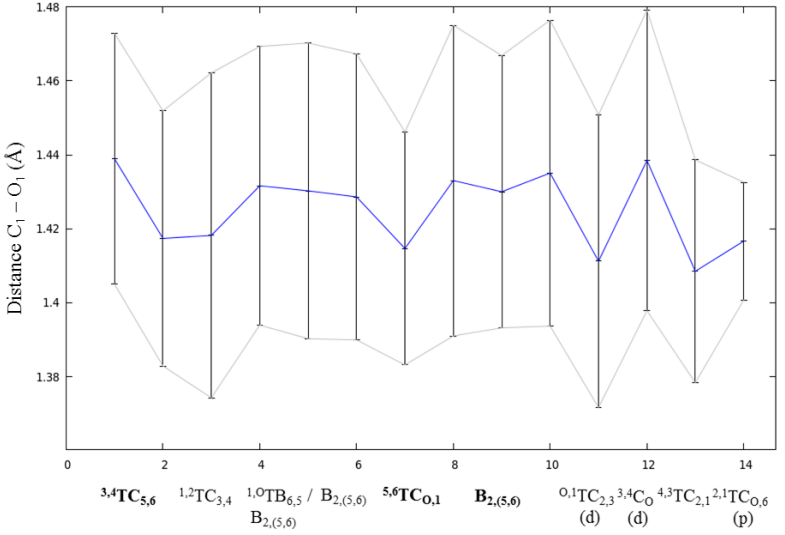


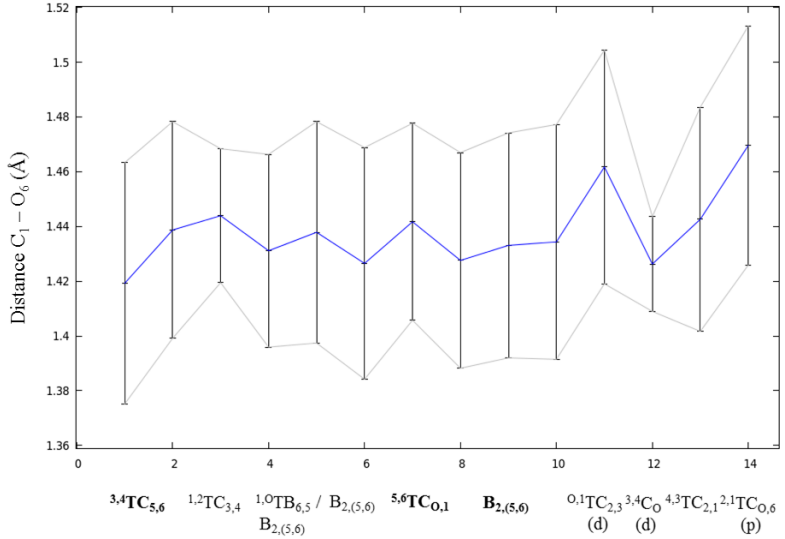
Figure E – 5. 3D representation (without hydrogen atoms) of the fourteen minima found over the conformational FEL of the α-D-glycero-D-idoseptanoside. Red atoms represent oxygen and cyan atoms represent carbon.


ESP Charges
C1
O6
O1
3,4TC5,6 1,2TC3,4 1,OTB6,5 / B2,(5,6) 5,6TCO,1 B2,(5,6) O,1TC2,3 3,4CO 4,3TC2,1 2,1TCO,6
B2,(5,6)
(d) (d) (p)

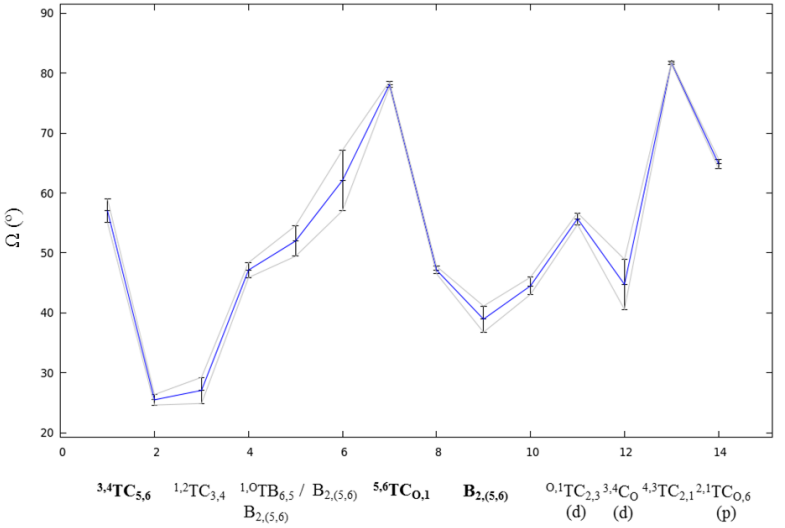
(a)


Distance C1 – O1 (Å)
3,4TC5,6 1,2TC3,4 1,OTB6,5 / B2,(5,6) 5,6TCO,1 B2,(5,6) O,1TC2,3 3,4CO 4,3TC2,1 2,1TCO,6
B2,(5,6)
(d) (d) (p)

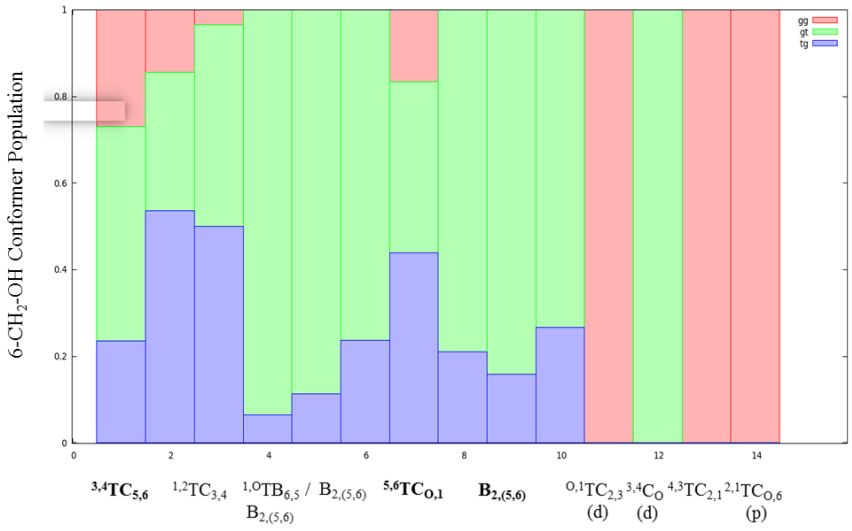
(b)


Distance C1 – O6 (Å)
3,4TC5,6 1,2TC3,4 1,OTB6,5 / B2,(5,6) 5,6TCO,1 B2,(5,6) O,1TC2,3 3,4CO 4,3TC2,1 2,1TCO,6
B2,(5,6)
(d) (d) (p)

(c)


Ω (°)
3,4TC5,6 1,2TC3,4 1,OTB6,5 / B2,(5,6) 5,6TCO,1 B2,(5,6) O,1TC2,3 3,4CO 4,3TC2,1 2,1TCO,6
B2,(5,6)
(d) (d) (p)

(d)



(e)

Figure E – 6. Preactivation analysis of the fourteen α-D-glycero-D-idoseptanoside conformers. (a) Calculated ESP charges of the C_1, O_1 and O_6 atoms, (b) calculated C_1-O_1 distances, (c) calculated C_1-O_6 distances, (d) calculated axial/equatorial C_1-O_1 Ω angle and (e) calculated relative population of the gg, gt and tg 6-CH_2-OH conformers.

Section VII – β-D-glycero-D-guloseptanoside

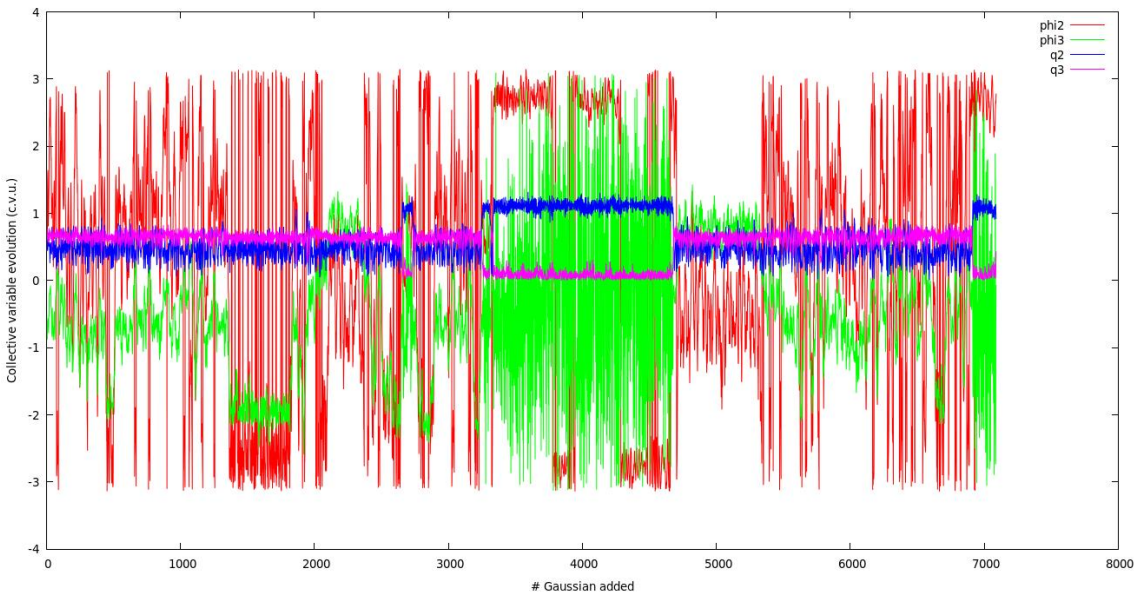
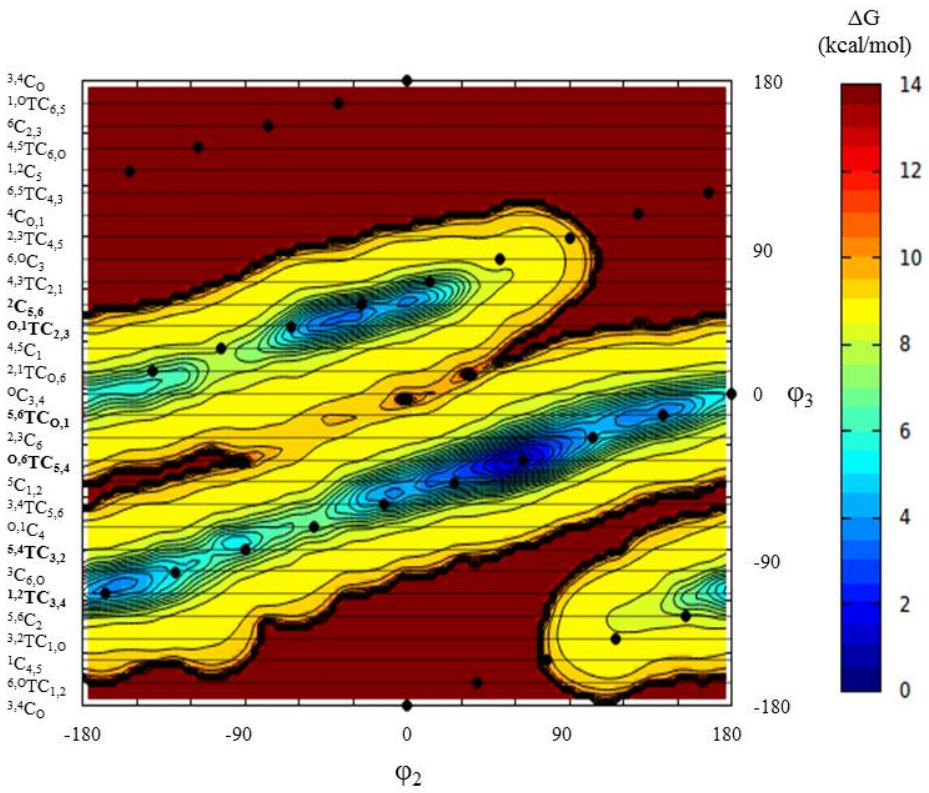
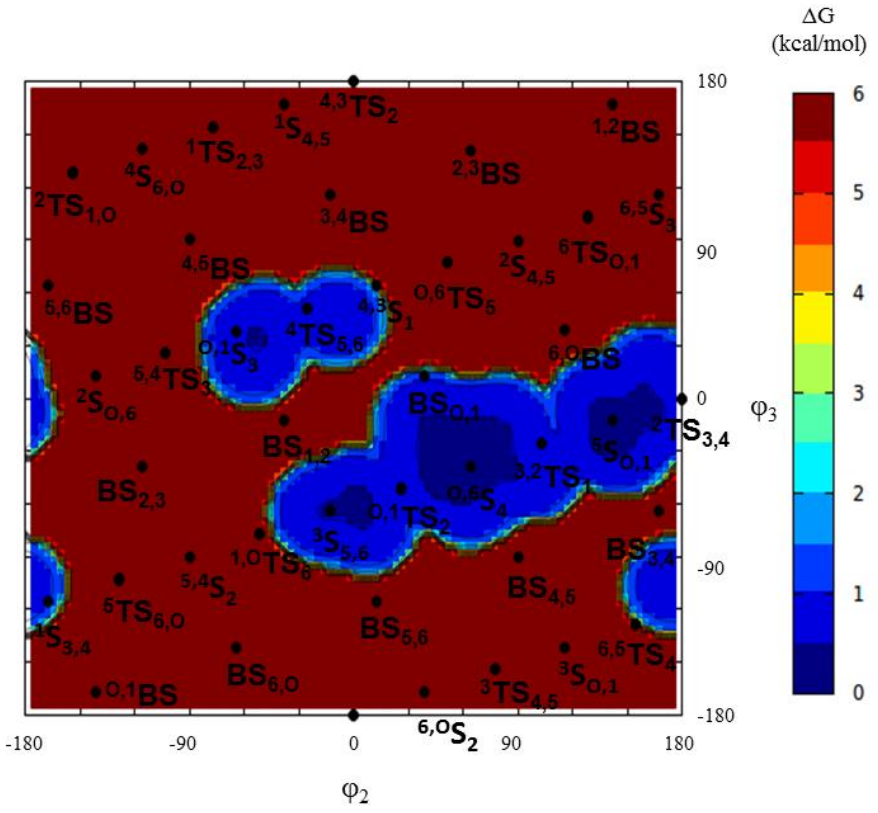


Figure E – 7. Collective variable evolution during the β-D-glycero-D-guloseptanoside conformational study using the complete set of puckering coordinates.


ΔG
(kcal/mol)
14
12
10
8
6
4
2
0
180
90
0
-90
-180
φ3
-180
-90
0
90
180
φ2
3,4CO
1,OTC6,5
6C2,3
4,5TC6,O
1,2C5
6,5TC4,3
4CO,1
2,3TC4,5
6,OC3
4,3TC2,1
2C5,6
O,1TC2,3
4,5C1
2,1TCO,6
OC3,4
5,6TCO,1
2,3C6
O,6TC5,4
5C1,2
3,4TC5,6
O,1C4
5,4TC3,2
3C6,O
1,2TC3,4
5,6C2
3,2TC1,O
1C4,5
6,OTC1,2
3,4CO


ΔG
(kcal/mol)
6
5
4
3
2
1
0
180
90
0
-90
-180
φ3
-180
-90
0
90
180
φ2
4,3TS2
1S4,5
1,2BS
1TS2,3
4S6,O
2,3BS
2TS1,O
3,4BS
6,5S3
4,5BS
2S4,5
6TSO,1
5,6BS
O,6TS5
4,3S1
4TS5,6
O,1S3
6,OBS
5,4TS3
2SO,6
BSO,1
2TS3,4
BS1,2
5SO,1
3,2TS1
BS2,3
O,6S4
O,1TS2
3S5,6
BS3,4
1,OTS6
5,4S2
BS4,5
5TS6,O
BS5,6
1S3,4
6,5TS4
BS6,O
3SO,1
O,1BS
3TS4,5
6,OS2

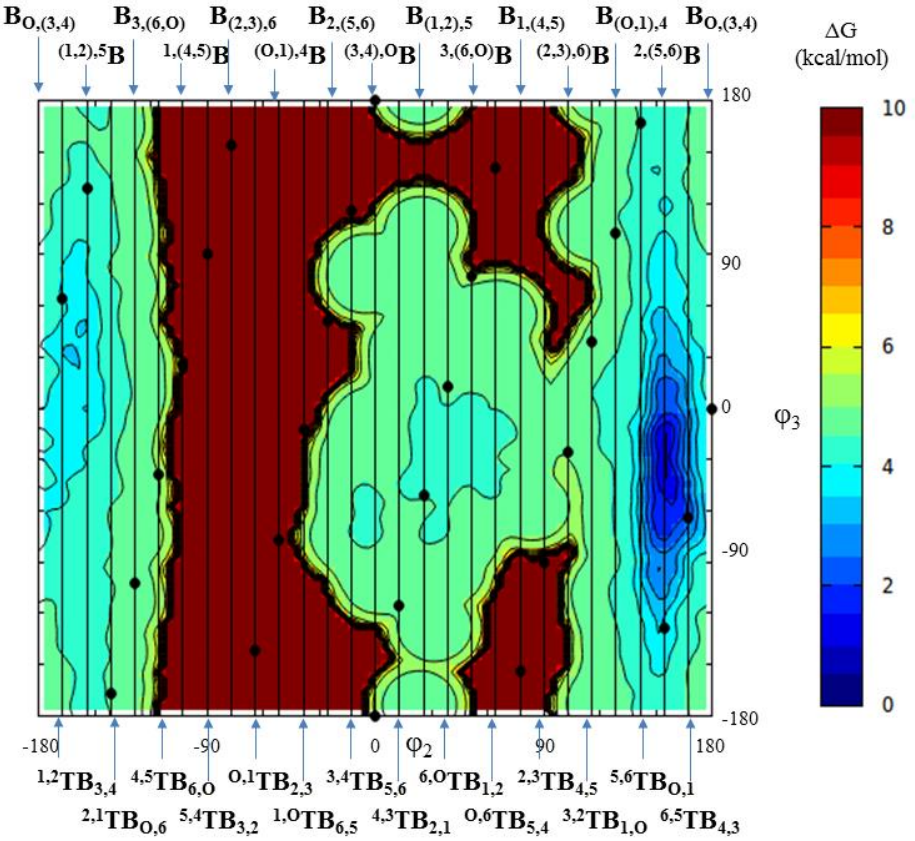


Figure E – 8. (φ_2, φ_3) projections of (up) TC/C plane (center) S/TS plane and (down) TB/B plane of the conformational FEL of β-D-glycero-D-guloseptanoside (One isoline each 0.5 kcal/mol).

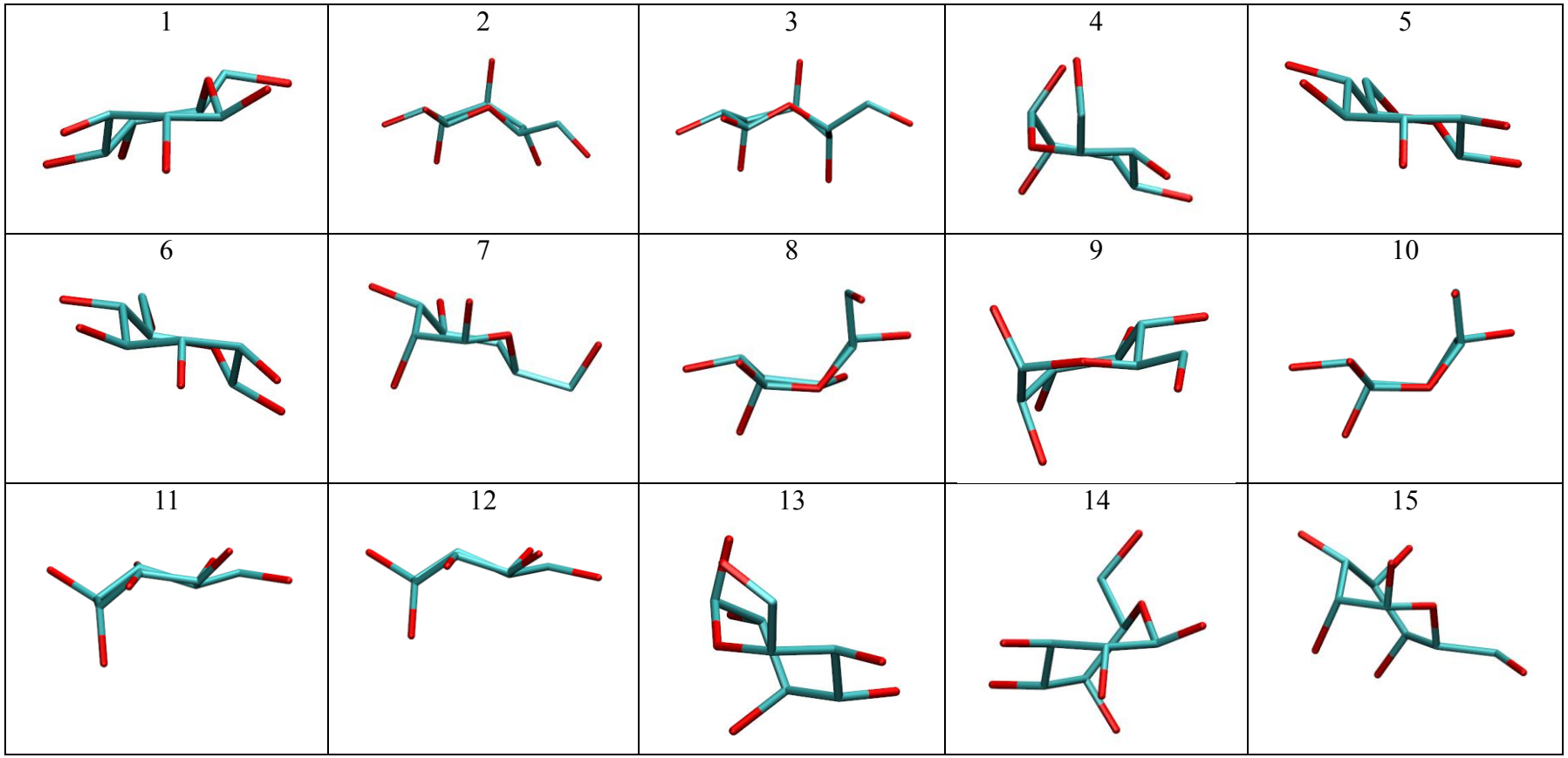
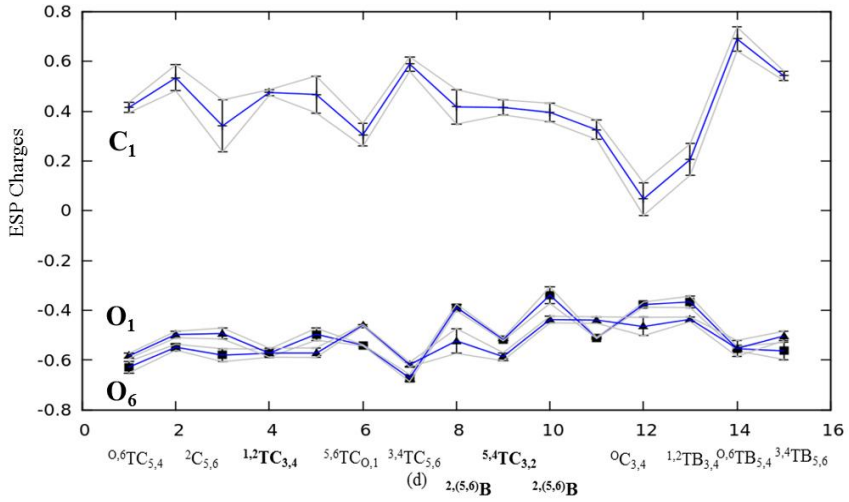
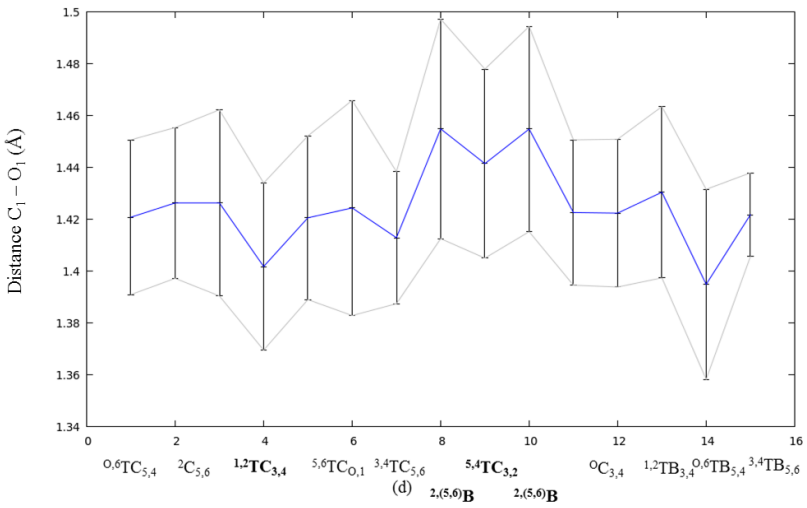


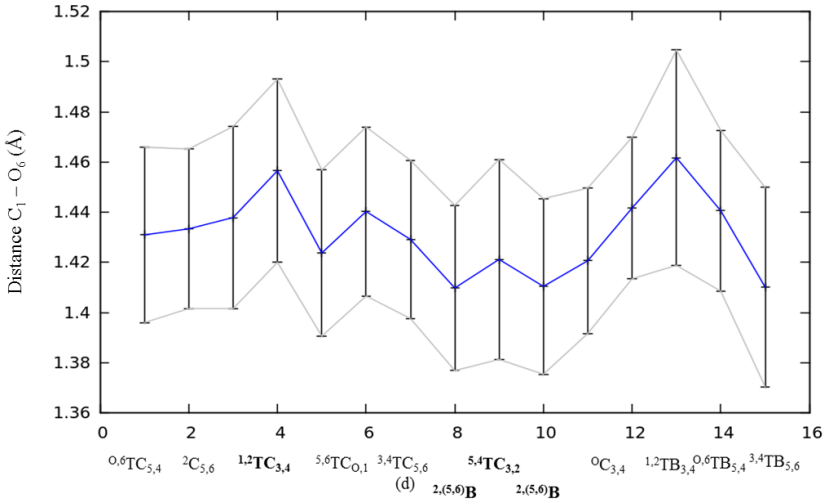
Figure E – 9. 3D representation (without hydrogen atoms) of the fourteen minima found over the conformational FEL of the β-D-glycero-D-guloseptanoside. Red atoms represent oxygen and cyan atoms represent carbon.


ESP Charges
C1
O1
O6
(d)
2,(5,6)B
2,(5,6)B

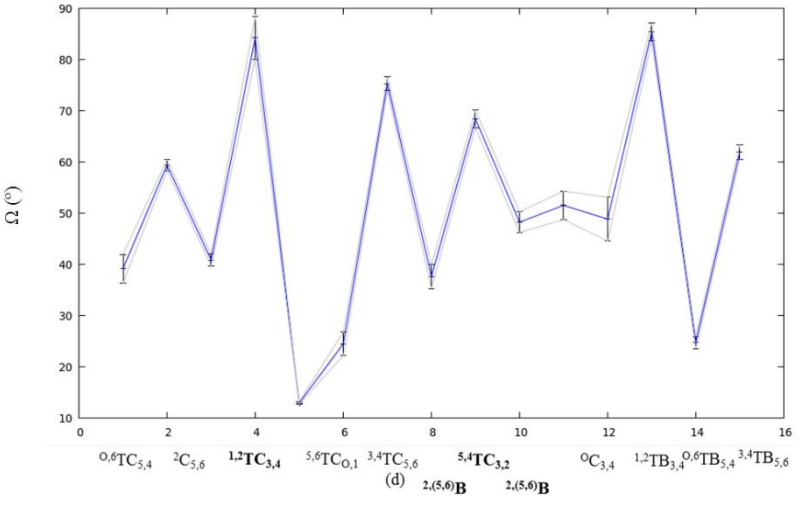
(a)


Distance C1 – O1 (Å)
(d)
2,(5,6)B
2,(5,6)B

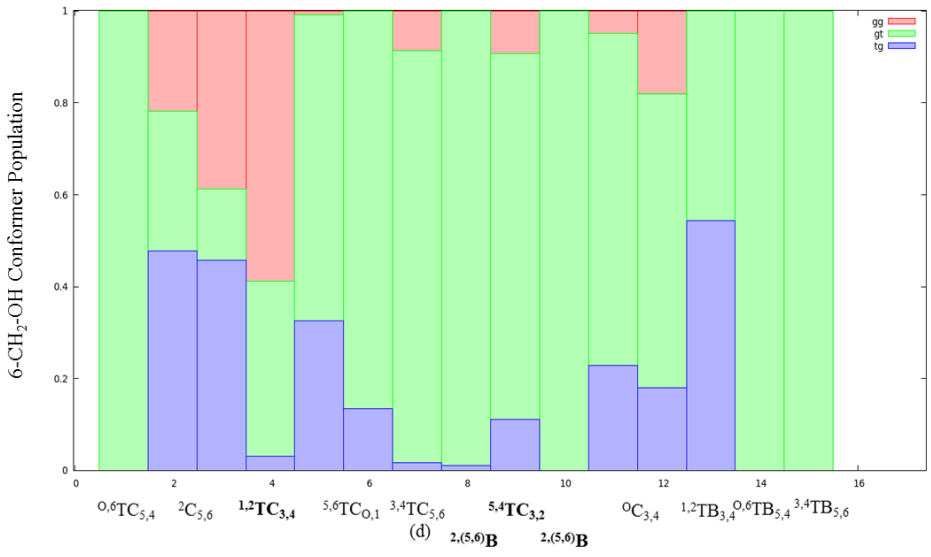
(b)


Distance C1 – O6 (Å)
(d)
2,(5,6)B
2,(5,6)B

(c)


Ω (°)
(d)
2,(5,6)B
2,(5,6)B

(d)



(e)

Figure E – 10. Preactivation analysis of the fourteen β-D-glycero-D-guloseptanoside conformers. (a) Calculated ESP charges of the C_1, O_1 and O_6 atoms, (b) calculated C_1-O_1 distances, (c) calculated C_1-O_6 distances, (d) calculated axial/equatorial C_1-O_1 Ω angle and (e) calculated relative population of the gg, gt and tg 6-CH_2-OH conformers.

Section VII – oxocarbenium septanoside cation

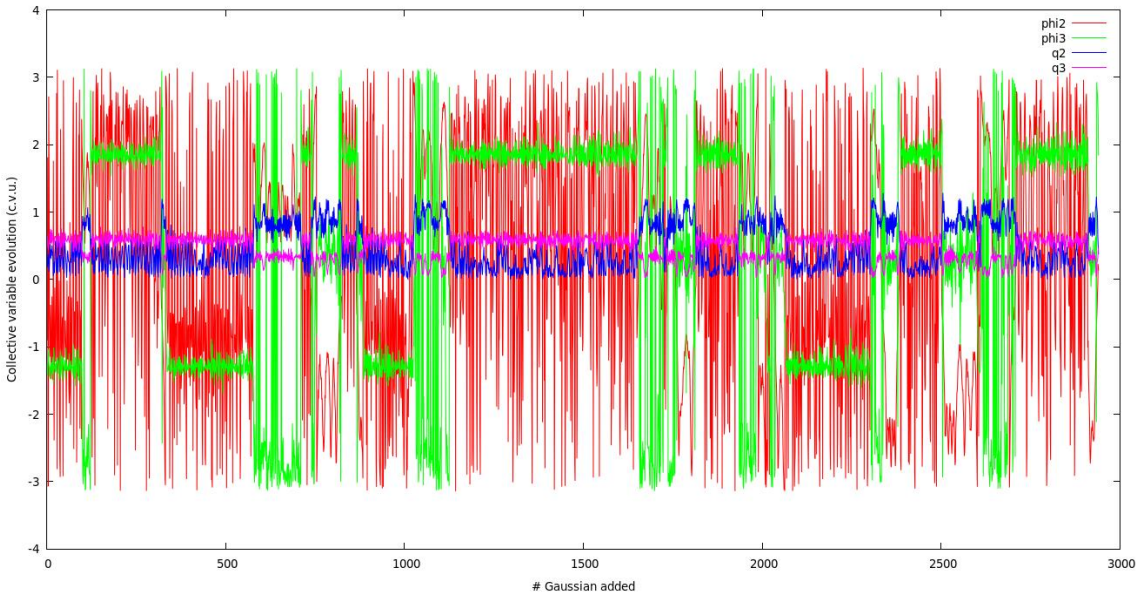


Figure E – 11. Collective variable evolution during the oxocarbenium septanoside cation conformational study using the complete set of puckering coordinates.

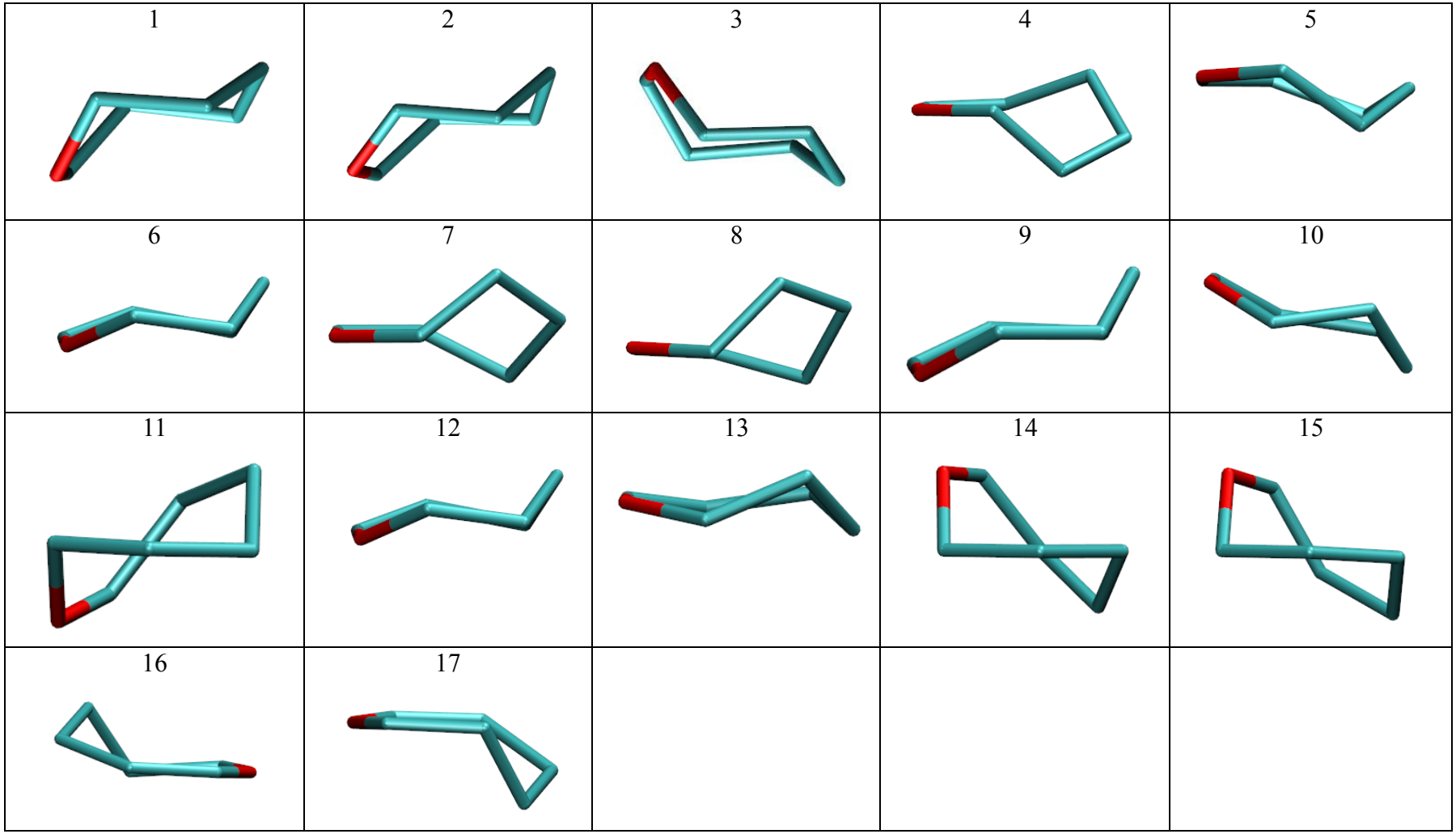
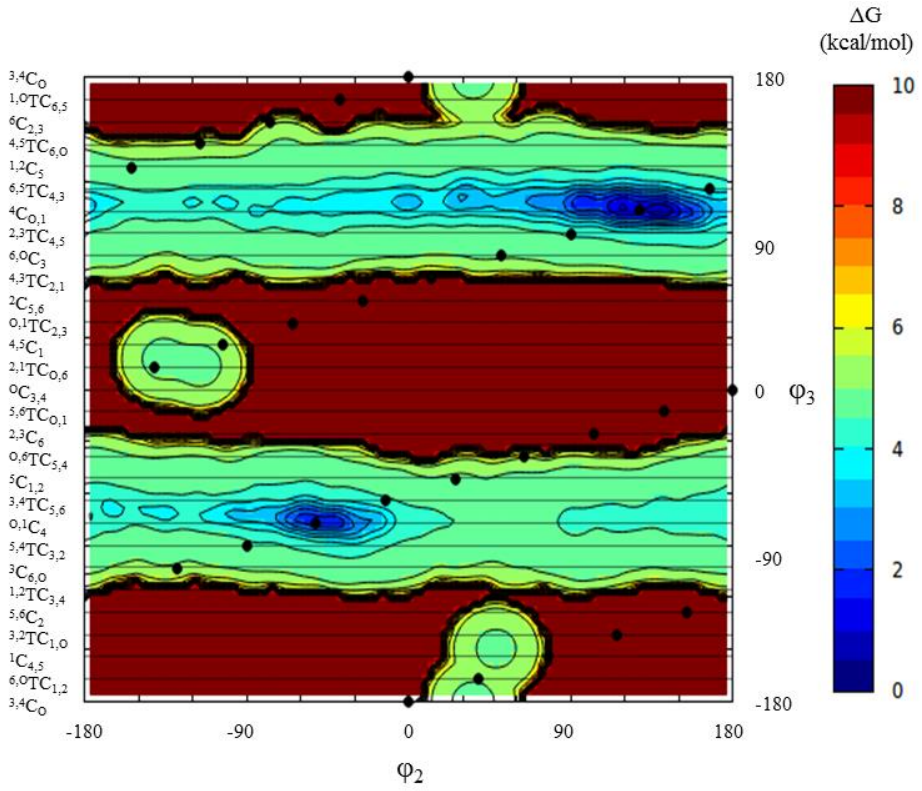
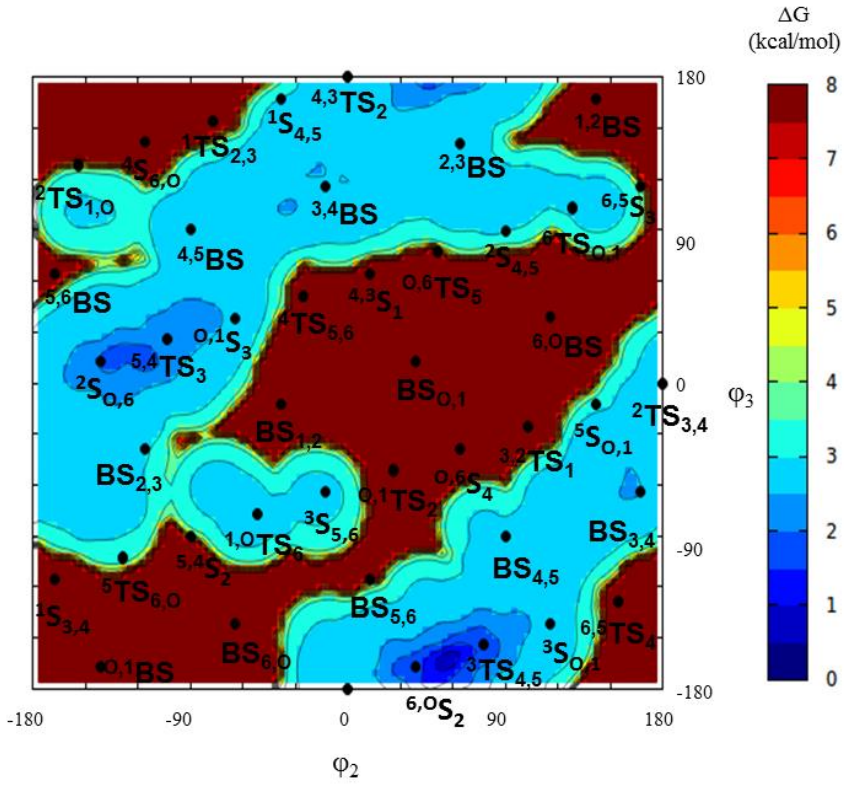


Figure E – 12. 3D representation (without hydrogen atoms) of the seventeen minima found over the conformational FEL of the oxocarbenium septanoside cation. Red atoms represent oxygen and cyan atoms represent carbon.


ΔG
(kcal/mol)
10
8
6
4
2
0
180
90
0
-90
-180
φ3
-180
-90
0
90
180
φ2
3,4CO
1,OTC6,5
6C2,3
4,5TC6,O
1,2C5
6,5TC4,3
4CO,1
2,3TC4,5
6,OC3
4,3TC2,1
2C5,6
O,1TC2,3
4,5C1
2,1TCO,6
OC3,4
5,6TCO,1
2,3C6
O,6TC5,4
5C1,2
3,4TC5,6
O,1C4
5,4TC3,2
3C6,O
1,2TC3,4
5,6C2
3,2TC1,O
1C4,5
6,OTC1,2
3,4CO


ΔG
(kcal/mol)
8
7
6
5
4
3
2
1
0
180
90
0
-90
-180
φ3
-180
-90
0
90
180
φ2
4,3TS2
1S4,5
1TS2,3
1,2BS
4S6,O
2,3BS
2TS1,O
3,4BS
6,5S5
4,5BS
2S4,5
6TSO,1
O,6TS5
4,3S1
5,6BS
4TS5,6
O,1S3
6,OBS
5,4TS3
2SO,6
BSO,1
2TS3,4
5SO,1
BS1,2
3,2TS1
BS2,3
O,6S4
O,1TS2
3S5,6
1,OTS6
BS3,4
5,4S2
BS4,5
5TS6,O
BS5,6
1S3,4
6,5TS4
BS6,O
3SO,1
O,1BS
3TS4,5
6,OS2

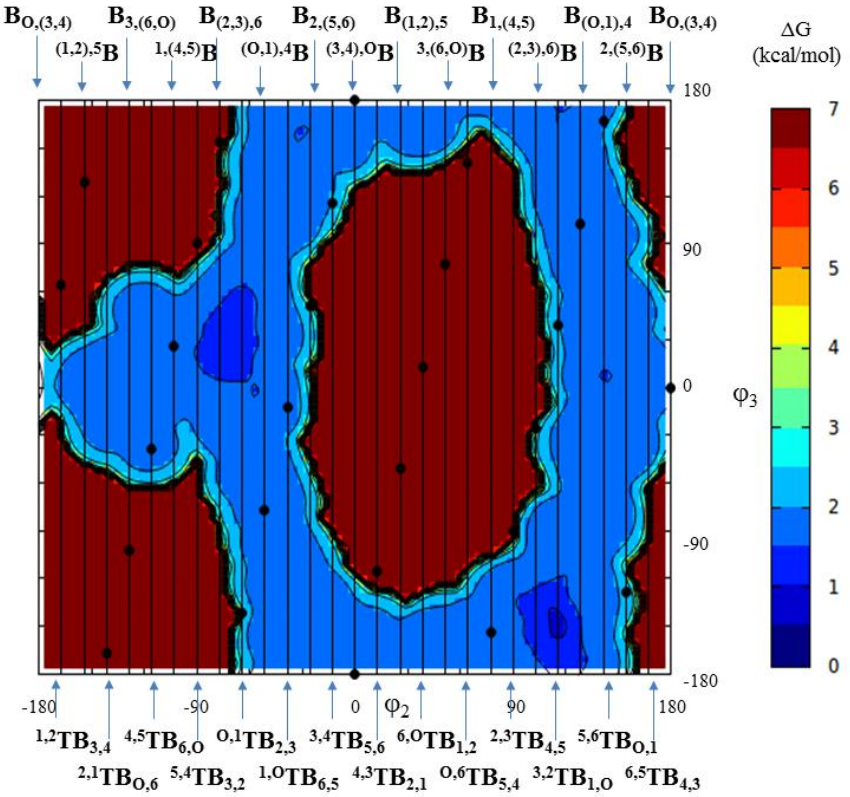


Figure E – 13. (φ_2, φ_3) projections of (up) TC/C plane (center) S/TS plane and (down) TB/B plane of the conformational FEL of oxocarbenium septanoside cation (One isoline each 0.5 kcal/mol).

Section VII – β-L-idoseptanoside

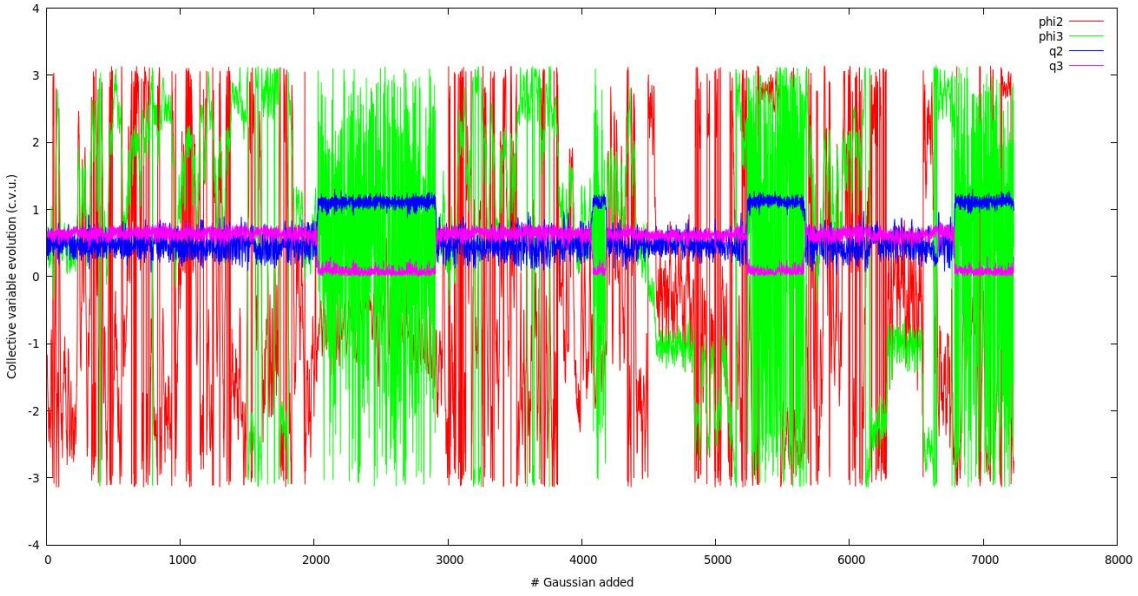


Figure E – 14. Collective variable evolution during the β-L-idoseptanoside conformational study using the complete set of puckering coordinates.

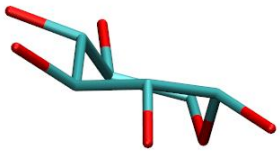
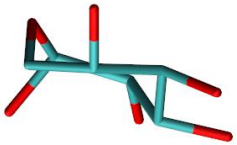
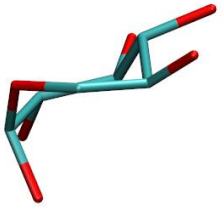
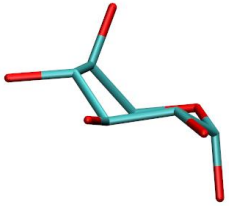
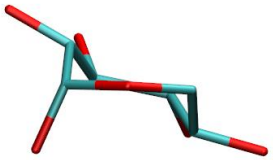
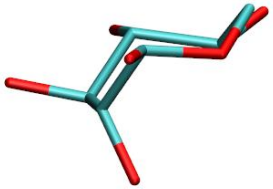
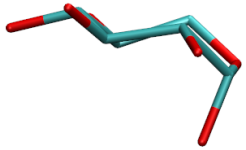
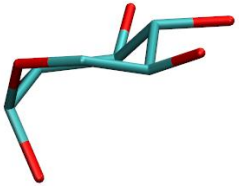
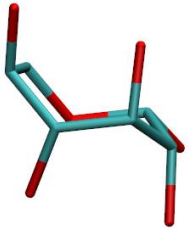
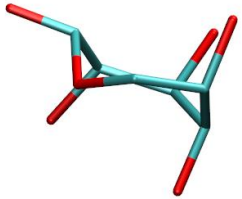
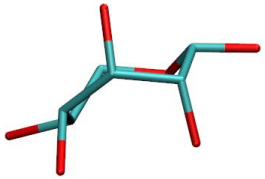
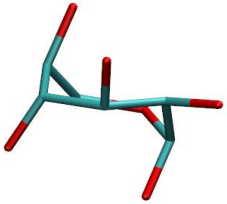
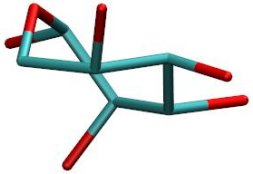
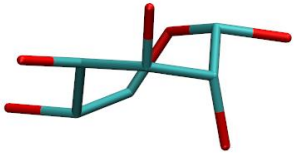
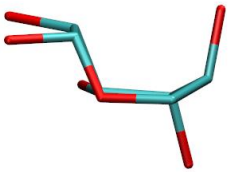
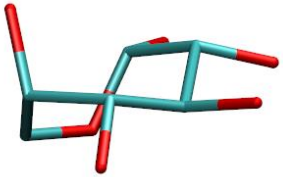
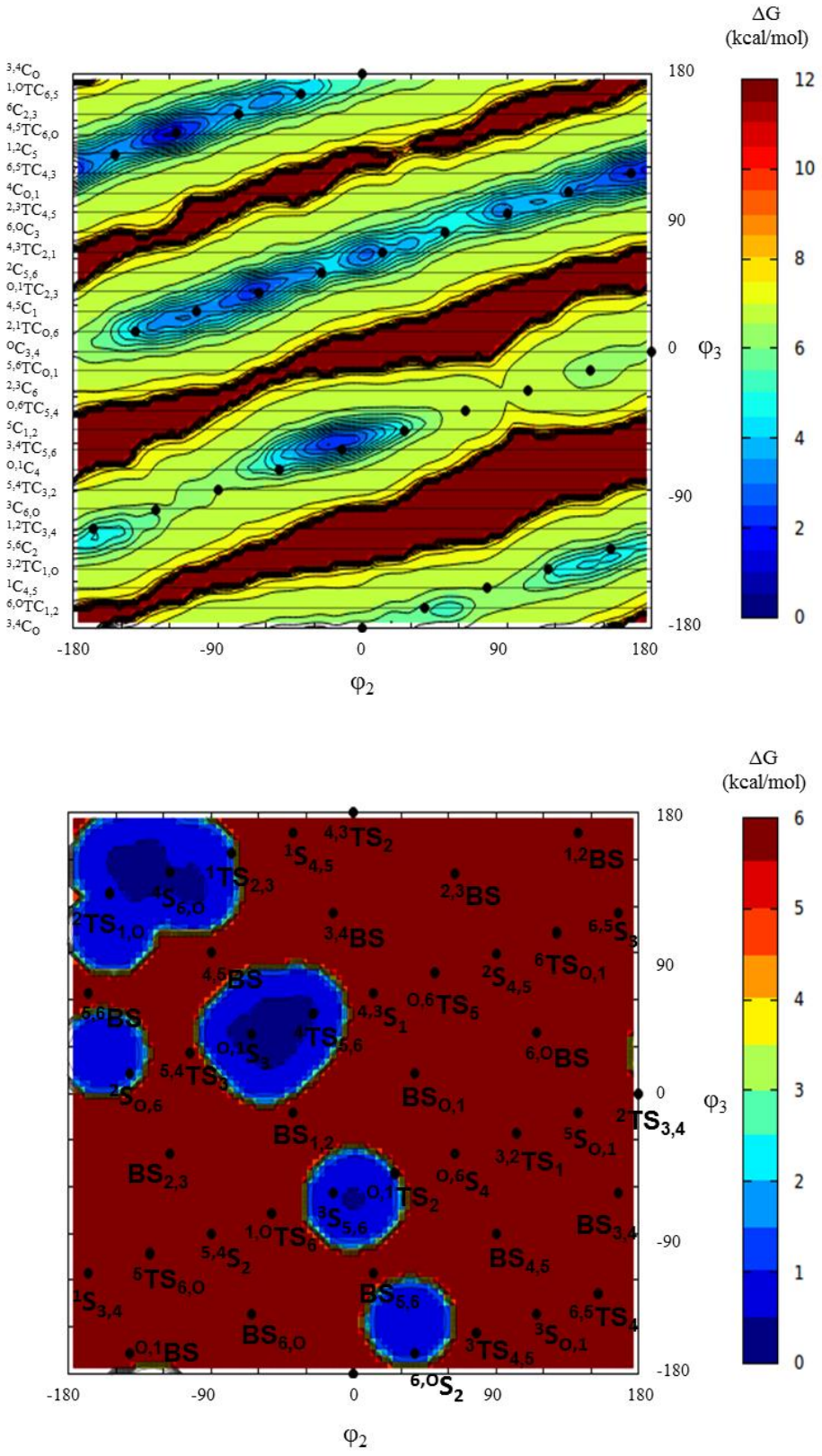
1 	2 	3 	4 	5 
6 	7 	8 	9 	10 
11 	12 	13 	14 	15 
16 	17 	18 		

Figure E – 15. 3D representation (without hydrogen atoms) of the eighteen minima found over the conformational FEL of the β-L-idoseptanoside. Red atoms represent oxygen and cyan atoms represent carbon.



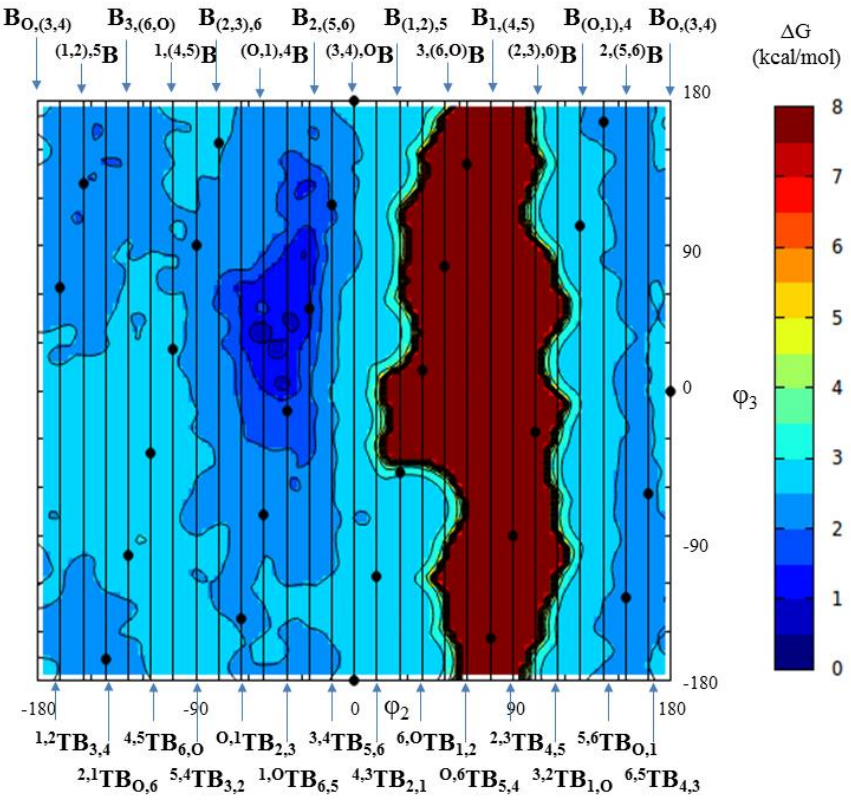
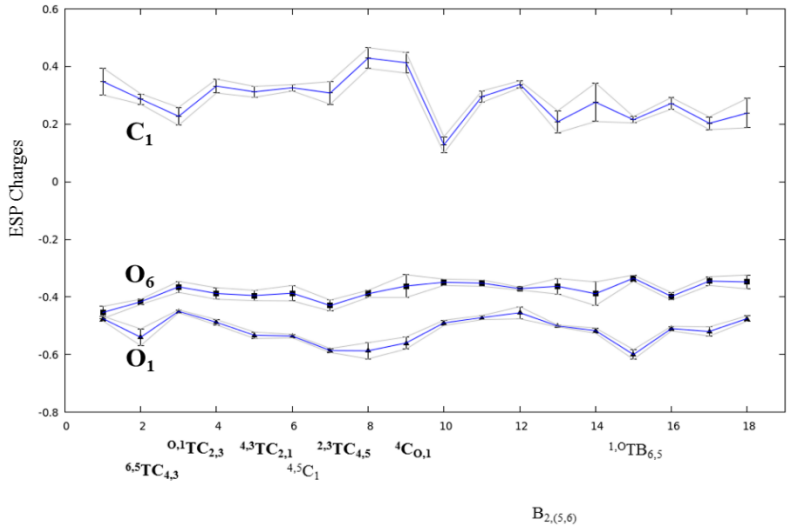
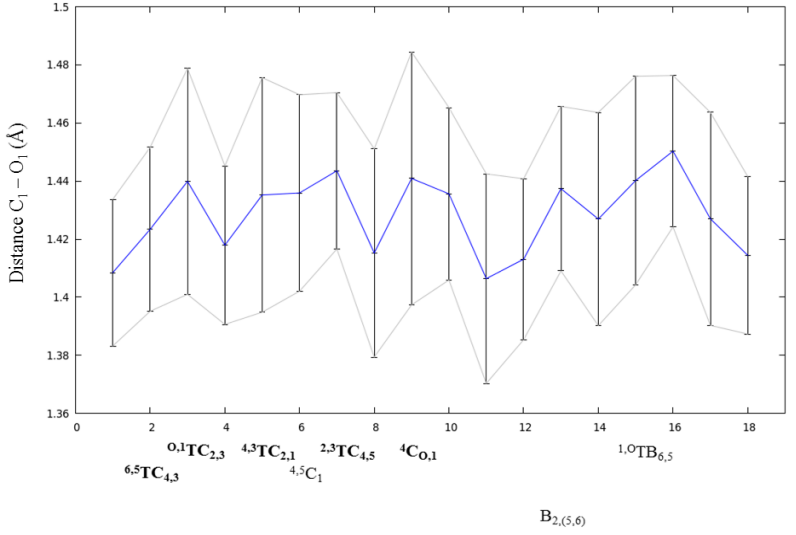


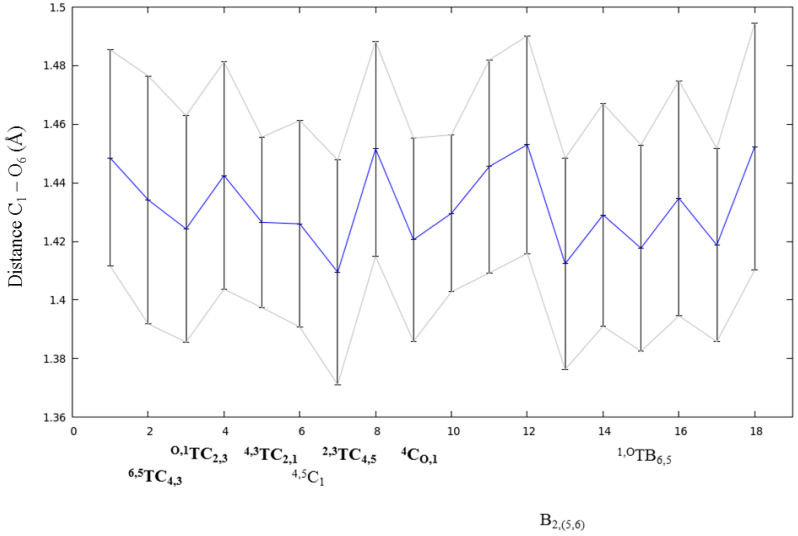
Figure E – 16. (φ_2, φ_3) projections of (up) TC/C plane (center) S/TS plane and (down) TB/B plane of the conformational FEL of β-L-idoseptanoside (One isoline each 0.5 kcal/mol).


ESP Charges
C1
O6
O1
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
0
2
4
6
8
10
12
14
16
18
O,1TC2,3
4,3TC2,1
2,3TC4,5
4Co,1
1,OTB6,5
6,5TC4,3
4,5C1
B2,(5,6)

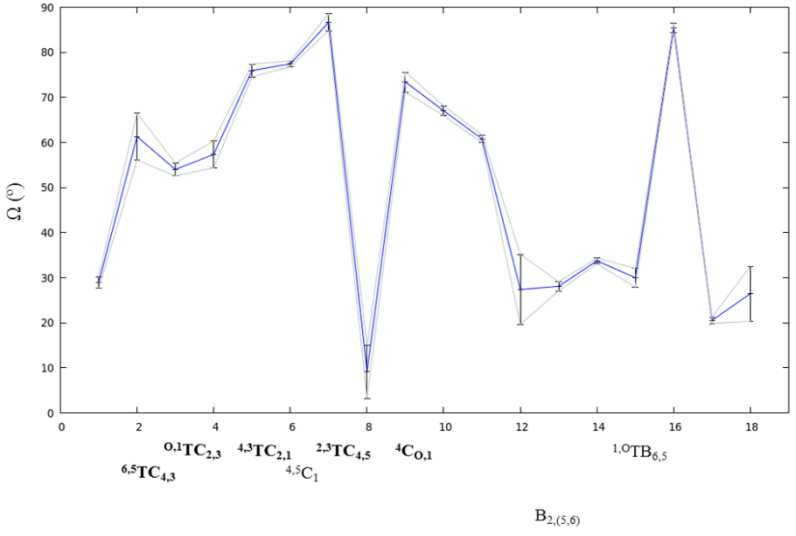
(a)


Distance C1 – O1 (Å)
1.5
1.48
1.46
1.44
1.42
1.4
1.38
1.36
0
2
4
6
8
10
12
14
16
18
O,1TC2,3
4,3TC2,1
2,3TC4,5
4Co,1
1,OTB6,5
6,5TC4,3
4,5C1
B2,(5,6)

(b)


Distance C1 – O6 (Å)
1.5
1.48
1.46
1.44
1.42
1.4
1.38
1.36
0
2
4
6
8
10
12
14
16
18
O,1TC2,3
4,3TC2,1
2,3TC4,5
4Co,1
1,OTB6,5
6,5TC4,3
4,5C1
B2,(5,6)

(c)


Ω (°)
90
80
70
60
50
40
30
20
10
0
0
2
4
6
8
10
12
14
16
18
O,1TC2,3
4,3TC2,1
2,3TC4,5
4Co,1
1,OTB6,5
6,5TC4,3
4,5C1
B2,(5,6)

(d)

Figure E – 17. Preactivation analysis of the fourteen β-D-glycero-D-guloseptanoside conformers. (a) Calculated ESP charges of the C_1, O_1 and O_6 atoms, (b) calculated C_1-O_1 distances, (c) calculated C_1-O_6 distances and (d) calculated axial/equatorial C_1-O_1 Ω angle.

Section VII – β-L-idoseptanosyl ring at the subsite -1 of the GH31

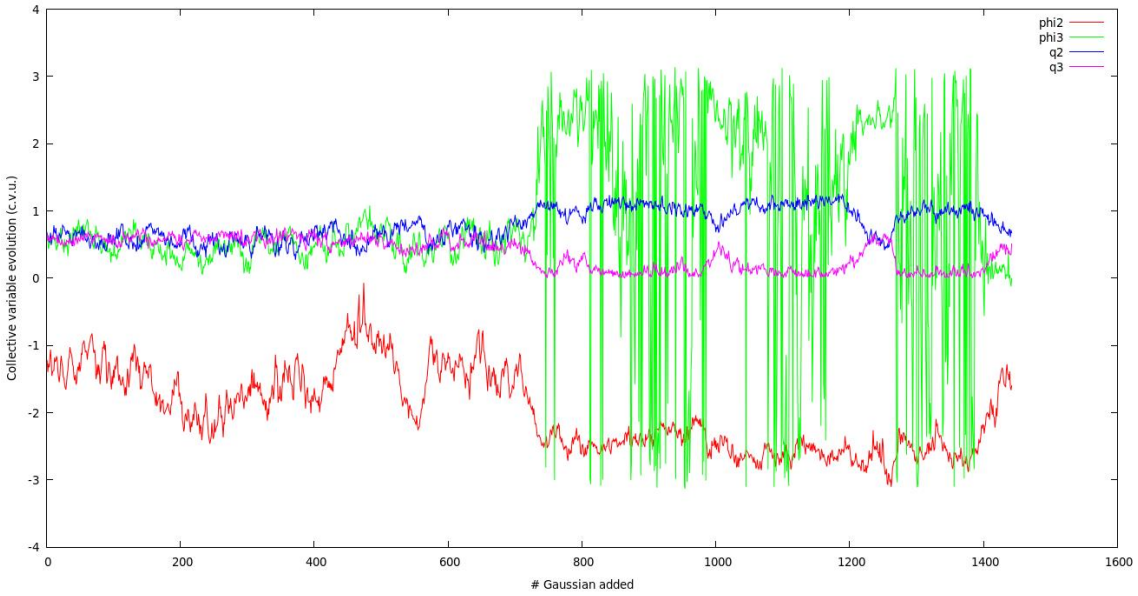
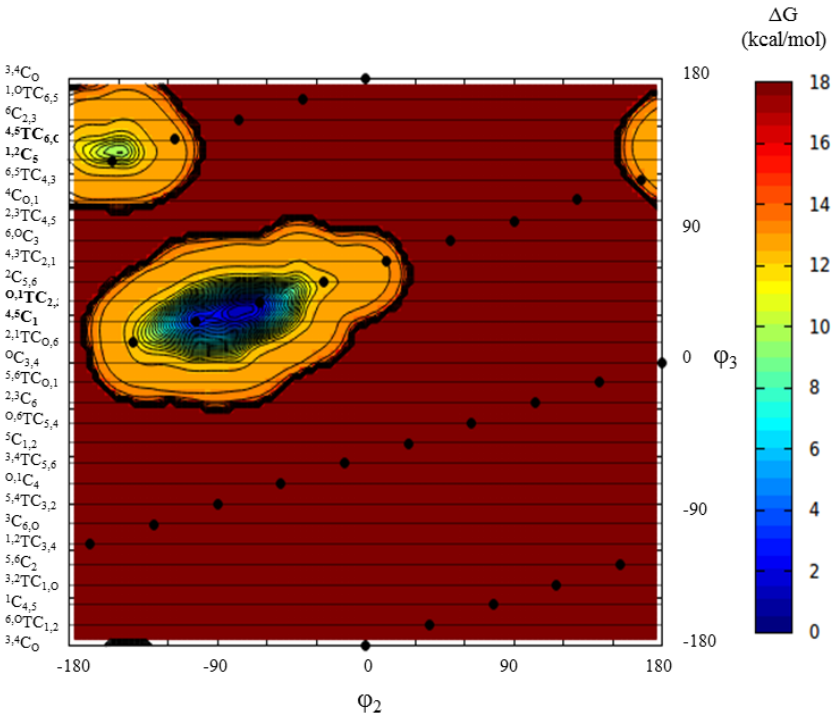


Figure E – 18. Collective variable evolution during the conformational study of the β-L-idoseptanosyl ring at the subsite -1 in the GH31 using the complete set of puckering coordinates.



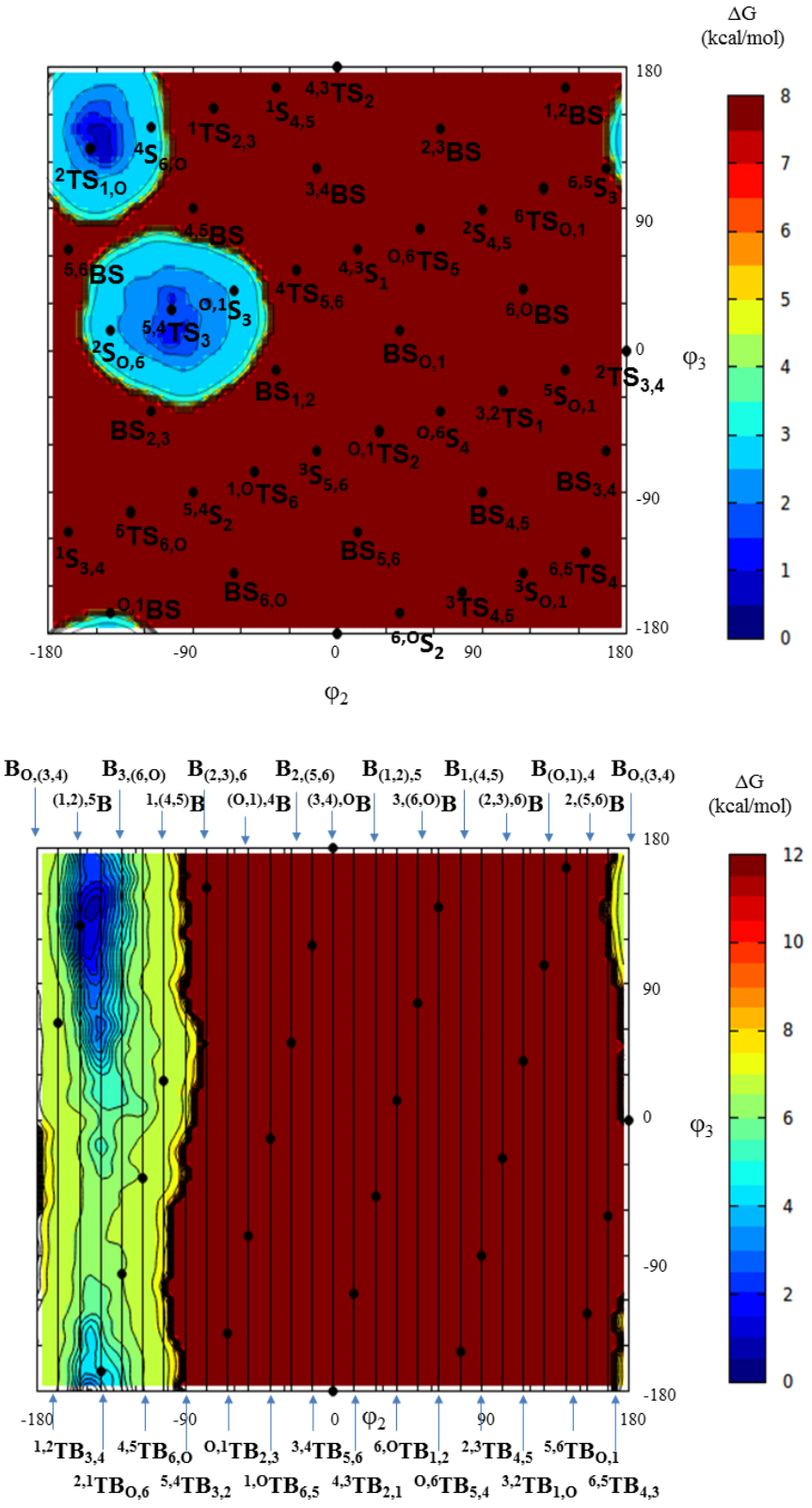
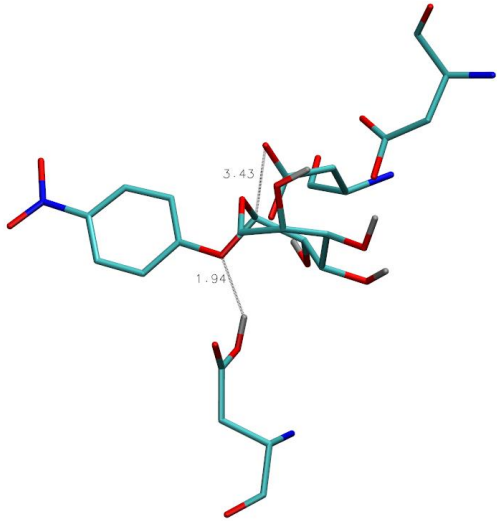
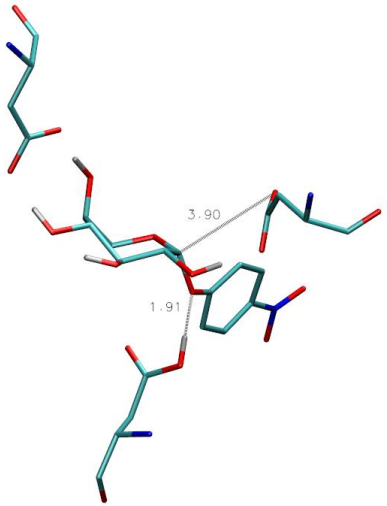


Figure E – 19. (φ_2, φ_3) projections of (up) TC/C plane (center) S/TS plane and (down) TB/B plane of the conformational FEL of the β-L-idoseptanosyl ring at the subsite -1 of the GH31 (One isoline each 0.5 kcal/mol).

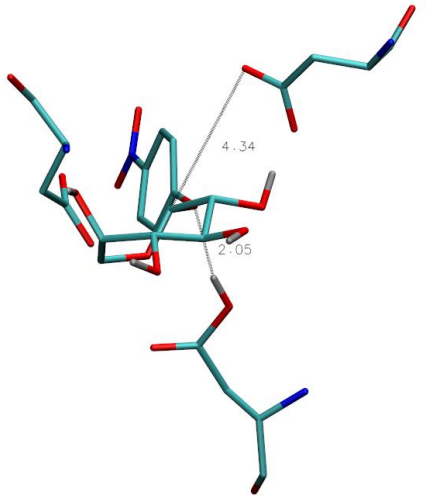
1



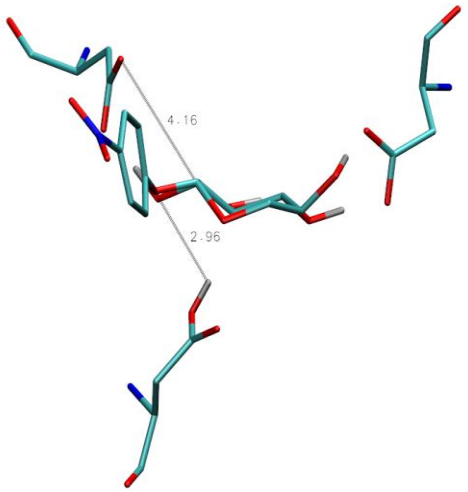
2



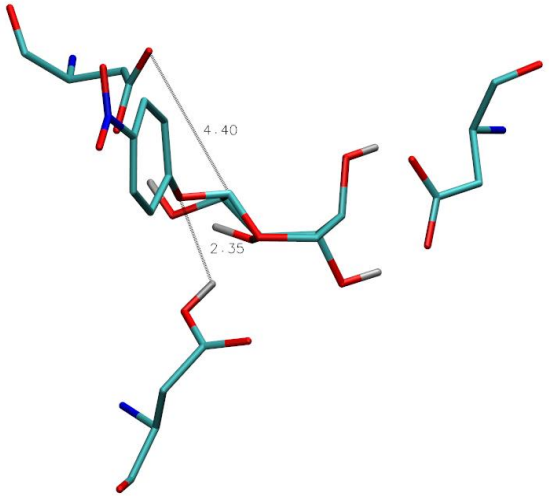
3



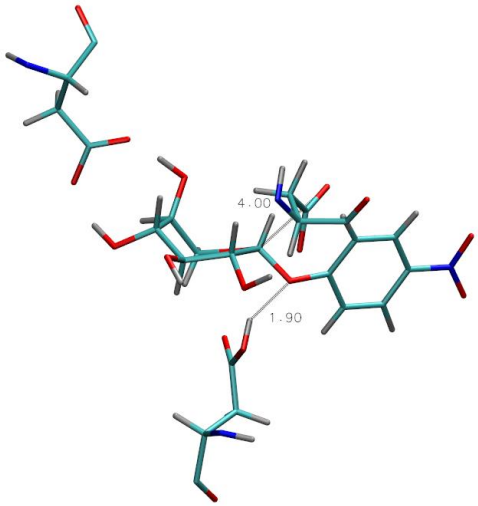
4



5



6



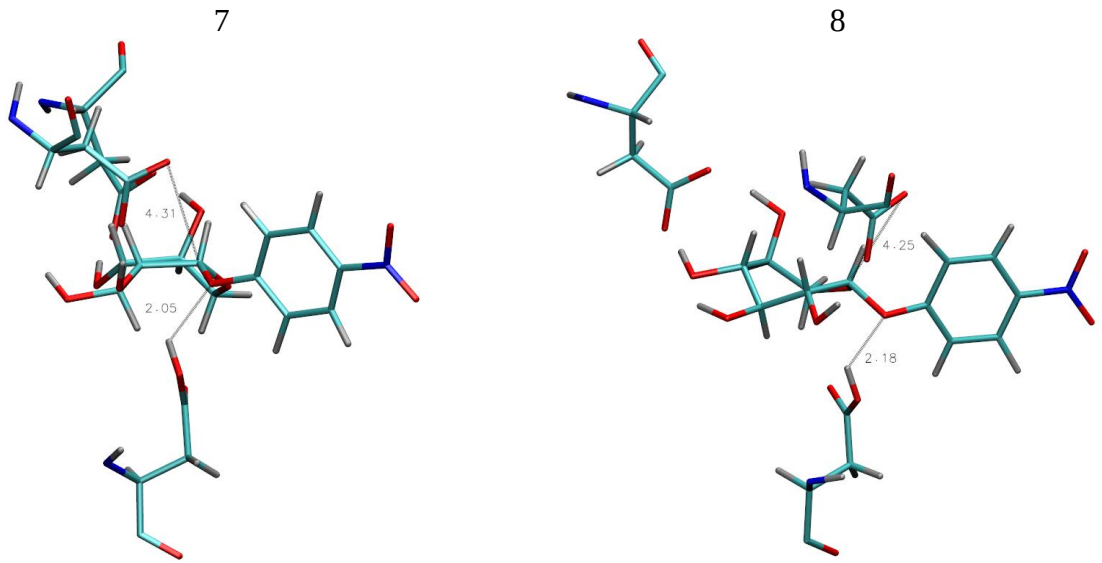


Figure E – 20. 3D representation of the eight minima found over the conformational FEL of the the β-L-idoseptanosyl ring at the subsite -1 of the GH31 (QM/MM). Red atoms represent oxygen, cyan atoms represent carbon, blue atoms represent nitrogen and grey ones hydrogen atoms.

Section VII - Glycosylation step simulation in the GH31 – pNP- β-L-idoseptanoside complex

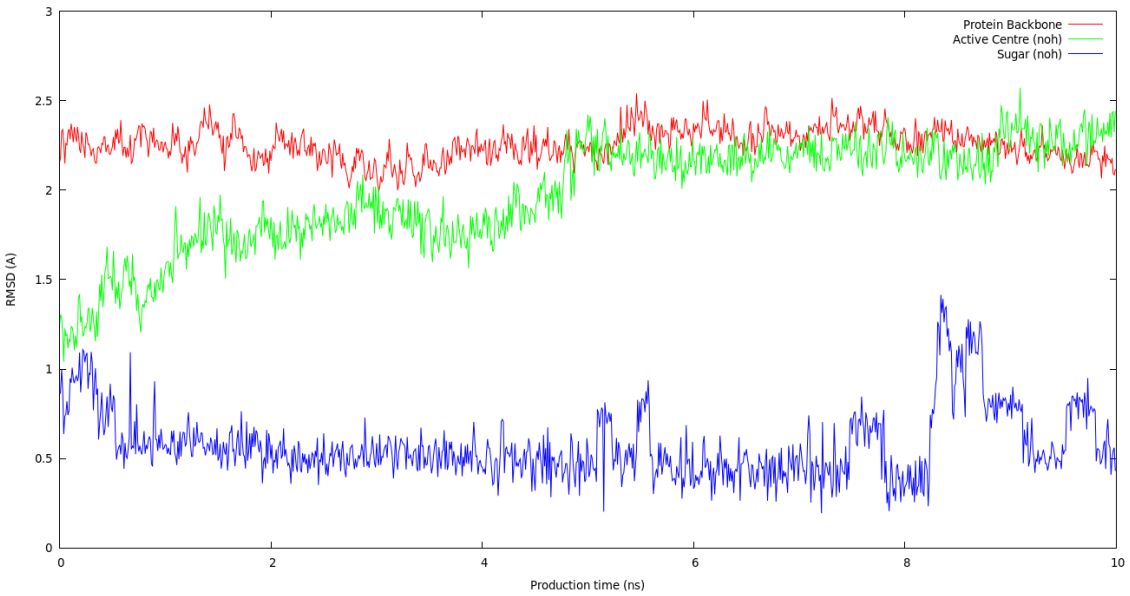


Figure E – 21. Representation of the RMSD of: Protein backbone (Red), active center residues (Green) and substrate (Blue) (Units = Å) during the NVT molecular dynamics simulation of the GH31 – pNP- β-L-idoseptanoside model.

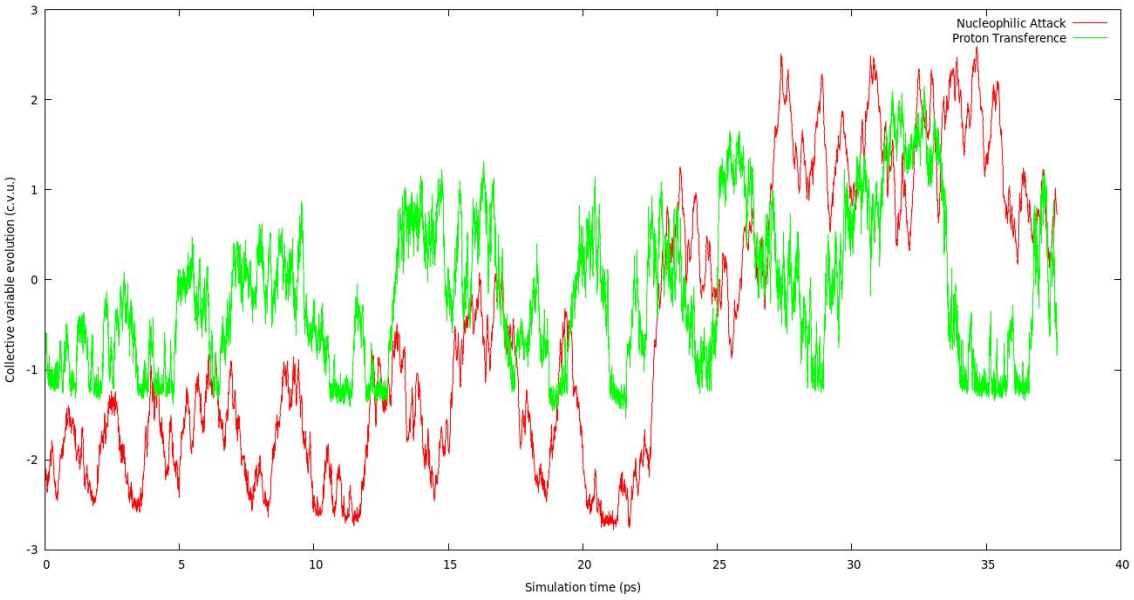


Figure E – 22. Collective variable evolution during the glycosylation step simulation in the GH31 – pNP- β-L-idoseptanoside complex.

Section VII - Glycosylation step simulation in the GH13 – α-D-gulosepta-1,4-α-glucose complex

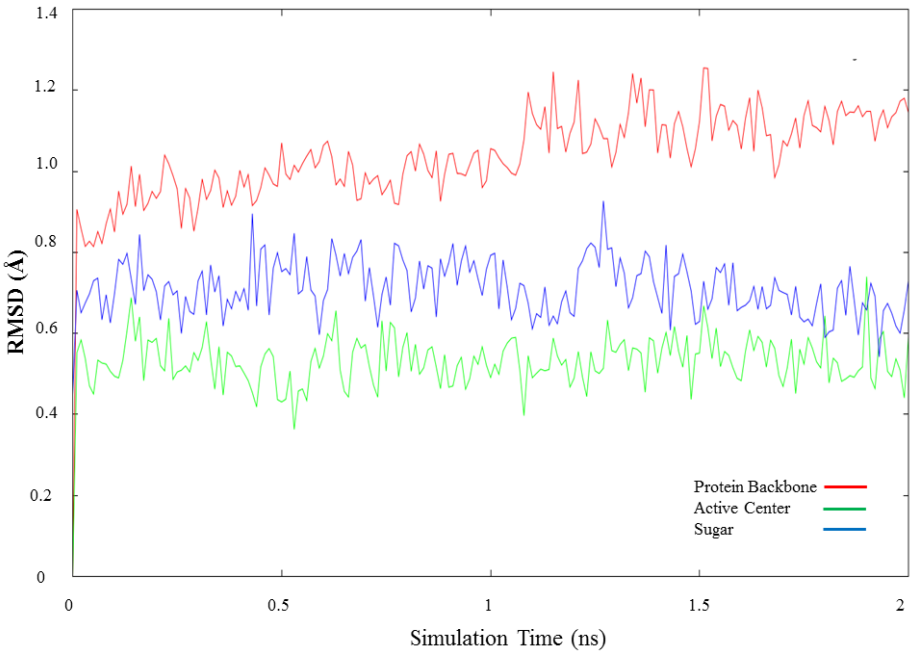


Figure E – 21. Representation of the RMSD of: Protein backbone (Red), active center residues (Green) and substrate (Blue) (Units = Å) during the NVT molecular dynamics simulation of the GH13 – α-D-gulosepta-1,4-α-glucose model.

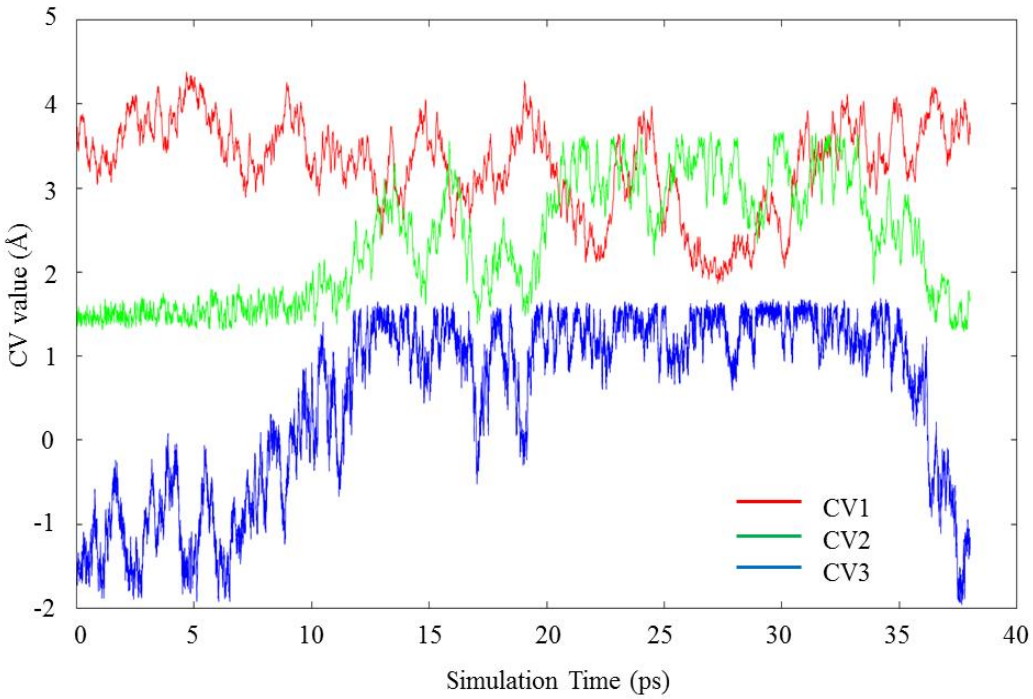


Figure E – 24. Collective variable evolution during the glycosylation step simulation in the GH13 – α-D-gulosepta-1,4-α-glucose complex.

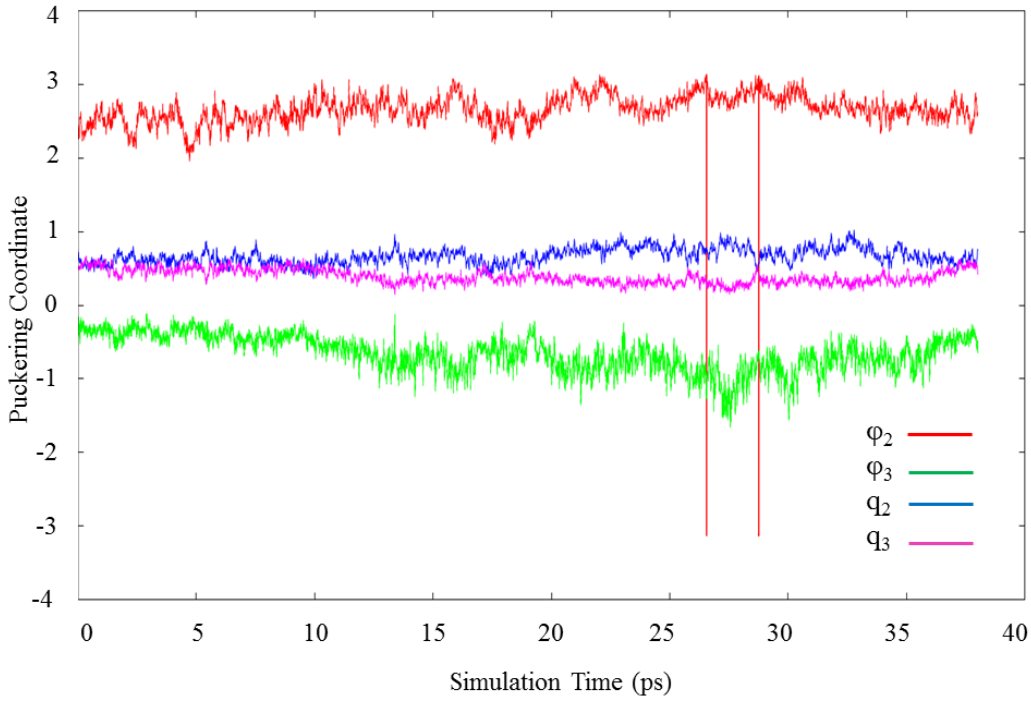


Figure E – 24. Puckering coordinates (φ_2, φ_3, q_2, q_3) evolution during the glycosylation step simulation in the GH13 – α-D-gulosepta-1,4-α-glucose complex.

Appendix F

Computational details: QM/MM, MD and metadynamics parameters selected for the presented simulations.

Chapter III – Amylosucrase. Computational details of QM/MM simulations. NN/MIX/ESP radii in Å. Nosé T Elec. in a.u. and cm^{-1}. Nosé T Nuc. in K and cm^{-1}. MD_0 τ is equilibration time in ps. MTD τ is MTD time in ps. Height of Gaussians (w) and bias-factor (γ) in kcal/mol. Deposition time (τ_D) in MD steps and width of Gaussians (δs) in c.v.u. Wall CVs (lower wall, upper wall in c.v.u., repulsion constant in kcal/mol).

Table III – F1. Conformational study of the sucrose in the *Np*AS.

Puckering	**NN/MIX/ESP**	6.35	6.35	7.94		**# Gaussians**	1971
Nosé T Elec.	0.011	10000	**MD_0 τ**	1.2	**Box**	45	atoms
Nosé T Nuc.	300	3500	**MTD τ**	95	13.1	14.7	15.4
w	0.75	**τ_D**	400	**γ**	10.0	**δs**	0.1

Table III – F2. Simulation of the glycosylation step in the *Np*AS.

Glycosylation	**NN/MIX/ESP**	5.29	5.29	10.58		**# Gaussians**	2284
Nosé T Elec.	0.012	10000	**MD_0 τ**	1.2	**Box**	60	atoms
Nosé T Nuc.	300	3500	**MTD τ**	83	17.5	17.2	16.2
w	1.00	**τ_D**	300	**δs(CV1)**	0.2	**δs(CV2/3)**	0.1
Wall CV1	-1.30	1.50	250.0	**Wall CV2**	Free	4.00	250.0
Wall CV3	Free	3.50	250.0	**Restraint**	H_2O	$H_{A/B}$	>3.0

Table III – F3. Simulation of the water approach process in the *Np*AS.

Water Approach	**NN/MIX/ESP**	5.29	5.29	10.58		**# Gaussians**	811
Nosé T Elec.	0.013	10000	**MD_0 τ**	1.2	**Box**	63	atoms
Nosé T Nuc.	300	3500	**MTD τ**	29	18.0	16.9	16.9
w	1.00	**τ_D**	300	**δs(CV1)**	0.1	**δs(CV2)**	0.1
Wall CV1	2.00	3.70	250.0	**Wall CV2**	2.60	6.00	250.0
				Restraint	C_1	O-fructose	>3.3

Table III – F4. Simulation of the deglycosylation step in the *Np*AS.

Deglycosylation	**NN/MIX/ESP**	5.29	5.29	10.58		**# Gaussians**	4947
Nosé T Elec.	0.013	10000	**MD_0 τ**	1.2	**Box**	63	atoms
Nosé T Nuc.	300	3500	**MTD τ**	148	18.0	16.9	16.6
w	1.00	**τ_D**	250	**δs(CV1)**	0.1	**δs(CV2/3)**	0.1
Wall CV1	2.00	4.00	250.0	**Wall CV2**	Free	3.30	250.0
Wall CV3	-1.30	1.10	250.0	**Restraint**	H_2O	$H_{A/B}$	>3.0

Table III – F5. Simulation of the transglycosylation step in the *Np*AS.

Transglyco.	**NN/MIX/ESP**	5.29	5.29	10.58		**# Gaussians**	3147
Nosé T Elec.	0.0145	10000	**MD_0 τ**	1.2	**Box**	61	atoms
Nosé T Nuc.	300	3500	**MTD τ**	94	17.5	17.2	16.2
w	1.00	**τ_D**	250	**δs(CV1)**	0.1	**δs(CV2/3)**	0.1
Wall CV1	2.00	4.00	250.0	**Wall CV2**	Free	3.70	250.0
Wall CV3	-1.70	1.30	250.0				

Chapter IV – GH125 α-mannosidase. Computational details of QM/MM simulations. NN/MIX/ESP radii in Å. Nosé T Elec. in a.u. and cm^{-1}. Nosé T Nuc. in K and cm^{-1}. MD_0 τ is equilibration time in ps. MTD τ is MTD time in ps. Height of Gaussians (w) and bias-factor (γ) in kcal/mol. Deposition time (τ_D) in MD steps and width of Gaussians (δs) in c.v.u. Wall CVs (lower wall, upper wall in c.v.u., repulsion constant in kcal/mol).

Table IV – F1. Conformational study of the isolated 1-thio-α-mannopyranose.

Puckering	**1-thio-α-mannose**					**# Gaussians**	2144
Nosé T Elec.	0.00570	10000	**MD_0 τ**	3.6	**Box**	24	atoms
Nosé T Nuc.	300	3500	**MTD τ**	103	12.0	12.3	13.2
w	0.5	**τ_D**	400	**δs**	0.1		

Table IV – F2. Conformational study of the 1S-thio-α-mannosyl group in the -1 subsite of the GH125 α-mannosidase.

Puckering S	**NN/MIX/ESP**	5.30	5.30	13.2		**# Gaussians**	2876
Nosé T Elec.	0.01233	10000	**MD_0 τ**	1.3	**Box**	45	atoms
Nosé T Nuc.	300	3500	**MTD τ**	138	17.6	13.2	15.0
w	1.00	**τ_D**	400	**γ**	10.0	**δs**	0.1

Table IV – F3. Conformational study of the natural substrate of the GH125 α-mannosidase.

Puckering O	**NN/MIX/ESP**	6.35	6.35	13.20		**# Gaussians**	1670
Nosé T Elec.	0.01258	10000	**MD_0 τ**	3.0	**Box**	45	atoms
Nosé T Nuc.	300	3500	**MTD τ**	80	17.5	12.9	14.1
w	0.75	**τ_D**	400	**γ**	15.0	**δs**	0.1

Table IV – F4. Simulation of the hydrolysis reaction in the GH125 α-mannosidase.

Hydrolysis	**NN/MIX/ESP**	5.29	5.29	10.58		**# Gaussians**	6430
Nosé T Elec.	0.01750	10000	**MD_0 τ**	1.2	**Box**	69	atoms
Nosé T Nuc.	300	3500	**MTD τ**	94	18.2	18.2	15.9
w	1.50	**τ_D**	250	**δs(CV1/2)**	0.1	**δs(CV3/4)**	0.1
Wall CV1	-0.65	2.00	250.0	**Wall CV2**	Free	3.50	250.0
Wall CV3	Free	3.50	250.0	**Wall CV4**	-2.0	Free	250.0

Chapter VI - Septanosides. Computational details of gas phase simulations. Nosé T Elec. in a.u. and cm^{-1}. Nosé T Nuc. in K and cm^{-1}. MD_0 τ is equilibration time in ps. MTD τ is MTD time in ps. Height of Gaussians (w) and bias-factor (γ) in kcal/mol. Deposition time (τ_D) in MD steps and width of Gaussians (δs) in c.v.u.

Table VI – F1. Conformational study of the isolated cycloheptane molecule.

Cycloheptane						**# Gaussians**	5208
Nosé T Elec.	0.00282	10000	**MD_0 τ**	4.70	**Box**	21	atoms
Nosé T Nuc.	300	3500	**MTD τ**	315	10.5	12.3	11.9
w	0.3	**τ_D**	500	**γ**	6.0	**δs**	0.1

Table VI – F2. Conformational study of the isolated α-D-idosepta molecule.

α-D-idosepta						**# Gaussians**	**6243**
Nosé T Elec.	0.00481	10000	**MD_0 τ**	4.70	**Box**	28	atoms
Nosé T Nuc.	300	3500	**MTD τ**	378	14.0	14.0	14.0
w	0.3	**τ_D**	500	**γ**	12.0	**δs**	0.1

Table VI – F3. Conformational study of the isolated β-D-gulosepta molecule.

β-D-gulosepta						**# Gaussians**	7089
Nosé T Elec.	0.0068	10000	**MD_0 τ**	4.70	**Box**	28	atoms
Nosé T Nuc.	300	3500	**MTD τ**	429	14.0	14.0	14.0
w	0.3	**τ_D**	500	**γ**	12.0	**δs**	0.1

Table VI – F4. Conformational study of the isolated β-L-idosepta molecule.

β-L-idosepta						**# Gaussians**	7229
Nosé T Elec.	0.005	10000	**MD_0 τ**	4.70	**Box**	24	atoms
Nosé T Nuc.	300	3500	**MTD τ**	437	13.0	13.0	13.0
w	0.3	**τ_D**	500	**γ**	10.0	**δs**	0.1

Table VI – F5. Conformational study of the isolated ($[C_6H_{11}O]^+$) cation.

($[C_6H_{11}O]^+$)						**# Gaussians**	2941
Nosé T Elec.	0.00275	10000	**MD_0 τ**	4.70	**Box**	18	atoms
Nosé T Nuc.	300	3500	**MTD τ**	178	13.3	13.3	13.3
w	0.3	**τ_D**	500	**γ**	10.0	**δs**	0.1

Chapter VI – Septanosides. Computational details of QM/MM simulations. NN/MIX/ESP radii in Å. Nosé T Elec. in a.u. and cm^{-1}. Nosé T Nuc. in K and cm^{-1}. MD_0 τ is equilibration time in ps. MTD τ is MTD time in ps. Height of Gaussians (w) and bias-factor (γ) in kcal/mol. Deposition time (τ_D) in MD steps and width of Gaussians (δs) in c.v.u. Wall CVs (Lower wall, Upper wall in c.v.u., Repulsion constant in kcal/mol).

Table VI – F6. Conformational study of the pNP – β-L-idosepta in the GH31 α-glucosidase.

Puckering	**NN/MIX/ESP**	5.29	5.29	10.58	**GH31**	**# Gaussians**	1443
Nosé T. Elec.	0.0075	10000	**MD_0 τ**	1.21	**Box**	36	atoms
Nosé T. Nuc.	300	3500	**MTD τ**	58	13.5	16.9	16.2
w	0.75	**τ_D**	333	**γ**	15.0	**δs**	0.1

Table VI – F7. Simulation of the glycosylation reaction of the pNP – β-L-idosepta in the GH31 α-glucosidase.

Glycosylation	**NN/MIX/ESP**	5.29	5.29	10.58	**GH31**	**# Gaussians**	1042
Nosé T. Elec.	0.0075	10000	**MD_0 τ**	2.10	**Box**	50	atoms
Nosé T. Nuc.	300	3500	**MTD τ**	38	14.8	17.2	16.2
w	1.00	**τ_D**	300	**δs(CV_1)**	0.14	**δs(CV_2)**	0.25
Wall CV_1	-2.40	2.40	250.0	**Wall CV_2**	-1.15	2.00	250.0

Table VI – F8. Simulation of the glycosylation reaction of the pNP – β-L-idosepta in the GH13 α-glucosidase.

Glycosylation	**NN/MIX/ESP**	7.94	7.94	11.64	**GH13**	**# Gaussians**	1056
Nosé T. Elec.	0.0125	10000	**MD_0 τ**	1.21	**Box**	64	atoms
Nosé T. Nuc.	300	3500	**MTD τ**	38	18.5	16.9	18.0
w	1.00	**τ_D**	300	**δs($CV_{1/2}$)**	0.10	**δs(CV_3)**	0.20
Wall CV_1	Free	4.00	150.0	**Wall CV_2**	Free	3.50	150.0
Wall CV_3	-1.90	1.50	150.0				

Appendix G

Publications.

S. Alonso-Gil, A. Males, P. Fernandes, S. J. Williams, G. J. Davies, C. Rovira. *J. Am. Chem. Soc.* **2017**, 139 (3), 1085 – 1088.

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Computational Design of Experiment Unveils the Conformational Reaction Coordinate of GH125 α-Mannosidases

Santiago Alonso-Gil,[†] Alexandra Males,[‡] Pearl Z. Fernandes,[§] Spencer J. Williams,[§] Gideon J. Davies,*[,‡] and Carme Rovira*[,†,∥]

[†]Departament de Química Inorgànica i Orgànica (Secció de Química Orgànica) & Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, 08028 Barcelona, Spain

[‡]York Structural Biology Laboratory, Department of Chemistry, The University of York, YO10 5DD York, United Kingdom

[§]School of Chemistry and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Melbourne, Victoria 3010, Australia

[∥]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

Supporting Information

ABSTRACT: Conformational analysis of enzyme-catalyzed mannoside hydrolysis has revealed two predominant conformational itineraries through $B_{2,5}$ or $^{3}H_{4}$ transition-state (TS) conformations. A prominent unassigned catalytic itinerary is that of *exo*-1,6-α-mannosidases belonging to CAZy family 125. A published complex of *Clostridium perfringens* GH125 enzyme with a non-hydrolyzable 1,6-α-thiomannoside substrate mimic bound across the active site revealed an undistorted $^{4}C_{1}$ conformation and provided no insight into the catalytic pathway of this enzyme. We show through a purely computational approach (QM/MM metadynamics) that sulfur-for-oxygen substitution in the glycosidic linkage fundamentally alters the energetically accessible conformational space of a thiomannoside when bound within the GH125 active site. Modeling of the conformational free energy landscape (FEL) of a thioglycoside strongly favors a mechanistically uninformative $^{4}C_{1}$ conformation within the GH125 enzyme active site, but the FEL of corresponding *O*-glycoside substrate reveals a preference for a Michaelis complex in an $^{O}S_{2}$ conformation (consistent with catalysis through a $B_{2,5}$ TS). This prediction was tested experimentally by determination of the 3D X-ray structure of the pseudo-Michaelis complex of an inactive (D220N) variant of *C. perfringens* GH125 enzyme in complex with 1,6-α-mannobiose. This complex revealed unambiguous distortion of the −1 subsite mannoside to an $^{O}S_{2}$ conformation, matching that predicted by theory and supporting an $^{O}S_{2} \rightarrow B_{2,5} \rightarrow {}^{1}S_{5}$ conformational itinerary for GH125 α-mannosidases. This work highlights the power of the QM/MM approach and identified shortcomings in the use of nonhydrolyzable substrate analogues for conformational analysis of enzyme-bound species.

The conformational itineraries employed by glycoside hydrolases to perform nucleophilic substitution reactions at the anomeric center of glycosides have been the topic of sustained interest since the mid-1990s.[1] Physical organic studies have provided compelling evidence that glycosidase-catalyzed glycoside cleavage occurs through oxocarbenium ion-like transition states (TSs) with significant partial double-bond character between the anomeric carbon and the ring oxygen.[2] Sinnott postulated that glycosidases must react through TSs in one of 4 major conformations: $^{4}H_{3}$ and $^{3}H_{4}$ half chairs (or their related envelopes) or $B_{2,5}$ and $^{2,5}B$ boats. The topological relationships of such conformations are conveniently visualized in a Mercator representation (Figure 1).

By the principle of least nuclear motion,[3] the conformations of the ground states of the enzymatic Michaelis complex, products, and (if relevant) associated intermediates must flank the TSs. Studies over the last 20 years have identified that all four major TS conformations are co-opted by various enzymes working across the breadth of stereochemically diverse carbohydrate substrates.[1] As TS mimicry provides a practical blueprint for the development of tight binding inhibitors, analysis of these reaction coordinates is invaluable in the design and application of TS mimics as mechanistic probes and therapeutic agents.[4]

The canvas upon which nature's treasure chest of glycosidases is depicted is the carbohydrate-active enzymes (CAZy) classification.[5] Enzymes are classified into families by amino-acid sequence (and hence 3D structural) similarity. Of particular interest are the diverse α- and β-mannanases and mannosidases that catalyze the sterically challenged reaction at the crowded anomeric carbon of mannose for which mechanistic insights can inform and enlighten key challenges involved in the chemical synthesis of mannosides.[6] The α- and β-mannanases are involved in glycan processing within important industrial and biological processes. In the latter, assorted α-mannosidases are involved in N-glycan maturation and processing,[7] fungal cell-wall biosynthesis[8] and catabolism,[9] and other cellular reactions of high interest for therapeutic intervention.

By the CAZy classification, α- and β-mannosidases (both *exo*- and *endo*-acting) populate a large number of GH families: α: 38, 47, 76, 92, 99, and 125; β: 2, 5, 26, 113, 130, and 134. Systematic analysis of the conformational itineraries of these enzyme

Received: November 3, 2016
Published: December 27, 2016

DOI: 10.1021/jacs.6b11247
J. Am. Chem. Soc. 2017, 139, 1085−1088

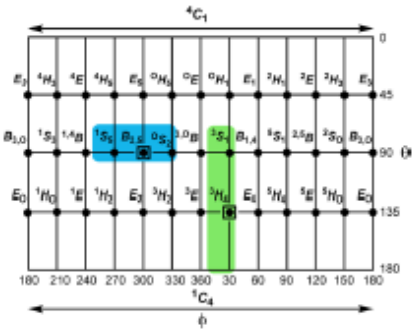


Figure 1. Mercator plot of major canonical conformations of a pyranose ring. The TS conformations (boxed) and associated ground-state conformations of mannosidase conformational itineraries through TSs with $B_{2,5}$ (blue) and $^{3}H_{4}$ (green) conformations.

families, primarily through crystallography of stable species flanking or mimicking the TS(s), has revealed two predominant strategies employed by these catalysts to overcome the challenges of mannoside chemistry (GH99 α-mannosidases are believed to react through an epoxide intermediate[10] and are not discussed further). One group of α- and β-mannosidases belonging to GH families 2,[11] 5,[12] 26,[13] 38,[14] 76,[15] 92,[16] 113,[17] and 130,[18] perform catalysis through a pathway around the $^{O}S_{2}$—$B_{2,5}$—$^{1}S_{5}$ region of the conformational space (Figure S2). The other group of GHs include the family GH47 α-mannosidases[19,20] and the GH134 β-mannanases,[21] which react in a "ring-flipped" (southern hemisphere) $^{3}S_{1}$—$^{3}H_{4}$—$^{1}C_{4}$ conformational arena (Figure S2).

In a seminal work, Gregg et al. reported the creation of GH family 125 based on the discovery of 1,6-α-mannosidase activity for enzymes from *Clostridium perfringens* (*Cp*GH125) and *Streptococcus pneumoniae*.[22] This family was shown to operate through an inverting mechanism, and insight into the active site residues was provided through X-ray structures of these enzymes in complex with the nonhydrolyzable substrate mimic 1,6-α-thiomannobiose (Protein Databank (PDB) ID: 3QT9) and deoxymannojirimycin (PDB ID: 3QRY). Despite the 1,6-α-thiomannobiose substrate mimic spanning the active site, the mechanistically informative −1 subsite mannose residue was observed in an undistorted, ground-state $^{4}C_{1}$ conformation, providing no insight into the conformational itinerary of this family of α-mannosidases. Intrigued by this surprising but uninformative result, we were motivated to investigate further. Although the distortion-free binding of the thiomannoside is surprising, it is not unprecedented. A similar situation was noted in the case of 1,2-α-thiomannobiose bound to a GH92 1,2-α-mannosidase; however, in that case a complex with the TS mimic mannoimidazole provided evidence in support of an $^{O}S_{2}$ → $B_{2,5}$ → $^{1}S_{5}$ conformational itinerary.[16] In the GH125 case the same approach cannot be applied as the general acid residue is not appropriately situated to allow lateral protonation of the basic mannoimidazole nitrogen, whereas in family GH92 enzymes the orientation of the general acid residue is "*anti*"[23] to the C1−O5 bond, which enables lateral protonation and binding of this inhibitor. The inability to assign a conformational itinerary to GH family 125 prevents rational application and design of conformationally locked or biased inhibitors selective for this family of biomedically important enzymes. To understand the conformational preferences of thioglycosides within the active site of *Cp*GH125, we first adopted a computational approach (*ab initio* QM/MM metadynamics)[24] to map the conformational free energy landscape (FEL) of the −1 mannoside ring as a function of the Cremer−Pople ring puckering coordinates,[25] an approach that has been applied to other GH families.[26] We first calculated the conformational FEL for isolated 1-thio-α-mannopyranose (see computational details in the Supporting Information). As previously found for α-mannopyranose,[19a] the sugar has a preference for $^{4}C_{1}$ conformation, but with other regions of the conformational energy surface energetically accessible most notably the region around the $^{O}S_{2}$ conformation (Figure S1).

When the QM/MM metadynamics approach was applied to the 1-thio-α-mannosyl residue of the complex of *Cp*GH125 with 1,6-α-thiomannobiose the FEL is transformed (Figure 2a) from

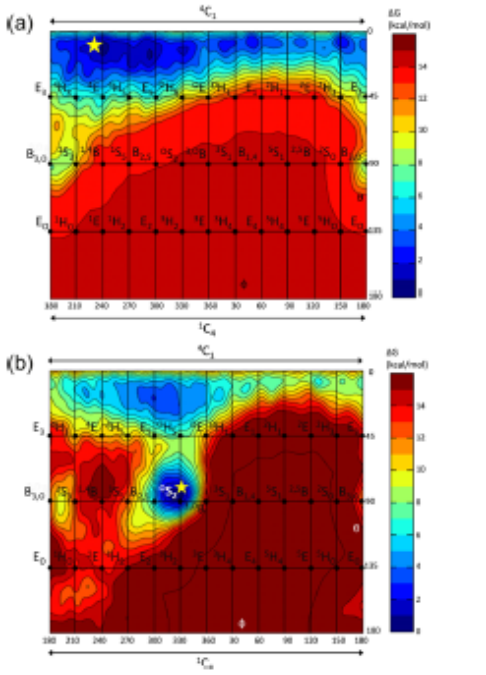


Figure 2. (a) Conformational FEL of the 1-thio-α-mannosyl residue at the −1 subsite of *Cp*GH125 in complex with 1,6-α-thiomannobiose. The star symbol plots the conformation observed experimentally.[22] (b) Conformational FEL of the α-mannosyl residue at the −1 subsite of *Cp*GH125 in complex with 1,6-α-mannobiose. The star symbol plots the new conformation subsequently observed experimentally (see below). Contour lines are every 1 kcal mol^{-1}.

that calculated for 1-thio-α-mannopyranose. The use of an S-linked substrate analogue results in a strong bias to a $^{4}C_{1}$ conformation, matching that observed in the original report of Gregg et al.,[22] with other more mechanistically relevant conformations not energetically accessible. In this case, while the thiomannoside substrate mimic is informative on the gross details of the catalytic apparatus and the ligand interactions, it is silent in terms of conformational insight.

We next sought to establish whether a solely computational approach could make testable predictions for the catalytic itinerary consistent with that previously observed for α- and β-

DOI: 10.1021/jacs.6b11247
J. Am. Chem. Soc. 2017, 139, 1085−1088

mannosidases. Starting with the experimentally determined *Cp*GH125 1,6-α-thiomannobiose complex, the glycosidic sulfur was substituted for oxygen *in silico* to generate a catalytically viable Michaelis complex, which was subjected to minimization to generate a lower energy form, followed by MD equilibration. The full conformational landscape of the −1 sugar ring of this competent substrate containing an *O*-glycosidic linkage was then calculated using the same procedure as in the case of the 1,6-α-thiomannobiose complex. Figure 2b shows that within this Michaelis complex (on-enzyme), an *O*-glycoside strongly favors an $^{O}S_{2}$ conformation, consistent with the α-mannosidase performing catalysis through an $^{O}S_{2} \rightarrow B_{2,5} \rightarrow {}^{1}S_{5}$ conformational itinerary. Subsequent QM/MM simulations of the reaction mechanism (Figures 3 and S2) starting from the $^{O}S_{2}$ conformation led to a $B_{2,5}$ TS, in a dissociative reaction pathway generating a β-mannose product bound to *Cp*GH125 with a $^{1}S_{5}/B_{2,5}$ conformation. This computational data, derived from the coordinate of the *Cp*GH125 1,6-α-thiomannobiose complex, matches that proposed for GH families 2, 5, 26, 38, 76, 92, 113, and 130.[1,4]

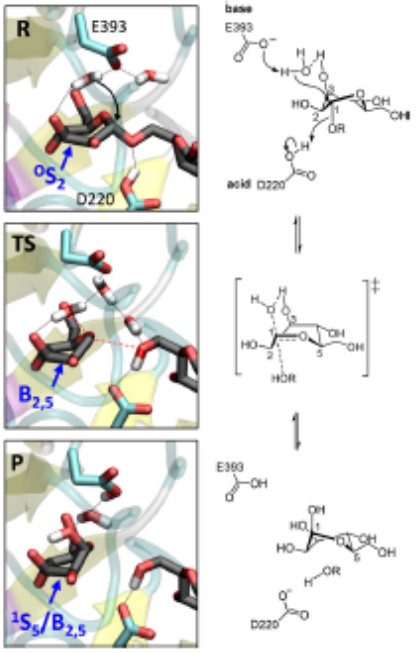


Figure 3. Reaction coordinates for *Cp*GH125 inverting 1,6-α-mannosidase obtained by QM/MM metadynamics with four collective variables. Most hydrogen atoms have been omitted for clarity.

In order to validate, experimentally, the *in silico* prediction, an inactive variant in which the general acid (D220) of *Cp*GH125 was mutated to a nonacidic asparagine residue was engineered. This catalytically inactive variant was crystallized and soaked with the native *O*-glycosides and 1,6-α-mannobiose and -mannotriose to obtain pseudo-Michaelis complexes. Comparison of the structures of the ligand-free *Cp*GH125 wildtype and ligand-bound D220N enzymes revealed no changes in the position of the amino acid side-chain or other residues, providing confidence that the observed ligand conformation was not a result of nonisomorphism. The *Cp*GH125 D220N complexes, solved at resolutions of 2.10 and 1.55 Å (Supporting Information Table 1), unambiguously reveal the −1 subsite mannoside distorted to a $^{O}S_{2}$ conformation (Figure 4a; for 1,6-α-mannotriose complex, see Figure S3), matching that predicted *a priori* by computation.

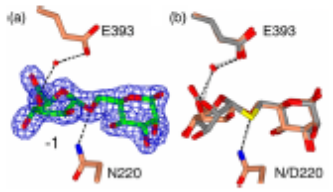


Figure 4. (a) Observed electron density ($2F_{obs} - F_{calc}$, σ_A, and maximum likelihood weighted) for D220N 1,6-α-mannobiose complex of *Cp*GH125, contoured at 0.31 electrons/Å^3. (b) Comparison of *Cp*GH125 complexes with 1,6-α-mannobiose (this work, brick red) with 1,6-α-thiomannobiose complex (gray, PDB 3QT9, ref 22). D220 is the general acid, E393 is the general base, and the proposed nucleophile water is shown.

The *Cp*GH125 complexes highlight the molecular basis for catalysis, with a water poised for in-line nucleophilic attack at the anomeric carbon and with E393 positioned to act as the Brønsted base in an inverting mechanism, essentially as proposed previously.[22] The nucleophilic water molecule is engaged in a hydrogen-bonding interaction with O3, rather than with O2, as was instead observed in the 1,6-α-thiomannobiose complex. This interaction with O3 is reminiscent of that seen for the nucleophilic residue for a GH family 76 retaining 1,6-α-mannanase from *Bacillus circulans*[15] and is thus a feature of the non-metal-dependent family 76 and 125 α-mannosidases. Overlay of the *Cp*GH125 D220N 1,6-α-mannobiose complex with the previously determined 1,6-α-thiomannobiose complex (Figure 4b) highlights the structural basis for the conformational differences: The +1 (leaving group) subsite mannoses are essentially identical in terms of conformation and interactions, but the −1 subsite mannoside moieties adopt different conformations and match those predicted by computation. One major contributor to these different conformations is the longer C−S bond (1.89 vs 1.48 Å for C−O); presumably as a result of this key structural difference, the −1 thiomannoside in a $^{4}C_{1}$ conformation with an axial O2 group makes similar interactions to the pseudoaxial O3 of the mannoside in an $^{O}S_{2}$ conformation (Figure S4). Both theory-based calculations and subsequent experimental observation support a conformational itinerary for the inverting GH125 α-mannosidases that proceeds through a (near) $B_{2,5}$ TS conformation. This TS is accessed following binding of the substrate in the ES complex in an $^{O}S_{2}$ conformation, Figure 3.

GH family 125 joins the growing list of mannose-active enzymes that follows a latitudinal pathway around a $B_{2,5}$ TS in which a key "feature" is the near-eclipsed 40° torsional angle between O3 and O2 that positions a *manno*-configured O2 pseudoequatorial and stabilized through H-bonding on-enzyme. There is a remarkable connection to Crich β-mannosylation methodology wherein judicious choice of a 4,6-*O*-benzylidene protecting group favors a similar pathway.[6]

Thiooligosaccharide substrate mimics have been widely used in X-ray crystallographic studies where they have provided mechanistically relevant insight into conformations possible on

DOI: 10.1021/jacs.6b11247
J. Am. Chem. Soc. 2017, 139, 1085−1088

enzyme, most notably in the case of distorted thiocellopentaoside bound to *Fusarium oxysporum* cellulase of GH family 7.[27] Could other thioglycoside complexes be misleading? In the case of another inverting α-mannosidase, from GH family 47, a 1,2-α-thiomannobioside was ring-flipped and distorted to the $^{3}S_{1}$ conformation suggesting that in this case it is mechanistically relevant.[19] Other GH47 complexes, notably with mannoimidazole in a $^{3}H_{4}$ conformation, kifunensine in a $^{1}C_{4}$ conformation, as well as subsequent QM/MM analysis of FEL of α-mannose "on-enzyme", collectively support the $^{3}S_{1} \rightarrow {}^{3}H_{4} \rightarrow {}^{1}C_{4}$ pathway for that enzyme.[19] In contrast, as discussed earlier, the 1,2-α-thiomannobioside complex reported for family GH92, as described here for GH125, was also observed undistorted and again silent to conformational pathways; in that case distortion of enzyme-bound mannoimidazole to a boat conformation allowed assignment of a $^{O}S_{2} \rightarrow B_{2,5} \rightarrow {}^{1}S_{5}$ pathway for that enzyme.[16]

This work highlights the power of computational methods to use preliminary enzyme−ligand complexes to explore conformational space and generate testable predictions that can provide mechanistic insight using X-ray structural methods. Here these approaches predicted ES distortion for *Cp*GH125 and informed an experimental approach that enabled direct observation of a distorted pseudo-Michaelis complex. This combined *in silico*-experimental approach could be applied to identify catalytic itineraries for other GH families that are presently unknown or for which unusual conformations have been proposed, guiding inhibitor design and leading to the development of mechanistic probes, cellular probes, and ultimately therapeutic agents. In this latter context, the ultimate goal is to obtain conformationally selective and thus specific inhibition of just one enzyme family, as has been achieved through inhibition of ($^{3}S_{1} \rightarrow {}^{3}H_{4} \rightarrow {}^{1}C_{4}$ pathway) GH47 α-mannosidases by kifunensine, a $^{1}C_{4}$ chair mimic.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b11247. Coordinates have also been deposited in the RCSB Protein Databank under codes 5M7Y and 5M7I.

Additional methods and experimental data (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*gideon.davies@york.ac.uk

*c.rovira@ub.edu

ORCID

Spencer J. Williams: 0000-0001-6341-4364

Carme Rovira: 0000-0003-1477-5010

Author Contributions

C.R., S.J.W., and G.J.D. designed experiments. S.A.-G. performed computational work, A.M. carried out structural work, and P.F. conducted synthesis. G.J.D., S.J.W., and C.R. wrote the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are supported by the Royal Society (Ken Murray Research professorship to G.J.D.), the Biotechnology and Biological Sciences Research Council (A.M.), the Australian Research Council (FT130100103) (S.J.W.), the Spanish Ministry of Economy and Competitiveness (CTQ2014-55174-P) (C.R. and S.A.-G.) and GENCAT (2014SGR-987) (C.R.). We thank Diamond Light Source for access to beamline I02 and I04 (mx-13587) and BSC-CNS for computer resources and technical support at MareNostrum (RES-QCM-2016-3-0017).

■ REFERENCES

(1) (a) Davies, G. J.; Planas, A.; Rovira, C. *Acc. Chem. Res.* **2012**, *45*, 308. (b) Speciale, G.; Thompson, A. J.; Davies, G. J.; Williams, S. J. *Curr. Opin. Struct. Biol.* **2014**, *28*, 1.

(2) (a) Koshland, D. E. *Biol. Rev.* **1953**, *28*, 416. (b) Sinnott, M. L. Glycosyl group transfer. In *Enzyme Mechanisms*; Page, M. I., Williams, A., Eds.; Royal Society of Chemistry: London, 1987; pp 259−297.

(3) Sinnott, M. L. *Adv. Phys. Org. Chem.* **1988**, *24*, 113.

(4) Davies, G. J.; Williams, S. J. *Biochem. Soc. Trans.* **2016**, *44*, 79.

(5) Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. *Nucleic Acids Res.* **2014**, *42*, D490.

(6) Crich, D. *Acc. Chem. Res.* **2010**, *43*, 1144.

(7) (a) Kornfeld, R.; Kornfeld, S. *Annu. Rev. Biochem.* **1985**, *54*, 631. (b) Helenius, A.; Aebi, M. *Annu. Rev. Biochem.* **2004**, *73*, 1019.

(8) Free, S. J. *Adv. Genet.* **2013**, *81*, 33.

(9) Cuskin, F.; et al. *Nature* **2015**, *517*, 165.

(10) Thompson, A. J.; et al. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 781.

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

(12) Vincent, F.; Gloster, T. M.; Macdonald, J.; Morland, C.; Stick, R. V.; Dias, F. M. V.; Prates, J. A. M.; Fontes, C. M. G. A.; Gilbert, H. J.; Davies, G. J. *ChemBioChem* **2004**, *5*, 1596.

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

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

(15) Thompson, A. J.; et al. *Angew. Chem., Int. Ed.* **2015**, *54*, 5378.

(16) Zhu, Y.; et al. *Nat. Chem. Biol.* **2010**, *6*, 125.

(17) Williams, R. J.; et al. *Angew. Chem., Int. Ed.* **2014**, *53*, 1087.

(18) Cuskin, F.; Baslé, A.; Ladevèze, S.; Day, A. M.; Gilbert, H. J.; Davies, G. J.; Potocki-Véronèse, G.; Lowe, E. C. *J. Biol. Chem.* **2015**, *290*, 25023.

(19) (a) Thompson, A. J.; et al. *Angew. Chem., Int. Ed.* **2012**, *51*, 10997. (b) Karaveg, K.; Siriwardena, A.; Tempel, W.; Liu, Z. J.; Glushka, J.; Wang, B. C.; Moremen, K. W. *J. Biol. Chem.* **2005**, *280*, 16197.

(20) Vallée, F.; Karaveg, K.; Herscovics, A.; Moremen, K. W.; Howell, P. L. *J. Biol. Chem.* **2000**, *275*, 41287.

(21) Jin, Y.; Petricevic, M.; John, A.; Raich, L.; Jenkins, H.; Portela De Souza, L.; Cuskin, F.; Gilbert, H. J.; Rovira, C.; Goddard-Borger, E. D.; Williams, S. J.; Davies, G. J. *ACS Cent. Sci.* **2016**, *21*, 896.

(22) Gregg, K. J.; Zandberg, W. F.; Hehemann, J. H.; Whitworth, G. E.; Deng, L.; Vocadlo, D. J.; Boraston, A. B. *J. Biol. Chem.* **2011**, *286*, 15586.

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

(24) (a) Barducci, A.; Bonomi, M.; Parrinello, M. *WIREs Comput. Mol. Sci.* **2011**, *1*, 826. (b) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562.

(25) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354.

(26) Ardèvol, A.; Rovira, C. *J. Am. Chem. Soc.* **2015**, *137*, 7528.

(27) Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schulein, M.; Davies, G. J. *Biochemistry* **1996**, *35*, 15280.